



Arnaldo Zaha
Henrique Bunselmeyer Ferreira
Luciane M. P. Passaglia

organizadores

Biologia Molecular Básica



5ª edição



B615 Biologia molecular básica [recurso eletrônico] / Organizadores, Arnaldo Zaha, Henrique Bunselmeyer Ferreira, Luciane M. P. Passaglia. – 5. ed. – Dados eletrônicos. – Porto Alegre : Artmed, 2014.

Editado também como livro impresso em 2014.
ISBN 978-85-8271-058-6

1. Biologia molecular. I. Zaha, Arnaldo. II. Ferreira, Henrique Bunselmeyer. III. Passaglia, Luciane M. P.

CDU 577.2

Arnaldo Zaha
Henrique Bunselmeyer Ferreira
Luciane M. P. Passaglia

organizadores

Biologia Molecular Básica

5ª edição

Versão impressa
desta obra: 2014



2014

© Artmed Editora Ltda., 2014

Gerente editorial: *Leticia Bispo de Lima*

Colaboraram nesta edição:

Preparação de originais: *Caroline Vieira, Carine Garcia Prates*

Capa: *Paola Manica*

Imagem da capa:

©iStockphoto.com/Theasis, 2010: *Model of an Enhanceosome Protein Complex Binding to DNA*

Modelo de um complexo proteico ligado ao reforçador (enhancer)

Editoração eletrônica: *Techbooks*

Nota

Assim como a medicina, a biologia molecular é uma ciência em constante evolução. À medida que novas pesquisas e a própria experiência clínica ampliam o nosso conhecimento, são necessárias modificações na terapêutica, onde também se insere o uso de medicamentos. Os autores desta obra consultaram as fontes consideradas confiáveis, num esforço para oferecer informações completas e, geralmente, de acordo com os padrões aceitos à época da publicação. Entretanto, tendo em vista a possibilidade de falha humana ou de alterações nas ciências médicas, os leitores devem confirmar estas informações com outras fontes. Por exemplo, e em particular, os leitores são aconselhados a conferir a bula completa de qualquer medicamento que pretendam administrar, para se certificar de que a informação contida neste livro está correta e de que não houve alteração na dose recomendada nem nas precauções e contraindicações para o seu uso. Essa recomendação é particularmente importante em relação a medicamentos introduzidos recentemente no mercado farmacêutico ou raramente utilizados.

Reservados todos os direitos de publicação à
ARTMED EDITORA LTDA., uma empresa do GRUPO A EDUCAÇÃO S.A.
Av. Jerônimo de Ornelas, 670 – Santana
90040-340 – Porto Alegre – RS
Fone: (51) 3027-7000 Fax: (51) 3027-7070

É proibida a duplicação ou reprodução deste volume, no todo ou em parte, sob quaisquer formas ou por quaisquer meios (eletrônico, mecânico, gravação, fotocópia, distribuição na Web e outros), sem permissão expressa da Editora.

Unidade São Paulo
Av. Embaixador Macedo Soares, 10.735 – Pavilhão 5 – Cond. Espace Center
Vila Anastácio – 05095-035 – São Paulo – SP
Fone: (11) 3665-1100 Fax: (11) 3667-1333

SAC 0800 703-3444 – www.grupoa.com.br

IMPRESSO NO BRASIL
PRINTED IN BRAZIL

Autores

Arnaldo Zaha

Professor titular do Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências e pesquisador do Centro de Biotecnologia, da Universidade Federal do Rio Grande do Sul (UFRGS). Doutorado pelo Programa de Pós-Graduação em Ciências Biológicas (Bioquímica) do Instituto de Química da Universidade de São Paulo (USP).

Henrique Bunselmeyer Ferreira

Professor associado do Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências e pesquisador do Centro de Biotecnologia, da UFRGS. Doutorado pelo Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS. Pós-doutorado na Yeshiva University, Nova Iorque, Estados Unidos.

Luciane M. P. Passaglia

Professora associada do Departamento de Genética, Instituto de Biociências, da UFRGS. Doutorado pelo Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS. Pós-doutorado na University of California, Berkeley, Estados Unidos.

Ana Tereza Ribeiro de Vasconcelos

Coordenadora do Laboratório de Bioinformática do Laboratório Nacional de Computação Científica (LNCC/MCT), Petrópolis, RJ. Professora visitante da Université Claude Bernarde Lyon I, Lyon, França. Doutorado pelo Programa de Pós-Graduação em Ciências Biológicas (Genética) da Universidade Federal do Rio de Janeiro (UFRJ). Pós-doutorado na University of Texas MD Anderson Cancer Center, Houston, Estados Unidos.

Augusto Schrank

Professor associado do Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências, e pesquisador do Centro de Biotecnologia da UFRGS. Doutorado pela University of Manchester Institute of Science and Technology (UMIST), Manchester, Inglaterra.

Charley Christian Staats

Professor do Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências e pesquisador do Centro de Biotecnologia, UFRGS. Doutorado pelo Programa de Pós-Graduação em Biologia Celular e Molecular da UFRGS.

Darcy Fontoura de Almeida

Professor titular emérito da UFRJ. Professor colaborador voluntário do Laboratório de Bioinformática do LNCC/MCT. Professor colaborador voluntário da Casa de Oswaldo Cruz/Departamento de Arquivo e Documentação da FIOCRUZ.

Elgion Loreto

Professor associado do Departamento de Biologia da Universidade Federal de Santa Maria (UFSM), RS. Doutorado pelo Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS.

Irene Silveira Schrank

Professora associada do Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências, e pesquisadora do Centro de Biotecnologia da UFRGS. Doutorado pela UMIST, Manchester, Inglaterra.

Luiza Amaral de Castro

Pós-doutoranda PNPD/CAPES do Programa de Pós-Graduação em Ciências Veterinárias da UFRGS. Doutorado pelo Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS.

Maria Lucia R. Rossetti

Professora adjunta da Universidade Luterana do Brasil. Pesquisadora do Centro de Desenvolvimento Científico e Tecnológico (CDCT), da Fundação Estadual de Produção e Pesquisa em Saúde (FEPPS). Doutorado pelo Programa de Pós-graduação em Ciências Biológicas (Bioquímica) da UFRGS. Pós-doutorado na Azienda Ospedaliera Universitaria Careggi, Firenze, Itália.

Marilene Henning Vainstein

Professora associada do Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências, e pesquisadora do Centro de Biotecnologia da UFRGS. Doutorado pela University of Nottingham, Nottingham, Inglaterra.

Sérgio Ceroni da Silva

Professor adjunto do Departamento de Patologia Clínica Veterinária, Faculdade de Veterinária da UFRGS. Doutorado pela Universidade de Glasgow, Glasgow, Escócia.

Vera Lúcia S. Valente Gaiesky

Professora titular do Departamento de Genética, Instituto de Biociências, da UFRGS. Doutorado pelo Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS.

Apresentação

Esta obra deve ser saudada por sua importância temática, qualidade e abrangência. Trata-se de um livro de biologia molecular escrito em uma linguagem clara e acessível, que será de grande utilidade não apenas para estudantes de graduação e aqueles que se preparam para ingresso na pós-graduação, mas também para profissionais que almejam reciclar-se nos conceitos da biologia molecular. Mesmo profissionais sem uma formação estrita em ciências biológicas irão se beneficiar dos conhecimentos aqui apresentados, considerando o papel central da biologia molecular como instrumento para as atividades em todas as áreas das ciências biológicas e grande parte das ciências médicas.

Ao longo da obra, os conceitos são formulados de modo que o leitor poderá avançar de acordo com o nível de complexidade do conteúdo abordado. São apresentados de maneira bastante simples os conceitos mais importantes da biologia moderna, e os conhecimentos adquiridos permitem entender os mecanismos moleculares envolvendo células simples e organismos mais complexos. Os principais termos técnicos são definidos, com a nomenclatura em inglês apresentada entre parênteses, estimulando o leitor a avançar para leituras mais aprofundadas em artigos de periódicos científicos ou em livros técnicos mais avançados.

O livro está organizado em 17 capítulos, nos quais são apresentados os fundamentos da célula, das macromoléculas que compõem as células e dos processos

e mecanismos envolvidos na organização molecular dos seres vivos. Os dois capítulos finais do livro são bastante interessantes, fornecendo a base para a compreensão das técnicas básicas de biologia molecular e das ferramentas de bioinformática para os modernos estudos de genômica. A preocupação com a didática permeia toda a obra, motivo pelo qual há resumos concluindo os temas abordados, permitindo ao leitor atentar para os conceitos mais importantes a serem apreendidos da leitura.

O fato de termos um livro de *Biologia molecular básica* escrito por autores brasileiros, pesquisadores de destaque em suas áreas de atuação e experientes professores de biologia molecular proporciona um grande diferencial a esta obra, que é o fato de os autores terem a vivência para identificar as maiores dificuldades dos estudantes, os grandes beneficiários da leitura deste livro. Destinado àqueles que têm um primeiro contato com a biologia molecular, ela também será útil aos que desejam adquirir conhecimentos fundamentais em biologia molecular.

Samuel Goldenberg

Pesquisador titular do Instituto Carlos Chagas/Fiocruz-Paraná.

Doutor em Ciências pela Universidade de Paris VII.

Prefácio

A primeira versão do livro *Biologia molecular básica*, publicada em 1996, tinha como objetivo fundamental reunir o conteúdo elementar dessa área do conhecimento em uma obra em língua portuguesa – objetivo esse que tem sido mantido desde então.

Na biologia molecular, os últimos 12 anos foram marcados por novidades tecnológicas espetaculares, que aumentaram muito os conhecimentos sobre como os genes funcionam e como suas atividades estão integradas em uma rede que permite o desenvolvimento e o funcionamento correto de um organismo completo, seja ele uma bactéria, um vegetal ou um animal. Graças ao desenvolvimento e à aplicação de novas tecnologias, milhares de organismos, incluindo o homem, tiveram o genoma completamente sequenciado, e hoje termos como genômica, transcritômica e proteômica fazem parte do dia a dia dos pesquisadores, professores, estudantes e interessados na área de biologia molecular.

Para acompanhar todos esses avanços, esta 5ª edição do livro *Biologia molecular básica* foi elaborada por uma equipe de 14 autores, em parceria com a Artmed Editora. Ela inclui, além de todas as reformulações da 4ª edição, que tornaram o livro mais didático, um capítulo que foi totalmente reescrito (Capítulo 11), agora intitulado *Mecanismos de processamento e maturação de RNA*. Esta nova edição conta com um projeto gráfico atraente e inclui aspectos que auxiliam na aprendizagem, como um

resumo ao final de cada capítulo, um glossário e o Material Complementar *online*. O resumo e o glossário, assim como o destaque de termos-chave ao longo do texto, facilitam ao leitor o reconhecimento do conteúdo essencial de cada tema abordado no livro. O conteúdo *online* possibilita ao leitor o aprofundamento de vários tópicos abordados no livro, inclusive com listas adicionais de leituras recomendadas.

Mesmo com a inclusão dos temas mais atuais, esta nova edição continua a apresentar, em todos os seus capítulos, os fundamentos de cada um dos temas abordados. Não houve a pretensão de incluir conteúdo de complexidade além daquela passível de ser aproveitada por leitores que terão, no livro, o seu primeiro contato com uma obra especificamente dedicada à biologia molecular. Ele é, portanto, indicado para estudantes e profissionais que procuram conteúdos básicos e que estão iniciando o contato com a literatura científica original de biologia molecular, do ensino à pesquisa básica ou aplicada.

Esperamos que esta 5ª edição contribua para o entendimento dos conceitos fundamentais da biologia molecular e estimule o leitor a interessar-se mais por esta área tão dinâmica e fascinante da biologia.

Os organizadores

Sumário

1	A Célula e seus Constituintes Moleculares	1
	Maria Lucia R. Rossetti	
2	Estrutura dos Ácidos Nucleicos	17
	Augusto Schrank	
3	Cromatina	37
	Henrique Bunselmeyer Ferreira	
4	Genes e Genomas Procarióticos	57
	Henrique Bunselmeyer Ferreira	
5	Genes e Genomas Eucarióticos	85
	Henrique Bunselmeyer Ferreira	
6	Replicação do DNA	111
	Irene Silveira Schrank	
7	Mutação e Reparação do DNA	133
	Luciane M. P. Passaglia	
8	Recombinação Genética	163
	Luciane M. P. Passaglia	
9	Elementos Genéticos Móveis	185
	Elgion Loreto Henrique Bunselmeyer Ferreira	

10	Transcrição.....	205
	Augusto Schrank	
11	Mecanismos de Processamento e Maturação de RNA.....	233
	Charley Christian Staats	
12	Código Genético e Síntese de Proteínas	255
	Irene Silveira Schrank Marilene Henning Vainstein	
13	Controle da Expressão Gênica em Procariotos.....	277
	Sérgio Ceroni da Silva Irene Silveira Schrank Luiza Amaral de Castro	
14	Controle da Expressão Gênica em Eucariotos	301
	Arnaldo Zaha	
15	Biologia Molecular do Desenvolvimento	319
	Vera Lúcia S. Valente Gaiasky	
16	Técnicas de Biologia Molecular	331
	Luciane M. P. Passaglia Arnaldo Zaha	
17	Bioinformática na Análise de Genes e Genomas	363
	Ana Tereza Ribeiro de Vasconcelos Darcy Fontoura de Almeida	
	Glossário.....	383
	Índice	395

Maria Lucia R. Rossetti

A Célula e seus Constituintes Moleculares

1. Estruturas celulares 2

- 1.1 Células de procariotos e eucariotos 3
- 1.2 Organelas 4

2. Constituintes moleculares 4

- 2.1 Aminoácidos e proteínas 4
 - 2.1.1 Estrutura tridimensional da proteína 8

2.2 Carboidratos 9

2.3 Lipídeos 10

- 2.3.1 Ácidos graxos 10
- 2.3.2 Fosfoacilgliceróis 12
- 2.3.3 Esteroides 12

2.4 Ácidos nucleicos 13

- 2.4.1 Ácido desoxirribonucleico 14
- 2.4.2 Ácido ribonucleico 14

Todos os organismos vivos são constituídos de pequenas estruturas denominadas **células**. Essas estruturas, que representam a menor unidade de vida, são bastante complexas e diversas, sendo que nelas estão contidas as características morfológicas e fisiológicas dos organismos vivos. As propriedades de um determinado organismo dependem de suas células individuais, cuja continuidade ocorre por meio de seu **material genético**. A forma mais simples de vida ocorre em células isoladas, que se propagam por **divisão celular**. Já os organismos superiores, como o próprio homem, são constituídos de **agregados celulares** que desempenham funções especializadas.

As células de diferentes organismos são muito similares quanto à estrutura e a constituintes moleculares, apesar das diferenças organizacionais fundamentais existentes. Ao analisar os constituintes moleculares, é importante considerar não apenas as propriedades individuais das moléculas, como também as interações existentes entre elas e a sua localização dentro da célula. Essa análise é ainda mais necessária quando se considera um organismo multicelular e os eventos que ocorrem em seu interior, a fim de produzir a diferenciação e o desenvolvimento desse organismo. Assim, este capítulo busca fornecer, uma breve revisão sobre a estrutura celular, seus constituintes moleculares e as interações realizadas por esses compostos.

zadas todas as funções necessárias à manutenção e à preservação da vida, foram feitas por Robert Hooke, em 1665. Todas as células, independentemente da complexidade do organismo, possuem uma mesma estrutura formada pela **membrana plasmática**, que circunda o conteúdo celular e o separa do meio extracelular pelo **citossol** e pelo **núcleo** (ou nucleóide). A **Figura 1.1** representa uma célula eucariótica e suas principais estruturas. O citossol é todo o volume interno celular, composto por uma solução aquosa complexa com várias partículas e moléculas dispersas. O tamanho e a forma da célula variam muito e não têm nenhuma relação com o tamanho do organismo. Algumas células vivem isoladas, como os **organismos unicelulares**, porém as dos **organismos pluricelulares**, em geral, se relacionam umas com as outras.

A membrana celular plasmática, que circunda todas as células, é formada basicamente por uma dupla camada de **lipídeos** da classe dos fosfolipídeos e, em quantidades variáveis, se associada com moléculas proteicas (ver Figura 1.1). O limite entre o meio intracelular e extracelular é definido pela membrana plasmática. Para entrar ou sair de uma célula, uma substância deve transpor a membrana celular; portanto, isso dependerá da **permeabilidade** da membrana. Essa bicamada lipídica é permeável a certos gases, como O₂ e CO₂, e impermeável a muitas substâncias, como açúcar, aminoácidos e íons inorgânicos (K⁺ e Cl⁻). A água pode difundir-se livremente através da célula. Muitas proteínas estão ligadas à membrana plasmática (permeases ou transportadoras), formando canais na bicamada lipídica e facilitando a passagem de certas substâncias. Dessa forma, todas as células, de todos os organismos, possuem características estruturais comuns, como a arquitetura de suas membranas e muitos

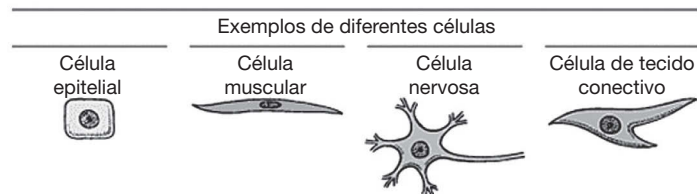
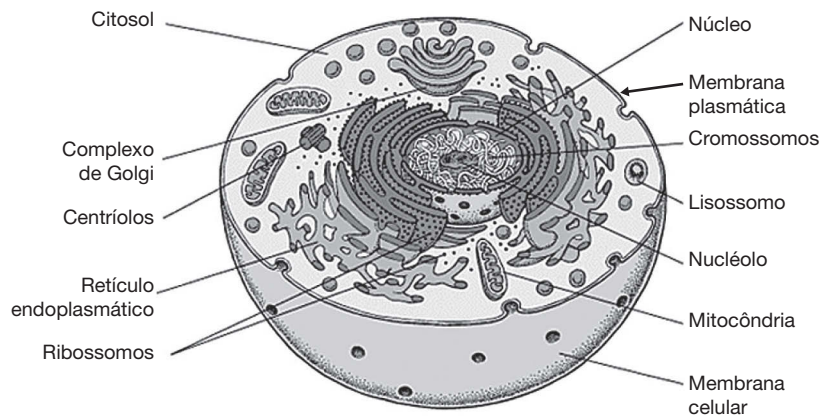
1. Estruturas celulares

As primeiras observações e a própria denominação de células para as unidades estruturais, nas quais são reali-

Figura 1.1

Representação esquemática de uma célula animal (célula de eucarioto) com suas principais estruturas e organelas celulares.

As características principais são a presença de um núcleo bem definido, que contém o genoma, e das organelas celulares, que compartimentalizam determinadas funções. Em destaque, diferentes tipos celulares.



processos metabólicos, até a replicação de DNA, a síntese proteica e a produção de energia química.

1.1 Células de procariotos e eucariotos

Apesar da similaridade existente entre as células que constituem os seres vivos, os organismos mantêm diferenças fundamentais em nível celular, podendo ser classificados em dois grandes grupos: os **procariotos** e os **eucariotos**. Os organismos procariotos são unicelulares e mais simples em sua organização, embora possam ocorrer associados a grupos, formando colônias com alguma diferenciação de funções. Os procariotos incluem as **bactérias** e as **arqueas** (bactérias que sobrevivem em ambientes não usuais, como lagos salinos, piscinas térmicas e pântanos). Os organismos eucariotos são mais complexos e incluem não somente plantas pluricelulares, animais e fungos, mas também protozoários e alguns organismos unicelulares, como leveduras e algas verdes.

A **Tabela 1.1** compara as principais características celulares dos organismos procariotos e eucariotos, possibilitando, então, evidenciar as semelhanças e as diferenças existentes.

A principal diferença entre procariotos e eucariotos é que, nos eucariotos, encontram-se organelas, principalmente o núcleo, que contém o genoma. As organelas são regiões delimitadas por membranas internas, que formam compartimentos, nos quais se realizam funções especializadas. Nos procariotos, a ausência de envoltório nuclear deixa o genoma em contato direto com o restante do citoplasma, em um espaço dentro da célula denominado **nucleoide**, ficando junto de ribossomos, outras partículas e uma grande variedade de moléculas dissolvidas (**Figura 1.2**).

As células de procariotos possuem, normalmente, além da membrana plasmática, uma parede celular, cuja função é proporcionar maior rigidez e proteção mecânica. Essa membrana também está presente em células eucarióticas vegetais. A composição química da parede celular de procariotos é bastante complexa, contendo moléculas de polissacarídeos, lipídeos e proteínas (camada de **pep-**

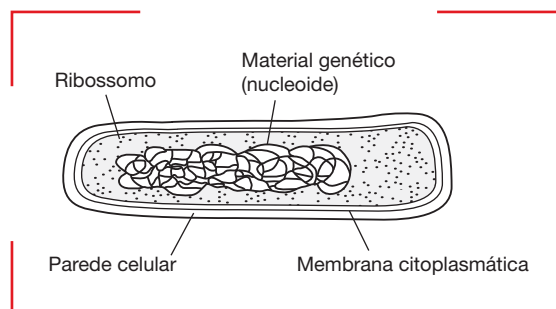


Figura 1.2

Representação esquemática da organização de uma célula de procarioto.

O nucleoide composto pelo genoma não está delimitado por uma membrana. Presença de uma parede celular.

ideoglicano). A parede celular dos vegetais contém celulose e outros polímeros. Células de fungos também estão circundadas por uma parede celular de composição diferente das de vegetais e bactérias. As bactérias gram-negativas possuem, ainda, uma membrana externa, que circunda a parede celular. Essa parede celular é permeável a muitas substâncias químicas com peso molecular superior a 1 kDa. Dentro da membrana plasmática, está o citoplasma, constituído pelo **citossol** – componente aquoso. O citoplasma de células eucarióticas difere do citoplasma de células procarióticas pela presença das organelas e de proteínas filamentosas, que constituem o chamado **citoesqueleto**. Entre essas proteínas estão os filamentos de actina e os microtúbulos, envolvidos na geração de movimentos celulares, na determinação da forma celular e na capacidade de arranjar as organelas.

Outra diferença fundamental observada entre procariotos e eucariotos é em relação ao material genético. A informação genética de organismos procarióticos apresenta-se, geralmente, em uma ou mais moléculas circulares de DNA. As bactérias são conhecidas por possuírem um único **cromossomo**. O arranjo de genes, dentro desse cromossomo, difere muito do arranjo em um cromos-

Tabela 1.1 Principais características celulares de procariotos e de eucariotos

	Procariotos	Eucariotos
Organização	Principalmente unicelular	Principalmente pluricelular
Membrana citoplasmática	Bicamada fosfolipídica; rara presença de esteróis	Bicamada fosfolipídica; presença de esteróis e carboidratos
Núcleo	Ausente	Definido pela membrana nuclear
Citoplasma	Sem citoesqueleto	Citoesqueleto constituído
Motilidade	Flagelos simples	Flagelos complexos; pseudópodes; outros órgãos de locomoção mais complexos
Organelas	Poucas ou nenhuma	Presentes: lisossomos, complexo de Golgi, retículo endoplasmático (RE), mitocôndria e cloroplastos
Parede celular	Contém glicopeptídeos, lipídeos, proteínas	Quando presente, contém quitina ou celulose

somo de células eucarióticas. Nas células eucarióticas, o DNA nuclear é dividido em dois ou mais um desses cromossomos. Cada um desses cromossomos é formado por uma molécula de DNA linear que, exceto durante a divisão celular, está confinada dentro do núcleo. Essas moléculas de DNA estão associadas a proteínas, chamadas de **histonas**, formando os **nucleossomos**, componentes da **cromatina**. O número e o tamanho dos cromossomos individuais variam muito entre os diferentes organismos eucarióticos. Os fungos, por exemplo, possuem de 12 a 18 cromossomos; células humanas contêm dois conjuntos de 23 cromossomos, cada um tem aproximadamente trinta vezes a quantidade de DNA presente em uma célula da bactéria *Escherichia coli*. Uma descrição mais detalhada sobre o tema será feita nos Capítulos 4 e 5.

O DNA não é encontrado apenas no núcleo, mas também na mitocôndria das células de animais, plantas e fungos e no cloroplasto das plantas. Essa é uma das evidências que sugere a evolução dessas organelas a partir de bactérias que sofreram **endocitose** por células ancestrais. O DNA dessas organelas contém genes que codificam proteínas específicas para o funcionamento da própria organela.

1.2 Organelas

Existem organelas comuns a todas as células eucarióticas (ver Figura 1.1), sendo talvez o núcleo a mais importante dessas organelas. O núcleo é envolto por uma membrana nuclear dupla, que possui como característica, além da cromatina, uma região rica em RNA, denominada **nucleólo**. No nucleólo, os RNAs são sintetizados a partir de um molde de DNA e, posteriormente, exportados para o citoplasma através da membrana nuclear. A **mitocôndria**, outra importante organela, possui enzimas especializadas em processos oxidativos que produzem energia para a célula. Além disso, o DNA e os ribossomos também são encontrados no interior da mitocôndria.

O **retículo endoplasmático** é uma estrutura formada por membranas distribuídas por todo o citoplasma e ligadas tanto à membrana celular como à membrana nuclear. O **RE granuloso** possui ribossomos ligados às suas membranas, que constituem a maquinaria molecular para a síntese proteica.

Já o **complexo de Golgi** é uma estrutura formada por membranas e vesículas, envolvido na modificação e na secreção de proteínas das células.

Existem organelas específicas para células vegetais e animais. As células animais, por exemplo, contêm **lisossomos** com a função de digestão, e as células vegetais possuem **cloroplastos**, local onde se realiza a fotossíntese. Outra característica, comum à maioria das células vegetais e a alguns microrganismos, é a presença de **vacúolo**, cuja função é a estocagem de nutrientes e metabólitos. Cada organela possui uma coleção própria de enzimas catalizadoras de reações específicas, desenvolvendo um papel único no crescimento e no metabolismo celular. As organelas celulares estão ligadas a uma rede de **mi-**

crotúbulos constituídos de proteínas filamentosas, denominado **citoesqueleto**. Essa estrutura é responsável por manter a forma da célula e auxiliar nos movimentos celulares.

2. Constituintes moleculares

Os constituintes moleculares são responsáveis pelas interações bioquímicas entre milhares de moléculas que permitem a vida celular. Essas reações químicas acontecem em meio aquoso, por isso, a **água**, com poucas exceções (célula óssea), é o componente encontrado em maior quantidade na célula, sendo indispensável para a atividade metabólica. A água, devido a sua natureza polar, serve como solvente natural para íons, minerais e outras substâncias e, também, como meio de dispersão para a estrutura coloidal do citoplasma. A presença de íons, como Cl^- , Na^+ e K^+ , é importante para manter a pressão osmótica e o equilíbrio acidobásico da célula. Alguns íons inorgânicos, como o magnésio, são necessários na função de **cofatores enzimáticos**. Outros, como o fosfato inorgânico, formam adenosina trifosfato (ATP), principal fonte de energia química dos processos vitais e os íons cálcio desempenham um papel regulador.

Além da água e dos elementos químicos citados, a célula é constituída por pequenas moléculas e macromoléculas. As pequenas moléculas, como aminoácidos, nucleotídeos, lipídeos e açúcares constituem os substratos e os produtos de vias metabólicas, fornecendo energia para a célula e podendo, também, ser as unidades formadoras das macromoléculas. Nessa estrutura de polímero biológico, essas moléculas são chamadas de **monômeros** ou **resíduos**.

As células são constituídas, basicamente, por três tipos de polímeros: **ácidos nucleicos** – formados pelos nucleotídeos (monômero); **proteínas** – constituídas pelos aminoácidos; e **carboidratos** ou polissacarídeos – cujos monômeros são os açúcares ou monossacarídeos. As células possuem uma grande quantidade de lipídeos que, diferentes dos demais, não são polímeros, sendo, de preferência, moléculas pequenas. O lipídeo mais simples e abundante é o **ácido graxo**, que participa da composição de outras moléculas de lipídeos mais complexas, como os triacilgliceróis. As estruturas maiores, como a bicamada das membranas biológicas e as fibras do citoesqueleto, são formadas pela associação dessas macromoléculas.

2.1 Aminoácidos e proteínas

As proteínas resultam da expressão da informação contida no gene. Por isso, é o gene que determinará a sequência de aminoácidos de uma proteína específica. Assim, toda proteína possui uma ordem definida de resíduos de aminoácidos, que, por sua vez, estabelece sua estrutura tridimensional ativa, denominada **conformação nativa** (Figura 1.3). A estrutura tridimensional da molécula,



Vírus: Parasitas celulares

Tanto as células eucarióticas como as procarióticas podem ser infectadas por partículas virais muito pequenas, que estão distribuídas na natureza. Essas partículas não celulares variam quanto à forma e à complexidade estrutural e, geralmente, são formadas por uma partícula central (*core*) de nucleoproteínas, envolta por um capsídeo composto de uma ou mais proteínas (**Figura Q.1**). Alguns vírus também possuem uma membrana lipoproteica envolvendo o capsídeo (envelope viral). O genoma viral pode ser de DNA ou RNA, mas não de ambos. A replicação viral ocorre somente em células vivas, devido à incapacidade genética de replicação de forma autônoma. A maioria dos vírus de DNA de plantas e animais necessita de enzimas nucleares das células hospedeiras para realizar a transcrição de mRNA e, assim, a síntese de proteínas virais. Em geral, os vírus de DNA replicam e agrupam-se no núcleo e os de RNA agrupam-se no citoplasma. Os vírus mais simples contêm RNA ou DNA suficiente para codificar quatro proteínas; os mais complexos podem codificar de 100 a 200 proteínas.

A maioria dos vírus infecta um número limitado de células, determinado pelo tipo de proteína presente na superfície viral, que se ligará especificamente a proteínas receptoras da célula hospedeira. Os vírus que infectam bactérias são conhecidos como bacteriófagos ou fagos. O genoma de alguns vírus pode ser integrado ao genoma da célula hospedeira, como

no caso do retrovírus, cujo genoma de RNA serve de molde para uma molécula de DNA. Alguns retrovírus possuem genes causadores de câncer (oncogenes), capazes de transformar a célula infectada em uma célula tumoral. As alterações celulares causadas pela presença de um vírus variam bastante, desde simples parasitas sem nenhum efeito, até altamente patogênicos, como, por exemplo, os que causam Aids, hepatites e câncer em seres humanos. Os vírus também possuem importância no controle biológico e como ferramenta de pesquisa.

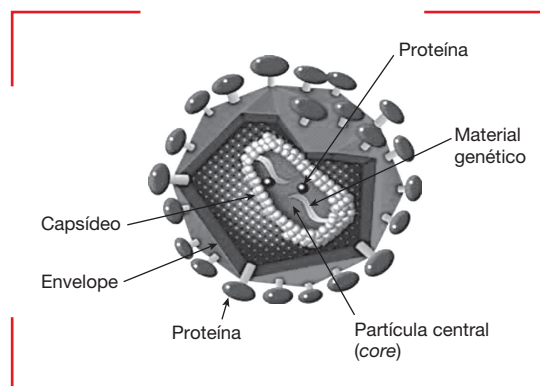


Figura Q.1

Desenho esquemático representativo da organização de uma partícula viral. O material genético do vírus está envolto por uma camada proteica (capsídeo) e por um envelope formado por uma dupla camada de fosfolípidos e proteínas, que reconhecem receptores celulares.

estabelecida conforme a própria sequência de aminoácidos, pode ser observada em experimentos de desnaturação de uma determinada proteína. As alterações em sua forma natural (nativa), por mudanças nas condições do meio (alterações de pH, temperatura, adição de solventes) onde se encontra a proteína, com conseqüente perda de sua função biológica, podem, algumas vezes, serem recuperadas. O restabelecimento da conformação nativa – a **renaturação proteica** – pode ocorrer quando as condições do meio em que a molécula se encontra são restabelecidas, possibilitando que os aminoácidos voltem a interagir. O enrolamento de uma proteína globular é um processo energeticamente favorável, sob condições fisiológicas, que permite as interações entre os grupos químicos.

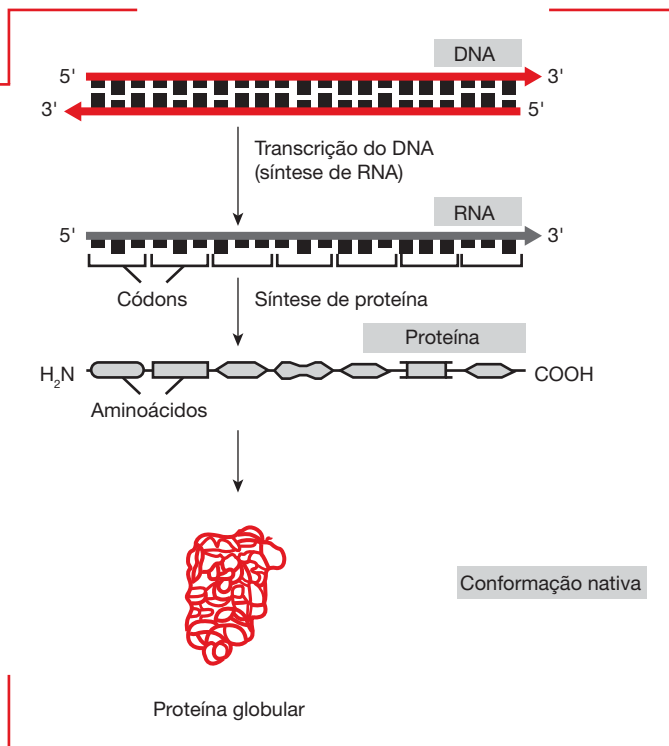
As proteínas são classificadas em duas classes principais: **fibrosas** e **globulares**. As proteínas fibrosas, em sua maioria, desenvolvem um papel estrutural nas células e nos tecidos animais. Nessa classe estão o colágeno, componente dos ossos e do tecido conectivo, e a α -queratina, presente em unhas e cabelo. As proteínas globulares são assim chamadas por possuírem uma estrutura enovela-

da e compactada, com formato globular; abundantes e essenciais, elas podem ser encontradas em quaisquer organismos. Um exemplo são as **enzimas**, eficientes catalisadores biológicos que aceleram as reações químicas. Com exceção de alguns RNAs (**ribozimas**), que possuem atividade catalítica, todas as enzimas são proteínas. Toda proteína globular tem uma estrutura única, enovelada de forma específica e de acordo com a função particular a ser executada.

Já que a estrutura de uma proteína determina a sua função, é importante conhecer as características estruturais dessa molécula. As proteínas são cadeias longas de aminoácidos e constituem mais da metade do peso seco de uma célula. Elas também são polímeros que desempenham inúmeras funções biológicas, além de determinarem a forma e a estrutura da célula. As proteínas são, ainda, conhecidas como moléculas que realizam o trabalho celular. Elas catalisam um extraordinário número de reações químicas, controlam a permeabilidade das membranas, regulam a concentração de metabólitos, reconhecem e ligam não covalentemente outras biomoléculas, proporcionam movimento e controlam a função gênica. Todo

Figura 1.3

Esquema de síntese de uma proteína globular a partir da informação contida no DNA e transcrita no RNA. A estrutura tridimensional funcional (proteína nativa) será determinada pela sequência de aminoácidos.



esse diverso número de funções é realizado por proteínas constituídas por apenas 20 aminoácidos, entre todos os aminoácidos possíveis, unidos por **ligações peptídicas**.

Como a conformação nativa, que permite à proteína realizar as suas funções, é uma consequência das propriedades individuais características dos aminoácidos presentes na molécula proteica, é importante revisar essas propriedades. Os **aminoácidos** são assim chamados por serem ácidos orgânicos, que possuem um átomo de carbono α (C_{α}) ligado a quatro grupamentos químicos diferentes. Um grupamento amínico ($-NH_2$), um grupamento carboxílico ($-COOH$), um átomo de hidrogênio ($-H$) e um grupamento variável, são denominados cadeia lateral ou radical ($-R$). A **Figura 1.4** apresenta a estrutura de todos os vinte aminoácidos conhecidos. Observando essa estrutura, é possível verificar que o grupamento R determina as diferenças estruturais entre os aminoácidos. Com exceção da glicina, que possui um átomo de hidrogênio também no radical, todos os demais aminoácidos possuem quatro grupamentos diferentes ligados ao C_{α} , dando origem a um carbono assimétrico. A presença desse carbono assimétrico gera duas moléculas de imagem especular não superpostas (**estereoisômeros**), denominadas, por convenção, isômero D (**dextro**) e L (**levo**). Com raras exceções, apenas os aminoácidos de forma L são encontrados nas proteínas. Em pH fisiológico, os grupamentos amínico e carboxílico dos aminoácidos são ionizados (NH_3^+ e COO^-), fazendo com que o aminoácido tenha cargas positiva e negativa na mesma molécula (**molécula dipolar**). A presença simultânea desses grupamentos determina o comportamento acidobásico dos

aminoácidos (moléculas **anfóteras**). O caráter acidobásico, bem como a carga elétrica do aminoácido, é determinado pelo pH do meio onde ele se encontra.

As cadeias laterais dos aminoácidos variam em tamanho, forma, carga elétrica, hidrofobicidade e reatividade. Os aminoácidos com cadeias laterais polares são hidrofílicos e tendem a se localizar na superfície da proteína, devido a suas interações com a água. Além disso, quanto mais aminoácidos polares estiverem presentes na proteína, mais solúvel em soluções aquosas ela será, porém os aminoácidos com grupamentos R apolares, por serem hidrofóbicos, tendem a estar presentes no interior das proteínas e provocarem sua insolubilidade em água. Na **Figura 1.4**, está representada a estrutura dos vinte aminoácidos em pH fisiológico (próximo de 7), que foram classificados conforme a solubilidade em água. A solubilidade varia com a polaridade de seus grupamentos R.

No grupo dos polares, arginina e lisina (polares básicos) são carregadas positivamente, e glutamato e aspartato (polares ácidos) são carregados negativamente, esses quatro aminoácidos são os principais responsáveis pelas cargas das proteínas. A histidina, que também tem carga positiva, auxilia na manutenção do pH (tampão fisiológico), pela sua capacidade de captar ou liberar prótons por meio do grupamento imidazol presente no radical. Pertencendo à classe dos aminoácidos polares neutros, a cisteína pode reagir com outros resíduos de cisteína por meio do grupamento tiol (SH), presente no radical para formar pontes dissulfeto (S-S) em uma reação de oxidação. Esta ligação tem uma participação importante na conformação das proteínas.

Os aminoácidos com cadeias laterais hidrofóbicas são quase insolúveis em água, consequência da presença de hidrocarbonetos nestes grupamentos. A fenilalanina, o triptofano e a tirosina possuem grupamentos aromáticos, responsáveis pela característica de absorção de luz ultravioleta das proteínas no comprimento de onda de 280 nm. Prolina é um aminoácido especial, uma vez que a sua cadeia lateral está ligada covalentemente ao nitrogênio do grupamento amínico, formando um anel rígido. A presença da prolina em uma cadeia proteica pode restringir a forma como a molécula irá se enovelar.

Durante a síntese da molécula proteica, os aminoácidos vão se unindo por uma ligação covalente (ligação peptídica) entre o grupamento carboxílico de um aminoácido com o grupamento amínico de outro aminoácido, ligados por uma reação de desidratação, com a perda de uma molécula de água. A molécula formada gera um peptídeo e mantém o seu caráter anfótero, já que sempre ficará um grupamento carboxílico livre em uma extremidade (C-terminal) e um grupamento amínico livre na outra extremidade (N-terminal) (ver **Figura 1.5**). A combinação de apenas dois aminoácidos forma um dipeptídeo; a união de poucos aminoácidos dá origem a oligopeptídeos. Um **polipeptídeo** é formado por muitos aminoácidos (às vezes, um número superior a 1.000).

A sequência de uma cadeia proteica é, por convenção, escrita com a extremidade N-terminal à esquerda e a C-terminal à direita. O tamanho de uma proteína é, em geral, expresso pela sua massa em daltons (Da). Existem também proteínas conjugadas, essencialmente importantes que, para realizarem sua atividade celular, necessitam estar ligadas a outras moléculas não proteicas, os denominados **grupos prostéticos**. A essa categoria pertencem as nucleoproteínas, as lipoproteínas e as cromoproteínas. As enzimas formam uma importante classe de proteínas que catalisa todas as reações químicas. Algu-

mas delas, quando envolvidas em reações de oxirredução, só possuem atividade quando ligadas covalentemente a uma coenzima (grupo prostético), como a nicotinamida adenina dinucleotídeo (NAD⁺), cuja estrutura é formada por um anel de nicotinamida, um anel de adenina e dois grupos de açúcares fosfatados unidos.

2.1.1 Estrutura tridimensional da proteína

A formação de uma cadeia polipeptídica, considerando a polimerização correta dos aminoácidos, é realizada no processo de tradução e determinada pela informação contida no RNA mensageiro (mRNA). A cadeia proteica sintetizada assume uma organização espacial precisa e necessária, para que a proteína desempenhe a sua função (conformação nativa). A **estrutura tridimensional** de uma proteína é a combinação de vários fatores, principalmente de interações entre os grupamentos químicos presentes nessa proteína e de limitações estereoquímicas impostas pela própria ligação peptídica, devido ao caráter de ressonância hídrica. Essa ressonância impede a rotação do carbono ligado ao nitrogênio, deixando todos os átomos envolvidos na ligação peptídica no mesmo plano (ver Figura 1.5). As proteínas são analisadas considerando os quatro níveis de organização estrutural. A **Figura 1.6** esquematiza esses níveis, do menor ao maior grau de complexidade, exemplificados pela estrutura primária e quaternária, respectivamente.

A **estrutura primária** é a primeira etapa na especificação da análise estrutural de uma proteína e refere-se à sequência de aminoácidos, ou seja, a ordem na qual os aminoácidos estão ligados para formar uma cadeia peptídica (Figura 1.6A). Nesta estrutura também estão localizadas as ligações peptídicas e as pontes S-S formadas entre os resíduos de cisteína. São as ligações peptídicas que estabilizam este tipo de estrutura. Cada proteína possui a sua estrutura primária específica, que, por sua vez,

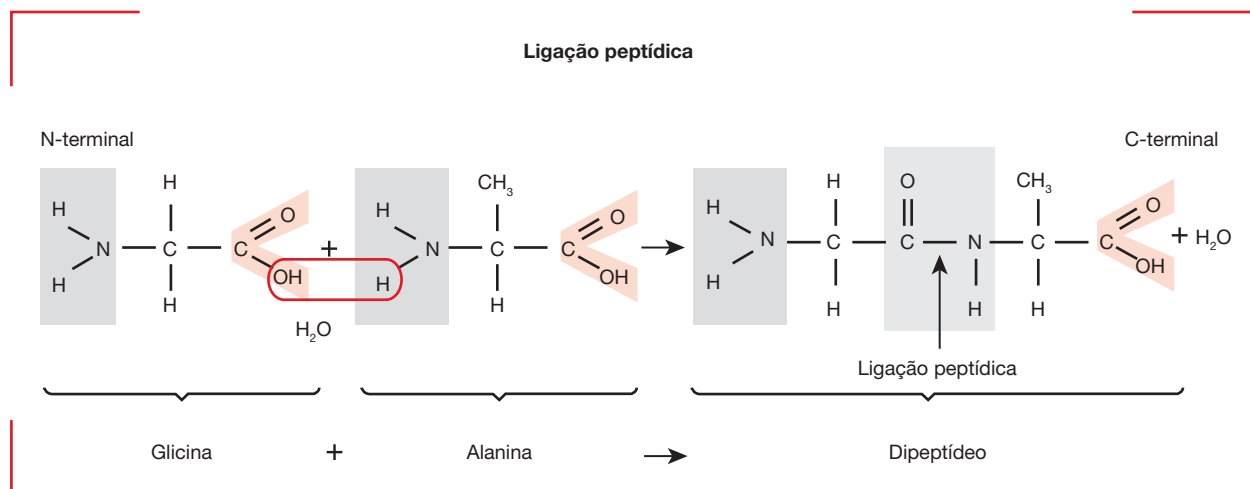


Figura 1.5

Desenho esquemático da formação de um peptídeo por meio da ligação entre o grupamento carboxílico de um aminoácido e o grupamento amínico de outro (ligação peptídica).

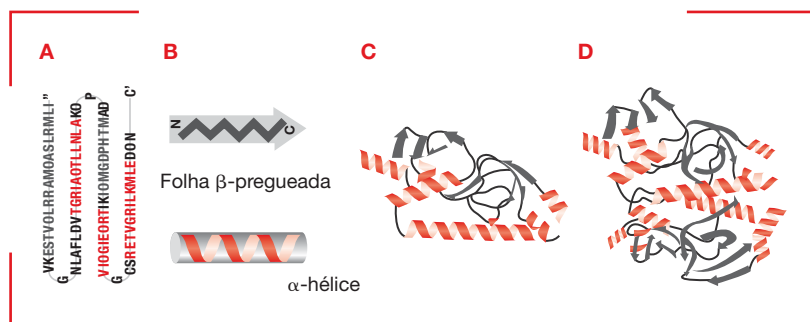


Figura 1.6

Desenho da estrutura tridimensional de uma proteína com os diferentes níveis organizacionais. As estruturas estão dispostas do nível menos complexo de organização para o mais complexo. **(A)** estrutura primária; **(B)** estruturas secundárias; **(C)** estrutura terciária; **(D)** estrutura quaternária.

determina a estrutura tridimensional. A importância biológica da sequência de aminoácidos é bem exemplificada na enfermidade hereditária humana, chamada de doença das células falciformes. Nessa doença, ocorrem mudanças biológicas profundas, provocadas pela substituição de um único aminoácido na molécula de hemoglobina (ver Capítulo 7).

A **estrutura secundária** se refere aos diversos arranjos espaciais de aminoácidos próximos na cadeia peptídica central, que provocam dobramentos. Tais dobramentos são denominados estruturas secundárias. Esses arranjos podem apresentar uma organização que se repete em intervalos regulares. As organizações de estruturas secundárias mais comuns são a α -hélice, a folha β -pregueada e as curvaturas (Figura 1.6B). As α -hélices, em que a cadeia peptídica se enrola em torno de um eixo imaginário, são estabilizadas por pontes de hidrogênio formadas entre o grupamento amínico da ligação peptídica de um aminoácido e o grupamento carboxílico da ligação peptídica do aminoácido situado quatro resíduos adiante na mesma cadeia polipeptídica. As folhas β -pregueadas ocorrem quando os aminoácidos assumem a conformação de uma folha pregueada. A folha β é estabilizada pela formação de pontes de hidrogênio entre os grupamentos amínico e carboxílico de cadeias polipeptídicas vizinhas, em vez de ocorrer dentro da própria cadeia, como é o caso das α -hélices. As curvaturas geralmente ocorrem na superfície da proteína, formando dobras definidas que redirecionam a cadeia. Essas curvaturas são compostas por três a quatro resíduos, e um deles é prolina. Essas dobraduras permitem a compactação de proteínas volumosas. Existem também, segmentos na proteína que não fazem ligações transversais não covalentes, formando configurações menos organizadas. Um polipeptídeo pode ser composto por um único tipo de estrutura secundária, como no caso da α -queratina, que é composta apenas por α -hélices, ou, ainda, a proteína pode possuir vários tipos de estruturas secundárias na mesma cadeia, como é o caso do citocromo C.

A **estrutura terciária** refere-se à forma como a cadeia polipeptídica está enovelada, incluindo o arranjo tridimensional de todos os átomos da molécula, inclusive os da cadeia lateral e do grupo prostético. Este nível estrutural é estabelecido quando diferentes estruturas secundárias se dispõem entre si. A estabilidade da estrutura

é mantida por pontes de hidrogênio entre grupos peptídicos não envolvidos na estrutura secundária, por pontes de hidrogênio entre grupos R, por interações hidrofóbicas, por ligações iônicas entre grupos carregados positiva e negativamente e por ligações covalentes do tipo dissulfeto (S–S) (Figura 1.6C). Nas proteínas globulares, as cadeias laterais dos aminoácidos mais hidrofóbicos tendem a se agregar no interior da molécula, e os grupamentos hidrofílicos se sobressaem da superfície da proteína. A estrutura tridimensional final pode ser composta pela combinação específica de estruturas secundárias, α -hélices, folhas β entre outras, que se enovelam, formando unidades globulares dobradas de forma compacta, chamadas de **domínios**. A estrutura terciária de proteínas maiores é subdividida em domínios, esses domínios possuem em torno de 100 a 150 aminoácidos e estão unidos pela cadeia peptídica.

Proteínas com mais de um polipeptídeo, formando subunidades (proteínas multiméricas), apresentam mais um nível estrutural, a **estrutura quaternária** (Figura 1.6D). Essa estrutura refere-se à disposição das subunidades proteicas que formam a molécula. O número de subunidades pode variar e a união entre elas ocorre de forma não covalente, por meio de interações eletrostáticas, pontes de hidrogênio e hidrofóbicas. Algumas proteínas, chamadas de alostéricas, exibem um efeito cooperativo entre as subunidades, de forma que uma alteração em uma dessas subunidades pode resultar em uma modificação em outra subunidade. Um bom exemplo é a molécula de hemoglobina, molécula tetramérica, composta de quatro cadeias polipeptídicas formando as subunidades. Cada subunidade se liga a uma molécula de oxigênio de forma cooperativa, ou seja, depois que uma molécula de oxigênio é ligada a uma subunidade, a ligação das demais moléculas é facilitada.

2.2 Carboidratos

Os carboidratos, ou monossacarídeos, são açúcares simples e representam uma das grandes classes de moléculas biológicas com uma variedade de funções celulares. Os polissacarídeos são polímeros com longas cadeias de unidades de monossacarídeos e constituem a principal fonte de energia celular. Eles são, também, constituintes estruturais importantes da parede celular, atuando como

sinais de reconhecimento específico e desempenhando um papel informacional. Além disso, são também substâncias intercelulares com função estrutural.

Os carboidratos são formados por ligações covalentes de carbono, em uma relação 1:1, e água $(CH_2O)_n$, em que n pode ser de 3 a 7. Eles são classificados de acordo com o número de átomos de carbono presentes na molécula: trioses (3), pentoses (5) ou hexoses (6). Todos os monossacarídeos podem conter vários grupamentos hidroxílicos e um grupamento aldeídico ou cetônico. Esses dois grupamentos podem reagir com um grupamento hidroxílico na mesma molécula, por meio de uma reação hemiacetal ou hemicetal, convertendo a estrutura linear para uma com formato de anel. O tipo de anel gerado será decorrente da ligação e de qual hidroxila se ligará com a carbonila. Um exemplo é a D-glicose, fonte primária de energia para a maioria das células. A estrutura da D-glicose pode se apresentar como uma cadeia linear ou como um anel hemiacetal, com duas estruturas diferentes (Figura 1.7). Quando o grupamento aldeídico do carbono 1 reage com o grupamento hidroxílico do carbono 5, o anel resultante possui 6 elementos, gerando uma D-glicopiranosose. Se a ligação hemiacetal ocorrer com o carbono 4, a estrutura é uma D-glicofuranose, cuja presença na natureza é muito mais rara. Todos os monossacarídeos, exceto a dihidroxiacetona, contêm um ou mais carbonos assimétricos gerando estereoisômeros opticamente ativos (D e L). A ciclização da estrutura linear gera novos isômeros, denominados anomeros α , por estarem ligados ao carbono anomérico.

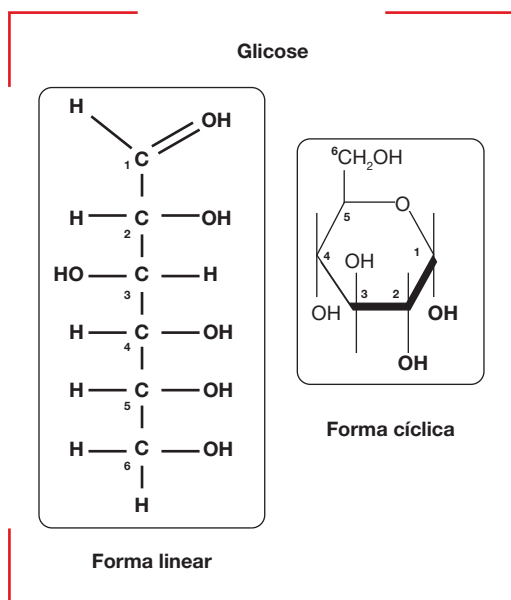


Figura 1.7

Estrutura linear e cíclica da glicose. A forma cíclica é resultante da ligação entre a carbonila do carbono 1 e a hidroxila do carbono 6, gerando uma glicopiranosose.

Os oligossacarídeos são moléculas formadas, na sua maioria, pela ligação de poucas unidades monoméricas. Um exemplo é a sacarose, um dissacarídeo formado pela união de uma molécula de glicose e uma de frutose que, após processado, produz o açúcar comum utilizado na alimentação (Figura 1.8). Os polissacarídeos mais importantes nos organismos vivos são o amido e o glicogênio, pois representam substâncias de reserva, ou seja, a forma de estocagem de energia nas células vegetais e animais. O glicogênio é um polissacarídeo formado pela ligação de várias moléculas de glicose (ver Figura 1.9). A celulose também é um importante polissacarídeo e é o principal elemento estrutural da parede celular da célula vegetal.

Os dissacarídeos, assim como os polissacarídeos, são formados por monossacarídeos, unidos covalentemente por **ligações glicosídicas**. Essas ligações são formadas quando um grupamento hidroxílico do **carbono anomérico** de um carboidrato reage com o grupamento hidroxílico de outro carboidrato (ver Figura 1.8). Esses grupamentos hidroxílicos livres podem, ainda, ligar-se com outros grupamentos amínico, sulfato e fosfato de diferentes moléculas, formando moléculas mais complexas, como os glicosaminoglicanos, principais componentes da matriz extracelular.

2.3 Lipídeos

Os lipídeos formam um grupo de compostos característico, que possuem múltiplas funções celulares e ocorrem com frequência na natureza. Geralmente, são moléculas pequenas que apresentam uma forte tendência a se associarem por meio de forças não covalentes, formando agregados lipídicos. Os lipídeos são, em geral, caracterizados por um tipo de estrutura própria, conforme é mostrado na Figura 1.10A para os ácidos graxos. Uma molécula de ácido graxo possui duas regiões distintas: uma região polar, hidrofílica, conectada a uma região apolar, hidrofóbica, constituída de uma cadeia de hidrocarboneto. Esse tipo de estrutura caracteriza os lipídeos como um grupo de compostos pouco solúveis em água e solúveis em solventes orgânicos. Essa característica molecular promove as associações do tipo **anfipáticas** – reuniões das moléculas lipídicas com interações não covalentes em meio aquoso. Essas interações possuem consequências consideráveis em nível celular, a mais importante delas é a tendência de os lipídeos formarem micelas e bicamadas, que constituem as membranas biológicas (ver Figura 1.10B e C). A estrutura exata, formada quando o lipídeo está em contato com a água, depende da estrutura molecular específica das regiões hidrofílicas e hidrofóbicas da molécula. A seguir, alguns dos principais lipídeos celulares.

2.3.1 Ácidos graxos

Os lipídeos mais simples são os **ácidos graxos**, também constituintes dos lipídeos mais complexos. Sua estrutura básica exemplifica a maioria das moléculas de lipídeos

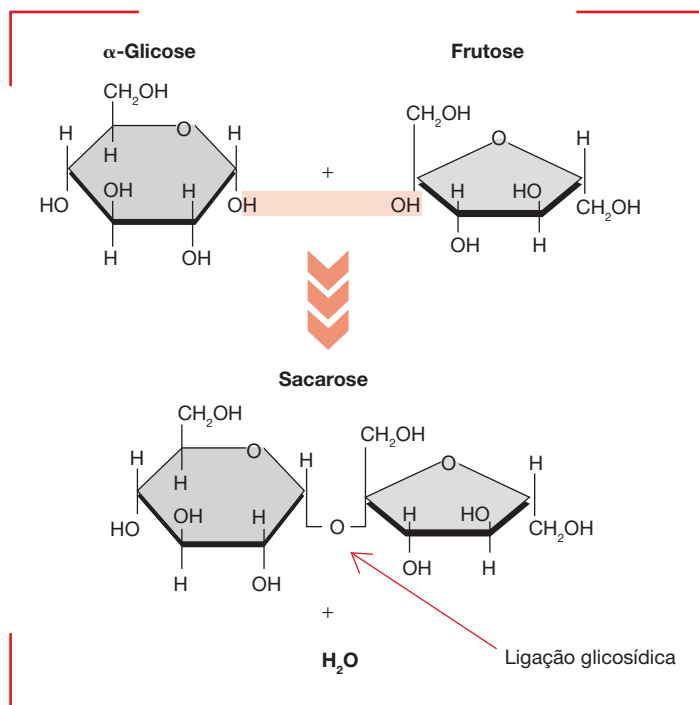


Figura 1.8

Estrutura química do dissacarídeo sacarose. A sacarose é formada pela ligação glicosídica entre uma molécula de glicose e uma molécula de frutose.

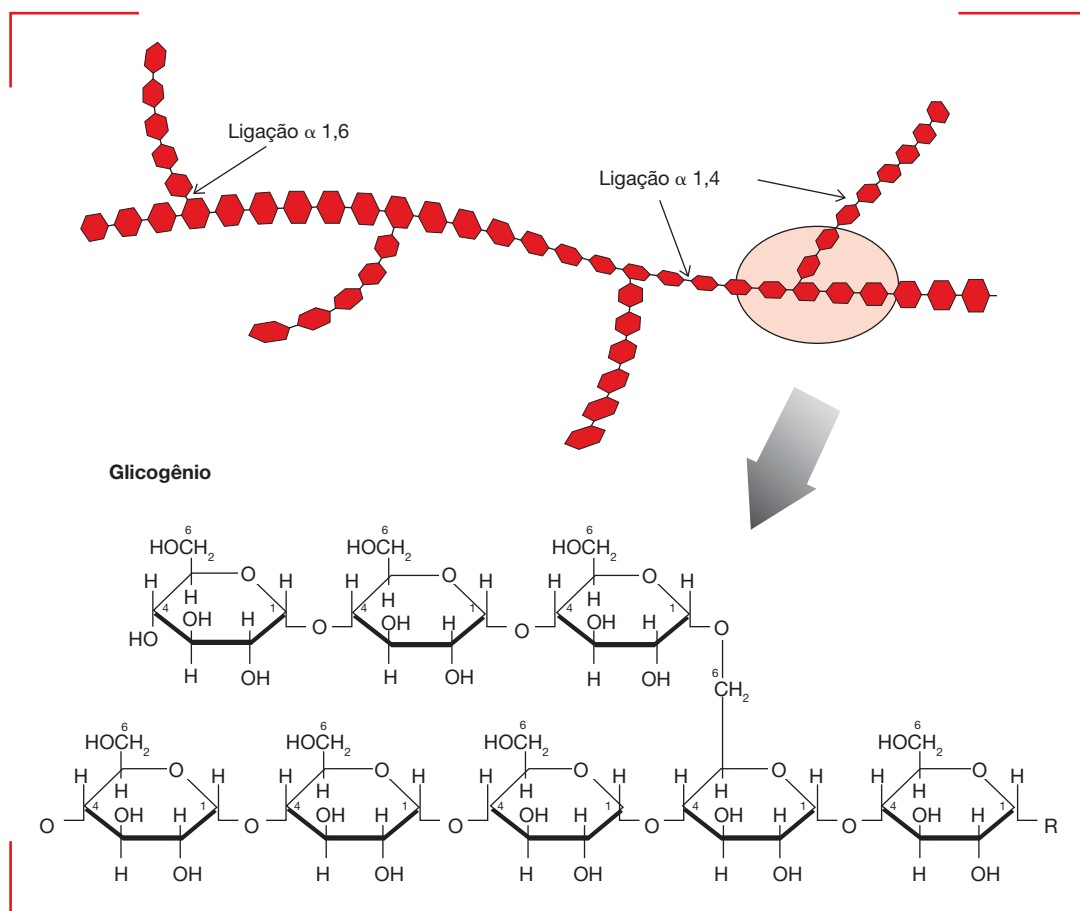


Figura 1.9

Desenho da estrutura polimérica do glicogênio. O glicogênio é um polímero ramificado de monômeros de glicose unidos por ligações glicosídicas, que ocorrem entre os carbonos 1 e 4, exceto onde a molécula se ramifica (1 e 6). A zona correspondente ao círculo é mostrada ampliada.

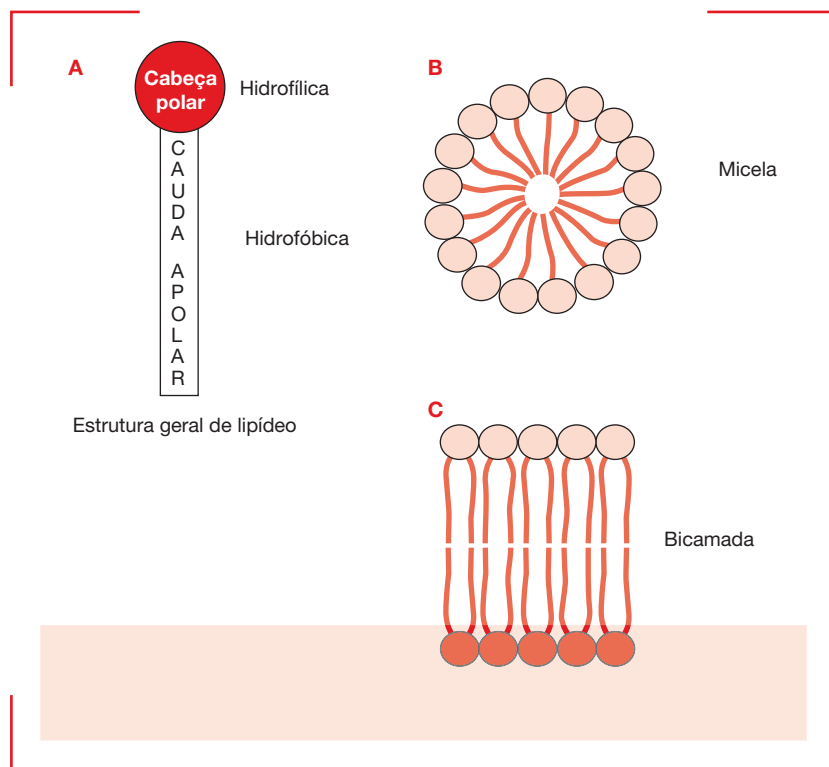


Figura 1.10

Desenho esquemático de uma estrutura de lipídeos (A) e formas de associações anfipáticas (B). A associação em bicamadas (C) é a base molecular das membranas.

encontrada em grande quantidade nas células humanas. A estrutura do ácido graxo é formada por uma longa cadeia hidrocarbonada, hidrofóbica e pouco reativa quimicamente. Em geral, os ácidos graxos encontrados nos organismos vivos contêm um número par de átomos de carbono, e sua cadeia de hidrocarboneto não é ramificada. Os ácidos graxos são classificados em saturados, insaturados ou poli-insaturados, dependendo das ligações entre os átomos de carbono. Nos ácidos graxos saturados, a cadeia contém apenas ligações simples, se existirem ligações duplas, os ácidos graxos são insaturados. Os ácidos graxos com mais do que uma ligação dupla são chamados de poli-insaturados. Dois ácidos graxos poli-insaturados, classificados como essenciais, são o ácido linoleico, com 18 carbonos e duas ligações duplas, e o linolênico, também com 18 carbonos, porém com três ligações duplas.

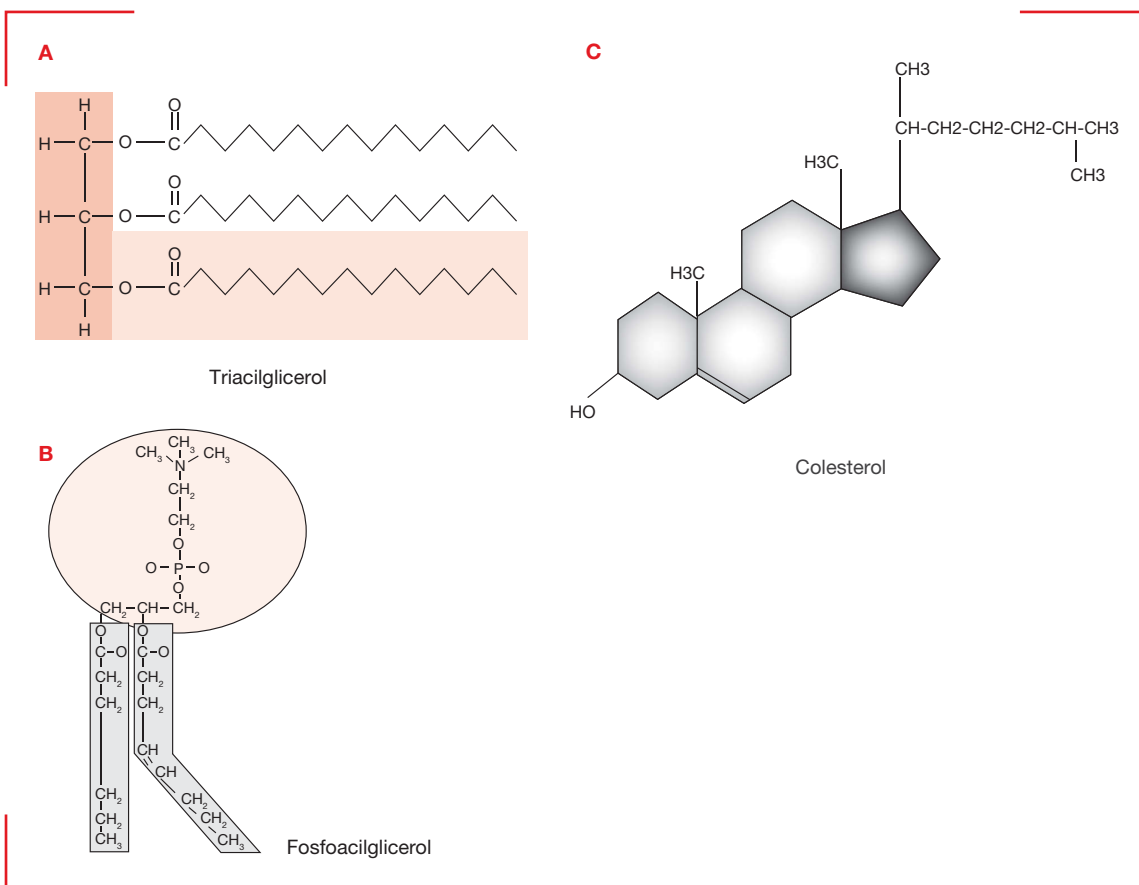
A molécula de ácido graxo é formada por um grupoamento carboxílico extremamente hidrofílico, ionizável em solução (COO⁻). Eles formam moléculas de triacilgliceróis (triglicerídeos), que são triésteres de ácido graxo e glicerol (ver **Figura 1.11A**), conhecidos como gorduras. Os triacilgliceróis são a forma de estocagem de lipídeos no citoplasma de muitas células, pois, devido à presença da cadeia carbonada, servem como excelente fonte de energia. Dessa forma, os triacilgliceróis são mais eficientes como estoque de energia que os carboidratos e, por essa razão, muito utilizados por vários organismos, incluindo os animais superiores.

2.3.2 Fosfoacilgliceróis

Os **fosfoacilgliceróis**, ou fosfolipídeos, são pequenas moléculas lipídicas, compostas por longas cadeias de ácido graxo e glicerol, ligadas a um grupo altamente polar (Figura 1.11B). Eles diferem dos triacilgliceróis por possuírem apenas duas moléculas de ácidos graxos unidas a uma molécula de glicerol, cuja terceira hidroxila está esterificada a um ácido fosfórico (ácido fosfatídico). Esse fosfato pode estar unido a uma molécula hidrofílica (colina, etanolamina, inositol ou serina), conforme o tipo de fosfoacil. A natureza anfipática dos fosfoacilgliceróis é responsável pelas associações moleculares, que formam a membrana celular e conferem muitas das suas propriedades. As membranas são basicamente duas camadas de fosfoacilgliceróis dispostas de forma que as regiões hidrofóbicas ficam voltadas para o interior e as regiões hidrofílicas situadas nas interfaces aquosas. Esse arranjo em bicamada é a unidade estrutural de quase todas as membranas biológicas. O núcleo hidrofóbico da estrutura atua como uma barreira de impermeabilidade.

2.3.3 Esteroides

Os **esteroides** são um grande grupo de moléculas, que agregam várias funções, e incluem um considerável número de hormônios, entre eles os hormônios sexuais de animais superiores. O colesterol é o esteroide de maior importância, fazendo parte de membranas de células, principalmente de animais (Figura 1.11C). Os esteroides derivam de uma estrutura geral que contém 3 anéis de 6 carbonos (anéis A, B e C) e um anel com 5 átomos. A

**Figura 1.11**

Desenho da arquitetura molecular de lipídeos. (A) Triacilglicerol; (B) Fosfoacilglicerol; (C) Colesterol.

molécula de colesterol é pouco anfipática, pois existe um grupamento hidroxílico localizado no final da molécula (ver Figura 1.11C). O restante da molécula de colesterol é solúvel no interior hidrofóbico das membranas. Os anéis cicloexanos fusionados dessa molécula formam uma estrutura bastante rígida e sua presença na membrana tende a romper a regularidade da estrutura, conferindo-lhe maior rigidez. Essa estrutura compacta também é responsável pelos efeitos danosos à saúde, como a aterosclerose – doença cardiovascular provocada pelo armazenamento de colesterol nos vasos sanguíneos.

2.4 Ácidos nucleicos

Os **ácidos nucleicos** são macromoléculas de grande importância biológica em todos os organismos vivos. A partir dos ácidos nucleicos as células recebem as informações sobre quais proteínas sintetizar, qual a sequência de aminoácidos de sua estrutura e qual a função dessas moléculas. Eles são, portanto, as moléculas que estocam e transmitem a informação genética na célula. Toda essa informação fica em unidades gênicas, localizadas nos cromossomos das células. Tal informação é decifrada por meio do código genético, cuja tradução resulta na síntese proteica.

Existem dois tipos de ácidos nucleicos: ácido desoxirribonucleico (DNA) e ácido ribonucleico (RNA). Ambos são polímeros lineares de nucleotídeos, unidos por ligações fosfodiéster. O número de monômeros em uma molécula de ácido nucleico é, na maioria dos casos, muito maior que o número de aminoácidos em uma proteína. Os RNAs variam em tamanho, podendo ter de dez a milhares de nucleotídeos. Tanto o DNA como o RNA consistem em apenas quatro diferentes tipos de nucleotídeos.

Cada nucleotídeo é composto por um grupamento fosfato, um açúcar (pentose) e uma base nitrogenada (púrica ou pirimídica) unidos por ligações covalentes (ver Figura 2.1 do Capítulo 2). As diferenças entre os dois ácidos nucleicos residem no tipo de açúcar e na composição de bases da molécula. No RNA, a pentose é sempre a ribose e, no DNA, é a desoxirribose. As bases que formam os ácidos são cinco: adenina (A), guanina (G) e citosina (C), encontradas tanto no DNA como no RNA; a base timina (T) presente apenas no DNA, e a base uracila (U), apenas no RNA (ver Figura 2.2 do Capítulo 2). Essas letras (A, C, G, T e U) são utilizadas para indicar uma sequência de nucleotídeos em um ácido nucleico. Além dessas cinco bases, existem bases não usuais com estrutura um pouco

diferenciada, que ocorrem principalmente no RNA, como a hipoxantina e a inosina.

O açúcar faz a ligação entre a base e o grupamento fosfato. Conforme mostrado na Figura 2.3 do Capítulo 2, a ligação é entre o C 1 do açúcar e o N 9 da purina ou N 1 da pirimidina. O grupamento hidroxílico do C 5 do açúcar é substituído por um grupamento fosfato. A combinação de uma base e um açúcar, sem o grupamento fosfato, constitui um nucleosídeo. Quando nucleotídeos polimerizam-se para formar um ácido nucleico, o grupamento hidroxílico ligado ao C 3 do açúcar de um nucleotídeo forma uma ligação éster com o fosfato de outro (ligação fosfodiéster), eliminando, assim, uma molécula de água e formando um esqueleto de unidades repetidas de fosfatopentose com as bases ligadas ao grupamento lateral.

Uma sequência de nucleotídeos possui uma orientação química de muita importância. Em uma fita de DNA ou RNA, em uma das extremidades, há um grupamento fosfato ligado ao C 5 (carbono 5') do açúcar (extremidade 5') e na outra extremidade há um agrupamento hidroxílico ligado ao C 3 (carbono 3') do açúcar (extremidade 3') (ver Figura 2.3 do Capítulo 2). Convencionou-se escrever e ler a sequência nucleotídica da esquerda para a direita, no sentido 5' → 3'. A cadeia polinucleotídica possui individualidade, determinada pela sequência de suas bases, conhecida como estrutura primária e é nessa estrutura primária que a informação genética está contida. O gene corresponde a uma sequência particular de DNA, codificadora de uma informação (proteína ou RNA).

2.4.1 Ácido desoxirribonucleico

O ácido desoxirribonucleico (DNA) encontra-se nos organismos vivos como moléculas, de alto peso molecular, por exemplo, *E. coli* que tem uma só molécula de DNA circular de $4,2 \times 10^6$ pb (pares de bases) e um comprimento total de 1,4 mm. A quantidade de DNA nos organismos superiores pode ser centenas de vezes maior (700 vezes no caso do homem); o DNA de uma só célula diploide humana, completamente estendido, pode ter um comprimento de 1,7 m.

Toda a informação genética de um organismo está acumulada na sequência linear das quatro bases. A estrutura primária de todas as proteínas (quantidade e sequência dos 20 aminoácidos) deve estar codificada por um alfabeto de quatro letras (A, T, G e C). Entre 1949 e 1953, Chargaff, estudando a composição de bases do DNA, demonstrou que embora a composição variasse de uma espécie para outra, em todos os casos a quantidade de adenina era igual à de timina (AT), e a de citosina era igual à de guanina (CG). Assim, o número total de purinas era igual ao de pirimidinas (A+G C+T). Por outro lado, a relação AT/GC variava, de forma considerável, entre as espécies.

Em 1953, com base em dados de difração por raios X, Watson e Crick propuseram um modelo para a estrutura do DNA. Esse modelo explicava as regularidades da

composição de bases, principalmente sua duplicação na célula. O modelo da estrutura da molécula do DNA está ilustrado na Figura 2.4, no Capítulo 2. O DNA é composto de duas cadeias polinucleotídicas associadas, que se enrolam para formar uma hélice dupla em torno de um eixo central com giro para a direita, na maioria das vezes. As bases estão no interior da hélice, em um plano perpendicular ao eixo helicoidal, interagindo através de pontes de hidrogênio que unem as duas cadeias. As pontes de hidrogênio, entre as bases de cadeias opostas, e as interações hidrofóbicas, entre as bases adjacentes na mesma cadeia, estabilizam essa estrutura. Uma vez que, entre os dois açúcares das cadeias opostas existe uma distância fixa, apenas alguns pares de bases podem acomodar-se dentro da estrutura. Assim, os únicos pares possíveis são A/T e C/G. É importante notar que entre A e T há formação de duas pontes de hidrogênio e entre C e G, três pontes; em consequência disso, o par C/G é mais estável que o par A/T. Além dessas pontes de hidrogênio, ainda existem interações hidrofóbicas que estabilizam a hélice dupla. A orientação das duas fitas de DNA é antiparalela, ou seja, o sentido 5' → 3' de cada uma é oposto ao da outra.

2.4.2 Ácido ribonucleico

O ácido ribonucleico (RNA) é uma molécula de ácido nucleico formada, em geral, por uma única cadeia com grande diversidade de conformações. A sequência de bases (estrutura primária) é similar à do DNA, exceto pela substituição da desoxirribose por ribose e de timina por uracila. Existem três classes principais de ácido ribonucleico: RNA mensageiro (mRNA), que contém a informação genética para a sequência de aminoácidos; RNA transportador (tRNA), que identifica e transporta as moléculas de aminoácidos até o ribossomo; e RNA ribossômico (rRNA), que representa 50% da massa dos ribossomos. O ribossomo proporciona as condições moleculares para a síntese dos polipeptídeos. Todos os tipos de RNAs intervêm na síntese proteica. A estrutura e a função dos RNAs serão detalhadas nos Capítulos 2 e 12.

Ainda que a molécula de RNA seja constituída por apenas uma cadeia polinucleotídica, sua estrutura não é simples. As moléculas de RNA possuem regiões onde as bases são complementares, nas quais se estabelecem pontes de hidrogênio entre pares A/U e G/C de diferentes regiões da mesma molécula. Como consequência dessas interações, a molécula se volta sobre si mesma, formando estruturas características, chamadas de grampos e alças. Nas regiões pareadas, a molécula se adapta em uma estrutura helicoidal, semelhante a do DNA. A estrutura compacta das moléculas de RNA tem importantes consequências biológicas. As diferenças de tamanho e conformações dos vários tipos de RNA permitem que eles desempenhem funções específicas. Os tRNAs possuem uma estrutura conformacional que permite o pareamento dos nucleotídeos do anticódon presentes na molécula com os nucleotídeos do códon do mRNA (ver Capítulo 12). Alguns RNAs, denominados ribozimas, têm atividade catalítica.



Resumo

Apesar da grande diversidade existente entre os organismos, todos são constituídos por pequenas unidades morfológicas, denominadas células, separadas do meio externo por uma membrana citoplasmática. Os organismos podem ser classificados em dois grandes grupos, conforme a sua organização celular. Os organismos eucarióticos possuem um núcleo delimitado por uma membrana que envolve o material genético e a compartimentalização de várias de suas funções em organelas. Os organismos procarióticos são estruturalmente mais simples, em geral, unicelulares e o seu cromossomo está condensado no nucleóide, porém, sem estar envolvido por uma membrana. A vida celular depende de milhares de interações químicas realizadas por moléculas que são comuns a todo tipo celular. A água é o componente encontrado em maior quantidade na célula e serve como solvente natural para que as reações químicas celulares aconteçam. Além da água, a célula é constituída de pequenas moléculas e de biopolímeros formados por muitas cópias de uma pequena molécula, ligada em cadeias por ligações covalentes. Os três tipos de biopolímeros presentes nas células são: os ácidos nucleicos – formados pelos nucleotídeos unidos por uma ligação fosfodiéster; as proteínas – constituídas pelos aminoácidos li-

gados através da ligação peptídica; e os carboidratos ou polissacarídeos – cujos monômeros são os açúcares ligados por ligação glicosídica. As células ainda possuem uma grande quantidade de lipídeos que são, preferencialmente, moléculas pequenas, como os ácidos graxos, apresentando como característica principal serem praticamente insolúveis em água. A estrutura geral de um lipídeo é de uma molécula anfipática com uma cauda apolar e uma cabeça polar. Esse tipo de estrutura é responsável pelas interações não covalentes realizadas pelos fosfolipídeos para formarem a dupla camada que origina as membranas celulares. As proteínas são sintetizadas no organismo conforme a informação contida nos genes que constituem o genoma. O gene corresponde a um segmento de ácido desoxirribonucleico (DNA), que é transcrito em diferentes tipos de ácidos ribonucleicos (RNAs) para a tradução da molécula de proteína. A estrutura tridimensional da proteína é determinada pela sequência de aminoácidos presente na molécula. Aminoácidos com cadeias laterais hidrofóbicas tendem a agregar-se no interior da proteína, evitando o ambiente aquoso. A preservação da conformação nativa de uma proteína é dependente das condições do meio, como pH e temperatura, em que a proteína se encontra.



Leituras recomendadas

Alberts B, Johnson A, Walter P, Lewis J, Raff M, Roberts K. Molecular biology of the cell. 5th ed. New York: Garland Science; 2008.

Campbell MK, Farrell SO, Tasks A. Bioquímica. 5. ed. São Paulo: Thomson; 2007.

Devlin TM. Manual de bioquímica com correlações clínicas. São Paulo: Edgard Blucher; 2007.

Lewin B. Genes IX. 9. ed. Porto Alegre: Artmed; 2009.

Lodish H, Baltimore D, Berk A, Zipursky L, Matsudaira PT. Biologia celular e molecular. 5. ed. Porto Alegre: Artmed; 2005.

Persing DH, editor. Molecular microbiology: diagnostic principles and practice. Washington: ASM Press; 2004.

Voet D, Voet JG. Bioquímica. 3. ed. Porto Alegre: Artmed; 2006.

Augusto Schrank

Estrutura dos Ácidos Nucleicos

1. DNA 18

- 1.1 Composição química 18
- 1.2 Hélice dupla 19
 - 1.2.1 Estrutura 19
 - 1.2.2 Propriedades químicas 23
- 1.3 Supertorção 25
- 1.4 Topoisomerases 27
- 1.5 Tipos de DNA 29

1.6 Outras estruturas do DNA 31

- 1.6.1 Curvatura 31
- 1.6.2 Estruturas cruciformes 31
- 1.6.3 Junções de Holliday 31

2. RNA 31

- 2.1 Composição química 32
- 2.2 Estrutura secundária 32
- 2.3 Classes de RNA 32

A estrutura dos ácidos nucleicos está relacionada à sua função. Essas moléculas são polímeros formados por cadeias de nucleotídeos, cuja composição (tipo e sequência) determina suas características químicas. Essas características definem a sua interação com outras macromoléculas na célula, em particular, com as proteínas e a sua conformação espacial (forma da molécula). A conformação espacial está diretamente relacionada à função e à atividade das macromoléculas na célula. Na natureza, são encontradas todas as formas de DNA e RNA: DNA de cadeia dupla, DNA de cadeia simples, RNA de cadeia dupla e RNA de cadeia simples.

A importância dos ácidos nucleicos na célula levou algum tempo para ser compreendida. A função do DNA, de armazenar a informação genética na sequência de suas quatro bases nitrogenadas componentes (**A**, **C**, **G** e **T**), foi, durante muito tempo, atribuída às proteínas. Devido à variedade de combinações possíveis entre os 20 aminoácidos diferentes com as proteínas (podendo, assim, conferir maior variabilidade e potencial para armazenar informações genéticas), e à dificuldade de obter preparações de DNA totalmente livres de proteínas, a comprovação de qual molécula armazenava a informação genética foi demorada. O DNA foi descoberto por Friedrich Miescher, em 1869, entretanto, as suas funções primordiais foram sugeridas pela primeira vez apenas 75 anos depois, em 1944, por Avery, MacLeod e McCarty. Esta proposição foi definitivamente comprovada em 1952, por Hershey e Chase e, no ano seguinte, Watson e Crick propuseram a estrutura de hélice dupla do DNA. Estes últimos autores também apresentaram estudos sobre como esta molécula seria duplicada e lançaram as bases para a transferência da informação genética para as proteínas, utilizando os RNAs como moléculas intermediárias neste processo. Em relação aos estudos dos RNAs, recentemente os genomas possibilitaram a descoberta de muitas novas classes de RNAs, revelando que essas novas classes têm participação fundamental na fisiologia da célula.

As descobertas mencionadas promoveram uma série de avanços na biologia. Atualmente, sabe-se que a molécula de DNA não é um mero depositário da informação genética, e que os RNAs representam muito mais do que apenas intermediários na transferência das informações às proteínas. Neste capítulo, serão abordados aspectos da estrutura do DNA e dos RNAs e de sua inter-relação com as proteínas.

1. DNA

1.1 Composição química

O **ácido desoxirribonucleico** (DNA) é um polímero composto por unidades de **desoxirribonucleotídeos**. Ele é uma macromolécula muito longa, para termo de comparação, enquanto uma célula eucariótica comum tem cerca de 50 μm , o menor cromossomo humano, se estendido, possui aproximadamente 14.000 μm , ou seja,

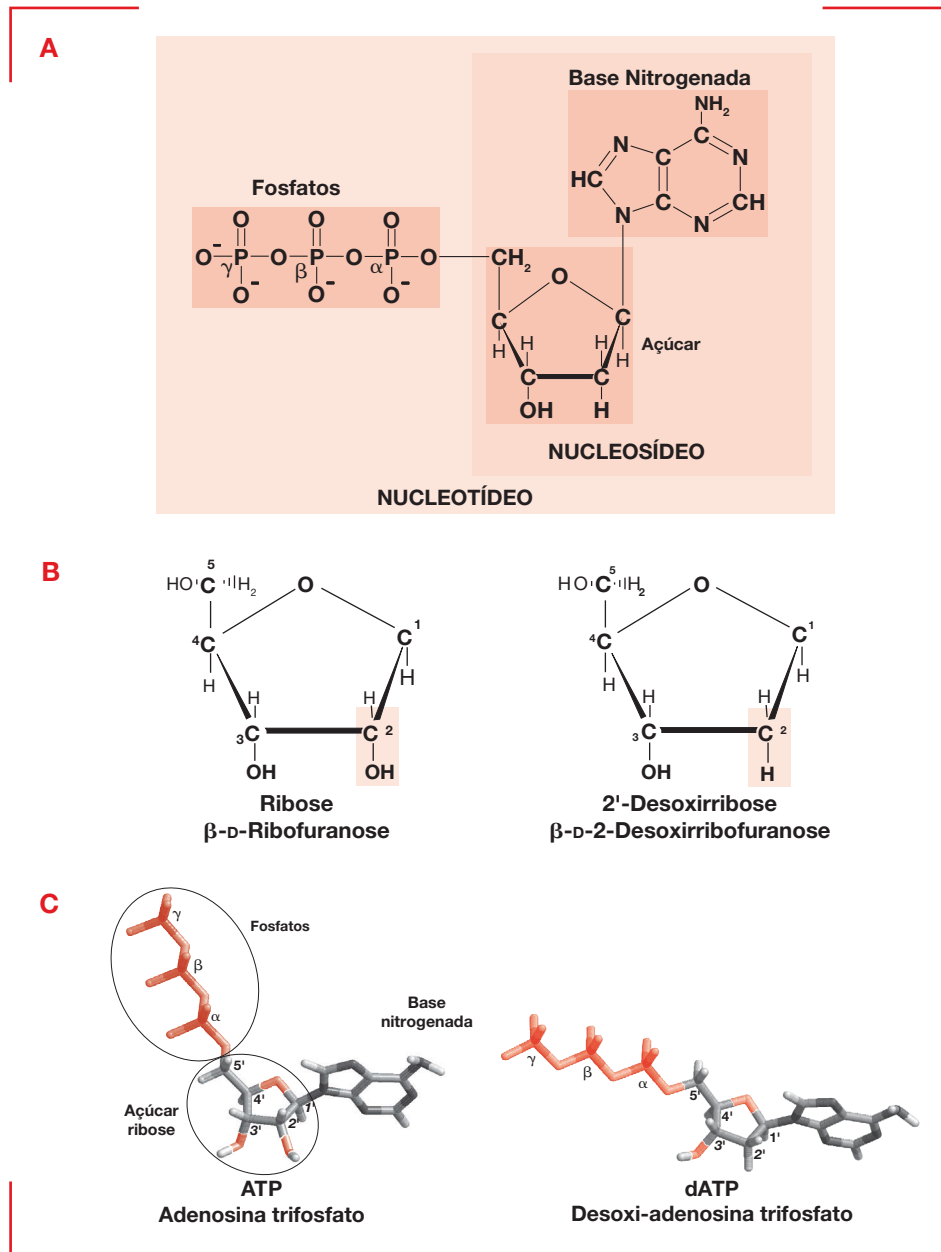
um único cromossomo humano é 280 vezes mais longo que o diâmetro da célula que o contém.

Os desoxirribonucleotídeos são compostos por: um açúcar, uma base nitrogenada e grupamentos fosfato (PO_4^-). O açúcar é a pentose **desoxirribose** (2'-desoxi-D-ribose); que está ligada pelo carbono 1' a uma base nitrogenada heterocíclica e pelo carbono 5' a grupamentos fosfato (**Figura 2.1**). As **bases nitrogenadas** podem ser de dois tipos: as bases **púricas** (derivadas das purinas), que são a adenina (A) e a guanina (G), e as bases **pirimídicas** (derivadas das pirimidinas), que são a citosina (C) e a timina (T) (**Figura 2.2**). A ligação entre a base nitrogenada e a pentose é denominada **ligação glicosídica**. A molécula composta pela base nitrogenada ligada ao açúcar, sem grupamentos fosfato, é denominada **nucleosídeo**. Quando um grupamento fosfato está ligado ao carbono 5' da pentose, a molécula é denominada **nucleotídeo**, que pode ser mono, di ou tri-fosfatado, dependendo do número de fosfatos a ele ligados. O primeiro grupamento fosfato ligado ao nucleotídeo é denominado α (alfa), o segundo β (beta) e o terceiro γ (gama). A molécula de DNA conterá sempre nucleotídeos mono-fosfatados e os precursores para a sua síntese serão sempre trifosfatados. Os desoxirribonucleotídeos são denominados de acordo com a base nitrogenada componente e com o número de grupamentos fosfato ligados à pentose (**Tabela 2.1**).

Na molécula de DNA, os desoxirribonucleotídeos formam cadeias, sendo unidos entre si por ligações covalentes, formando **pontes fosfodiéster** estabelecidas entre o grupamento fosfato ($5'\text{-PO}_4$, ou, simplificadamente $5'\text{-P}$) e o grupamento hidroxílico ($3'\text{-OH}$) do carbono 3' do nucleotídeo adjacente (**Figura 2.3**). Isso confere às cadeias polinucleotídicas uma propriedade importante, que é a **direcionalidade**. O “primeiro” nucleotídeo da cadeia terá uma extremidade 5' com um grupamento fosfato, e o “último” nucleotídeo na extremidade 3' terá um grupamento hidroxílico. O próximo nucleotídeo a ser adicionado à cadeia, obrigatoriamente, será adicionado na extremidade 3'-OH da cadeia polinucleotídica (ver Capítulo 5). As cadeias polinucleotídicas são, por convenção, representadas na orientação $5' \rightarrow 3'$, e apenas as letras indicativas das bases nitrogenadas são representadas, como no exemplo:

$5'\text{-AACGTTGCTATCGT-3'}$, ou, mais comumente, **AACGTTGCTATCGT**

Estudos detalhados sobre a estrutura cristalográfica do DNA demonstraram que os ângulos das ligações covalentes e as interações entre as nuvens eletrônicas conferem certa rigidez à conformação espacial das cadeias do DNA, mesmo quando somente uma das cadeias (fitas) é formada. O maior grau de rotação ocorre nas ligações entre o oxigênio e o fósforo (ligação fosfodiéster) e entre a pentose e a base nitrogenada (ligação glicosídica). Entretanto, mesmo nessas posições, existem localizações preferenciais, demonstrando que as moléculas de DNA assumem conformações espaciais não aleatórias. Estas características químicas das cadeias de DNA contribuem para a formação da hélice dupla.

**Figura 2.1**

Estrutura dos nucleotídeos. (A) Estrutura geral. (B) Os açúcares componentes; em detalhe, o grupo OH na posição C2 da ribose e o H na desossirribose. (C) Representação das estruturas cristalográficas do ATP e do dATP. Observe a grande diferença no posicionamento espacial dos átomos pela simples presença, ou ausência do oxigênio no carbono 2' da pentose.

1.2 Hélice dupla

1.2.1 Estrutura

A elucidação da estrutura secundária da molécula de DNA foi fundamental na sua definição como molécula armazenadora da informação genética. A principal propriedade para esta função é a capacidade de “autorreplicação” (Capítulo 6). Além de muitos estudos sobre a química do DNA, a utilização de métodos de **difração por raios X** foi necessária para a descoberta da estrutura do DNA. A difração permite que seja atribuída a localização espacial (tridimensional, nos eixos x , y e z) dos átomos em uma molécula cristalizada e, assim, definir a conformação espacial da molécula. Em 1953, James Watson e Francis

Crick propuseram um modelo de estrutura tridimensional do DNA com base nos estudos de difração por raios X, conduzidos por Rosalind Franklin e Maurice Wilkins, e em outros estudos químicos da molécula.

O modelo proposto por Watson e Crick mostrou que o DNA é uma hélice dupla e que as suas duas fitas se enrolam em torno do eixo da hélice. As desossirriboses, ligadas umas às outras pelos grupamentos fosfato, ficam externas em relação às bases nitrogenadas, como se fossem o corrimão de uma escada circular, expostas ao meio aquoso. As ligações fosfodiéster nas duas fitas estão em direções opostas – uma fita na direção $5' \rightarrow 3'$ e a outra na direção $3' \rightarrow 5'$, sendo, portanto, ditas **anti-paralelas**. O pareamento das bases é fundamental para

Figura 2.2

Estrutura das bases nitrogenadas púricas (A) e pirimídicas (B). Estão representadas a forma planar e a representação em varetas.

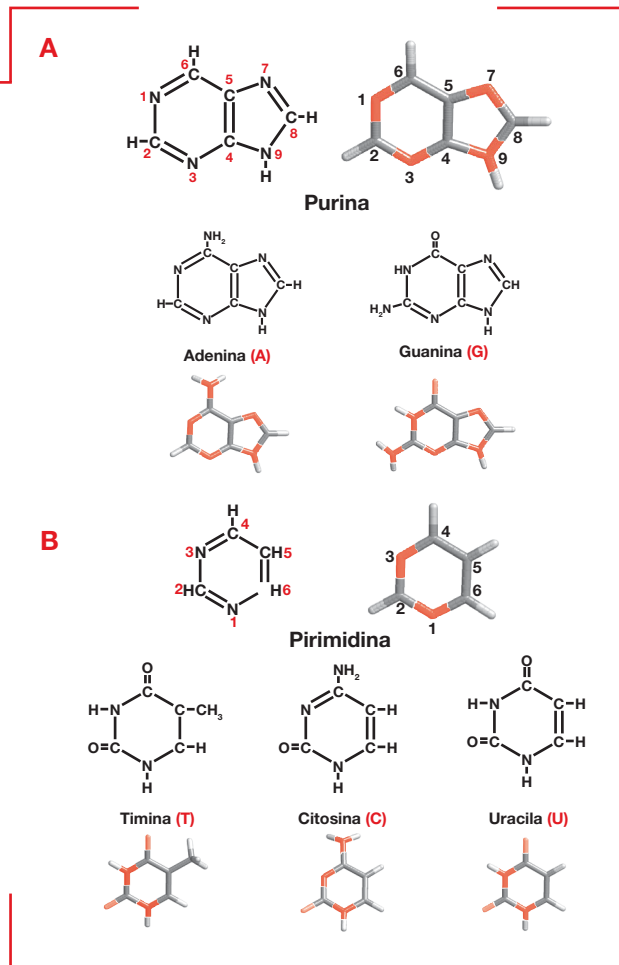
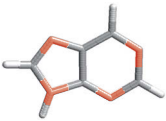
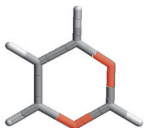
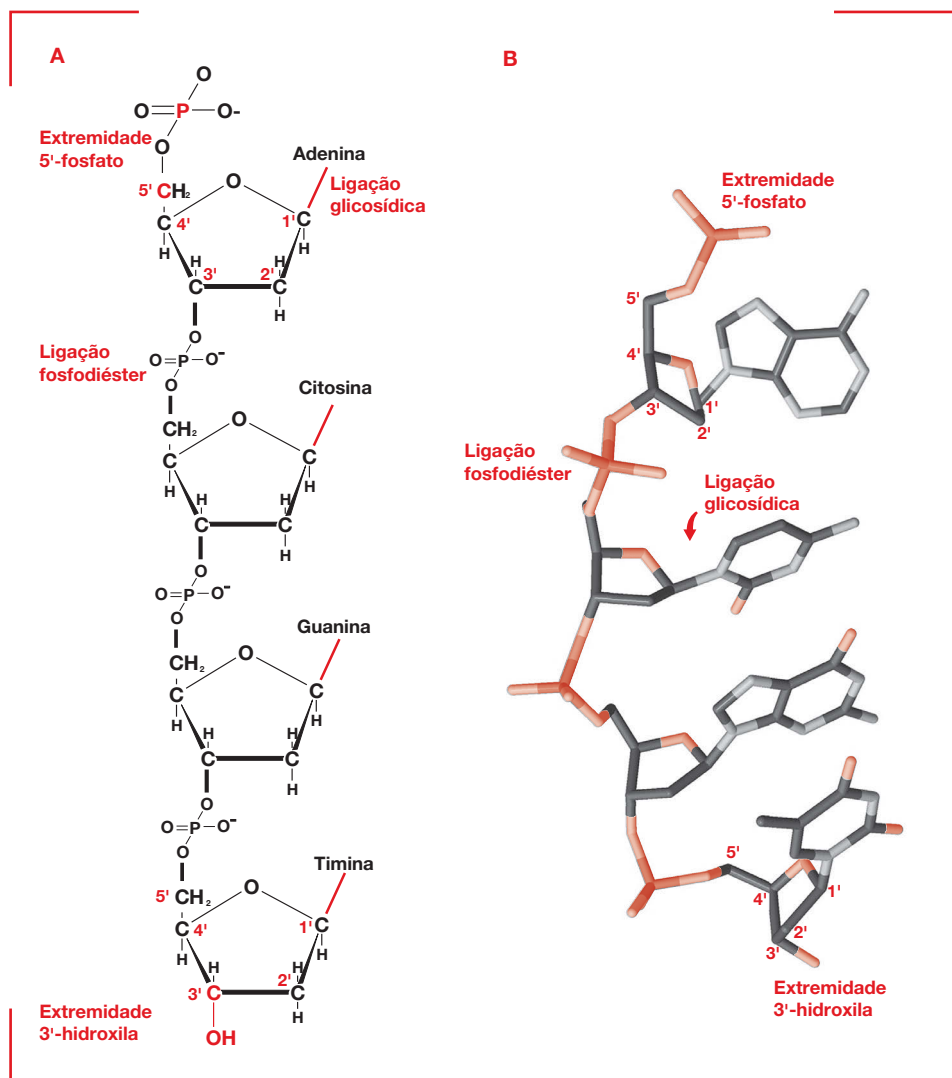


Tabela 2.1 Nomenclatura dos ácidos nucleicos

Bases	Nucleosídeo	Nucleotídeo	Ácido nucleico
Purinas			
Adenina	Adenosina Desoxiadenosina	Adenilato Desoxiadenilato	RNA DNA
Guanina	Guanosina Desoxiguanosina	Guanilato Desoxiguanilato	RNA DNA
Pirimidinas			
Citosina	Citidina Desoxicitidina	Citidilato Desoxicitidilato	RNA DNA
Timina	Timidina	Timidilato	DNA
Uracila	Uridina	Uridilato	RNA

**Figura 2.3**

Ligações químicas na cadeia de DNA. (A) Representação planar do DNA (fita simples), (B) representação tridimensional das ligações no DNA (fita simples). Observe as extremidades 5'-fosfato e 3'-OH, que conferem direcionalidade à cadeia; as ligações fosfodiéster entre os nucleotídeos e as ligações glicosídicas entre os açúcares e as bases nitrogenadas. A representação em (B) é a forma tridimensional obtida com a cristalografia por raios X, que é a forma espacial real da molécula. As bases nitrogenadas ficam perpendiculares e voltadas para o interior da cadeia fosfodiéster.

a manutenção da hélice dupla. Os anéis aromáticos das bases nitrogenadas são hidrofóbicos e ficam orientados para o interior da hélice dupla, quase perpendiculares ao seu eixo, como se fossem os degraus de uma escada. Cada base nitrogenada de uma das cadeias está pareada com a base complementar na outra cadeia de DNA (**Figura 2.4**). Este pareamento ocorre pela formação de **pontos de hidrogênio** entre as duas fitas, mantendo a estrutura da molécula.

Um átomo de hidrogênio, em geral, forma uma ligação covalente com apenas outro átomo. Entretanto, se esta ligação ocorre com um átomo que é um doador eletronegativo, o hidrogênio pode formar associações fracas adicionais, como as pontes de hidrogênio, com outro átomo aceptor. A presença de grupamentos **ceto** ($C=O$) e **amínico** ($C-NH_2$) (**Figura 2.5**) permite a formação de pontes de hidrogênio entre as bases. Dessa forma, **T** e **U**, que contêm um grupamento ceto, podem parar com **A**, que contém um grupamento amínico,

por meio de uma ponte de hidrogênio. **C** e **G**, que contêm um grupamento ceto e um grupamento amínico, podem formar duas pontes de hidrogênio. Além disso, uma ponte de hidrogênio adicional pode ser formada entre os nitrogênios dos anéis aromáticos em todos os pares. Assim, entre **T** ou **U** e **A** são formadas **duas** pontes de hidrogênio, e entre **C** e **G** são formadas **três** pontes de hidrogênio (**Figura 2.5**). Isto confere à molécula de DNA a complementaridade, ou seja, sempre que existir uma **A** em uma fita haverá uma **T** pareada, e quando houver uma **C** na outra fita haverá uma **G**. Esta é uma propriedade fundamental do DNA e a base para a replicação (Capítulo 6), para a transcrição (Capítulo 10) e para outros processos celulares.

As bases nitrogenadas possuem dois tamanhos, sendo as pirimídicas menores do que as púricas (ver **Figura 2.2**). O pareamento sempre ocorre entre uma base púrica e uma pirimídica e, portanto, os pares **AT** e **CG** têm aproximadamente o mesmo tamanho e di-

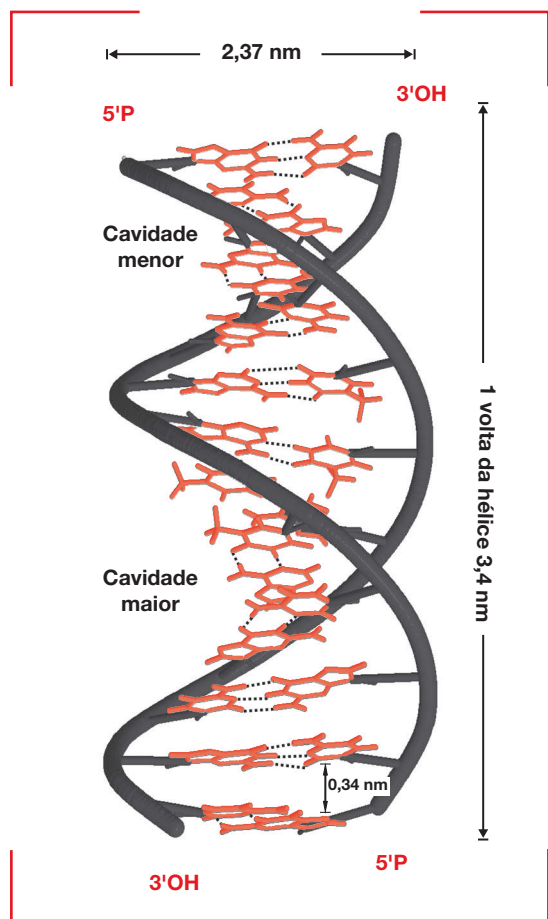


Figura 2.4

Modelo da hélice dupla do DNA proposta por Watson e Crick (DNA tipo B). As ligações fosfodiéster da hélice foram salientadas com o traço de cor roxa para reforçar a visão de enrolamento da hélice. Em rosa-claro estão representadas as desoxirribose. As bases nitrogenadas, voltadas para o interior da hélice, estão representadas na forma de varetas, e as pontes de hidrogênio estão representadas por linhas vermelhas tracejadas. Observe que as bases nitrogenadas são perpendiculares ao eixo da hélice (veja na Figura 2.7 que, se a molécula for vista de seu eixo longitudinal da hélice, as bases todas voltadas para dentro da hélice e “empilhadas” mantêm-se por forças de Van der Waals). As setas pretas indicam distâncias em nanômetros. São mostradas as duas cavidades formadas na hélice.

mensões semelhantes (Figura 2.5). Desse modo, os dois pares ocupam o mesmo espaço tridimensional, conferindo uma uniformidade ao longo da molécula de DNA. Por isso, não existe nenhuma restrição quanto à sequência de nucleotídeos na molécula de DNA. Essas

características de pareamento explicam o fato de que, em qualquer sequência de DNA fita dupla, a relação molar entre A/T é igual a 1,0, ocorrendo o mesmo com a relação C/G, embora as concentrações molares entre AT e CG variem com a sequência de DNA analisada. As pontes de hidrogênio (Figura 2.5) **CG** são mais fortes (necessitam de mais energia para serem rompidas) que as ligações **AT**.

Além dos pareamentos **AT** e **CG**, outros pareamentos raros podem ser encontrados. O pareamento de Hoogsteen, que ocorre em fitas triplas de RNA e DNA e os pareamentos GA, CA e TG são exemplos desses raros eventos.

As ligações glicosídicas no DNA, entre as desoxirribose e as bases nitrogenadas, não estão diretamente opostas na hélice dupla, gerando duas cavidades desiguais em seu contorno (ver Figura 2.4). As duas cavidades são denominadas **cavidade maior** e **cavidade menor**. Nessas cavidades, principalmente na maior, as bases, estão expostas ao meio solvente e são quimicamente distinguíveis. Assim, moléculas (em geral proteínas), que interagem com sequências específicas de bases, podem identificar essas sequências sem romper a estrutura da hélice dupla.

Em princípio, as fortes cargas negativas dos grupos fosfato das cadeias fosfodiéster das duas fitas de DNA tendem à repulsão. Entretanto, um conjunto de forças age para estabilizar a estrutura da hélice dupla do DNA. É fundamental que essas forças sejam fortes o suficiente para manter a integridade da hélice dupla, mas devem permitir a flexibilidade conformacional, que é essencial para a sua atividade. Como veremos, durante a replicação e a transcrição do DNA (Capítulos 6 e 10), a hélice dupla deve ser separada para que estes processos vitais ocorram.

Além das ligações covalentes, que unem os átomos nas moléculas, outras forças mais fracas atuam, estabilizando a estrutura de hélice dupla do DNA. Estas forças são:

1. Os **efeitos hidrofóbicos**, que estabilizam o pareamento entre as bases. Os anéis das purinas e das pirimidinas, que estão voltados para o interior da hélice dupla, são mantidos por coesão interna de moléculas de água, e os sítios hidrofílicos das bases ficam expostos ao solvente nas cavidades.
2. O **empilhamento das bases** no interior da hélice dupla (observe em uma visão do eixo da hélice, Figura 2.4) permite o estabelecimento de forças de Van der Waals entre os anéis aromáticos de bases adjacentes. Estas forças são fracas, mas aditivas na manutenção da estrutura final.
3. As **cadeias de açúcar-fosfato**, que são carregadas negativamente, interagem com cátions (principalmente Mg^{+2}) em solução, neutralizando a repulsão entre as duas cadeias e estabilizando a hélice dupla.

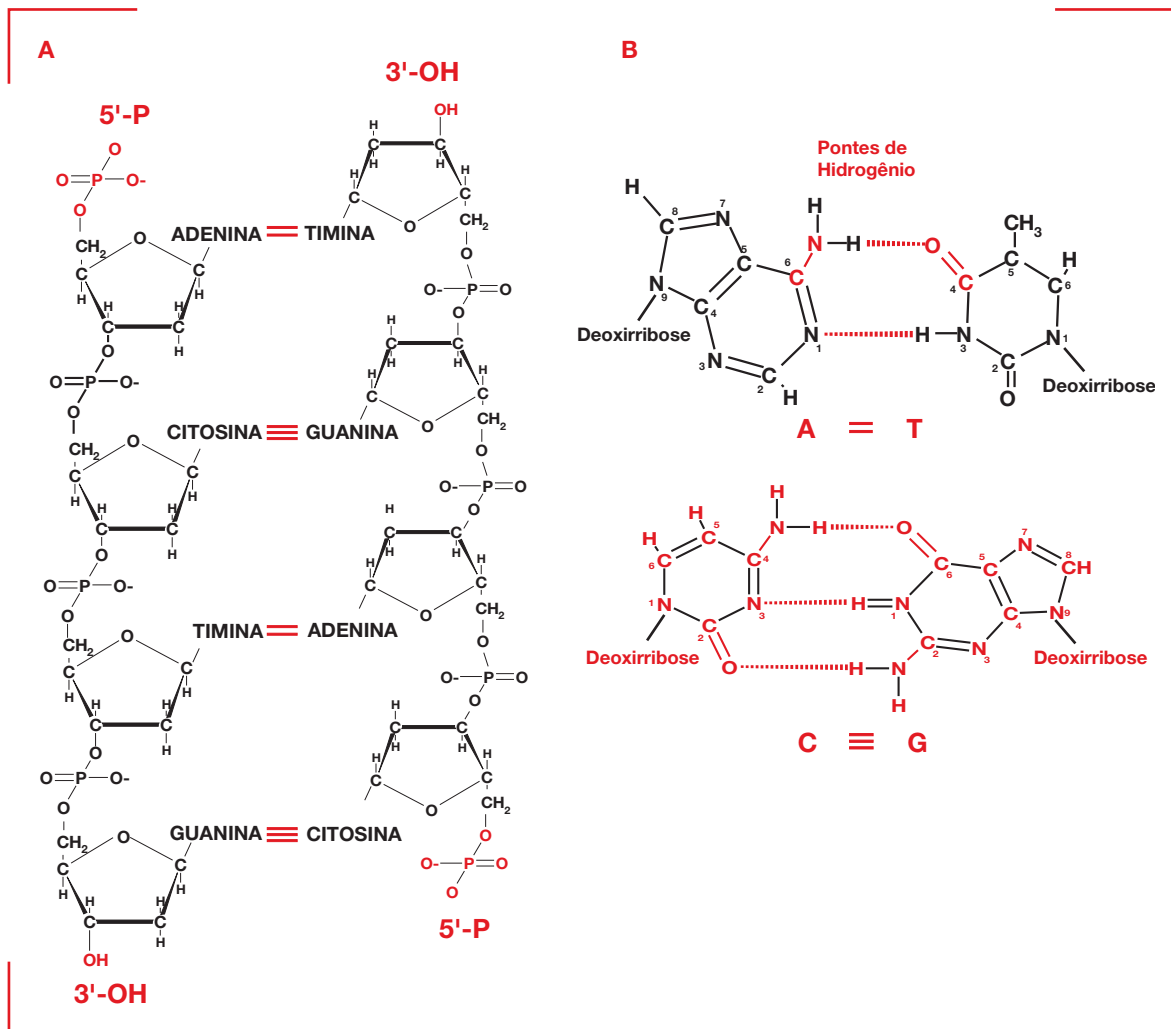


Figura 2.5

Cadeia dupla e pontes de hidrogênio na molécula de DNA. (A) Representa a forma planar das duas cadeias (fitas) de DNA, mostrando a formação das pontes de hidrogênio, (=) representa duas pontes de hidrogênio, (≡) representa três pontes de hidrogênio, 5'-P (extremidades 5' fosfato), 3'-OH (extremidades 3' hidroxila). (B) Representa a forma planar dos pareamentos entre adenina e timina e entre citosina e guanina. As pontes de hidrogênio estão representadas em vermelho.

Na célula, a molécula de DNA quase sempre está ligada a proteínas estruturais, como as histonas, ou a proteínas com diferentes funções. Essas interações com as proteínas também contribuem, de forma majoritária, para manutenção e alterações na estrutura do DNA. A formação da cromatina é o principal tipo de estruturação das moléculas de DNA na célula, e sua estrutura está apresentada no Capítulo 3. Embora o DNA seja visto, na maioria das vezes, como uma cadeia dupla, o DNA de cadeia única ou de fita simples é encontrado na natureza representando genomas de vírus e em alguns plasmídeos bacterianos.

1.2.2 Propriedades químicas

As propriedades químicas do DNA mais importantes para as funções de armazenamento da informação genética

são a **complementaridade** entre as bases nitrogenadas das duas cadeias, o **antiparalelismo**, que confere direcionalidade às cadeias, e a capacidade de **desnaturação** e **renaturação** da hélice dupla. Na seção anterior, foram analisadas as duas primeiras propriedades e, agora, será descrita a propriedade de desnaturação/renaturação.

A desnaturação (ou fusão) ocorre quando as pontes de hidrogênio entre as cadeias complementares do DNA, que formam a hélice dupla, são rompidas e as fitas se separam. O processo inverso, o restabelecimento das pontes de hidrogênio entre as bases complementares (formação da hélice dupla) é a renaturação (ou hibridização ou anelamento). Sempre que a molécula de DNA é replicada, transcrita, ou realiza recombinação (Capítulos 6, 10 e 8, respectivamente) as cadeias de DNA são desnaturadas e

renaturadas. Após a renaturação, todas as propriedades originais da hélice dupla são restabelecidas.

É possível desnaturar o DNA *in vitro*, em solução, por aumento da temperatura, por titulação com ácidos ou álcalis e por agentes desnaturantes, como a formamida, a ureia e o dimetilsulfóxido (DMSO), por exemplo. Os ácidos protonizam os anéis nitrogenados de A, G e C, e os álcalis deprotonizam os anéis nitrogenados de G e T. Esses tratamentos geram grupamentos carregados no interior da hélice dupla, o que leva, também, ao rompimento das pontes de hidrogênio entre as bases complementares. Como as ligações glicosídicas nas purinas são sensíveis em pH ácido, a desnaturação por ácidos tem pouca aplicação prática, pois provoca a **depurinação** das bases, o que altera a molécula de DNA de forma definitiva.

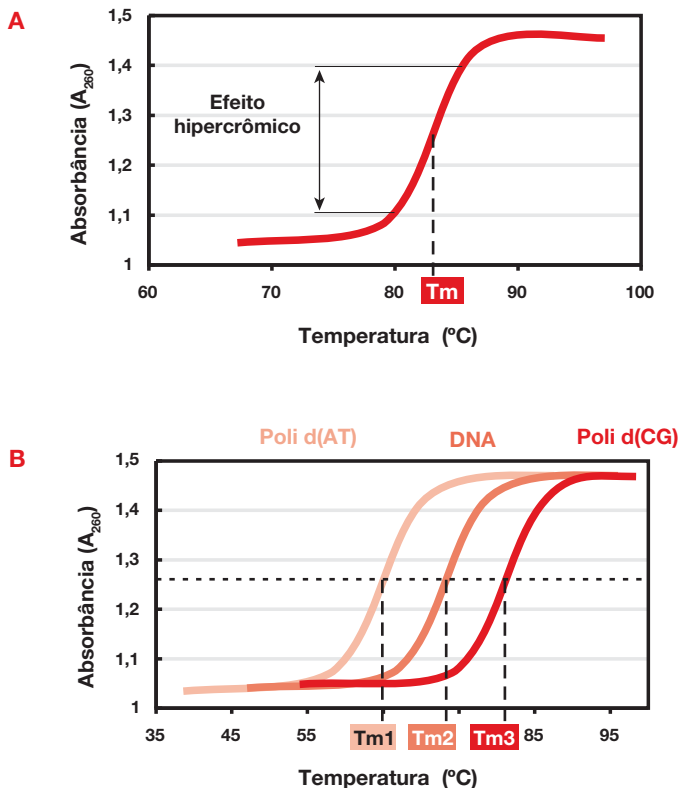
A desnaturação/renaturação do DNA pode ser acompanhada pela medida, em espectrofotômetro, da **absorbância** da luz ultravioleta (UV, **Figura 2.6**). As bases nitrogenadas absorvem luz UV, atingindo o máximo em comprimento de onda de 260 nm (A_{260}). Na molécula de DNA, as bases nitrogenadas são responsáveis pela maior absorção da luz UV, que é máxima quando as fitas de uma hélice dupla de DNA estão completamente separadas, pois as bases estão, em sua totalidade, expostas ao meio. Quando as fitas de uma molécula de DNA estão totalmente separadas, a absorção da luz UV é 37% maior do que aquela obtida com o DNA em sua forma nativa

(hélice dupla). Isso pode ser demonstrado se uma solução de DNA for submetida a um aumento de temperatura, e a absorbância da solução for medida a 260 nm. O gráfico da Figura 2.6 mostra o resultado obtido em tal experimento. A absorbância aumenta com a temperatura até o momento em que as fitas se separam completamente, quando se registra a absorbância máxima (essa alteração brusca chama-se efeito hiperacrômico) e permanece constante mesmo que a temperatura continue aumentando, pois as bases já se encontram expostas e têm absorção máxima da luz UV. A temperatura na qual 50% do DNA está desnaturado é denominada **temperatura média de fusão (T_m)**, Figura 2.6). O efeito inverso – a renaturação – também pode ser acompanhado desta forma. Assim, se esta solução de DNA totalmente desnaturado for resfriada de forma gradativa, as cadeias complementares de DNA voltarão a formar a hélice dupla e a absorbância da luz UV diminuirá, pois as bases nitrogenadas ficarão menos expostas ao meio e, portanto, absorverão menos luz UV. A redução da absorbância cessa quando a hélice dupla estiver completamente renaturada.

Mesmo quando as duas cadeias do DNA estão completamente separadas, o processo pode ser revertido, pois as fitas complementares se reassociam. Esse **anelamento** ou **hibridização** ocorre, em geral, a uma temperatura 25°C abaixo da T_m . No início, essa renaturação ocorre devagar, mas, à medida que algumas bases complementares se associam, a velocidade do processo au-

Figura 2.6

Curvas de absorbância a 260 nm de soluções de DNA submetidas à elevação da temperatura. (A) Temperatura média de fusão (T_m) e efeito hiperacrômico. **(B)** A T_m varia de acordo com a composição GC% ou AT% do DNA ($T_{m3} > T_{m2} > T_{m1}$). AT é uma sequência de DNA apenas com adeninas e timinas (portanto, apenas com duas pontes de hidrogênio entre as fitas complementares). CG é uma sequência de DNA apenas com citosinas e guaninas (portanto, com três pontes de hidrogênio entre as fitas complementares).



menta de forma drástica, tornando este processo muito rápido. Lembre-se de que o pareamento entre as bases nitrogenadas das cadeias do DNA ocorre por complementaridade das sequências e, portanto, é dependente da sequência de bases das fitas, o que torna este processo preciso. Em outras palavras, apenas duas fitas que apresentarem sequências complementares e forem antiparalelas (uma fita na orientação 5'→3' e a outra complementar na orientação 3'→5') irão reassociar-se. Se ocorrer um resfriamento abrupto, as fitas do DNA colapsam e a renaturação pode não ocorrer.

Os pareamentos por pontes de hidrogênio entre as bases no DNA não apresentam a mesma estabilidade em relação aos agentes desnaturantes, devido à presença de duas pontes de hidrogênio entre os pares **AT** e três pontes entre os pares **CG**. Portanto, para o rompimento de um par **CG** são necessárias temperaturas mais elevadas, pH mais alcalino, ou maiores concentrações de agentes desnaturantes do que para a separação de um par **AT**. Assim, a T_m de um DNA depende da proporção de **AT** em relação a **GC** (lembre-se que as moléculas de DNA podem ter diferentes proporções de **AT** e **GC**, dependendo do organismo ou da região do DNA analisada). Quanto maior for a porcentagem de **GC** (índice de **GC** ou **GC%**), maior será a temperatura necessária para desnaturar a molécula de DNA e, portanto, maior será o valor de T_m (Figura 2.6). Na natureza, isso pode ser observado no genoma de alguns organismos. Por exemplo, bactérias que habitam ambientes com temperaturas elevadas tendem a ter um genoma com maior índice **GC** que, potencialmente, teria maior estabilidade em temperatura mais elevada. Outro exemplo são regiões do genoma onde a hélice dupla é separada para o início da replicação ou da transcrição que tendem a ter índice **AT** maior (**GC** menor), facilitando a abertura (separação das cadeias da hélice dupla).

É possível calcular o conteúdo de **GC** ou **AT** de um dado DNA a partir do valor da sua T_m , determinado, de forma experimental, utilizando as equações:

$$T_m (^{\circ}C) = 69,3 + 0,41 (\text{GC}\%) \text{ e } \text{AT}\% = 1 - (\text{GC}\%)$$

Além da medida da absorbância de luz UV, outras técnicas podem ser utilizadas para acompanhar a desnaturação/renaturação do DNA, como, por exemplo: a sedimentação por centrifugação; o tratamento com enzimas específicas para fita simples (nuclease S_1 , por exemplo); e a microscopia eletrônica.

É possível determinar o tamanho de um genoma utilizando medidas da renaturação do DNA. A velocidade de renaturação do genoma de um determinado organismo depende do tamanho desse genoma. A seguinte fórmula define essa relação:

$$\frac{C}{C_0} = \frac{1}{1+k_2 C_0 t}$$

onde:

C é a concentração de fita simples em um tempo t

C_0 é a concentração de fita simples em um tempo zero

k_2 é uma constante t é tempo

Quando a fração de reassociação C/C_0 é igual a 50%, então $C_0 t = 1/k_2$. Esse valor, denominado $(C_0 t)_{1/2}$, é proporcional a N , que é a medida direta do tamanho do DNA analisado.

Assim, o tamanho de um determinado genoma pode ser definido comparando seu $C_0 t_{1/2}$ com o de um DNA padrão com tamanho conhecido. Por exemplo: o cromossomo da bactéria *E. coli* corresponde a $4,2 \times 10^6$ pb. Então, a seguinte fórmula pode ser utilizada:

$$\frac{(C_0 t)_{1/2} \text{ do DNA desconhecido}}{(C_0 t)_{1/2} \text{ do DNA de } E. coli} = \frac{\text{tamanho do genoma}}{4,2 \times 10^6 \text{ pb}}$$

A cinética de renaturação é muito importante para a caracterização do tamanho e da complexidade dos genomas, principalmente de organismos eucarióticos.

O **reanelamento (renaturação)** do DNA também pode ser empregado na identificação de sequências de interesse. Esse processo, denominado hibridização, utiliza a propriedade dos ácidos nucleicos de fita simples de parearem (formarem pontes de hidrogênio) com fitas complementares, mesmo na presença de grandes quantidades de DNA não complementar. Esse tipo de metodologia tem uma vasta aplicação prática e será discutida no Capítulo 16.

1.3 Supertorção

Além da estrutura secundária da hélice dupla, o DNA assume uma estrutura terciária, denominada supertorcida, superenrolada ou super-helicoidal (**Figura 2.7**). Essa outra estrutura é basicamente o enrolamento da hélice dupla sobre si mesma. Lembre-se que a molécula de DNA é muito longa. A estrutura supertorcida é a conformação predominante do DNA na célula e é fundamental para o empacotamento do DNA nos genomas e na estrutura dos nucleossomos (Capítulo 3). Além disso, o superenrolamento está envolvido na replicação, na transcrição e na recombinação, controlando a expressão de alguns genes.

Os métodos comuns de isolamento de DNA, em geral, alteram a estrutura tridimensional da molécula e, quando o superenrolamento foi descrito pela primeira vez, acreditava-se que ele ocorria somente em moléculas circulares pequenas, como genomas virais, plasmídeos, DNA mitocondrial e de cloroplastos. Atualmente, sabe-se que esta é uma característica de quase todos os DNAs circulares e lineares e que é de extrema importância para a sua funcionalidade.

As estruturas de DNA superenrolado são estudadas em um ramo da matemática denominado **topologia**. Para essas considerações, a molécula de DNA de hélice dupla pode ser representada como duas fitas elásticas

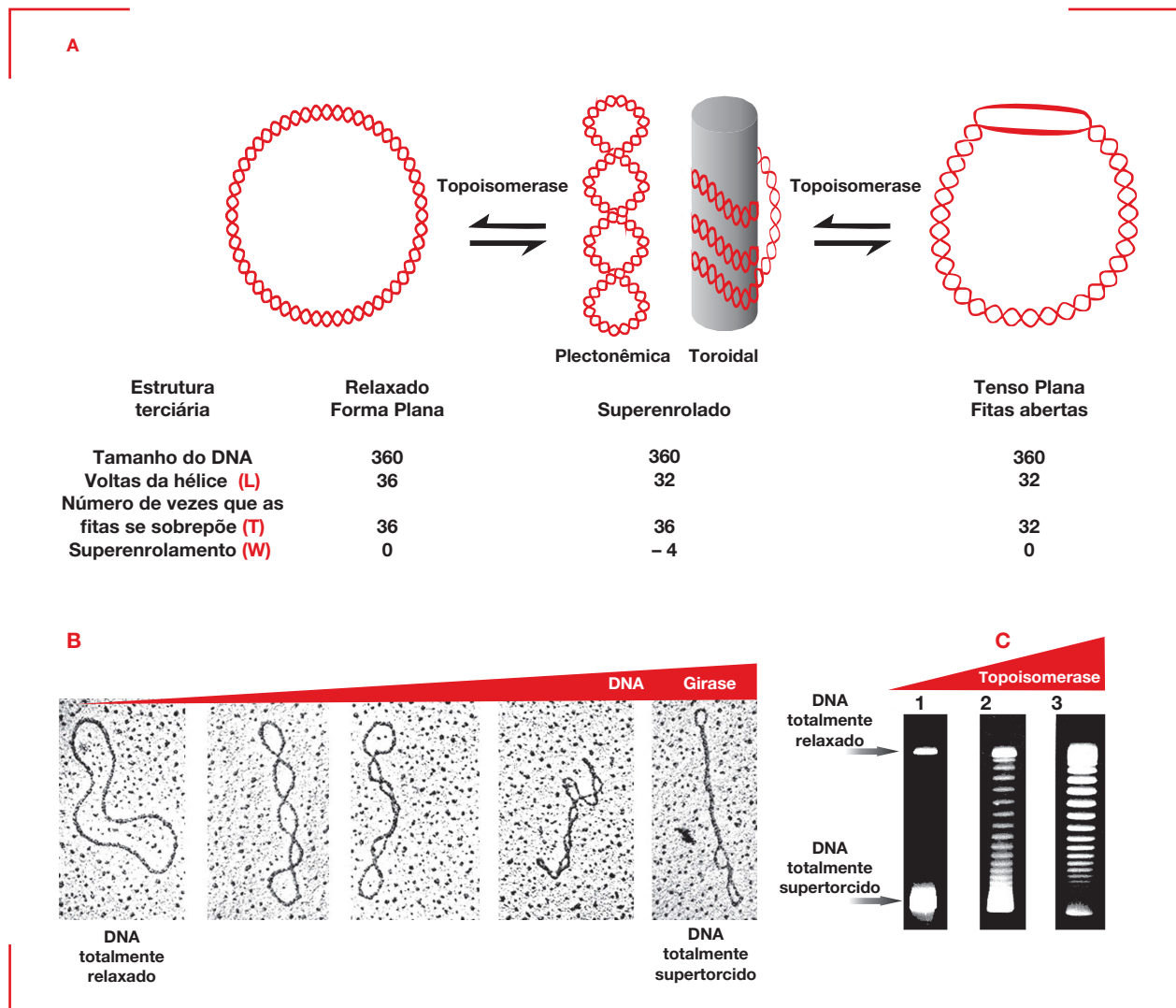


Figura 2.7

Superenrolamento do DNA. (A) Topologia do superenrolamento e definição dos parâmetros. O superenrolamento pode ser representado da forma plectonêmica ou, como ocorre nos nucleossomos, na forma toroidal. Medidas do superenrolamento: (B) Moléculas individuais de DNA observadas por microscopia eletrônica, após tratamento por diferentes tempos, com topoisomerase. (C) Moléculas de DNA fracionadas por eletroforese em gel de agarose, coradas com intercalante fluorescente, que foram tratadas com topoisomerase I em diferentes tempos: (1) tempo zero; (2) 5 min e (3) 30 min. As bandas de DNA intermediárias entre as formas de DNA totalmente relaxado e superenrolado representam diferentes graus de superenrolamento.

que se enrolam, formando a hélice – estrutura secundária. Quando as extremidades das fitas são unidas, uma estrutura circular plana é formada (ver Figura 2.7). Como a orientação das fitas na hélice dupla é antiparalela, elas devem ser corretamente ligadas, ou seja, a extremidade 5'-P é ligada à extremidade 3'-OH. Quando essa estrutura fica perfeitamente assentada em uma superfície plana, sem alterar a geometria da hélice dupla, ela é denominada relaxada (ver Figura 2.7). Se, entretanto, antes de unir as extremidades, uma das fitas for girada sobre si mesma uma ou mais vezes, em voltas de 360° na direção do desenrolamento da hélice dupla, ao passo que a outra fita permanece fixa, a molécula circular formada após a união

das extremidades não estará mais relaxada. Esta nova situação poderá ser acomodada de três maneiras (acompanhe observando as Figuras citadas):

1. A tensão gerada poderá ser distribuída ao longo da hélice dupla, diminuindo o número de vezes que uma fita se enrola na outra. Dessa forma, aumenta o número de bases pareadas por volta da hélice, ou há concentração de tensão em um ponto específico da molécula circular, que ficará não pareada, (fita simples), mantendo o mesmo número de pares de bases por volta da hélice (Figura 2.7). Nos dois casos, a estrutura continuará plana.

2. A tensão produzida poderá ser anulada, se a hélice enrolar-se sobre si mesma (superenrolamento, Figura 2.7). Dessa forma, o número de pares de bases por volta da hélice permanecerá praticamente o mesmo, mas a estrutura não será mais plana, mas superenrolada de forma tridimensional. Quando esse superenrolamento é gerado pelo desenrolamento da hélice dupla, ele é chamado de **superenrolamento negativo**. Esse superenrolamento negativo é encontrado em todos os organismos. As arqueas são exceções a esta regra e têm seu DNA, com um superenrolamento positivo. O superenrolamento pode assumir dois tipos: o **plectonêmico** e o **toroidal** (é o tipo de espiral dos fios de telefone, por exemplo). O primeiro é o encontrado no DNA em solução, e o outro, no DNA enrolado em proteínas como as histonas, que formam os nucleossomos (Capítulo 3).
3. O desenrolamento das hélices pode induzir à formação de estruturas alternativas, como estruturas cruciformes, de triplex ou de DNA do tipo Z. Essas alterações também podem absorver um pouco da tensão criada, gerando um número menor de superenrolamentos.

O superenrolamento pode ser descrito matematicamente como:

$$Lk = Tw + Wr, \text{ onde:}$$

Lk é o **número de ligação**, ou seja, é o número total das vezes em que uma das fitas do DNA passa sobre a outra;

Tw é o **número de voltas** de uma fita em torno da outra, ou seja, é o número de voltas da hélice;

Wr é o **número de superenrolamentos**.

Lk é uma propriedade topológica da molécula dupla circular e permanece inalterado, independentemente do que ocorre com o DNA, desde que as cadeias açúcar-fosfato não sejam rompidas. Esse valor será alterado apenas se as cadeias do DNA forem rompidas e novamente seladas. Existem enzimas que realizam esse processo, denominadas **topoisomerases**. Tw e Wr são parâmetros geométricos e podem variar. Moléculas com o mesmo Lk podem ter diferentes Tw e Wr e, portanto, diferentes conformações tridimensionais (topologia). Moléculas com sequências e tamanhos idênticos, que diferem apenas na sua topologia, são denominadas **topoisômeros**.

Em uma molécula de DNA do tipo B relaxada, o único enrolamento presente é o das voltas da hélice dupla e, portanto, o número de vezes que uma fita passa sobre a outra (Lk) é igual ao número de voltas da hélice (Tw), que, por sua vez, é igual ao número de pares de bases da molécula dividido pelo número de pares de bases em cada volta da hélice (10,5 para B-DNA, ver Figura 2.4). Portanto, $Lk = Tw$ e $Wr = 0$ para a molécula relaxada. Quando uma ou as duas fitas do DNA são rompidas e uma rotação é aplicada a uma delas, sendo as fitas novamente seladas, o valor de Lk é alterado. Em consequência, também os

valores de Tw e Wr serão alterados para acomodar a nova situação (Figura 2.7).

O grau de superenrolamento e a sua geração e remoção em moléculas de DNA podem ser medidos por diversos métodos. A **eletroforese em gel de agarose**, que permite separar as moléculas de DNA em razão da sua massa molecular (tamanho), forma e compactação (as duas últimas são funções diretas do superenrolamento, pois uma molécula superenrolada é mais compacta, Figura 2.7), é uma dessas técnicas (ver Capítulo 16). A **velocidade de sedimentação** é outro exemplo, pois as moléculas de DNA com diferentes graus de enrolamento sedimentam com velocidades diferentes e, portanto, podem ser separadas por centrifugação em gradientes. Finalmente, o superenrolamento também pode ser visualizado por **microscopia eletrônica** ou de **força atômica**, pela observação direta de moléculas com diferentes graus de superenrolamento.

O grau de superenrolamento do DNA na célula é fundamental para a sua funcionalidade. O DNA é mantido em uma condição de superenrolamento negativo (lembre-se que apenas em arqueas o superenrolamento é positivo), que fornece energia para a desnaturação da hélice dupla permitindo o acesso de enzimas, que promovem a replicação, a transcrição, a reparação e a recombinação do DNA. O conhecimento atual mostra que esse superenrolamento é mantido pela célula em um sistema homeostático e que, além da estrutura primária do DNA (sequência de bases), o superenrolamento também armazena informação. Essa informação é importante para a sua funcionalidade. Além disso, o superenrolamento permite que regiões dos genomas que estão muito afastadas umas das outras estejam espacialmente próximas.

1.4 Topoisomerases

Para a manutenção da homeostasia do superenrolamento e da conformação do DNA na célula, existe um sistema complexo de enzimas, as topoisomerases, atuando de forma coordenada. As topoisomerases são enzimas que catalisam a interconversão de topoisômeros de DNA e, portanto, permitem alterações no grau de superenrolamento. Essas enzimas promovem a quebra transitória de ligações fosfodiéster da fita de DNA, gerando uma forma intermediária, em que a proteína permanece ligada covalentemente ao DNA e permite que as fitas do DNA passem umas sobre as outras ou sofram torção, alterando, assim, o superenrolamento da molécula. Desse modo, essas enzimas introduzem ou removem superenrolamentos no DNA e resolvem estruturas, como os nós e os concatêmeros, geradas nos processos celulares (**Figura 2.8**). Todas as células procarióticas e eucarióticas estudadas possuem topoisomerases, que são fundamentais nos processos de replicação do DNA (Capítulo 6), na transcrição (Capítulo 10), na recombinação (Capítulo 8) e no remodelamento da cromatina (Capítulos 3 e 14).

A evolução desenvolveu duas estratégias gerais para a alteração da topologia do DNA pelas topoisomerases.

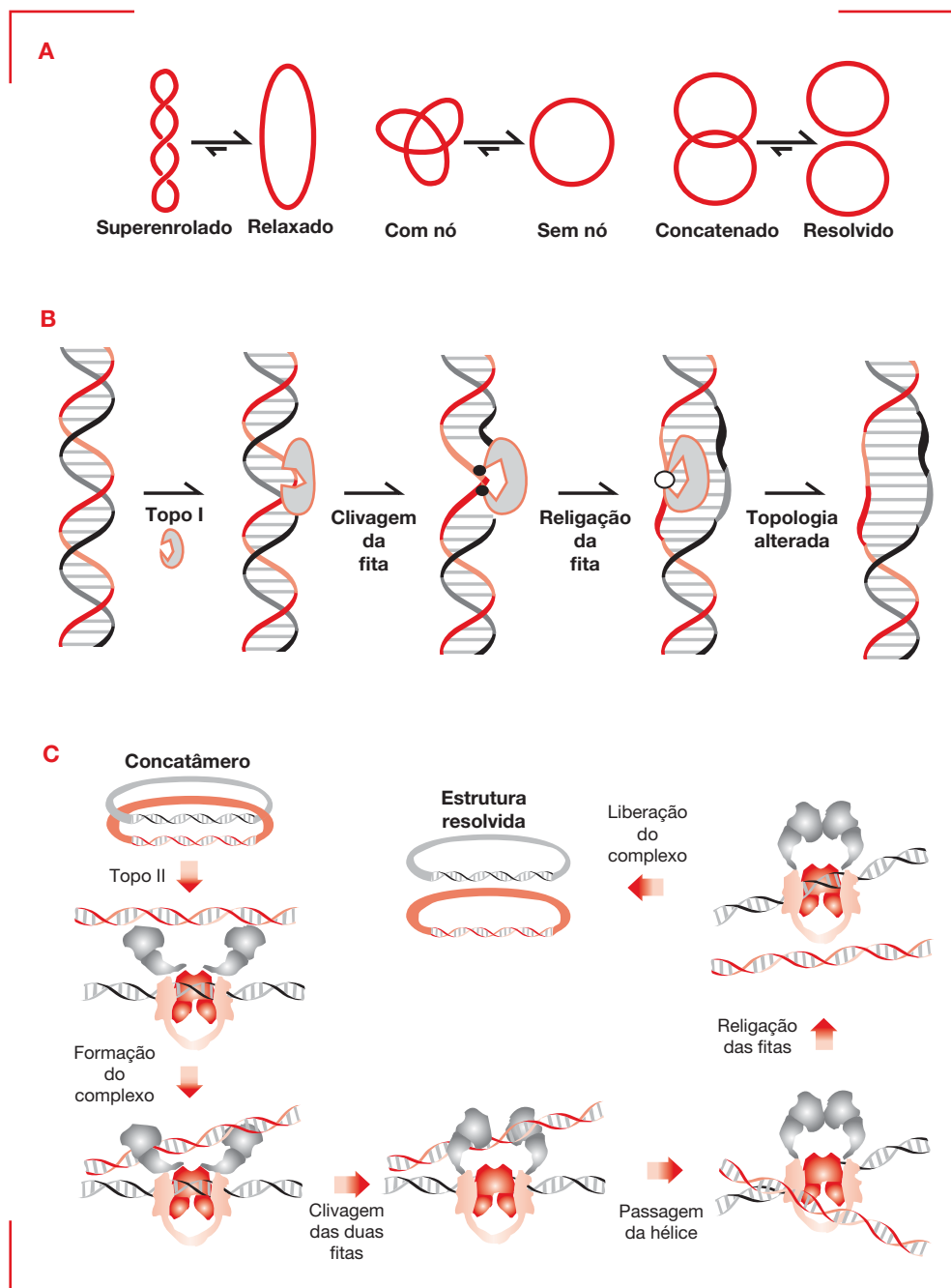


Figura 2.8

Topologia do DNA e topoisomerases. (A) As estruturas topológicas do DNA mais comuns encontradas na célula. (B) Representação do mecanismo de ação das topoisomerases da família Tipo I. A forma em cinza representa um complexo de topoisomerase. Apenas uma das fitas do DNA é clivada, sofre rotação, alterando o superenrolamento, e ocorre a religação da cadeia fosfodiéster. (C) Representação do mecanismo de ação das topoisomerases da família Tipo II. O substrato inicial é um concatâmero, onde duas moléculas de DNA estão entrelaçadas. Para efeito de visualização, apenas uma parte da molécula é apresentada como hélice dupla. O complexo Topo II se liga a uma das moléculas de DNA, depois, interage com a outra molécula de DNA. Com as moléculas ligadas ao complexo, uma das hélices é clivada nas duas fitas. A outra molécula de DNA é passada por dentro das fitas clivadas, alterando o superenrolamento. O DNA não clivado é liberado do complexo. A outra hélice é religada e liberada. O produto da reação é o concatâmero resolvido. As moléculas, antes entrelaçadas, agora estão livres.

Uma delas é o **mecanismo de rotação**, em que a topoisomerase cliva a ligação fosfodiéster de uma das cadeias do DNA e permite que uma das extremidades da cadeia aberta gire em torno da hélice. Em seguida, a cadeia clivada é religada e a molécula de DNA resultante fica com um superenrolamento diferente do inicial (topoisomerasas da família tipo I). A outra estratégia envolve a clivagem de uma ou das duas cadeias de DNA da hélice dupla, causando o rompimento de algumas pontes de hidrogênio entre bases adjacentes, seguida da passagem de uma segunda hélice dupla pela região clivada (topoisomerasas da família tipo II). Para facilitar a visualização desta segunda estratégia, acompanhe a Figura 2.8 e lembre-se de que o superenrolamento é uma característica espacial.

O entendimento da importância da manutenção do superenrolamento do DNA para a funcionalidade das células levou a estudos detalhados das topoisomerasas desde a sua descoberta, em 1970. Esses estudos foram também impulsionados pela descoberta de que estas enzimas são alvos importantes para a ação de muitas substâncias com aplicação em medicina. Assim, atualmente, são conhecidas e caracterizadas muitas topoisomerasas em diversos organismos.

De forma geral, existem muitas subfamílias de cada uma das duas famílias de topoisomerasas, classificadas a partir da sequência de aminoácidos e do mecanismo envolvido na reação de formação dos topoisômeros. As topoisomerasas da família tipo I estão divididas em três subfamílias: IA, IB e IC. As da família tipo II estão divididas em duas subfamílias: IIA e IIB. Em todos os organismos estudados é encontrada pelo menos uma topoisomerase de cada família. As topoisomerasas de cada organismo receberam nomes como: topoisomerase I, girase, topoisomerase IV, etc., o que pode criar alguma confusão. A **Tabela 2.2** mostra uma classificação resumida das topoisomerasas, em que organismos são encontradas cada uma das subfamílias e algumas de suas atividades.

1.5 Tipos de DNA

Os tipos de DNA encontrados em condições fisiológicas são: DNA tipo B, DNA tipo A e DNA tipo Z (**Figura 2.9**).

Esses tipos foram identificados por análises da estrutura cristalográfica e são gerados pela composição de bases, pelo meio em que se encontram ou devido a sua ligação com determinadas proteínas. Nestes três tipos de DNA, as cadeias formam a hélice dupla como visto até aqui. As regras de pareamento se aplicam e as alterações na conformação não alteram a informação contida na sequência de bases. Entretanto, alterações na conformação podem facilitar ou dificultar a interação do DNA com as proteínas.

O que diferencia estes três tipos de DNA é a sua espessura, o número de pares de base por volta da hélice e a exposição das bases nitrogenadas ao meio externo. O DNA tipo B é o fisiológico e o majoritariamente encontrado na célula. O DNA tipo A ocorre em condições de umidade muito baixa e quando fitas mistas de DNA-RNA e fitas duplas de RNA são formadas. Híbridos DNA-RNA são formados na transcrição (Capítulo 10) e fitas duplas de RNA são encontradas em genomas de alguns vírus e, na célula, nos RNAs transportadores e ribossômicos. O DNA tipo Z ocorre em regiões curtas do DNA, onde existem sequências de citosinas e guaninas seguidas.

O DNA tipo B apresenta a forma de hélice dupla clássica, descrita por Watson e Crick (ver Figura 2.4). Esse é o tipo de DNA que ocorre na presença de umidade relativa elevada (92%) e em soluções de baixa força iônica. Ele apresenta as seguintes características:

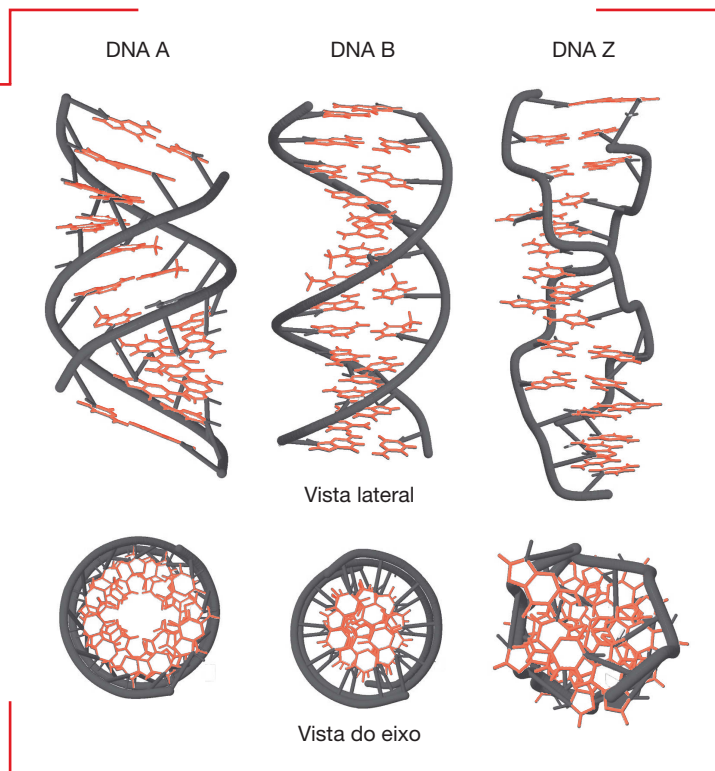
1. A hélice dupla gira para a direita e a rotação entre os dois pares de bases adjacentes é de $34,6^\circ$. Dessa forma, a hélice dupla perfaz uma volta completa a cada 10,4 pb.

Tabela 2.2 Topoisomerasas: classificação resumida

Família	Subfamília	Organismo/enzima	Atividade
Tipo I, Cliva uma das fitas do DNA	IA	Bactéria e arqueas/topo I e III Eucariotos/topo III Mimivírus/topo IA	Desfaz nós e concatâmeros Relaxa superenrolamento (-) Girase reversa de bactérias hipertermófilas, induz superenrolamento (+)
	IB	Eucariotos/topo I Vírus e algumas bactérias	Relaxa superenrolamento (-) e (+)
	IC	Somente um gênero de arqueas/topo V	Relaxa superenrolamento (-) e (+)
Tipo II Cliva as duas fitas do DNA	IIA	Eucariotos, vírus e bacteriófagos/ topo II Bactérias/topo IV Bactérias, arqueas e cloroplasto, DNA girase	Relaxa superenrolamento (-) e (+). Desfaz nós e concatâmeros Relaxa superenrolamento (+). Induz superenrolamento (-) Desfaz nós e concatâmeros
	IIB	Arqueas, plantas e algumas algas/ topo IV	Relaxa superenrolamento (-) e (+) Desfaz nós e concatâmeros

Figura 2.9

Tipos mais comuns de DNA (A, B e Z). Representação da estrutura tridimensional das moléculas, vista lateral. Foi traçada uma fita ligando os grupos fosfatos para evidenciar o enrolamento das cadeias em torno do eixo da hélice. As bases nitrogenadas estão representadas em forma de varetas e em vermelho. Observe as alterações no diâmetro das hélices e a alteração nas cavidades maior e menor. Abaixo de cada hélice está a representação transversal das moléculas, mostrando a posição das bases nitrogenadas em relação ao eixo das hélices.



2. Uma volta da hélice percorre uma distância de 3,40 nm, que é o passo da hélice. Cada par de bases adiciona à hélice uma distância de 0,33 nm.
3. O diâmetro da hélice dupla é de 2,37 nm.

Se o DNA tipo B em solução for desidratado, por exemplo, em 75% de etanol, ou se o conteúdo de sal for diminuído, a molécula, que é fina e comprida, se encurtará e engrossará gerando o DNA tipo A (Figura 2.9). A cavidade maior se tornará mais estreita e mais profunda, e a cavidade menor mais larga e rasa. O passo da hélice reduz-se de 3,40 nm para 2,60 nm por volta, e o número de pares de bases por volta aumenta para 11.

Tanto no DNA tipo B quanto no tipo A, a desoxirribose e a base nitrogenada estão em lados opostos da ligação glicosídica, ou seja, na conformação *anti*. Nessa situação, a repulsão estérica entre a base e o açúcar está minimizada. Na presença de altas concentrações de cátions, alguns nucleotídeos assumem a conformação *syn*, ficando o açúcar e a base do mesmo lado da ligação glicosídica. Nessas condições, uma estrutura diferente, o DNA tipo Z, é formada (Figura 2.9). A hélice dupla neste tipo de DNA está enrolada para a esquerda e alterna as conformações *syn* e *anti*, sendo comum em sequências onde os nucleotídeos G e C se alternam. As pirimidinas conservam a conformação padrão *anti* com o açúcar na forma C2'-*endo* e as purinas estão em *syn* e C3'-*endo*. A hélice dupla gira para a esquerda nessas condições. Devido à aparência em ziguezague da cadeia, essa forma, descrita em oligonucleotídeos sintéticos, foi denominada Z.

O DNA tipo Z é mais longo e fino que o tipo B. Uma volta completa da hélice ocorre a cada 12 pb, e o passo da hélice aumenta para 4,56 nm, com um diâmetro de 1,84 nm. A cavidade maior desaparece e a menor torna-se muito profunda, formando uma espiral em torno da estrutura. Algumas regiões do DNA *in vivo* podem estar nessa forma. Alguns fatores estabilizam a formação do DNA tipo Z, como: a metilação ou a bromatação das bases e o estresse torcional, como ocorre em regiões do DNA superenrolado negativamente ou quando determinadas proteínas estão ligadas ao DNA. Dois fatores determinam a possibilidade da ocorrência de regiões de DNA tipo Z: a sequência de nucleotídeos e a estrutura global da hélice dupla. Se tais fatores forem propícios, o DNA pode converter-se do tipo B para o tipo Z em condições naturais. Isto parece ocorrer em regiões do genoma que estão sendo ativamente transcritas. O DNA tipo Z pode ter importância na regulação da expressão gênica e, também, estar envolvido em eventos de processamento de DNA e em instabilidade genética. Existem proteínas que se ligam especificamente a DNA tipo Z. Entretanto, ainda não são bem conhecidas as funções deste tipo de DNA na fisiologia celular.

Embora a hélice dupla seja a forma predominante do DNA na natureza, moléculas de DNA de fita simples compõem os genomas de alguns vírus de animais e vegetais, bem como os de alguns bacteriófagos, que são vírus bacterianos. Alguns bacteriófagos que infectam a bactéria *E. coli*, por exemplo, os bacteriófagos M13 e ϕ X174, apresentam DNA de fita simples como genoma. Esse DNA produz uma forma replicativa (DNA de fita dupla), que é

o molde para a síntese de DNA dos bacteriófagos replicados no hospedeiro. Genomas de procariotos e eucariotos são discutidos no Capítulos 4 e 5, respectivamente.

1.6 Outras estruturas do DNA

1.6.1 Curvatura

Longas cadeias de DNA são muito flexíveis, entretanto, sequências menores que 100 pb são relativamente rígidas para serem curvadas. Algumas sequências específicas e proteínas especiais tornam determinadas regiões do DNA mais suscetíveis à curvatura. Essas regiões mais flexíveis à curvatura são importantes em sequências envolvidas com o controle de processos como a replicação, a transcrição e a recombinação.

Sequências de DNA constituídas por regiões de adeninas têm sua curvatura facilitada, principalmente se tais sequências são intercaladas e se repetem várias vezes. A ligação de proteínas também pode facilitar a curvatura. Um exemplo clássico é o nucleossomo, no qual um segmento de DNA de 145 pb se enrola quase duas vezes em torno do octâmero de histonas (Capítulo 3). As cavidades em contato com a proteína devem ficar mais estreitas, e as cavidades opostas mais largas.

Muitas outras proteínas, além das histonas, promovem curvatura no DNA: fatores de transcrição, enzimas de recombinação, enzimas que modificam o DNA, entre outras. Essas proteínas curvam o DNA com diferentes magnitudes. Existem diferentes funções para a curvatura do DNA, como:

1. condensar e empacotar DNA, como nos nucleossomos;
2. expor a sequência de bases internalizada na hélice dupla;
3. aproximar sítios de ligação distantes no DNA linear;
4. formar estruturas especiais proteína-DNA para alinhar determinados sítios, como no caso de recombinação sítio-específica;
5. forçar a molécula de DNA para estimular a clivagem ou a desnaturação.

1.6.2 Estruturas cruciformes

Existem sequências no DNA que apresentam simetria, contendo regiões repetidas e invertidas. Por exemplo:

ATTTCGCGTAGTAGACATAGCTGACATAGTCAG-
CTATGTCTGCTC

Sequências como essas podem formar estruturas onde o pareamento entre as duas fitas é substituído pelo pareamento entre as bases complementares na mesma fita.

ATTTCGCGTAGTAGACATAGCTGACA
:::.....:T
CTCGTCTGTATCGACTGA

Essa alteração reduz o número de voltas da hélice dupla na região, removendo o grau de superenrolamento negativo, em uma unidade para cada dez bases de pareamento. Essas estruturas ocorrem *in vivo* e são fundamentais para determinados processos.

1.6.3 Junções de Holliday

As junções de Holliday são estruturas formadas por quatro fitas de DNA, sendo intermediárias importantes na recombinação genética (Capítulo 8). Proteínas e enzimas específicas reconhecem essas estruturas e determinam como elas serão clivadas, originando os produtos da recombinação.

2. RNA

Uma das moléculas que gerou uma grande quantidade de novos conhecimentos em biologia molecular na última década foi o RNA. Descoberto na mesma época em que o DNA, os estudos com o RNA tiveram, a princípio, um desenvolvimento mais lento. O número de classes diferentes de RNA e o seu envolvimento em inúmeros processos na célula, antes desconhecidos, foram, os principais responsáveis pela demora na sua caracterização e na descoberta das suas funções. Atualmente, o número de funções biológicas em que os RNAs estão envolvidos ainda está aumentando.

Há pouco mais de dez anos, aos RNAs eram atribuídas funções, não sem importância, primordialmente na síntese de proteínas. O **RNA mensageiro (mRNA)** era visto como uma molécula intermediária na transferência da informação contida no genoma para a síntese das proteínas. A sua conformação parecia ser aleatória e sem importância. A conformação do **RNA ribossômico (rRNA)**, gerada por suas estruturas secundária e terciária, parecia ser um mero arcabouço para interações com as proteínas ribossômicas, com a função de catalisar a síntese das proteínas. Os **RNAs transportadores (tRNAs)** eram mais estudados em termos de estrutura terciária, mas com vistas a sua capacidade de carregar os aminoácidos e interagir com os ribossomos.

Atualmente, os RNAs e a sua conformação espacial são bastante estudados. Várias novas classes de RNA foram descobertas e estão envolvidas no controle da expressão gênica (Capítulo 14). Foram também descobertas novas possibilidades de interações dos RNAs com as proteínas, com o DNA e com outros RNAs, a partir de uma diversidade de conformações espaciais. Quanto às possíveis atividades bioquímicas dos RNAs, a descoberta de que eles podem catalisar reações levou à formulação do, atualmente denominado “o mundo do RNA”. Uma molécula de RNA linear pode, em sua sequência de bases, armazenar informação genética (semelhante ao DNA); realizar um pareamento interno, formando uma estrutura secundária, expondo ou escondendo determinadas sequências e, assim, interagir com outras moléculas; pode

assumir uma conformação espacial, estrutura terciária, contendo superfícies que podem interagir com outras moléculas ou regiões internas que possam criar sítios de ligação com metais; e promover catálise, como as enzimas. Esta maleabilidade de funções não é exibida por nenhuma outra macromolécula na célula, permitindo, assim, propor que o RNA tenha sido a **molécula primordial** e que evoluiu para mais funções ainda, como na regulação da expressão gênica, por exemplo. Portanto, o conhecimento da estrutura e da conformação do RNA é ainda mais relevante.

2.1 Composição química

Existem semelhanças entre a estrutura do RNA e a do DNA. Ambos são polímeros lineares de subunidades de nucleotídeos ligadas entre si por ligações fosfodiéster 5'→3' (**Figura 2.10A**). Entretanto, na molécula de RNA, o açúcar presente é a ribose, e a base nitrogenada timina (T) é substituída pela base nitrogenada uracila (U). Portanto, o RNA é composto por adenina, citosina, guanina e uracila. A presença da ribose origina a denominação de ácido ribonucleico. O RNA é predominantemente de cadeia simples, mas forma regiões de cadeia dupla por pareamentos internos. Estas regiões variam muito em sua extensão, dependendo da classe de RNA e da sua função.

O RNA é mais reativo que o DNA. A diferença aparentemente sutil da presença da hidroxila na posição 2' da ribose (**Figura 2.10A**) tem reflexos importantes na estrutura e na reatividade do RNA. Assim, o DNA é estável em presença de álcali, pois, apesar das fitas serem desnaturadas com o rompimento das pontes de hidrogênio, as ligações fosfodiéster nas cadeias permanecem inalteradas. O RNA, na presença de álcali, é hidrolisado, originando uma mistura de nucleosídeos monofosfatados 2' e 3'. Na presença de íons hidroxila, o grupamento 2'-hidroxílico da ribose é convertido no **ânion alcóxi**. Esse é um potente agente nucleofílico que reage com o grupamento 3'-fosfodiéster, formando um **nucleotídeo cíclico**, 2',3'-nucleosídeo monofosfato. Nesse processo, a ligação fosfodiéster 5'→3' é rompida. Em seguida, o nucleotídeo cíclico é submetido a uma segunda hidrólise, em que um íon hidroxila rompe o ciclo fosfodiéster, formando uma mistura de nucleosídeos monofosfatados 2' e 3'.

2.2 Estrutura secundária

Pareamentos entre C e G e entre A e U ocorrem entre regiões complementares da mesma molécula do RNA ou, com menos frequência, entre moléculas distintas. A presença da hidroxila no carbono 2' da ribose, também acarreta uma diferença espacial grande em relação à posição do grupamento fosfato em relação à base nitrogenada. Observe na **Figura 2.1C** a diferença da conformação es-

pacial entre o ATP e o dATP, por exemplo. Na formação da cadeia de RNA, esta diferença também tem seus efeitos. No RNA, ainda, ocorre a formação de hélice dupla, em regiões onde existe pareamento entre duas sequências. Neste caso, a sua estrutura assemelha-se ao DNA tipo A. A estrutura do DNA tipo B não é formada no RNA pareado devido ao impedimento espacial representado pelo grupamento hidroxílico na posição 2' da ribose. Híbridos RNA-DNA são formados em diferentes processos na célula, por exemplo, na transcrição, e sua estrutura é do tipo A. As estruturas secundárias e terciárias do RNA são importantes nas suas funções e no reconhecimento proteína-RNA.

A **Figura 2.10** mostra alguns tipos de estruturas secundárias dos RNAs. Os denominados “grampos” são estruturas muito frequentes, compostas pelo pareamento interno de uma cadeia simples de RNA formando uma alça onde não existe pareamento. A estabilidade destes grampos é muito variável, dependendo da sequência do pareamento e da sequência presente na alça.

2.3 Classes de RNA

Em contraste com o DNA, diferentes tipos de RNA estão presentes nas células e possuem funções específicas. Os RNAs são classificados de acordo com a sua localização e a sua função na célula. Os avanços nos últimos anos, principalmente na área de caracterização de genomas e os métodos de alto desempenho para avaliar a transcrição, levaram à descoberta de várias novas classes de RNAs. Algumas dessas classes ainda têm funções não entendidas em sua totalidade e outras classes provavelmente serão descritas. Portanto, é importante aqui apresentar uma classificação geral dos RNAs descritos até o momento (**Tabela 2.3**).

Os principais tipos de RNA encontrados nas células, envolvidos diretamente na síntese de proteínas são os seguintes:

1. mRNA (RNA mensageiro), que transfere a informação genética do DNA aos ribossomos, local onde ocorre a síntese das proteínas. Representa 1 a 5% do RNA total da célula.
2. rRNA (RNA ribossômico), que é o componente majoritário dos ribossomos. Representa cerca de 75% do RNA total da célula formando fitas duplas por pareamentos internos.
3. tRNA (RNA transportador), que transporta os resíduos de aminoácidos até os ribossomos para a síntese das proteínas. Representa 10 a 15% do RNA total da célula.

A descoberta das outras classes de RNA levou a uma subdivisão mais ampla dos RNAs, em codificadores e não

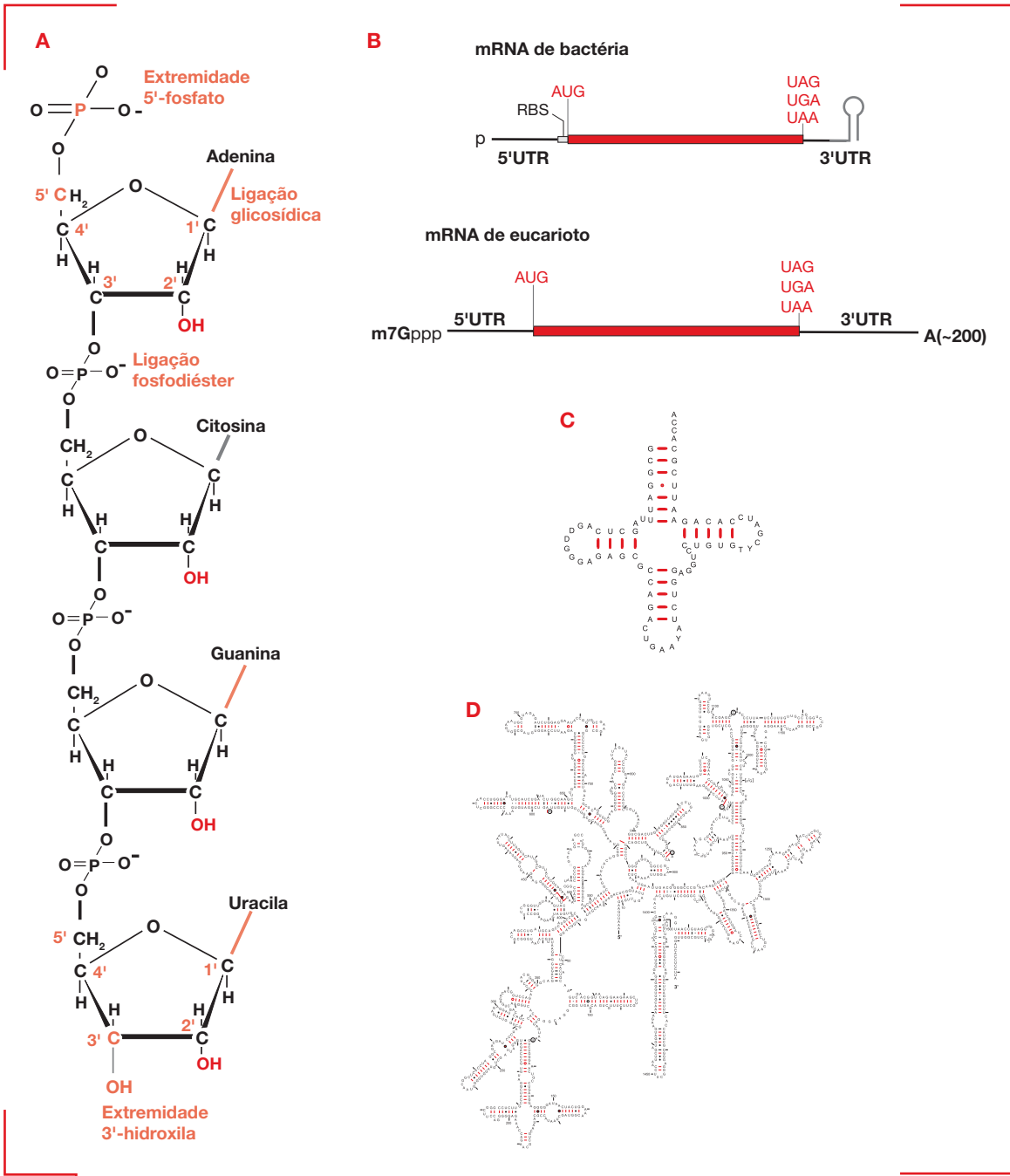


Figura 2.10

Estrutura do RNA. (A) Formação da cadeia do RNA. Note a hidroxila na posição 2' da ribose e a substituição da timina (no DNA) pela uracila no RNA. (B) Organização de mRNA em bactérias e em eucariotos. Em vermelho a região que codifica a proteína com os códons de início e término da tradução. UTR: região não traduzida (de *untranslated sequences*); m7G: metil-7-guanosina (5' CAP); A₍₋₂₀₀₎: cauda de poli(A). (C) Estrutura secundária do tRNA (RNA de transferência). As regiões pareadas são os grampos e a regiões não pareadas são as alças. (D) Estrutura secundária complexa do rRNA 16S de bactéria (RNA ribossômico). Observe que moléculas é uma cadeia única de RNA com regiões internas de pareamento.

Tabela 2.3 Classes de RNAs, processos celulares que participam, distribuição e função^a

Tipo	Processo celular	Classe	Procarioto	Eucarioto	Função
Codificador	Síntese de proteínas	mRNA (mensageiro)	P	P	Contém a sequência para a síntese das proteínas
Não codificador	Síntese de proteínas	rRNA (ribossômico)	P	P	Forma junto com as proteínas ribossômicas os ribossomos.
		tRNA (transportador)	70S (50S+30S)	80S (60S+40S)	Local da síntese das proteínas
	Processamento do rRNA e tRNA	snoRNA (RNA nucleolar pequeno)	P	P	Transporta os aminoácidos ao local da síntese das proteínas
	Processamento do pré-RNA	snRNA (RNA nuclear pequeno)	A	P	Maturação de ribossomos e RNAs transportadores
	Regulação da expressão gênica	sRNA (RNA pequeno)	P	P	Síntese do mRNA eucariótico
		CRISP RNA (RNA CRISP)	Liga-se em proteínas e em RNAm 100 a 300nt	P	Modula a tradução e a estabilidade de mRNAs
		miRNA (micro RNA)	P	A	Interfere com a infecção por bacteriófagos e com a conjugação
		siRNA (RNA de interferência pequeno)	Homologia com bacteriófagos e plasmídeos	P	Silenciamento gênico. Reprime a tradução ou cliva mRNAs alvo
		piRNA (RNA que liga a proteína PIWI)	A	P	Silenciamento gênico. Cliva mRNAs (vírus de retroelementos ou sequências repetitivas)
		lncRNA (RNA não codificador longo)	A	P	Silenciamento gênico em células germinativas
				P	Muito transcrito. "Transcrição generalizada". Existem subclasses
				Núcleo e citoplasma	

^a Classes de RNA descritas e com função reconhecida até 2009.

A = ausente; P = presente.

codificadores. **RNAs codificadores** são exclusivamente aqueles que contêm a informação genética para a síntese das proteínas. São os mRNAs das células procarióticas e eucarióticas. No caso das células eucarióticas, os mRNAs são produzidos, a princípio, como precursores denominados **pré-mRNAs**, que passam por vários processos de modificação para produzir o mRNA final ativo.

Todos os demais RNAs são denominados **não codificadores** e incluem o rRNA e o tRNA. Estas três classes (mRNA, rRNA e tRNA) participam diretamente da síntese de proteínas e são denominados **RNAs de manutenção do metabolismo celular básico (housekeeping)**. Nas células eucarióticas, além destes RNAs de manutenção, as organelas, as mitocôndrias e os cloroplastos têm seus mRNAs, rRNAs e tRNAs com

características semelhantes aos RNAs procarióticos, que realizam a síntese de proteínas a partir dos genomas das organelas.

Nos procariotos, foram descritas mais duas classes de RNAs não codificadores, os **sRNAs** (de *small*) e os **CRISP RNAs** (de *clustered regulatory interspaced short palindromic repeats*). Os sRNAs são pequenos (entre 100 e 300 nt), podem parear com mRNAs específicos, podem ligar-se com proteínas e regular a expressão de determinados genes por alteração da estabilidade do mRNA ou por modulação da tradução. Os CRISP RNAs têm sequências homólogas com genomas de bacteriófagos (vírus de bactérias) e plasmídeos e interferem com a infecção por bacteriófagos e com a conjugação de plasmídeos.

Nos organismos eucarióticos, além dos RNAs não codificadores (rRNA e tRNA) ainda são encontrados os RNAs nucleares pequenos, **snRNAs** (de *small nuclear*), os RNAs citoplasmáticos pequenos, **scRNAs** (de *small cytoplasmic*) e os RNAs nucleolares (do nucléolo) pequenos, **snoRNAs** (de *small nucleolar*). Estes RNAs também estão envolvidos com a manutenção básica do metabolismo da célula. Os snRNAs são encontrados no núcleo, formando complexos com proteínas, as **ribonucleoproteínas (RNPs)**, e participam do processo de remoção dos íntrons. Os scRNAs são encontrados no nucléolo (região do núcleo onde ocorre a síntese do rRNA) e participam do processamento dos ribossomos.

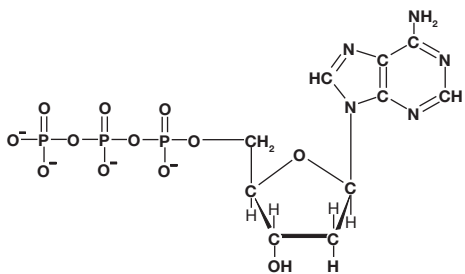
As demais classes de RNA de eucariotos estão envolvidas na regulação da expressão dos genes. São RNAs não codificadores **ncRNA** (de *non-coding*) que são divididos em dois grupos: curtos e longos. Os ncRNAs curtos são: os microRNAs, **miRNAs** (de *micro*), os RNAs de interferência pequenos, **siRNA** (de *small interfering*, algumas vezes somente iRNA) e os **piRNAs**. Esses RNAs ligam-se com duas classes de proteínas: as Argonautas (mi e siRNA) e as proteínas PIWI (piRNA). Os miRNAs e os siRNAs são encontrados em todas as células eucarióticas. Os piRNAs são encontrados apenas em células germinativas de animais. Os ncRNAs longos, **lncRNA** (de *long non-coding*), são encontrados tanto no núcleo como no citoplasma, sendo, em geral, transcritos e tendo funções regulatórias ainda não totalmente elucidadas.



Representação gráfica de moléculas e átomos

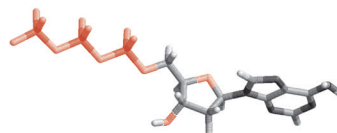
Para representar átomos e moléculas, existem pelo menos três formas:

1. A representação de fórmula elementar: os átomos são representados por letras, e as ligações químicas, por traços.

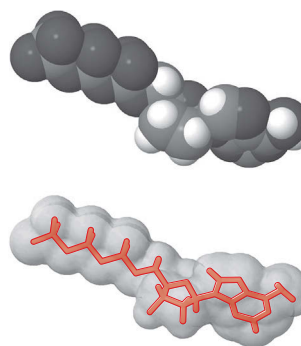


2. A representação em forma de varetas (*stick*): as ligações químicas são representadas como bastões,

estando os átomos nas extremidades e nos vértices representados em diferentes cores.



3. A representação de preenchimento (ou superfícies de Van der Waals): são representadas as superfícies das nuvens eletrônicas, ou seja, uma aproximação do espaço ocupado pelo átomo ou molécula.





Resumo

Os ácidos nucleicos – DNA e RNA – são polímeros de nucleotídeos, desempenham funções vitais na célula e são encontrados em todos os organismos vivos. O DNA tem como função principal armazenar, em sua sequência de bases, a informação genética. A capacidade de autorreplicação do DNA permite que esta informação seja transmitida às gerações seguintes. Moléculas de DNA são polímeros muito longos, e os nucleotídeos componentes contêm a desoxirribose como açúcar e as bases nitrogenadas adenina, timina, citosina e guanina. O DNA forma uma cadeia dupla que tem seu pareamento mantido por pontes de hidrogênio entre bases complementares (AT e CG), formando uma hélice, sendo as cadeias antiparalelas. Os três principais tipos de DNA são: tipo B, que é o predominante, tipo A e tipo Z. Estas cadeias podem ser separadas por ruptura das pontes de hidrogênio, que pode ser provocada pelo aumento da temperatura, por agentes desnaturantes ou *in*

vivo e pelas proteínas ou enzimas. As cadeias podem ser renaturadas, voltando a ter exatamente as mesmas propriedades, em um processo denominado hibridização. O DNA de hélice dupla, em geral, sofre um superenrolamento, que forma a sua estrutura terciária. O superenrolamento é mantido nas células por enzimas denominadas topoisomerases e é fundamental para vários processos. O RNA é composto pela ribose e pelas bases nitrogenadas adenina, uracila, citosina e guanina. O RNA normalmente é de cadeia simples, mas forma muitas estruturas secundárias por pareamentos internos. Existem várias classes de RNAs nas células diretamente envolvidas na síntese de proteínas, no processamento dos RNAs, que normalmente são produzidos como precursores não funcionais, e na regulação da expressão gênica. As classes básicas são o RNA mensageiro (mRNA), o ribossômico (rRNA) e o transportador (tRNA).



Leituras recomendadas

Brosnan CA, Voinnet O. The long and the short of noncoding RNAs. *Curr Opin Cell Biol.* 2009;21(3):416-25.

Carninci P. Molecular biology: the long and short of RNAs. *Nature.* 2009;457:974-75.

Carthew RW, Sontheimer EJ. Origins and mechanisms of miRNAs and siRNAs. *Cell.* 2009;136(4):642-55. Review.

Champoux JJ. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem.* 2001;70:369-413. Review.

Cooper TA, Wan L, Dreyfuss G. RNA and disease. *Cell.* 2009;136(4):777-93. Review.

Draper DE, Biophys J. RNA folding: thermodynamic and molecular descriptions of the roles of ions. 2008;95(12):5489-95.

Ellington AD, Chen X, Robertson M, Syrett A. Evolutionary origins and directed evolution of RNA. *Int J Biochem Cell Biol.* 2009;41(2):254-65. Review.

Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. *Nat Rev Genet.* 2009;10(2):94-108. Review.

Hendrix DK, Brenner SE, Holbrook SR. RNA structural motifs: building blocks of a modular biomolecule. *Q Rev Biophys.* 2005;38(3):221-43.

Jacquier A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nat Rev Genet.* 2009;10(12):833-44.

Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes

and a possible function for pervasive transcription. *Science.* 2007;316(5830):1484-8.

Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet.* 2007;8(6):413-23.

Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol.* 2009;10(2):126-39. Review.

Li PT, Viereggs J, Tinoco I Jr. How RNA unfolds and refolds. *Annu Rev Biochem.* 2008;77:77-100. Review.

Liu Z, Deibler RW, Chan HS, Zechiedrich L. The why and how of DNA unlinking. *Nucleic Acids Res.* 2009;37(3):661-71. Review.

Malone CD, Hannon GJ. Small RNAs as guardians of the genome. *Cell.* 2009;136(4):656-68. Review.

Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell.* 2009;136(4):629-41. Review.

Schoeffler AJ, Berger JM. DNA topoisomerases: harnessing and constraining energy to govern chromosome topology. *Q Rev Biophys.* 2008;41(1):41-101.

Sharp PA. The centrality of RNA. *Cell.* 2009;136(4):577-80.

Svoboda P, Di Cara A. Hairpin RNA: a secondary structure of primary importance. *Cell Mol Life Sci.* 2006;63(7-8):901-8. Review.

Witz G, Stasiak A. DNA supercoiling and its role in DNA decatenation and unknotting. *Nucleic Acids Res.* 2010;38(7):1-15.

Henrique Bunselmeyer Ferreira

Cromatina

1. Cromatina procariótica 38

- 1.1 Nucleoide e cromatina bacteriana 38
- 1.2 Nucleoide e cromatina de arqueas 42

2. Cromatina eucariótica 46

- 2.1 Composição molecular da cromatina eucariótica 46
- 2.2 Nucleossomos eucarióticos 48

2.3 Níveis de organização mais complexos da cromatina eucariótica 48

- 2.3.1 Fibra de 10 nm 50
- 2.3.2 Fibra de 30 nm 52
- 2.3.3 Alças de cromatina e fibra de 300 nm 53
- 2.3.4 Estruturas de cromatina de ordem superior e cromossomos metafásicos 54

3. Cromatina de organelas 54

- 3.1 Nucleoides mitocondriais 55
- 3.2 Nucleoides de plastídeos 55

Um princípio geral é evidente na organização do material genético de qualquer forma de vida, seja ela acelular, como um vírus, ou celular, como um organismo procariótico ou eucariótico. Em qualquer caso, as moléculas de RNA ou DNA devem ser compactadas para sua acomodação no compartimento que ocupam. Isso ocorre porque a extensão das moléculas de ácido nucleico, que constituem os genomas, excede em muito as dimensões do espaço a elas destinado, seja ele proporcionado por um capsídeo viral, por uma célula procariótica ou pelo núcleo de uma célula eucariótica. Por exemplo, o capsídeo do vírus do mosaico do tabaco (TMV, de *tobacco mosaic virus*), com 0,0008 μm de diâmetro e 0,3 μm de comprimento, acomoda um genoma de RNA que tem 2 μm de extensão. Já a bactéria *Escherichia coli*, um procarioto típico, com uma célula cilíndrica de 2 μm de comprimento e 1 μm de diâmetro, abriga um genoma de DNA circular de 1.600 μm (1,6 mm) de circunferência. Entre eucariotos, as discrepâncias entre as dimensões do compartimento que abriga o genoma e a extensão de DNA total são ainda maiores. O núcleo esférico de 6 μm de diâmetro de uma célula humana, por exemplo, abriga uma extensão total de DNA de 1,8 m.

É fundamental que a compactação do material genético aconteça de forma organizada, pois somente assim pode ser viabilizada a ocorrência de processos funcionais (da replicação e segregação de cromossomos até a expressão gênica). Essa compactação funcional é alcançada pelas interações organizadas das moléculas de ácido nucleico com as proteínas específicas. A identidade dessas proteínas e a mecânica molecular da compactação genômica diferem entre os diversos organismos.

Em alguns sistemas virais, diferentes estratégias de montagem de capsídeos em torno de moléculas de DNA ou RNA são utilizadas. Exemplos de processos de montagem de capsídeo são discutidos no **Material Complementar Online 3.1**. Em sistemas celulares, por outro lado, tanto no caso de procariotos como no de eucariotos, o DNA que constitui o genoma é compactado por sua associação com proteínas em um arranjo mais complexo, cujo grau de compactação varia de acordo com seu estado funcional. Esse arranjo nucleoproteico organizado e dinâmico é chamado de **cromatina** e, ao longo deste capítulo, serão analisadas as configurações típicas da cromatina de procariotos e de eucariotos, descrevendo sua organização, unidades constituintes básicas e discutindo como elas podem ser arranjadas em estruturas progressivamente mais compactas.

1. Cromatina procariótica

Todos os procariotos (bactérias e arqueas) têm genomas estruturados em um **nucleoide**, uma massa de cromatina ocupando cerca de 1/4 do volume celular total. Por não serem delimitados por qualquer tipo de envoltório e serem estruturas funcionalmente dinâmicas, os nucleoides procarióticos não têm uma forma fixa, embora possa

ser dito que eles são cilíndricos. Em *E. coli*, por exemplo, as dimensões do nucleoide giram em torno de 1,7 μm de comprimento e 0,65 μm de diâmetro, determinando um volume próximo a 0,5 μm^3 . Esse volume contrasta com os 200 μm^3 , que seriam ocupados pelo DNA genômico da bactéria, caso estivesse livre. Por isso, cromossomos procarióticos, no contexto do nucleoide, precisam ser compactados algumas centenas de vezes, formando uma estrutura de cromatina organizada e funcional.

Apesar de semelhantes em muitos aspectos de suas respectivas organizações gênica e genômica (ver Capítulo 4), bactérias e arqueas diferem em outros aspectos moleculares básicos, dentre os quais se inclui a estrutura da cromatina. Isso ocorre porque diferentes proteínas interagem com o DNA em nucleoides de bactérias e arqueas, e os complexos nucleoproteicos formados definem configurações de cromatina distintas, entre esses dois clados procarióticos. Nas duas próximas seções, será descrita a estrutura geral do nucleoide e as interações nucleoproteicas que caracterizam a cromatina procariótica.

1.1 Nucleoide e cromatina bacteriana

A estrutura geral do nucleoide depende de forças inerentes à própria estrutura do DNA cromossômico e de interações entre o DNA e outras moléculas, podendo ser identificados diferentes níveis de compactação e organização estrutural e funcional na cromatina bacteriana. No espaço ocupado pelo nucleoide em uma célula, o DNA encontra-se em uma concentração elevada (da ordem de 50 a 100 mg/mL), o que também acontece com proteínas e íons que têm afinidade pelo DNA.

Algumas poliaminas (cátions polivalentes, como a espermidina e a putrescina) e outros cátions, juntamente com o DNA, configuram o nucleoide como um ambiente eletricamente carregado. Nesse ambiente, a estrutura polianiônica do DNA e a autorrepulsão proporcionada por ela, contrabalançada por interações com contraíons, levam à ocorrência de mudanças de trajetórias aleatórias no DNA, a cada 150 pb (equivalentes a uma extensão de 50 nm), que definem o primeiro nível de compactação do DNA, cromossômico (**Figura 3.1A**).

Um nível maior de compactação do cromossomo é proporcionado pela formação de alças, que formam **domínios topológicos** independentes (Figura 3.1B). Cada alça que constitui um domínio tem, em média, 10 kb de extensão total, sua base e o seu ápice são mantidos por interações entre o DNA e as proteínas cromatínicas. Ao longo de cada alça, um nível adicional de compactação da cromatina é obtido pela introdução do **superenrolamento negativo** no DNA (ver Capítulo 2), o que também depende de interações com proteínas específicas (detalhe da Figura 3.1B). Além disso, no contexto da organização em domínios topológicos, há níveis adicionais de organização estrutural do cromossomo, que definem os chamados **macrodomínios**.

Os macrodomínios são superestruturas que abrangem de dezenas a centenas de domínios topológicos,

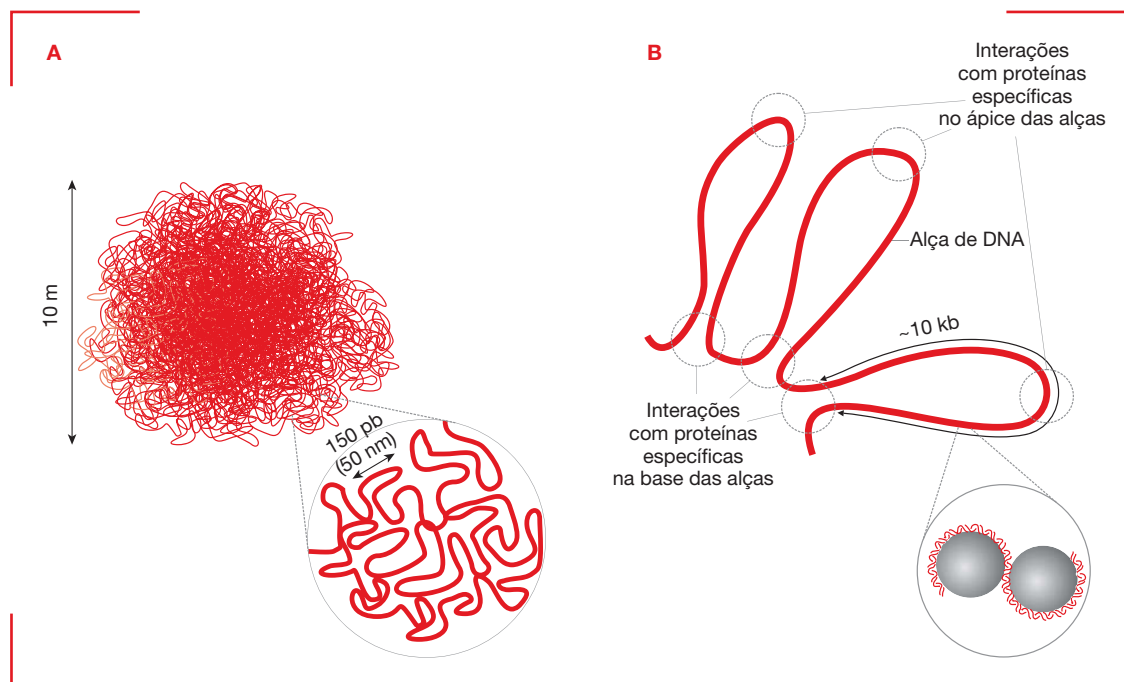


Figura 3.1

Níveis progressivos de compactação do DNA no nucleóide bacteriano. (A) Um cromossomo bacteriano em solução, independentemente de interações com proteínas, forma uma massa de DNA enrolado de forma aleatória, que ocuparia um espaço esférico com um diâmetro em torno de 10 μm ; isso ocorre porque forças entrópicas determinam que o DNA mude de trajetória (dobre-se) aproximadamente a cada 150 pb ou 50 nm (detalhe). **(B)** Na célula, interações com proteínas definem a formação de alças que constituem domínios topológicos independentes, sendo cada uma estabilizada na sua base e no seu ápice; a compactação adicional é conseguida pelo superenrolamento localizado do DNA nas alças, o que depende de interações com proteínas (detalhe).

compartimentalizando o cromossomo em grandes segmentos com funções distintas (ver Capítulo 4). Do ponto de vista funcional, os macrodómnios podem diferir quanto à probabilidade de seu envolvimento em eventos de recombinação, ou à frequência de transcrição nos genes ali presentes. Do ponto de vista estrutural, pouco se sabe sobre a organização da cromatina nos macrodómnios, mas acredita-se que sejam interações específicas com proteínas que delimitam fisicamente cada um desses grandes segmentos cromossômicos, ao mesmo tempo em que determinam suas diferentes propriedades funcionais.

A estrutura do nucleóide bacteriano é dinâmica e varia conforme o estado fisiológico da célula. Por exemplo, quando uma bactéria está na chamada fase exponencial de sua multiplicação, o nucleóide se organiza em um grande número (de centenas a milhares) de domínios topológicos e possui DNA superenrolado negativamente em cada um deles (**Figura 3.2A e B**). Os limites e a extensão de cada domínio variam muito, podendo haver a formação de alças com até 30 kb ou mais de extensão e com o nucleóide podendo passar de uma estrutura mais compacta (**Figura 3.2A**), que agrupa certos genes altamente transcritos, como os genes de rRNA, para uma mais estendida (**Figura 3.2B**), que desfaz esses agrupamentos, quando há diminuição do nível geral de transcrição. As

proteínas que delimitam as alças (ver **Figura 3.1B**) atuam como barreiras que impedem a livre difusão de superenrolamento ao longo do DNA, permitindo que, em cada domínio, o DNA possa apresentar diferentes graus de superenrolamento. Já na fase estacionária de multiplicação caracterizada por desaceleração do processo de divisão celular e diminuição acentuada da atividade transcricional, o nucleóide passa a ter um número menor de domínios e, em cada um deles, o DNA apresenta uma estrutura mais relaxada, isto é, com menor grau de superenrolamento (**Figura 3.2C**).

Um grupo heterogêneo com mais de uma dezena de proteínas que se ligam ao DNA, chamadas de **proteínas associadas ao nucleóide (NAPs, de nucleoid-associated proteins)** (**Tabela 3.1**), é o principal responsável pela estruturação do nucleóide. As NAPs desempenham diferentes papéis na organização da cromatina bacteriana e a atividade de cada proteína depende do(s) efeito(s) que ela exerce sobre o trecho do DNA ao qual está ligada. Os efeitos mais evidentes das NAPs são: o dobramento ou o enrolamento localizado de uma molécula de DNA e a interligação (*bridging*) de duas porções de uma mesma molécula de DNA ou de duas moléculas de DNA diferentes (formando uma ponte DNA-proteína-DNA). Na **Figura 3.3**, estão representadas as NAPs bacterianas mais conhecidas

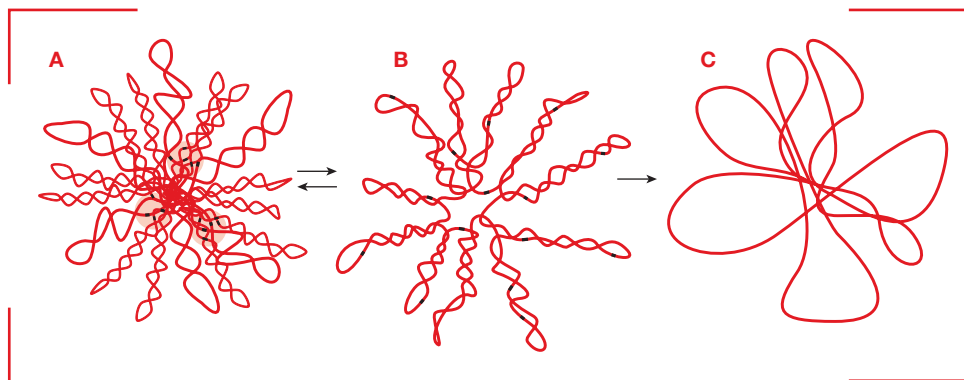


Figura 3.2

A dinâmica de alteração estrutural do nucleóide bacteriano nas fases exponencial ou estacionária de multiplicação. Na fase exponencial de multiplicação, o nucleóide pode adotar uma conformação mais compacta (A), com maior número de domínios (alças) de DNA superenrolado, ou uma conformação mais estendida (B), com um número menor desses domínios topológicos. A transição de uma conformação para a outra é reversível, como indicado pelas setas, e está correlacionada com o nível de transcrição geral na célula. Em situações de multiplicação mais acelerada (altos níveis transcripcionais), é adotada a conformação compacta, que aproxima fisicamente genes altamente transcritos (representados em ■), formando o que se convencionou chamar de “fábricas de transcrição” (áreas indicadas por círculos em ■). Em situações de multiplicação mais lenta (níveis transcripcionais menores), é adotada a conformação estendida, que desfaz as fábricas de transcrição mantendo os genes transcritos mais dispersos. Quando a bactéria chega a uma fase estacionária de multiplicação, o nucleóide passa a ter uma estrutura ainda mais estendida (C), em que o número de domínios topológicos é menor e a estrutura do DNA é mais relaxada (com menor conteúdo de superenrolamento).

Tabela 3.1 Proteínas associadas ao nucleóide bacteriano e suas propriedades

Proteína ^a	Massa molecular	Estado oligomérico funcional	Sítios de ligação no DNA ^b	Efeito(s) sobre o DNA ^c
CbpA	33 kDa	Homodímero	DNA curvado	n.d.
CbpB	33 kDa	Monômero	DNA curvado	n.d.
EbfC	11 kDa	Homodímero	GTNAC	Provavelmente interligação
Dps/MrgA	17-19 kDa	Monômero ou dodecâmero	n.d.	n.d.
Fis	22 kDa	Homodímero	Segmentos com repetições de AT ou A, especialmente (A) ₆	Dobramento, enrolamento, interligação
Hlp	21 kDa	Monômero	n.d.	n.d.
H-NS	15,5 kDa	Homodímero ou Heterodímero (H-NS–StpA)	TCGATAAATT e sequências ricas em AT	Interligação
HU	9,5 kDa (HU α) 9,5 kDa (HU β)	Heterodímero (HU α -HU β)	Preferencialmente sequências ricas em AT ou DNA curvado	Dobramento, enrolamento
IHF	11 kDa (IHF α) 9,5 kDa (IHF β)	Heterodímero (IHF α -IHF β)	(A/T)ATCAANNNTT(A/G)	Dobramento
Lrp/AsnC	15-18 kDa	Homodímero, Octâmero ou Hexadecâmero	n.d.	Enrolamento, interligação
Lsr2	12 kDa	Homodímero	Sequências ricas em AT	Interligação
SMC/MukB	150-200 kDa	Homodímero	Preferencialmente DNA de fita simples	Interligação
MvaT	13,6 kDa	Homodímero	Sequências ricas em AT	Interligação
StpA	15,3 kDa	Homodímero ou Heterodímero (H-NS–StpA)	Sequências ricas em AT	Interligação

^a Informações sobre a nomenclatura das proteínas listadas estão disponíveis no **Material Complementar Online 3.2**.

^b Nas sequências de DNA, N = qualquer nucleotídeo (A, C, G ou T).

^c Apenas efeitos já comprovados por experimentos *in vitro* ou *in vivo* estão listados.

n.d. = não determinado.

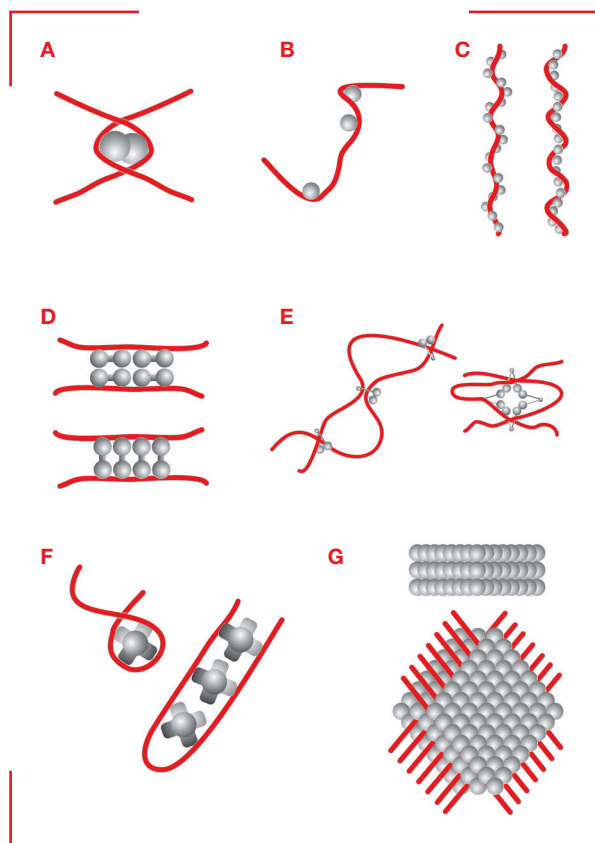


Figura 3.3

Ligação e efeitos no DNA de algumas NAPs. Nos modelos apresentados, o dúplice de DNA está representado em ■ e as proteínas, nos respectivos estados oligoméricos funcionais, em ■.

(A) Dobramento e formação de um nó (interligação de dois segmentos do DNA) devido à ligação de dois dímeros (cada um representado por uma esfera) da Fis. **(B)** Dobras no DNA induzidas por ligação da HU (em baixas concentrações) ou da IHF. **(C)** Em concentrações mais elevadas, a HU leva à formação de filamentos super-helicoidais de DNA relativamente rígidos; dois modelos de filamentos HU-DNA são mostrados, com o da esquerda correspondendo a um filamento menos rígido e o da direita a um filamento mais rígido. **(D)** Dois modelos possíveis para a interligação de DNA mediada por moléculas da H-NS. **(E)** Entrelaçamento de segmentos de DNA, devido à ligação de moléculas da SMC/MukB; os dímeros da proteína podem estar associados individualmente ao DNA (esquerda) ou na forma de multímeros (direita). **(F)** Enrolamento (esquerda) e interligação (direita) do DNA mediados pela Lrp. **(G)** A Drp forma dodecâmeros (representados por esferas) que se associam em uma estrutura hexagonal de 3 camadas (em cima, em uma vista lateral) com poros; nesse complexo proteico, o DNA é arranjado em uma estrutura de segmentos paralelos, que passam entre as camadas e através dos poros (embaixo, em uma vista laterofrontal).

estrutural e funcionalmente, sendo apresentados alguns dos efeitos exercidos por elas sobre o DNA.

Todas as NAPs têm a capacidade de compactar o DNA, mas, na organização da cromatina, a contribuição individual de cada tipo de NAP é relativamente modesta. Os graus de compactação e organização atingidos são o resultado do somatório das atividades de várias NAPs. Além disso, há sobreposição parcial de funções entre as NAPs, o que é evidenciado pela manutenção da estrutura da cromatina mesmo na ausência de uma dessas proteínas (em caso de mutação deletéria no gene correspondente), demonstrando que as atividades de uma ou mais das outras NAPs podem compensar funcionalmente essa ausência.

Apesar do pouco conhecimento sobre as funções específicas desempenhadas pelas NAPs, algumas dessas proteínas podem ser associadas à compactação global do nucleóide, à formação de determinadas estruturas da cromatina bacteriana ou à definição do grau de superenrolamento do DNA. Existem evidências experimentais do envolvimento de pelo menos seis NAPs (Dps/MrgA, Fis, IHF, H-NS, HU e SMC/MukB) nos mecanismos gerais de compactação da cromatina, que dependeriam apenas de suas capacidades de ligação ao DNA. A manutenção de certas estruturas da cromatina, como as bases e os ápices das alças de DNA (ver Figura 3.B), por sua vez, parece envolver a atuação mais específica de algumas NAPs. Na

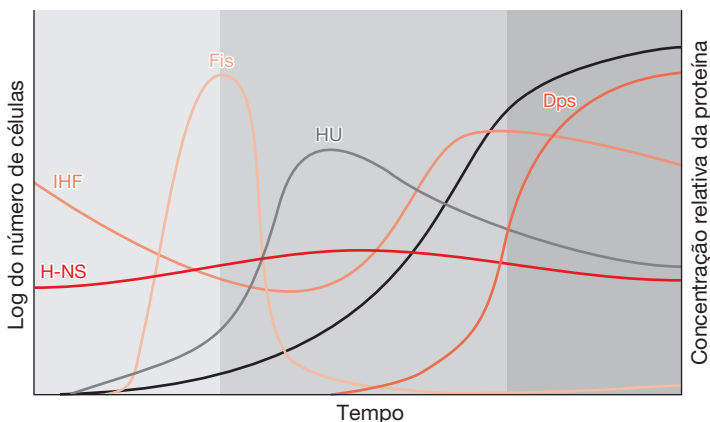
estabilização das bases das alças, estariam envolvidas NAPs capazes de fazer a interligação de diferentes segmentos do DNA (H-NS, Lrp e SMC/MukB), ao passo que NAPs capazes de induzir o dobramento do DNA, (Fis, IHF e HU) definiriam o ápice de cada alça. Por fim, o grau de superenrolamento do DNA nos domínios topológicos independentes requer a atuação de NAPs como a SMC/MukB, que introduz e estabiliza superenrolamentos negativos no DNA, e a H-NS, que delimita segmentos de DNA com diferentes graus de superenrolamento.

O repertório de NAPs expressas na célula varia em função do estado fisiológico da bactéria, sugerindo uma relação de causa e efeito com as alterações estruturais observadas no nucleóide (ver Figura 3.2). A **Figura 3.4** apresenta os padrões de expressão diferencial de algumas NAPs, em função do andamento do processo de multiplicação de *E. coli*, demonstrando que diferentes composições de proteínas que dobram e interligam o DNA modulam a estrutura do nucleóide em cada fase. As diferentes composições proteicas e o estado de compactação do DNA modulam funcionalmente a cromatina bacteriana, que, em fase exponencial de multiplicação, é muito mais ativa do ponto de vista da transcrição e da replicação do que no estágio de multiplicação estacionário.

A mecânica molecular dos processos de regulação da expressão gênica é influenciada pela estrutura da cromatina, de modo que muitos genes podem ter a sua trans-

Figura 3.4

Níveis de expressão de algumas NAPs em diferentes fases da multiplicação de *E. coli*. A curva sigmoide típica que representa o processo de multiplicação celular, obtida a partir da plotagem do logaritmo (Log) do número de células *versus* o tempo, está indicada em ■ e as diferentes fases do processo (*lag*, exponencial e estacionária) estão indicadas, respectivamente, por áreas coloridas em tons progressivamente mais escuros de ■. As concentrações relativas das diferentes NAPs estão indicadas por curvas coloridas em ■ (Fis), ■ (H-NS), ■ (HU), ■ (IHF) e ■ (Dps).



crição regulada por NAPs. Esse tipo de regulação pode ser de dois tipos: indireto, em razão de alterações no grau de supertorção de um domínio cromossômico (alça), que afetam positiva ou negativamente a transcrição de genes naquela região; e direto, pois algumas NAPs podem se ligar a regiões reguladoras e ativar ou reprimir a transcrição de alguns genes.

Existe outras proteínas que podem influenciar a estrutura do nucleóide. Dentre essas proteínas, destacam-se a DnaA, a IciA e a SeqA, envolvidas no controle da replicação (ver Capítulo 6), CAP e outros reguladores transcricionais (ver Capítulo 13). Todas essas proteínas são capazes de remodelar o DNA em certo grau, o que fez com que algumas chegassem a ser classificadas como NAPs. Entretanto, a remodelação introduzida por qualquer uma delas na cromatina está restrita a regiões reguladoras da replicação (como a origem de replicação (OriC) e sítios metilados ou hemimetilados no DNA), ou da transcrição (como promotores). Portanto, a contribuição dessas proteínas para o estado de compactação da cromatina bacteriana é apenas moderada, justificando sua exclusão do conjunto de NAPs.

Acredita-se também que algumas enzimas, como a DNA-girase (ver Capítulo 6) e a RNA-polimerase (ver Capítulo 10), desempenhem algum papel na organização da cromatina bacteriana. A ligação ao DNA e as atividades de polimerase e topoisomerase, respectivamente, dessas enzimas, alteram de forma localizada a distribuição de superenrolamentos no DNA e constituem barreiras topológicas de natureza transitória na cromatina. Assim, elas podem ser responsáveis, em parte, pela definição dos limites entre domínios ou entre macrodomínios e pelas transições nos estados de superenrolamento do DNA nas estruturas cromatínicas de ordem superior.

1.2 Nucleóide e cromatina de arqueas

Nas arqueas, o material genético está, assim como nas bactérias, organizado em um nucleóide, que consiste em uma massa de cromatina cujo grau de compactação varia

conforme o estado fisiológico da célula. Entretanto, as fibras de cromatina que compõem o nucleóide em arqueas diferem estruturalmente das bacterianas, pois apresentam sistemas de compactação distintos, alguns dos quais lembram mais estruturas da cromatina eucariótica (Seção 2). Essa organização “híbrida” da cromatina de arqueas é reflexo da história evolutiva do domínio Archaea, que se separou inicialmente da linhagem bacteriana e, em um segundo momento, da linhagem eucariótica (ver Capítulo 4). Com isso, as arqueas atuais apresentam um amplo repertório de NAPs, incluindo tanto proteínas exclusivas do domínio Archaea, como proteínas ortólogas de NAPs bacterianas ou de proteínas cromatínicas eucarióticas.

As NAPs de arqueas e algumas de suas principais propriedades estruturais e funcionais estão listadas na **Tabela 3.2**. A maioria das NAPs de arqueas é capaz de polimerizar sobre o DNA e, assim, formar fibras nucleoproteicas. Essas fibras cromatínicas de arqueas podem apresentar diferentes configurações, dependendo da espécie, em função do repertório de NAPs presente e da quantidade relativa de cada uma dessas proteínas. É importante salientar, contudo, que a distribuição das NAPs não é homogênea no domínio Archaea, pois várias delas têm distribuição restrita a um determinado filo ou a certas classes ou gêneros. Além disso, o número de genes parálogos que codifica cada tipo de NAP varia bastante (podendo chegar até 6), implicando em variação correspondente no repertório de isoformas para pelo menos algumas dessas proteínas.

Uma análise filogenética sobre a distribuição dos genes codificadores de NAPs demonstra que, em diferentes linhagens evolutivas do domínio Archaea, diversos subconjuntos de NAPs são utilizados para a compactação da cromatina (**Figura 3.5**). Considerando espécies com genomas já sequenciados, os dois tipos de NAPs com maior distribuição entre arqueas são Alba e histonas. (ver Capítulo 4).

As proteínas Alba (em uma ou duas isoformas) estão presentes nos 3 filos de arqueas (Euryarchaeota, Nono-

Tabela 3.2 Proteínas associadas ao nucleóide em arqueas e suas propriedades

Proteína ^a	Massa molecular	Estado oligomérico funcional	Sítios de ligação no DNA	Efeito(s) sobre o DNA ^b
Alba ^c	10 kDa	Homodímero ou Heterodímero	n.d.	Interligação
Cren7	6,6 kDa	Monômero	Afinidade genérica por DNA de fita dupla (liga-se nas cavidades menores do DNA)	Dobramento
CC1	6 kDa	Monômero	n.d.	n.d.
Histonas ^c	7,5 kDa	Homodímero ou Heterodímero	(A/T) ₃ NN(G/C) ₃ NN ^d	Enrolamento
HU/HTa	10 kDa	Homodímero	n.d.	Dobramento
Lrp	15-22 kDa	Homodímero	n.d.	Enrolamento, interligação
MC1	10 kDa	Homodímero	Sequências ricas em AT	Dobramento
Sul7d	7 kDa	Monômero	n.d.	Dobramento

^a Informações sobre a nomenclatura das proteínas listadas estão disponíveis no **Material Complementar Online 3.3**.

^b Apenas efeitos já comprovados por experimentos *in vitro* ou *in vivo* estão listados.

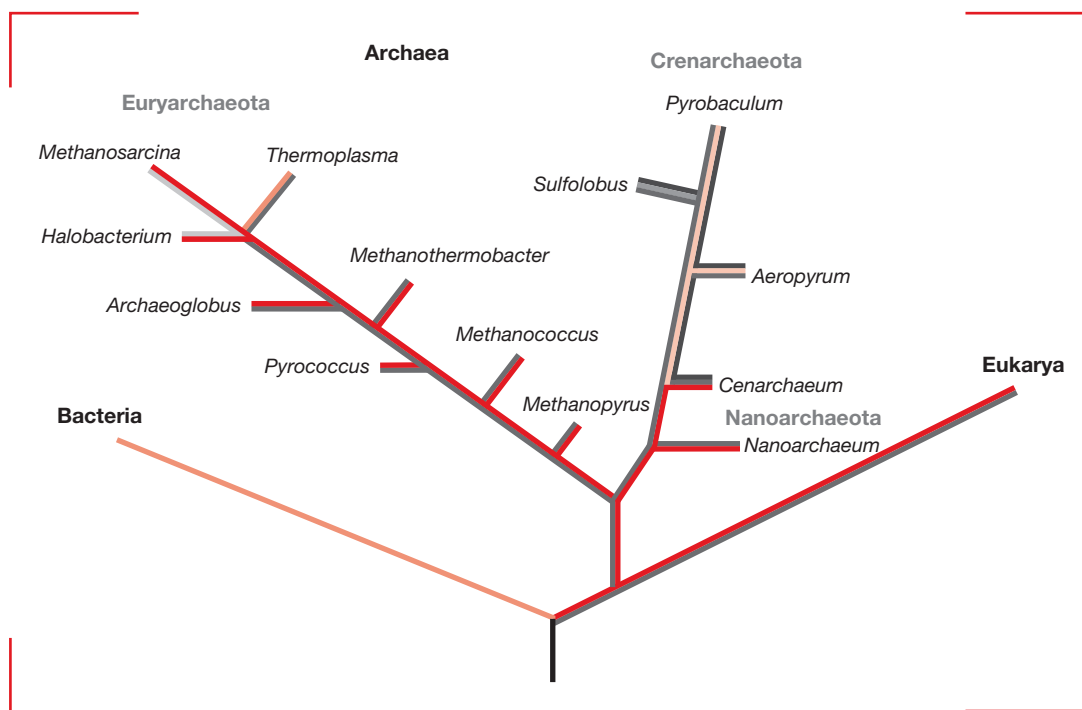
^c Proteínas Alba e histonas podem estar presentes em mais de uma isoforma, dependendo da espécie.

^d Na sequência de DNA, N = qualquer nucleotídeo (A, C, G ou T).

n.d. = não determinado.

archaeota e Chrenarchaeota), não tendo sido identificadas apenas em euriarqueotas das classes Halobacteria e Methanomicrobia. As histonas (de uma a seis isoformas), são também encontradas nos 3 filos de arqueas, mas estão ausentes em euriarqueotas da classe Thermoplasmata e em crenarqueotas hipertermofílicas. As outras proteí-

nas cromatínicas de arqueas, por sua vez, têm distribuições mais restritas. Por exemplo, a MC1 é encontrada apenas nas classes Halobacteria e Methanomicrobia e a HU/HTa e a Sul7d estão restritas à classe Thermoplasmata e à ordem Sulfolobales de crenarqueotas, respectivamente. É importante notar que, em todas as linhagens

**Figura 3.5**

Árvore filogenética mostrando a distribuição de proteínas associadas ao nucleóide em arqueas. As proteínas representadas são: Alba ■, histonas ■, CC1 ■, Cren 7 ■, HU/HTa ■ e MC1 ■ e Sul7d ■. Na representação simplificada da figura, a ordem das ramificações é correta, mas a extensão dos ramos não corresponde às distâncias evolutivas reais.

de arqueas, pelo menos dois tipos de NAPs são expressos. Nas espécies que possuem Alba, há também a presença de histonas (Cren7 ou HU/HTa), podendo ocorrer ainda, em espécies de crenarqueotas, um terceiro tipo de NAP (CC1 ou Sul7d). Em espécies que não possuem proteínas Alba, há, em contrapartida, histonas e MC1.

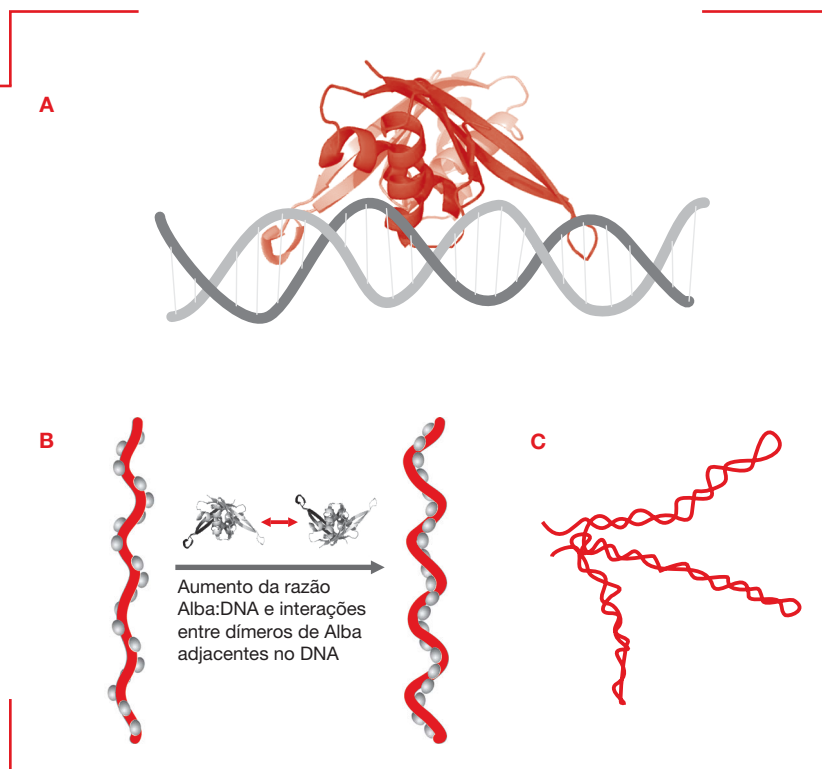
Estrutural e funcionalmente, as proteínas Alba e as histonas são mais conhecidas e, por isso, serão aqui utilizadas para apresentação de algumas características da cromatina de arqueas. As **proteínas Alba** em arqueas são codificadas por um ou dois genes e, em casos de presença de dois genes parálogos, as isoformas de Alba produzidas (Alba1 e Alba2), podem formar homodímeros Alba1-Alba1 e heterodímeros (Alba1-Alba2). Dímeros Alba2-Alba2 não ocorrem, pois a isoforma Alba2 representa apenas 5% do total de proteínas Alba na célula. Dímeros de Alba associam-se ao DNA (**Figura 3.6A**), cobrindo-o em arranjos nucleoproteicos em tandem. Com isso, são formadas fibras nucleoproteicas helicoidais estendidas (similares às formadas pela HU bacteriana, representadas na Figura 3.3C), cujo grau de compactação aumenta proporcionalmente ao número de dímeros presente (**Figura 3.6B**). Isso ocorre porque, quando Alba está presente em concentrações mais elevadas, há a associação de dímeros ocupando posições adjacentes ao longo do DNA. Heterodímeros Alba1-Alba2, em qualquer concentração, e homodímeros de Alba1, quando em altas concentrações, são também capazes de mediar a interligação de diferentes segmentos do DNA, o que leva à formação de estruturas altamente compactas e ramificadas na cromatina (**Figura 3.6C**).

As **histonas** são proteínas básicas, com elevada afinidade pelo DNA, que em arqueas, são tipificadas pelas duas isoformas presentes em *Methanothermus fervidus*, denominadas HMfA e HMfB. As histonas foram inicialmente caracterizadas como componentes da cromatina eucariótica (ver Seção 2), mas a demonstração da sua presença e ampla distribuição entre arqueas (ver Figura 3.5) evidenciou aspectos comuns na organização da cromatina também nos domínios Archaea e Eukarya. As histonas de arqueas, apesar de menores, são estruturalmente muito similares a histonas eucarióticas. Elas consistem de uma cadeia de 65 a 75 aminoácidos enovelada em 3 α -hélices separadas por alças curtas (**Figura 3.7A**), em uma estrutura homóloga àquela do domínio característico (*histone fold*) das histonas centrais eucarióticas (ver **Figura 3.8**). De um a seis genes codificadores de histonas são encontrados nos genomas de espécies de arqueas, que possuem essas proteínas, são estruturalmente estabilizadas a partir da formação de homo ou de heterodímeros (**Figura 3.7B**). Na presença do DNA, os dímeros das histonas associam-se para formar tetrâmeros, em torno dos quais o ácido nucleico se enrola para formar **nucleossomos** (**Figura 3.7C**), que são muito similares aos seus equivalentes eucarióticos (ver Seção 2.2). Na cromatina, os nucleossomos aparecem organizados em tandem, proporcionando a formação de fibras compactas que lembram um colar de contas (**Figura 3.7D**).

Em espécies de arqueas com duas ou mais isoformas de histonas, diversas combinações de isoformas podem ser encontradas nos tetrâmeros. Essas diferentes combi-

Figura 3.6

Interações das proteínas Alba com o DNA em arqueas. (A) Dímero de Alba1 (com um monômero em ■ e outro em ■) associado ao DNA (em ■). **(B)** Fibras formadas por dímeros de Alba (representados por formas ovais em ■) associados em arranjos em tandem no DNA (com o dúplice representado em ■); as fibras nucleoproteicas helicoidais formadas são compactadas devido a interações entre dímeros de Alba adjacentes, que ocorrem quando a razão Alba:DNA aumenta. Heterodímeros Alba1-Alba2, em qualquer concentração, e homodímeros de Alba1, em concentrações mais elevadas, interligam diferentes segmentos do DNA, levando à formação de estruturas altamente ramificadas e compactas na cromatina, como o segmento de DNA associado a proteínas Alba representado em **(C)**.



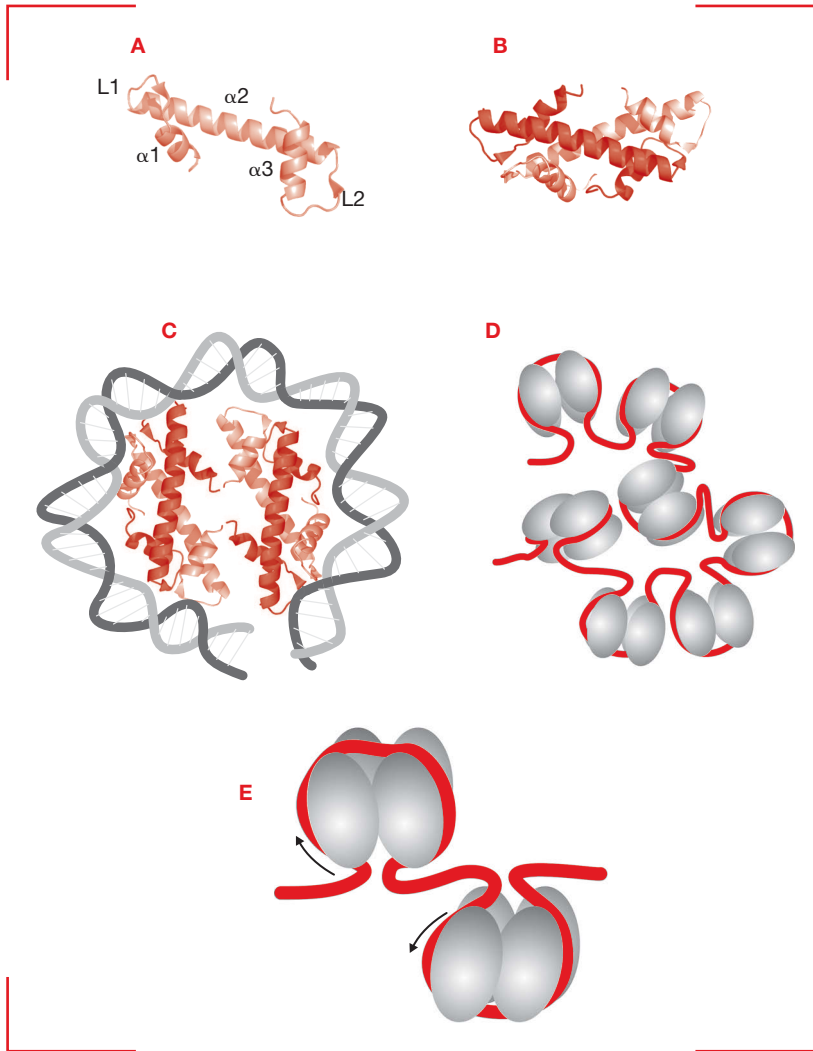


Figura 3.7

Histonas e nucleossomos de arqueas. (A) As histonas de arqueas apresentam um enovelamento típico (*histone fold*) dessa família de proteínas, compartilhado com as histonas centrais de eucariotos. Cada histona consiste, em 3 α -hélices ($\alpha 1$, $\alpha 2$ e $\alpha 3$), unidas por duas alças curtas (L1 e L2). (B) Dímero de histonas, com um monômero representado em ■ e outro em ■. (C) Na presença de DNA, dímeros de histonas associam-se para formar tetrâmeros, em torno dos quais aproximadamente 90 pb se enrolam para formar um nucleossomo. (D) Ao longo do DNA, os nucleossomos aparecem organizados em tandem, em fibras de cromatina com forma similar à de um colar de contas. (E) Dois nucleossomos adjacentes com orientações dímero-dímero distintas (representadas por inclinações distintas do dímero anterior em relação ao dímero posterior em cada um dos nucleossomos representados). A variação do ângulo formado entre os dímeros altera o sentido do enrolamento do DNA em cada nucleossomo (indicado por uma seta), gerando superenrolamento negativo em um caso, e positivo em outro.

nações de histonas determinam várias propriedades de associação ao DNA para cada tipo de tetrâmero, o que leva a modificações na estrutura das fibras de cromatina. Por exemplo, as combinações de histonas determinam alterações na orientação de um dímero em relação a outro, na estrutura do tetrâmero, levando a mudanças no sentido de enrolamento do DNA no nucleossomo (Figura 3.7E) que pode levar, à introdução de superenrolamento negativo ou positivo na molécula.

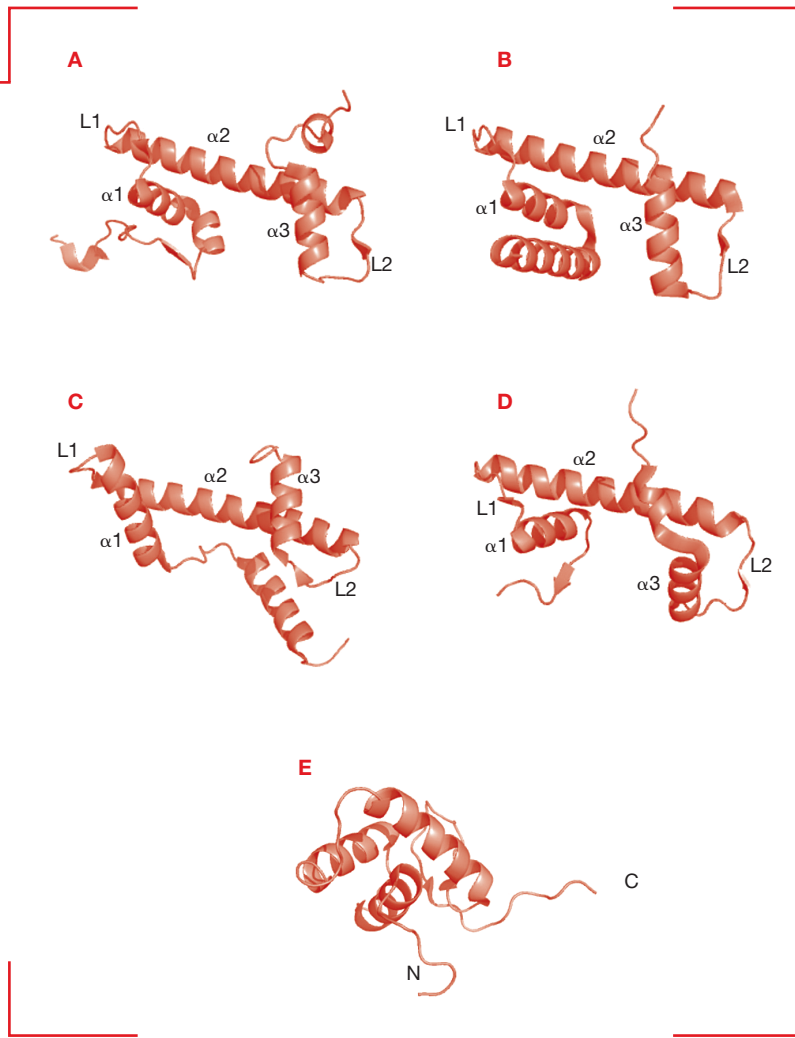
De 60 a 90 pb associam-se a histonas para formar um nucleossomo em arqueas e essa extensão de DNA não chega a dar uma volta completa no tetrâmero (ver Figura 3.7C). O número exato de pares de bases de DNA, em cada nucleossomo, depende da composição de isoformas das histonas presente no tetrâmero. A distribuição de nucleossomos ao longo do DNA na cromatina das arqueas, por sua vez, é mais uniforme. Há uma preferência para a montagem de nucleossomos sobre segmentos de DNA que estejam de acordo com a sequência consensual $(A/T)_3NN(G/C)_3NN$ (sendo N qualquer base nitrogena-

da), os quais têm uma ampla distribuição ao longo de genomas de arqueas. Acredita-se, também, que tetrâmeros de arqueas com diferentes composições de histonas possam apresentar variações quanto à afinidade por certas sequências no DNA.

Devido à maior simplicidade dos nucleossomos em arqueas (menor conteúdo de histonas), o grau de compactação atingido na fibra cromatina é menor do que o atingido em fibras cromatínicas eucarióticas. Isso determina diferenças fisiológicas importantes entre a cromatina de arqueas e a de eucariotos, pois, nas arqueas, a configuração menos compacta da cromatina proporciona à maquinaria transcricional alta acessibilidade ao genoma, o que não ocorre nos eucariotos. Essa acessibilidade proporciona às arqueas a capacidade de responder, com rapidez, a mudanças ambientais (em termos de expressão gênica), o que é uma exigência comum para procariotos, pois, em geral, estão em contato direto com os fatores ambientais (bióticos e abióticos) que os condicionam.

Figura 3.8

Estrutura de histonas centrais e de histonas de ligação eucarióticas. Todas as histonas centrais (A-D) apresentam o domínio típico (*histone fold*) dessa família de proteínas, que consiste de 3 α -hélices (α_1 , α_2 e α_3) unidas por duas alças curtas (L1 e L2). Nas histonas H2A (A), H2B (B), H3 (C) e H4 (D), esse domínio é flanqueado por extensões (caudas) amino e carboxiterminais. As histonas representadas são as de nucleossomos de galinha (Harp *et al.*), cujas estruturas são representativas de cada uma das quatro classes de histonas centrais. (E) Domínio central globular, contendo 3 α -hélices, característico das histonas de ligação H1 e H5. O domínio representado é da histona H1 da levedura *S. cerevisiae* (Ali *et al.*) e as caudas amino e carboxiterminais da proteína, não são mostradas, estando indicadas apenas por N e C, respectivamente.



2. Cromatina eucariótica

A cromatina eucariótica, assim como a de arqueas que contém histonas (ver Seção 1.2), possui nucleossomos como sendo seus componentes estruturais básicos. Nucleossomos eucarióticos, contudo, possuem uma estrutura mais complexa que a dos nucleossomos de arqueas, pois são formados por um número maior de histonas. Esse aumento de complexidade dos nucleossomos em eucariotos proporciona maiores variações estruturais, o que implica em organização estrutural e dinâmica funcional mais complexas para a sua cromatina. Neste capítulo, serão discutidos os principais aspectos estruturais da cromatina de eucariotos e os diferentes estados de compactação passíveis de serem atingidos por fibras de cromatina. Os aspectos funcionais relacionados serão discutidos nos Capítulos 6, 10 e 14.

2.1 Composição molecular da cromatina eucariótica

A cromatina eucariótica é um complexo formado por várias moléculas de DNA linear (uma por cromossomo) associadas a proteínas. A massa de cromatina eucariótica chega a conter duas vezes mais proteínas do que o DNA e as suas principais proteínas estão listadas na **Tabela 3.3**. A massa de RNA desse complexo é 10 vezes menor do que a massa de DNA e é constituída, basicamente, por cadeias de RNA nascentes, ainda associadas à fita-molde de DNA.

As **histonas** são os constituintes proteicos principais da cromatina em quase todos os eucariotos, com exceção dos dinoflagelados (filó Dinoflagellata, supergrupo Chromalveolata), cuja cromatina é estruturada de maneira ainda pouco conhecida. Em geral, a massa combinada de histonas é aproximadamente igual à do DNA total presente na cromatina eucariótica.

Tabela 3.3 Proteínas associadas ao DNA na cromatina eucariótica e suas propriedades

Proteína ^a	Massa molecular	Estado oligomérico funcional	Sítios de ligação no DNA	Efeito(s) sobre o DNA ^b
Histonas centrais (H2A, H2B, H3 e H4)	11-14 kDa	Homodímero (parte do nucleossomo)	Preferencialmente regiões contendo dinucleotídeos TA repetidos a cada 10 pb e intercalados com dinucleotídeos GC (com 5 pb de distância entre cada TA e cada GC)	Enrolamento
Histonas de ligação (H1 e H5)	~21 kDa	Homodímero	Sequências ricas em AT	Interligação
Proteínas HMG	11-38 kDa	Homodímero ou heterodímero	Trechos ricos em AT	Dobramento
Proteínas SMC	~140 kDa	Heterodímero (p. ex. SMC2-SMC4, parte da condensina)	Sequências ricas em AT capazes de formar	Interligação estruturas secundárias

^a Informações sobre a nomenclatura das proteínas listadas estão disponíveis no **Material Complementar Online 3.4**.

^b Apenas efeitos já comprovados por experimentos *in vitro* ou *in vivo* estão listados.

Há muitas isoformas de histonas em células eucarióticas, mas será analisado, neste capítulo, apenas as chamadas **histonas canônicas** – envolvidas na compactação geral da cromatina. Essas histonas são divididas em dois grandes grupos: o das **histonas centrais**, que formam o complexo proteico central dos nucleossomos eucarióticos, e o das **histonas de ligação**, que se associam externamente aos nucleossomos e participam de interações entre eles.

As **histonas centrais (H2A, H2B, H3 e H4)**, são caracterizadas por um domínio constituído por 3 α -hélices separadas por alças curtas, chamado de **histone fold** (enovelamento de histonas) (Figura 3.8A-D). Entretanto, por serem um pouco maiores (possuem em torno de 100 aa de extensão), as histonas centrais apresentam extensões amino e carboxiterminais não encontradas em histonas de arqueas.

As **histonas de ligação (H1 e H5)**, por sua vez, são proteínas maiores que as histonas centrais (com 190 a 250 aminoácidos de extensão) e possuem uma estrutura típica de 3 domínios, na qual um domínio globular central (com aproximadamente 80 aminoácidos) é flanqueado por uma cauda aminoterminal mais curta (com até 40 aminoácidos) e uma cauda carboxiterminal mais longa (próximo a 100 aminoácidos). O domínio globular central (Figura 3.8E), com 3 α -hélices, é essencial para a ligação da histona ao nucleossomo, estando também a cauda carboxiterminal envolvida em interações com o DNA.

Tanto as histonas centrais como as histonas de ligação possuem um caráter básico acentuado, pois contêm uma elevada proporção (de 20 a 30%) de aminoácidos carregados positivamente (lisina e arginina) na sua constituição. A presença desses aminoácidos favorece a interação das histonas com o DNA de fita dupla, carregado negativamente.

As 5 classes de histonas canônicas (H2A, H2B, H3, H4 e H1/H5) podem ser reconhecidas em praticamente todos os eucariotos, como será visto no Capítulo 5. É co-

mo, principalmente no caso de histonas centrais, os membros de cada classe sejam codificados por famílias de genes parálogos, que, em muitas espécies, aparecem organizados em agrupamentos e têm expressão associada ao processo de replicação.

As histonas centrais são proteínas bastante conservadas ao longo da escala evolutiva, como H3 e H4, por exemplo, que possuem sequências de aminoácidos praticamente idênticas mesmo em organismos pouco relacionados como mamíferos e plantas dicotiledôneas. As histonas de ligação, por outro lado, são proteínas bem mais variáveis, com mais de uma centena de subtipos já descritos. Há variações consideráveis entre as formas de H1, encontradas em diferentes espécies, e entre isoformas expressas em diferentes tipos celulares ou tecidos de uma mesma espécie. As isoformas mais divergentes encontradas em eritrócitos de aves, como a galinha (onde foram inicialmente descritas), constituem a subclasse H5.

Além de histonas, a cromatina eucariótica contém também algumas outras proteínas, classificadas genericamente como **proteínas não histônicas**, correspondentes sempre a uma fração menor do que a constituída pelas histonas. Como regra geral, as histonas respondem pelos níveis de organização básicos do material genético, ao passo que as proteínas não histônicas estão associadas com a estruturação da cromatina em níveis de organização superiores. São também classificadas como proteínas não histônicas as várias DNA e RNA-polimerases, assim como todas as proteínas reguladoras dos processos de replicação e transcrição.

Dentre as proteínas não histônicas relevantes para a estruturação da cromatina eucariótica podem ser destacadas as proteínas HMG e SMC. As **proteínas HMG** (de *high mobility group* = grupo de alta mobilidade) constituem um grupo de proteínas cromossômicas não histônicas abundantes, de tamanho pequeno (em geral, entre 11 e 38 kDa), de elevada mobilidade eletroforética e com uma longa cauda carboxiterminal negativamente carregada. As proteínas HMG são divididas em 3 famílias,

HMGA, HMGB e HMGN, com base em diferenças nos seus motivos de ligação ao DNA e nos substratos a que se ligam. As proteínas da família HMGA ligam-se, por meio de um motivo chamado *AT-hook* (gancho de AT), aos trechos curtos ricos em AT no DNA. As proteínas da família HMGB possuem domínios de ligação ao DNA com aproximadamente 80 aa (*HMG box*) e interagem com ele sem especificidade de sequência. As proteínas da família HMGN são as únicas que se ligam especificamente aos nucleossomos, sendo caracterizadas por um domínio de ligação nucleossômico (NBD, de *nucleosomal binding domain*) com 30 aa.

As **proteínas SMC** (de *structural maintenance of chromosomes* = manutenção estrutural de cromossomos) constituem uma família de proteínas com propriedades de ligação ao DNA, que formam diferentes complexos multiproteicos envolvidos na organização da cromatina eucariótica. Esses complexos são formados por heterodímeros de proteínas SMC (SMC1-SMC3 ou SMC2-SMC4), que se associam a uma subunidade da superfamília das cleisininas (do grego, κλεισιμο = *kleismo* = fechamento) para formar uma estrutura em anel. Anéis de SMC1-SMC3-cleisina constituem as chamadas **coesinas** e anéis SMC2-SMC4-cleisina constituem as chamadas **condensinas**. Devido à forma dos complexos de coesinas e condensinas, acredita-se que eles atuam na estruturação da cromatina, formando anéis em torno de uma ou mais fitas duplas do DNA.

2.2 Nucleossomos eucarióticos

A unidade estrutural básica da cromatina eucariótica é uma partícula nucleoproteica, formada por DNA e histonas, chamada de **nucleossomo**. Como possuem um número maior de histonas em sua composição, os nucleossomos eucarióticos são maiores e mais complexos que os de arqueas. Cada nucleossomo eucariótico é constituído por um octâmero proteico, formado por duas cópias de cada uma das histonas centrais (H2A, H2B, H3 e H4), e por uma extensão de DNA suficiente para dar duas voltas sobre o octâmero. A esse complexo está ainda associada, externamente, uma molécula de histona de ligação (H1 ou H5). Embora ainda restem dúvidas sobre o local preciso e a forma de associação da molécula da histona de ligação com o nucleossomo, acredita-se que as principais interações entre H1 ou H5 com o DNA ocorram próximas dos pontos de “entrada” e “saída” do ácido nucleico no complexo.

O termo **partícula central** é utilizado para descrever o octâmero de histonas e a extensão de DNA associada a ele. Em ensaios *in vitro*, o DNA presente na partícula central tem uma extensão de 145 a 147 pb, já *in vivo*, a extensão do DNA na partícula central pode variar dentro de limites mais amplos, entre 100 e 170 pb. Essa variação é dinâmica e responde a influências externas, exercidas por outras proteínas nucleares que interagem com a cromatina. É importante salientar que a associação de uma molécula de histona de ligação já determina que de 150 a 250 pb de DNA fiquem estavelmente associados ao nucle-

ossomo, dependendo da variante de H1 ou H5 presente, em cada espécie ou em diferentes tecidos.

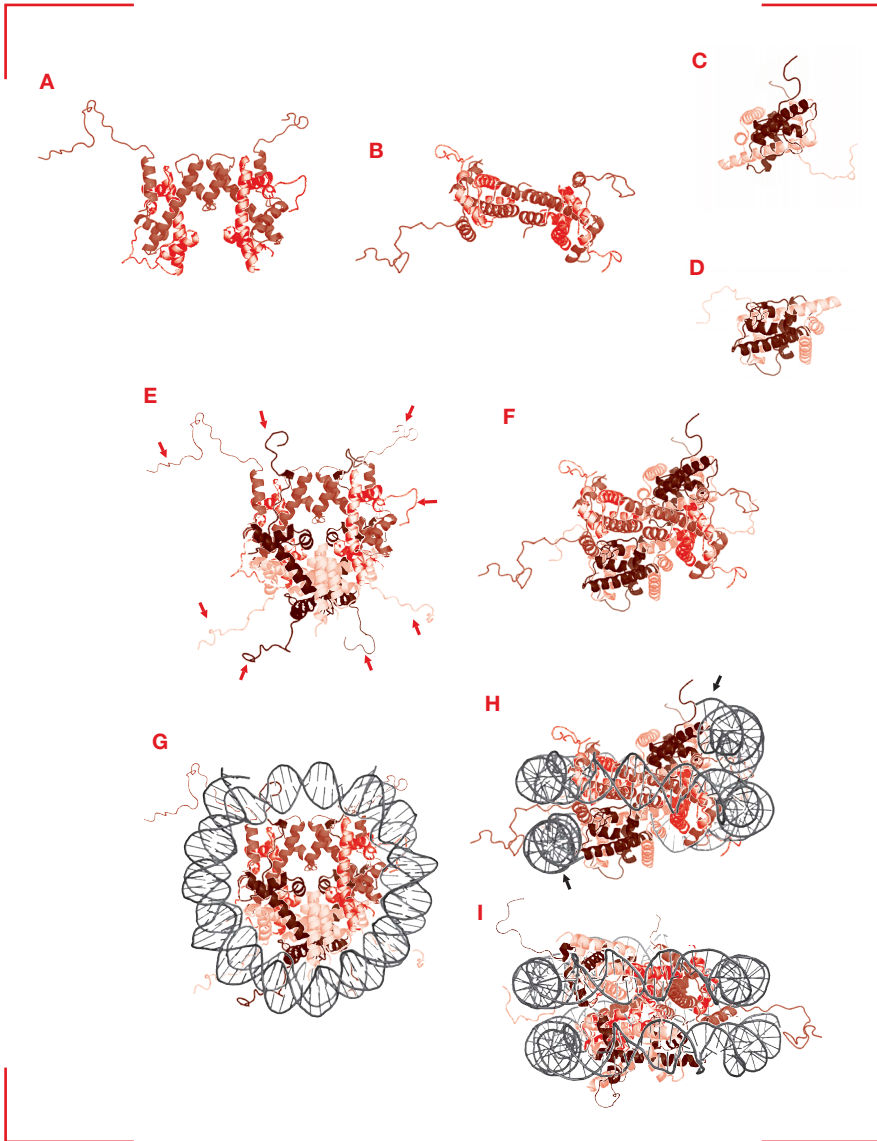
A estrutura molecular da partícula central do nucleossomo está representada na **Figura 3.9**. Três agregados, considerados estáveis, formam o octâmero proteico central: o tetrâmero (H3)₂(H4)₂ (Figura 3.9A-B) e os dois dímeros H2A-H2B (Figura 3.9C-D). O tetrâmero (H3)₂(H4)₂, formado por dois dímeros H3-H4, tem os seus componentes arranjados em uma posição central e aproximadamente em um mesmo plano, acima ou abaixo do qual estão posicionados cada um dos dois dímeros H2A-H2B (Figura 3.9E-F). Essa organização forma uma espécie de super-hélice proteica, sobre a qual o DNA se enrola (Figura 3.9G-I). Essa estrutura confere ao nucleossomo uma forma típica, de cilindro achatado, pelo qual se projetam, até certo ponto, as caudas das histonas centrais, representadas por 15-30 resíduos aminoterminais de cada uma delas. As caudas aminoterminais das histonas também interagem com o DNA, auxiliando na sua estabilização em torno do nucleossomo.

Um importante aspecto estrutural do nucleossomo é o chamado **posicionamento rotacional** do DNA no complexo, que diz respeito à posição da hélice dupla do DNA em relação à superfície do octâmero das histonas centrais (**Figura 3.10**). No nucleossomo, apenas uma das faces da molécula de DNA associada ao octâmero das histonas fica exposta para interagir com outras proteínas. Assim, a interação de uma proteína que se liga a uma sequência nucleotídica específica no DNA, como um fator de transcrição (ver Capítulo 14), depende do posicionamento do respectivo sítio de ligação (sua sequência-alvo), na face exposta da hélice dupla.

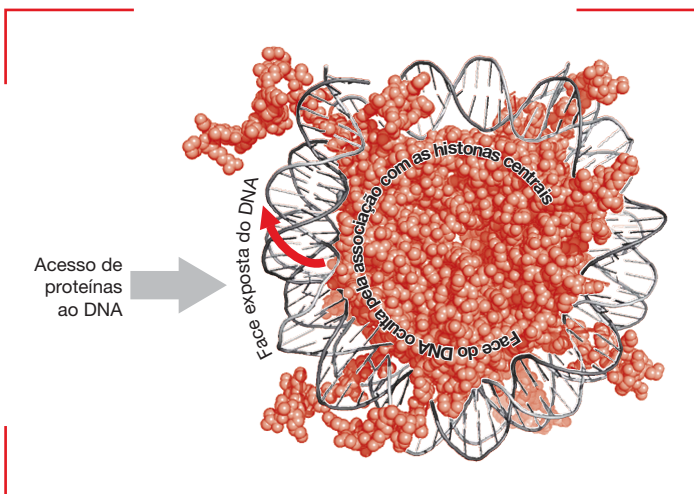
O posicionamento rotacional do DNA no nucleossomo é influenciado por muitos fatores, incluindo certas sequências nucleotídicas (que especificam estruturas na molécula de DNA mais ou menos favoráveis à interação com as histonas), contatos entre nucleossomos vizinhos e ação de complexos de remodelação da cromatina. Sobre a influência de sequências nucleotídicas, já foi demonstrado que a ocorrência de dinucleotídeos purínicos (AG ou GA), regularmente espaçados a cada 10 pb em um segmento de DNA, é um determinante de posicionamento rotacional nucleossômico importante. Esse tipo de organização periódica de dinucleotídeos purínicos é encontrado, por exemplo, em torno dos sítios de início de transcrição de muitos genes, sendo responsável pelo posicionamento rotacional do DNA nos nucleossomos dessas regiões, que incluem os sítios de ligação de muitos fatores de transcrição.

2.3 Níveis de organização mais complexos da cromatina eucariótica

A partir da sua subunidade estrutural básica (nucleossomo), a cromatina eucariótica é estruturada em conformações progressivamente mais complexas, que vão desde estruturas fibrilares menos compactas até estruturas mais enoveladas e compactas, cuja organização em nível

**Figura 3.9**

Estrutura típica da partícula central do nucleossomo. O tetrâmero formado por duas moléculas de H3 (em ■) e duas moléculas de H4 (em ■) determinam um plano, que, em (A), está posicionado perpendicularmente à superfície da página e, em (B), tem uma inclinação de aproximadamente 30°. Dois dímeros H2A-H2B, mostrados em (C) e (D), com H2A em ■ e H2B em ■, se posicionam um acima e outro abaixo do plano definido por $(H3)_2(H4)_2$, formando o octâmero mostrado em (E) (vista superior), e em (F) (vista lateral). Em volta do octâmero de histonas enrolam-se 146 pb de DNA, em aproximadamente duas voltas, formando a partícula central do nucleossomo (G-I). Em (G), a partícula central é apresentada (em uma vista superior), equivalente à do octâmero mostrado em (E). Em (H), a partícula central é apresentada em uma vista lateral, equivalente à do octâmero mostrado em (F); nesta vista, ficam evidentes os pontos de “entrada” e “saída” do DNA no nucleossomo (indicados por setas em ■). A vista lateral em (I) mostra o lado oposto da partícula central mostrada em (H), evidenciando as duas voltas que o DNA dá em torno do octâmero de histonas. Em todas as estruturas mostradas, ficam evidentes as caudas aminoterminais das histonas, que, em (E), estão indicadas por setas em ■.

**Figura 3.10**

Posicionamento rotacional do DNA no nucleossomo. A hélice dupla do DNA (em ■) fica, no nucleossomo, com uma de suas faces (a oculta) associada ao octâmero das histonas centrais (em ■) e a outra (a exposta) acessível para interações com proteínas que se ligam especificamente a sequências do DNA (indicadas pela seta em ■). Se a sequência nucleotídica reconhecida por uma determinada proteína estiver na face oculta do DNA, a interação apenas ocorrerá se o segmento do DNA que a contém for para a face exposta, em função de uma alteração no grau de torção da hélice dupla (rotação sobre seu próprio eixo longitudinal, representada pela seta em ■).

molecular ainda não foi esclarecida por completo. A organização da cromatina é dinâmica, pois se altera conforme o estágio do ciclo celular e de acordo com a sua atividade.

Durante a interfase, dois tipos básicos de cromatina podem ser observados em nível de microscopia óptica: a **eucromatina**, que representa a cromatina no seu estado de menor compactação; e a **heterocromatina**, uma forma de cromatina densamente compactada. Durante a divisão celular (mitose ou meiose), toda a cromatina encontra-se muito mais condensada, formando os **cromossomos mitóticos ou meióticos**, que são unidades discretas e bastante compactas. Quanto ao seu estado funcional, a eucromatina é considerada a cromatina ativa (em termos de expressão gênica), ao passo que a heterocromatina e os cromossomos constituem suas formas inativas.

O nível de compressão do DNA em cada nível de organização da cromatina pode ser quantificado pela chamada **razão de compactação** (*packing ratio*). Essa grandeza é definida pela razão entre o comprimento de uma molécula de DNA e o comprimento da unidade que a contém. Por exemplo, o menor dos cromossomos humanos (cromossomo 21) possui em torno de $4,7 \times 10^7$ pb de DNA e, estendido, teria um comprimento de aproximadamente 1,4 cm. Assim, como esse cromossomo tem, na mitose, um comprimento de 2 μm , a sua razão de compactação chega a 7.000. A razão de compactação, contudo, pode variar de cromossomo para cromossomo e de espécie para espécie e é ainda mais difícil de ser estimada para estruturas amorfas, como a da cromatina interfásica. Em geral é aceito que cromossomos mitóticos têm razões de compactação de 5.000 a 20.000 e que a cromatina interfásica tem uma estrutura de 5 a 10 vezes menos compacta, com razões de compactação oscilando entre 1.000 e 2.000.

A organização da cromatina em níveis progressivamente mais complexos e compactos, desde as suas fibras mais simples até as estruturas cromossômicas, está sumarizada nas **Figuras 3.11 e 3.12**.

2.3.1 Fibra de 10 nm

O DNA livre tem um diâmetro de aproximadamente 2 nm (ver Capítulo 2), mas, no contexto da cromatina, é quase inexistente. Devido a sua associação com histonas, a forma menos compacta do DNA na cromatina corresponde a um arranjo de nucleossomos em tandem, formando uma estrutura similar a um colar de contas (Figura 3.11A). Essa estrutura linear, cujo diâmetro corresponde ao do próprio nucleossomo (de aproximadamente 10 nm), representa o primeiro nível de compactação da cromatina e é chamada de **fibra de 10 nm**.

Nesse primeiro nível de compactação da cromatina, os nucleossomos adjacentes estão conectados entre si pela continuidade da molécula de DNA que, entre nucleossomos e livre da associação com histonas, é chamado de **DNA de ligação** (*linker DNA*). A estrutura da fibra de 10 nm depende apenas de interações entre nucleossomos

adjacentes, em que a extensão de cada trecho do DNA de ligação pode variar conforme a espécie ou até mesmo o tecido ou tipo celular. Extensões de 20 a 80 pb ocorrem em leveduras e em espermatozoides de equinodermos, e extensões médias de aproximadamente 35 pb são típicas para vertebrados. A estrutura da fibra de 10 nm é afetada pelas histonas de ligação (H1 ou H5), que, por mediarem interações entre os nucleossomos, são determinantes do espaçamento entre eles.

A fibra de 10 nm da cromatina tem duas propriedades importantes. A primeira é a **periodicidade**, que decorre da distribuição regular dos nucleossomos a cada 200 ± 50 pb ao longo de praticamente toda a extensão do DNA nuclear. Essa periodicidade pode ser demonstrada a partir de experimentos simples, nos quais é feita uma clivagem parcial da cromatina com uma desoxirribonuclease (**ver Material Complementar Online 3.5**). Esses experimentos são fundamentados na sensibilidade diferencial do DNA à ação da enzima, quando associado ou não aos nucleossomos. A outra propriedade estrutural importante, identificada na fibra de 10 nm, é o **posicionamento em fase** dos nucleossomos. Essa propriedade determina que, ao longo do genoma de uma espécie, a maior parte dos nucleossomos esteja posicionada da mesma maneira, em relação à sequência nucleotídica de todas as células. Em *S. cerevisiae*, por exemplo, estima-se que uma fração de 80% do genoma é ocupada por nucleossomos posicionados em fase.

O posicionamento em fase de nucleossomos pode resultar da associação específica das histonas ou de seus complexos (dímero, tetrâmero ou octâmero – ver Seção 2.2) com determinadas sequências nucleotídicas, caso em que é chamado de **intrínseco**. Isso ocorre porque, apesar da aparente inespecificidade da interação das histonas com o DNA, os nucleossomos se formam preferencialmente sobre determinadas sequências de DNA, e evitam outras. Por exemplo, os nucleossomos têm maior probabilidade de se formarem sobre regiões com curvatura orientada (sempre na mesma direção) no DNA, definidas por segmentos ricos em pares de bases AT considerados longos (mais de 8 pb). Em contrapartida, trechos homopoliméricos de nucleotídeos de desoxiadenosina em uma das fitas da hélice dupla parecem excluir nucleossomos. Motivos de sequência como esses, disseminados em genomas eucarióticos, definiriam uma espécie de “código genômico”, parcialmente responsável pelo posicionamento de nucleossomos em fase.

O posicionamento em fase, também pode acontecer de maneira **extrínseca**, quando apenas um nucleossomo de determinada região é posicionado em um sítio preferencial, por afinidade com alguma sequência ou estrutura na molécula de DNA, sendo os demais montados a partir dele, obedecendo a periodicidade característica de cada um. Isso ocorre, também, a partir de regiões adjacentes aos sítios em que a montagem do DNA em um nucleossomo é evitada, pela ligação (geralmente reversível) de uma proteína não histônica a uma sequência nucleotídica específica (como no caso da ligação de fatores de trans-

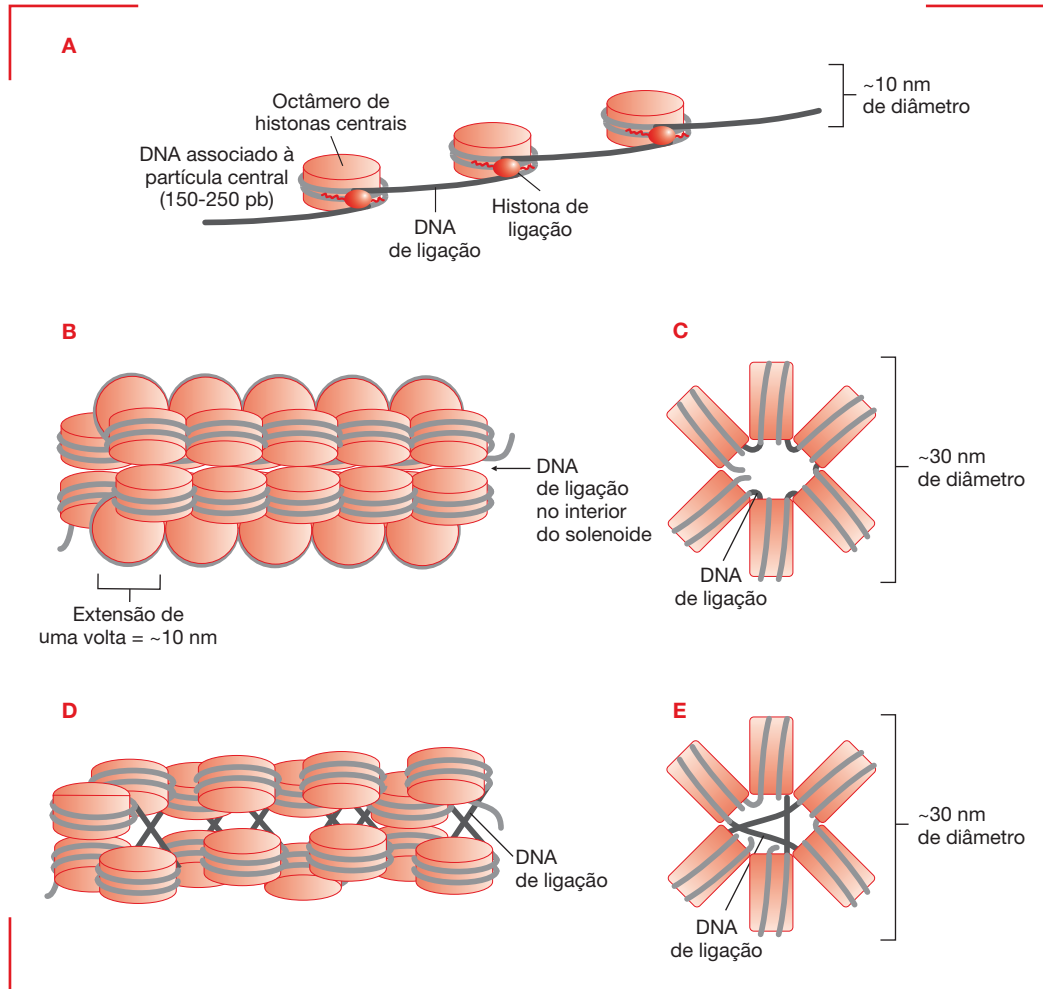


Figura 3.11

Fibras de 10 nm e 30 nm da cromatina eucariótica. (A) Representação esquemática de nucleossomos organizados em uma fibra de 10 nm de diâmetro. O octâmero de histonas centrais de cada nucleossomo (representado como um cilindro achatado em ■) e o DNA associado a ele (representado em ■) constituem a chamada partícula central. Uma molécula de uma histona de ligação (H1 ou H5, em ■) fica associada ao nucleossomo externamente, na região de “entrada” e “saída” do DNA da partícula. O DNA de ligação (representado em ■), entre nucleossomos adjacentes, pode ter extensões variáveis (de 20 a 80 pb) na fibra de 10 nm. (B-E) Modelos da fibra de 30 nm (para simplificação, as moléculas de histonas de ligação foram omitidas de todas as representações). No modelo em solenoide (B-C), é proposta uma estrutura super-helicoidal, com 6 nucleossomos por volta. No modelo solenoidal, o DNA de ligação não cruza o eixo longitudinal central da fibra, como pode ser observado em uma vista transversal (C). No modelo em zig-zague (D-E), os nucleossomos estão arranjados em duas fileiras alternadamente. O DNA de ligação tem, nesse modelo, uma trajetória mais ou menos retilínea e cruza várias vezes o espaço entre as duas fileiras de nucleossomos, como pode ser melhor observado em uma vista transversal (E).

crição a regiões promotoras) ou pela ligação do DNA a estruturas nucleares, como a própria matriz nuclear (ver Seção 2.3.3). Essas regiões definiriam limites (posições específicas no genoma), a partir dos quais os nucleossomos seriam posicionados em fase.

Os mecanismos de posicionamento dos nucleossomos possuem importantes implicações funcionais na fisiologia da cromatina. Um arranjo em fase, assim

determinado, expõe certas sequências de DNA e oculta outras, facilitando, ou dificultando, sua interação com as proteínas reguladoras de processos como a replicação e a transcrição. Na caracterização da acessibilidade de uma sequência a proteínas reguladoras no contexto da fibra de 10 nm, pode ser definido, além do posicionamento rotacional inerente à estrutura de cada nucleossomo (ver Seção 2.2), o chamado **posicionamento translacional**, que descreve um determinado

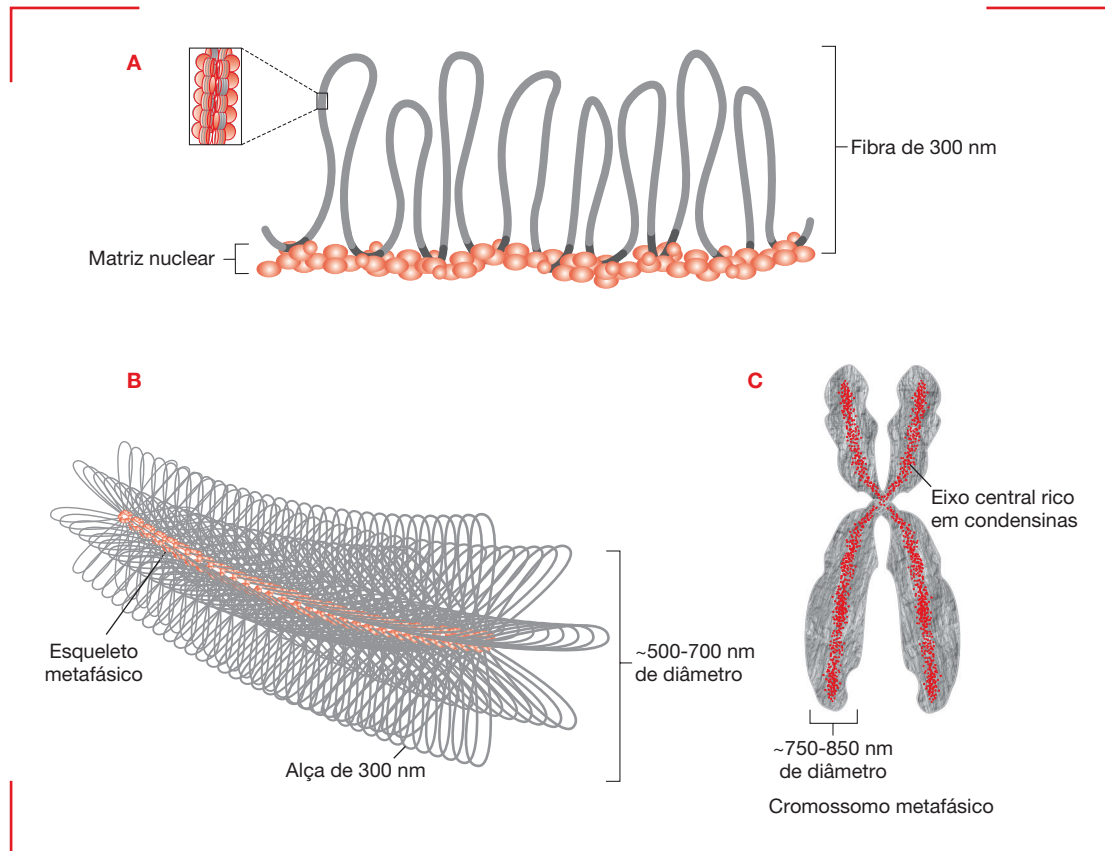


Figura 3.12

Estruturas de cromatina de ordem superior. (A) Representação esquemática das alças de cromatina formadas a partir da associação da fibra de 30 nm (em ■) com proteínas da matriz nuclear (em ■). A estrutura da fibra de 30 nm é mostrada no detalhe, e os seus segmentos, incluindo MARS, estão representados em ■. (B) Na condensação da cromatina, as alças da fibra de 300 nm são compactadas de modo a formar fibras de cromatina de ordem superior. Como representado na figura, as alças de cromatina formam um arranjo helicoidal em torno do esqueleto metafásico axial, com várias alças de 300 nm por volta, de modo a formar fibras com diâmetros de 500 a 700 nm. (C) A estrutura do cromossomo metafásico formada a partir do enovelamento da fibra de 300 nm e de outras fibras de ordem superior apresenta em cada braço do cromossomo aproximadamente 700 a 850 nm de diâmetro. A região axial de cada braço cromossômico é rica em condensinas (complexos protéicos que incluem proteínas SNC), representadas na figura por pontos em ■.

segmento do DNA em relação aos limites externos da partícula central. Isto é, dependendo do seu posicionamento translacional, uma sequência de DNA é parte de um nucleossomo ou do DNA de ligação. Para esse processo são também úteis os experimentos de clivagem parcial da cromatina, discutidos no **Material Complementar Online 3.5**.

Limitações diferenciais de acesso, determinadas pela cromatina para a ação da enzima, permitem demonstrar que a cromatina pode ter uma configuração aberta ou uma configuração fechada. A **configuração aberta**, mais acessível à ação de nucleases (mas também de quaisquer proteínas que interajam com o DNA) é a da fibra de 10 nm, ao passo que a **configuração fechada** é encontrada na fibra de 30 nm, ou em estruturas mais compactas da cromatina.

Proteínas reguladoras podem ter diferentes exigências quanto à posição do DNA em relação ao nucleossomo, inclusive na configuração aberta da fibra de 10 nm. Algumas proteínas são capazes de interagir apenas com o DNA livre da associação com histonas (no DNA de ligação, entre dois nucleossomos), sendo, para isso, fundamental o posicionamento translacional. Outras proteínas, como discutido na Seção 2.2, são capazes de interagir com sequências específicas do DNA expostas na superfície do nucleossomo, sendo fundamental, neste caso, o posicionamento rotacional.

2.3.2 Fibra de 30 nm

É provável que a fibra de 10 nm só ocorra *in vivo* e em regiões específicas, nas quais haveria uma remoção localizada das moléculas de histonas de ligação dos nucle-

ossomos. Em regiões nas quais os nucleossomos estão completos, com uma histona de ligação associada externamente a cada um deles, a cromatina tende a adotar uma conformação mais complexa, com um diâmetro de 30 nm em média.

A estrutura tridimensional dessa fibra, chamada de **fibra de 30 nm**, é controversa. Alguns modelos, como o apresentado na Figura 3.11B-C, propõem que ela teria um arranjo solenoidal, no qual cada espira conteria 6 nucleossomos organizados radialmente. Já outros modelos, como o apresentado na Figura 3.11D-E, propõem um padrão em zigue-zague para a organização dos nucleossomos na fibra. No modelo em zigue-zague, são formadas duas fileiras de nucleossomos ao longo da fibra, que podem ser torcidas em diferentes arranjos helicoidais. A acessibilidade do DNA para a interação com as proteínas também difere nos dois modelos da fibra de 30 nm. No modelo solenoidal, o DNA de ligação fica pouco exposto e relativamente inacessível no interior da fibra, pois não passa através do seu eixo longitudinal (ver Figura 3.11B). Já no modelo em zigue-zague, o DNA de ligação cruza o eixo longitudinal da fibra e, por isso, fica mais exposto (ver Figura 3.11E).

Tanto o modelo solenoidal como o em zigue-zague se baseiam em observações experimentais e são coerentes com propriedades já demonstradas para a fibra de 30 nm. Esses modelos, bem como suas variações, não são mutuamente exclusivos e podem representar diferentes configurações para a fibra de 30 nm no contexto da cromatina. Sabe-se que, *in vivo*, a estrutura da fibra de 30 nm não é homogênea, pois possui uma organização mais frouxa e uma aparência mais irregular do que as propostas em qualquer outro modelo, com um diâmetro variável ao longo de toda a sua extensão.

A estabilidade estrutural da fibra de 30 nm depende tanto das histonas centrais como das histonas de ligação dos nucleossomos. As histonas centrais contribuem para a integridade da fibra graças a suas caudas aminoterminais, que promovem interações adicionais entre os nucleossomos. As histonas de ligação, por sua vez, são fundamentais para a definição da extensão do DNA de ligação (um dos fatores que afetam a conformação da fibra) e também para o estabelecimento de interações entre os nucleossomos adjacentes. Na dinâmica da fibra de 30 nm, acredita-se que há participação importante das proteínas HMG. Membros das 3 famílias das proteínas HMG (HMGA, HMGB e HMGN) competem com as histonas de ligação pela associação aos nucleossomos ou ao DNA de ligação e determinam uma dinâmica competitiva, responsável pela modulação da estrutura local da fibra de cromatina e da acessibilidade do DNA a ela associado para interação com outras proteínas.

2.3.3 Alças de cromatina e fibra de 300 nm

A organização da cromatina em fibras de 10 e 30 nm permite que sejam atingidas razões de compactação de no máximo 40, muito aquém do necessário para acomodar um genoma eucariótico no interior do núcleo. Assim, são

necessários níveis de compactação adicionais, nos quais as fibras de 10 e 30 nm se organizam em estruturas de ordem superior. Ainda se sabe pouco, em nível molecular, sobre a condensação da cromatina em níveis de compactação superiores ao da fibra de 30 nm, mas geralmente é aceito que esta fibra forme **alças de cromatina** (Figura 3.12A).

As bases de cada alça de cromatina ficam ancoradas em estruturas proteicas, podendo variar em composição e estrutura conforme o estado fisiológico da cromatina (eucromatina ou heterocromatina), ou o estágio do ciclo celular (na interfase ou durante a divisão). Durante a interfase, as alças de cromatina ficam ancoradas na chamada **matriz nuclear**, que é a estrutura proteica filamentosa que preenche o núcleo. Durante a divisão celular, as alças aparecem ancoradas em um arcabouço proteico mais estruturado (ou, pelo menos, mais evidente citologicamente), chamado de **esqueleto metafásico**, no caso de cromossomos metafásicos.

A extensão de DNA em cada alça pode variar conforme o tipo celular, o tecido, ou a espécie considerada, mas, em geral, oscila entre 40 e 90 kb. Isso corresponde de 200 a 450 nucleossomos, que, assumindo a organização da fibra de 30 nm, com 6 nucleossomos a cada 10 nm (extensão de uma volta do solenoide – ver Figura 3.11B), equivalem a uma extensão linear aproximada de 300 a 700 nm para cada alça. Por isso, costuma-se definir a fibra formada por essas alças como **fibra de 300 nm** (de diâmetro). Na fibra de 300 nm, o DNA fica de 15 a 20 vezes mais compactado do que na fibra de 30 nm, ou seja, ele atinge razões de compactação da ordem de 600 a 700.

A estruturação das alças na fibra de 300 nm depende de segmentos do DNA que estabelecem interações com a matriz nuclear ou com o esqueleto metafásico (Figura 3.12A-B). Esses segmentos são operacionalmente definidos como sequências de DNA que se associam a proteínas da matriz nuclear ou do esqueleto metafásico no contexto da cromatina, por isso, são chamados de **MARs** (de *matrix attachment regions* = regiões de ligação à matriz) ou **SARs** (de *scaffold attachment regions* = regiões de ligação ao arcabouço) respectivamente. As SARs e as MARs são regiões consideradas extensas, ricas em pares de bases AT (representando até 70% das suas bases), mas sem qualquer sequência consensual específica. Mais precisamente, elas são compostas por inúmeras sequências repetidas de poli(A), espaçadas (sem regularidade) ao longo de várias centenas de pares de bases. Assim, é provável que as proteínas organizadoras da cromatina não reconheçam sequências específicas nas SARs e MARs, mas sim certas características estruturais determinadas por elas no DNA, como as regiões de curvatura acentuada ou mais propensas à separação das fitas.

A quantidade de MARs ou SARs ao longo de qualquer genoma eucariótico é muito grande, estima-se, por exemplo, um número de pelo menos 150.000 para o genoma do camundongo, considerando estudos realizados em hepatócitos. Sabe-se, entretanto, que o número de MARs ou SARs efetivamente envolvido em interações com a matriz

nuclear varia conforme o estado fisiológico da cromatina, o tipo celular ou o tecido. Essas variações implicam também, em modificações dinâmicas para o número e o tamanho médio das alças de cromatina formadas e as razões de compactação alcançadas pelo DNA na estrutura. Por isso, foi introduzido recentemente o conceito funcional de **LARs** (de *loop attachment regions* = regiões de ligação de alças), para definir, dentre todas as potenciais MARs ou SARs existentes em um genoma, as que possuem envolvimento efetivo no ancoramento das alças de cromatina, em análise de um determinado tipo celular, tecido ou situação fisiológica.

A composição proteica da matriz nuclear é a principal responsável pela definição da estrutura da cromatina durante a interfase. A caracterização estrutural da cromatina, contudo, é dificultada por ela ser essencialmente amorfa e ter na sua composição, pelo menos, algumas centenas de proteínas diferentes (mais de 400, segundo estimativas recentes). Dentre os componentes proteicos principais da matriz nuclear, cabe citar os **SMFs** (de *scaffold attachment factors* = fatores de ligação ao arcabouço) e a topoisomerase II.

Os SMFs constituem uma família de proteínas da matriz nuclear com afinidade por MARs e SARs, propriedade que faz com que alguns autores classifiquem esses fatores como componentes da cromatina. Os SMFs, além de serem propostos como elementos organizadores da cromatina, participam de vários outros processos nucleares, incluindo a regulação transcricional. A topoisomerase II (ver Capítulo 2), por sua vez, é uma enzima cuja presença é significativa em estruturas de cromatina de ordem superior. Isso sugere que mudanças no superenrolamento do DNA podem ser críticas no processo de compactação da cromatina.

2.3.4 Estruturas de cromatina de ordem superior e cromossomos metafásicos

A organização da cromatina em estruturas mais compactas que a da fibra de 300 nm ocorre na transição de conformações interfásicas para aquelas representadas pelos cromossomos individualizados na divisão celular. Essa transição envolve mudanças tanto nas sequências de DNA como nas proteínas envolvidas, pois, tanto o repertório de LARs como o de proteínas às quais elas se ancoram mudam quantitativa e qualitativamente. Em geral, assume-se que, no processo de compactação envolvido, a fibra de 300 nm seja espiralada em estruturas (fibras) de ordem superior, que teriam uma estrutura similar à representada na Figura 3.12B. Entretanto, sabe-se que o enrolamento da fibra de 300 nm em estruturas mais complexas não é tão simétrico e homogêneo, de modo que as fibras de ordem superior da cromatina podem ter conformações variáveis.

Modelos recentes para a estrutura dos cromossomos propõem que a fibra de 30 nm se enovelaria, de forma progressiva e até certo ponto desordenada, para formar estruturas com diâmetros de 100 a 150, 200 a 250 e 500

a 750 nm (Figura 3.12C). Estruturas cromossômicas alternativas, como as de **cromossomos plumosos**, observados em anfíbios, e **cromossomos politênicos**, encontrados em insetos dípteros, são discutidas no **Material Complementar Online 3.6**.

O arcabouço proteico que sustenta as estruturas de cromatina de ordem superior, chamado de esqueleto metafásico, apresenta uma composição proteica distinta daquela da matriz nuclear, embora ambas as estruturas compartilhem algumas proteínas em comum, como a própria topoisomerase II. Assim como a matriz nuclear, o esqueleto metafásico é bastante complexo e dinâmico. Ele é composto por dezenas de proteínas diferentes e a sua composição proteica pode variar ao longo do ciclo de divisão celular. Com isso, são determinadas as diferentes estruturas e níveis de compactação observados nos processos de formação e de desestruturação de cromossomos, que ocorrem entre a prófase e a metáfase e após a telófase.

Dentre as proteínas que participam da organização de estruturas de cromatina de ordem superior, inclusive de cromossomos metafásicos, pode-se salientar HMGs da família HMGA e proteínas SMC (ver Seção 2.1). As proteínas HMGA ligam-se a sequências com características de SARs (ricas em AT e repetidas) e estudos indicam que são componentes integrais da cromatina. Em cromossomos metafásicos, proteínas dessa família aparecem em posições definidas e podem estar ativamente envolvidas nas alterações estruturais dinâmicas que ocorrem na cromatina durante o processo de condensação cromossômica a cada ciclo de divisão celular. As proteínas SMC, por sua vez, formam, juntamente com outras proteínas, os complexos chamados de condensinas e coesinas, que participam dos processos de condensação cromossômica e de adesão de cromátides-irmãs. Em cromossomos metafásicos (ver Figura 3.12C), complexos de condensina são observados nos eixos axiais da estrutura, evidenciando sua importância para a manutenção da estrutura cromossômica. Além disso, já foi demonstrado (em experimentos) que a presença de condensinas e de proteínas SMC é fundamental para que os cromossomos sejam condensados de maneira correta.

3. Cromatina de organelas

Plastídeos e mitocôndrias são organelas celulares que possuem genomas próprios, constituídos por moléculas de DNA circulares pequenas, em geral com dezenas ou poucas centenas de quilobases e, que estão presentes em múltiplas cópias em cada organela (Capítulo 4). Nas organelas, as múltiplas cópias dos genomas estão associadas com proteínas, constituindo uma forma de cromatina, organizadas em nucleóides. Entretanto, é importante salientar que, apesar da origem evolutiva procariótica das organelas, seus nucleóides não são meros remanescentes da organização da cromatina bacteriana ancestral. Os nucleóides de plastídeos e mitocôndrias são sistemas

nucleoproteicos complexos e com arquitetura própria, estabelecidos, ao longo de sua evolução, a partir da extensa interação entre genomas organelares e o genoma nuclear.

3.1 Nucleoides mitocondriais

O DNA mitocondrial (**mtDNA**), presente em múltiplas cópias por mitocôndria, forma nucleoides discretos no interior da organela a partir da sua associação com proteínas estruturais específicas. O número de nucleoides por mitocôndria varia de dezenas a centenas, dependendo da espécie, do tipo celular, do estágio de diferenciação da célula ou do estágio de desenvolvimento do organismo. Cada nucleóide, por sua vez, contém de 2 a 10 cópias do mtDNA e comporta-se como um elemento genético de estrutura estável. O grau de compactação do mtDNA em um nucleóide mitocondrial é equivalente ao de um nucleóide bacteriano.

Mais de 50 proteínas diferentes podem ser associadas ao mtDNA em nucleoides mitocondriais, mas, dentre elas, apenas algumas seriam relevantes para a compactação e estabilização do mtDNA. É possível que o componente proteico estrutural mais importante para a cromatina mitocondrial seja uma proteína relacionada ao grupo HMG (ver Seção 2.1). Esse grupo está representado em mitocôndrias por proteínas como a **TFAM**, de mamíferos, caracterizada pela presença de domínios de ligação ao DNA do tipo *HMG box*, encontrados em proteínas HMG da família HMGB. A TFAM (de *transcription factor A of mitochondria* = fator de transcrição A de mitocôndrias), foi inicialmente caracterizada como um fator de transcrição mitocondrial, mas acredita-se que ela, assim como as proteínas ortólogas **Abf2p**, de leveduras, e **Glom** (de *agglomeration of mitochondrial DNA* = aglomeração de DNA mitocondrial), de *Dictyostelium discoideum* (Mycetozoa; Amoebozoa), têm como função primária a formação e a estabilização da cromatina mitocondrial. Todas essas proteínas formam homodímeros que se ligam ao mtDNA a cada 35 ou 40 pb e são essenciais para a compactação do mtDNA.

Outra proteína, que possivelmente está envolvida na organização estrutural do mtDNA é a proteína de ligação ao DNA de fita simples. Essa **mtSSB** (de *mitochondrial single-stranded DNA-binding protein* = proteína mitocondrial de ligação a DNA de fita simples) teria uma participação mais localizada que a da proteína HMG mitocondrial. Devido a sua especificidade estrutural de ligação ao DNA de fita simples, ela seria responsável, pelo menos em parte, pela estabilização da estrutura da alça em D (*D-loop*) do mtDNA, um elemento de fita simples importante para a sua replicação.

A participação de produtos do genoma nuclear na organização e na manutenção do nucleóide mitocondrial é evidenciada pela presença da proteína Dna2 entre aquelas que são identificadas como ligantes diretas do mtDNA. A **Dna2**, codificada por um gene nuclear, é uma DNA-helicase importante, tanto para a replicação como para o reparação do mtDNA. Assim como a DNA-helicase **Twinkle**, codificada por um gene mitocondrial, a Dna2 está entre as proteínas mais representadas em nucleoides de mitocôndrias.

3.2 Nucleoides de plastídeos

O DNA de cloroplastos e outros plastídeos (**ptDNA**), assim como o mtDNA, associa-se a proteínas específicas para formar nucleoides no interior das organelas. Em cada um dos diversos nucleoides que ocorrem por organela, estão organizadas de 3 a mais de 10 cópias do ptDNA. Os nucleoides de plastídeos podem variar em morfologia e distribuição na organela, conforme o tecido ou o estágio de desenvolvimento de uma planta. Em células meristemáticas (indiferenciadas), cada proplastídeo (precursor de um cloroplasto ou outro tipo de plastídeo) tem apenas um nucleóide, ocupando uma posição central na organela. Durante a diferenciação de proplastídeos em cloroplastos, os nucleoides, então, em grande número, aparecem posicionados na periferia da organela em contato com a sua membrana interna. Em cloroplastos maduros, os nucleoides ficam livres, como pequenas partículas discretas no interior da organela. Por fim, em células senescentes, o número de nucleoides por organela, assim como o número de cópias do ptDNA por nucleóide, tende a decrescer.

Pelo menos dois sistemas distintos foram estabelecidos para a organização dos nucleoides plastidiais ao longo das linhagens evolutivas dos 4 supergrupos de eucariotos, que hoje apresentam filós com espécies contendo plastídeos (Plantae, Chromalveolata, Excavata e Rhizaria – ver Capítulo 5). Um desses sistemas ocorre em algumas espécies de protozoários e algas e o outro, em eucariotos fotossintetizantes. A distinção entre eles é feita com base nas proteínas responsáveis pela compactação do ptDNA.

Apicoplastos (plastídeos especializados) de espécies do filo Apicomplexa (Chromoalveolata), como os do gênero *Plasmodium*, e plastídeos de algas como *Cyanidios-chyzon merolae* (Rodophyta, Plantae) e *Guillardia theta* (Cryptophyta, Chromalveolata), têm como principal elemento estruturante de seus nucleoides uma proteína ortóloga da HU bacteriana (ver Seção 1.1). A HU plastidial pode ser codificada tanto pelo ptDNA como pelo genoma nuclear, dependendo da espécie. Por exemplo, em *C. merolae* e *G. theta*, a HU é produto de um gene do ptDNA, ao passo que, em *Plasmodium falciparum*, a HU é produto de um gene nuclear, mas que, após sua síntese, é direcionada para o apicoplasto. Independentemente da localização do gene que a codifica, qualquer HU plastidial liga-se ao ptDNA sem especificidade de sequência e ocasiona o seu dobramento e compactação, sendo essencial para a formação de nucleoides plastidiais, em organismos que possuem essa proteína.

Plastídeos de eucariotos fotossintetizantes, por sua vez, incluindo os de algas verdes (Chlorophyta, Plantae) e plantas terrestres (Embryophyta, Plantae), desenvolveram, ao longo de sua evolução, um sistema de compactação do ptDNA sem a participação da HU. Dentre as várias proteínas que se ligam ao ptDNA e que possivelmente estão envolvidas nesse sistema, podem ser destacadas três: a enzima sulfito-redutase, a protease CND41 e a proteína PEND. A **sulfito-redutase (SiR)**, além de atuar no estroma de cloroplastos no processo de assimilação de enxofre, liga-se ao ptDNA e induz a sua compactação de forma reversível. Essa enzima é a proteína de ligação ao DNA mais

abundante em nucleoides plastidiais. A protease **CND41** (de *41-kDa chloroplast nucleoid DNA-binding protein* = proteína de ligação a nucleoides de cloroplastos de 41 kDa) também é considerada uma proteína bifuncional, como a SiR, pois, além de atuar com sua atividade proteolítica de aspartil-peptidase ácida nos processos de senescência foliar, liga-se ao ptDNA e o compacta, reprimindo a transcrição gênica plastidial. A CND41 contém motivos de ligação ao DNA do tipo hélice-volta-hélice e dedo de zinco (ver Capítulo 14) e associa-se ao ptDNA sem especificidade

de sequência. A **PEND** (de *plastid envelope DNA-binding* = ligação do DNA ao envoltório plastidial), por sua vez, é uma proteína que se associa simultaneamente à membrana interna do envoltório plastidial e ao ptDNA, ancorando os nucleoides de plastídeos em desenvolvimento. A PEND é uma proteína de 70 kDa, com um domínio de ligação ao DNA, (cbZIP), na sua extremidade aminoterminal, e um domínio hidrofóbico transmembranar, na sua região carboxiterminal. Ela forma homodímeros que se ligam a sequências TAA-GAAGT no ptDNA.



Resumo

As moléculas de DNA que constituem os genomas celulares ocupam, em sua forma estendida, volumes maiores do que os dos compartimentos que as contêm. Para que essas moléculas sejam acomodadas em células procarióticas, no núcleo, ou em organelas de células procarióticas, elas precisam ser compactadas, constituindo o que se chama de cromatina. Na cromatina, o DNA é compactado de maneira organizada e reversível, de modo que, conforme a demanda fisiológica da célula, ele pode ser descompactado para permitir a ocorrência de processos como a replicação e a transcrição.

Diferentes sistemas de compactação do DNA são utilizados em bactérias, arqueas e no núcleo e organelas de células eucarióticas. Em bactérias, o complexo nucleoproteico da cromatina é formado pela associação do cromossomo com as chamadas proteínas associadas ao nucleóide (NAPs). As NAPs organizam o nucleóide bacteriano em múltiplas alças, com o DNA associado a proteínas e superenrolado negativo em cada uma delas. Em arqueas, o repertório de NAPs inclui tanto proteínas exclusivas do domínio Archaea, como proteínas ortólogas de NAPs bacterianas ou

de proteínas cromatínicas eucarióticas. Algumas das NAPs de arqueas polimerizam sobre o DNA e formam fibras nucleoproteicas similares à da cromatina eucariótica. Em células eucarióticas, a cromatina nuclear está organizada em nucleossomos, que são partículas formadas por um octâmero de histonas em torno das quais o DNA dá duas voltas. Uma histona de ligação associa-se externamente ao complexo octâmero-DNA e envolve-se em interações entre nucleossomos.

Durante o ciclo celular, há uma transição entre diferentes estados de compactação da cromatina, desde a sua conformação mais aberta, na interfase, até a formação de cromossomos metafásicos, que representam seu estado de condensação máxima. Nessa dinâmica de compactação, as fibras de cromatina formadas pela sequência linear de nucleossomos condensam-se progressivamente e formam fibras de maior diâmetro e alças cromossômicas. Mitocôndrias e plastídeos têm os seus genomas compactados em nucleoides por proteínas, que são exclusivas de cada organela e diferem tanto das NAPs procarióticas como das proteínas da cromatina nuclear.



Leituras recomendadas

Dillon SC, Dorman CJ. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol.* 2010;8(3):185-95.

Luijsterburg MS, Noom MC, Wuite GJ, Dame RT. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J Struct Biol.* 2006;156(2):262-72.

Luijsterburg MS, White MF, van Driel R, Dame RT. The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol.* 2008;43(6):393-418.

Richmond TJ, Davey CA. The structure of DNA in the nucleosome core. *Nature.* 2003;423:145-50.

Sandman K, Reeve JN. Archaeal chromatin proteins: different structures but common function? *Curr Opin Microbiol.* 2005;8(6):656-61.

Sakai A, Takano H, Kuroiwa T. Organelle nuclei in higher plants: structure, composition, function, and evolution. *Int Rev Cytol.* 2004;238:59-118.

Segal E, Widom J. What controls nucleosome positions? *Trends Genet.* 2009;25(8):335-43.

Spelbrink JN. Functional organization of mammalian mitochondrial DNA in nucleoids: history, recent developments, and future challenges. *IUBMB Life.* 2010;62(1):19-32.

Woodcock CL, Ghosh RP. Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol.* 2010;2(5):a000596.

Henrique Bunselmeyer Ferreira

Genes e Genomas Procarionóticos

1. Genes procarionóticos 58

- 1.1 Estrutura básica 58
- 1.2 Tamanho 60
- 1.3 Homologia entre genes procarionóticos 60

2. Genomas procarionóticos 62

- 2.1 Tamanho 62
- 2.2 Forma, número e organização em replicons de cromossomos procarionóticos 63
- 2.3 Densidade gênica 64
- 2.4 Unidades organizacionais de genomas procarionóticos: características e representatividade 65
 - 2.4.1 Motivos 65
 - 2.4.2 Repetições 66
 - 2.4.3 Genes 67
 - 2.4.4 Óperons 69
 - 2.4.5 Domínios e macrodomínios 70
 - 2.4.6 Ilhas genômicas 71

- 2.5 Conteúdo de G+C 72

2.6 Distribuição dos genes e de outras sequências em genomas procarionóticos 74

- 2.6.1 Orientação e localização de genes em relação à origem e ao sentido da replicação 74
- 2.6.2 Desvios de GC 75
- 2.6.3 Sintenia: conservação da ordem dos genes 76
- 2.6.4 Superóperons e ultraóperons 77

3. Dinâmica evolutiva dos genomas procarionóticos 79

- 3.1 Principais processos evolutivos 79
 - 3.1.1 Processos inovadores 79
 - 3.1.2 Processos que determinam perda de sequências 81
- 3.2 Principais tendências evolutivas 82

Sistemas de classificação relativamente recentes dividem os seres vivos celulares em três grandes clados, denominados domínios: **Bacteria**, **Archaea** e **Eukarya** (ver **Material Complementar Online 4.1**). Os dois primeiros domínios compõem o superclado **Prokarya**, que inclui organismos unicelulares caracterizados pela ausência de organelas e chamados genericamente de **procariotos**. Em células procarióticas, o material genético (DNA) encontra-se compactado em uma região denominada **nucleoide** (ver Capítulos 1 e 3), que não é separada do restante do citoplasma por qualquer tipo de membrana. O domínio Bacteria é formado pelas bactérias verdadeiras, que às vezes são chamadas de eubactérias. As bactérias são separadas das **arqueas**, constituintes do domínio Archaea, com base em uma série de diferenças genéticas e bioquímicas importantes. O terceiro domínio – Eukarya – inclui todos os **eucariotos**, um grupo heterogêneo de organismos que inclui desde protozoários e algas unicelulares até fungos, vegetais e animais. As células eucarióticas, caracterizadas principalmente pela presença de organelas definidas, como o **núcleo**, têm uma estrutura e uma organização funcional mais complexa do que as células procarióticas (ver Capítulo 1).

Apesar da aparente simplicidade e da semelhança superficial entre bactérias e arqueas, os procariotos em geral são organismos tão ou mais heterogêneos do ponto de vista de suas constituições genéticas do que os organismos eucarióticos (discutidos no Capítulo 5). Essa grande diversidade genética fica evidente a partir de análises genômicas comparativas abrangentes, possíveis agora devido ao número cada vez maior de espécies procarióticas que estão tendo seus genomas completamente sequenciados. O primeiro genoma procariótico a ser inteiramente sequenciado foi o da bactéria patogênica *Haemophilus influenzae*, em 1995, e o primeiro sequenciamento completo de um genoma de uma arquea foi o de *Methanococcus jannaschii* (espécie hoje denominada *Methanocaldococcus jannaschii*), concluído em 1996. Desde então, o número de genomas procarióticos completamente sequenciados vem crescendo de maneira quase exponencial e, apesar das espécies estudadas ainda representarem uma fração mínima da biodiversidade procariótica na Terra (< 0,1%), já existe uma representatividade amostral de todos os filos de bactérias e arqueas. No momento da revisão final deste capítulo, em setembro de 2011, 1.643 genomas de bactérias e 117 genomas de arqueas haviam sido completamente sequenciados, havendo ainda 5.140 genomas bacterianos e 90 genomas de arqueas em processo de sequenciamento. Dados atualizados sobre o progresso do sequenciamento de genomas procarióticos podem ser obtidos na internet, em páginas institucionais especializadas, cujos endereços eletrônicos podem ser encontrados no final deste capítulo.

Apesar da grande heterogeneidade genética, fisiológica e ecológica das espécies procarióticas, os genes e os genomas tanto de bactérias como de arqueas possuem alguns princípios organizacionais comuns e fa-

cilmente discerníveis à luz de análises comparativas de suas sequências nucleotídicas. A seguir, serão descritas as principais características estruturais que definem genes e genomas procarióticos. Além disso, será feita uma correlação dos aspectos estruturais descritos com a funcionalidade e a dinâmica evolutiva de genomas de bactérias e de arqueas.

1. Genes procarióticos

1.1 Estrutura básica

Antes de discutir a organização de genes e genomas de procariotos, é importante conceituar o que é um gene e qual é a estrutura básica dos genes de bactérias e arqueas. Simplificadamente, um **gene**, seja ele de um procarioto ou de um eucarioto, pode ser definido, do ponto de vista molecular, como um segmento de DNA que inclui todas as sequências nucleotídicas necessárias e suficientes para a síntese de pelo menos um produto correspondente, que pode ser um mRNA a ser traduzido em uma cadeia polipeptídica (proteína) ou outros tipos de RNA (como rRNAs ou tRNAs, entre outros). Em sua estrutura básica, cada gene é constituído por uma **região codificadora**, que é a sequência nucleotídica que codifica o seu produto (de RNA ou polipeptídico), e pelas **sequências reguladoras**, que a flanqueiam (as distâncias são variáveis de gene para gene) controlando a sua expressão (**Figura 4.1**).

A **expressão** de um gene procariótico inclui a sua transcrição em um RNA (ver Capítulo 10) e a tradução do transcrito (se for um mRNA) em uma cadeia polipeptídica (ver Capítulo 12). Genes que codificam rRNAs, tRNAs ou outras classes menores de RNA são transcritos, mas não são traduzidos. A tradução só ocorre com mRNAs, os quais são transcritos a partir de genes que codificam cadeias polipeptídicas. Etapas de maturação de produtos de transcrição primários para a produção de transcritos maduros (ver Capítulo 11) também fazem parte do processo de expressão gênica, mas, em procariotos, essas etapas estão restritas à maturação de rRNAs e tRNAs.

O processo de síntese de RNA (transcrição), que progride de 5' para 3', define a **orientação** do gene. Assim, cada gene está orientado de 5' para 3', considerando a orientação da fita de DNA cuja sequência estará representada no RNA (denominada **fita codificadora**). A outra fita de DNA, complementar à fita codificadora e com orientação inversa (de 3' para 5'), é denominada **fita-molde**, pois serve de molde para a síntese do RNA. O RNA transcrito é complementar à fita-molde e, por conseguinte, tem a mesma orientação e sequência correspondente à da fita codificadora.

Em um gene procariótico típico, as sequências reguladoras que controlam a sua transcrição encontram-se em posições adjacentes à região codificadora (ver

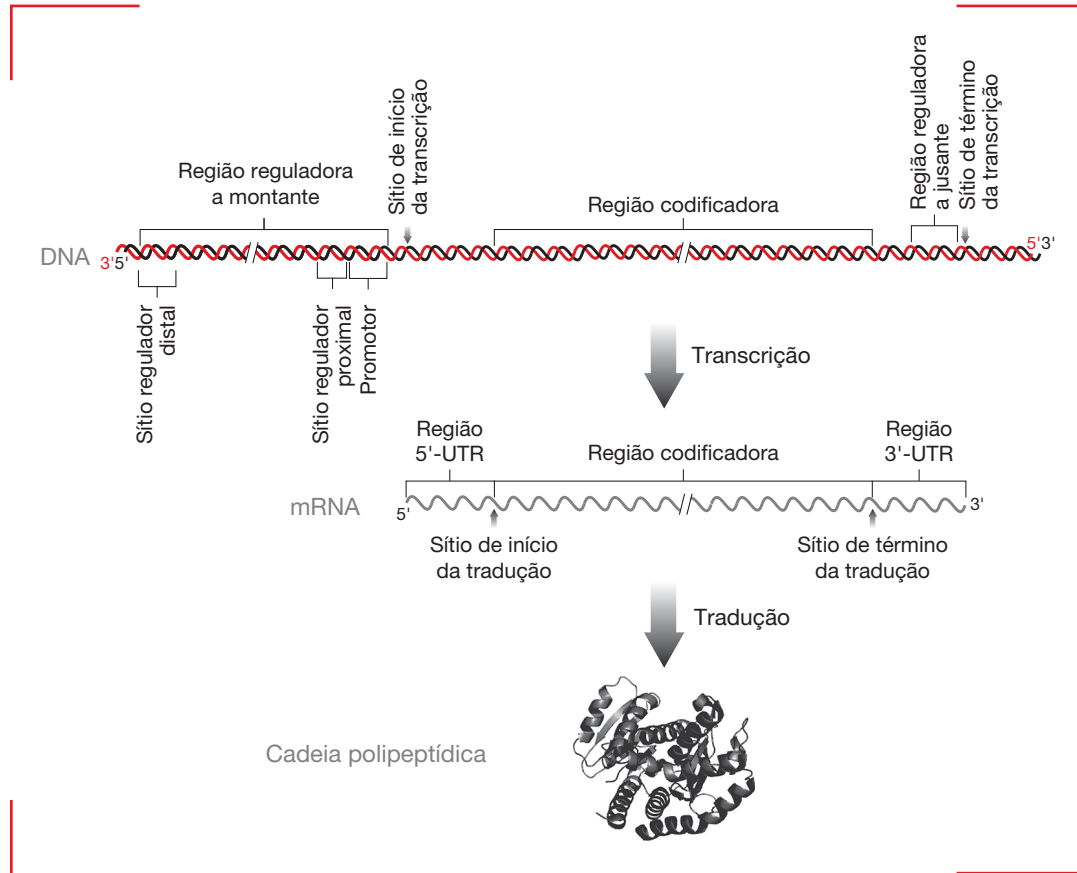


Figura 4.1

Esquema representativo da estrutura e da expressão de um gene procaríótico que codifica uma proteína. O gene é constituído basicamente por uma região codificadora (de seu produto) e por regiões reguladoras (da sua expressão). Por convenção, o gene é representado de 5' para 3', considerando sua fita codificadora (em ■). A fita complementar (em ■), de orientação oposta (de 3' para 5'), é utilizada como molde para a síntese de um mRNA de mesma orientação e sequência de bases (salvo a troca de T por U) que a fita codificadora. A região codificadora é flanqueada pelas regiões reguladoras. A região reguladora a montante da região codificadora inclui o promotor, e pode incluir um ou mais sítios de ligação de proteínas ativadoras ou repressoras da transcrição. A jusante da região codificadora, há uma sequência reguladora que determina o final do processo de transcrição do gene. Quando o gene é expresso, a RNA-polimerase o transcreve em um mRNA, cuja síntese inicia alguns pares de bases antes da região codificadora e termina alguns pares de bases depois do final dessa região. As regiões não codificadoras do RNA são chamadas de regiões 5' ou 3' não traduzidas ou UTR (de *untranslated region*). Na região 5'-UTR do mRNA transcrito, há sequências importantes para o início da sua tradução. Nos ribossomos, a região codificadora do RNA é interpretada de 5' para 3' e traduzida em uma cadeia polipeptídica, cuja sequência de aminoácidos é sintetizada da sua extremidade aminoterminal para a sua extremidade carboxiterminal. Em genes que codificam apenas RNAs, como os de rRNA e tRNA, a etapa de tradução não ocorre.

Figura 4.1). As sequências reguladoras da transcrição mais fundamentais são o promotor, situado na região flanqueadora 5', a montante da região codificadora, e o terminador, situado na região flanqueadora 3', a jusante da região codificadora. O promotor é o sítio no DNA onde a RNA-polimerase, enzima responsável pela transcrição do gene, se liga. O terminador é a sequência nucleotídica que determina a dissociação da RNA-polimerase da fita-molde e o final do processo de transcrição. Além das referidas sequências, podem estar

presentes outros elementos reguladores, que são sítios de ligação de proteínas ativadoras ou repressoras da transcrição. Esses sítios reguladores adicionais podem estar situados em uma posição adjacente ao promotor e até sobrepostos a ele, mas também podem estar em posições mais distais, de até poucas centenas de pares de bases a montante do promotor. O Capítulo 13 apresenta uma discussão detalhada sobre a organização e as funções das sequências reguladoras de genes procaríóticos.

Uma característica típica de genes procarióticos é a colinearidade existente entre cada gene e o seu produto (seja uma cadeia polipeptídica ou um RNA). Isso significa dizer que há uma exata correspondência entre a sequência nucleotídica da região codificadora do gene (DNA) e a sequência de aminoácidos da cadeia polipeptídica ou a sequência de nucleotídeos de um rRNA ou tRNA, por exemplo. Assim, considerando que, em um gene codificador de uma cadeia polipeptídica, cada códon de três nucleotídeos codifica um aminoácido, para codificar uma proteína com um número n de aminoácidos é necessária uma região codificadora no gene correspondente com $3n$ pares de bases. Essa relação, embora pareça óbvia, representa uma diferença básica entre a estrutura de praticamente todos os genes de procariotos e a estrutura de um grande número de genes de eucariotos, que podem apresentar sequências intervinientes (os chamados íntrons) interrompendo aquelas sequências de DNA que devem estar representadas no produto final de expressão do gene (ver Capítulo 5). O sequenciamento de centenas de genomas e milhares de genes revelou que a maioria dos genes procarióticos não apresenta íntrons e é colinear com os respectivos produtos, apesar de exceções a essa regra já terem sido identificadas em genes tanto de bactérias como de arqueas.

1.2 Tamanho

O tamanho de um gene está vinculado ao número e à extensão das sequências nucleotídicas que compõem as suas regiões reguladoras e à extensão da sua região codificadora, que varia de acordo com o tamanho do produto codificado. Entretanto, os genes procarióticos, por não apresentarem regiões reguladoras muito extensas nem serem interrompidos por íntrons, variam em tamanho dentro de uma faixa mais restrita do que aquela dos genes eucarióticos.

As sequências de função reguladora de genes procarióticos são curtas, com menos de 20 pb, estando em geral posicionadas a distâncias relativamente pequenas (de poucas dezenas a poucas centenas de pares de bases), a montante ou a jusante das regiões codificadoras. Por isso, elas representam uma fração da extensão total do gene que oscila, em média, entre 5 e 10%, sendo o restante correspondente à região codificadora. Para genes procarióticos que codificam polipeptídeos (os quais ocupam mais de 95% do espaço do genoma em qualquer espécie de bactéria ou arquea), a extensão das regiões codificadoras oscila entre pouco menos de 100 pb até quase 10.000 pb, com a grande maioria tendo em torno de 1.000 pb.

No genoma da bactéria *Escherichia coli* (linhagem K-12), por exemplo, o tamanho médio das regiões codificadoras dos aproximadamente 4.300 genes identificados é de 951 pb (equivalente a 317 aminoácidos). A região codificadora mais longa tem 7.149 pb e codifica uma proteína de 2.383 aminoácidos. Há ainda quatro regiões

codificadoras com tamanhos entre 4.500 e 5.100 pb (codificando proteínas com 1.500 a 1.700 aminoácidos), 51 com tamanhos entre 3.000 e 4.500 pb (codificando proteínas com 1.000 a 1.500 aminoácidos) e 381 com tamanho menor que 300 pb (codificando proteínas com menos de 100 aminoácidos).

1.3 Homologia entre genes procarióticos

Muitos genes (e também outros elementos genômicos estruturais e funcionais) de procariotos (e também de eucariotos, ver Capítulo 5) apresentam similaridades evidentes em suas sequências nucleotídicas, as quais são genericamente referidas como **homologias**. De grau variável conforme o caso, essas homologias decorrem dos processos evolutivos que deram origem às sequências envolvidas. Os processos evolutivos determinantes do surgimento de sequências homólogas, como a duplicação seguida de divergência, a transferência gênica horizontal e a especiação, serão discutidos, em um contexto genômico procariótico, na Seção 3 deste capítulo. Aqui, será apresentado a definição de alguns conceitos e parâmetros para o reconhecimento de parentesco evolutivo entre genes e outras sequências, os quais são aplicáveis tanto para os procariotos como para os eucariotos.

Assume-se que quaisquer genes ou sequências nucleotídicas com homologia (isto é, com identidade parcial ou até total) tenham uma origem evolutiva comum. O grau de conservação necessário para que duas sequências nucleotídicas possam ser consideradas homólogas é quantitativo e arbitrário. Alguns autores, por exemplo, consideram homólogos genes que têm pelo menos 30% de identidade em suas sequências ao longo de pelo menos 60% de suas extensões. Genes com menor grau de identidade são considerados não relacionados (pelo menos em termos evolutivos), e genes com um grau de identidade muito mais elevado, próximo a 100%, mas não necessariamente idênticos, são considerados cópias, pois se assume que codifiquem produtos sem distinção funcional.

Genes homólogos surgem ao longo da evolução a partir da ocorrência de diferentes processos evolutivos e, em razão disso, são estabelecidas diferentes categorias de homologia e de sequências homólogas (**Figura 4.2**). Quando o evento evolutivo determinante da presença de dois genes (ou sequências) homólogos é uma duplicação de uma região cromossômica, os genes duplicados são chamados de **parálogos**. Em um primeiro momento, uma duplicação gera **cópias** idênticas do segmento cromossômico envolvido. Contudo, com o passar do tempo e dependendo das pressões seletivas atuantes, diferentes situações podem ocorrer. As cópias podem, por exemplo, serem fixadas como tal (**Figura 4.2A**), em situações nas quais os genes duplicados supririam uma demanda elevada do mesmo produto gênico. Alternativamente, uma das cópias pode ser fixada, suprimindo a demanda pelo produto original, e a outra pode divergir em sua sequência (**Figura**

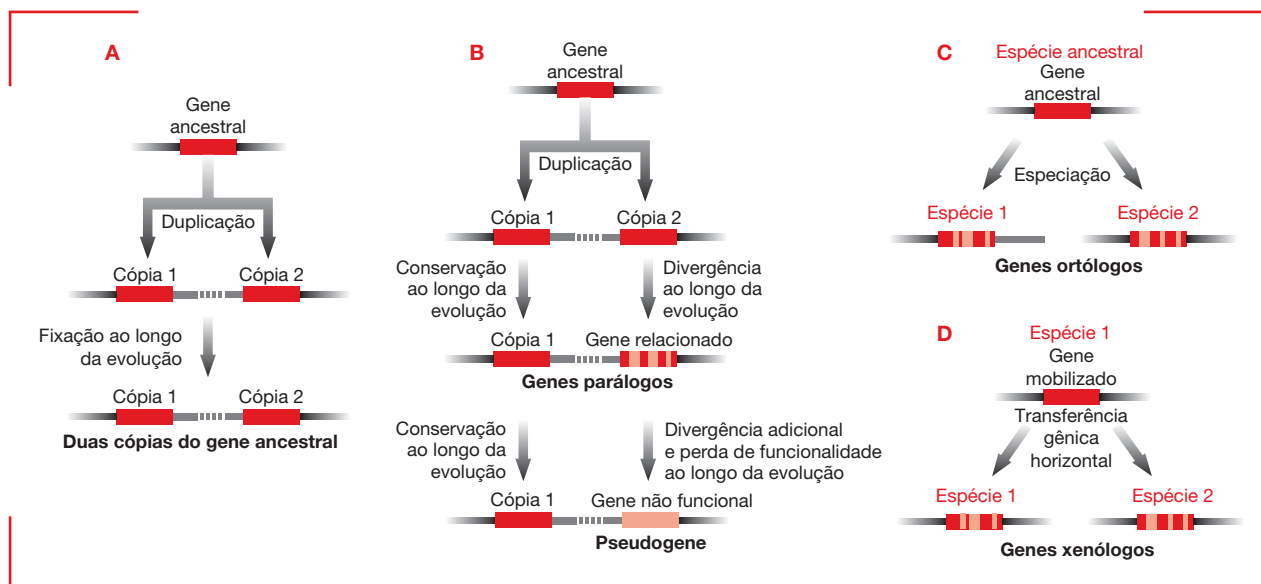


Figura 4.2

Eventos evolutivos que geram cópias de genes, genes parálogos, pseudogenes, genes ortólogos e genes xenólogos. (A) Uma simples duplicação de um segmento cromossômico pode gerar duas cópias de um mesmo gene; se esse aumento do número de cópias conferir alguma vantagem adaptativa, ele poderá ser fixado ao longo da evolução. (B) Após um evento de duplicação, uma das cópias geradas pode manter a função ancestral, ao passo que a outra pode divergir e, eventualmente, assumir outra função relacionada. Os genes gerados por duplicação e posterior divergência são chamados de parálogos; genes duplicados que, ao divergirem, perdem sua funcionalidade, mas não a homologia com seu(s) parálogo(s), são chamados de pseudogenes. (C) Genes com sequências nucleotídicas similares e codificando proteínas com funções correspondentes em duas espécies diferentes podem ser derivados de um gene originalmente presente em um ancestral comum àquelas duas espécies; esses genes, relacionados não por duplicação, mas pela ancestralidade comum, são chamados de ortólogos. (D) Genes adquiridos independentemente por duas espécies por meio de eventos de transferência gênica horizontal envolvendo uma terceira espécie são chamados de xenólogos. Na figura, as barras em ■ representam as sequências ancestrais; as barras em ■ representam sequências que divergiram ao longo da evolução.

4.2B), dando origem a um gene parálogo codificador de um produto relacionado, mas não idêntico ao original, podendo inclusive ter funcionalidade distinta. Em alguns casos, a progressiva divergência associada à uma relativa falta de pressão seletiva para manutenção da sequência de uma das cópias de um gene pode determinar que essa cópia se torne não funcional, o que fará dela um **pseudogene**.

Quando o determinante da presença de homologia entre dois genes (ou sequências) são eventos de especiação (surgimento de novas espécies ao longo da evolução), os genes em questão são chamados de **ortólogos** (Figura 4.2C). Nesse caso, em genomas de espécies distintas, os genes ortólogos (que tenham um ancestral comum em sua história evolutiva) são identificáveis como sequências similares, que codificam produtos também similares e com funções correspondentes em cada uma das espécies. Quanto maior for o número de genes ortólogos presentes em genomas de espécies diferentes, maior é o grau de parentesco (ou a proximidade) desses organismos na escala evolutiva. Por exemplo, o número de genes ortólogos identificáveis na comparação de genomas de duas espécies quaisquer de bactérias será

muito maior do que o número identificável na comparação do genoma de cada uma dessas bactérias com o genoma de uma arquea ou de um eucarioto. A implicação óbvia destas comparações é que as duas espécies bacterianas são mais aparentadas entre si do que com espécies dos domínios Archaea e Eukarya, justificando a separação desses organismos em diferentes clados. A identificação de genes ortólogos, mesmo quando comparando genomas de organismos muito distantes evolutivamente, como o de uma bactéria com o de um mamífero, por exemplo, é uma evidência de que, em algum momento ao longo da evolução, tais organismos compartilharam um mesmo ancestral.

A transferência gênica horizontal, mediada por elementos genéticos móveis (Capítulo 9), também é reconhecida como um processo marcante na evolução de organismos procarionóticos e capaz de levar ao surgimento de genes ortólogos (Figura 4.2D). Neste caso, duas ou mais espécies distintas podem adquirir, de forma independente, um mesmo gene a partir de eventos de transferência gênica horizontal da espécie de origem do gene. Os genes ortólogos, adquiridos dessa forma, são chamados de genes **xenólogos**.

2. Genomas procarióticos

2.1 Tamanho

Os genomas procarióticos já sequenciados, que, como visto anteriormente, constituem uma amostragem representativa de todos os filos de bactérias e de arqueas, variam de tamanho dentro de uma faixa ampla, que vai de pouco mais de 150.000 pb (150 kb ou 0,15 Mb) até pouco mais de 13.000.000 pb (13.000 kb ou 13 Mb) (Figura 4.3). Entre bactérias, são exemplos extremos o genoma de 159.662 pb de *C. rudii* (uma endossimbionte de insetos hemípteros da família Psyllidae) e o genoma de 13.033.779 pb da mixobactéria de solo *S. cellulosum*. Entre as arqueas, o menor genoma conhecido é o de *N. equitans*, uma endossimbionte obrigatória de outra arquea (*Ignicoccus hospitalis*), com 490.885 pb, e o maior é o de *M. acetivorans*, uma metanógena anaeróbia, com 5.751.492 pb. Esses limites extremos de tamanho chegam a proporcionar sobreposições com os maiores genomas virais de um lado (o do bacteriófago φKZ tem 280.334 pb e o do mimivírus APMV tem 1,2 Mb, p. ex.) e os menores genomas eucarióticos de outro

(microsporídios têm genomas da ordem de 2,9 Mb, e os menores genomas de leveduras oscilam entre 10 e 12 Mb, ver Capítulo 5).

Entre os extremos menores e maiores, tanto as espécies de bactérias como as de arqueas apresentam distribuições características no tamanho de seus genomas. Para bactérias, predominam espécies com genomas com cerca de 2 e 5 Mb, o que resulta em uma distribuição bimodal, apesar de haver muitas espécies com genomas de tamanho intermediário. Já para arqueas, há uma clara predominância de espécies com genomas de aproximadamente 2 Mb, coincidindo com o primeiro pico bacteriano. Genomas de arqueas muito menores ou maiores que 2 Mb são uma minoria e, dentre eles, há uma clara predominância dos maiores que 2 Mb em relação aos menores que 2 Mb.

Quando é feita uma análise comparativa dos tamanhos dos genomas entre espécies, linhagens ou isolados de bactérias ou arqueas dentro de uma mesma categoria taxonômica, observa-se uma considerável heterogeneidade. Por exemplo, *E. coli* K-12 tem um genoma de 4.639.221 pb (pouco mais de 4,6 Mb), mas já foram

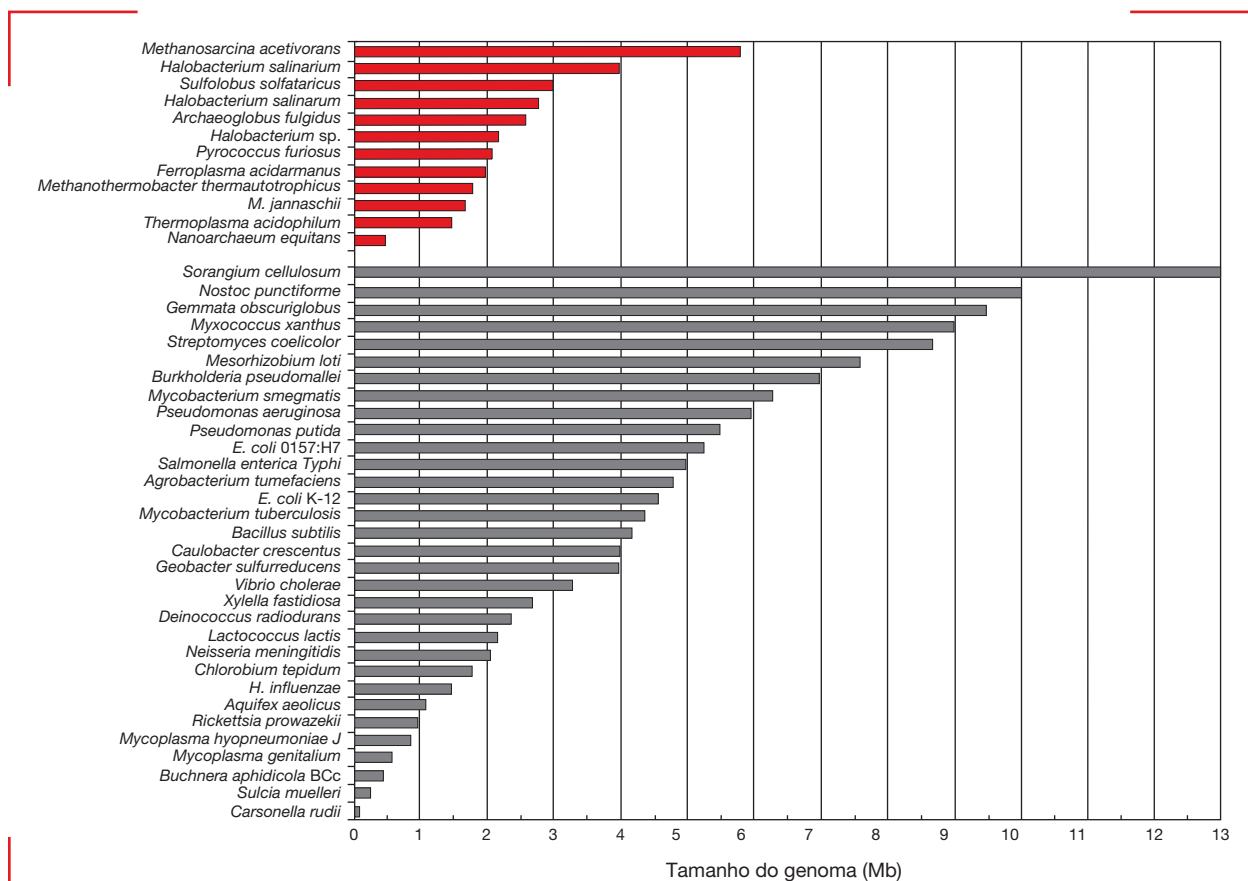


Figura 4.3

Gráfico representativo do tamanho dos genomas de algumas espécies de arqueas (barras em ■) e bactérias (barras em ■).

caracterizados genomas de outras linhagens de *E. coli* com tamanhos entre 4,5 e 5,5 Mb, o que implica em diferenças de até 1 Mb entre genomas de indivíduos de uma mesma espécie. Situações similares foram descritas também para ecotipos de *Prochlorococcus marinus* (filo Cyanobacteria), com genomas com tamanhos entre 1,6 Mb e 2,4 Mb, e para linhagens de *S. coelicolor* (filo Actinobacteria), com genomas de tamanhos entre 8,7 Mb e 9,7 Mb.

Quando comparações como essas são estendidas para níveis taxonômicos mais elevados, como o de filo, observa-se que o tamanho dos genomas de espécies de um mesmo filo pode variar em até 10 vezes. Por exemplo, em Bacteria, o filo Cyanobacteria apresenta espécies com genomas de 1,6 a 9,0 Mb, o filo Actinobacteria apresenta espécies com genomas de 0,9 a 9,7 Mb, e o filo Firmicutes (bactérias gram-positivas), espécies com genomas de 0,6 a 6 Mb. Em Archaea, a variação de tamanho entre genomas de diferentes espécies de um mesmo filo também é evidente, apesar de restrita a limites mais estreitos. Por exemplo, espécies do filo Crenarchaeota têm genomas de 1,3 a 3 Mb e espécies do filo Euryarchaeota têm genomas de 1,6 a 5,8 Mb. Tanto em bactérias como em arqueas há uma considerável sobreposição entre as faixas de tamanho dos genomas de espécies de filios diferentes, o que torna o tamanho dos genomas um parâmetro inadequado para o estabelecimento de relações filogenéticas. Isso ocorre porque o tamanho do genoma de qualquer espécie procariótica é potencialmente muito plástico, podendo variar com relativa facilidade ao longo da evolução sob a influência de processos determinantes de aquisição ou perda de sequências, os quais serão discutidos na Seção 3.

Os menores genomas encontrados na natureza para organismos procarióticos ou eucarióticos, heterotróficos ou autotróficos e de vida livre ou dependentes de uma espécie hospedeira suscitam questões sobre qual seria o conteúdo mínimo de genes (e, conseqüentemente, o tamanho mínimo de genoma) necessário para cada uma dessas situações. Isso resultou na proposição do conceito de **genoma mínimo**, que descreveria o menor genoma capaz de abrigar o número mínimo necessário de genes para a sobrevivência de uma forma de vida celular (ver **Material Complementar Online 4.2**).

2.2 Forma, número e organização em replicons de cromossomos procarióticos

Do ponto de vista estrutural molecular básico, todos os genomas procarióticos, independentemente do tamanho, são constituídos por DNA de fita dupla. Essa universalidade, contudo, não se estende a alguns outros aspectos estruturais importantes, como a forma e o número de cromossomos e a maneira como os cromossomos estão organizados em unidades de replicação.

No que diz respeito à forma e ao número de moléculas de DNA constituintes do genoma, a maioria das espécies de bactérias e de arqueas conhecidas tem seus genomas representados por um único **cromossomo**, na forma de uma molécula de DNA de fita dupla circular covalentemente fechada. Em bactérias, essa situação pode ser exemplificada por *E. coli*, cujo único cromossomo é constituído por uma molécula de DNA circular que, como visto na Seção 2.1, tem mais de 4,5 Mb. Há, contudo, alguns poucos casos descritos de espécies bacterianas cujo genoma é formado por mais de uma molécula de DNA, incluindo situações nas quais uma ou mais dessas moléculas são lineares, em vez de circulares (**Tabela 4.1**). Essas diferentes configurações genômicas parecem estar restritas a algumas espécies ou gêneros, havendo evidências sugerindo que as variações estruturais hoje observadas surgiram de modo independente ao longo da evolução em cada grupo taxonômico que as apresenta. Em cada gênero ou espécie com genoma multimolecular, as moléculas constituintes podem ser tratadas de acordo com seu tamanho e conteúdo gênico. Em geral, as moléculas maiores, da ordem de centenas ou milhares de quilobases, e contendo predominantemente genes essenciais são chamadas de cromossomos. Quando se trata de moléculas de DNA menores, da ordem de dezenas a poucas centenas de quilobases, e contendo um número mais limitado de genes essenciais, algumas delas podem ser denominadas plasmídeos ou megaplasmídeos, apesar de ainda serem consideradas como parte do genoma.

Em arqueas, a amostragem de genomas estudados é menor, mas as evidências disponíveis apontam para uma situação similar a de bactérias no que diz respeito à organização cromossômica. Assim, para arqueas, a situação típica é a de genomas constituídos por um único cromossomo circular (ver **Tabela 4.2**), embora haja, também, como em bactérias, casos de genomas multimoleculares, como o de *Haloarcula marismortui* (com dois cromossomos) e o de *M. jannaschii* (com um cromossomo e dois elementos extracromossômicos). Ainda não foram descritos casos de cromossomos lineares em arqueas.

Outro aspecto estrutural importante de genomas procarióticos refere-se à organização dos cromossomos em unidades de replicação, que pode diferir entre bactérias e arqueas, e é bastante distinta da encontrada em eucariotos (Tabela 4.2 e **Figura 4.4**). As unidades de replicação, ou **replicons**, são os segmentos cromossômicos que replicam a partir de uma origem de replicação (ver Capítulo 6) e, em bactérias, o conceito de replicon se confunde com o de cromossomo, pois cada cromossomo bacteriano possui apenas uma origem de replicação (Figura 4.4A). Em arqueas, por outro lado, há pelo menos duas situações distintas (Figura 4.4B), que foram evidenciadas a partir dos genomas até agora estudados quanto a este aspecto. Existem espécies de arqueas com cromossomos com uma única origem de replicação e constituin-

Tabela 4.1 Algumas espécies bacterianas com genomas formados por mais de uma molécula de DNA

Espécies	Tipos de moléculas constituintes do genoma*	Tamanho (kb)	Forma
<i>S. coelicolor</i>	Cromossomo	8.667	Linear
	Plasmídeo	356	Linear
	Plasmídeo	31	Circular
<i>S. coelicolor</i>	Cromossomo	8.667	Linear
	Plasmídeo	356	Linear
	Plasmídeo	31	Circular
<i>Borrelia burgdorferi</i>	Cromossomo	911	Linear
	Plasmídeos (11)	9-54	Circular/Linear
<i>Brucella melitensis</i>	Cromossomo	2.117	Circular
	Cromossomo	1.178	Circular
<i>Clostridium acetobutylicum</i>	Cromossomo	3.941	Circular
	Megaplasmídeo	192	Circular
<i>D. radiodurans</i>	Cromossomo	2.649	Circular
	Cromossomo	412	Circular
	Megaplasmídeo	177	Circular
	Plasmídeo	46	Circular
<i>Ralstonia solanacearum</i>	Cromossomo	3.716	Circular
	Megaplasmídeo	2.095	Circular
<i>S. enterica</i> Typhi	Cromossomo	4.809	Circular
	Plasmídeo	218	Circular
	Plasmídeo	107	Circular
<i>Sinorhizobium meliloti</i>	Cromossomo	3.654	Circular
	Megaplasmídeo	1.683	Circular
	Megaplasmídeo	1.354	Circular
<i>V. cholerae</i>	Cromossomo	2.941	Circular
	Cromossomo	1.072	Circular
<i>Yersinia pestis</i>	Cromossomo	4.654	Circular
	Plasmídeos (3)	10-96	Circular

* A denominação de cada molécula componente do genoma está de acordo com sua descrição original para a espécie. A definição como cromossomo, megaplasmídeo ou plasmídeo tem como base, para todas as espécies, o tamanho da molécula e a presença de genes essenciais, mas os parâmetros quantitativos foram estabelecidos arbitrariamente para cada espécie e, portanto, podem diferir entre elas.

do, portanto, um só replicon, como em bactérias. Há, contudo, outras arqueas com cromossomos com duas ou até três origens de replicação, os quais estão, portanto, organizados em dois ou três replicons, respectivamente. Essa situação em genomas de arqueas parece ser intermediária entre a de bactérias, de um cromossomo: um replicon, e a de eucariotos, nos quais cada cromossomo está organizado em centenas ou milhares de replicons (Figura 4.4C).

2.3 Densidade gênica

Fatores como o tamanho relativamente reduzido dos genomas, a extensão das regiões codificadoras e reguladoras dos genes e o número de genes por genoma, determinam que os genomas procarióticos apresentem uma elevada **densidade gênica**, principalmente em comparação com genomas eucarióticos. A análise do

número de genes por quilobase em genomas procarióticos evidencia que a maioria das espécies procarióticas tem genomas com densidades gênicas de 0,9 a 1,1 gene a cada 1.000 pb ou, em média aproximada, 1 gene/kb. Os genomas de arqueas, contudo, tendem a ser um pouco mais compactos, pois predominam genomas com densidades gênicas entre 1 e 1,1 gene/kb, ao passo que, para genomas bacterianos, os valores de densidade gênica predominantes são da ordem de 0,9 gene/kb.

Outra análise interessante de ser feita é a da extensão das regiões intergênicas, que como um dos componentes que influenciam a densidade gênica, também evidencia o grau elevado de compactação da informação em genomas procarióticos. Tanto em bactérias como em arqueas, as regiões intergênicas são bastante curtas, sendo, via de regra, muito menores que os genes ou as suas regiões codificadoras. A extensão das re-

Tabela 4.2 Forma e organização em replicons de cromossomos de Bacteria, Archaea e Eukarya. A situação de Archaea é representada por algumas das espécies que já tiveram seus genomas estudados quanto ao número de origens de replicação. Para Bacteria e Eukarya, foram feitas generalizações baseadas em evidências experimentais concordantes, disponíveis para diversas espécies

Domínio	Espécie	Forma do(s) cromossomo(s)	Número de cromossomos	Número de origens de replicação (ou de replicons) por cromossomo
Bacteria	Todas	Circular ^a	1 ^b	1
Archaea	<i>A. fulgidus</i>	Circular	1	1
	<i>Halobacterium</i> sp. NRC-1	Circular	1 (e 2 plasmídeos)	2
	<i>M. thermotrophicus</i>	Circular	1	1
	<i>Methanosarcina mazei</i>	Circular	1	1
	<i>Pyrococcus abyssi</i>	Circular	1 (e 1 plasmídeo)	1
	<i>P. furiosus</i>	Circular	1	1
	<i>M. jannaschii</i>	Circular	1 (e 2 elementos extracromossômicos)	2
	<i>Sulfolobus acidocaldarius</i>	Circular	1	3
	<i>S. solfataricus</i>	Circular	1	2
Eukarya	Todas	Linear	Vários	Centenas ou milhares

^a Casos de cromossomos lineares são considerados exceções.

^b Casos de genomas multimoleculares são considerados exceções.

giões intergênicas em genomas procarióticos pode variar desde nenhum par de base até pouco mais de 1.000 pb. Entretanto, tanto em bactérias como em arqueas, há uma evidente distribuição bimodal das extensões observadas, com predominância de regiões intergênicas menores que 2 a 5 pb e em torno de 100 pb. Essa distribuição em dois picos é consequência da organização de muitos genes procarióticos em óperons, unidade organizacional discutida na Seção 2.4.4. Em óperons, a justaposição de regiões codificadoras sob controle de uma região reguladora comum explica a elevada frequência de regiões intergênicas menores que 5 pb, ao passo que regiões intergênicas da ordem de 100 pb corresponderiam, predominantemente, a regiões entre óperons.

2.4 Unidades organizacionais de genomas procarióticos: características e representatividade

Os genes são apenas alguns dos componentes de sequência presentes em um genoma, e, em genomas procarióticos, outras unidades organizacionais de sequência podem ser identificadas. Essas unidades organizacionais podem ser hierarquizadas do ponto de vista estrutural, pois variam em complexidade, desde simples motivos de sequência curtos até segmentos cromossômicos inteiros (Figura 4.5). Contudo, o mais importante é que cada uma delas pode ter implicações funcionais e/ou evolutivas e que, no conjunto, elas definem a arquitetura estrutural e funcional de qualquer

genoma bacteriano ou de arqueas. A seguir, serão discutidas as características fundamentais de cada um dos principais tipos de unidades organizacionais encontrados em genomas procarióticos.

2.4.1 Motivos

Motivos são sequências nucleotídicas curtas, em geral de dois até algumas dezenas de pares de bases de extensão, que podem ser relativamente comuns em genomas procarióticos. Essas sequências podem estar dispersas ou agrupadas no genoma e, em muitos casos, são unidades formadoras de repetições (ver Seção 2.4.2). Alguns motivos, contudo, podem ser mais do que meras unidades estruturais e constituir elementos de sequência importantes do ponto de vista funcional. Por exemplo, os **sítios Chi** (de *crossover hot spot instigator*, instigador de sítios preferenciais de recombinação), com as sequências consensuais 5'-GCTGGTGG-3', em *E. coli*, e 5'-GNTGGTGG-3' e 5'-G(G/C)TGGAGG-3', em *H. influenzae*, são motivos superrepresentados em genomas bacterianos (1.008 cópias no genoma de *E. coli* K-12) e importantes para a recombinação genética, assunto que será tratado no Capítulo 8.

Outro exemplo interessante são os **motivos KOPS** (de *FtsK orienting polar sequences*, sequências polares orientadoras de FtsK), orientados a partir da origem em direção ao sítio de término da replicação do cromossomo. Os motivos KOPS têm a sequência 5'-GGGNAGGG-3' em *E. coli* e outras bactérias, e o reconhecimento deles pela DNA-translocase FtsK faz parte do mecanismo de separação (resolução) em mo-

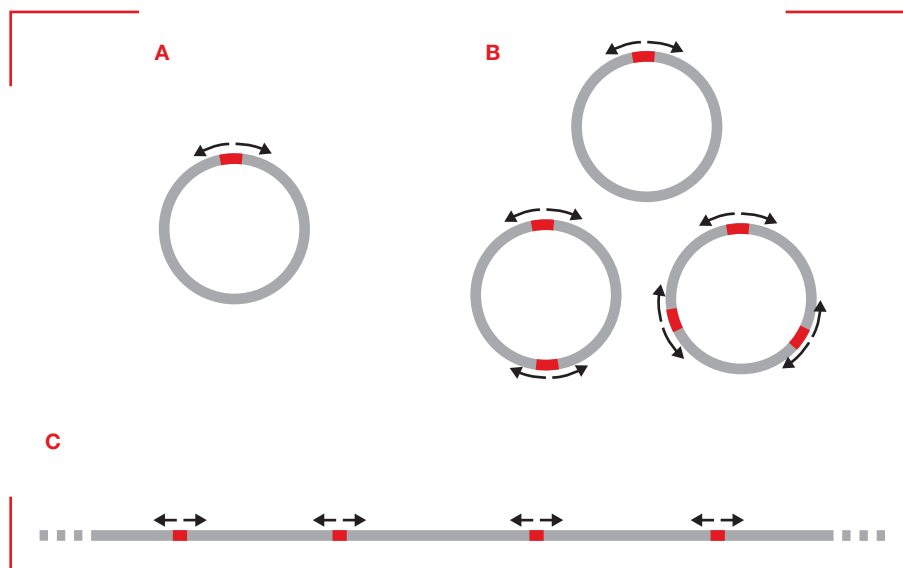


Figura 4.4

Organização em replicons de cromossomos de bactérias, arqueas e eucariotos. Cromossomos bacterianos (A), tipicamente circulares, apresentam uma única origem de replicação e, portanto, constituem um único replicon. Cromossomos de arqueas (B) também são circulares, mas, dependendo da espécie, podem ter uma, duas ou três origens de replicação, constituindo um, dois ou três replicons, respectivamente. Cromossomos eucarióticos (C), por sua vez, são lineares e estão organizados em centenas ou milhares de replicons. As origens de replicação estão representadas em ■; a cada ciclo de divisão celular, um processo de replicação bidirecional (indicado por um par de setas em ■ divergentes) inicia a partir de cada origem.

números dos dímeros de cromossomos gerados na replicação (ver Capítulo 6). No genoma de *E. coli* K-12, os motivos KOPS estão superrepresentados na fita replicada continuamente (fita-líder do cromossomo), com 335 ocorrências (contra apenas 33 na fita tardia), o que representa uma cópia do motivo a aproximadamente cada 13 kb.

2.4.2 Repetições

Sequências repetidas, apesar de serem muito mais abundantes em genomas eucarióticos (ver Capítulo 5),

também estão presentes em genomas procarióticos. Essas sequências repetidas são compostas por **unidades de repetição** em geral curtas, com menos de 10 pb, mas que, no conjunto, podem representar uma fração significativa de um genoma procariótico, embora sempre minoritária em relação à fração correspondente às **sequências únicas** (não repetidas). Menos de 2% de qualquer genoma procariótico são compostos de sequências repetidas, o que contrasta com a representatividade desse tipo de sequência em genomas eucarióticos, que pode chegar a mais de 50%.

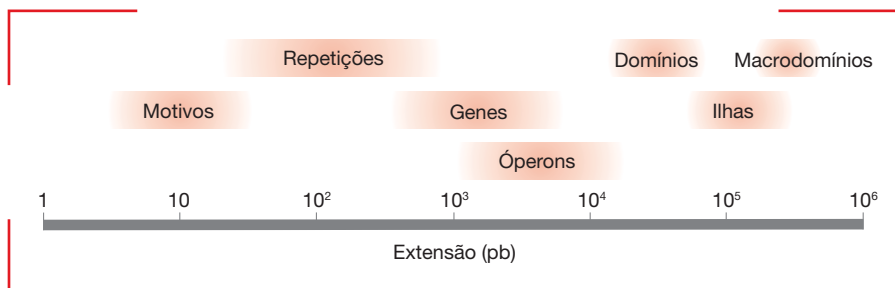


Figura 4.5

Principais unidades organizacionais de genomas procarióticos. As barras em ■ indicam as extensões aproximadas típicas para cada tipo de unidade organizacional, conforme a escala exponencial indicada na barra inferior.

Em genomas procarióticos, diferentes **famílias de sequências repetidas** estão presentes, com números de cópias que podem variar de menos de 5 até algumas centenas ou até milhares por genoma, dependendo da família. As unidades de repetição de uma determinada família podem estar distribuídas de maneira aparentemente aleatória no genoma ou em **arranjos em tandem** (agrupadas lado a lado). Unidades de repetição agrupadas também podem estar arrançadas em uma mesma orientação ou apresentarem orientações invertidas (convergentes ou divergentes), umas em relação às outras. Em *E. coli* K-12, por exemplo, pelo menos sete famílias de sequências repetidas estão representadas no genoma, sendo o número e a distribuição dos elementos repetidos variáveis de acordo com a família. As sequências repetidas mais longas têm de 5,7 a 9,6 kb e pertencem à **família Rhs**, que representa 0,8% do genoma. A família mais representada do ponto de vista do número de unidades de repetição é a dos **elementos REP**. Essas **sequências palindrômicas** (idênticas quando lidas da esquerda para a direita ou da direita para a esquerda) têm aproximadamente 40 pb e estão presentes em 314 unidades organizadas em 12 arranjos em tandem, representando 0,54% do genoma.

A presença de muitas das famílias de sequências repetidas identificadas em genomas procarióticos é explicada pela presença e atividade de elementos genéticos móveis (ver Capítulo 9). Processos envolvendo integração e excisão de genomas de bacteriófagos ou elementos transponíveis em genomas procarióticos geram sequências repetidas, que podem se acumular ao longo da evolução. Além disso, independentemente da origem, algumas repetições de genomas procarióticos podem estar associadas a funções importantes. Dentre elas, cabe salientar, além das repetições de sítios *Chi* e de motivos KOPS, discutidas anteriormente (Seção 2.4.1), as **repetições CRISPR** (de *clustered regularly interspaced short palindromic repeats*, repetições palindrômicas curtas agrupadas e regularmente espaçadas). As CRIS-PRs são componentes de um sistema de defesa antiviral (contra bacteriófagos), e já foram identificadas em cerca de 40% dos genomas bacterianos e 90% dos genomas de arqueas analisados até o momento. Elas são constituídas por unidades de repetição palindrômicas, com extensão e organização em arranjos que variam de espécie para espécie. O tamanho das unidades de repetição das CRISPRs varia de 24 a 47 pb e, nos arranjos, elas estão presentes em um número de cópias que pode variar de 2 até 249 (identificado na bactéria gram-negativa *Verminiphrobacter eisenial*). O tamanho das sequências espaçadoras entre as repetições dentro de um mesmo arranjo de CRISPR varia de 26 a 72 pb, conforme a espécie. Muitos genomas, principalmente os de bactérias, contêm apenas um ou poucos arranjos (ou lócus) CRISPR, mas em genomas de arqueas, vários lócus CRISPR podem estar presentes (até um número máximo de 18, identificados na arquea *M. jannaschii*), chegando a representar até 1% das sequências do genoma.

2.4.3 Genes

Como discutido anteriormente (Seção 2.3), a maior parte do espaço disponível em um genoma procariótico é ocupada pelos genes. Aqui, serão tratados os aspectos mais amplos, ligados às características do repertório de genes encontrado em genomas procarióticos.

Em um genoma de qualquer espécie, há um **repertório de genes** que deve suprir todos os produtos gênicos (RNAs e proteínas) necessários à sobrevivência do organismo no(s) hábitat(s) por ele ocupado(s). Entretanto, para espécies procarióticas, além de os hábitats serem altamente variáveis (em condições como disponibilidade de nutrientes, temperatura, salinidade, pH, etc.), o impacto das condições ambientais é mais direto sobre o organismo, por ele ser unicelular e desprovido de compartimentalização intracelular (ao contrário de organismos e células eucarióticas). Assim, para sobrevivência em ambientes variáveis e relativamente hostis, foi fundamental que os organismos procarióticos adquirissem, ao longo da evolução, variabilidade também em nível de repertório genético entre diferentes populações ou linhagens de uma mesma espécie.

Hoje, a partir de estudos comparativos do repertório genético entre linhagens de uma mesma espécie bacteriana, observa-se que entre 20 a 35% dos genes podem ser linhagem-específicos, ou seja, estão no genoma de uma linhagem, mas não nos genomas de outras linhagens da mesma espécie. Isso dificulta uma distinção clara do ponto de vista genético entre espécies muito relacionadas ou linhagens muito divergentes de uma mesma espécie, o que levou à proposição do conceito de **pangenoma**, que descreve o repertório genético total de uma determinada espécie. O pangenoma incluiria um conjunto de genes compartilhado por todos os isolados, populações ou linhagens da espécie (chamado de genoma central) e um conjunto de genes presentes em algumas, mas não todas as linhagens da espécie (chamado de genoma dispensável) (**Figura 4.6**). Em teoria, o genoma central corresponde ao repertório de genes indispensáveis para a sobrevivência da espécie em qualquer ambiente, ao passo que o genoma dispensável corresponde ao repertório de genes determinantes de características fenotípicas linhagem-específicas, que podem ser adaptativas e necessárias à sobrevivência em algumas condições ambientais, mas não em outras.

Exemplos de genes que fariam parte da fração “dispensável” do pangenoma de uma espécie bacteriana seriam os que conferem resistência a agentes antimicrobianos e os que determinam sobrevivência e multiplicação em uma determinada espécie hospedeira. No primeiro caso, alguns genes do repertório “dispensável” codificam produtos que só seriam necessários à sobrevivência da espécie em condições de exposição a um agente antimicrobiano específico e que, portanto, seriam dispensáveis para linhagens que vivem em condições “normais”, livres da exposição a compostos bactericidas ou bacteriostáticos. No segundo caso, alguns genes do repertório “dis-

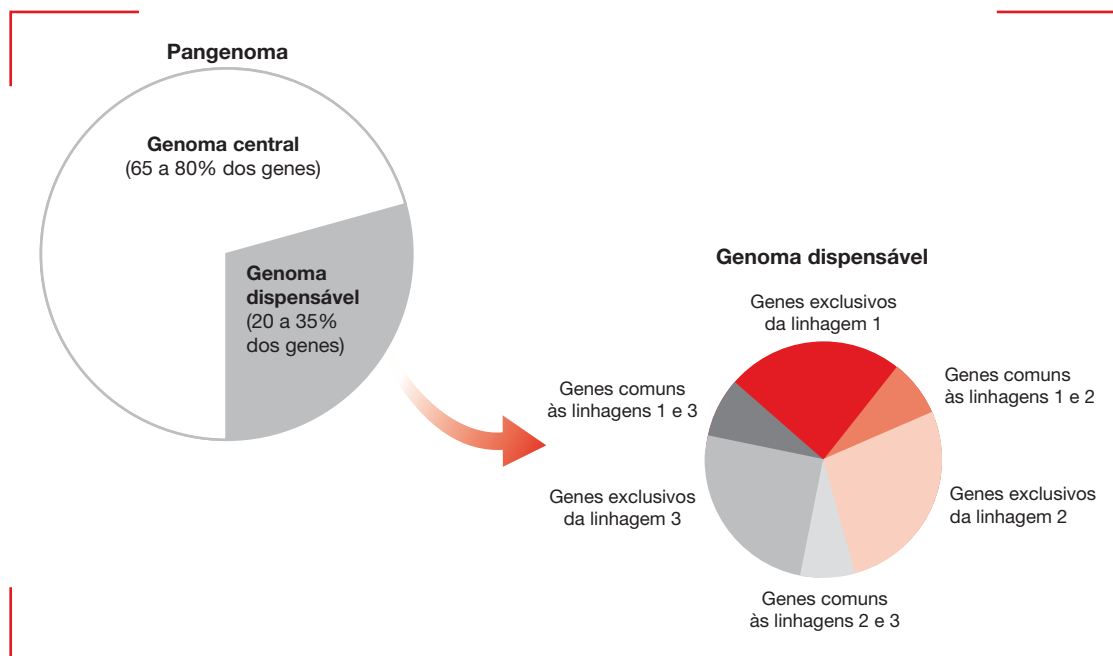


Figura 4.6

Diagrama representativo do conceito de pangenoma procariótico. No diagrama, é considerada a situação hipotética de uma espécie procariótica para a qual foram identificadas três linhagens. Do total de genes identificados nos genomas sequenciados das três linhagens, a maior fração (que pode ser de 65 a 80% deles) é comum a todas elas e constitui a fração do pangenoma chamada de genoma central. Os genes restantes (que podem ser de 20 a 35% do total) estão presentes em uma ou duas linhagens, porém ausentes na(s) linhagem(s) restante(s), constituindo a fração do pangenoma chamada de genoma dispensável.

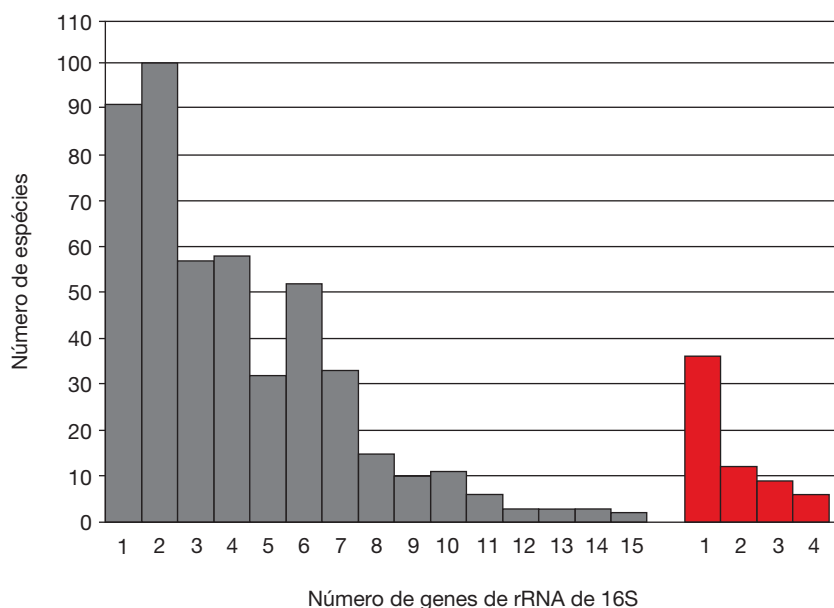
pensável” fazem com que uma ou mais linhagens sejam capazes de colonizar uma determinada espécie hospedeira, causando (se a bactéria for parasita/patogênica) ou não (se a bactéria for comensal ou simbiótica) um quadro patológico no hospedeiro. Linhagens sem os genes responsáveis pela capacidade de colonização seriam incapazes de colonizar a espécie hospedeira, mas manteriam (com os genes do genoma central e outros genes do genoma dispensável) a capacidade de sobreviver independentemente de um hospedeiro (em vida livre) ou em outra(s) espécie(s) de hospedeiro. Os mecanismos evolutivos determinantes da variabilidade de repertório gênico intraespecífico em procariotos e, portanto, da formação de repertórios gênicos pangenômicos em espécies de bactérias e arqueas serão discutidos na Seção 3.

Uma característica típica do repertório gênico de genomas procarióticos é a de que a maioria dos genes está presente em apenas uma cópia. Esse baixo grau de **redundância gênica** ocorre apesar da frequência relativamente elevada de eventos de duplicação gênica em genomas procarióticos, o que sugere que as duplicações produzidas são, em sua maioria, eliminadas (ver Seção 3). Os raros casos de redundância gênica observados em genomas de bactérias ou arqueas são, com frequência, explicáveis pelo fato de que o maior número de cópias dos genes envolvidos viabiliza o atendimento de uma demanda elevada dos produtos gênicos correspondentes

por parte da célula. Um bom exemplo disso são os genes codificadores de rRNAs, cujo número pode chegar a 15 em bactérias e 4 em arqueas (**Figura 4.7**). Assume-se que a existência de múltiplas cópias, associada a elevados níveis de transcrição, é uma estratégia para o suprimento da grande demanda de moléculas de rRNA para a montagem dos ribossomos.

No repertório gênico de genomas procarióticos também é significativo o número de genes parálogos e ortólogos, definidos na Seção 1.3. Em qualquer genoma bacteriano ou de arqueas pode ser identificado um número considerável de **famílias de genes parálogos**, que podem chegar a mais de 200 dependendo da espécie, com cada família podendo conter desde dois até algumas dezenas de membros. A presença de muitas famílias de genes parálogos em bactérias e arqueas, embora não constitua redundância – uma vez que os genes parálogos codificam produtos relacionados, mas funcionalmente distintos –, é considerada uma evidência de que a duplicação gênica seguida de divergência foi um mecanismo evolutivo importante para o aumento do repertório gênico em procariotos (ver Seção 3).

Quando a busca de homologia entre sequências gênicas é feita pela comparação entre genomas de diferentes espécies, também é marcante, em procariotos, a ocorrência de **grupos de genes ortólogos**. Para a maioria

**Figura 4.7**

Número de cópias de genes de rRNA de 16S em espécies de bactérias e de arqueas. Os dados representados no gráfico correspondem a uma análise de 476 espécies de bactérias (em ■) e 63 espécies de arqueas (em ■). A análise tem como base apenas genes de rRNA de 16S, mas pode ser extrapolada para os demais genes de rRNA (23S e 5S), assumindo-se a sua organização em um único óperon, comum à maioria das espécies procarióticas.

dos genes do genoma de qualquer espécie de bactéria ou arquea, há ortólogos identificáveis em outras espécies procarióticas, mesmo quando as comparações são feitas entre espécies pouco aparentadas, como em casos de comparação de uma bactéria com uma arquea. O número de genes ortólogos identificáveis na comparação de diferentes espécies e o grau de similaridade (homologia) entre eles são, em geral, diretamente proporcionais ao grau de parentesco evolutivo das espécies comparadas. Entretanto, desvios no resultado desse tipo de análise podem advir da presença frequente de genes xenólogos, adquiridos de forma independente, por transferência horizontal de espécies mais ou menos aparentadas do que as que estão sendo comparadas.

2.4.4 Óperons

Um **óperon** é uma unidade funcional do genoma, na qual duas ou mais regiões codificadoras de um produto gênico (de RNA ou proteico) ocupam posições adjacentes, estão coorientadas e têm a sua transcrição controlada por um mesmo conjunto de sequências reguladoras (**Figura 4.8**). Muitas vezes, por simplificação, as diferentes regiões codificadoras presentes em um óperon são referidas como “genes estruturais” ou apenas “genes”, e é comum que os diferentes “genes” de um óperon codifiquem produtos funcionalmente relacionados em algum grau (p. ex., em óperons que codificam diferentes tipos de rRNA ou em óperons que codificam diferentes enzimas que participam de uma mesma via metabólica).

As diferentes regiões codificadoras presentes em um óperon são **cotranscritas** em um único produto de RNA, que é por isso chamado de **policistrônico**. A nomenclatura de policistrônico é derivada de **cístron**, um conceito genético clássico equivalente a um gene (ou uma

região codificadora). A cotranscrição de mais de uma região codificadora, a partir de um mesmo conjunto de sequências reguladoras, representa uma otimização da maquinaria reguladora da célula, que assegura a expressão coordenada (ao mesmo tempo e nos mesmos níveis) dos produtos gênicos de cada óperon. Além disso, a organização em óperons proporciona uma economia de espaço no DNA, o que é relevante para os genomas relativamente pequenos e compactos de procariotos. O número de regiões codificadoras presentes varia de óperon para óperon, assim como também é variável o número de pares de bases que separa uma região codificadora da seguinte. A extensão das sequências entre as regiões codificadoras adjacentes de um óperon, chamadas de **regiões inter-cistrônicas**, varia tipicamente de um a poucas dezenas de pares de bases. Em casos extremos, pode ocorrer a sobreposição de duas regiões codificadoras, sendo a individualização de seus produtos de expressão definida graças ao controle independente para cada uma delas da etapa de tradução (discutida no Capítulo 12).

A presença de óperons é típica de genomas procarióticos, e o número de óperons em um genoma pode ser estimado por análise *in silico* (ver Capítulo 17), a partir da identificação de sequências codificadoras coorientadas e muito próximas umas das outras. Pode-se assumir que a maioria dos óperons preditos *in silico* (também chamados de **directons**) corresponde realmente a regiões codificadoras organizadas, quanto a sua funcionalidade, em unidades transcricionais policistrônicas (o que também pode ser confirmado experimentalmente). O número de sequências codificadoras organizadas em óperons em um genoma varia de espécie para espécie, sem qualquer correlação evidente com grupos taxonômicos (**Figura 4.9**). O que se pode afirmar é que a organização

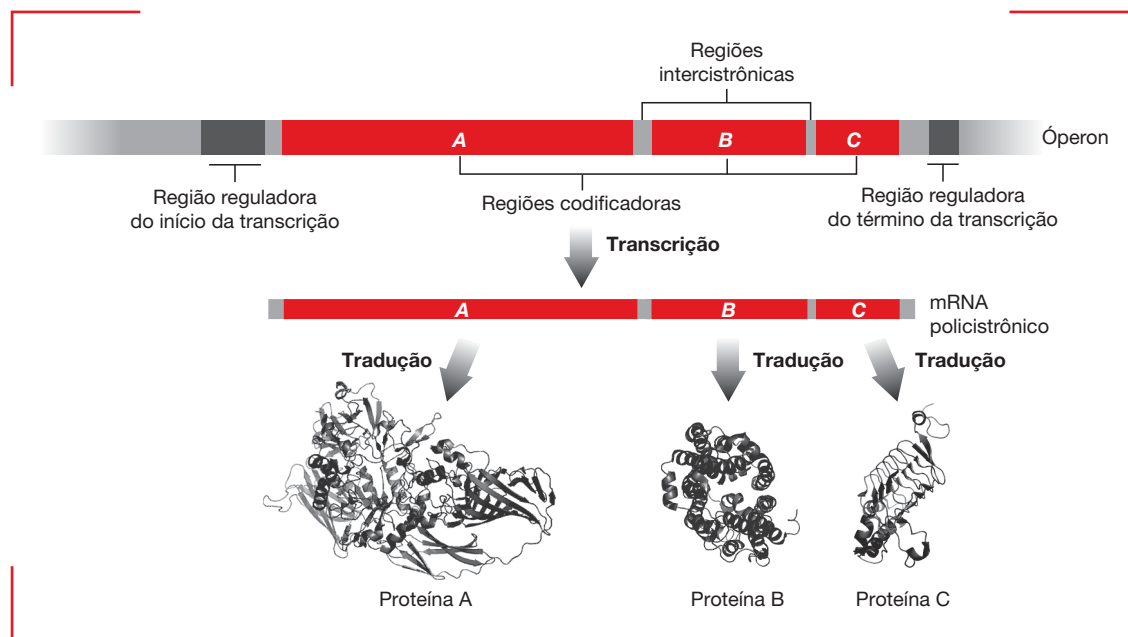


Figura 4.8

Representação esquemática da estrutura típica de um operon. O operon ilustrado contém três sequências codificadoras de proteínas (A, B e C), separadas por regiões intercistrônicas curtas, cujas extensões variam entre 1 e poucas dezenas de pares de bases em operons típicos. As três regiões codificadoras têm sua transcrição controlada por regiões reguladoras comuns, as quais determinam a produção de um mRNA único, que é por isso chamado de policistrônico. As três proteínas codificadas (A, B e C) pelo mRNA policistrônico são depois sintetizadas de forma independente nos ribossomos.

em operons é prevalente tanto em genomas de bactérias como de arqueas e que a fração de sequências codificadoras em operons é maior do que a daquelas transcritas individualmente. Com base nas estimativas disponíveis, representam os extremos da distribuição os genomas da arquea *A. pernix* e da bactéria *T. maritima*, com 54 e 90%, respectivamente, de suas sequências codificadoras organizadas em operons. Para *E. coli* K-12, são preditos em torno de 700 operons, com um número médio de 3 a 4 sequências codificadoras por operon, o que corresponde a mais de 50% do total estimado de genes no genoma dessa bactéria.

2.4.5 Domínios e macrodomínios

Os **domínios topológicos**, ou simplesmente **domínios**, de cromossomos procarióticos são regiões consideradas extensas, definidas, do ponto de vista estrutural, por interações com proteínas cromatínicas (ver Capítulo 3) e por padrões de superenrolamento do DNA (ver Capítulo 2). Aqui, o mais importante a salientar é que, em cromossomos bacterianos, esses domínios topológicos têm correlação com a distribuição de pelo menos alguns grupos de genes, os quais, em função disso, podem ter padrões transcripcionais até certo ponto coordenados. Em arqueas, pelo menos até agora, correlações estruturais e funcionais como essas não são tão evidentes, pois a estrutura da cromatina de arqueas é mais relacionada a da cromatina eucariótica (ver Capítulo 3).

As primeiras evidências da existência de domínios topológicos vieram de imagens de microscopia eletrônica de cromossomos bacterianos isolados, nas quais puderam ser visualizadas alças de DNA superenrolado, delimitadas por interações com proteínas cromatínicas. Dados experimentais recentes sugerem que, em cromossomos bacterianos, um domínio teria aproximadamente 10 kb de extensão, o que determinaria, por exemplo, a existência de mais de 400 domínios no cromossomo de *E. coli*. Entretanto, o número e os limites de cada domínio poderiam variar, dentro de limites não muito amplos, em diferentes momentos e em diferentes células. Essa dinâmica determinaria a ocorrência eventual de domínios de até 30 kb ou mais (já experimentalmente detectados), dependendo de interações transitórias com proteínas estruturais que atuariam como barreiras à livre difusão de superenrolamento ao longo do DNA.

Em uma escala hierarquicamente superior de domínios estruturais e funcionais foram definidos os chamados **macrodomínios**, que são superestruturas (conjuntos) de domínios topológicos que podem abranger boa parte de um cromossomo bacteriano. As barreiras estruturais (na forma de proteínas e estruturas de superenrolamento de DNA) entre macrodomínios seriam mais fortes do que aquelas que delimitam os domínios. No cromossomo de *E. coli*, por exemplo, foram identificados quatro macrodomínios, além de duas regiões me-

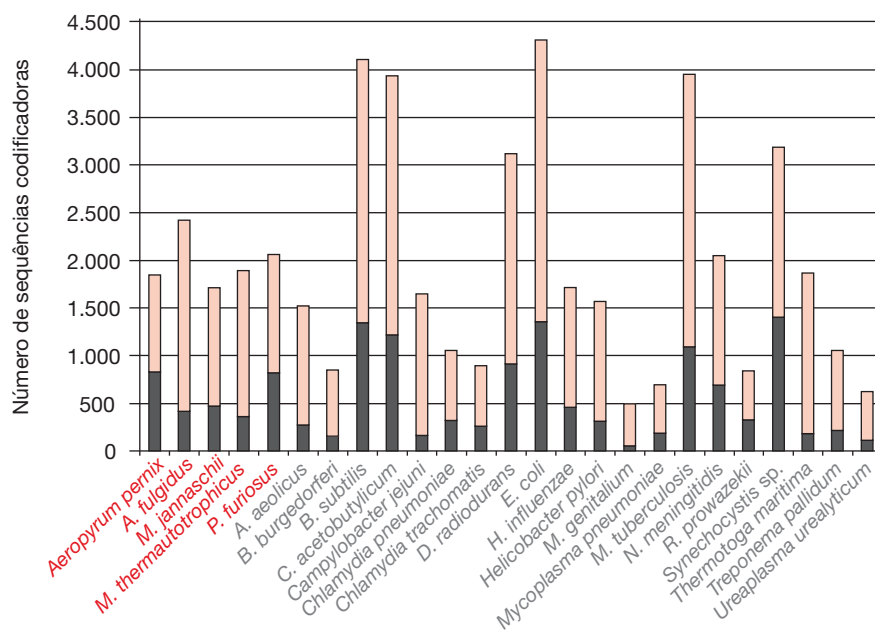


Figura 4.9

Estimativa do número de sequências codificadoras organizadas em óperons em genomas de algumas espécies de bactérias e arqueas. Os nomes das espécies de arqueas estão em ■ e os das bacterianas, em ■. Cada barra representa o número total de sequências codificadoras identificadas no genoma de uma espécie, estando representada em ■ a fração das sequências codificadoras organizadas em óperons e, em ■, a fração de sequências codificadoras individuais. A predição de organização em óperons tem como base análises *in silico*.

nos estruturadas (sem características de macrodomínios) (Figura 4.10).

Os macrodomínios foram identificados em *E. coli* a partir da detecção, ao longo do cromossomo, de grandes regiões (de até centenas de milhares de pares de bases) que são, pelo menos até certo ponto, compartimentalizadas funcionalmente. Por exemplo, a probabilidade de ocorrência de eventos de recombinação é maior entre dois sítios que estão a uma determinada distância um do outro, mas que estão em um mesmo macrodomínio, do que entre dois sítios com distância similar entre eles, mas situados em macrodomínios diferentes. Além disso, há semelhanças evidentes entre as atividades (expressão) de genes dentro de um mesmo macrodomínio, cujo padrão difere daquele de outros domínios.

Do ponto de vista da distribuição dos genes em um cromossomo bacteriano, cada domínio ou macrodomínio definiria uma região, na qual um grupo de genes teria a sua expressão restringida (ou não) por uma determinada situação topológica (grau de superenrolamento do DNA, p. ex.), que seria diferente da encontrada em outros domínios ou macrodomínios em razão da atuação diferencial de diversas topoisomerasas (discutidas no Capítulo 2). Assim, genes distribuídos em um determinado domínio devem ter sua transcrição adequada às condições estruturais do domínio/macrodomínio topológico onde estão posicionados. Por exemplo, no cromossomo de *E. coli*, é evidente que a distribuição dos genes em relação à organização do cromossomo em domínios e macrodomínios não é aleatória. Há uma concentração importante de genes cuja transcrição é reprimida pelo relaxamento do superenrolamento próximo à origem de replicação. Em contrapartida, há um grande grupo de genes posiciona-

dos junto à região de término da replicação, cuja transcrição é estimulada pelo relaxamento do superenrolamento do DNA.

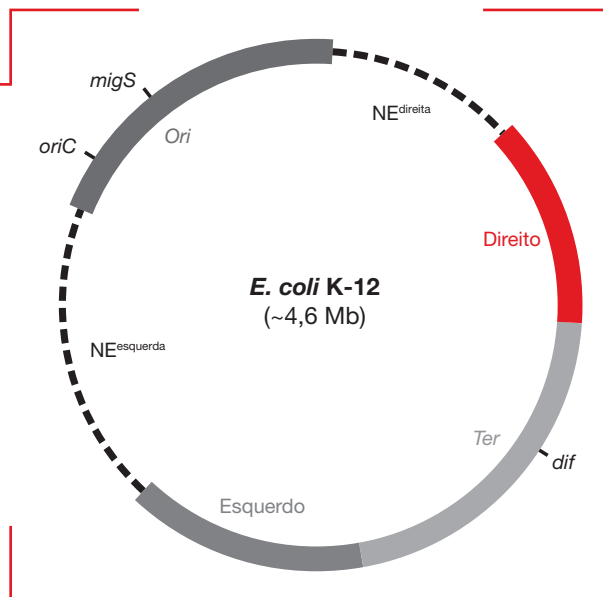
2.4.6 Ilhas genômicas

As **ilhas genômicas**, ou simplesmente **ilhas**, são segmentos de genomas procarionóticos que diferem em algumas características do restante do genoma, como em conteúdo de GC e padrões de sintenia e orientação gênicas (discutidos nas Seções 2.5 e 2.6). Essas ilhas podem estar presentes ou não em diferentes isolados ou linhagens de uma mesma espécie ou em diferentes espécies relacionadas (de um mesmo gênero, p. ex.). A estrutura e a composição gênica das ilhas podem ser bem variáveis, mas em geral elas estão associadas em algum grau a elementos genéticos móveis, como bacteriófagos e elementos conjugativos ou transponíveis (ver Capítulo 9), e a eventos de transferência gênica horizontal.

Do ponto de vista estrutural, algumas generalizações podem ser feitas em relação às ilhas genômicas. Elas são relativamente extensas, podendo variar de 10 até mais de 200 kb. Regiões menores que 10 kb com outras características de ilhas genômicas são chamadas de **ilhotas** por alguns autores. No que diz respeito à composição nucleotídica, em geral o conteúdo de GC e outros parâmetros (ver Seções 2.5 e 2.6.2) das ilhas são diferentes do restante do genoma. O local de inserção de uma ilha em um genoma pode ser variável, mas é comum a inserção em genes de tRNA, o que é associado a uma origem ligada à mobilidade de elementos conjugativos integrativos (discutidos no Capítulo 9). A associação das ilhas a elementos genéticos móveis também é evidenciada pela frequente presença de sequências flanqueadoras repetidas diretas, de 16 a 20 pb,

Figura 4.10

Representação esquemática dos macrodomínios do cromossomo de *E. coli*. Os macrodomínios *Ori*, *Ter*, direito e esquerdo estão identificados por barras em diferentes cores (em ■, ■, ■ e ■, respectivamente). As linhas tracejadas interrompidas indicam as regiões não estruturadas em macrodomínios esquerda (*NE*^{esquerda}) e direita (*NE*^{direita}). Para referência, as posições dos loci *oriC*, *migS* e *dif* estão indicadas. O macrodomínio *Ori* contém a origem de replicação (*oriC*) e está centrado no locus *migS*, uma sequência análoga a um centrômero que está envolvida no posicionamento bipolar (segregação) da *oriC* logo depois dela ter sido replicada. O macrodomínio *Ter*, que inclui a região de encontro das forquilhas de replicação, está centrado em *dif*, um locus de recombinação sítio-específica que permite a resolução em monômeros dos dímeros de cromossomos resultantes da replicação. Ver Capítulo 6 para mais informações sobre o processo de replicação do cromossomo de *E. coli*.



que geralmente resultam de eventos de integração sítio-específica (nesse caso, da ilha no seu sítio-alvo). Tais sítios podem funcionar como sítios-alvo de enzimas capazes de levar à excisão da ilha do genoma e, muitas vezes, as próprias enzimas envolvidas em processos de excisão ou integração, como as integrases (Capítulo 8), podem ser codificadas por genes presentes nas próprias ilhas.

Quanto à composição gênica, as ilhas contêm genes considerados não essenciais, mas que, em determinadas condições, podem conferir alguma vantagem adaptativa à célula. Um exemplo comum é o das chamadas **ilhas de patogenicidade**, que ocorrem em muitas espécies bacterianas patogênicas, como *H. pylori*, *S. enterica* ou *V. cholerae*. Ilhas de patogenicidade contêm genes cujos produtos conferem à célula capacidade de infectar e causar doença em alguma espécie hospedeira, por exemplo, genes que codificam toxinas, proteínas imunomoduladoras ou proteínas que medeiam a adesão a determinados tipos de células. A capacidade codificadora das ilhas não fica limitada, contudo, a determinantes de patogenicidade (ou virulência) e pode ser bastante diversa. Podem ser citadas, dentre outras, ilhas genômicas com genes associados ao metabolismo de compostos incomuns, como compostos aromáticos; à síntese de sideróforos; à resistência a metais pesados; à resistência a agentes antimicrobianos e à adaptação a um modo de vida simbiótico.

Para exemplificar o conceito de ilha genômica será analisada uma ilha de patogenicidade da bactéria gram-negativa *H. pylori*, que coloniza o estômago de seres humanos e pode causar diversas doenças gástricas, desde gastrites crônicas até úlceras pépticas e linfomas de tecido linfóide associado à mucosa gástrica. O genoma de linhagens virulentas de *H. pylori* apresenta uma ilha de patogenicidade chamada de *cag*, que consiste em uma região cromossômica de aproximadamente 37 kb contendo 31 genes (Figura 4.11). A principal proteína codificada

na ilha *cag* é o antígeno CagA (produto do gene *cagA*, situado próximo da extremidade direita, ou 3', da ilha). CagA é uma proteína secretada, e sua translocação para células da mucosa gástrica leva a alterações morfológicas importantes na célula hospedeira, as quais podem resultar na transformação maligna. A maioria dos produtos dos outros genes *cag* está envolvida com o sistema de secreção e translocação da CagA para o citoplasma de células do epitélio gástrico. Algumas proteínas codificadas pela ilha, ainda, induzem células dos hospedeiros a produzirem citocinas (proteínas de sinalização com função imunomoduladora), que alteram a resposta imune do hospedeiro de modo favorável ao patógeno.

Além dos genes codificadores de fatores de virulência, a ilha *cag* também apresenta outras características típicas de ilhas genômicas. Ela é flanqueada por uma repetição direta de 31 pb e contém, a jusante do gene *cagA*, duas sequências de inserção (IS, de *insertion sequences*) pertencentes a uma família de elementos transponíveis (IS605) comum no genoma de *H. pylori*. As repetições diretas e as IS estão associadas a eventos de recombinação e mobilidade genética da ilha *cag*.

2.5 Conteúdo de G+C

O **conteúdo médio de G+C** (ou, simplesmente, o **conteúdo de GC**) de um genoma é a fração (porcentagem) de nucleotídeos de guanina e citosina que ele apresenta, que é complementar à fração de nucleotídeos de adenina e timina. O conteúdo de GC (GC%) pode então ser calculado pela fórmula:

$$GC(\%) = (G+C/A+T+G+C) \times 100$$

Onde G, A, T e C representam a quantidade de nucleotídeos de guanina, adenina, timina e citosina, respectivamente, no genoma.

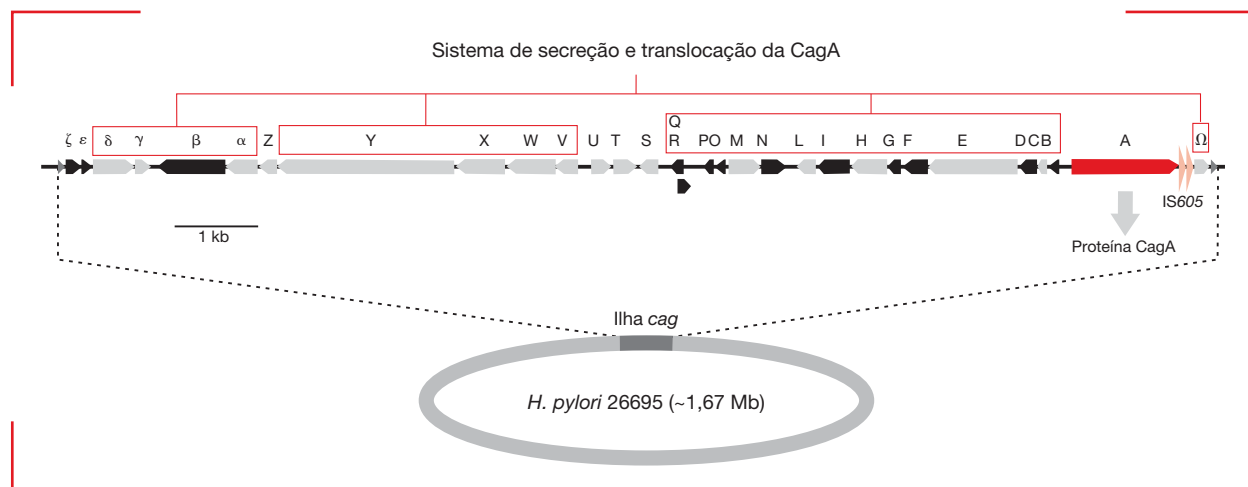


Figura 4.11

Ilha de patogenicidade *cag* de *H. pylori*. A configuração da ilha *cag* no cromossomo da linhagem 26695 de *H. pylori* está representada. Os genes codificadores de proteínas e suas respectivas orientações estão representados por setas em ■, ■ ou ■, e a letra (grega ou romana) que designa cada um deles está representada abaixo de cada seta. O gene *cagA* (em ■) codifica a proteína secretada CagA. A maioria dos demais genes (nomes assinalados por retângulos em ■) codifica proteínas essenciais para a secreção e a translocação de CagA para células do hospedeiro, sendo que as proteínas codificadas pelos genes *cag*, *cag*, *cagY*, *cagX*, *cagV*, *cagE* e *cagC* são ortólogas de proteínas do sistema de secreção do tipo IV do patógeno de plantas *A. tumefaciens*. Dezesesseis genes da ilha (setas em ■) codificam produtos que levam células da mucosa gástrica a produzirem interleucina 8 (IL-8, uma citocina), o que modula a resposta imune do hospedeiro de modo favorável à bactéria. As pontas de seta menores (em ■, em cada uma das extremidades da ilha) representam as duas cópias da repetição direta de 31 pb e as pontas de seta maiores (em ■) representam as duas cópias da IS605; as pontas de seta também indicam as orientações de cada um desses elementos de sequência.

O conteúdo de GC varia de espécie para espécie e, em procarionotos, essa variação é ampla, indo de 22,5%, no genoma da *Wigglesworthia glossinidia* (proteobactéria endossimbionte da mosca tsé-tsé), até 72,1%, no genoma de *S. coelicolor* A3(2) (uma actinobactéria de solo). Embora os exemplos extremos até agora conhecidos sejam de bactérias, os genomas de arqueas também variam bastante em termos de conteúdo de GC, com distribuição entre extremos de pouco mais de 30% até pouco menos de 70%.

Os exemplos extremos mencionados ilustram também um outro fato interessante: há certa correlação entre o tamanho do genoma e o conteúdo de GC, com genomas menores (como o genoma de *W. glossinidia*, com 0,69 Mb) tendendo a ser menos ricos em GC (ou mais ricos em AT) do que genomas maiores (como o de *S. coelicolor*, com 8,7 Mb).

Comparações do conteúdo de GC entre genomas de espécies de diferentes filos bacterianos também sugerem algumas tendências, como a de genomas de actinobactérias serem mais ricos em GC do que genomas de firmicutes, por exemplo. Entretanto, as tendências ou correlações mais marcantes com o conteúdo de GC parecem não estar associadas a grupos taxonômicos específicos, mas sim ao estilo de vida das espécies. Espécies parasitas ou simbióticas, que também tendem a ter genomas menores (ver Seção 2.1), possuem genomas mais ricos em AT (e pobres em GC), ao passo que espécies de

vida livre tendem a ter genomas maiores e mais ricos em GC. O conteúdo de GC médio para genomas de espécies dependentes de hospedeiro seria de cerca de 38%, e o conteúdo de GC médio para genomas de espécies de vida livre, de 49%.

Não há ainda explicações definitivas para as correlações existentes entre o conteúdo de GC, o tamanho dos genomas e o estilo de vida das espécies. Dentre as explicações possíveis está o fato de nucleotídeos de G e C serem energeticamente mais custosos para a célula e, por isso, a presença deles seria mais limitada em espécies com metabolismo mais deficiente, como parasitas e simbiotes. Outra explicação residiria no fato da maioria das espécies parasitas e simbiotes (e com genomas reduzidos) ser deficiente em pelo menos alguns genes responsáveis pela reparação de mutações (que é assunto do Capítulo 7). Assim, como as mutações pontuais mais frequentes são as de C para T ou de G para A, haveria uma tendência de acúmulo de AT em relação a GC nos genomas dessas espécies.

Cabe ainda salientar que o conteúdo de GC discutido representa uma média da composição nucleotídica de um genoma. Porém, nenhum genoma tem um conteúdo de GC homogêneo ao longo de toda a sua extensão. Por exemplo, em qualquer genoma procarionótico, sequências de origens de replicação e promotoras da transcrição tendem a ser mais ricas em AT e, mesmo

em genomas pequenos de espécies parasitas ou sim-biontes, a tendência a um maior conteúdo de AT é maior em regiões inter-gênicas do que nas regiões codificadoras dos genes. Isso ocorre porque, dependendo das funções que cada tipo de região exerce, diferentes pressões seletivas atuam sobre elas (ver Seção 3), determinando também diferenças nos respectivos conteúdos de GC.

2.6 Distribuição dos genes e de outras sequências em genomas procarióticos

Os genes (e também outras sequências) de procariotos são, até certo ponto, independentes do ponto de vista funcional, pois cada gene (ou pequeno grupo de genes, no caso de óperons – ver Seções 1.1 e 2.4.4) é autônomo na regulação de sua expressão. Apesar disso, quando analisado o posicionamento relativo dos genes e de outras sequências em genomas procarióticos, são identificáveis padrões de distribuição não aleatórios. Esses padrões de distribuição, que definem a **arquitetura** de cada genoma, foram estabelecidos e são mantidos ou modificados ao longo da evolução em razão de diversos fatores, desde a mecânica molecular do processo de replicação do DNA (Capítulo 6) até a frequência de rearranjos cromossômicos ou de transferência gênica horizontal na história evolutiva de uma espécie (Seção 3). Aqui, serão descritos alguns padrões típicos encontrados na arquitetura de genomas procarióticos e discutido as suas possíveis implicações funcionais.

2.6.1 Orientação e localização de genes em relação à origem e ao sentido da replicação

Como discutido na Seção 2.2, a maioria dos genomas procarióticos contém uma única origem de replicação, a partir da qual o genoma é replicado bidirecionalmente (ver Capítulo 6). Essa origem de replicação única representa o sítio a partir do qual há uma troca entre aquela que é a fita-líder e a que é a fita tardia em cada metade do cromossomo, definindo a região cromossômica que será replicada em primeiro lugar. Esses fatores parecem ter sido, ao longo da evolução, determinantes de padrões de distribuição e de orientação de genes em cromossomos procarióticos.

Quanto à orientação dos genes em relação à origem (ou ao sentido) de replicação, foi observada uma predominância de genes na fita-líder (ou seja, genes cuja sequência codificadora está na fita replicada continuamente de cada metade do cromossomo). Esse padrão, denominado **tendência gênica de fita** (de *gene strand bias*), implica dizer que muitos genes estão coorientados em relação ao sentido da replicação, estando no sentido horário, quando situados na metade do cromossomo replicado nesse mesmo sentido, e, no sentido anti-horário, quando na outra metade do cromossomo (**Figura 4.12**). Essa tendência de coorientação dos genes em relação ao sentido da replicação é mais evidente em genomas bacte-

rianos, mas também é observada em genomas de arqueas, principalmente naqueles com uma única origem de replicação.

A explicação para a coorientação dos genes e da replicação reside no fato desta organização evitar ou minimizar o efeito negativo de colisões entre a maquinaria proteica de replicação (DNA-polimerase e proteínas a ela associadas) e RNA-polimerases que estão transcrevendo genes. Colisões entre a maquinaria de replicação e a RNA-polimerase em situações nas quais elas se movimentam em direções opostas podem levar à interrupção de eventos de transcrição, bem como levariam a um significativo atraso da replicação se fossem muito frequentes. Em contrapartida, quando o sentido de deslocamento das RNA-polimerases em transcrição é o mesmo da maquinaria de replicação para a maioria dos genes, as eventuais colisões (em razão da movimentação mais rápida da DNA-polimerase) não são abortivas para eventos transcricionais, e a replicação não é atrasada.

Inicialmente, foi considerado que isso só significaria uma vantagem importante para genes com elevados níveis de transcrição, como genes relacionados aos processos de síntese de ácidos nucleicos ou de proteínas, para os quais a coorientação entre transcrição e replicação foi evidenciada primeiro (ver Figura 4.12). Entretanto, foi demonstrado que, pelo menos para espécies bacterianas dos filos Firmicutes e Proteobacteria, a coorientação entre transcrição e replicação ocorre para a maioria dos genes e está mais correlacionada com a essencialidade de cada gene do que com seu nível de expressão. Por exemplo, em *B. subtilis*, em torno de 95% dos genes essenciais (ou seja, aqueles indispensáveis para a sobrevivência em qualquer condição ambiental) estão coorientados com a replicação, bem como 75% dos demais genes, independentemente do nível de transcrição de cada um.

Outro determinante de organização gênica em genomas procarióticos associado à replicação é o número de eventos de replicação que são iniciados a cada ciclo de divisão celular. *E. coli*, por exemplo, divide-se aproximadamente a cada 20 minutos em condições ideais, mas um processo de replicação completo de seu cromossomo dura até três vezes mais. Por isso, mais de um processo de replicação deve ser iniciado a cada ciclo celular, para que as células-filhas, ao herdarem cromossomos parcialmente replicados, possam concluir um ciclo de replicação antes da próxima divisão celular. Logo, um cromossomo herdado por uma célula-filha já pode estar parcialmente duplicado uma ou mais vezes, o que determina um **efeito de dosagem gênica associado à replicação**. Nessas situações, genes mais próximos da origem de replicação apresentam um número de cópias mais elevado do que o número de cópias de genes próximos à região de término da replicação, o que garante níveis de expressão aumentados aos primeiros (**Figura 4.13A**). Apesar de eventual e variável em intensidade dependendo das condições ambientais, o efeito de dosagem gênica associado à replicação parece ter

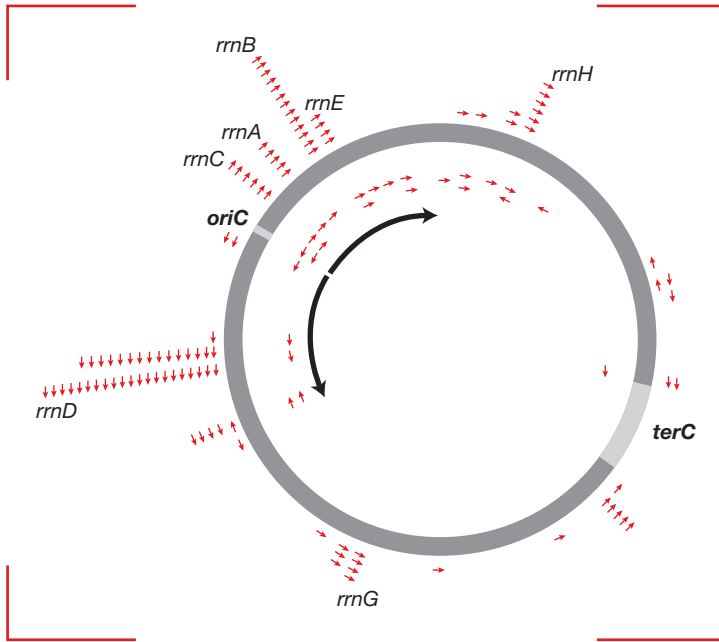


Figura 4.12

Orientação de alguns genes no cromossomo de *E. coli*, mostrando a tendência de coorientação com o sentido da replicação. Cada seta em ■ corresponde a um gene e indica a direção da sua transcrição; apenas genes transcritos em níveis elevados estão representados. As setas do lado externo do círculo correspondem a genes que fazem parte da maquinaria de síntese proteica, incluindo aqueles organizados nos sete ôperons responsáveis pela síntese de rRNA (*rrnA, B, C, D, E, G e H*); as setas no interior do círculo correspondem a genes necessários para a síntese de DNA (replicação) ou RNA (transcrição). As posições da origem (*oriC*) e da região de término (*terC*) da replicação estão assinaladas, e as setas em ■ indicam o sentido da replicação em cada metade do cromossomo.

vido um fator evolutivo importante para a distribuição de alguns genes ao longo do cromossomo, pelo menos em genomas de espécies procarióticas que se multiplicam com rapidez. Os genes para os quais a localização preferencial em regiões próximas à origem é mais evidente são os relacionados à transcrição (como os que codificam subunidades da RNA-polimerase) ou à tradução (como genes de rRNA ou proteínas ribossômicas), o que pode ser observado na arquitetura do genoma de *E. coli* (Figura 4.13B).

2.6.2 Desvios de GC

Na maioria dos genomas procarióticos, há uma assimetria marcante entre as fitas-líder e tardia em cada metade do cromossomo, no que diz respeito à composição nucleotídica. Essa assimetria, conhecida como **desvio de GC**, decorre do fato da fita-líder ser mais rica em nucleotídeos de guanina do que de citosina e a fita tardia ser mais rica em nucleotídeos de citosina do que de guanina. Em uma determinada fita (ou segmento de fita) de DNA, seja ela replicada continuamente (fita-líder) ou descontinuamente (fita tardia), o desvio de GC (GC_{desvio}) pode ser calculado pela fórmula:

$$GC_{\text{desvio}} = (G - C)/(G + C)$$

Onde G e C representam a quantidade de nucleotídeos de guanina e citosina, respectivamente, na fita (ou em qualquer segmento dela considerado).

O cálculo do desvio de GC ao longo de toda uma das fitas de um cromossomo circular procariótico pode ser útil, por exemplo, para identificar sítios de início e término da replicação (Figura 4.14). Na representação gráfica dos resultados de uma análise *in silico* como essa, fica evidente um padrão bifásico determinado por dois pontos

de transição, um de valores de GC_{desvio} positivos para negativos e outro de valores de GC_{desvio} negativos para positivos. Esses pontos de transição correspondem aos sítios de término e origem da replicação, pois mostram quando a fita analisada passa de uma composição um pouco mais rica em nucleotídeos de G (típica da metade do cromossomo na qual ela foi replicada continuamente) para uma composição um pouco mais rica em nucleotídeos de C (típica da metade do cromossomo na qual ela foi replicada descontinuamente).

Os desvios de GC em genomas procarióticos parecem ser consequência de múltiplos fatores. Eles podem ser explicados, pelo menos em parte, pela ação diferencial de forças mutacionais sobre a fita-líder e sobre a fita tardia, em razão de seus modos distintos de replicação (contínuo para a fita-líder e descontínuo para a fita tardia). O molde (uma fita-líder) da fita tardia, por ficar mais tempo exposto na forma de fita simples durante a replicação, seria mais vulnerável às mutações mais frequentes. Como as mutações mais frequentes (de modo especial em DNA de fita simples) são justamente as desaminações de citosinas que originam timinas (C→T), haveria então uma tendência de diminuição da quantidade de nucleotídeos de citosina em relação aos de guanina na fita-líder. Além disso, haveria ainda uma pressão seletiva maior contra mutações frequentes (como C→T) na fita-líder, já que ela contém as fitas codificadoras da maioria dos genes essenciais (ver Seção 2.6.1) e, por isso, haveria uma maior probabilidade de que eventuais mutações nessa fita fossem deletérias, ao contrário da situação na fita tardia. Finalmente, outra causa de desvios de GC seria a presença preferencial na fita-líder de motivos repetidos e orientados, ricos em nucleotídeos de guanina, como é o caso dos sítios *Chi*, discutidos na Seção 2.4.1.

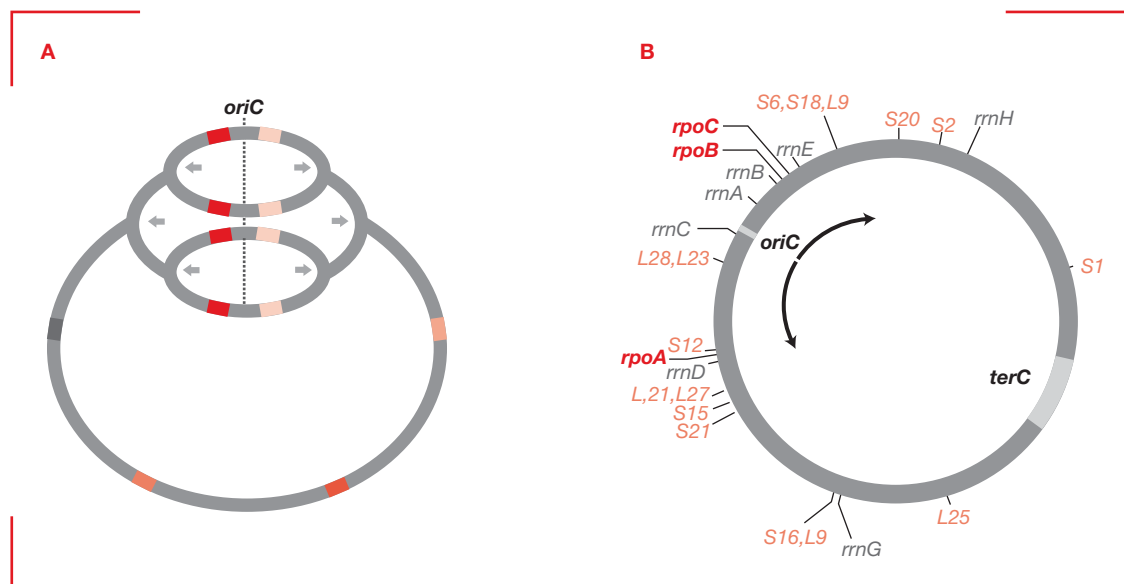


Figura 4.13

O efeito de dosagem gênica associado à replicação. (A) Representação esquemática de um cromossomo procariótico típico no qual três processos de replicação bidirecionais (indicados por setas divergentes em ■) foram iniciados durante um mesmo ciclo celular. Em situações como essa, genes posicionados mais próximos à origem de replicação (*oriC*), representados em ■ e ■, tendem a apresentar transitariamente um número maior de cópias do que genes localizados mais distantes dela, representados em outras cores. (B) Representação do cromossomo de *E. coli* K-12 mostrando a posição dos óperons de genes de rRNA (em ■) e de alguns dos genes que codificam proteínas ribossômicas (em ■) ou subunidades da RNA-polimerase (em ■). Pode-se observar uma maior tendência de localização desses genes em posições mais próximas a *oriC* do que à região de término da replicação (*terC*). As setas em ■ indicam o sentido da replicação em cada metade do cromossomo.

2.6.3 Sintenia: conservação da ordem dos genes

Em estudos genômicos, **sintenia** é o termo que descreve situações de conservação da ordem de genes e outras sequências ortólogas entre dois ou mais genomas. A análise da sintenia exige, portanto, comparações entre genomas de diferentes espécies, ou entre diferentes linhagens de uma mesma espécie, e depende da identificação das sequências ortólogas nos genomas comparados. Uma das maneiras mais simples para demonstrar sintenia entre dois genomas é o alinhamento das sequências e a plotagem do resultado da comparação na forma de uma matriz de pontos (*dot plot*), que representa graficamente a posição de cada segmento de sequência de um genoma *versus* a posição de seu segmento ortólogo no outro genoma (Figura 4.15). Quando espécies ou linhagens próximas são assim comparadas, a plotagem gera uma linha diagonal contínua, o que indica uma situação de colinearidade (mesma ordenação) para a grande maioria das sequências (e, consequentemente, dos genes) nos dois genomas (Figura 4.15A). Em alinhamentos de genomas de espécies um pouco menos relacionadas (moderadamente divergentes) (Figura 4.15B e C), o padrão mais observado nas matrizes de pontos é o de duas diagonais

truncadas e de orientações contrárias. Esse padrão em X é resultante de uma inversão principal, centrada na origem de replicação, do cromossomo de uma espécie em relação ao cromossomo da outra. Inversões menores determinam as descontinuidades observadas nas duas diagonais principais. Na comparação de genomas de espécies mais divergentes entre si (Figura 4.15D), há uma perda quase total de sintenia, e os pontos correspondentes aos pares de sequências ortólogas ficam espalhados em toda a área da matriz.

A acumulação de inversões ao longo do período evolutivo decorrido desde a separação das linhagens ou espécies comparadas de seu ancestral comum mais recente parece ser o determinante principal da perda progressiva de sintenia entre procariotos. Dentre as inversões, aquelas simétricas, centradas na origem de replicação e abrangendo grande parte do cromossomo (principais responsáveis pelo padrão em X das matrizes de pontos), são bastante comuns. A predominância dessas inversões, chamadas de recíprocas, e de outras inversões menores sobre outros tipos de rearranjos cromossômicos, como translocações, pode ter várias explicações plausíveis. As inversões poderiam, por exemplo, ter uma probabilidade natural maior de ocorrência e/ou ser mais facilmente fixadas no genoma de uma espécie ao longo da evolução. Um fator importante pode ser o

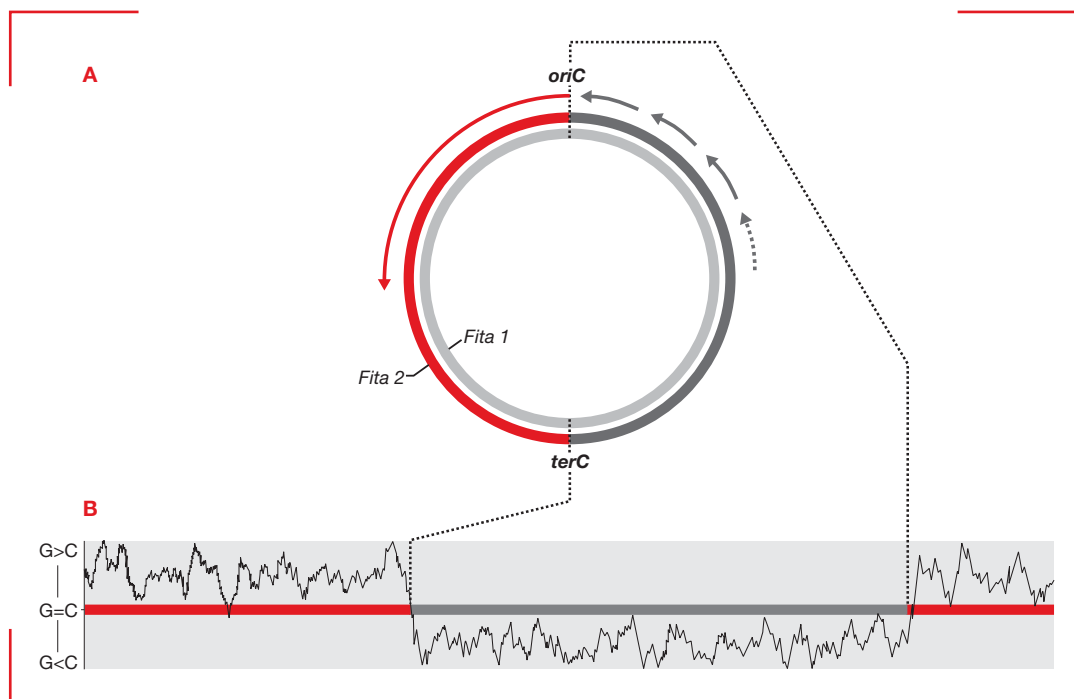


Figura 4.14

Análise de desvios de GC para identificação dos sítios de início (*oriC*) e término (*terC*) da replicação em um cromossomo procarionótico típico. (A) Representação esquemática da estrutura e da replicação do cromossomo de *E. coli*; como a análise de desvios de GC é feita para cada fita de forma individual, apenas uma delas (arbitrariamente denominada *Fita 1*) será considerada aqui, sendo a outra fita (denominada *Fita 2*) mostrada em cinza apenas para referência. A metade da *Fita 1* representada em vermelho corresponde à sua porção replicada continuamente (como indicado pela seta em vermelho contínua) e, portanto, a uma região de fita-líder; a outra metade, representada em cinza, corresponde à porção da *Fita 1* replicada descontinuamente (como indicado pelas setas consecutivas em cinza) e, portanto, a uma região de fita tardia. (B) Gráfico resultante da análise de desvios de GC (GC_{desvio}) da *Fita 1*, com as regiões de fita-líder e de fita tardia representadas por barras em vermelho e cinza, respectivamente. Na análise *in silico* dos desvios de GC, segmentos de tamanho definido (em geral de 1 kb ou 10 kb), sucessivos e parcialmente sobrepostos têm os respectivos valores de GC_{desvio} calculados; depois, os valores calculados são plotados em um gráfico em função da posição de cada segmento ao longo do cromossomo. Na análise do gráfico, é observada uma clara transição de valores de GC_{desvio} negativos (quantidade de G menor que a de C) para valores de GC_{desvio} positivos (quantidade de G maior que a de C) na posição do genoma correspondente à transição da metade da fita replicada descontinuamente para a metade da fita replicada continuamente, o que indica a posição da *oriC*. Uma segunda transição, de valores de GC_{desvio} positivos para valores de GC_{desvio} negativos, é observada na transição da metade da fita replicada continuamente para a metade da fita replicada descontinuamente, o que indica a posição de *terC*.

fato de as inversões recíprocas não alterarem as extensões de cada metade do cromossomo circular, nem a distância ou a orientação dos genes em relação à origem de replicação. Assim, essas inversões não causariam qualquer desequilíbrio na dinâmica da replicação bidirecional, na preferência gênica de fita, ou na dosagem gênica associada à replicação.

2.6.4 Superóperons e ultraóperons

Como discutido na Seção 2.6.3, a sintenia gênica observada na análise comparativa de genomas tende a diminuir de forma proporcional ao grau de divergência evolutiva entre as espécies. Quando alguns casos localizados de

conservação de sintenia ainda são observados entre genomas predominantemente não sintênicos, em geral eles correspondem a óperons (Seção 2.4.4). Entretanto, mesmo os óperons apresentam uma considerável plasticidade organizacional ao longo da evolução, manifestada na forma de quebras de óperons maiores (com mais genes) em óperons menores (com menos genes) e de fusões de óperons menores para a formação de óperons maiores, o que determina limitações adicionais à sintenia entre genomas. Por exemplo, excluindo-se casos de espécies muito aparentadas, apenas 5 a 25% dos genes de bactérias e arqueas estão organizados em óperons compartilhados por dois ou mais genomas.

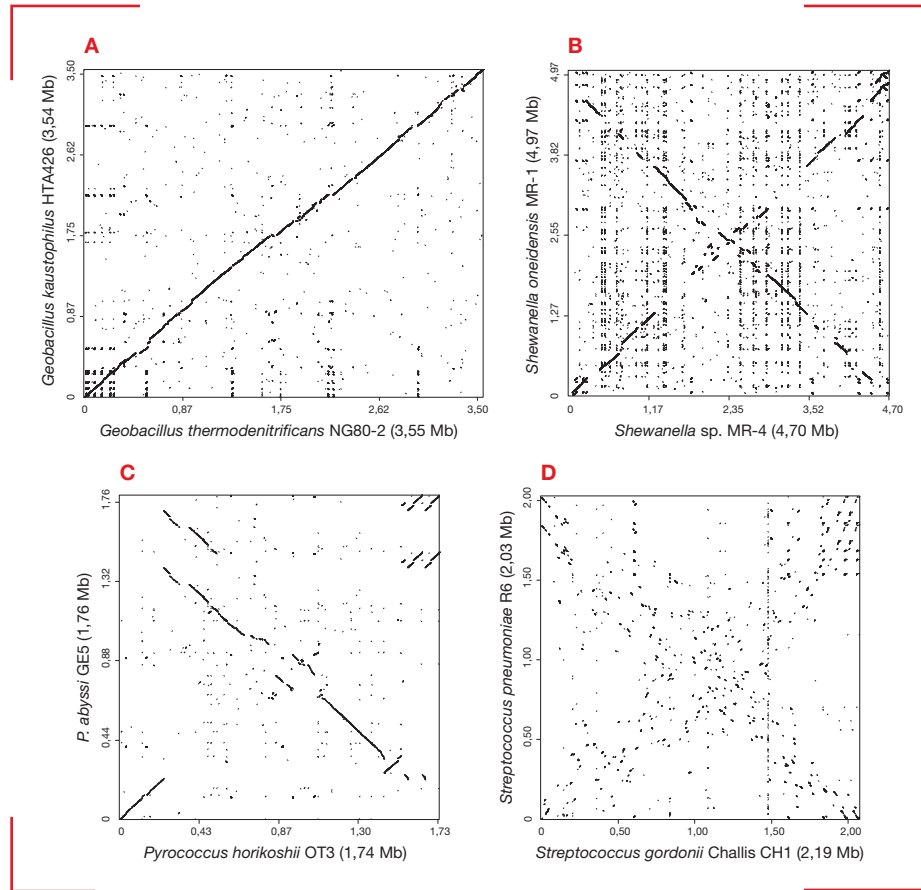


Figura 4.15

Matrizes de pontos para alinhamento e identificação de sintenia entre genomas de espécies com diferentes graus de parentesco evolutivo.

Em todas as matrizes, as posições ao longo de cada genoma comparado estão indicadas (escala em Mb) no eixo das abscissas (primeira espécie) e no eixo das ordenadas (segunda espécie); cada ponto em uma matriz indica um pareamento entre sequências homólogas das duas espécies comparadas. **(A)** Na comparação do genoma de duas espécies proximamente relacionadas (no caso, as bactérias *G. thermodenitrificans* NG80-2 e *G. kaustophilus* HTA426), um elevado grau de coincidência na ordenação dos genes ortólogos (sintenia) é evidenciado por uma linha diagonal praticamente sem interrupções. **(B)** Na comparação dos genomas de duas espécies moderadamente divergentes (no caso, as bactérias *Shewanella* sp. MR-4 e *S. oneidensis* MR-1), o padrão em X evidencia a ocorrência de uma inversão centrada na origem de replicação e envolvendo mais de dois terços do cromossomo de uma espécie em relação ao da outra. **(C)** Situação similar à descrita em (B), mas envolvendo duas espécies de arqueas moderadamente divergentes (*P. horikoshii* OT3 e *P. abyssi* GE5). **(D)** Na comparação dos genomas de duas espécies com um maior grau de divergência entre elas (como as bactérias *S. gordonii* Challis CH1 e *S. pneumoniae* R6), os padrões de sintenia são praticamente eliminados, o que é evidenciado pela ausência de linhas diagonais definidas.

Entretanto, mesmo com a pouca conservação da sintenia global e da estrutura de óperons individuais, alguns padrões organizacionais de ordem superior emergem quando são feitas análises comparativas mais abrangentes entre genomas de espécies procarióticas. Esses padrões são complexos e ficam mais evidentes quando são analisados óperons de genes que codificam proteínas funcionalmente relacionadas, como é o caso dos óperons

que codificam proteínas ribossômicas, os quais incluem, em seu conjunto, mais de 50 genes, presentes em todos os genomas de bactérias e de arqueas até agora sequenciados. Em cada espécie, o conjunto de óperons que engloba esses mais de 50 genes configura o que se convencionou chamar de um **superoperon**, cuja presença ubíqua em procariotos é explicável pela pressão seletiva favorável à manutenção dos genes e à expressão de todas as pro-

teínas por eles codificadas, em quantidades equilibradas entre si e com os rRNAs para a montagem adequada dos ribossomos.

A partir de estudos comparativos da organização dos superóperons de proteínas ribossômicas e de outros grupos menores de óperons parcialmente conservados em procariotos, emergiu o conceito mais amplo de **ultraoperon** (*überoperon*), que descreve um grupo de genes que aparecem nos genomas de diferentes espécies associados uns aos outros em um conjunto de óperons, mesmo que o número e a organização (em termos de combinação e ordem de genes) de óperons individuais variem de espécie para espécie. O ultraoperon é uma entidade estrutural e funcional supragenômica, porque sua existência e organização só podem ser evidenciadas a partir da comparação dos genomas de várias espécies. Em cada espécie, a organização dos óperons vinculados a um ultraoperon representa apenas uma das alternativas funcionais possíveis para a distribuição dos genes envolvidos.

No contexto de um ultraoperon, há pressões seletivas para que os genes permaneçam, ao longo da evolução, estrutural e funcionalmente associados uns aos outros em **vizinhanças gênicas conectadas** (*connected gene neighbourhoods*), expressão às vezes utilizada como sinônimo de ultraoperon. Uma pressão seletiva favorável à associação estrutural leva à inclusão e à manutenção desses genes em óperons parcialmente conservados, os quais têm sempre alguns genes em comum com os óperons relacionados (ortólogos) de outras espécies. Uma pressão seletiva favorável à associação funcional determina que padrões de regulação compartilhados ou coordenados entre si sejam adquiridos e mantidos por todos os óperons envolvidos. Em conjunto, essas forças evolutivas determinam que, em cada espécie, independentemente do número de óperons e do número e da ordem dos genes em cada um deles, todos os produtos gênicos de um ultraoperon sejam produzidos nas proporções adequadas às demandas quantitativas e funcionais da célula.

À luz do conceito de ultraoperon, o superoperon de proteínas ribossômicas, discutido anteriormente, ganha outra dimensão, pois esse conjunto de óperons relacionados inclui, além dos mais de 50 genes de proteínas ribossômicas, mais outros 20 a 30 genes, dependendo da espécie procariótica considerada. Dentre esses genes adicionais, estão alguns também relacionados à síntese proteica, como os que codificam alguns fatores de tradução. A presença de genes de fatores de tradução no contexto do ultraoperon é explicável pela correlação funcional de seus produtos com as proteínas ribossômicas no contexto da síntese proteica. Entretanto, vários outros genes sem correlação mais direta com a síntese proteica também estão presentes, como os que codificam proteínas associadas à transcrição (subunidades da RNA-polimerase e fatores de terminação ou anti-terminação transcricional) e os que codificam enzimas como a lactato-desidrogenase (associada à conversão de energia), a enolase (associada ao metabolismo central)

e a mevalonato-quinase (associada ao metabolismo de lipídeos). A presença desses e de outros genes no ultraoperon não é explicável por associações funcionais óbvias, mas pode ser explicada pela necessidade comum de um elevado nível de expressão, proporcionado pela correção transcricional com os genes que codificam proteínas ribossômicas.

Além do ultraoperon de proteínas ribossômicas, muitos outros exemplos de ultraóperons já foram evidenciados por estudos *in silico*, tanto em bactérias como em arqueas (**ver Material Complementar Online 4.3**). Esses estudos, mesmo que limitados à análise de uma fração do número total de espécies com genomas completamente sequenciados, já demonstram a prevalência da organização de genes em ultraoperons em procariotos. Além disso, os ultraoperons identificados constituem a base a partir da qual serão delineadas abordagens experimentais para a caracterização de muitas interações funcionais entre os produtos dos genes envolvidos, o que deverá levar à elucidação de muitos processos celulares procarióticos hoje ainda parcial ou totalmente desconhecidos.

3. Dinâmica evolutiva dos genomas procarióticos

Organismos procarióticos apresentam diversos mecanismos moleculares que proporcionam variabilidade genética, são altamente proliferativos e estão em geral expostos a condições ambientais bastante variáveis. Por isso, procariotos evoluem de forma rápida e, na dinâmica evolutiva de espécies de bactérias e de arqueas, a possibilidade do genoma ter o seu conteúdo gênico e a sua estrutura frequentemente alterados de maneira a gerar variantes genotípicas e fenotípicas com potencial adaptativo é uma característica fundamental. Essa característica, chamada de **plasticidade genômica**, resulta da grande suscetibilidade dos genomas procarióticos à ação de diferentes processos evolutivos, alguns dos quais serão discutidos a seguir.

3.1 Principais processos evolutivos

Na dinâmica evolutiva de genomas procarióticos (**Figura 4.16**), vários processos distintos podem ser reconhecidos. Esses processos podem ser divididos, para fins didáticos, em dois grupos principais, um deles incluindo processos que introduzem inovações nos genomas, em termos de aquisição de novas sequências ou de reorganização estrutural de sequências preexistentes, e o outro incluindo processos que levam à perda de sequências. Esses dois grupos de processos evolutivos serão tratados separadamente nas duas próximas seções.

3.1.1 Processos inovadores

São incluídos no grupo de processos evolutivos inovadores aqueles que levam ao surgimento de novas se-

quências ou de novas combinações de sequências preexistentes, aumentando a complexidade de um genoma. A aquisição de novas sequências por parte de genomas procarióticos decorre de três processos: a duplicação seguida de divergência, a transferência gênica horizontal e a fusão de replicons. Na **duplicação seguida de divergência**, uma segunda cópia de uma região genômica é gerada, e os pares de genes duplicados (inicialmente cópias um do outro) podem então divergir pelo acúmulo diferencial de mutações ao longo do tempo. Isso, além de aumentar o número de genes (e também de outras sequências), determina que mesmo mutações funcionalmente desfavoráveis possam ocorrer em uma das cópias de um gene sem maiores prejuízos para a adaptabilidade do organismo, desde que a outra cópia se mantenha funcional. Como discutido na Seção 1.3, a duplicação seguida de divergência pode levar à geração de genes parálogos com funções idênticas (em casos de divergência mínima) ou similares (em casos de divergência moderada), à geração de genes relacionados, mas com funções distintas (em casos de divergência mais acentuada), ou até a parálogos não funcionais (pseudogenes).

Na **transferência gênica horizontal**, novas sequências são incorporadas ao genoma de uma espécie

por processos mediados por elementos genéticos móveis, como plasmídeos conjugativos, bacteriófagos e elementos transponíveis (ver Capítulo 9). Essa transferência é chamada de “horizontal” porque pode ocorrer entre espécies sem uma relação linear (“vertical”) de descendência entre elas. Elementos genéticos móveis são encontrados praticamente em todas as espécies de bactérias e arqueas, constituindo uma fração significativa e muito dinâmica do repertório genético procariótico, chamada de **mobiloma**. O mobiloma tem componentes celulares, tanto cromossômicos (epissomos, incluindo elementos transponíveis, plasmídeos integrativos e bacteriófagos lisogênicos), como extracromossômicos (plasmídeos e bacteriófagos não integrativos), e também componentes extracelulares (vírions de bacteriófagos). A natureza dinâmica do mobiloma e a alta prevalência dos eventos de transferência gênica horizontal são hoje reconhecidas como aspectos fundamentais da evolução de procariotos e, à luz desse reconhecimento, o modelo inicialmente proposto para a evolução dos seres vivos em três domínios (Bactéria, Archaea e Eukarya) vem sendo revisto. Conforme modelos evolutivos mais recentes, eventos de transferência gênica horizontal teriam ocorrido com bastante frequência entre as células primitivas que compunham

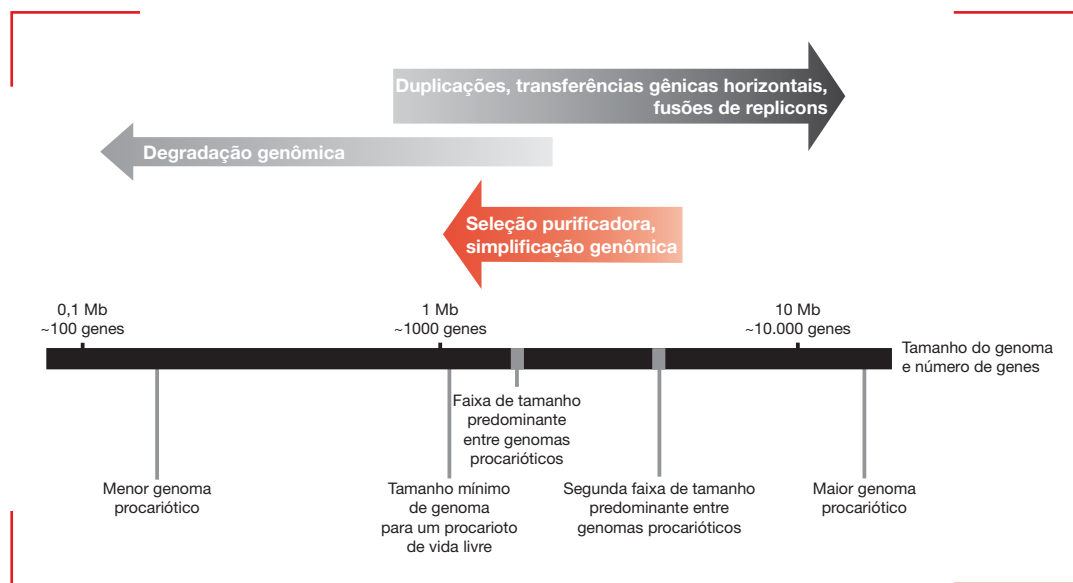


Figura 4.16

Esquema representativo das principais forças atuantes na evolução de genomas procarióticos. A barra de escala representa (em escala logarítmica) os tamanhos dos genomas procarióticos e o número aproximado correspondente de genes. Estão também indicados na barra, para referência, o tamanho do menor genoma procariótico conhecido (0,16 Mb, de *C. rudii*), o limite mínimo de tamanho conhecido para uma espécie procariótica de vida livre (1,3 Mb, de *Pelagibacter ubique* SAR11), a faixa de tamanho mais frequente para genomas procarióticos (~2 Mb), a segunda faixa de tamanho mais frequente para genomas procarióticos (~5 Mb) e o tamanho do maior genoma procariótico conhecido (13 Mb, de *S. cellulosum*). As principais forças evolutivas atuantes estão indicadas por setas que apontam para a esquerda, quando o efeito é de diminuição de tamanho dos genomas, ou para a direita, quando o efeito é de aumento de tamanho dos genomas; cada seta está posicionada sobre a faixa de tamanho de genomas na barra de escala na qual os efeitos das forças evolutivas por ela representadas são mais pronunciados.

a comunidade ancestral comum aos domínios Bacteria, Archaea e Eukarya e também ao longo de todo o período evolutivo subsequente, de modo especial entre os ramos correspondentes a linhagens bacterianas e de arqueas. Os eventos de transferência gênica horizontal ocorridos envolveram também genes cromossômicos, eventualmente associados a elementos genéticos móveis (em processos de conjugação, transdução ou transposição, p. ex.) e, com isso, é provável que a transferência gênica horizontal tenha contribuído para o aumento do repertório gênico de muitas espécies procarióticas, da mesma forma que a duplicação gênica.

A **fusão de replicons** é a união covalente e estável a longo prazo (em termos evolutivos) de dois replicons originalmente independentes, sendo também um processo importante para a aquisição de novas sequências por parte de genomas procarióticos. É proposto, por exemplo, que a fusão de replicons primordiais de tipo plasmidial originou os primeiros cromossomos procarióticos. Além disso, nos genomas de grande parte das espécies procarióticas atuais, é fácil identificar sequências originalmente plasmidiais ou virais que agora são componentes cromossômicos estáveis, em uma situação distinta daquela que ocorre em eventos de integração transitórios de elementos genéticos móveis. Há evidências também de fusões de cromossomos (em casos de espécies com ancestrais portadores de mais de um cromossomo), que seriam eventos bem mais raros do que fusões entre cromossomos e elementos do mobiloma.

A **reorganização de sequências** é um processo que introduz inovações na configuração estrutural de genomas procarióticos, embora, ao contrário da duplicação, da transferência gênica horizontal e da fusão de replicons, não leve à aquisição de novas sequências. Rearranjos cromossômicos que levam a reorganizações estruturais ocorrem com mais frequência em genomas relativamente mais ricos em sequências repetidas, cuja presença aumenta a probabilidade de ocorrerem, por exemplo, eventos de recombinação. A reorganização de sequências não muda o repertório de genes de um genoma, pois não implica em aquisição de novos genes ou perda de genes preexistentes. Mesmo assim, seu impacto evolutivo é marcante, já que ela proporciona a ocorrência de novas interações funcionais entre os genes, que, ao serem deslocados de uma posição para outra no cromossomo, podem ser integrados a novos sistemas de regulação da expressão. A reorganização de sequências é de particular importância quando envolve óperons, que podem ser criados ou desfeitos, ou podem ter o número e a ordem dos genes que os integram alterados. É provável que essa **reorganização de óperons** (*operon shuffling*) tenha sido um dos principais determinantes evolutivos para o surgimento e a manutenção de estruturas organizacionais e funcionais mais complexas em genomas procarióticos, como no caso dos ultraóperons (ver Seção 2.6.4).

3.1.2 Processos que determinam perda de sequências

Em sentido oposto ao de processos evolutivos inovadores, atuam processos determinantes da perda de sequências genômicas. Pode-se dizer que estes últimos levam a uma redução geral da complexidade dos genomas, que pode se dar pela diminuição de tamanho (contração genômica), pela perda de arranjos estruturais e funcionais e/ou pela eliminação de variabilidade genética. Em procariotos, três processos evolutivos se destacam como determinantes de redução de complexidade genômica: a seleção purificadora, a simplificação genômica e a degradação genômica.

A **seleção purificadora** é um tipo de seleção natural determinante de redução da variabilidade genética, pois atua eliminando de uma população variantes genéticas não adaptativas. Apesar de a maioria das mutações ser deletéria, o efeito de cada mutação individual em termos de redução da adaptabilidade de um indivíduo é muito pequeno e, por isso, os efeitos da seleção purificadora podem ser relativamente sutis. Na evolução de procariotos, a seleção purificadora parece ter sido mais preponderante em casos de diminuição de complexidade genética associados à adaptação de espécies de vida livre a ambientes mais estáveis. Nesses casos, a seleção purificadora, atuando juntamente com forças evolutivas determinantes de contração genômica, resultou em espécies com genomas menores e com pouca variação entre indivíduos de uma mesma população (o que é uma condição adaptativa para ambientes pouco variáveis).

A **simplificação genômica** (*genome streamlining*) é um processo de seleção negativo contra sequências supérfluas presentes em um genoma. A simplificação decorre de uma tendência natural (mas não muito acentuada) à perda de sequências nucleotídicas não funcionais, pois a replicação dessas sequências representa uma sobrecarga metabólica para a célula sem qualquer contrapartida adaptativa. Desse modo, há, em genomas de procariotos, uma tendência natural à perda, por exemplo, de sequências intergênicas (sem função codificadora, reguladora ou estrutural) e sequências repetidas (incluindo genes duplicados cuja presença em mais de uma cópia não oferece vantagem para a célula).

O processo de **degradação genômica** , por sua vez, engloba vários componentes que, em conjunto, determinam contração genômica. Dentre as forças evolutivas componentes de processos de degradação genômica de procariotos, podem ser destacadas a própria simplificação genômica, discutida anteriormente, a deriva genética (de modo particular em casos de populações de pequeno tamanho), baixas frequências de recombinação, altas taxas de mutação, capacidade reduzida para a aquisição de novas sequências (por transferência gênica horizontal, p. ex.) e uma predominância de eventos de deleção de sequências em detrimento de eventos de inserção. Quantos ou quais desses componentes atuam em um processo de degradação genômica ao longo de uma linhagem evolu-

tiva é algo que varia de caso para caso, pois depende da espécie considerada, de seu modo de vida e das condições ambientais vigentes em cada etapa de sua evolução.

3.2 Principais tendências evolutivas

Quando é feita uma análise genômica comparativa abrangente, incluindo um grande número de espécies procarionóticas com diferentes graus de parentesco filogenético entre elas, algumas tendências evolutivas gerais podem ser identificadas. A tendência geral mais evidente é a de aumento de **entropia** (desorganização), associada a eventos frequentes de perda, aquisição ou rearranjo de sequências, que levam à perda parcial de organização estrutural e funcional (**Figura 4.17**). Ao aumento da entropia se contrapõem, geralmente em menor proporção, as forças seletivas determinantes de arquiteturas genômicas mais estruturadas e funcionais, os eventos de transferência horizontal de grupos de genes organizados (óperons) e as restrições impostas pelo sistema de replicação e pela estruturação em domínios do cromossomo. O relativo predomínio das forças entrópicas determina que a configuração presente de qualquer genoma, apesar de funcional e adaptativa para as condições vigentes, está sempre aquém da ideal, o que abre espaço para o constante surgimento de novas configurações.

Genomas procarionóticos também podem apresentar tendências de **contração** ou **expansão**, sendo a dinâmica de variação de tamanho determinada por desequilíbrios entre as forças evolutivas atuantes no sentido da aquisição de sequências e aquelas que atuam no sentido oposto (ver Figura 4.16). A ação contrária e em diferentes proporções dessas forças explica os dois maiores picos de distribuição de tamanho de genomas procarionóticos, um correspondente a espécies com genomas de aproximadamente 2 Mb e outro correspondente a espécies com genomas de aproximadamente 5 Mb (ver Seção 2.1). A maior frequência de genomas com tamanho em torno de 2 Mb, especialmente vista em arqueas, mas também observada em bactérias, seria resultado da ação predominante de forças que determinaram a contração de genomas ancestrais maiores. Já o pico secundário de frequência, correspondente a genomas com tamanho próximo a 5 Mb, seria resultado da ação predominante de forças que levaram genomas ancestrais menores a uma expansão.

No **Material Complementar Online 4.4** são discutidos alguns aspectos adicionais da dinâmica evolutiva de genomas procarionóticos. Nele são apresentados alguns casos de maior ou menor estabilidade estrutural e de contração ou expansão em genomas de procarionotos, considerando os principais fatores evolutivos determinantes de cada situação.

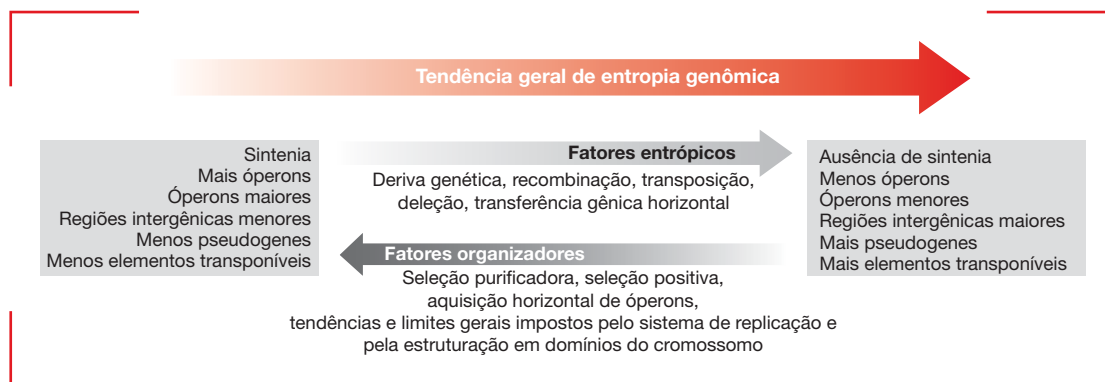


Figura 4.17

Principais fatores determinantes do grau de organização de genomas procarionóticos. A tendência geral à entropia (desorganização) observada em genomas procarionóticos é resultante de uma relativa predominância dos fatores entrópicos sobre os fatores organizadores. As principais consequências do aumento ou da diminuição da entropia em um genoma procarionótico estão listadas nos quadros à direita e à esquerda, respectivamente.



Resumo

Procariotos (bactérias e arqueas) compartilham algumas características celulares importantes, como a ausência de organelas e a estrutura genômica relativamente simples (em comparação à de eucariotos). Os genes procarióticos são compostos por uma região codificadora (de um produto de RNA ou proteína) e pelas regiões reguladoras responsáveis por sua expressão. A orientação de cada gene é definida pelo sentido da sua transcrição. Os genomas procarióticos em geral se apresentam na forma de um cromossomo circular único, embora haja espécies com mais de um cromossomo ou com cromossomos lineares. O tamanho dos genomas pode variar de menos de 0,16 Mb até mais de 13 Mb, mas a maioria dos procariotos tem genomas de 2 a 5 Mb. Os genomas procarióticos são compactos, com uma densidade de quase 1 gene/kb, regiões intergênicas curtas e poucas sequências repetidas. Cada genoma possui diferentes unidades organizacionais, que vão desde repetições de sequência curtas e genes até ilhas genômicas e domínios cromossômicos. Além disso, os genes podem ter a sua ordem no cromossomo

conservada entre espécies aparentadas (sintenia), e alguns grupos de genes podem ter coorientação e coregulação definidas em óperons, ou em arranjos maiores estruturados de acordo com a replicação e com os domínios cromossômicos. Em termos evolutivos, os genomas procarióticos apresentam uma tendência à instabilidade e à desorganização. São frequentes os eventos de perda, aquisição e rearranjo de sequências, muitos dos quais envolvem elementos genéticos móveis e transferência gênica horizontal. Em casos de populações pequenas, mesmo modificações genéticas não adaptativas podem ser fixadas por deriva genética. Padrões organizados são estabelecidos e mantidos por seleção natural, pela aquisição lateral de grupos ordenados de genes e pelas tendências impostas pela estrutura e pelo funcionamento de domínios e replicons. Dependendo do equilíbrio entre as diferentes forças evolutivas, um cromossomo procariótico pode manter-se razoavelmente estável ou pode tender à contração ou à expansão.



Leituras recomendadas

Bentley SD, Parkhill J. Comparative genomic structure of prokaryotes. *Annu Rev Genet.* 2004;38:771-92.

Center for Biological Sequence Analysis. Genome Atlas Database [Internet]. [capturado 23 ago. 2011]. Disponível em: <http://www.cbs.dtu.dk/services/GenomeAtlas/>.

Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP. The bacterial species challenge: making sense of genetic and ecological diversity. *Science.* 2009;323(5915):741-6.

Koonin EV, Wolf YI. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 2008;36(21):6688-719.

Koonin EV, Wolf YI. The fundamental units, processes and patterns of evolution, and the tree of life conundrum. *Biol Direct.* 2009;4:33.

Koonin EV. Evolution of genome architecture. *Int J Biochem Cell Biol.* 2009;41(2):298-306.

Moya A, Gil R, Latorre A, Peretó J, Pilar Garcillán-Barcia M, de la Cruz F. Toward minimal bacterial cells: evolution vs. design. *FEMS Microbiol Rev.* 2009;33(1):225-35.

National Center for Biotechnology Information. Entrez Genome Project: complete microbial genomes [Internet]. [capturado 23 ago. 2011]. Disponível em: <http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>.

Rocha EP. The organization of the bacterial genome. *Annu Rev Genet.* 2008;42:211-33.

Rocha EP. Order and disorder in bacterial genomes. *Curr Opin Microbiol.* 2004;7(5):519-27.

Henrique Bunselmeyer Ferreira

Genes e Genomas Eucarióticos

1. Genes eucarióticos 86

- 1.1 Estrutura básica e ocorrência de genes interrompidos em eucariotos 86
- 1.2 Tamanho 88
- 1.3 Domínios transcricionais complexos 88
- 1.4 Famílias multigênicas eucarióticas 89

2. Genomas nucleares eucarióticos 90

- 2.1 Tamanho 90
- 2.2 Forma, número e organização em replicons de cromossomos eucarióticos 92
- 2.3 Composição de sequências 92
 - 2.3.1 Sequências gênicas 92
 - 2.3.1.1 Número de genes e densidade gênica 94
 - 2.3.1.2 Famílias de parálogos, genes únicos e pseudogenes 95
 - 2.3.2 Sequências intergênicas 99
 - 2.3.2.1 Sequências repetidas simples 99
 - 2.3.2.2 Elementos transponíveis 101

- 2.4 Sintenia 103

- 2.5 Agrupamentos organizados de genes 104

- 2.5.1 Óperons eucarióticos 104

- 2.5.2 Agrupamentos de genes funcionalmente relacionados 104

- 2.5.3 Agrupamentos de genes coexpressos 105

3. Dinâmica evolutiva de genes e genomas nucleares eucarióticos 105

- 3.1 Evolução dos genomas eucarióticos 105
- 3.2 Evolução dos genes interrompidos 106

4. Genomas de organelas 107

- 4.1 Genomas mitocondriais 107
- 4.2 Genomas de plastídeos 109

O domínio **Eukarya** abrange uma ampla variedade de organismos uni ou pluricelulares, cuja taxonomia passou por grandes reformulações nos últimos anos. A divisão clássica dos eucariotos nos reinos Protista, Metazoa (ou Animalia), Fungi e Plantae deu lugar a uma divisão em seis **supergrupos**, com base em análises filogenéticas multigênicas e características morfológicas e bioquímicas (**ver Material Complementar Online 5.1**). Dos seis supergrupos propostos, quatro (**Amoebozoa**, **Chromalveolata**, **Excavata** e **Rhizaria**) são essencialmente microbianos, sendo formados por diferentes grupos de protozoários, alguns grupos de algas uni ou pluricelulares (mas sem tecidos verdadeiros) e micetozoários. Animais (metazoários) e fungos típicos (Dikarya), juntamente com alguns grupos de eucariotos unicelulares, como os dos microsporídeos e o dos coanoflagelados, formam o supergrupo **Opisthokonta**. Os vegetais terrestres fazem parte do supergrupo **Plantae**, assim como, as algas verdes, vermelhas e glaucófitas.

A classificação dos eucariotos em seis supergrupos ainda não é definitiva, pois alguns dos clados e relações filogenéticas propostos carecem de um maior embasamento, uma vez que foram estabelecidos a partir da análise de números ainda relativamente pequenos de espécies e/ou sequências nucleotídicas. Apesar das incertezas filogenéticas e taxonômicas ainda existentes, a característica mais evidente e distintiva de qualquer célula eucariótica é a presença de um **núcleo**, que é o compartimento celular que abriga a maior parte do DNA celular (o **genoma nuclear**). As células eucarióticas também são caracterizadas pela presença de outras **organelas** além do núcleo, como as mitocôndrias, os retículos endoplasmáticos liso e rugoso, o complexo de Golgi, os plastídeos (típicos de células vegetais) e os lisossomos (típicos de células animais) (ver Capítulo 1). As mitocôndrias e os plastídeos possuem pequenos genomas especializados, que codificam as funções específicas dessas organelas.

O número disponível de sequências completas de genomas eucarióticos é considerado pequeno, se comparado com o número de genomas sequenciados de procariotos e com a enorme biodiversidade do domínio Eukarya. Em setembro de 2011, já haviam sido disponibilizadas ao público as sequências completas dos genomas de 37 espécies eucarióticas, estando ainda em andamento os projetos de sequenciamento dos genomas de outras 1.178 espécies. Essa amostragem restrita limita o alcance da genômica comparativa, como ferramenta para o estudo da organização e da evolução de genomas eucarióticos, de modo que muitas das generalizações feitas, ainda, se baseiam na análise dos genomas de algumas poucas espécies-modelo, como a levedura *Saccharomyces cerevisiae*, a mosca-da-fruta *Drosophila melanogaster*, o nematódeo *Caenorhabditis elegans* e o camundongo *Mus musculus*.

Estudos genômicos abrangentes apresentam dificuldades pelo grande tamanho e a elevada complexidade dos genomas eucarióticos. Por isso, o foco das investigações é muitas vezes dirigido apenas à fração expressa de cada

genoma pela utilização de metodologias para a análise em massa de produtos de transcrição (principalmente mRNA) ou tradução (ver Capítulo 16).

O primeiro genoma eucariótico a ser completamente sequenciado foi o de *S. cerevisiae*, publicado em 1996. Em 1998, foi publicada a primeira sequência genômica completa de um metazoário (*C. elegans*) e a ela se seguiram as sequências dos genomas de *D. melanogaster*, em 2000, e do camundongo, em 2002. Uma primeira versão (~90% concluída) do genoma humano foi publicada em 2001, mas a versão completa foi publicada apenas em 2006. O primeiro genoma de um vegetal a ser completamente sequenciado foi o da crucífera *Arabidopsis thaliana*, em 2000, e o de um fungo verdadeiro o de *Neurospora crassa*, em 2003. Informações sobre projetos de sequenciamento de genomas eucarióticos concluídos ou em andamento e acesso aos dados de sequência podem ser obtidos por meio de páginas institucionais especializadas na internet, referenciadas na Seção de Leituras Recomendadas ou no **Material Complementar Online 5.5**.

A seguir, serão analisados os aspectos estruturais típicos de genes e genomas eucarióticos. A princípio, serão discutidos os genes interrompidos e o caráter complexo das unidades transcricionais eucarióticas. Depois, será estudada a forma como genes e outras sequências estão distribuídos nos genomas de diferentes organismos eucarióticos e a hipótese sobre como ocorreu a evolução de genes e genomas nucleares. Por último, será vista a origem evolutiva das organelas e de seus genomas, descrevendo brevemente a estrutura de genomas mitocondriais e de cloroplastos.

1. Genes eucarióticos

O conceito básico de gene e as descrições de seus componentes essenciais de sequência, apresentados no Capítulo 4 no contexto de genomas procarióticos, também são válidos para genomas eucarióticos. Genes procarióticos e eucarióticos compartilham a mesma estrutura básica, sendo formados por uma região codificadora e pelas regiões reguladoras associadas a ela. Entretanto, as sequências reguladoras ou codificadoras de genes de eucariotos, em geral, possuem estruturas mais complexas que as de genes procarióticos, sendo que essa maior complexidade determina refinamentos funcionais que explicam, pelo menos em parte, o maior grau de sofisticação de células e organismos eucarióticos.

1.1 Estrutura básica e ocorrência de genes interrompidos em eucariotos

Como pode ser visto na **Figura 5.1**, as regiões codificadoras e reguladoras de genes eucarióticos podem apresentar estruturas bem mais complexas que as de genes procarióticos. Por exemplo, em genes eucarióticos típicos,

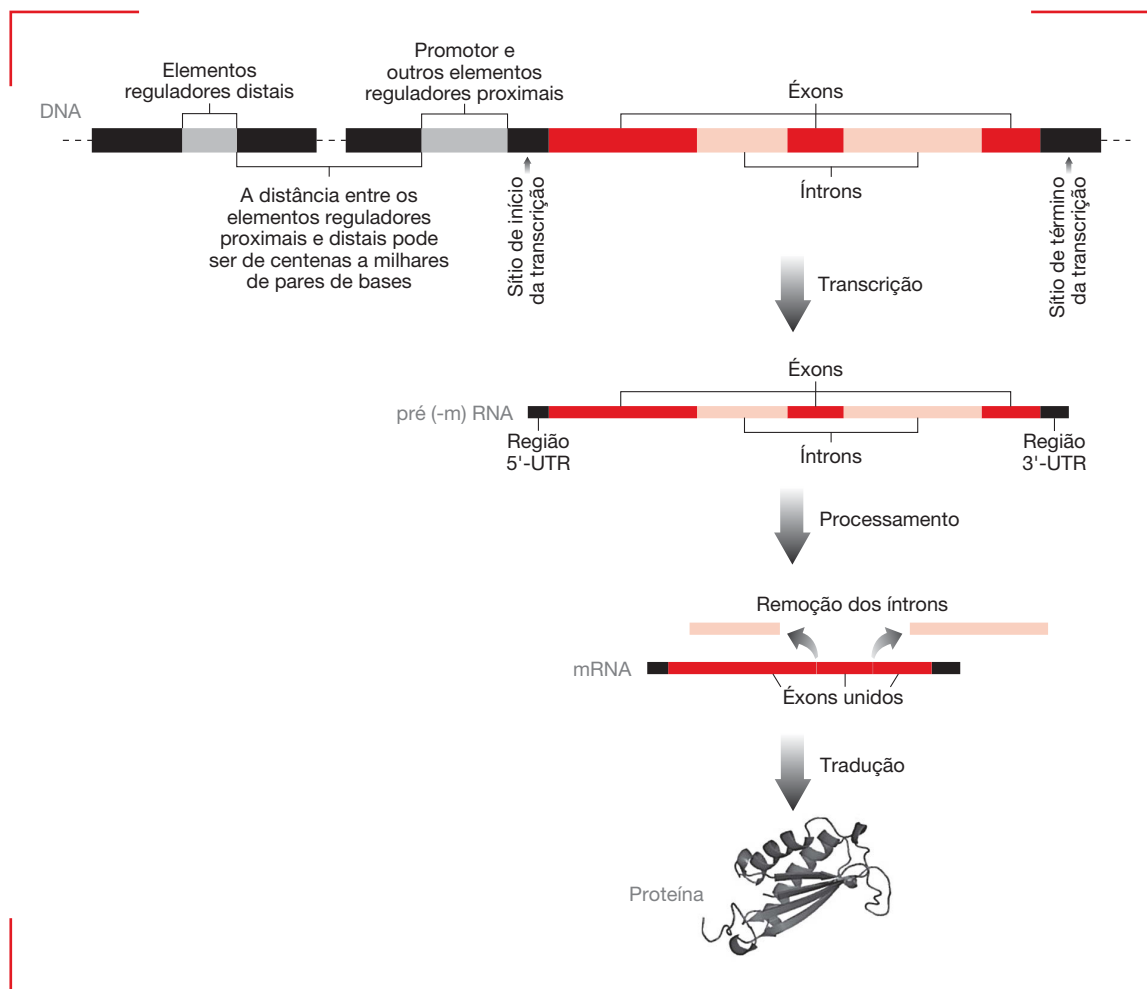


Figura 5.1

Esquema representativo da estrutura e da expressão de um gene eucariótico que codifica uma proteína.

A figura representa um gene interrompido (com íntrons); a presença de íntrons em genes eucarióticos, apesar de típica (principalmente em eucariotos superiores), não é obrigatória. Como explicado na Figura 4.3 (ver Capítulo 4), o gene está representado de 5' para 3', considerando a sua fita codificadora, o mesmo acontecendo com o RNA transcrito a partir dele. A região reguladora é constituída por um promotor e outros elementos reguladores proximais, situados logo a montante do sítio de início da transcrição, e por elementos reguladores distais, que podem estar situados a muitos milhares de pares de bases a montante ou a jusante do promotor e da região codificadora. A região codificadora é dividida em segmentos (éxons), que são separados por sequências intervenientes (íntrons); nesse caso, o gene possui três éxons (em ■) e dois íntrons (em ■). O gene é transcrito em um pré-mRNA, que inclui as sequências dos íntrons. Durante o processamento do pré-mRNA, ocorre a remoção dos íntrons e a união dos éxons (no processo de *splicing*). O mRNA maduro gerado é, então, exportado do núcleo para o citoplasma da célula, onde pode ser traduzido em uma proteína.

a região reguladora a montante da região codificadora, incluindo o promotor (sítio de ligação da RNA-polimerase), é, em geral, bem mais extensa e contém um número de elementos reguladores (sequências nucleotídicas reconhecidas por proteínas reguladoras específicas) muito maior do que o observado em genes procarióticos. Além disso, um gene eucariótico pode conter elementos reguladores distais situados a milhares de pares de bases da sua região codificadora, tanto a montante como a jusante, o que contrasta com os elementos reguladores distais de

genes procarióticos, em geral, situados a algumas centenas de pares de bases a montante da região codificadora.

Outra característica marcante, embora não exclusiva, de genes eucarióticos é a presença de sequências intervenientes. Essas sequências, denominadas **íntrons**, interrompem a região transcrita do gene e a dividem em vários segmentos, denominados **éxons** (ver Figura 5.1). As sequências correspondentes aos íntrons estão presentes no DNA genômico, são transcritas em um RNA precursor (pré-RNA), porém removidas durante o processamento

do RNA, que dá origem ao RNA maduro. Em genes codificadores de proteínas, os íntrons em geral interrompem a região codificadora. Além disso, também podem ocorrer íntrons que interrompem regiões flanqueadoras (5'-UTR e 3'-UTR), que são transcritas, mas não traduzidas (UTR – *untranslated region* = região não traduzida).

Os íntrons de genes eucarióticos diferem dos de genes procarióticos, quanto a estrutura, essas diferenças têm reflexos funcionais importantes. Em razão das sequências nucleotídicas presentes, íntrons eucarióticos, em geral, não adotam estruturas secundárias (por pareamento intracadeia) no contexto de transcritos de RNA. Os íntrons procarióticos, por sua vez, formam estruturas secundárias relativamente complexas, e isso é um determinante da maneira como eles são removidos durante o processamento do RNA. Por isso, os íntrons eucarióticos dependem de um complexo ribonucleoproteico (spliceossomo) para a sua remoção, ao passo que os íntrons procarióticos são removidos por meio de processos autocatalíticos (ver Capítulo 11).

A ocorrência de genes interrompidos é rara em procariotos (ver Capítulo 4), mas, entre eucariotos, ela é bastante comum. Há, contudo, uma ampla gama de variação na frequência de genes interrompidos, quando comparados aos genomas de diferentes espécies eucarióticas. Por exemplo, em algumas espécies de eucariotos unicelulares, como as dos gêneros *Giardia* e *Trichomonas* (Excavata, filo Metamonada), *Trypanosoma* (Excavata, filo Euglenozoa), *Cryptosporidia* (Chromalveolata, filo Apicomplexa) e *Entamoeba* (Amoebozoa), os íntrons estão ausentes, ou o percentual de genes interrompidos é muito baixo (<1%). Em outras espécies, também unicelulares, como *Plasmodium falciparum*, *Babesia bovis* e *Theileria parva* (todas Chromalveolata, filo Apicomplexa), a fração de genes interrompidos no genoma chega a 53,9, 73,6 e 61,5%, respectivamente. Uma variação ampla no percentual de genes com íntrons também ocorre em leveduras, indo de 5%, em *S. cerevisiae*, a 50%, em *Schizosaccharomyces pombe*. Em eucariotos mais complexos, desde fungos multicelulares até vertebrados, a grande maioria dos genes têm íntrons, e o percentual de genes interrompidos chega a 83%, em *D. melanogaster*, e a 94% ou mais, em mamíferos.

1.2 Tamanho

Para eucariotos, a delimitação precisa da extensão das regiões reguladoras dos genes não é simples, pois são muitos os elementos de sequência envolvidos na regulação da expressão de genes eucarióticos e, com frequência, esses elementos estão muito distantes do(s) gene(s) por eles regulado(s) (ver Capítulo 14). Por isso, para avaliações comparativas de tamanho entre genes eucarióticos, como as descritas a seguir, são consideradas apenas as regiões transcritas (éxons e íntrons) de cada gene.

O tamanho dos genes eucarióticos varia dentro de limites amplos mas, em grande parte dos casos, eles são maiores que o de genes procarióticos, que têm, em sua

maioria, em torno de 1 kb e dificilmente ultrapassam os 10 kb de extensão (ver Capítulo 4). O *S. cerevisiae*, um dos eucariotos mais simples (com apenas 5% de genes interrompidos por íntrons), possui genes com extensões que variam de 120 pb até mais de 14 kb, com um tamanho médio de 1,4 kb. Entretanto, é importante considerar que os genes interrompidos são predominantes entre eucariotos, e isso determina que muitos genes eucarióticos, sejam bem maiores do que as regiões codificadoras de seus produtos. Assim, em *D. melanogaster* ou em mamíferos, por exemplo, a maioria dos genes tem de 2 a 100 kb.

Sendo a maioria dos genes eucarióticos interrompidos, é imprescindível que, em avaliações gerais de tamanho, sejam considerados o número e o tamanho dos íntrons e dos éxons componentes desses genes. Em relação a íntrons, há, entre genes eucarióticos interrompidos, uma tendência geral de aumento no número de íntrons por gene e no tamanho dos íntrons individuais, à medida que se avança na escala evolutiva. Essas tendências determinam que o tamanho médio dos genes seja maior em espécies eucarióticas complexas. No homem, por exemplo, os genes possuem um tamanho médio de 27 kb, com 1,35 kb (5%) correspondendo a éxons e 25,65 kb (95%) a íntrons. A média do número de íntrons por gene no homem (9) é um pouco superior à dos vertebrados (5 a 8 íntrons por gene em geral). Excedendo as médias, há alguns casos, como o do gene humano que codifica a distrofina, uma proteína muscular com mais de 3.500 aminoácidos. Esse gene, o maior até agora conhecido, tem um total de 2.400 kb (2,4 Mb), dos quais 14 kb correspondem aos seus 79 éxons e o restante corresponde aos seus 78 íntrons. Nas regiões codificadoras, a mais longa conhecida é a do gene humano da titina (outra proteína muscular), com quase 81 kb, que tem também o maior número de éxons (178) e o maior íntron (17,1 kb) até agora conhecidos entre eucariotos.

1.3 Domínios transcricionais complexos

Análises recentes de genomas e transcritomas de muitas espécies eucarióticas revelaram que a fração transcrita de cada genoma excede em muito a fração correspondente a genes codificadores de mRNA/proteínas ou das outras duas classes majoritárias de RNA (rRNA e tRNA). Por exemplo, o *S. cerevisiae*, *C. elegans*, o camundongo e o homem têm, respectivamente, pelo menos 85, 70, 63 e 93% de seus genomas transcritos. Foi dado o nome de **transcrição disseminada** (de, *pervasive transcription*) a esse fenômeno, e o repertório de **RNAs não codificadores (ncRNAs, de, noncoding RNAs)** gerado inclui classes diferentes de transcritos (ver Capítulo 10), algumas das quais com função reguladora da expressão gênica (ver Capítulo 14) e outras ainda sem atribuições funcionais claras.

Independentemente das questões funcionais, os ncRNAs tornaram a definição dos genes eucarióticos ainda mais complexa. Enquanto alguns desses genes são transcritos a partir de regiões antes consideradas como

intergênicas, outros são transcritos a partir de regiões que se sobrepõem total ou parcialmente aos genes codificadores de proteínas, tanto em porções exônicas como intrônicas, tanto no mesmo sentido (de 5' para 3' em relação à fita codificadora) como no sentido inverso (antissenso). Assim, existem lócus cromossômicos, aqui referidos como **domínios transcricionais complexos**, a partir dos quais são gerados múltiplos tipos de transcritos independentes. Um domínio transcricional complexo pode incluir um ou mais genes de estrutura “tradicional”, como a descrita na Seção 1.1, e regiões

inter e intragênicas transcritas em diferentes ncRNAs (**Figura 5.2**).

1.4 Famílias multigênicas eucarióticas

Como descrito no Capítulo 4 (Seção 1.4), a presença de genes parálogos em um genoma é a evidência de processos evolutivos de duplicação gênica. Os eventos de duplicação gênica e a consequente presença de genes parálogos são características comuns a genomas procarionóticos e eucarióticos. Contudo, há ainda, em procario-

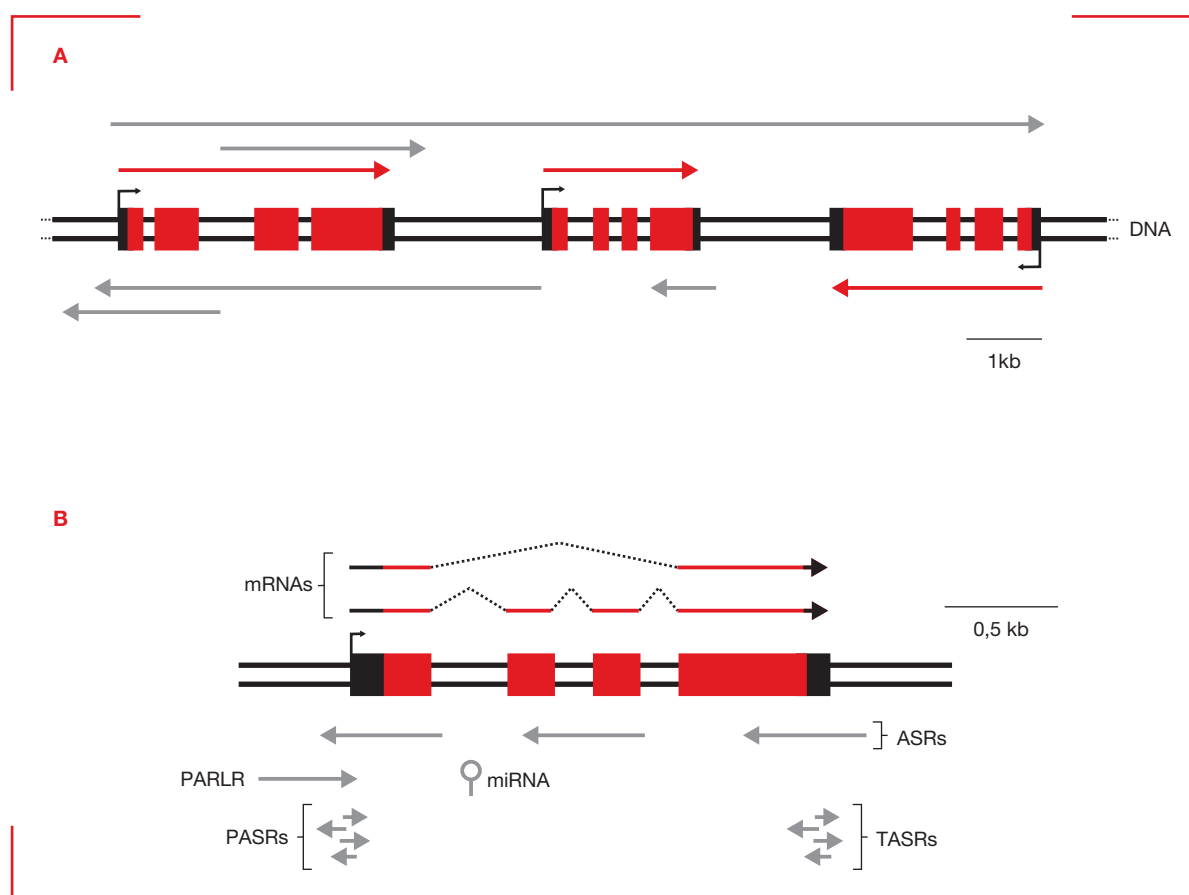


Figura 5.2

A transcrição disseminada e o repertório de RNAs produzido de um domínio transcricional complexo. (A) O lócus cromossômico eucariótico representado (DNA) inclui três genes de estrutura “tradicional”, cujos sítios de início (e o sentido, de 5' para 3') de transcrição e os éxons estão representados, respectivamente, por setas menores em preto e por retângulos (com as porções codificadoras em ■ e as regiões 5'-UTR e 3'-UTR em ■); os diferentes transcritos primários produzidos a partir desse lócus estão representados por setas (que indicam o sentido de transcrição) em ■ (pré-mRNAs) ou em ■ (outros RNAs); (B) Ampliação da região correspondente ao gene central do lócus representado em (A), mostrando algumas das várias classes de RNA produzidas a partir do processamento dos transcritos primários. Os diferentes transcritos codificadores de proteína (mRNAs maduros), produzidos por *splicing* alternativo, estão em ■, com as regiões 5'-UTR e 3'-UTR em preto e as regiões removidas durante o processamento (íntrons e éxons alternativos) indicadas por linhas pontilhadas. Os transcritos não codificadores (ncRNAs) estão em ■, com indicação da respectiva classificação (ASRs: RNAs antissenso, complementares a diferentes porções dos transcritos codificadores; PALR: RNA longo associado ao promotor; PASRs: RNAs curtos associados ao promotor; TASRs: RNAs curtos associados ao terminador; miRNA: micro-RNA produzido a partir de íntron e formando estrutura secundária). Cada RNA indicado na figura encontra-se alinhado horizontalmente com a região genômica a partir da qual ele foi gerado.

tos, forças evolutivas que atuam no sentido oposto e tendem a limitar uma maior expansão de famílias de genes parálogos. Já em genomas eucarióticos, principalmente naqueles maiores, de espécies mais complexas, essas restrições são menores e, assim, ao longo da evolução, foram constituídas algumas **famílias multigênicas**, típicas de eucariotos. Cada uma dessas famílias possui vários membros, cada um deles codificando uma variante de um determinado tipo de proteína. As proteínas codificadas pelos genes de uma família multigênica são similares, quanto a sua estrutura e funcionalidade, mas as diferenças existentes entre elas determinam certo grau de especialização funcional. Além disso, é comum a diversificação dos padrões de expressões espaciais e/ ou temporais em cada gene da família, o que pode definir que certas proteínas sejam expressas apenas em um determinado tipo celular, de estágio do desenvolvimento, ou de tecido.

Em cada família multigênica, o número de genes que a compõe e a distribuição desses genes ao longo dos cromossomos pode variar de espécie para espécie. Há, uma tendência de que espécies mais complexas na escala evolutiva (como os mamíferos) tenham, para uma determinada família, um número maior de membros do que aquele encontrado em espécies mesmo evoluídas (como os invertebrados). Isso provavelmente aconteça, em razão da ocorrência de sucessivos eventos de duplicação gênica dos genes daquela família ao longo de uma linhagem evolutiva (como a dos metazoários, p. ex.).

A origem das famílias multigênicas a partir de eventos de duplicação gênica também determina que os membros da família mais recentes se encontrem ainda próximos (agrupados) em uma mesma região cromossômica. Existe uma tendência, de que, ao longo do tempo, rearranjos cromossômicos (como inversões e translocações) levem à dispersão desses membros em um mesmo cromossomo ou em cromossomos diferentes. Entretanto, em oposição a essa tendência, forças seletivas podem atuar para a manutenção de agrupamentos gênicos, em razão da ocorrência de mecanismos de regulação coordenada (corregulação) da expressão dos genes (ver Seção 2.5.2).

Existem muitas famílias multigênicas típicas de eucariotos, embora nem todas elas sejam comuns a todas as espécies eucarióticas. Como exemplos de famílias multigênicas ubíquas em eucariotos, podem ser citadas as de genes codificadores de proteínas da cromatina, como as histonas, e as de citoesqueleto, como as actinas, as tropomiosinas e as tubulinas. Outras famílias são típicas de espécies eucarióticas mais complexas, como vertebrados ou plantas superiores. Dentre essas, estão as famílias multigênicas que codificam hemoglobinas, imunoglobulinas, interferons e fatores de transcrição homeóticos. No **Material Complementar Online 5.3**, são discutidos alguns aspectos adicionais sobre famílias multigênicas eucarióticas, utilizando como exemplos a família de genes codificadores de actinas, ubíqua em eucariotos, e a família de genes codificadores de globinas de vertebrados.

2. Genomas nucleares eucarióticos

2.1 Tamanho

Uma das características mais evidentes que diferencia genomas eucarióticos de genomas procarióticos é o tamanho, já que os de espécies eucarióticas em geral são maiores que os de espécies procarióticas. Para procariotos, como visto no Capítulo 4, a amostragem de genomas completamente sequenciados é grande o suficiente para estimativas abrangentes sobre as variações de tamanho existentes entre espécies de bactérias e de arqueas. Para eucariotos, contudo, o número de espécies com genomas completamente sequenciados é considerado pequeno e a amostragem disponível é pobre em representatividade para muitos grupos taxonômicos. Assim, as análises comparativas mais amplas para eucariotos se baseiam não apenas em genomas sequenciados, mas também em dados obtidos por meio de outras metodologias, como a densitometria ou a citometria de fluxo, que determinam o conteúdo total de DNA nuclear (**ver Material Complementar Online 5.4**).

A **Figura 5.3** mostra a ampla variação de tamanho de genomas que ocorre entre eucariotos, comparando o conteúdo de DNA haploide, chamado de **valor C**.

No outro extremo da distribuição, é interessante notar que os maiores genomas eucarióticos conhecidos também ocorrem entre espécies unicelulares, pelo menos com base em quantificações densitométricas ou citométricas do conteúdo de DNA nuclear. Para algumas amebas (Amoebozoa), por exemplo, as dos gêneros *Polychaos* e *Chaos*, foram feitas estimativas de conteúdo genômico da ordem de 600.000 Mb a 1.400.000 Mb. Entretanto, essas estimativas, que estendem muito a faixa de variação de tamanho dos genomas entre espécies protistas, devem ser consideradas com cautela. Muitas amebas são altamente poliploides, podendo conter 100 ou mais conjuntos cromossômicos completos, o que inviabiliza comparações diretas com estimativas de tamanho genômico haploide.

Entre animais (metazoários), o menor genoma conhecido é o do nematódeo parasita de plantas *Pratylenchus coffeae*, com tamanho próximo a 20 Mb, e o maior é o do peixe pulmonado *P. aethiopicus*, com 130.000 Mb. São notavelmente grandes, também, os genomas de alguns anfíbios, como o de *N. lewisi* (ordem Caudata, família Proteidae), com 120.000 Mb. O genoma humano, com pouco mais de 3.000 Mb, ocupa uma posição intermediária entre os genomas de mamíferos, cujos tamanhos oscilam entre 1.700 Mb (para o morcego *M. schreibersi*) e 8.200 Mb (para o roedor *Tympanoctomys barrerae*). Em qualquer filo considerado há uma ampla variação de tamanho, que, entre cordados (representados por peixes, anfíbios, répteis, aves e mamíferos), chega a mais de 300 vezes do menor ao maior genoma e, entre artrópodes (representados por insetos e crustáceos), chega a mais de 600 vezes. Apesar disso, considerando-se os

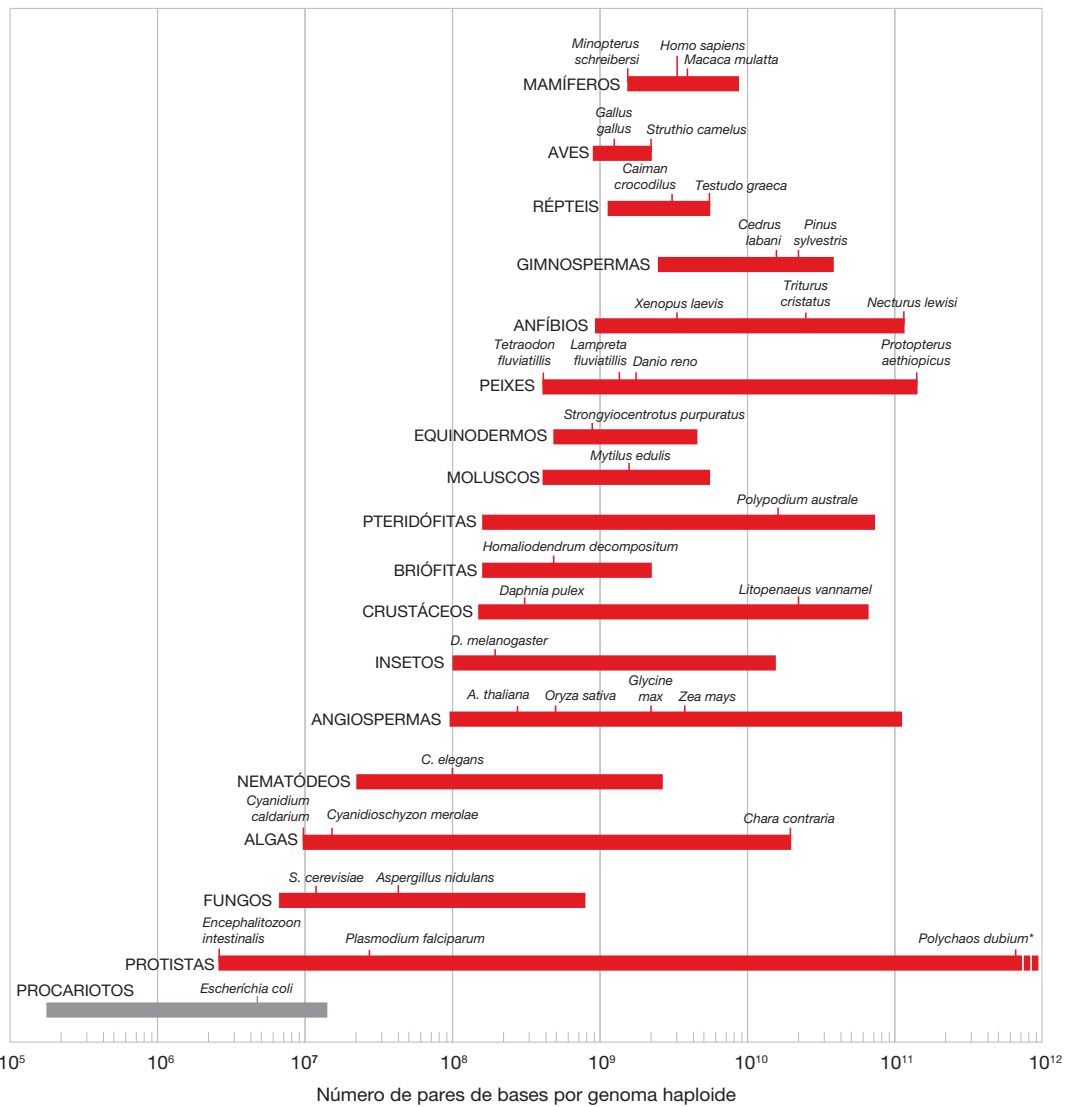


Figura 5.3

Tamanhos de genoma em diferentes grupos taxonômicos. Para cada grupo taxonômico indicado, a amplitude na variação de tamanho do genoma encontrada entre espécies a ele pertencentes está indicada por uma barra em ■. O tamanho do genoma de uma ou mais espécies representativas do grupo está posicionado na escala. Grupos taxonômicos de diferentes hierarquias estão representados e, por simplificação, espécies unicelulares pertencentes aos supergrupos Opisthokonta, Amoebozoa, Chromalveolata, Excavata e Rhizaria foram reunidas como protistas. Como algas, estão classificadas apenas as espécies pertencentes ao supergrupo Plantae. Os dados de procariotos (barra em ■) foram incluídos para referência. Todos os dados representados referem-se ao conteúdo de DNA do genoma haploide (valor C), com exceção de alguns protistas (*), cujo genoma apresenta alto grau de poliploidização.

tamanhos mínimos de genoma observados em cada filo, parece haver uma tendência ao aumento de tamanho que acompanha, até certo ponto, a crescente complexidade anatômica e funcional observada entre metazoários, desde nematódeos até cordados.

Entre vegetais, os menores genomas conhecidos são os de algumas algas unicelulares, como *C. caldarium*, com menos de 10 Mb, e os maiores, de algumas angios-

permas, chegam a mais de 100.000 Mb. De algas do supergrupo Plantae até plantas superiores (gimnospermas e angiospermas), passando por briófitas e pteridófitas, também parece haver uma certa tendência ao aumento no tamanho mínimo ou médio de genoma para cada grupo taxonômico, que acompanha, parcialmente, o aumento da complexidade das espécies. Entretanto, assim como acontece para os animais, o tamanho do genoma pode

oscilar muito entre as espécies de um mesmo grupo taxonômico. Isso é evidente, por exemplo, em angiospermas, em que a diferença entre o menor genoma (da rosácea *Fragaria viridis*, com em torno de 98 Mb) e o maior (da liliácea *Fritillaria assyriaca*, com perto de 120.000 Mb) chega a 2.000 vezes.

Fungos, em comparação com animais e vegetais, possuem genomas menores e também apresentam uma menor amplitude na variação de tamanho. O menor genoma de fungos, até agora conhecido, é o do patógeno pulmonar de seres humanos *Pneumocystis jiroveci* (antes denominado *Pneumocystis carinii*), com 6,5 Mb, e o maior é o de *Scutellospora castanea* (ordem Diversisporales), com 795 Mb, o que significa uma amplitude de variação de apenas pouco mais de 120 vezes. Entre esses extremos, cerca de 90% das espécies de fungos têm genomas entre 10 e 60 Mb e, portanto, são menores que os genomas da grande maioria dos metazoários e dos vegetais (excluindo as algas).

Um aspecto evidente em qualquer análise mais abrangente sobre a variação de tamanho entre genomas eucarióticos é o denominado **paradoxo do valor C**. Esse fenômeno pode ser observado em muitos grupos taxonômicos e descreve uma frequente falta de correlação entre a complexidade biológica dos organismos e os tamanhos dos respectivos genomas. Organismos pertencentes a uma mesma classe, por exemplo, a dos anfíbios, podem ter variações de mais de 120 vezes no tamanho dos seus genomas, ao passo que grupos com graus de complexidade anatômica, fisiológica e comportamental muito distintas, como nematódeos, insetos e mamíferos, podem ter espécies com genomas de tamanhos similares.

O paradoxo do valor C resulta, em essência, da presença, nos genomas de algumas espécies eucarióticas, de uma grande quantidade de sequências não associadas a genes, isso faz com que o tamanho do genoma não reflita diretamente o número de genes nele presente. Essas sequências intergênicas, com representatividade variada nos genomas de diferentes espécies, são constituídas em grande parte por elementos transponíveis e outras classes de DNA repetitivo.

2.2 Forma, número e organização em replicons de cromossomos eucarióticos

O DNA nuclear de espécies eucarióticas está organizado em **cromossomos**, cada um deles composto por uma única molécula de **DNA linear**, o que contrasta com a situação dos cromossomos circulares, típicos de procariontos. A maioria dos organismos eucarióticos é **diploide**, isto é, possui dois conjuntos completos de cromossomos em cada célula somática. Contudo, espécies que são **poliploides** (com 3 ou mais cópias de cada cromossomo por célula) ocorrem entre eucariotos unicelulares, fungos, animais e vegetais. Há espécies, ainda, que apresentam fases de desenvolvimento ou formas **haploides** (com um

único conjunto de cromossomos por célula), como as células vegetativas haploides de leveduras e os machos haploides de insetos himenópteros.

O número e o tamanho dos cromossomos variam muito, dependendo da espécie considerada. Quanto ao número, há espécies como o nematódeo *Parascaris uni-valens* e a formiga *Myrmecia pilosula*, que possuem apenas um cromossomo em seus complementos cromossômicos haploides ($n = 1$), e, por outro lado, a pteridófita *Ophioglossum reticulatum*, com seus 630 cromossomos no complemento haploide ($n = 630$). Entre animais, o maior número haploide de cromossomos até agora encontrado foi de $n = 127$, no caranguejo *Eupagurus ochotensis*. O tamanho do DNA dos cromossomos eucarióticos lineares típicos pode variar desde pouco mais de 0,2 ou 0,3 Mb, como no microsporídio *Encephalitozoon cuniculi*, até dezenas ou centenas de megabases. Em seres humanos, por exemplo, o menor cromossomo (cromossomo 21) tem cerca de 47 Mb e o maior (cromossomo 1) aproximadamente 245 Mb.

Em relação à replicação, cada cromossomo eucariótico está organizado em múltiplos replicons em tandem (**Figura 5.4A**), em uma situação distinta daquela, em que cromossomos procarióticos estão organizados em um ou poucos replicons (ver Capítulo 4, Seção 2.2). Quanto a estrutura, pode-se afirmar que cada replicon eucariótico tem entre 5 e 15 kb, sendo essa extensão definida a partir do espaçamento (nem sempre regular) das múltiplas origens de replicação ao longo de cada molécula de DNA que constitui um cromossomo (Figura 5.4B). Porém, a cada ciclo celular, nem todas as origens de replicação são ativadas, o que determina, do ponto de vista funcional, que a extensão de cada replicon seja bem maior, podendo chegar a mais de 1 Mb (Figura 5.4C).

Em microrganismos eucarióticos, como espécies de leveduras (Opisthokonta), amebas (Amoebozoa) e tripanossomatídeos (Euglenozoa; Excavata), também ocorrem **plasmídeos nucleares**, que são circulares como os plasmídeos procarióticos típicos. Esses plasmídeos têm replicação autônoma e cópias múltiplas, como muitos plasmídeos procarióticos. Pelo menos em alguns casos, como em espécies do gênero *Entamoeba* (Amoebozoa), esses plasmídeos podem conter genes essenciais, como os de rRNA.

2.3 Composição de sequências

2.3.1 Sequências gênicas

Em eucariotos, os genes podem ocupar uma fração variável do genoma, dependendo da espécie. Em espécies de genoma reduzido, como o microsporídio *E. cuniculi*, essa fração chega a mais de 80%, e, em *S. cerevisiae*, os genes ocupam cerca de 70% do genoma. O homem exemplifica o outro extremo, pois a fração gênica é de 25%, considerando éxons e íntrons, e cai para 1% se forem considerados apenas os éxons (que constituem a chamada fração codificadora do genoma). A fração gênica, contudo,

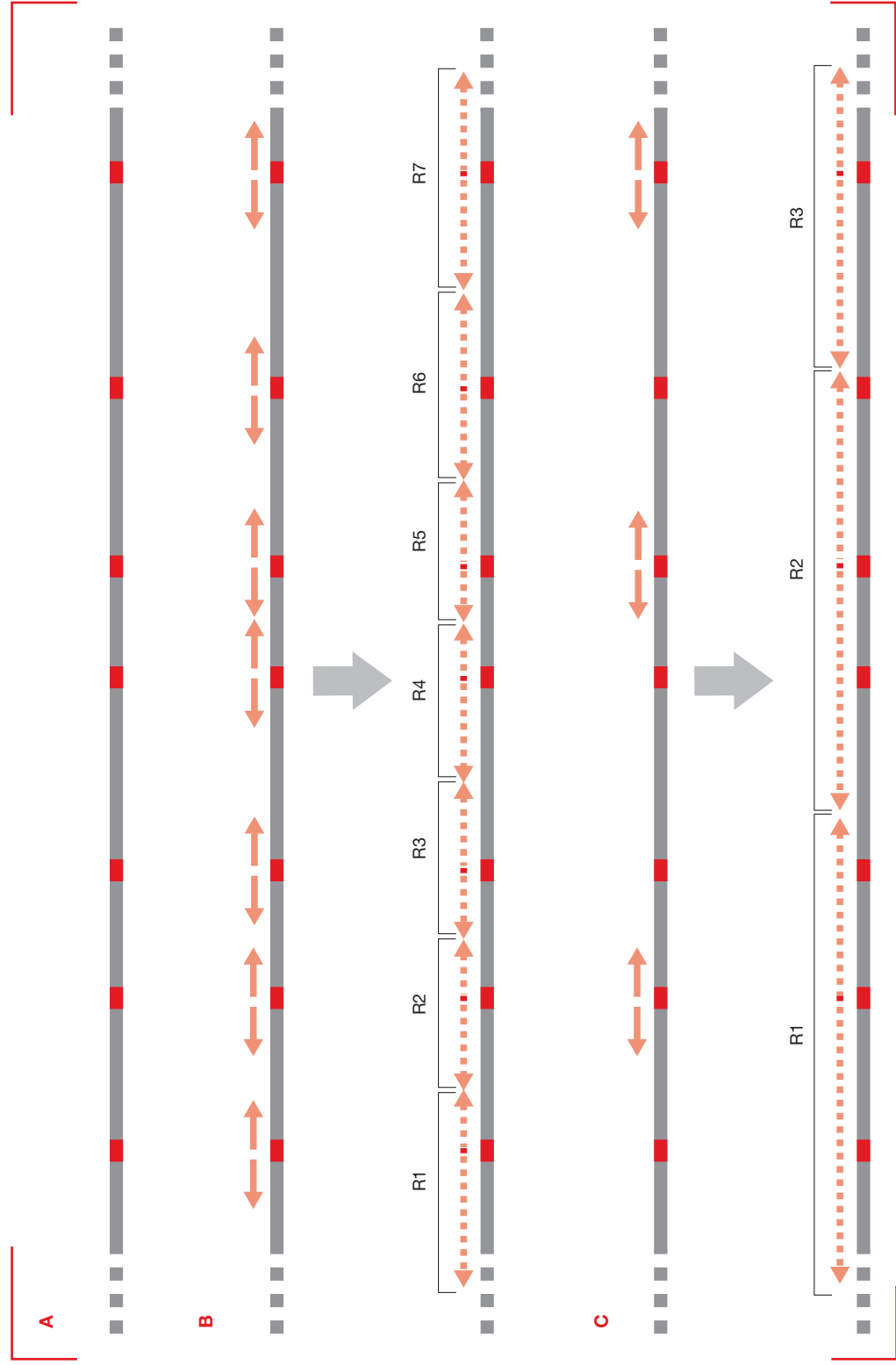


Figura 5.4

Representação esquemática da organização estrutural e funcional de replicons em cromossomos eucarióticos. (A) Em um cromossomo eucariótico típico, as origens de replicação (em ■) estão distribuídas a cada 5 a 15 kb. **(B)** Se todas as origens de replicação fossem simultaneamente ativadas a cada ciclo celular, como indicado pelas setas divergentes em ■, o tamanho dos replicons (R1-R7), determinado pelos encontros dos processos de replicação que avançam de forma bidirecional, a partir de cada origem (indicados pelas setas pontilhadas em ■), seria também de 5 a 15 kb. Entretanto, esse tipo de situação não ocorre, pois, como demonstrado em **(C)**, a cada ciclo celular, apenas algumas das potenciais origens de replicação são ativadas, de modo que a distância entre origens ativadas e, em consequência, o tamanho efetivo dos replicons (R1R3) são muito maiores.

independentemente da sua representatividade, é sempre, do ponto de vista qualitativo, muito complexa e, a seguir, serão discutidos alguns aspectos básicos da organização dos genes em genomas eucarióticos.

2.3.1.1 Número de genes e densidade gênica

Estimar o **número de genes** em um genoma eucariótico não é tarefa fácil, pois, além dos genomas serem muito grandes, os genes podem variar muito em tamanho e estrutura e não ter uma distribuição uniforme ao longo dos cromossomos. Mesmo o sequenciamento completo de um genoma eucariótico pode não ser definitivo para a determinação do número total de genes nele presente.

No caso de espécies mais complexas, a predominância de genes interrompidos, associada ao tamanho pequeno dos éxons e a ocorrência de íntrons de grande tamanho, dificulta sobremaneira a delimitação de genes individuais. Além disso, com a descoberta da transcrição disseminada e de várias novas classes de ncRNAs, a própria definição de gene ficou mais complexa, o que tornou a tarefa de determinar o número total de genes em um genoma ainda mais difícil. Por isso, as estimativas gerais se restringem à quantificação de genes codificadores de proteínas, cuja análise estrutural e funcional pode ser considerada mais simples.

A **Tabela 5.1** apresenta estimativas do **número de genes codificadores de proteínas** nos genomas de algumas espécies eucarióticas representativas em diferentes níveis de complexidade genômica. Como pode

ser observado, o número total de genes tende a aumentar conforme o grau de complexidade das espécies, pois é inferior a 10.000 para eucariotos unicelulares, fica entre 10.000 e 20.000 para invertebrados e é maior que 20.000 para o homem (e outros vertebrados) e para plantas superiores. Um número menor de genes em *D. melanogaster* (artrópode) do que em *C. elegans* (nematódeo) parece contraditório, já que a mosca é morfológica e fisiologicamente mais complexa. Da mesma forma, as diferenças de apenas 4 vezes entre o número de genes no homem e em *S. cerevisiae* ou de apenas 5.000 genes (20%) entre o homem e *C. elegans* estão bem aquém do esperado, a partir das diferenças de complexidade em todos os níveis entre essas espécies. Entretanto, tais discrepâncias podem ser inerentes à grande variabilidade de tamanhos de genomas (e possivelmente de número de genes) em muitos grupos taxonômicos (ver Figura 5.3) e/ou decorrentes de diferenças funcionais relacionadas à expressão gênica.

Um dos aspectos funcionais relevantes para a análise quantitativa de genes codificadores de proteínas diz respeito à possibilidade de um mesmo gene poder gerar mais de uma proteína diferente. Isso pode ocorrer por meio da utilização de códons de iniciação ou término de tradução alternativos (ver Capítulo 12) e/ou por *splicing* alternativo (ver Capítulos 11 e 14). Como exemplo pode-se citar o gene EPB41 humano, que codifica a proteína do citoesqueleto de eritrócitos 4.1R. Esse gene é capaz de produzir pelo menos 18 isoformas da proteína, a partir da utilização de dois códons de início de tradução alternati-

Tabela 5.1 Estimativas de tamanho, densidade gênica e proporção de genes únicos em famílias de parálogos para genomas de diferentes espécies eucarióticas¹

Espécies	Tamanho do genoma (Mb)	Número total estimado de genes ^a	Fração do genoma ocupada pelos genes ^b	Densidade gênica ^c	Genes únicos	Genes em famílias
<i>E. coli</i> ^d	4,6	4.300	90%	1,1	>90%	<10%
<i>S. cerevisiae</i>	12	6.000	73% (72%)	2	72%	28%
<i>P. falciparum</i>	23	5.300	62% (53%)	4,3	n.d.	n.d.
<i>C. elegans</i>	97	20.000	53% (27%)	5	55%	45%
<i>D. melanogaster</i>	180	13.600	24,5% (13,4%)	13	72%	28%
<i>H. sapiens</i>	3.000 ^e	25.000	25% (1%)	120	20-25%	75-80%
<i>A. thaliana</i>	157 ^f	25.500	>40%(n.d.)	6,1	10%	90%
<i>G. max</i>	1.100	46.000	n.d.	24	27%	73%
<i>Z. mays</i>	2.300	32.000	5% (2,3%)	72	16%	84%

^a Estimativas de número total de genes codificadores de proteínas. Nas estimativas, apenas genes funcionais são considerados, excluindo-se pseudogenes.

^b Para genomas eucarióticos, considerando éxons e íntrons, com a fração ocupada apenas pelos éxons entre parênteses; no caso de *S. cerevisiae* (com apenas 5% dos genes com íntrons e íntrons de pequeno tamanho), as duas estimativas são praticamente idênticas, pois a fração do genoma representada pelos íntrons não chega a 1%. Para *E. coli* (sem íntrons), a estimativa considera apenas a fração codificadora de proteínas do genoma.

^c Espaçamento médio entre genes (kb); uma densidade gênica de n kb significa um gene a cada n quilobases.

^d Os dados da bactéria *E.coli* foram incluídos para comparação com procariotos.

^e Estimativa do tamanho total da porção eucromática do genoma.

^f Estimativa com base em citometria de fluxo; essa estimativa é mais recente e é considerada mais confiável que a estimativa de 125 Mb, feita a partir do sequenciamento do genoma.

n.d. = dados não disponíveis.

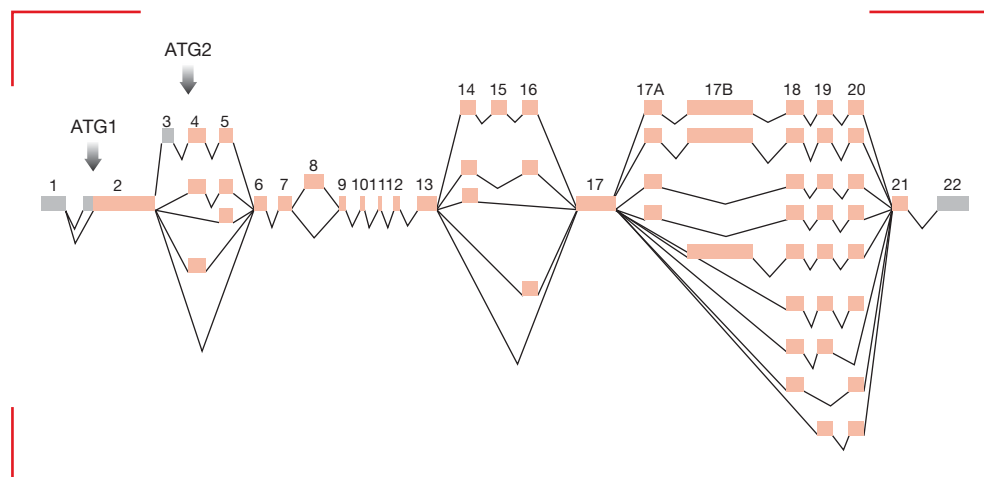


Figura 5.5

Geração de múltiplos mRNAs a partir do gene humano EPB41, que codifica a proteína 4.1R. O gene, com mais de 90 kb de extensão, é transcrito em um único pré-mRNA, que gera pelo menos 18 isoformas diferentes da proteína. Isso é possível a partir da utilização de dois códons de início de tradução alternativos (ATG1 e ATG2, indicados por setas) e do *splicing* alternativo de 12 dos 23 éxons (representados por barras e numerados) do gene. As regiões codificadoras dos éxons estão representadas em ■, e as não codificadoras, em ■; as sequências dos íntrons não estão representadas. As linhas ligam os éxons unidos por *splicing*; os eventos de *splicing* alternativo estão representados em paralelo.

vos e por *splicing* alternativo de pelo menos 12 dos seus 23 éxons (Figura 5.5). Esse tipo de fenômeno pode expandir significativamente a quantidade de proteínas produzidas a partir de um genoma e explicar, pelo menos em parte, porque genomas de espécies mais complexas não têm um número de genes maior, em proporcionalidade, que os de espécies menos complexas.

A fração de genes que sofrem *splicing* alternativo é maior em organismos mais complexos, sendo de aproximadamente 10% em *C. elegans* e subindo para mais de 40% em *D. melanogaster* e mais de 70% no homem. A análise de sequências transcritas dos genomas de *A. thaliana* e do arroz indica que também entre plantas superiores o *splicing* alternativo é comum e responsável, como entre animais, por boa parte da diversidade proteica entre espécies mais complexas.

Outro parâmetro importante relacionado às estimativas de número total de genes em genomas eucarióticos é a **densidade gênica**, definida como a distância média entre genes individuais ao longo de um genoma (considerando também genes codificadores de proteínas). A densidade gênica depende da proporção de sequências intergênicas no genoma, com tendência a aumentar mais que o número de genes à medida que avança na escala evolutiva. Assim, a densidade gênica pode variar muito entre diferentes espécies e apresenta propensão geral inversa à do número de genes, pois tende a diminuir em eucariotos mais complexos (ver Tabela 5.1). Em microsporídios de genoma reduzido, como *E. cuniculi* (genoma de 2,9 Mb), a distância entre os genes é apenas de 100 a 200 pb e a

densidade é de um gene por quilobase, comparável à de genomas procarióticos. Em *S. cerevisiae*, com genoma quatro vezes maior que *E. cuniculi* e o triplo de genes, a densidade cai para um gene a cada 2 kb.

A densidade gênica diminui em proporções ainda maiores quando são analisados organismos com genomas progressivamente maiores, variando, por exemplo, desde um gene a cada 4 ou 5 kb, para *P. falciparum* e *C. elegans*, até um gene a cada 120 kb no homem.

É importante lembrar que a densidade gênica é um parâmetro que não contempla eventuais diferenças na distribuição dos genes ao longo dos cromossomos (frequentes em genomas maiores). Por exemplo, no genoma de *D. melanogaster*, a densidade gênica varia entre 0 e 30 genes a cada 50 kb, dependendo da região considerada e, no homem, ela varia de 23 genes/Mb, no cromossomo com maior densidade (cromossomo 19) a apenas 5 genes/ Mb, no cromossomo com menor densidade (cromossomo 13). Entre os vegetais, a soja é também um exemplo de falta de uniformidade da densidade gênica, pois quase 80% dos seus genes encontram-se nas extremidades dos cromossomos, representando menos de 50% do genoma.

2.3.1.2 Famílias de parálogos, genes únicos e pseudogenes

A ocorrência de eventos de duplicação gênica ao longo da evolução determina que muitos genes façam parte de **famílias de parálogos**, cujo grau de parentesco, definido pela similaridade de sequências, pode variar desde apenas 30 até quase 100% ao longo de todo o gene ou, pelo

menos, em um ou mais de seus éxons. Em contra-partida, os genes “órfãos”, que não fazem parte de nenhuma família em um genoma, são os chamados de **genes únicos**. A proporção de genes funcionais únicos e em famílias no genoma varia de espécie para espécie, como pode ser visto na Tabela 5.1.

A partir de comparações entre procariotos e eucariotos e das espécies eucarióticas entre si, pode-se identificar uma tendência ao aumento da fração de genes em famílias (com a correspondente diminuição da fração de genes únicos) nos genomas à medida que se avança na escala evolutiva. Assim, a fração de genes únicos é maior que 50% em *S. cerevisiae*, *D. melanogaster* ou *C. elegans*, ao passo que metazoários e vegetais mais complexos, como mamíferos e angiospermas, têm de 70 a 90% de seus genes em famílias e, em consequência, apenas de 10 a 30% de seus genes são únicos.

A origem de famílias gênicas por duplicação e divergência por mutação leva também à formação de muitos parálogos não funcionais, chamados de **pseudogenes**. O número de pseudogenes desse tipo (pseudogenes duplicados ou não processados), estruturalmente distintos dos pseudogenes processados (gerados por retrotransposição), também varia muito nos genomas de diferentes espécies eucarióticas. Estima-se, por exemplo, que a quantidade total de pseudogenes (excluindo os processados) em *D. melanogaster*, *C. elegans* e no homem seja de aproximadamente 70, 2.000 e 12.000, respectivamente. A **Figura 5.6** apresenta a proporção de genes únicos, de genes em famílias e de pseudogenes no genoma humano, o que ilustra uma situação considerada típica dos genomas eucarióticos de maior tamanho, encontrados em espécies mais complexas.

A tendência ao aumento do número de genes em famílias de eucariotos mais complexos é acompanhada também pelo aumento do número de famílias diferentes e do número de genes em cada família. Por exemplo, em *S. cerevisiae*, o número de famílias é de aproximadamente 4.000, e as famílias com 4 ou mais membros representam menos de um terço do total. Em metazoários ou plantas superiores, o número de famílias gênicas é de 10.000 a 14.000, e 50% ou mais dos genes estão em famílias com 4 ou mais membros.

Na fração de genes organizados em famílias, os casos mais extremos são representados por genes idênticos (com 100% de identidade ou muito próximo disso), presentes em **cópias múltiplas**. Entre as famílias de genes de cópias múltiplas de eucariotos, serão considerados 3 exemplos marcantes, que são os dos genes codificadores de rRNAs, tRNAs e histonas. A presença de múltiplas cópias para esses genes representa uma estratégia para suprir a demanda elevada dos produtos gênicos correspondentes, que há em células eucarióticas. A maior parte do RNA celular consiste em rRNA e em tRNA, sendo o rRNA o produto de transcrição predominante, representando de 80 a 90% da massa celular total de RNA.

As histonas, por sua vez, representam aproximadamente 50% da massa da cromatina e 2% da massa total de proteínas de uma célula eucariótica típica. Repetições de genes de rRNA e tRNA ocorrem também em procariotos, mas, em eucariotos, esses genes atingem níveis de reiteração bem mais pronunciados. Os genes de histonas constituem uma família compartilhada por eucariotos e arqueas, mas os eucariotos apresentam esses genes em múltiplas cópias e organizados em arranjos que lhe são peculiares. A seguir, serão discutidos os padrões de or-

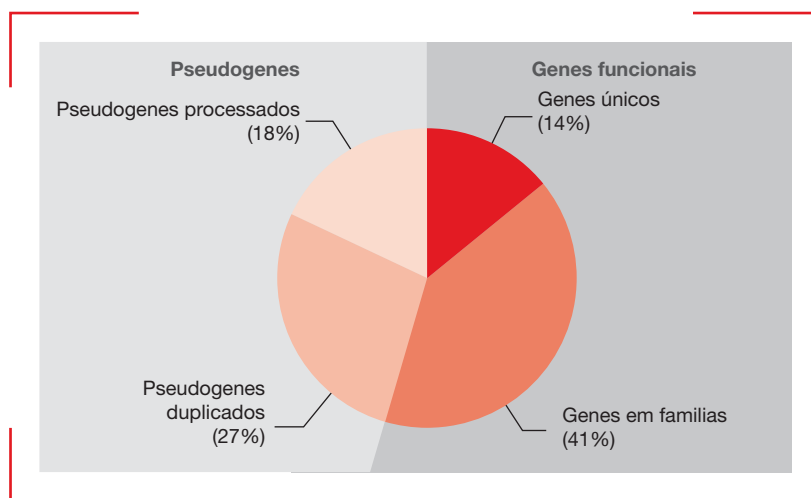


Figura 5.6

Proporção de genes únicos, genes em famílias e pseudogenes no genoma humano. Os genes únicos e os genes em famílias de parálogos constituem a fração gênica funcional do genoma. Os pseudogenes são cópias não funcionais, que podem ser geradas por eventos de duplicação seguidos por mutações (pseudogenes duplicados ou não processados) ou por eventos de retrotransposição (pseudogenes processados).

ganização típicos de eucariotos para os genes de rRNA, tRNA e histonas, lembrando que as estimativas de número de genes nem sempre são precisas, pois mesmo para genomas sequenciados a quantificação de sequências repetidas está sujeita a erros.

Os **genes que codificam rRNAs** em eucariotos, em geral, se dividem em uma unidade de transcrição, que inclui as regiões codificadoras dos rRNAs de 18S, 5,8S e 28S, e no gene de rRNA de 5S, que é transcrito individualmente (**Figura 5.7**). Os genes de rRNA de 18S, 5,8S e 28S (**Figura 5.7A**) são cotranscritos e formam, portanto, um óperon (ver Seção 2.5.1), equivalente ao óperon dos rRNAs de 16S, 23S e 5S que ocorre em genomas de procaríotos. O óperon de rRNAs de 18S, 5,8S e 28S é geralmente encontrado em repetições múltiplas em tandem (**Figura 5.7B**), cujo número varia de espécie para espécie. Sendo assim, ele sempre será muito maior do que o número de repetições encontrado para o óperon equivalente em procaríotos (menor que 10). Por exemplo, em torno de 150 cópias estão presentes no genoma de *S. cerevisiae*, aproximadamente 400 ocorrem no genoma humano e o número pode chegar a muitos milhares nos genomas de algumas espécies de plantas. Nos genomas da maioria

das espécies eucarióticas, os agrupamentos em tandem do óperon de rRNA estão distribuídos entre diferentes cromossomos, embora haja algumas espécies, como *S. cerevisiae*, em que todas as cópias estão em um único locus cromossômico.

O gene de rRNA de 5S encontra-se transcricionalmente dissociado do óperon de rRNA de 18S, 5,8S e 28S em genomas de eucariotos, em uma situação distinta da encontrada em muitas espécies de bactérias e arqueas, em cujos genomas o gene de rRNA de 5S é parte do óperon de rRNA. Há certas espécies, como *S. cerevisiae* e *Dictyostelyum discoideum* (Mycetozoa; Amoebozoa), em que os genes de rRNA de 5S individuais são encontrados em posições adjacentes ao óperon de rRNA, mas, na maioria dos genomas de espécies eucarióticas, os genes de rRNA de 5S encontram-se fisicamente distantes dos genes dos demais tipos de rRNA. É comum que os genes de rRNA de 5S de genomas eucarióticos se organizem em um (como em *D. melanogaster* e na galinha [*G. gallus*]), ou em poucos agrupamentos (como em *Drosophila virilis* ou *A. thaliana*) (**Figura 5.7C**). Pode ocorrer também uma situação mista, como no homem, que possui dois agrupamentos e vários genes individuais espalhados pelo

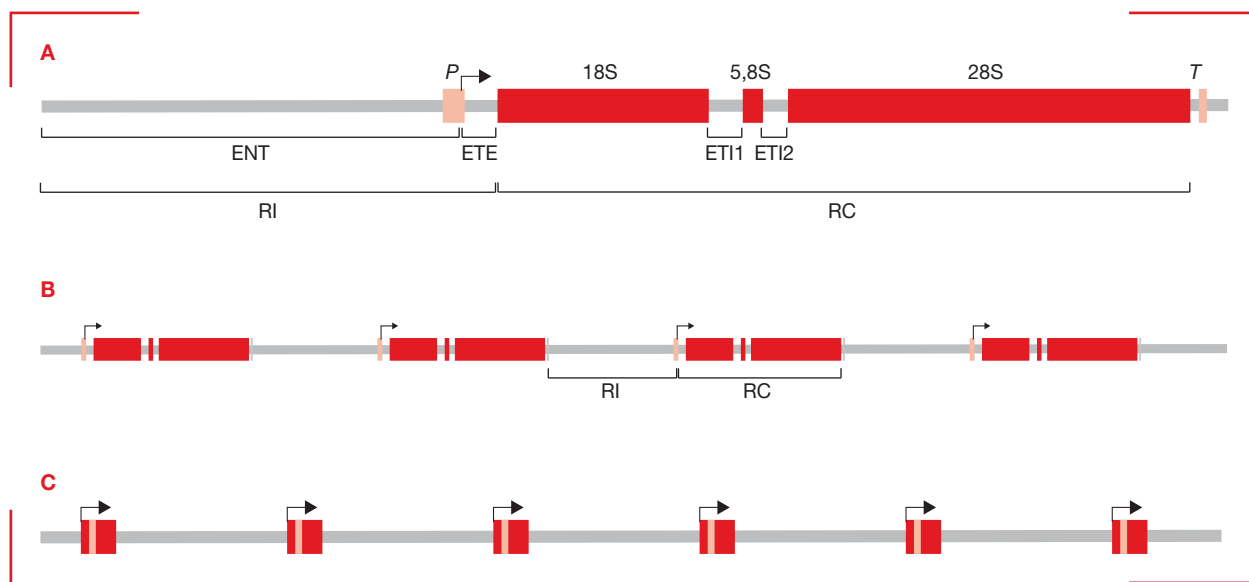


Figura 5.7

Organização típica dos genes que codificam rRNAs em genomas de eucariotos. (A) Uma unidade de repetição (óperon) dos genes de rRNA de 18S, 5,8S e 28S é composta pelas regiões codificadoras (RC) dos 3 rRNAs (em ■) e por uma região intergênica (RI). Na porção transcrita da unidade de repetição, há, além das 3 regiões codificadoras, um espaçador transcrito externo (ETE) e dois espaçadores transcritos internos (ETI1 e ETI2). A região intergênica consiste essencialmente no chamado espaçador não transcrito, embora ela inclua também as porções não transcritas dos elementos de regulação de início e término da transcrição; esses elementos regulatórios estão representados na figura (em ■), com *P* indicando o promotor e *T* indicando o terminador, e o sítio de início da transcrição está representado por uma seta. (B) Um arranjo em tandem típico de unidades de repetição de genes de rRNA de 18S, 5,8S e 28S, como a descrita em (A). Esse arranjo pode ter de dezenas até milhares de unidades de repetição regularmente espaçadas e, dependendo da espécie, dois ou mais desses arranjos podem estar presentes em diferentes cromossomos. (C) Arranjo em tandem típico de genes de rRNA de 5S individuais. Em genes de rRNA de 5S, o promotor (em ■) é interno à região codificadora (em ■). Nesse arranjo, cada gene é transcrito individualmente, estando os sítios de início da transcrição de cada um indicados por setas. Um ou mais desses agrupamentos, cada um contendo de dezenas a milhares de genes de rRNA de 5S, podem estar presentes no genoma, dependendo da espécie.

genoma. Uma distribuição exclusivamente espalhada, sem agrupamentos, é mais rara, mas já foi encontrada, por exemplo, no genoma de *S. pombe*.

O número de genes de rRNA de 5,8S pode variar desde algumas poucas centenas, como em *S. cerevisiae* (100 a 200) e no homem (em torno de 400), até alguns milhares, como no roedor *Cricetulus griseus* (em torno de 3.000) ou na ervilha *Pisum sativum* (em torno de 5.000). O número de genes em cada agrupamento também é bastante variável, por exemplo, no genoma humano há um agrupamento com 5 a 10 genes e outro com 100 a 150 genes e, no genoma da ervilha, cada um de seus 3 agrupamentos tem mais de 1.000 genes.

Os **genes de tRNA**, embora também repetidos em genomas de eucariotos, apresentam uma organização distinta da dos genes de rRNA. O número de genes diferentes varia entre as espécies, podendo chegar a 50, havendo no genoma humano, por exemplo, 38 genes diferentes de tRNA. O número de vezes que cada um dos genes se repete em um genoma também é variável conforme a espécie, assim como também é variável o número total de cópias de genes de tRNA. A distribuição desses genes ocorre geralmente em pequenos agrupamentos, que se encontram dispersos no genoma, embora possam, em alguns casos, estar concentrados próximos a regiões centroméricas, como em *S. pombe* e *C. elegans*.

As menores quantidades de genes de tRNA são encontradas nos genomas de algumas espécies unicelulares e parasitas, como *E. cuniculi*, com 46 genes de tRNA, e de *Trypanosoma brucei* (Excavata), com 70 a 120 genes de tRNA, distribuídos em pequenos agrupamentos de 2 a 10 genes. Esses números chegam a ser inferiores aos de muitos genomas procarióticos, como o de *E. coli*, que tem 86 genes de tRNA. Um número maior de genes de tRNA é encontrado nos genomas da grande maioria das outras espécies, variando desde 186, em *S. pombe*, até mais de 12.000, no peixe *D. rerio* (zebrafish).

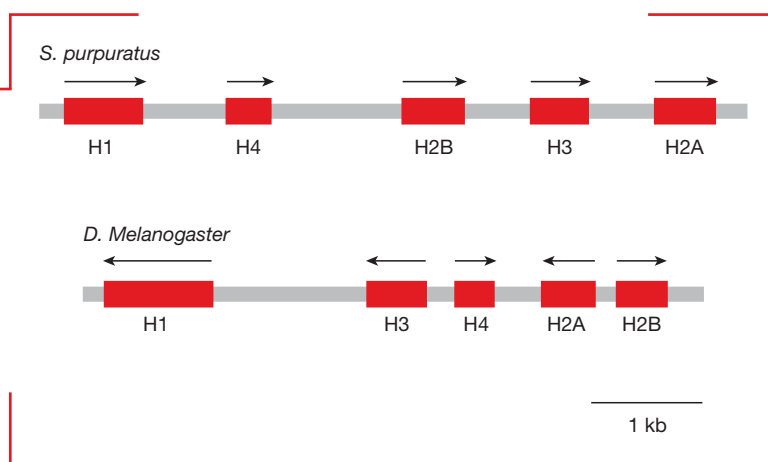
Os **genes de histonas** também constituem uma família de genes repetidos típica de genomas eucarióticos.

Essa família inclui 5 genes, cada um deles codificando um tipo diferente de histona (H1, H2A, H2B, H3 ou H4; ver Capítulo 3). A organização dessa família é heterogênea, havendo uma variação considerável não só no número de cópias, como também na distribuição dos genes no genoma. O menor número de cópias conhecido para a família é o de *S. cerevisiae*, em que o genoma possui apenas 2 unidades de repetição dos genes H2A, H2B, H3 e H4, distantes uma da outra, e um único gene da histona H1. Nos genomas de muitas outras espécies, os genes de histonas estão organizados em arranjos em tandem, com a unidade de repetição incluindo todos ou alguns dos 5 genes diferentes.

Arranjos de agrupamentos que incluem todos os 5 genes organizados em tandem no genoma ocorrem, por exemplo, em diferentes espécies de ouriço-do-mar, em que a unidade de repetição é reiterada de 300 a 1.000 vezes, e em *D. melanogaster*, em que a unidade de repetição está presente em pouco mais de 100 cópias (Figura 5.8). É comum também a ocorrência simultânea de agrupamentos que contêm todos os 5 genes e de agrupamentos que incluem apenas os genes para as histonas centrais na organização do nucleossomo (H2A, H2B, H3 e H4). Esses dois tipos de agrupamentos, organizados no genoma em arranjos em tandem, ocorrem, por exemplo, em *D. virilis* e na rã *Xenopus borealis*.

Em outros eucariotos, os genes para histonas estão organizados em agrupamentos com um número variável de genes (podendo chegar a algumas dezenas), que estão dispersos pelo genoma e não formam arranjos em tandem. No genoma do camundongo, por exemplo, há dois agrupamentos, um no cromossomo 3 (com 10 genes) e um no cromossomo 13 (com 45 genes), sendo que o agrupamento maior está dividido em 3 agrupamentos menores (um com 12, um com 15 e outro com 18 genes), separados entre si por aproximadamente 500 kb. Situações similares ocorrem também nos genomas de outros vertebrados, como a galinha (com 43 genes organizados em dois agrupamentos principais), e no de *C. elegans*, que possui 11 agrupamentos distribuídos em diferentes cromossomos. No milho,

Figura 5.8
Organização dos agrupamentos de genes de histonas nos genomas de *S. purpuratus* e *D. melanogaster*. Cada agrupamento constitui uma unidade de repetição que inclui cinco genes (em ■), um para cada classe de histonas; as sequências intergênicas estão representadas em □. As setas indicam o sentido da transcrição de cada um dos genes.



os múltiplos genes de histonas presentes no genoma não formam agrupamentos e estão dispersos individualmente.

2.3.2 Sequências intergênicas

As **sequências intergênicas** são todas as sequências genômicas que não estejam vinculadas diretamente a um gene. A delimitação dessas regiões “entre genes” ou “não gênicas” pode ser principalmente complicada em genomas eucarióticos, pois neles os limites físicos dos genes nem sempre são claros. São fatores complicadores as extensas, complexas e heterogêneas regiões reguladoras dos genes (a maioria ainda não caracterizada), o grande número de íntrons e sua grande extensão (em especial, em genomas de eucariotos mais complexos) e o recente conhecimento de que muitas das regiões antes tratadas como intergênicas podem ser alvo de transcrição. Assim, por simplificação, são em geral tratadas como intergênicas todas as sequências que não sejam parte de éxons ou íntrons de genes funcionais de estrutura típica, como a descrita na Seção 1.1. Essa simplificação é útil para as análises mais amplas do conteúdo de regiões intergênicas, como a realizada neste capítulo.

Grande parte do genoma da maioria das espécies eucarióticas é formada por sequências intergênicas. A quantidade dessas sequências tende a aumentar proporcionalmente mais do que a quantidade de sequências gênicas nos genomas de eucariotos mais complexos, explicando o chamado “paradoxo do valor C” (ver Seção 2.1). Nos primeiros estudos, essa grande quantidade de DNA intergênico em genomas eucarióticos maiores foi tratada como “lixo genômico” (do inglês, “*junk DNA*”), devido a sua natureza predominantemente repetitiva e por não ter qualquer função mais evidente. Hoje, contudo, se reconhece que essa fração de genomas eucarióticos inclui

diversos tipos de elementos de sequência importantes para a fisiologia e para a evolução dos genomas eucarióticos e não pode ser tratada como um mero componente estrutural. A partir desse reconhecimento, foi proposto, em 2001, que o antigo “paradoxo do valor C” vigente desde a década de 1970 fosse substituído pelo **enigma do valor C**, que engloba uma série de questões referentes à organização e às funções dessas sequências em genomas eucarióticos.

A **Figura 5.9** mostra a composição de sequências gênicas e intergênicas do genoma humano. Essa composição pode ser considerada típica de genomas eucarióticos maiores, encontrados em eucariotos mais complexos, e ilustra a heterogeneidade e a representatividade das sequências intergênicas. A seguir, serão discutidas as duas classes mais representativas de sequências intergênicas eucarióticas – a das sequências repetidas simples e a dos elementos transponíveis.

2.3.2.1 Sequências repetidas simples

A fração intergênica de genomas eucarióticos, principalmente os de espécies mais complexas, apresenta um grande número de sequências curtas que podem estar repetidas de milhares a milhões de vezes. Elas são chamadas, de forma geral, de **sequências repetidas simples** e podem constituir uma fração significativa de um genoma. Em mamíferos, as sequências repetidas simples representam menos de 10% do genoma, mas em outras espécies, a representatividade dessas sequências pode chegar a quase 50%, como em *D. virilis*.

O tamanho da unidade de repetição de cada família de sequências repetidas simples varia, e diferentes famílias são encontradas nos genomas de diferentes espécies. São mais comuns unidades de repetição de 5 a 10 pb, mas

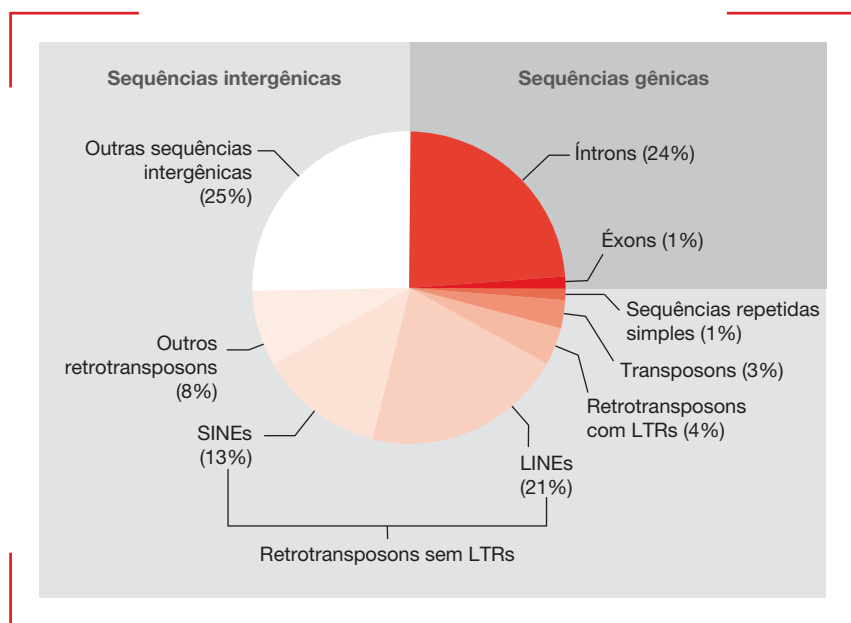


Figura 5.9

Composição de sequências gênicas e intergênicas do genoma humano. A fração gênica é representada, pelas sequências transcritas (éxons e íntrons) de genes codificadores de proteínas. As regiões intergênicas são compostas pelas sequências repetidas simples, pelos elementos transponíveis (transposons e retrotransposons) e por outras sequências, incluindo pseudo-genes.

muitas famílias com unidades de repetição maiores, com até mais de 300 pb também já foram identificadas, por exemplo, em genomas de insetos, mamíferos e vegetais superiores. As unidades de repetição, em sua maioria, estão organizadas em vários arranjos em tandem, cada um deles podendo chegar a dezenas ou centenas de quilobases de extensão. A **Tabela 5.2** mostra alguns exemplos de DNA de sequência simples que ocorrem em diferentes espécies eucarióticas.

Os arranjos em tandem de sequências repetidas simples muitas vezes formam frações do genoma com conteúdo de G+C distinto do conteúdo da média do restante do genoma. Por isso, essas sequências eram isoladas de DNA genômico fragmentado (por ultrassom, p. ex.) em experimentos de sedimentação diferencial em gradientes de densidade. O gradiente então separava um componente (banda) principal, correspondente à fração gênica e a outros elementos de sequência mais variados, e um ou mais componentes (famílias) adicionais, de sequências repetidas simples. Esses componentes adicionais recebiam o nome de “satélites” e, por essa razão histórica, as sequências repetidas simples também podem ser chamadas de **DNA-satélite**. Repetições ainda menores, de 2 ou 3 pb, são chamadas de **microsatélites**, sendo importante salientar que alguns microsatélites podem ocorrer inclusive dentro de éxons e íntrons gênicos.

Quanto à distribuição, os arranjos em tandem de uma determinada família de sequências repetidas simples podem estar tanto dispersos pelo genoma como associados a regiões específicas, tanto de eucromatina como de heterocromatina (ver Tabela 5.2). É comum, por exemplo, a associação de famílias de sequências repetidas simples com regiões específicas de cromossomos metafásicos, principalmente centrômeros e telômeros. Isso acontece, por exemplo, com as sequências alfoides do homem e com o DNA-satélite telomérico de *D. virilis*.

Em relação à origem, acredita-se que as sequências repetidas simples podem ter tido suas primeiras cópias originadas, a partir de eventos de transposição, que duplicam os sítios-alvo de inserção do elemento móvel (ver Capítulo 9). Um número inicial dessas cópias, gerado em espécies ancestrais, teria depois expandido, ao longo da evolução, por meio de sucessivos eventos de recombinação desigual (ver Capítulo 8). Dentro de uma mesma família, as sequências das unidades de repetição são muito conservadas. Entretanto, o número de unidades de repetição presente em cada arranjo em tandem é bastante variável, inclusive entre diferentes indivíduos de uma mesma espécie. Essa variação é resultado da ocorrência de recombinação desigual envolvendo tais sequências, durante divisões meióticas que dão origem aos gametas.

Tabela 5.2 Sequências repetidas simples de DNA em diferentes espécies eucarióticas^a

Espécie	Extensão da unidade de repetição (pb)	Sequência da unidade de repetição ^b	Localização cromossômica
<i>C. elegans</i> (verme)	24	CeRep25B	Dois agrupamentos em regiões eucromáticas do cromossomo III
<i>D. melanogaster</i> (mosca)	359	359-bp repeat	Heterocromatina do cromossomo X
<i>D. virilis</i> (mosca)	7	ACAAACT (satélite I)	Centrômeros
	7	ATAAACT (satélite II)	Centrômeros
	7	ACAAATT (satélite III)	Centrômeros
	370	Satélite telomérico	Telômeros
<i>Cancer borealis</i> (crustáceo)	2	AT	n.d.
<i>Pagurus pollicaris</i> (crustáceo)	3	CAG	n.d.
	4	CCTA	n.d.
<i>Cavia porcellus</i> (roedor)	6	CCCTAA	Centrômeros
<i>Dipodomys ordii</i> (roedor)	10	ACACAGCGGG	Centrômeros
<i>Cercopithecus aethiops</i> (primata)	172	Sequências α	Centrômeros
<i>H. sapiens</i> (ser humano)	171	Sequências alfoides	Centrômeros
<i>Hordeum vulgare</i> (angiosperma)	3	AAC	Heterocromatina, Heterocromatina
	3	AAG	Centrômeros
	3	ACG	Eucromatina (cromossomos 2, 3, 4, 5 e 6)
	3	ACT	Heterocromatina
		AGG	Centrômeros
		CAC	Centrômeros (cromossomos 3 e 4)
		CAT	Heterocromatina (cromossomos 4 e 5)

^a Apenas um ou poucos exemplos das várias famílias de sequências repetidas simples de cada espécie são apresentadas.

^b Apenas as sequências de unidades de repetição ≤ 10 pb são apresentadas; as sequências maiores estão identificadas apenas pelo nome genérico da família.

A expansão das sequências repetidas simples em genomas eucarióticos ao longo da evolução sugere pressões seletivas favoráveis à manutenção dessas repetições. Essas pressões podem decorrer de alguns papéis importantes dessas sequências na estruturação e no funcionamento de cromossomos. Por exemplo, as sequências – e alfoides de primatas definem a estrutura dos centrômeros de modo a viabilizar a sua interação com microtúbulos do fuso mitótico ou meiótico. Em *D. virilis*, a recombinação mediada por sequências repetidas simples é responsável pelo alongamento e pela manutenção das sequências teloméricas dos cromossomos. Além disso, a recombinação desigual proporcionada por sequências repetidas simples é passível de incluir sequências adjacentes às repetições e pode incrementar também de forma qualitativa a variabilidade genética de uma espécie, aumentando a sua adaptabilidade ao longo da evolução. Variações em microssatélites são capazes de determinar alterações no nível de expressão ou no produto de genes adjacentes, com efeitos fenotípicos evidentes. Exemplos disso são as diversas doenças humanas, causadas por variações no número de repetições em microssatélites associadas a certos genes.

2.3.2.2 Elementos transponíveis

Os **elementos transponíveis** são sequências de DNA capazes de se integrarem (por transposição) em outros locos genômicos, o que, em geral, implica aumento do seu número de cópias no genoma. Os elementos transponíveis se dividem em dois grupos principais: o dos transposons de DNA, que são mobilizados por meio de intermediários de DNA, e o dos retrotransposons, cuja mobilização se dá por um mecanismo que envolve a transcrição reversa de um intermediário de RNA (ver Capítulo 9).

Na maioria das espécies eucarióticas, os elementos transponíveis constituem a maior parte do DNA intergênico e, em alguns casos, a maior parte do genoma. Entretanto, o conteúdo de elementos transponíveis em genomas eucarióticos pode variar desde a ausência total, como no caso do genoma reduzido de *E. cuniculi*, até mais de 80%.

Os **transposons de DNA** representam, em geral, uma fração menor dos elementos transponíveis eucarióticos, em uma situação oposta à observada em genomas procarióticos, em que os transposons são a classe principal. No homem, os transposons constituem 3% (90 Mb) do genoma, o que, ainda assim, é 3 vezes o total de sequências gênicas codificadoras (éxons) (ver Figura 5.9). A maior parte dos transposons de mamíferos acumulou muitas mutações ao longo da evolução, de modo que é improvável haver qualquer um ainda ativo (funcional) nessa classe de vertebrados. O mesmo acontece entre plantas superiores, nas quais a fração do genoma ocupada por transposons fica entre 10 e 20%.

Dois das famílias de transposons eucarióticas mais conhecidas e com cópias ainda ativas são a dos elementos *Ac/Ds*, do milho, e a dos elementos P, de algumas espécies de moscas do gênero *Drosophila*. Os elementos *Ac*

possuem 4,5 kb e codificam sua própria transposase (enzima que medeia a transposição, ver Capítulo 9), ao passo que os elementos *Ds* são elementos *Ac* truncados, cuja transposição depende de elementos *Ac* completos. De 5 a 12 cópias de elementos *Ac/Ds* são encontradas no genoma do milho, dependendo da linhagem. Os elementos P de *Drosophila* sp., por sua vez, têm um tamanho de 2,9 kb para cópias autônomas (para transposição), que, assim como os elementos *Ac*, codificam sua própria transposase. Ocorrem também cópias truncadas, chamadas de elementos P não autônomos, dependentes de elementos P autônomos para transposição. Em linhagens de moscas portadoras do elemento P podem ser encontradas até 50 cópias do elemento P espalhadas em diferentes cromossomos, com aproximadamente dois terços delas sendo não autônomas.

Os **retrotransposons**, por sua vez, constituem a maioria dos elementos genéticos móveis encontrados em genomas eucarióticos e, conseqüentemente, representam também a maior parte das sequências intergênicas. Em comparações intra e interfiléticas, a quantidade de retrotransposons está, pelo menos em linhas gerais, correlacionada com o tamanho do genoma, o que torna os retrotransposons os maiores responsáveis pelo incremento no tamanho dos genomas observado em eucariotos mais complexos. No homem, por exemplo, eles constituem 56% das sequências intergênicas e 42% de todo o genoma (ver Figura 5.9).

Os retrotransposons são divididos em duas categorias principais, uma de retrotransposons contendo longas repetições terminais (LTRs, de *long terminal repeats*), que corresponde aos retrotransposons virais (semelhantes a retrovírus), e outra de retrotransposons sem LTRs, que inclui os retrotransposons não virais.

Os **retrotransposons com LTRs** variam em tamanho, de algumas centenas de pares de bases até mais de 10 kb, com as LTRs tendo, em geral, de poucas centenas a alguns milhares de pares de bases de extensão. Eles são encontrados dispersos nos genomas de todos os eucariotos e são abundantes, por exemplo, em leveduras e em plantas superiores. Em *S. cerevisiae*, 377 kb ou 3,1% do genoma, é composto por retrotransposons, o que é bastante significativo para um genoma compacto. Na soja e no milho, eles constituem 42 e 70% do genoma, respectivamente. Em mamíferos, eles são menos comuns que os retrotransposons sem LTRs, embora também estejam presentes em número significativo. No homem, por exemplo, retrotransposons com LTRs constituem aproximadamente 4% do genoma.

Os elementos Ty de *S. cerevisiae* com 6,3 kb, quando íntegros, formam 5 famílias (Ty1-5), totalizando um número de cópias (completas ou incompletas) que varia de poucas dezenas até mais de 300, conforme a linhagem. Na linhagem que primeiramente teve o genoma sequenciado, foram encontradas 331 cópias, sendo 217 de Ty1, 34 de Ty2, 41 de Ty3, 32 de Ty4 e 7 de Ty5. Do total dos elementos Ty presentes no genoma dessa linhagem, 85%

correspondem a cópias truncadas. O elemento *Grande*, por sua vez, é um retrotransposon com LTR abundante no genoma de espécies do gênero *Zea*, como o milho. Esse elemento deve o seu nome à longa extensão que possui, de 13,8 kb, ele também tem distribuição dispersa por todos os cromossomos. O número de cópias de elementos *Grande* varia de 5.700 a 11.600 por genoma haploide, conforme a espécie ou linhagem e, em média, eles representam 3% do genoma.

Os **retrotransposons sem LTRs** de genomas eucarióticos são representados por duas classes principais, primariamente distinguidas pelos seus tamanhos. A primeira é constituída por elementos mais longos, com 6 a 7 kb quando íntegros, chamados de LINEs (do inglês, *long interspersed nuclear elements* = elementos nucleares longos dispersos). A segunda é constituída por elementos mais curtos, geralmente com 100 a 300 pb, chamados de SINEs (do inglês, *short interspersed nuclear elements* = elementos nucleares curtos dispersos). Como o próprio nome de cada tipo de elemento indica, LINEs e SINEs são encontrados como cópias dispersas ao longo de qualquer genoma onde eles ocorram.

Os **LINEs**, quando íntegros, codificam pelo menos uma transcriptase reversa e uma endonuclease (RNase). Na extremidade 3', eles contêm uma cauda de poli(A), o que evidencia a transcrição típica pela RNA-polimerase II (ver Capítulo 10), e a reintegração do elemento no genoma. Os LINEs são considerados ubíquos em eucariotos, pois estão representados em protistas, fungos, invertebrados, vertebrados e vegetais, e é provável que sejam a classe (do ponto de vista evolutivo) mais antiga de retrotransposons. Eles são abundantes em genomas de mamíferos, onde são reconhecidas mais de 10 famílias distintas, cada uma constituída por milhares de membros. Por exemplo, o genoma humano contém de 20.000 a 40.000 LINEs, o correspondente a mais de 21% das suas sequências. A família mais representativa de LINEs em mamíferos é chamada de LINE-1 (ou simplesmente L1), seus membros representam 17% do genoma humano e 20% do genoma do camundongo.

Os **SINEs** são elementos originários da transcrição reversa e da integração entre algumas sequências inicialmente transcritas pela RNA-polimerase III (ver Capítulo 10), como tRNAs, rRNA de 5S ou RNA 7SL (componente da partícula de reconhecimento de sinal, um elemento ribonucleoproteico que auxilia na secreção de polipeptídeos recém-sintetizados através das vesículas do retículo endoplasmático). Os SINEs não codificam transcriptase reversa e dependem de outros retrotransposons, como os LINEs, para sua retrotransposição. Esses elementos são abundantes em mamíferos, mas pouco frequentes em outros grupos taxonômicos, como em insetos e em plantas.

Dentro de um mesmo grupo taxonômico, a presença de SINEs no genoma também pode ser bastante variável. Entre insetos, por exemplo, não há SINEs na porção sequenciada (eucromatina) do genoma de *D. melanogaster*, mas há várias famílias distintas, algumas com mais de 7.000 cópias, no genoma do mosquito *Aedes aegypti*.

Em primatas, os SINEs mais abundantes são as chamadas sequências *Alu*, que se caracterizam por um alto grau de homologia (de até 80%) com o RNA 7SL e pela usual presença de um sítio de clivagem para a endonuclease de restrição *AluI*. No homem, há em torno de 1,5 milhão de cópias de sequências *Alu*, que representam 13% do genoma.

Um aspecto bastante interessante dos retrotransposons é a capacidade que alguns deles possuem de mobilizarem não só a si próprios, mas também outras sequências, por retrotransposição (isto é, transcrição reversa de um RNA em DNA e integração da cópia de DNA no genoma). Por exemplo, LINEs da família L1 podem mobilizar SINEs da família *Alu*. É interessante considerar também o fato de que transcritos de mRNA maduros podem sofrer o mesmo processo, de modo a terem suas sequências integradas ao genoma em diferentes sítios (**Figura 5.10**). Essas sequências são destituídas do promotor e de outras sequências reguladoras da expressão original e, por isso, as cópias integradas não são em geral funcionais, constituindo pseudogenes. Entretanto, como as sequências integradas sofrem *splicing*, para remoção de íntrons, e adição de cauda de poli(A) (ver Capítulo 11), elas claramente se distinguem dos pseudogenes parálogos, gerados a partir de cópias duplicadas (ver Seção 2.3.1.2), sendo então chamados de **pseudogenes processados**.

Os pseudogenes processados foram até agora observados apenas em genomas de metazoários e de plantas angiospermas. Eles são abundantes em mamíferos, mas menos comuns em outras espécies. Por exemplo, no genoma humano, estima-se que, do total de 20.000 pseudogenes, 8.000 sejam pseudogenes processados. O genoma do camundongo, por sua vez, tem 14.000 pseudogenes, dos quais 5.000 são pseudogenes processados. Em *C. elegans* e *D. melanogaster*, o número de pseudogenes processados cai para 208 (2.200 pseudogenes no total) e 34 (110 pseudogenes no total), respectivamente.

A mobilidade dos elementos transponíveis e o aumento do número de cópias por recombinação desigual explicam mecanisticamente a disseminação dessas sequências em genomas eucarióticos. Entretanto, o acúmulo desses elementos em genomas eucarióticos maiores, em proporções superiores à da fração gênica, exige considerações em termos funcionais e evolutivos. Na variação dos tamanhos dos genomas entre espécies eucarióticas, há uma evidente contribuição dos elementos transponíveis. Em plantas, por exemplo, a proliferação ou a eliminação de elementos transponíveis são, juntamente à poliploidização, os principais fatores responsáveis por variações de tamanho de genoma em diferentes espécies. A abundância de elementos transponíveis em um genoma também é um fator de desestabilização, seja por eventos de mobilização frequentes ou por rearranjos cromossômicos proporcionados por recombinação entre os elementos repetidos. Com isso, os elementos transponíveis são uma força importante para a rápida reestruturação de genes e cromossomos e, consequentemente, para a evolução de genomas eucarióticos.

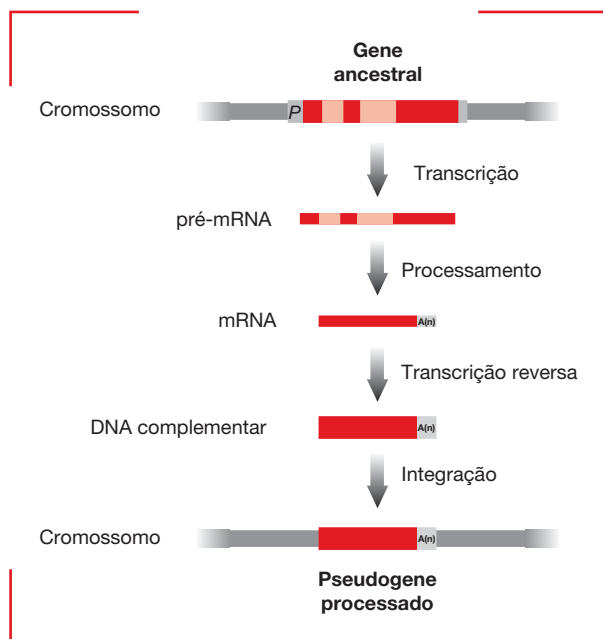


Figura 5.10

A origem de um pseudogene processado. Os pseudogenes processados são originados a partir de um transcrito de pré-mRNA que normalmente é processado. Na figura, as sequências gênicas estão representadas pelos éxons (em ■), íntrons (em ■) e regiões reguladoras (em ■), com P indicando o promotor. Por ação de enzimas codificadas por retrotransposons ou retrovírus, o mRNA processado é transcrito reversamente (por uma transcriptase reversa) em um DNA complementar, que depois é integrado no genoma (por uma integrase). Note que o pseudogene processado não tem íntrons ou promotor, mas tem uma cauda de poli(A), indicada por A(n).

Os elementos transponíveis podem alterar os padrões de expressão de alguns genes ou até dar origem a novos genes. Elementos transponíveis ativos são agentes mutagênicos, e a sua inserção em promotores ou em regiões próximas pode levar a alterações no padrão de expressão dos genes controlados por essas regiões reguladoras. Há elementos transponíveis potencialmente ativos em qualquer espécie, sendo que esses elementos são, por exemplo, responsáveis por 10% das mutações em roedores. Além disso, inserções em éxons podem resultar na incorporação da sequência do elemento transponível à região codificadora de uma proteína, causando modificações estruturais e funcionais nos respectivos produtos gênicos, e alguns mRNAs retrotranspostos podem ser inseridos próximos a promotores e tornarem-se genes processados funcionais (retrogenes). É possível considerar que, no genoma humano, de 50 a 100 genes codificadores de proteína evoluíram a partir de sequências de transposons ou retrotransposons. Por exemplo, o gene *RAG1*, que medeia a recombinação das sequências codificadoras de regiões variáveis de proteínas do sistema imune, foi originalmente o gene de uma transposase de um elemento transponível em um ancestral de vertebrados.

Por fim, os elementos transponíveis podem ter também papéis relevantes na estruturação dos cromossomos e na fisiologia da cromatina. Por exemplo, os telômeros de cromossomos de moscas do gênero *Drosophila* são mantidos por dois retrotransposons sem LTRs (Het-A e TART), que se retrotranspõem especificamente para as extremidades dos cromossomos. Em um cenário mais amplo, os elementos transponíveis podem ser determinantes de processos epigenéticos (que envolvem fatores não genéticos que alteram a expressão gênica). Isso é evidenciado pela associação de sequências de elementos transponíveis, por exemplo, ao posicionamento de nu-

cleossomos, ao recrutamento de modificações no DNA (como a metilação), à cromatina (como as modificações de histona) e à delimitação de regiões de heterocromatina ou de eucromatina.

2.4 Sintenia

A comparação da ordem dos genes nos genomas entre diferentes espécies eucarióticas revela níveis de sintenia significativos e sua preservação, pelo menos em parte, mesmo ao longo de distâncias evolutivas consideráveis. As relações de sintenia genômica são evidentes em praticamente todos os grupos taxonômicos eucarióticos já analisados quanto a esse aspecto, mas os níveis de sintenia variam em cada grupo. Entre insetos, por exemplo, há sintenia de 99% entre genomas de diferentes espécies do gênero *Drosophila*, porém ela cai para apenas 10% quando os genomas dessas moscas (ordem Diptera) são comparados com o da abelha comum (*Apis mellifera*, ordem Hymenoptera). Entre vertebrados, parece haver uma maior conservação de sintenia, sendo evidente mesmo em comparações de espécies mais distantes em termos de parentesco evolutivo. Por exemplo, quando se compara os genomas do homem e da galinha, separados há 310 milhões de anos, ou do homem e de peixes, como o *zebrafish* e o fugu (*Takifugu niphobles*), separados do homem há 450 milhões de anos, podem ser detectados blocos sintênicos que variam de 1,8 Mb até 70 Mb de extensão. Blocos sintênicos que se estendem ao longo de 90% dos genomas são observados em comparações entre o homem e o camundongo, duas espécies de mamíferos separadas há menos tempo (91 milhões de anos).

Entre plantas, altos níveis de sintenia genômica já foram evidenciados, por meio do alinhamento de marcadores genéticos, entre espécies de diferentes gêneros de

uma mesma família. Comparações de sequências genômicas completas evidenciam sintenia mesmo entre espécies como o arroz (monocotiledônia) e *A. thaliana* (dicotiledônia), separadas há 200 milhões de anos. Essas duas espécies mantêm muitos blocos genômicos sintênicos, incluindo uma região que abrange 119 genes.

O grau de conservação sintênica, mesmo ao longo de distâncias evolutivas da ordem de 100 milhões de anos ou mais, demonstra que a perda de sintenia genômica em eucariotos ocorre de forma mais lenta do que em procariotos. É certo que, em bactérias ou arqueas, quando o grau de identidade global entre sequências de aminoácidos de duas espécies é da ordem de 90%, já não se observa mais qualquer vestígio de sintenia, ao passo que em eucariotos, a sintenia só se perde por completo quando essa identidade é de 50% ou menos. Isso faz com que, em eucariotos, blocos de sintenia possam ser observados mesmo quando são comparadas espécies de classes distintas (mamíferos e peixes, p. ex.), ao passo que entre procariotos, as sintenias só podem, em geral, ser observadas no máximo entre duas espécies diferentes de um mesmo gênero. Uma provável explicação para isso é a aceleração da perda de sintenia por mecanismos determinantes de inversões pericêntricas, os quais são muito ativos em procariotos (ver Capítulo 4, Seção 2.6.3), mas não ocorrem em eucariotos.

2.5 Agrupamentos organizados de genes

Como visto no Capítulo 4 (Seção 2.4), podem ser identificadas em genomas de procariotos diversas unidades organizacionais envolvendo grupos de genes estrutural e/ou funcionalmente relacionados, como óperons, domínios e ilhas genômicas. Em eucariotos, agrupamentos organizados de genes dessa ordem são mais escassos ou, pelo menos, de identificação bem mais difícil, devido à natureza individualizada da regulação da expressão da grande maioria dos genes e aos maiores tamanho e complexidade dos genomas. Mesmo assim, alguns óperons e outros agrupamentos organizados podem ser reconhecidos em genomas eucarióticos e os principais tipos serão descritos a seguir.

2.5.1 Óperons eucarióticos

Os genes eucarióticos típicos estão organizados em unidades de transcrição individuais, com cada gene sendo transcrito em um produto de RNA monocistrônico (ver Seção 1.1). Portanto, ao contrário do que ocorre em bactérias e arqueas (ver Capítulo 4), é menos comum em eucariotos a organização em óperons, com dois ou mais genes (ou regiões codificadoras) sendo transcritos em um único RNA policistrônico. Há, contudo, várias exceções conhecidas, sendo o caso mais extremo o de euglenozoários (Excavata) da ordem Kinetoplastida (incluindo tripanossomatídeos), que têm a maioria de seus genes organizados em unidades de transcrição policistrônicas. Entretanto, ao contrário do que acontece em óperons procarióticos, os transcritos policistrônicos não são traduzi-

dos de forma direta, sendo convertidos antes em mRNAs monocistrônicos por *trans-splicing* (ver Capítulo 11), e a regulação da expressão gênica se dá principalmente em nível pós-transcricional. A ocorrência de óperons em genomas de nematódeos, também é marcante como no de *C. elegans*, que tem aproximadamente 3.000 (15%) de seus genes organizados em 1.000 unidades de transcrição policistrônicas, cada uma delas incluindo de 2 a 8 genes.

Em nematódeos, os transcritos policistrônicos também são processados por *trans-splicing*, embora, do ponto de vista evolutivo, óperons de nematódeos e de tripanossomatídeos não estejam relacionados entre si, assim como não estão relacionados aos óperons procarióticos. Casos mais pontuais de ocorrência de óperons também foram descritos em espécies de celenterados, platelmintos, insetos, cordados primitivos e plantas.

2.5.2 Agrupamentos de genes funcionalmente relacionados

O agrupamento eventual de parálogos durante a evolução de um genoma é relativamente comum em eucariotos, pois as duplicações de segmentos cromossômicos parecem ser eventos recorrentes na história evolutiva de muitas linhagens eucarióticas (ver Seção 3). Entretanto, a longo prazo em termos evolutivos, esses genes parálogos mantêm-se agrupados apenas se houver pressão seletiva para tanto. Caso contrário, os parálogos se dispersam, como ocorre, por exemplo, com as famílias de genes de actina (**ver Material Complementar Online 5.3**).

Há, contudo, muitos casos de famílias multigênicas eucarióticas cujos membros são mantidos agrupados em função das vantagens funcionais proporcionadas por esse tipo de organização. Exemplos marcantes disso são os agrupamentos de genes de globina de vertebrados e os genes de fatores transcricionais homeóticos (genes *Hox*) de metazoários, cujos padrões de expressão espacial e temporal dependem da organização de cada agrupamento. Os genes de globina estão organizados em agrupamentos e são expressos de maneiras diferentes ao longo do desenvolvimento de vertebrados em eritrócitos (**ver Material Complementar Online 5.3**). O padrão de expressão espacial (eritrócito-específico) e temporal (estágio-específico) de cada gene depende, pelo menos em parte, de sua posição no agrupamento e da proximidade relativa com elementos reguladores distais (discutidos no Capítulo 14), como reforçadores (do inglês, *enhancers*) ou regiões controladoras de locus (LCRs, do inglês *locus control regions*). Os genes *Hox* organizados em agrupamentos, por sua vez, determinam o padrão de diferenciação anteroposterior em metazoários. A ordem desses genes em cada agrupamento (que têm em geral de 4 a 10 genes, conforme a espécie e o agrupamento) corresponde à ordem temporal da expressão ao longo do desenvolvimento do organismo e à ordem espacial de expressão de cada gene ao longo do eixo anteroposterior. Em outras palavras, genes *Hox* anteriores no agrupamento são expressos primeiro e em posição mais anterior do corpo que os genes subsequentes (ver Capítulo 15). Esse padrão de

expressão espaço-temporal vinculado à posição no agrupamento é regulado, pelo menos em parte, pelo silenciamento seletivo de parte do agrupamento por compactação da cromatina.

Há ainda muitos outros agrupamentos de genes funcionalmente relacionados que viabilizam mecanismos de geração de variabilidade. Em genomas de vertebrados, são exemplos disso os agrupamentos dos genes que codificam os vários componentes do sistema imune e os agrupamentos dos genes de receptores olfatórios. Os agrupamentos dos genes do sistema imune viabilizam o mecanismo de recombinação que proporciona variabilidade a anticorpos e imunorreceptores. Os milhares de genes de receptores olfatórios, por sua vez, são organizados em múltiplos agrupamentos, nos quais o número de membros diferentes e o número de cópias de cada membro podem variar de indivíduo para indivíduo. Essa variabilidade, resultante dos eventos de duplicação e deleção proporcionados pelo agrupamento dos genes similares, é responsável pela evolução dessa família e pela adaptação do repertório de receptores às necessidades olfatórias de cada espécie.

Em genomas de plantas, são exemplos de agrupamentos de genes funcionalmente relacionados aqueles de diferentes classes de genes que codificam resistência a patógenos (genes R). Muitos genes R estão organizados em agrupamentos de vários membros derivados de uma mesma família (parálogos), ou em agrupamentos nos quais estão colocalizados vários genes R de diferentes famílias. O número de genes R em um agrupamento pode chegar a algumas dezenas, como no caso do agrupamento Rp1 do milho, que pode chegar a incluir 52 genes. Sequências gênicas e intergênicas repetidas em cada um dos agrupamentos R proporcionam a ocorrência de eventos de recombinação desigual, podendo resultar em variações no número de cópias e até na fusão de diferentes genes e eventual formação de uma variante funcional. Isso é muito interessante do ponto de vista adaptativo e explica o porquê desses genes serem mantidos em agrupamentos, apesar de poderem ser expressos independentemente.

2.5.3 Agrupamentos de genes coexpressos

Além de óperons e de agrupamentos de genes parálogos, ocorrem também em genomas de diferentes espécies eucarióticas agrupamentos de genes sem correlação funcional. As evidências disponíveis até agora sugerem que muitos genes individuais (que não fazem parte de qualquer óperon ou agrupamento de parálogos) podem não ter uma distribuição aleatória no genoma, pois estão posicionados, em geral, próximos a outros genes coexpressos, isto é, compartilhando um determinado padrão de expressão espacial ou temporal. Foi demonstrado, que em torno de 25% dos genes de *S. cerevisiae*, expressos no mesmo estágio mitótico do ciclo celular, estão agrupados no genoma, isto é, têm pelo menos um vizinho imediato no cromossomo expresso no mesmo estágio. Não há, contudo, uma correlação funcional mais evidente entre os genes nos agrupamentos identificados em cada espécie,

embora haja casos de agrupamentos de genes que codificam enzimas de uma mesma rota metabólica. A principal justificativa do agrupamento parece ser mesmo o compartilhamento de um mesmo padrão de expressão.

Independentemente da função, genes coexpressos agrupados no genoma de uma determinada espécie não estão necessariamente agrupados, mesmo quando também coexpressos, no genoma de outra espécie. Isso é um indicativo de evolução que independe dos padrões de agrupamento em cada espécie, embora os mecanismos que embasam a coexpressão devam ser comuns às diferentes espécies. Os mecanismos de regulação transcripcional associados ao fenômeno de agrupamento de genes coexpressos podem ser tanto de ordem local, envolvendo promotores bidirecionais ou reforçadores comuns, como globais, envolvendo alterações estruturais na cromatina (ver Capítulo 14).

3. Dinâmica evolutiva de genes e genomas nucleares eucarióticos

3.1 Evolução dos genomas eucarióticos

A linhagem evolutiva eucariótica descende de procariotos do domínio Archaea e surgiu há cerca de 1,5 bilhão de anos. Análises genômicas comparativas sugerem que as arqueas primordiais deram origem a eucariotos, que não pertenceriam a qualquer dos dois principais filos de arqueas atuais (Euryarchaeota e Crenarchaeota), podendo pertencer a um filo menor, inclusive já extinto. Além disso, há evidências de que outros procariotos ancestrais, ainda não bem definidos, também contribuíram para o repertório gênico eucariótico por meio de eventos posteriores de transferência gênica horizontal (ver Capítulo 4).

Os eucariotos atuais podem ser divididos, do ponto de vista evolutivo, em dois grandes grupos: o dos eucariotos unicelulares e o dos eucariotos multicelulares. A maioria dos eucariotos unicelulares tem genomas que não conseguiram ultrapassar o chamado “limiar de complexificação” que permitiria a evolução para formas multicelulares. Assim, eles ainda possuem genomas que, em certos aspectos, lembram os de procariotos. Esses eucariotos permanecem com um número pequeno de genes, com regiões intergênicas curtas, com um pequeno número de elementos transponíveis e outras sequências repetidas. Os genomas de eucariotos multicelulares, que ultrapassaram o limiar de complexificação, têm como características principais a fixação de duplicações e a proliferação de elementos transponíveis, sendo essa última a possível responsável pela maior parte do incremento de tamanho nos genomas de eucariotos mais complexos.

O papel evolutivo dos elementos transponíveis teria ido além da mera contribuição para o aumento de tamanho na evolução dos genomas eucarióticos. É possível inclusive que eventos de especiação tenham sido, diversas

vezes, desencadeados ao longo da evolução em períodos nos quais houve relaxamento no controle da mobilização desses elementos em algumas linhagens eucarióticas. Isso teria ocorrido porque a transposição disseminada, em espaços de tempo evolutivo considerados curtos, levaria ao rápido surgimento e divergência de novas espécies.

Em genomas eucarióticos, assim como nos procariontes (ver Capítulo 4), as expansões de tamanho foram acompanhadas por um aumento do nível de desorganização (entropia). Essa maior entropia é evidenciada pelo menor número de arranjos organizados de genes que podem ser observados em eucariotos. Em sua maioria, os genes ficam dispersos em meio a outras sequências, o que é mais exacerbado em genomas eucarióticos maiores, onde as próprias características de repetição e mobilidade de boa parte das regiões intergênicas favorecem rearranjos estruturais e, conseqüentemente, a dispersão gênica.

A manutenção dos genes, como unidades funcionais individualizadas, foi acompanhada pelo aumento da complexidade dos genes e seus produtos proteicos. A associação desses dois fatores viabilizou um aumento da complexidade estrutural e funcional de células e organismos eucarióticos. Genes codificando proteínas multifuncionais, sendo independentemente regulados, podem ser expressos em diferentes combinações, locais e momentos, permitindo a evolução de relações bioquímicas mais complexas e a diversificação de funções, passo fundamental para o surgimento de organismos multicelulares, com especialização de células e organização de tecidos e órgãos. Em procariontes, variações combinatórias da expressão gênica são limitadas, pelo menos até certo ponto, pela coregulação de boa parte dos genes no contexto de óperons.

Apesar das vantagens oferecidas pela regulação individual dos genes em organismos mais complexos, há ainda, em genomas eucarióticos, muitos arranjos multi-gênicos com algum tipo de coordenação funcional, inclusive óperons e agrupamentos de genes coexpressos ou com expressão coordenada. Comparações genômicas, contudo, evidenciam que, em geral, esses arranjos não são conservados por períodos evolutivos mais longos. Isso indica que os arranjos genômicos mais complexos seriam, em sua maioria, produtos primários de processos evolutivos mais ou menos neutros (independentes de seleção), como a simplificação genômica (discutidos no Capítulo 4), estando as pressões seletivas envolvidas apenas na modulação dos resultados desse processo. Os arranjos genômicos mais complexos seriam, então, relativamente efêmeros, sendo formados de forma contínua, rearranjados (podendo expandir ou contrair) e perdidos ao longo de uma linhagem evolutiva. A expansão ou a perda de um arranjo dependeria do equilíbrio entre pressões seletivas favoráveis à sua manutenção e de eventos desorganizadores aleatórios, como mutações, recombinação e atividade de elementos transponíveis.

A história evolutiva eucariótica também não pode ser dissociada da ocorrência de duplicações maiores e me-

nores, as quais tiveram grande impacto no aumento do número de genes e no surgimento de famílias de genes parálogos. Há um aumento da fração de genes em famílias (com a correspondente diminuição da fração de genes únicos) nos genomas à medida que se avança na escala evolutiva. A expansão do número de famílias e do número de genes em cada família foi um fator preponderante para a diversificação e o aumento da complexidade de estruturas e processos em eucariotos ao longo da evolução. Por exemplo, diferentes grupos de famílias tiveram maior expansão nas linhagens evolutivas de leveduras, *Drosophila* sp. e mamíferos.

Em leveduras, as famílias de maior expansão foram as de genes de resposta a estresse, metabolismo e floculação; em espécies do gênero *Drosophila*, tiveram particular expansão as famílias de genes de resposta defensiva, proteólise e resposta a estímulos químicos; e, em mamíferos, dentre as famílias de maior expansão, estão as de genes de resposta imune, desenvolvimento neuronal e transporte e comunicação intercelular. Nos roedores, em particular, ocorreram dezenas de expansões locais em famílias gênicas, envolvendo principalmente genes associados à reprodução, à imunidade e ao olfato, o que sugere que esses sistemas fisiológicos foram o foco de maior inovação evolutiva na linhagem. Seguindo essa lógica de raciocínio, pode-se, a partir de uma avaliação da dinâmica evolutiva de aquisição e expansão (por duplicação) ou perda e contração (por deleção ou pseudogenização) de famílias gênicas em diferentes linhagens evolutivas eucarióticas, identificar aspectos importantes para o surgimento, a adaptação e a diversificação de cada linhagem.

3.2 Evolução dos genes interrompidos

Como mencionado, genes interrompidos típicos, com íntrons removidos por spliceossomo, são exclusivos de eucariotos, mas tanto bactérias como arqueas possuem, embora com baixa frequência, genes interrompidos por íntrons capazes de autorremocção (ver Capítulo 11). Além disso, em genomas de eucariotos, o número de genes interrompidos e o número médio de íntrons por gene interrompido varia muito de espécie para espécie (**ver Seção 1.2 e Material Complementar Online 5.2**). A ocorrência desses tipos diferentes de íntrons e a sua distribuição heterogênea, observada em procariontes e eucariotos, tornam controversa a questão sobre a origem evolutiva dos genes interrompidos. A origem dos íntrons eucarióticos é discutida no Capítulo 11.

Independentemente da origem, acredita-se que a estrutura de genes interrompidos e o seu processamento por *splicing* tenham contribuído muito para a evolução de eucariotos. Isso aconteceu pelo surgimento de novas proteínas, mais complexas, a partir de combinações de éxons codificadores de diferentes domínios estruturais ou funcionais. O mais importante, ainda, teriam sido as alternativas adicionais de regulação da expressão gênica. Sabe-se que, pelo menos em *C. elegans*, genes interrompidos, que devem sofrer *splicing*, são expressos em níveis mais elevados que versões sem íntrons dos

mesmos genes e são inúmeros os exemplos de regulação pós-transcricional da expressão que envolvem componentes do spliceossomo. O próprio *splicing* alternativo em geral não é constitutivo, sendo as diferentes formas de mRNA maduro produzidas de maneira diferencial e controlada em diferentes tipos celulares ou tecidos, ou em diferentes momentos do desenvolvimento de um organismo (ver Capítulo 11).

4. Genomas de organelas

Mitocôndrias e **plastídeos** são organelas que apresentam genomas próprios, que diferem tanto em estrutura como em composição gênica dos genomas nucleares eucarióticos. Esses genomas são chamados de **genomas extranucleares** e contêm essencialmente um elenco de genes relacionado às funções de cada organela, centradas na síntese de ATP como fonte de energia; no caso de mitocôndrias, na fixação de dióxido de carbono atmosférico em compostos orgânicos às expensas de luz solar; no caso de **cloroplastos**, que são plastídeos fotossintetizantes e na síntese e/ou armazenamento de pigmentos, amido, lipídeos ou proteínas; e no caso de plastídeos não fotossintetizantes (**cromoplastos**, **amiloplastos**, **elaioplastos** e **proteoplastos**, respectivamente).

A presença de genomas próprios é resquício da origem evolutiva dessas organelas. Mitocôndrias e plastídeos evoluíram a partir de bactérias que foram internalizadas, inicialmente como **endossimbiontes**, em células eucarióticas ancestrais. A identidade das bactérias envolvidas pode ser inferida a partir de estudos comparativos entre genomas de organelas e genomas bacterianos. As mitocôndrias descendem de bactérias da classe α -proteobactéria (filo Proteobacteria), embora o momento da evolução eucariótica no qual isso ocorreu ainda seja controverso. A partir do reconhecimento recente de que todos os eucariotos atuais possuem mitocôndrias (ou, pelo menos, algum tipo de organela delas derivada, como **hidrogenossomos** ou **mitossomos**), acredita-se que o evento de endossimbiose, que deu origem às mitocôndrias atuais, se deu em uma célula eucariótica ancestral de todos os grupos eucarióticos atuais.

Os plastídeos, por sua vez, originaram-se de uma bactéria do filo Cyanobacteria, primeiramente de vida livre, e sua célula hospedeira inicial foi a de um eucarioto não fotossintetizante, com núcleo, citoesqueleto e mitocôndria(s). É provável que a célula hospedeira eucariótica primordial foi um ancestral comum de algas verdes, vermelhas e glaucófitas (supergrupo Plantae). Eventos de endossimbiose secundários e terciários, entre eucariotos, explicariam a disseminação atual dos plastídeos em linhagens de outros três supergrupos de eucariotos (Chromalveolata, Excavata e Rhizaria).

Do ponto de vista estrutural, os genomas de organelas são constituídos por moléculas de DNA de fita dupla, em geral circulares, chamadas, no caso das mi-

tocôndrias, de **mtDNA** e, no dos plastídeos, de **ptDNA** (denominação utilizada para qualquer tipo de plastídeos ou, às vezes, especificamente para plastídeos não fotossintetizantes), ou **cpDNA** (denominação utilizada para plastídeos fotossintetizantes). É importante salientar que o genoma é o mesmo para todos os tipos diferentes de plastídeos presentes em células de uma determinada espécie vegetal. Além do fato de serem geralmente circulares, os genomas das organelas possuem outras semelhanças com os genomas procarióticos: são pequenos, contêm genes que codificam funções relacionadas, organizados em óperons e muitas das proteínas por eles codificadas (como as proteínas ribossômicas) são mais similares às proteínas correspondentes de bactérias do que às codificadas por genomas nucleares eucarióticos. Essas semelhanças são explicadas pela origem endossimbionte dessas organelas.

O repertório gênico de genomas de mitocôndrias e plastídeos é hoje deficiente, produzindo apenas uma fração menor das proteínas necessárias à biogênese e ao funcionamento dessas organelas. Isso ocorre porque, durante a evolução tanto de mitocôndrias como de plastídeos, muitos genes foram transferidos horizontalmente, para o núcleo da célula hospedeira em eventos de **transferência gênica endossimbionte**, ou perdidos. Assim, na biogênese de novas mitocôndrias e plastídeos (que acontece a partir da divisão de organelas pré-existentes), a grande maioria das proteínas necessárias é codificada pelo genoma nuclear e importada para as organelas depois da síntese em ribossomos citosólicos.

Os processos de replicação de genomas de organelas e de expressão (transcrição e tradução) de seus genes são autônomos, isto é, ocorrem nas próprias organelas. Entretanto, apesar das proteínas que medeiam esses processos genéticos serem exclusivas das organelas, a maioria delas também é codificada pelo genoma nuclear.

Apesar das semelhanças ditadas pela origem endossimbionte das organelas, os genomas de mitocôndrias e plastídeos têm algumas características distintas, que vão além dos seus repertórios gênicos especializados. Além disso, o número de cópias, o tamanho e a composição gênica dos genomas de cada tipo de organela podem variar muito de espécie para espécie, como demonstrado pelo sequenciamento e por análises comparativas de genomas organelares. Até outubro de 2010, 2.253 genomas mitocondriais e 186 genomas de plastídeos já haviam sido completamente sequenciados. Dados atualizados sobre o sequenciamento de novos genomas organelares podem ser obtidos a partir de páginas institucionais especializadas na internet, algumas das quais são referenciadas na Seção de Leituras recomendadas ou no **Material Complementar Online 5.5**. A seguir, serão discutidas as principais características dos genomas de mitocôndrias e plastídeos.

4.1 Genomas mitocondriais

Mitocôndrias são organelas que, em geral, possuem genomas próprios, havendo até agora, apenas um caso des-

crito – o de *Cryptosporidium parvum* (Apicomplexa; Alveolata) – de mitocôndria sem genoma. Esse caso, assim como os de hidrogenossomos e de mitossomos, que são organelas relacionadas, mas caracteristicamente sem genomas, são decorrentes da degeneração dessas organelas durante a evolução.

Os genomas das mitocôndrias típicas são formados por um cromossomo circular único. Há, contudo, exceções, como genomas mitocondriais constituídos por mais de um cromossomo ou por um cromossomo linear. Por exemplo, o fungo *Spizellomyces punctatus* (Chytridiomycota) tem seu mtDNA na forma de 3 cromossomos distintos, e o fungo *P. jiroveci (carinii)* (Dikarya) tem um mtDNA linear. Normalmente, há várias cópias do genoma em cada mitocôndria, além de haver muitas mitocôndrias por célula, podendo chegar a 1.000 ou 2.000 em alguns casos, como em hepatócitos de mamíferos. Isso determina que, apesar dos genes mitocondriais serem de cópia única no genoma, todo o mtDNA seja parte da fração de DNA repetitivo da célula.

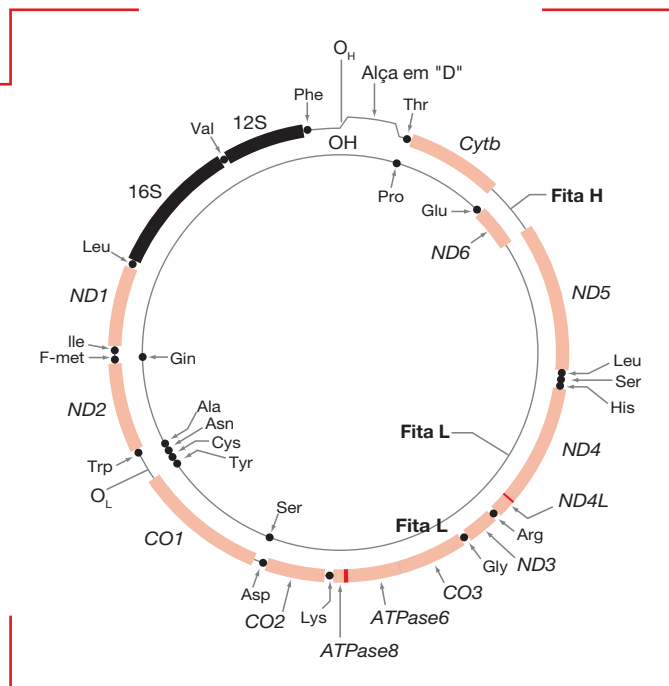
Os tamanhos dos genomas mitocondriais podem variar muito de espécie para espécie. Os maiores são encontrados em plantas angiospermas, com tamanhos que oscilam entre 220 kb a 2,9 Mb, dependendo da espécie. Entre metazoários, os genomas mitocondriais são menores, tendo tamanhos entre 11 kb (no rotífero *Brachionus plicatilis*) e 24 kb (no placozoário *Trichoplax adhaerens*). O genoma mtDNA humano, por exemplo, tem 16,6 kb (Figura 5.11). Entre fungos, os extremos de tamanho até agora conhecidos são os dos mtDNAs dos *Dikaria Hanseniaspora uvarum* (18 kb) e *Moniliophthora perniciosa* (109 kb), com a levedura *S. cerevisiae* tendo

um mtDNA de 86 kb. Os mtDNAs de leveduras e outros micro-organismos, maiores que os de quaisquer metazoários, demonstram não haver correlação direta entre o tamanho do genoma mitocondrial e a complexidade da espécie. Isso acontece porque as diferenças de tamanho dos genomas mitocondriais entre espécies muitas vezes se devem à presença ou não de íntrons nos genes e/ou às extensões das regiões intergênicas. Por exemplo, boa parte da diferença de tamanho entre os mtDNAs de *S. cerevisiae* e humano se deve à presença, no primeiro, de um número considerável de íntrons (de tipo removível por *auto-splicing*; ver Capítulo 11), que estão ausentes no segundo. Já no caso das plantas, o grande tamanho dos mtDNAs em comparação com os de animais, fungos e microrganismos é explicado, principalmente, pelo acúmulo de sequências repetidas curtas e, até, de sequências de cpDNA. Nas regiões intergênicas de genomas mitocondriais maiores também podem ser encontrados pseudogenes.

O conteúdo gênico dos genomas mitocondriais é variável, pois o número de proteínas e RNAs codificados pode ser diferente em cada espécie e não está diretamente correlacionado com o tamanho do genoma mitocondrial. Assim, enquanto o mtDNA humano, com seus 16,5 kb, codifica 13 proteínas (ver Figura 5.11), o de *S. cerevisiae*, com 86 kb, codifica 8. Além disso, estão presentes, em geral, 2 ou 3 genes de rRNA e um número bastante variável de genes de tRNA, que pode ir de 0 até 2 ou 3 dezenas, dependendo da espécie. No mtDNA humano, além de 2 rRNAs (de 12S e de 16S, o primeiro para a subunidade menor e outro para a subunidade maior de ribossomos mitocondriais) e de 13 proteínas componentes do aparato

Figura 5.11

O genoma mitocondrial humano. Cada uma das fitas do mtDNA humano circular, com 16.569 pb, codifica um conjunto diferente de genes. A chamada fita H (de *heavy* = pesada), representada pelo círculo externo, codifica 12 proteínas, 14 tRNAs e 2 rRNAs. A outra fita, chamada de fita L (de *light* = leve), representada pelo círculo interno, codifica apenas uma proteína e 8 tRNAs. Os genes que codificam proteínas estão representados em e regiões de sobreposição entre eles estão representadas em . Aproximadamente 60% da capacidade do mtDNA humano é utilizada na codificação de sete subunidades da enzima NADH-Q-redutase (genes *ND1*, 2, 3, 4, 4L, 5 e 6), o primeiro de três complexos bombeadores de prótons da membrana mitocondrial interna; são codificadas também uma subunidade da citocromo-redutase (gene *Cytb*), as três subunidades da citocromo-oxidase (genes *CO1*, 2 e 3) e as duas subunidades da ATP-sintetase (genes *ATPase6* e 8). As posições dos genes de tRNAs e dos genes que codificam os rRNAs de 16S e de 12S estão assinaladas por círculos e barras em , respectivamente. O_H e O_L designam as origens de replicação das fitas H e L. A região de formação da alça em "D", estrutura importante para o início do processo de replicação, está assinalada.



da cadeia respiratória, são codificados 22 tRNAs, o mínimo necessário para a tradução de todos os seus códons.

Um aspecto interessante da codificação de proteínas em genomas mitocondriais de muitas espécies, desde leveduras até seres humanos, é a utilização de **códigos genéticos alternativos**, distintos do nuclear (ver Capítulo 12). Isso demanda tRNAs específicos para a tradução de determinados códons nas proteínas por eles codificadas. Além disso, são considerados frequentes em mitocôndrias eventos de processamento pós-transcricional de RNAs, como a editoração de RNA e o *trans-splicing* (ver Capítulo 11).

A editoração de RNA, envolvendo mRNAs ou tRNAs, já foi descrita em mitocôndrias de plantas, mamíferos e de alguns outros organismos, como *Acanthamoeba castellanii* (Amoebozoa), *Physarum polycephalum* (Myxozoa; Amoebozoa) e tripanossomatídeos (Euglenozoa; Excavata). Os mecanismos de editoração em cada um dos casos são distintos, o que é tido como evidência de aquisição recente e independente em cada grupo taxonômico. Genes mitocondriais interrompidos também podem ser processados por *trans-splicing*. O mecanismo de *trans-splicing* envolve éxons individuais, flanqueados a montante e a jusante por sequências intrônicas, que são transcritos independentemente e depois unidos.

4.2 Genomas de plastídeos

Todos os plastídeos até hoje estudados contêm seu próprio genoma. Os genomas de plastídeos típicos são circulares e formados por um único cromossomo, cujo número de cópias por organela pode variar de algumas dezenas até pelo menos 1.000. Em plantas superiores, o número de cópias do ptDNA ou cpDNA por organela é maior em células com divisão mais ativa (que têm menos plastídeos) e menor em células maduras, que se dividem com menor frequência (e que em geral acumulam um maior número de plastídeos).

Os genomas de plastídeos são maiores e contêm mais genes que os genomas mitocondriais. Quanto ao tamanho, os genomas de plastídeos possuem, em sua maioria, entre 100 e 200 kb, mas podem variar desde pouco menos de 35 kb, como no caso de **apicoplastos**, que são plastídeos especializados de protistas do filo Apicomplexa (Chromalveolata), até mais de 420 kb, como no caso de cloroplastos da alga *Volvox carteri* (Chlorophyta; Plantae). O tamanho de genomas de plastídeos, assim como acontece no caso do mtDNA, não tem correspondência direta com a complexidade da espécie, pois boa parte das sequências de genomas, principalmente as grandes, não são gênicas. No cpDNA de *V. carteri*, por exemplo, apenas 24% das sequências correspondem a genes, e 80% das regiões intergênicas (61% de todo o genoma) são formadas por sequências repetidas curtas.

O repertório gênico dos genomas de plastídeos é dedicado à fotossíntese, a alguns outros aspectos do metabolismo dessas organelas e também à expressão gênica

(transcrição e tradução). O número de genes no ptDNA ou no cpDNA oscila entre 80 e 200, conforme a espécie. Um número em torno de 80 genes codificadores de proteínas é comum ser encontrado no cpDNA de plantas superiores (angiospermas e gimnospermas), ao passo que cpDNAs com aproximadamente 220 genes codificadores de proteínas são comuns entre algas vermelhas (Rhodophyta; Plantae). O número de proteínas codificadas, contudo, pode variar desde apenas 25, no cpDNA de *Epifagus virginiana* (angiosperma dicotiledônea), até 273, no cpDNA de *Pinus koraiensis* (gimnosperma).

Um número muito superior de genes codificadores de proteínas (867) é encontrado na organela fotossintetizante de *Paulinella chromatophora* (Cercozoa; Rhizaria), que tem um genoma de pouco mais de 1 Mb. Contudo, esse tipo de organela, chamado de **crimatóforo**, teve origem em um evento de endossimbiose primária independente e mais recente do que o que deu origem aos plastídeos típicos.

A quantidade de proteínas codificada nos genomas de plastídeos representa apenas uma fração menor do número total de proteínas de cada organela, que, no caso de cloroplastos de plantas é de 2.000 a 5.000. Essencialmente, as proteínas codificadas podem ser divididas em duas categorias. A primeira é a de proteínas relacionadas às funções específicas da organela, que, no caso de cloroplastos, inclui diferentes subunidades dos quatro complexos envolvidos na fotossíntese e do complexo da NADH-desidrogenase. A segunda é a de proteínas relacionadas à expressão gênica (transcrição e tradução) da organela, que inclui uma RNA-polimerase e, em geral, mais de duas dezenas de proteínas ribossômicas. São tipicamente codificados pelo ptDNA ou pelo cpDNA de 2 a 4 rRNAs e um repertório de 22 a 50 tRNAs.

Em genomas de plastídeos é comum a presença de genes interrompidos por íntrons. A representatividade dos íntrons em ptDNAs ou cpDNAs, em termos da extensão que ocupam ou em número de genes interrompidos, varia de espécie para espécie. O tipo dos íntrons encontrados são de procariotos (removidos por *auto-splicing*) e ocorrem tanto em genes codificadores de proteínas como em genes codificadores de tRNAs.

Ao contrário do que acontece em muitos genomas mitocondriais, os mRNAs codificados por genomas de plastídeos são traduzidos de acordo com o código genético universal. Na expressão dos genes de plastídeos, contudo, ocorrem, assim como em mitocôndrias, eventos de editoração de RNAs e de *trans-splicing*. Esses dois tipos de processamento pós-transcricional são bastante comuns em plastídeos de praticamente qualquer espécie, estando o *trans-splicing*, porém, ausente em apicoplastos de espécies do filo Apicomplexa e em plastídeos de espécies de Cercozoa (Rhizaria). Os mecanismos moleculares dos eventos de editoração de RNA e de *trans-splicing* de plastídeos são os mesmos dos eventos correspondentes que ocorrem em mitocôndrias de algas e plantas.



Resumo

Os eucariotos são organismos cujas células são compartimentalizadas em organelas. Os genomas nucleares eucarióticos são constituídos por vários cromossomos, cada um deles formado por uma única molécula de DNA linear. Genomas eucarióticos são, em geral, muito maiores que os procarióticos, mas há desde genomas com menos de 3 Mb (menores que os de diversos procariotos), até genomas com mais de 100.000 Mb de extensão, dependendo da espécie. Os genes de eucariotos são tipicamente interrompidos por íntrons. Espécies mais simples têm uma frequência menor de genes com íntrons, mas eles são predominantes em genomas de espécies mais complexas. Boa parte dos genes em um genoma eucariótico pertence a famílias de parálogos, cujo número e tamanho tendem a aumentar em espécies mais complexas.

O número de genes e o tamanho dos genomas não aumentam, necessariamente, de maneira proporcional ao aumento de complexidade das espécies eucarióticas. Isso acontece porque, em eucariotos mais complexos, a maior parte dos genes pode dar origem a vários produtos proteicos por processamento alternativo e porque a maior parte de seus genomas é constituída por sequências intergênicas. Algumas famílias

de genes parálogos podem formar agrupamentos com expressão coordenada e muitos genes não relacionados quanto a sua função, mas que são coexpressos, também se encontram agrupados.

A fração intergênica de genomas eucarióticos é, em sua maioria, formada por sequências repetidas, que podem ser sequências repetidas simples, mais curtas e com maior nível de repetição, ou ainda, elementos transponíveis, que podem ser transposons ou retrotransposons ativos ou inativos. Essas sequências repetidas podem ter funções importantes para a estruturação e o funcionamento dos cromossomos e desempenharam um papel vital como elementos geradores de variabilidade genômica ao longo da evolução dos eucariotos. Mitocôndrias e plastídeos são organelas que possuem genomas próprios, com características similares às de genomas procarióticos, devido à origem endossimbiótica. Os repertórios gênicos dos genomas de mitocôndrias e plastídeos estão associados às funções específicas de cada organela. Muitos dos genes necessários ao funcionamento de mitocôndrias ou plastídeos, contudo, estão no genoma nuclear, pois, em algum momento durante a evolução, foram transferidos horizontalmente das organelas para o núcleo.



Leituras recomendadas

Archibald JM. The puzzle of plastid evolution. *Curr Biol.* 2009;19(2):R81-8.

Demuth JP, Hahn MW. The life and death of gene families. *Bioessays.* 2009;31(1):29-39.

Embley TM, Martin W. Eukaryotic evolution, changes and challenges. *Nature.* 2006;440(7084):623-30.

Gingeras TR. Origin of phenotypes: genes and transcripts. *Genome Res.* 2007;17(6):682-90.

Jurka J, Kapitonov VV, Kohany O, Jurka MV. Repetitive sequences in complex genomes: structure and evolution. *Annu Rev Genomics Hum Genet.* 2007;8:241-59.

Kim E, Magen A, Ast G. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res.* 2007;35(1):125-31.

Koonin EV. Evolution of genome architecture. *Int J Biochem Cell Biol.* 2009;41(2): 298-306.

Michalak P. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics.* 2008;91(3):243-8.

National Center for Biotechnology Information. Entrez Genome Project: eukaryotic genomes sequencing projects [Internet]. [capturado 23 ago. 2011]. Disponível em: <http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi>.

National Center for Biotechnology Information. Organelle Genome Resources [Internet]. [capturado 23 ago. 2011]. Disponível em: <http://www.ncbi.nlm.nih.gov/genomes/GenomesHome.cgi?taxid52759&hopt5html>.

Irene Silveira Schrank

Replicação do DNA

- 1. Replicação unidirecional e bidirecional 113**
- 2. Proteínas envolvidas na replicação 113**
 - 2.1. DNA-polimerases 113
 - 2.2. Outras proteínas presentes no replissomo 115
 - 2.3. Síntese das fitas contínua e descontínua do DNA 115
 - 2.4. Estrutura das DNA-polimerases 117
- 3. Replicação no contexto da cromatina eucariótica 119**
- 4. Origem da replicação 120**
 - 4.1. Características das origens da replicação 120
 - 4.2. Ciclo celular e origem da replicação 124
- 5. Dinâmica da replicação: *Escherichia coli* como modelo 125**
- 6. Término da replicação 126**
 - 6.1. Genomas circulares 126
 - 6.2. Genomas lineares 127
 - 6.2.1. Telômeros e telomerase 128
 - 6.2.2. Sistemas iniciadores alternativos 129
- 7. Sistemas de replicação por círculo rolante 130**

O processo de divisão celular, que ocorre em todos os organismos, é normalmente descrito como uma sequência de eventos que resultam na duplicação do genoma, denominado replicação do DNA. Após a divisão celular, tanto em organismos eucarióticos como procarióticos, o mesmo material genético deverá estar presente em todas as células. Para que as células mantenham o mesmo material genético, todo o genoma deverá ser precisamente replicado antes de cada evento de divisão celular. A replicação do DNA necessita de várias atividades enzimáticas que devem agir de forma integrada e sequencial atuando de maneira coordenada nos estágios de início, alongamento (ou crescimento) da cadeia e término. O conjunto de proteínas que atuam na replicação é denominado **replissomo**.

Como princípio básico da replicação, pode-se ressaltar que ela ocorrerá apenas se existir uma molécula de DNA a ser copiada (fita-molde), com uma região pareada contendo uma extremidade 3'-OH livre onde será iniciada a síntese da nova fita de DNA. A molécula

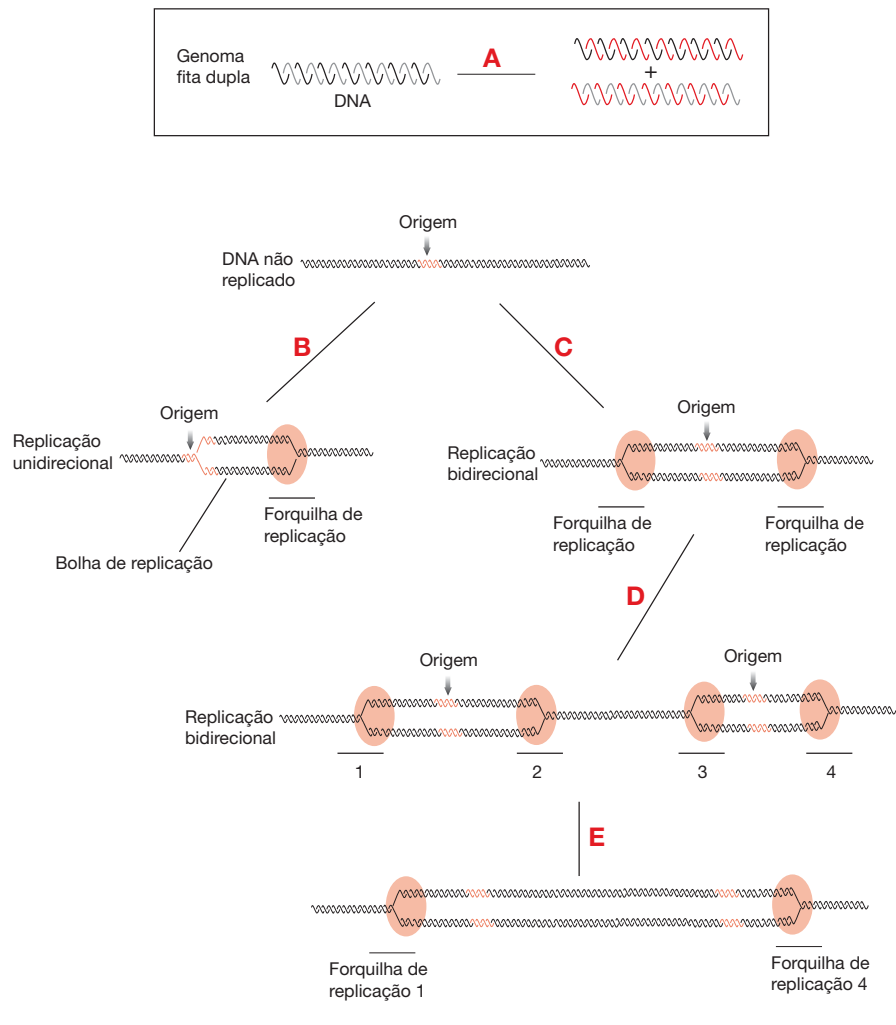
de DNA a ser replicada pode ser tanto de fita dupla (**Figura 6.1A**) como de fita simples. Para a replicação de DNA de fita simples, como no caso de alguns genomas virais, é necessária a formação de um intermediário de fita dupla antes que o processo de replicação possa ocorrer.

Além dessas características, existem algumas regras básicas para o processo de replicação que podem ser aplicadas aos genomas de todos os organismos procarióticos e eucarióticos: o processo é semiconservativo e semidescontínuo; a adição dos nucleotídeos para o crescimento da cadeia é sempre no sentido de 5'-P para 3'-OH (5'→3'); e a replicação inicia sempre em sequências específicas chamadas de origens.

Discutiremos nas seções a seguir o processo de replicação de *Escherichia coli*, modelo de organismos procarióticos bastante estudado. Serão também ressaltados aspectos característicos de outros organismos, com ênfase nas diferenças em relação ao sistema procariótico.

Figura 6.1

Replicação do DNA de fita dupla. (A) A replicação ocorre de forma semiconservativa, gerando duas novas moléculas de DNA, cada uma contendo uma fita de DNA parental (■) e uma fita de DNA recém-sintetizada (■). **(B)** Na replicação unidirecional, uma única forquilha de replicação (elipse em ■) se movimenta em somente um sentido, indicado pela seta horizontal. A região de DNA replicada, onde as duas novas fitas já foram sintetizadas (em ■), é visualizada na forma de uma bolha de replicação. **(C)** Na replicação bidirecional, duas forquilhas de replicação (elipse em ■) partem da origem e se movimentam em sentidos opostos, indicados pelas setas horizontais divergentes. **(D)** Na replicação de genomas com mais de uma origem da replicação, as origens da replicação são ativadas em diferentes momentos e as forquilhas de replicação que se movimentam a partir de cada uma delas se comportam como em **(C)**. **(E)** As forquilhas de replicação 2 e 3, que se movimentam em sentidos opostos, se encontram, finalizando com a fusão das duas bolhas de replicação nesta região.



1. Replicação unidirecional e bidirecional

O processo de replicação inicia na origem e prossegue sequencialmente ao longo da fita de DNA, em uma ou ambas as direções até o término, formando uma bolha de replicação (ver Figura 6.1). Nessa bolha de replicação encontra-se a região do DNA que já foi replicada e flanqueada por regiões onde a síntese das novas fitas ainda não foi realizada.

Na replicação unidirecional, uma forquilha de replicação (replissomo) parte da origem e segue replicando o DNA em uma só direção, ao passo que na bidirecional, duas forquilhas de replicação deixam a origem em direções opostas (Figura 6.1B e C). No caso de genomas contendo múltiplas origens de replicação, duas forquilhas de replicação se movimentando em orientações opostas podem se encontrar permitindo a fusão de duas bolhas de replicação (Figura 6.1D).

Em 1958, os experimentos de Meselson e Stahl demonstraram que a replicação de DNA ocorre de forma **semiconservativa**, sendo cada fita de DNA copiada no sentido 5'→3', originando novas fitas. Nesta forma de replicação, as duas fitas de DNA parental devem ser separadas e cada uma servirá de molde para a síntese das novas fitas de DNA. Este processo é considerado semiconservativo, pois as moléculas de DNA geradas, após a replicação (ver Figura 6.1A), contêm somente uma das fitas recém-

-sintetizada e a outra corresponde à fita do DNA parental (molde). Durante a replicação a manutenção da estrutura básica do DNA deve ser respeitada. Como discutido no Capítulo 2, a molécula de DNA de fita dupla é mantida por meio do pareamento de bases complementares e da polaridade inversa, sendo uma fita orientada no sentido 5'→3' e a outra no sentido 3'→5' (antiparalelismo). Na **Figura 6.2** está representada, de forma esquemática, a síntese das duas fitas de DNA a partir de um molde. Cada fita nova sintetizada é complementar (pareamento AT e GC) à fita-molde e apresenta conformação antiparalela.

A conformação antiparalela do DNA e a síntese no sentido 5'→3' apresentam um desafio para a maquinaria de replicação, pois cada replissomo se movimenta em um sentido ao longo das duas fitas de DNA. Estes problemas foram superados pela dinâmica do complexo da holoenzima que realiza a replicação (ver Seção 5).

2. Proteínas envolvidas na replicação

2.1. DNA-polimerases

As **DNA-polimerases** são enzimas necessárias, em todos os organismos, tanto para a replicação do DNA como

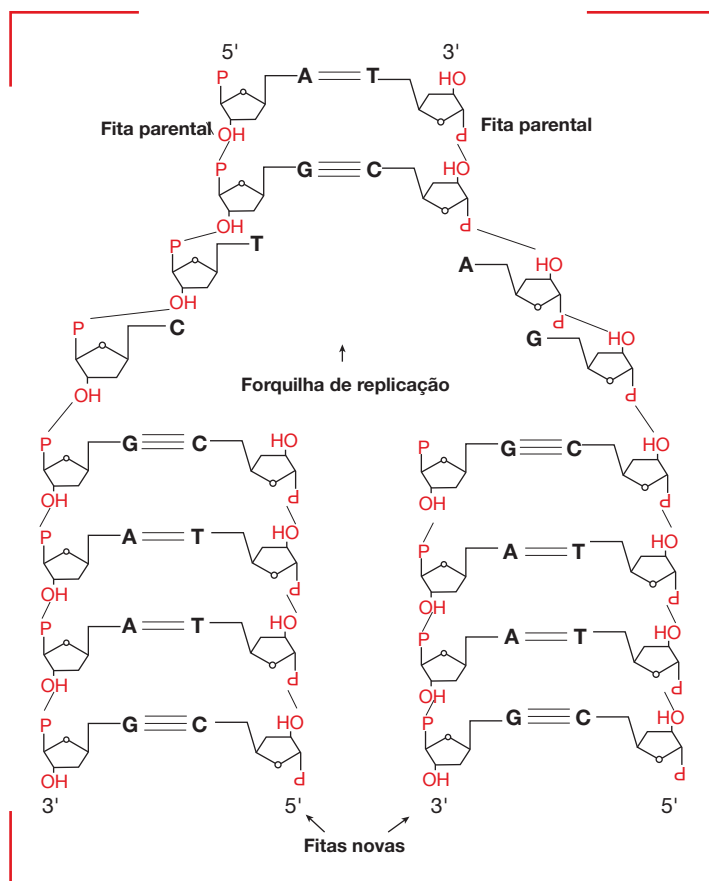


Figura 6.2

Princípios básicos da replicação: antiparalelismo e complementaridade de bases. As duas fitas novas que estão sendo sintetizadas apresentam complementaridade de bases e antiparalelismo quando comparadas com as fitas-molde.

para os processos de reparação do DNA. No entanto, diferentes processos metabólicos são realizados por DNA-polimerases específicas, sendo algumas delas proteínas multifuncionais. A presença de várias DNA-polimerases foi observada em organismos procarióticos e eucarióticos. No entanto, somente algumas estão envolvidas com o processo de replicação, sendo a denominação **DNA-replicase** utilizada para a DNA-polimerase que realiza a síntese das duas fitas novas de DNA.

As DNA-polimerases de todos os organismos realizam a mesma reação enzimática, que consiste na incorporação de nucleotídeos trifosfatados na extremidade 3'-OH livre, sintetizando a fita no sentido 5'→3' (**Figura 6.3**). Portanto, para que a DNA-polimerase possa realizar a incorporação de bases no DNA é necessária a síntese prévia de pequenas sequências de ribonucleotídeos, denominados **iniciadores** (*primers*), pela proteína primase (ver Seção 2.2). A seleção do nucleotídeo a ser adicionado depende exclusivamente da complementaridade de bases com a fita-molde. Assim, a ligação fosfodiéster é realizada entre a extremidade 5'-P do nucleotídeo incorporado com a extremidade 3'-OH do nucleotídeo já pareado.

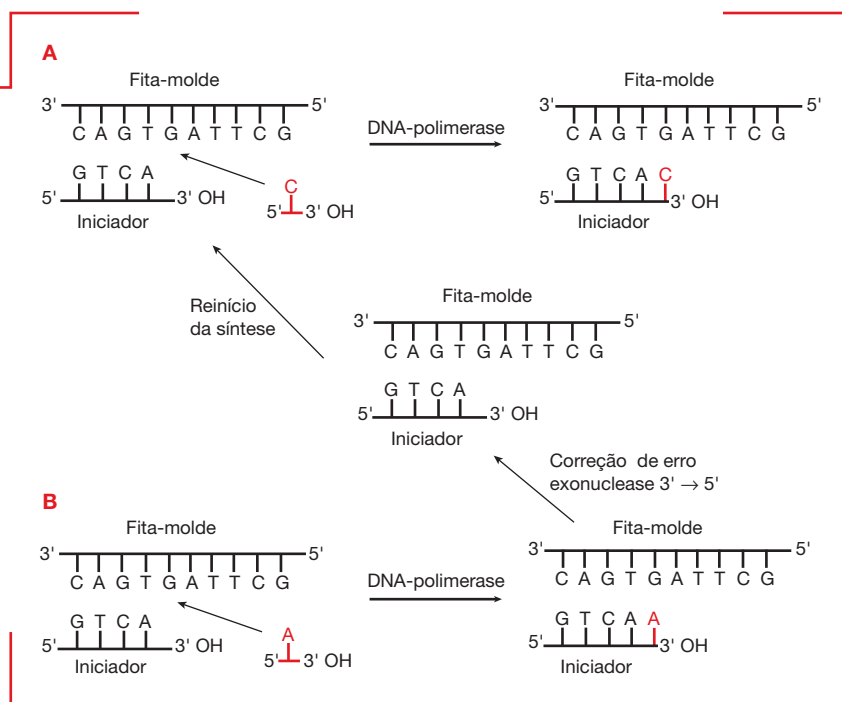
Os estudos já realizados identificaram cinco DNA-polimerases em *E. coli*, sendo denominadas DNA-polimerases I, II, III, IV e V. As DNA-polimerases de diversos organismos foram classificadas em seis famílias, com base na similaridade das sequências de aminoácidos. As DNA-polimerases I (família A) e III (família C) são enzimas que participam do processo de replicação, ao passo que as DNA-polimerases II (família X), IV (família Y) e V (família Y) estão relacionadas aos eventos de reparação de lesões do DNA (**Material Complementar Online 6.1**). Neste capítulo, serão analisadas, em detalhes, somente as DNA-

-polimerases I e III, por estarem diretamente envolvidas com a replicação.

Todas as DNA-polimerases bacterianas, além da atividade de **polimerização** (síntese do DNA no sentido 5'→3'), apresentam uma atividade de **exonuclease** (remoção de nucleotídeos a partir da extremidade da molécula de DNA). Nas DNA-replicases existe uma atividade de exonuclease no sentido de 3'-OH para 5'-P (3'→5'), que tem a função de remover nucleotídeos adicionados incorretamente na fita nova do DNA. Este mecanismo, ilustrado na Figura 6.3B, é chamado de “correção de erro”, pois permite que a própria DNA-polimerase, que está realizando a síntese, remova o nucleotídeo com pareamento incorreto. A atividade de exonuclease 3'→5' é muito importante para que o processo de replicação ocorra livre de erros, sendo a principal função da atividade de exonuclease 3'→5' reconhecer e clivar um pareamento incorreto na extremidade 3'-OH da fita nova. Portanto, na replicação do genoma de *E. coli*, por exemplo, o processo de correção de erro pode ser detalhado da seguinte forma: (1) um nucleotídeo é adicionado na cadeia nova crescente pela atividade de polimerização da DNA-polimerase III, formando o pareamento das bases; (2) a enzima faz o movimento de uma base para adicionar o próximo nucleotídeo de um novo pareamento; (3) sendo o pareamento de bases anterior (1) incorreto, a enzima fica impedida de continuar a polimerização e, movendo um nucleotídeo para trás, coloca-se em contato com o nucleotídeo que deve ser removido com a atividade de exonuclease 3'→5'; e (4) o nucleotídeo com pareamento de bases incorreto é removido e um novo nucleotídeo pode ser adicionado na região onde ocorreu a correção de erro (ver Figura 6.3B).

Figura 6.3

Síntese de uma nova fita de DNA.
(A) Adição do nucleotídeo trifosfatado (■) na extremidade 3'-OH da cadeia nascente de DNA (iniciador). A formação da ligação fosfodiéster é catalisada pela DNA-polimerase, com liberação de pirofosfato (PPi) e o correto pareamento das bases (CG). A seguir, o próximo nucleotídeo (T) é incorporado. **(B)** Representação esquemática da atividade de correção de erro realizada pela DNA-polimerase. Quando ocorre a incorporação de um nucleotídeo incorreto (não complementar, representado pelo nucleotídeo A em ■), a DNA-polimerase o remove pela sua atividade de exonuclease 3'→5'. A correção do erro permite então que a síntese seja reiniciada.



Outra atividade de exonuclease ($5' \rightarrow 3'$), presente em algumas DNA-polimerases, remove blocos de nucleotídeos (aproximadamente 10 nucleotídeos) de uma das fitas no sentido oposto ao da exonuclease de correção de erro. Portanto, se no DNA ocorre uma quebra de ligação fosfodiéster (*nick*), ou há falta de alguns nucleotídeos (*gap*), os seguintes eventos podem acontecer: (1) a atividade de exonuclease $5' \rightarrow 3'$ remove blocos de nucleotídeos no sentido $5' \rightarrow 3'$, a partir da região alterada no DNA; e (2) a atividade de polimerização sintetiza as regiões que foram removidas, originando uma fita nova de DNA nessa região.

O mecanismo de replicação em organismos eucarióticos é muito similar ao descrito para *E. coli*. Uma das principais diferenças está relacionada à presença de um maior número de DNA-polimerases envolvidas com a replicação. A DNA-polimerase α é uma proteína complexa contendo além da atividade de polimerização uma atividade de primase (ver Seção 2.2). Portanto, existem evidências indicando a função dessa polimerase na síntese dos iniciadores nas duas novas fitas de DNA (fita contínua e fita descontínua; ver Seção 2.3). O prosseguimento da síntese da fita contínua seria de responsabilidade da DNA-polimerase δ , porém, ainda é controverso o tipo de DNA-polimerase que realizaria a síntese da fita descontínua. Evidências sugerem a presença da DNA-polimerase ϵ na síntese da fita descontínua, mas a DNA-polimerase δ também poderia ser a enzima responsável por essa síntese. As DNA-polimerases δ e ϵ pertencem à família B, mas a DNA-polimerase α é um membro da família C, que até o momento é considerada uma família exclusiva de bactérias.

2.2 Outras proteínas presentes no replissomo

Para que ocorra o processo de replicação do DNA é necessária a presença, no replissomo, de várias proteínas com funções específicas, dentre elas DNA-polimerases, helicases, proteínas SSB, primases, ligase e topoisomerase. Algumas dessas enzimas já foram descritas, como as topoisomerases (Capítulo 2) e as DNA-polimerases (Seção 2.1).

A forquilha de replicação deve conter uma **helicase**, com função de quebrar as pontes de hidrogênio entre as bases, separando as duas fitas de DNA. Essa abertura das fitas é necessária para que a forquilha de replicação possa se movimentar e, portanto, todas as DNA-polimerases replicando DNA de fita dupla dependem da ação inicial de uma helicase (**Figura 6.4**). A helicase DnaB se encontra no replissomo de *E. coli*, apresenta-se como um hexâmero (forma de anel) e utiliza principalmente a energia do ATP para romper as ligações (pontes de hidrogênio) entre as fitas de DNA. Essa enzima realiza um movimento ao longo da fita do DNA no sentido $5' \rightarrow 3'$ (Figura 6.4B). Existem, em procariotos, outras helicases participando de diferentes processos celulares, como reparação, recombinação, transferência gênica (conjugação) e trans-

crição. Em arqueas e eucariotos, a helicase presente no replissomo, também em estrutura de anel com seis subunidades, é denominada de MCM (de *minichromosome maintenance* = manutenção de minicromossomo).

Outro grupo de proteínas presentes no replissomo são as proteínas que se ligam à fita simples do DNA ou proteínas **SSB** (de *single strand binding* = ligação à fita simples). Essas proteínas têm alta afinidade pelo DNA na forma de fita simples, ocorrendo ligação sempre de forma cooperativa (Figura 6.4A). A presença das proteínas SSB durante a replicação é muito importante, pois, ao se ligarem às regiões de fita simples do DNA, elas evitam que aquela região sofra torções, induzindo uma conformação do DNA ideal para a replicação e o pareamento de bases, além de protegerem as fitas simples da degradação por nucleases.

Como mencionado anteriormente, as DNA-polimerases não possuem capacidade para iniciar uma cadeia de nucleotídeos, necessitando de uma região pareada (fita dupla) com uma extremidade $3'$ -OH livre. Portanto, para cada início de síntese de uma das fitas de DNA, se faz necessária a presença de um iniciador. Os iniciadores utilizados na replicação dos genomas celulares são pequenas sequências de RNA sintetizados por uma **primase** (ver Figura 6.4A).

Em *E. coli*, foi caracterizado o produto do gene *dnaG* como sendo uma primase. A primase de *E. coli* é composta por um polipeptídeo de 60 kDa, presente nas células em número de 50-100 cópias/célula. O mecanismo de ação dessa primase, na origem de replicação, ocorre pela síntese de uma sequência de 11 a 12 nucleotídeos (iniciador), complementar àquela presente na fita-molde de DNA, na presença de ribonucleosídeos trifosfatados. Em geral, a primase de *E. coli* está associada com a atividade de DnaB e outras proteínas, fazendo parte do replissomo na forquilha de replicação. Tanto a primase como a DnaB são enzimas necessárias durante todo o processo de replicação do DNA.

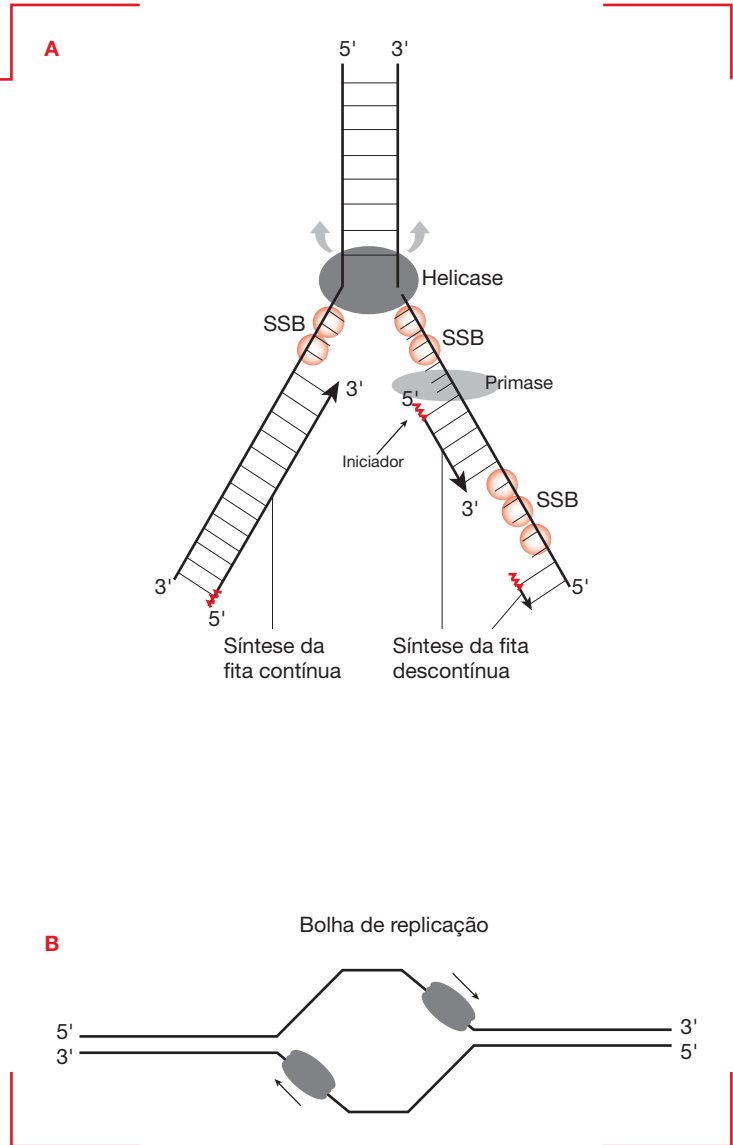
Além da primase codificada pelo gene *dnaG*, outras primases já foram caracterizadas, como as codificadas pelos genomas dos bacteriófagos T7 e T4, entre outros. Em organismos eucarióticos, a atividade da primase está principalmente presente na DNA-polimerase α .

2.3 Síntese das fitas contínua e descontínua do DNA

Durante o processo de replicação do DNA, dois fatores devem ser considerados: as fitas do DNA têm polaridades opostas ($5' \rightarrow 3'$ e $3' \rightarrow 5'$) e as DNA-polimerases sintetizam apenas no sentido $5' \rightarrow 3'$. Dessa forma, as novas fitas de DNA crescem em sentidos opostos, sendo uma das fitas sintetizada de forma contínua e a outra de forma descontínua (**Figura 6.5**). Portanto, o processo de replicação é considerado **semidescontínuo**. Considerando esse modelo de replicação, a DNA-replicase necessitaria somente um iniciador inicial para a síntese da fita contínua. No en-

Figura 6.4

Forquilha de replicação. (A) Demonstração da ação coordenada das proteínas helicase, primase e SSB durante a replicação. As setas em representam a abertura das fitas-molde e o movimento da forquilha de replicação. E os iniciadores da fita contínua e descontínua estão em . **(B)** Representação da ação da helicase durante a replicação bidirecional. Um hexâmero da helicase em envolve uma das fitas de DNA (5'→3') para realizar a quebra das pontes de hidrogênio entre as bases. As setas representam a direção do movimento das helicases na forquilha de replicação.



tanto, diversas evidências já demonstraram que a síntese da fita contínua pode ser interrompida por diferentes fatores. Para que a replicação reinicie, a cada interrupção, é necessária a síntese de iniciadores adicionais pela primase.

A síntese da fita descontínua, por sua vez, é realizada a partir de vários iniciadores, sendo polimerizada em diversos fragmentos (fragmentos de Okazaki), que possuem entre 1.000 e 2.000 nucleotídeos em organismos procarióticos e entre 100 a 200 nucleotídeos em eucariotos. A síntese de cada fragmento de Okazaki prossegue até encontrar a região 5'-P do fragmento anterior, quando então a DNA-polimerase se dissocia da fita descontínua do DNA. Durante este processo, a fita-molde da fita descontínua parece formar uma alça, resultando na coorientação da síntese da fita descontínua com o movimento do replissomo (ver Seção 5).

Um outro aspecto fundamental na replicação do DNA é a coordenação da síntese das duas fitas novas, a

contínua e a descontínua. Essa coordenação, que demanda uma série de reações enzimáticas sequenciais, sincroniza a síntese dos iniciadores nas duas fitas, a reciclagem da ligação da DNA-polimerase na fita descontínua e a produção dos **fragmentos de Okazaki**.

Após a abertura das fitas do DNA na origem, o movimento da forquilha de replicação ocorre no mesmo sentido em que a enzima helicase realiza a separação das fitas. Durante este processo, dois fatores são importantes e merecem ser ressaltados. A velocidade de síntese dos iniciadores pela primase é muito menor do que a velocidade de polimerização realizada pela DNA-polimerase III na forquilha de replicação, e a reciclagem da fita descontínua pela DNA-polimerase III, também é mais lenta se comparada à velocidade de polimerização dos nucleotídeos. Esses eventos sugerem que, antes da formação da alça, a primase pode atuar como um “freio” e causar uma parada temporária no

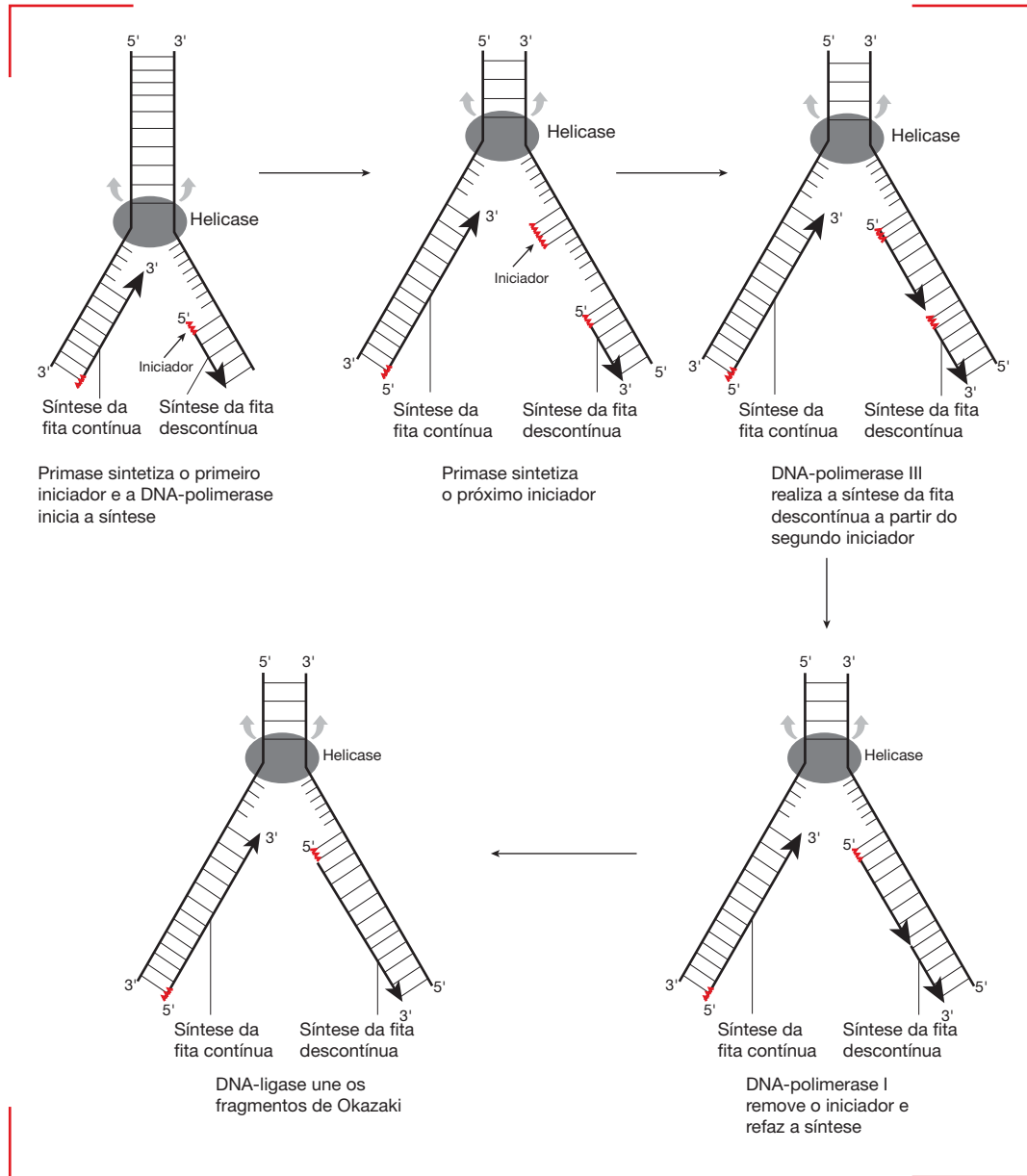


Figura 6.5

Síntese das fitas contínua e descontínua durante a replicação do DNA. A fita descontínua é sintetizada de forma interrompida, formando os chamados fragmentos de Okazaki. Os iniciadores estão em ■.

movimento do replissomo. Este mecanismo de ação seria o responsável por impedir que o crescimento da fita contínua ultrapassasse excessivamente a síntese da fita descontínua. A presença do complexo da holoenzima, que utiliza duas DNA-polimerases, em eucariotos, ou uma DNA-polimerase com dois núcleos catalíticos (DNA-polimerase III), em procariotos, para sintetizar as duas fitas novas de DNA também apresenta uma maneira de superar os problemas da síntese coordenada da fita contínua e descontínua do DNA.

2.4 Estrutura das DNA-polimerases

As DNA-polimerases caracterizadas até o momento apresentam uma estrutura conservada, contendo diferentes domínios de atividades. A enzima DNA-polimerase I de *E. coli* foi a primeira DNA-polimerase caracterizada por Arthur Kornberg, em 1956. Ela é formada por uma única cadeia polipeptídica, contendo 928 aminoácidos, e participa dos processos celulares de replicação e reparação do DNA. A DNA-polimerase I apresenta três domínios de atividade, sendo que na região carboxiterminal está localizada a atividade

de polimerização e, adjacente a ela, está a atividade de exonuclease 3'→5'. A região aminoterminal contém a terceira atividade, que é a de exonuclease 5'→3' (**Figura 6.6A**). Essa cadeia polipeptídica pode ser clivada por proteases, originando um fragmento pequeno na região aminoterminal (contendo a atividade de exonuclease 5'→3') e um fragmento grande ou **fragmento de Klenow**, com as duas outras atividades. O fragmento de Klenow pode ser utilizado *in vitro* para síntese de DNA e remoção de nucleotídeos pela atividade de exonuclease 3'→5'.

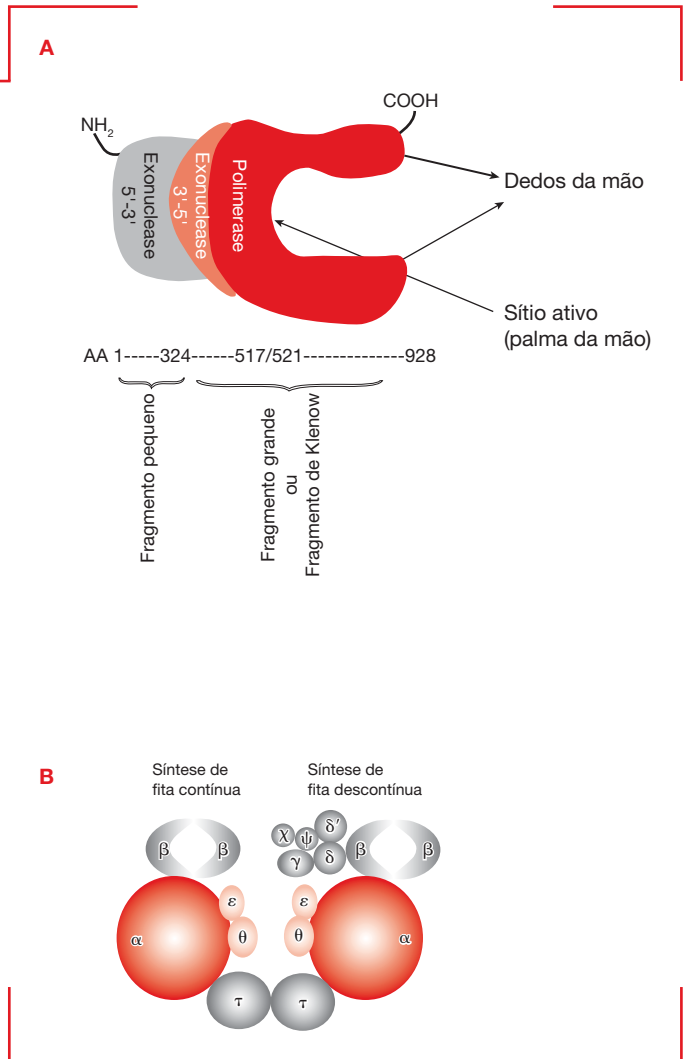
A definição da estrutura das DNA-polimerases, pelas técnicas de cristalografia, demonstrou uma organização que pode ser comparada por analogia com a mão humana. Na palma da mão, estaria localizado o sítio ativo (atividade de polimerização) e os dedos seriam responsáveis pelo correto posicionamento do DNA molde no sítio ativo (ver Figura 6.6A). A região equivalente ao sítio ativo (palma da mão) apresenta conservação (ou seja, semelhança na sequência de aminoácidos entre as proteínas correspondentes em diferentes espécies) nas diferentes famílias de DNA-polimerases.

A DNA-polimerase III é composta por várias cadeias polipeptídicas, por isso é definida como uma holoenzima. Ela é responsável pelo processo de replicação (DNA-replicase) do cromossomo de *E. coli*. A estrutura dessa DNA-polimerase (Figura 6.6B) apresenta dois núcleos catalíticos heterotriméricos, cada um deles formado pelas subunidades α , ϵ e θ . Na subunidade α , está localizada a atividade de polimerização e, na subunidade ϵ , a atividade de correção de erro. A função da subunidade θ ainda não está elucidada por completo, mas ela parece estimular a atividade de exonuclease. Como já discutido, a subunidade α da DNA-polimerase III pertence à família C (específica de bactérias) de DNA-polimerases e não apresenta similaridade de sequência de aminoácidos com outras DNA-polimerases. A estrutura dimérica dos dois núcleos catalíticos é estabilizada pela presença de duas subunidades τ .

Associada a cada núcleo catalítico encontra-se a subunidade β , que é responsável por recrutar o complexo DNA-polimerase $\alpha\epsilon\theta$, pela sua capacidade de se envolver ao DNA. A subunidade β apresenta-se como um homo-

Figura 6.6

Representação esquemática da estrutura das DNA-polimerases I e III de *E. coli*. (A) Localização das atividades de polimerização e de correção de erro da DNA-polimerase I. Os números representam as posições dos aminoácidos da DNA-polimerase I e as setas indicam, na estrutura, as regiões análogas a uma “mão humana”. (B) Estrutura esquemática da holoenzima DNA-polimerase III. As letras gregas representam a denominação de cada subunidade. A atividade de polimerização está presente na subunidade alfa (α) e a de correção de erro na subunidade épsilon (ϵ). Acima de cada metade da holoenzima está indicada sua função na síntese da fita contínua ou da fita descontínua do DNA.



dímero em forma de anel que fixa a DNA-polimerase III ao DNA, garantindo sua **processividade** (capacidade de permanecer ligada ao DNA por um período de tempo maior). Na estrutura final da DNA-polimerase III, cada núcleo catalítico é responsável pela síntese de uma das fitas novas do DNA. Essa enzima apresenta uma estrutura assimétrica, pois apenas um “núcleo catalítico” associa-se a um grupo de subunidades, denominadas complexo γ . O complexo γ é um pentâmero formado por diferentes subunidades, cuja função seria a de permitir a ligação e o desligamento da subunidade β do DNA.

Como já discutido, a replicação em organismos eucarióticos é realizada por três diferentes DNA-polimerases. Essas enzimas são formadas por várias subunidades e seu tamanho muito grande vem dificultando a resolução de suas estruturas. A conservação, tanto filogenética como de sequência, tem sido demonstrada na subunidade catalítica entre as três DNA-polimerases (α , δ , ϵ) de diferentes organismos eucarióticos. A análise das outras subunidades, presentes no multicomplexo proteico, demonstra que somente a subunidade “B” é compartilhada entre as três DNA-polimerases, e a subunidade catalítica também é similarmente conservada em diferentes organismos eucarióticos. A subunidade “B” foi, também, encontrada nas DNA-polimerases de arqueas como a única cadeia polipeptídica adicional da replicase. A subunidade catalítica e a subunidade “B” interagem para formar o núcleo destas DNA-polimerases, sendo a função da subunidade “B” estabelecer interações múltiplas e simultâneas com as diversas subunidades do complexo das replicases.

As DNA-polimerases eucarióticas também apresentam subunidades com funções equivalentes às da subunidade β e do complexo γ da DNA-polimerase III de *E. coli*. Estudos estruturais do PCNA (de *proliferating cell nuclear antigen* = antígeno nuclear de células proliferativas), equivalente à subunidade β , revelam que esse componente apresenta a mesma estrutura em anel, com o seu espaço interno permitindo o encaixe de uma molécula de DNA de fita dupla. Similaridades tanto estruturais como funcionais são também observadas entre a proteína RFC (de *replication factor C* = fator C de replicação) de eucarióticos com o complexo γ da DNA-polimerase III.

A **Tabela 6.1** apresenta um resumo comparativo entre os componentes do replissomo conhecidos nos três domínios da vida (Bacteria, Archaea e Eukarya).

3. Replicação no contexto da cromatina eucariótica

Nesta seção, será analisada a relação da cromatina com o processo de replicação, pois a replicação está inter-relacionada com a organização local da cromatina, que permite o acesso de fatores reguladores específicos capazes de interagir com o DNA. Será utilizada a cromatina eucariótica como modelo, pois existem maiores informações sobre sua estrutura e função, se comparada com a cromatina procariótica.

Em eucariotos, durante a fase S, o DNA e a cromatina devem ser duplicados, e a cromatina deve sofrer alterações para permitir a passagem do replissomo. Para a síntese das novas fitas, os nucleossomos do DNA parental devem ser transitoriamente alterados, sendo novamente estruturados nas duas novas moléculas de DNA sintetizadas.

O octâmero do nucleossomo é formado por dois dímeros das histonas H3 e H4, que representam o núcleo central, sendo flanqueados por dois dímeros compostos pelas histonas H2A, H2B e H1 (ver Capítulo 3). As histonas H3, H4, H2A e H2B podem estar presentes na célula na forma de dímeros associados a **chaperonas** (proteínas responsáveis pelo correto enovelamento de outras proteínas). Logo após a passagem do replissomo ocorre a associação do tetrâmero de H3 e H4 ao DNA, seguida da associação dos dímeros de H2A e H2B para a formação do nucleossomo (**Figura 6.7**). Portanto, na dinâmica de formação de novos nucleossomos, podem participar tanto dímeros parentais (originalmente presentes no DNA parental) como novos dímeros. Durante o processo de replicação ocorre a síntese de novas histonas (H1, H2, H3 e H4). Existem evidências sugerindo que os tetrâmeros de H3 e H4, formados com as histonas recém-sintetizadas (novas), se associam somente ao

Tabela 6.1 Composição do replissomo nos três domínios da vida

Função	Bacteria	Archaea	Eukarya
Helicase	DnaB	MCM	MCM
Proteínas que estabilizam a fita simples de DNA	SSB	RPA	RPA
Síntese do iniciador (primase)	DnaG	Pri	DNA-polimerase α /primases
Replicação	DNA-polimerase III	DNA-polimerase B	DNA-polimerase δ DNA-polimerase ϵ
Processividade	Subunidade β da DNA-polimerase III	PCNA	PCNA
Ligação da DNA-polimerase ao DNA na fita descontínua (complexo acessório)	Complexo γ da DNA-polimerase III	RFC	RFC

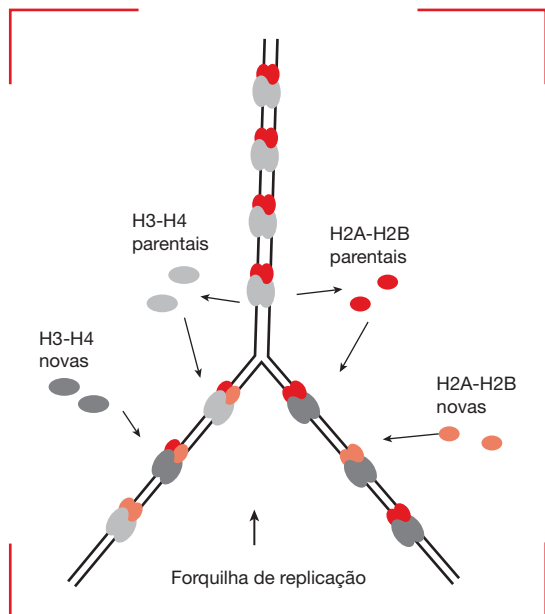


Figura 6.7

Montagem dos nucleossomos durante a replicação em eucariotos. Os dímeros H2A-H2B parentais e novos podem se misturar na formação dos nucleossomos, tanto nas fitas novas sintetizadas como nas parentais. No entanto, os tetrâmeros (H3)₂-(H4)₂ parecem ser formados somente por histonas H3 e H4 parentais ou por novas histonas. Na figura, está apresentada a distribuição aleatória dos tetrâmeros (H3)₂-(H4)₂ e dos dímeros H2A-H2B durante a replicação.

DNA já replicado. No entanto, as outras histonas (H2A, H2B ou H1) podem se associar tanto ao DNA não replicado como ao já replicado.

A estruturação dos nucleossomos no DNA, durante a replicação, necessita de uma chaperona auxiliar denominada CAF-1 (de *chromatin assembly factor* = fator de montagem da cromatina), formada por aproximadamente 5 subunidades responsáveis pela montagem do tetrâmero H3-H4 no DNA. Outra chaperona, denominada proteína de montagem de nucleossomo 1 (Nap1, de *nucleosome assembly protein 1*), parece ser responsável pela montagem no DNA dos dímeros de H2A e H2B. Finalmente, os novos nucleossomos devem sofrer uma maturação, iniciando com a ligação da histona H1; seguido da distribuição no DNA para adquirirem o espaçamento característico (com a participação de chaperonas e do fator de montagem e remodelação de cromatina); e, posteriormente, ocorre a modificação das histonas H4 (desacetilada) e H2A (fosforilada). Na forquilha de replicação, o PCNA recruta CAF-1, estabelecendo desta forma uma conexão entre o movimento do replissomo e a estruturação dos nucleossomos.

4. Origem da replicação

A replicação ao longo de uma molécula de DNA ocorre a partir de um ponto inicial denominado **origem da replicação**. Origens da replicação já foram isoladas em diversas espécies, apresentando a característica comum de serem formadas por sequências ricas em nucleotídeos AT. As principais diferenças entre os organismos são relativas ao número de origens de replicação no seu genoma. O DNA de células bacterianas, de vírus ou de plasmídeos contém somente uma origem de replicação, ao passo que múltiplas origens são encontradas em genomas de eucariotos e de algumas espécies de arqueas (ver Capítulo 4).

Em todos os organismos o início da replicação é caracterizado pelo reconhecimento da origem por proteínas específicas, que posteriormente recrutam proteínas adicionais formando então o replissomo.

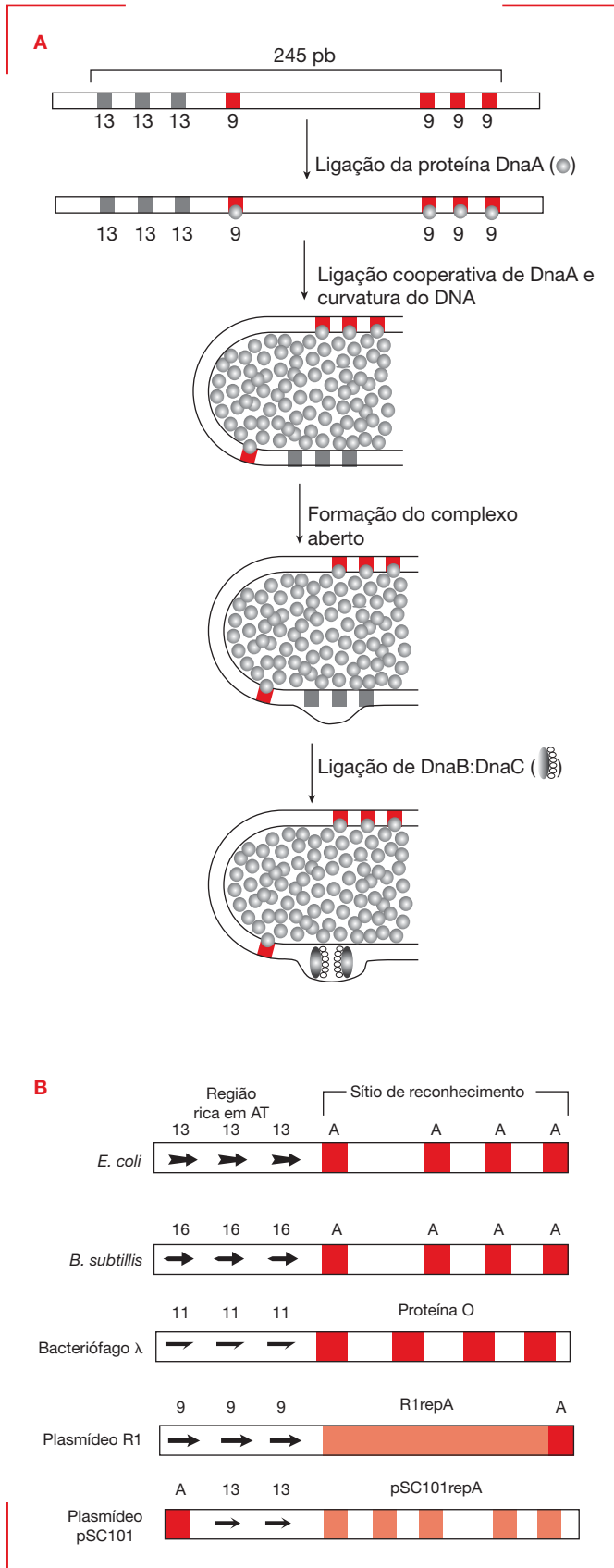
4.1 Características das origens da replicação

O isolamento e a caracterização da origem da replicação (*oriC*) do genoma de *E. coli* proporcionou um entendimento dos mecanismos que possibilitam o início da replicação e serviu de base para a comparação da estrutura das origens da replicação em diferentes espécies. Uma região de 245 pb do cromossomo de *E. coli* foi previamente determinada como essencial à replicação. Dentro dessa região, foram caracterizados dois grupos de sequências envolvidos no reconhecimento da origem da replicação: (1) quatro blocos contendo sequências repetidas de 9 pb, cujo consenso é 5'TTAT(C/A)CA(C/A)A3'; e (2) três regiões com sequências repetidas de 13 pb ricas em nucleotídeos A ou T, localizadas adjacentes às sequências acima (**Figura 6.8**).

As sequências de 9 pb são os sítios-específicos de reconhecimento da proteína iniciadora **DnaA**. A DnaA é um monômero de 52 kDa que possui a capacidade de ligar-se especificamente a cada região repetida de 9 pb. Sua ligação à *oriC* é cooperativa, originando a ligação de 20-40 monômeros nessa região da *oriC* e formando uma região central onde o DNA se enrola. A DnaA possui três funções durante o processo de início da replicação: (1) reconhecimento da sequência nucleotídica da origem, determinando o lugar onde deve iniciar a replicação; (2) indução da abertura das duas fitas de DNA, por meio da sua ligação cooperativa; e (3) possibilitar a ligação da helicase DnaB na *oriC*.

O processo de início da replicação na *oriC* pode ser dividido em três estágios (Figura 6.8A):

1. Formação do complexo inicial: neste estágio, ocorre a ligação dos monômeros da proteína DnaA nos quatro sítios de 9 pb, e, posteriormente, a ligação cooperativa de 20-40 monômeros, favorecendo a alteração na estrutura do DNA nesta região;

**Figura 6.8**

Mecanismo de reconhecimento da origem da replicação em bactérias. (A) Apresenta as etapas de ligação da proteína DnaA nas regiões repetidas de 9 nucleotídeos (9), que representam os sítios de ligação dessa proteína ao DNA no genoma de *E. coli*. Adjacentes a estes sítios encontram-se sequências de 13 nucleotídeos (13) ricas em bases AT, onde ocorre a abertura das fitas para ligação da helicase e o início da replicação. (B) Esquema comparativo de origens da replicação de diferentes DNAs (genômicos, virais ou plasmidial). Os retângulos coloridos representam a localização das sequências que constituem o sítio de reconhecimento das proteínas iniciadoras. As setas representam as sequências de nucleotídeos ricas em AT e a letra A representa o sítio de ligação da proteína DnaA de *E. coli*.

2. Formação do complexo aberto: a formação do complexo DnaA:oriC produz alterações estruturais no DNA, resultando na separação das duas fitas na região das sequências de 13 pb. Neste estágio, estão envolvidas proteínas de função estrutural na cromatina bacteriana, como HU ou IHF (ver Capítulo 3), que, provavelmente, estimulam a reação por estabilizarem a curvatura do DNA;
3. Formação do complexo pré-priming: a formação da bolha da replicação, pela separação das fitas, permite o posicionamento da helicase DnaB. A DnaA recruta a DnaB na forma de um complexo DnaB-DnaC, que se posiciona para dar início a cada forquilha de replicação. O complexo DnaB-DnaC tem a função de transferir a DnaB para a fita de DNA correta, utilizando a hidrólise do ATP.

A ligação de DnaB na *oriC* representa o início da replicação em resposta à ligação de DnaA, pois recruta a primase para dar início à formação do replissomo.

As sequências de DNA que constituem a origem da replicação de outras espécies procarióticas apresentam características semelhantes à da *oriC* de *E. coli*. Atualmente, com o sequenciamento de diversos genomas de espécies procarióticas (ver Capítulo 4), foi possível observar que as regiões de reconhecimento da proteína iniciadora, localizadas próximas a sequências ricas em AT, estão presentes em outras bactérias, demonstrando a conservação desse sistema de reconhecimento da origem.

No genoma do bacteriófago λ , também foram encontradas na região da origem da replicação quatro sequências repetidas, reconhecidas como sítios de ligação da proteína iniciadora (proteína O) codificada pelo fago, próximas a três regiões ricas em AT (Figura 6.8B). Alguns plasmídeos de *E. coli* apresentam, em suas origens da replicação, sequências reconhecidas pela DnaA e por proteínas codificadas pelo próprio plasmídeo, além das sequências ricas em AT (ver Figura 6.8B). Os exemplos aqui relacionados sugerem que as características encontradas na *oriC* de *E. coli* não são específicas do genoma dessa espécie, podendo estar presentes em outros genomas procarióticos, plasmídeos ou bacteriófagos.

A caracterização de origens da replicação nos cromossomos de organismos eucarióticos apresenta dificuldades devido à complexidade desses genomas. Em cromossomos de algumas espécies de leveduras e de alguns fungos filamentosos foram isoladas sequências com replicação autônoma ARS (de *autonomously replicating sequence* = sequência autônoma de replicação), ricas em bases AT, que atuam como origens da replicação. No entanto, no estudo de origens da replicação em outros organismos, como em mamíferos, tanto o mapeamento genético destes elementos como o estabelecimento de sua função tem apresentado certa dificuldade. Estudos recentes de mapeamento de origens da replicação no genoma de células HeLa (células cancerosas imortais mantidas em cultura *in vitro*) revelaram que muitas origens estão sobrepostas a regiões de ligação de fatores que regulam

a transcrição, não existindo, no entanto, qualquer conexão entre a regulação da expressão gênica e a origem da replicação. Durante os últimos anos, diversos estudos demonstraram a importância da estrutura da cromatina na regulação do início da replicação em eucariotos. Apesar destes esforços, pouco se conhece sobre a influência da cromatina na ativação de origens de replicação.

Existem evidências que sugerem a presença de proteínas iniciadoras em todos os domínios da vida (Bactéria, Archaea e Eukarya). Proteínas relacionadas ao início da replicação em espécies eucarióticas e em arqueas formam um complexo denominado **ORC** (de *origin recognition complex* = complexo de reconhecimento de origem da replicação). As proteínas do ORC foram inicialmente caracterizadas em leveduras como um complexo de seis proteínas diferentes, sendo que suas subunidades ORC1 a ORC5 parecem estar presentes em todos os eucariotos. Apesar do ORC atuar, assim como o oligômero de DnaA, como um iniciador da replicação, estes dois complexos proteicos não apresentam identidade entre as respectivas sequências de aminoácidos. No entanto, estudos considerando a estrutura conformacional da DnaA e de proteínas do ORC demonstram a existência de domínios funcionais similares, como também a utilização do mesmo mecanismo de ligação (enrolamento) ao DNA.

O mecanismo de atuação do ORC foi proposto inicialmente para leveduras. Nesses organismos, o complexo de subunidades ORC1-6 encontra-se ligado às origens da replicação no DNA, durante todas as fases do ciclo celular (Figura 6.9). Portanto, a função principal do ORC seria a de servir como base para a associação sequencial de várias proteínas que formam o complexo de pré-replicação, incluindo a Cdc6 (de *cell division cycle 6* = proteína 6 do ciclo de divisão celular), a Cdt1 (de *Cdc10-dependent transcript* = proteína Cdc10-dependente de transcritos) e a helicase Mcm. A Cdc6 e a Cdt1 são fatores de replicação altamente regulados, presentes na célula somente na fase G1 do ciclo celular, promovendo a conexão entre o ORC e os outros fatores do complexo de pré-replicação. Uma característica interessante da Cdc6 é sua similaridade com a ORC1, e estudos filogenéticos sugerem que os genes codificadores dessas proteínas poderiam ser parálogos.

O processo de replicação nos eucariotos, utilizando levedura como modelo, pode ser separado nas seguintes etapas (ver Figura 6.9A):

1. no início da fase G1, ORC está ligado às origens da replicação;
2. durante a fase G1 ocorre a expressão da Cdc6, que reconhece a presença do ORC em cada origem e liga-se ao complexo;
3. no final da fase G1, ORC/Cdc6 recrutam Mcm para a formação do complexo de pré-replicação;
4. a célula passa, então, para a fase S (fase de síntese de DNA, replicação) na qual ocorre a replicação. Após o início da replicação, tanto Cdc6 como Mcm se desli-

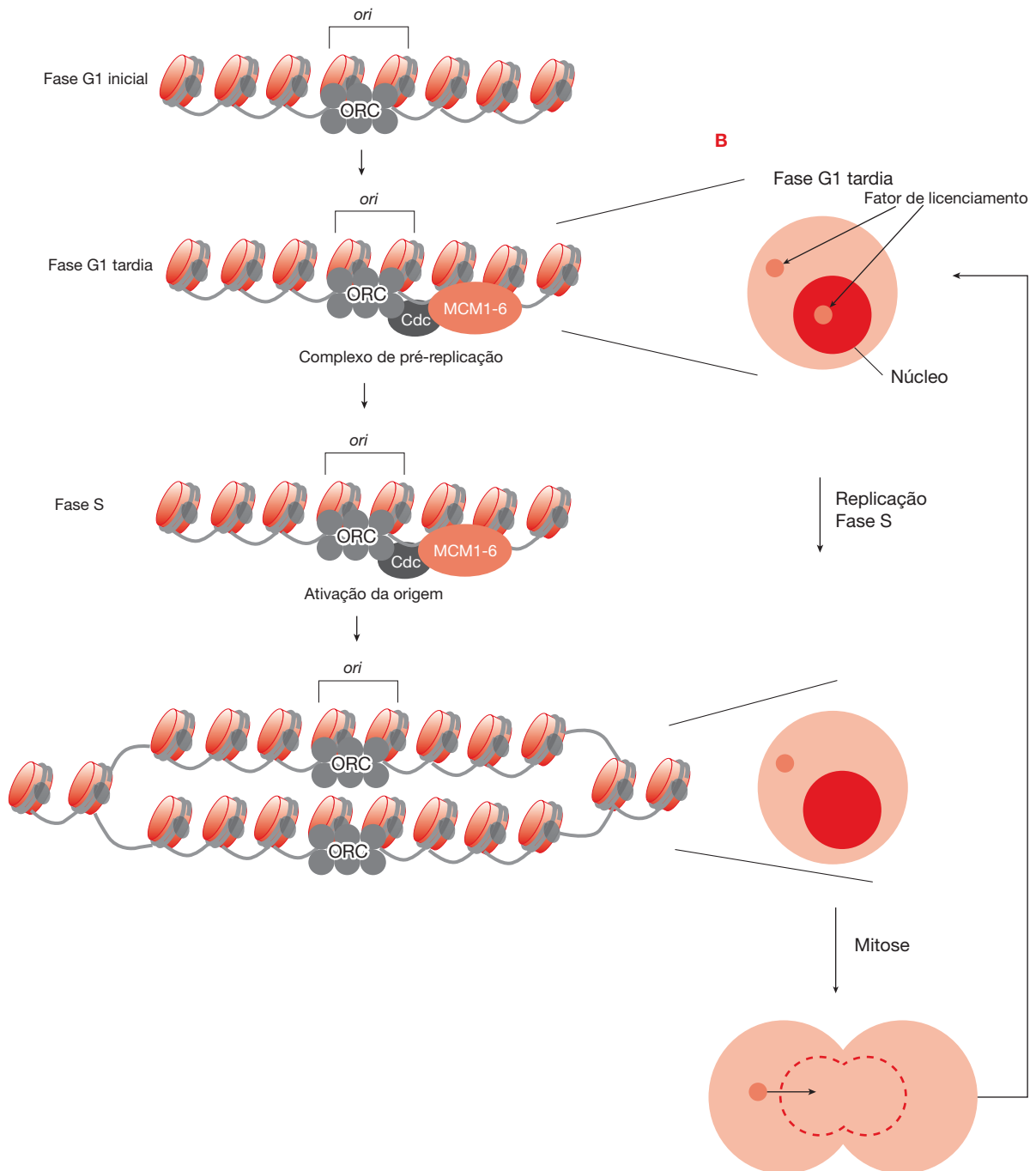




Figura 6.9

Início da replicação em organismos eucarióticos. (A) Representação da interação do ORC com outras proteínas do complexo de pré-replicação. Os cilindros achatados em  representam os octâmeros de histonas aos quais o DNA (linha em ) está associado para formar os nucleossomos. A origem da replicação está indicada (*ori*). Na fase S, as origens (há várias ao longo do cromossomo eucariótico) são ativadas e ocorre a replicação do DNA. (B) Esquema de uma célula demonstrando a presença do fator de licenciamento durante as fases do ciclo celular. Na fase G1 tardia, o fator de licenciamento está no interior do núcleo, permitindo que, na fase S, as origens de replicação sejam ativadas. Após a replicação, esse fator é inativado e só aparece novamente no núcleo depois da mitose.

gam da origem. Nesta fase, somente o ORC permanece ligado à origem;

5. as proteínas Cdc6 e Mcm são inativadas não podendo mais se ligar à origem.

Este mecanismo garante que as origens de replicação de eucariotos sejam ativadas apenas uma vez a cada ciclo celular.

4.2 Ciclo celular e origem da replicação

Os mecanismos que regulam a replicação do DNA e a divisão celular são, em geral, analisados de forma independente. No entanto, replicação e ciclo celular devem estar conectados de forma que a frequência dos ciclos de replicação se ajuste à velocidade do crescimento da célula, e o término da replicação seja relacionado ao final da divisão celular. Devem existir, portanto, sinais celulares como velocidade de síntese das proteínas ou resposta à falta de aminoácidos ou fontes de carbono, que sejam responsáveis pela indução dos ciclos de replicação e de divisão celular. A relação entre os dois ciclos é necessária, uma vez que, no processo de divisão de uma célula bacteriana, por exemplo, devem existir dois cromossomos completos e a célula deve ter atingido um tamanho mínimo necessário.

Células de *E. coli*, multiplicando-se em condições ideais, possuem um tempo de geração de aproximadamente 20 min. Nessas condições, não haveria tempo suficiente para ocorrer um ciclo de replicação, e o septo de separação das células também não estaria formado por completo. Portanto, os ciclos de replicação, assim como do início da formação do septo, devem ocorrer antes do processo de divisão celular. Neste modelo, as células, logo após a divisão celular, devem conter o DNA em que um novo processo de replicação foi iniciado, originando cromossomos com várias forquilhas de replicação durante os diferentes estágios de desenvolvimento do ciclo.

Um aspecto importante na inter-relação entre replicação e ciclo celular é o momento em que o processo de duplicação do cromossomo deve ser iniciado. Existem evidências apontando para o papel de uma proteína agindo como molécula reguladora da ativação da origem da replicação e, também, para a existência de uma ligação física entre o DNA e a membrana celular.

A ligação do DNA na membrana celular foi demonstrada com a região contendo a origem da replicação do cromossomo de *E. coli* e extratos de membranas celulares bacterianas. Essa região do DNA é rica em **sequências GATC**, que são sítios nos quais a adenina é passível de metilação nas duas fitas do DNA. As regiões metiladas podem influenciar no processo de ligação com a membrana celular. Quando o DNA inicia o processo de replicação, durante um período de aproximadamente 8 min, a nova fita sintetizada ainda não foi metilada, originando um DNA hemimetilado na região de *oriC* (**Figura 6.10A**). Nessas condições, o DNA é capaz de ligar-se à membrana, ao passo que o DNA totalmente metilado não liga. Uma explicação possível para a importância desse evento seria

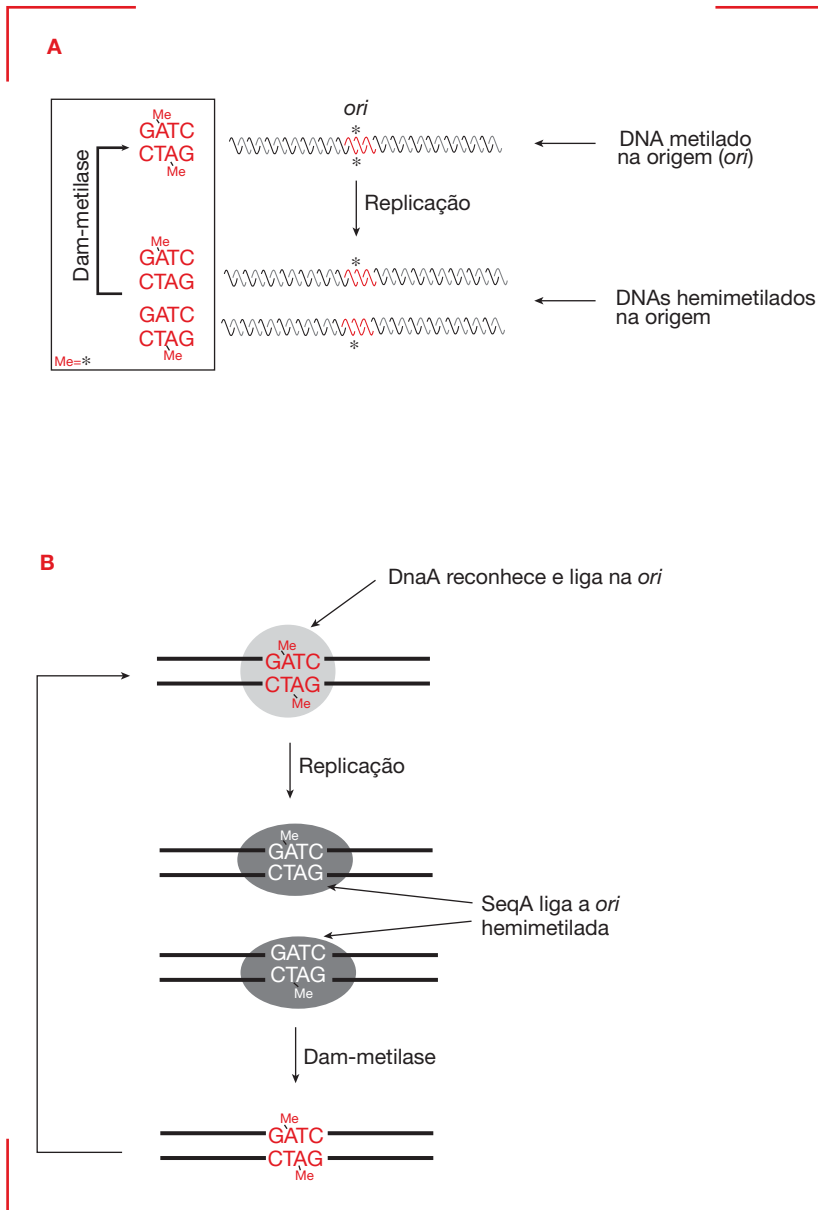
o fato de que a ligação do DNA à membrana, no início da replicação, permitiria a segregação de um cromossomo para cada uma das células após o processo de divisão, uma vez que a coordenação destes dois processos (divisão celular e segregação dos cromossomos) está relacionada a uma DNA-translocase (FtsK) localizada no septo.

Contudo, devem existir outros mecanismos de regulação da reativação da origem da replicação, pois mutantes de *E. coli* deficientes no processo de metilação do DNA são ainda capazes de divisão celular normal. Uma explicação provável para a metilação tardia de uma das fitas do DNA seria o sequestro da origem da replicação, impedindo a ação da proteína que realiza a metilação (Dam-metilase). Uma proteína, denominada **SeqA**, apresenta maior afinidade de ligação ao DNA hemimetilado, sugerindo seu papel como proteína sequestradora de *oriC* e, portanto, como reguladora negativa do circuito de ativação de *oriC*. A SeqA estabelece um complexo muito estável com as múltiplas sequências GATC hemimetiladas, que surgem na replicação, impedindo transitoriamente a formação do DNA metilado (**Figura 6.10B**). Além desta função, a SeqA é encontrada na forquilha de replicação onde, provavelmente, é responsável pela reestruturação (formação dos nucleossomos) do DNA replicado.

Até o momento, foi demonstrada a existência de sequências GATC metiladas na origem da replicação e da proteína SeqA apenas em *E. coli* e espécies relacionadas. Estudos em outras bactérias sugerem a existência de sistemas diferenciados para regulação da reativação da origem da replicação. Nos organismos eucarióticos, o início da replicação deve ser regulado com exatidão, para assegurar que as origens serão ativadas somente uma vez a cada divisão celular, durante a fase S.

As células eucarióticas, em geral, iniciam seu ciclo como diploides (2n), permanecendo nesse estágio durante a fase G1. O tempo de duração da fase G1 é o mais variável entre as diferentes espécies. Assim que a célula inicia a fase G1, os outros estágios do ciclo celular seguem progressivamente. Na fase S, o início das origens da replicação ocorrem em diferentes momentos, possibilitando a duplicação do material genético. No final dessa fase, a célula é tetraploide (4n), permanecendo nessa condição durante a fase seguinte (G2). Na fase G2, o núcleo está se preparando para divisão e não ocorre mais replicação do DNA. Por fim, na mitose ou fase M, acontece a redução do número dos cromossomos, originando duas células diploides.

Os mecanismos que regulam o ciclo de replicação em células eucarióticas ainda não foram totalmente elucidados. Existem evidências que sugerem a existência de um **“fator liberador”** da replicação, limitando a ativação das origens da replicação do DNA uma vez a cada ciclo celular (ver **Figura 6.9B**). Como discutido na Seção 4.1, os ORCs encontram-se associados às origens da replicação durante todo o ciclo celular, ao passo que a Mcm somente se liga na fase G1, pela ação da Cdt1 e da Cdc6, que são consideradas como os fatores liberadores

**Figura 6.10**

Metilação e sequestro da origem da replicação. (A) Representação do processo de metilação do DNA, após a replicação. A replicação do DNA, que apresenta ambas as fitas metiladas marcadas com (*), origina duas moléculas de DNA de fita dupla com apenas a fita parental metilada. A metilação da nova fita sintetizada deve ser realizada pela enzima Dam-metilase. A sequência GATC em destaque (■) representa o sítio de metilação (Me=*) marcado em ■ na fita de DNA. (B) Esquema do mecanismo de sequestro de origem da replicação em *E. coli*. A proteína SeqA liga-se a DNA hemimetilado, impedindo a ação imediata da Dam-metilase. Após a metilação, a proteína sequestradora SeqA é removida do DNA e a DnaA pode então reconhecer e ligar-se à origem e iniciar um novo ciclo de replicação.

da replicação. No momento em que a célula inicia a fase S e as origens são ativadas, o complexo de pré-iniciação se separa do DNA. Durante esta fase, mecanismos que regulam os diferentes fatores liberadores não realizam a reiniciação das origens, o que limita a síntese do DNA a um evento por ciclo celular. Os mecanismos de regulação da expressão dos fatores liberadores Cdt1 e Cdc6, durante o ciclo celular, constituiriam uma etapa decisiva no processo de replicação, pois a presença de Cdt1/Cdc6 seria responsável pela liberação das origens da replicação na fase G₁ e, a inibição de sua função impediria a reativação das origens na fase S. Portanto, para a montagem do complexo de pré-replicação, ocorreria uma ligação coordenada dos fatores liberadores, que resultaria no recrutamento da helicase Mcm em cada forquilha de replicação.

5. Dinâmica da replicação: *Escherichia coli* como modelo

O processo de replicação, em todos os organismos, inicia com o reconhecimento da origem por uma proteína específica. A seguir, ocorre o recrutamento de outras proteínas, que terão a função de dar continuidade na abertura das fitas formando a bolha de replicação em que serão montados os replissomos responsáveis pela síntese das duas fitas de DNA (contínua e descontínua), um em cada forquilha de replicação.

Em *E. coli*, o reconhecimento da origem de replicação (*oriC*) é realizado pela proteína DnaA (ver Figura

6.8), que inicia a abertura das fitas recrutando o complexo DnaB/DnaC, com função helicase, que dará continuidade à quebra das pontes de hidrogênio entre as bases, permitindo a formação da bolha de replicação com duas forquilhas de replicação. Essas forquilhas de replicação irão se movimentar em direções opostas, convergindo para a região de término da replicação. Um único replissomo estaria localizado em cada forquilha, sendo responsável pelo desenrolamento das fitas de DNA, pela formação dos iniciadores e pela síntese simultânea das fitas contínua e descontínua. Portanto, estariam presentes no replissomo, além da DNA-girase e helicase, as proteínas SSB, a primase e a DNA-polimerase III.

A proposta de síntese coordenada das duas fitas, durante o avanço de uma forquilha de replicação, tem embasamento nas características estruturais e funcionais da holoenzima DNA-polimerase III. Essa holoenzima tem sido isolada como um dímero assimétrico (ver Figura 6.6B). Cada um dos seus “núcleos catalíticos” sintetiza uma das fitas novas do DNA, apresentando diferentes capacidades de dissociação do DNA. Na síntese das duas fitas, a holoenzima se movimenta continuamente, no entanto, na síntese descontínua a fita-molde deve ser “puxada” gerando uma alça de DNA (Figura 6.11). Assim, a síntese da fita contínua gera uma região de DNA de fita simples na fita original complementar, que servirá de molde para a síntese da fita descontínua. A estrutura da subunidade β , que consiste em duas subunidades formando um anel que circunda a fita-molde do DNA, permitiria que ela tivesse a função de “prender e soltar” a fita-molde de DNA durante a síntese descontínua.

A ligação da subunidade β envolvendo o DNA é realizada pelo complexo proteico γ , que está associado a síntese da fita descontínua. Este mecanismo poderia ser separado em três etapas: (1) o complexo γ , na presença de ATP, se ligaria a subunidade β abrindo sua estrutura em anel; (2) o complexo γ -ATP se ligaria a seguir na região do iniciador (onde deveria iniciar a síntese), induzindo a hidrólise do ATP; e (3) após a hidrólise do ATP o complexo γ se dissociaria do DNA. Por fim, a DNA-polimerase III seria recrutada para a região do iniciador, interagindo com a subunidade β e iniciando a incorporação dos nucleotídeos. A DNA-polimerase III termina a síntese de um fragmento de Okazaki ao encontrar a extremidade 5'-P do fragmento de Okazaki anteriormente sintetizado. Neste momento, o núcleo catalítico da DNA-polimerase III, responsável pela síntese da fita descontínua, se dissocia da subunidade β e se reassocia a uma nova subunidade β , que está localizada em um novo iniciador, recém-sintetizado (ver Figura 6.11).

A helicase que se encontra na frente da forquilha de replicação, realizando a abertura das fitas de DNA, recruta a primase. Para a síntese dos iniciadores na fita descontínua, a primase deve movimentar-se em sentido oposto ao do movimento da forquilha. Portanto, a primase deve se movimentar ao longo da fita simples até o início do iniciador, no mesmo sentido do avanço da forquilha. A formação da alça no DNA poderia ser utiliza-

da pela fita descontínua para coordenar o movimento da primase, que realiza a síntese no sentido oposto, com a replicação simultânea das duas fitas pela holoenzima (ver Figura 6.11).

Este mecanismo de replicação pode também ser utilizado por células eucarióticas. O modelo sugerido para eucariotos apresenta a forquilha de replicação contendo um complexo de DNA-polimerase α /primase e dois complexos de DNA-polimerase δ (ou uma DNA-polimerase δ e outra ϵ). Desta forma, a DNA-polimerase α seria responsável pela síntese dos iniciadores, e os dois complexos de DNA-polimerase δ (ou δ/ϵ) teriam o mesmo mecanismo de ação da DNA-polimerase III de *E. coli*, sendo um dos complexos responsável pela síntese da fita contínua e o outro pela síntese da fita descontínua.

Uma visão mais ampla do processo de replicação sugere a existência de uma alternância de polimerases na forquilha de replicação. Portanto, uma única holoenzima não seria responsável pela replicação de todo o cromossomo de *E. coli*, podendo ocorrer alternância de DNA-polimerase III durante o processo. Este modelo de troca de DNA-polimerases seria decisivo na superação de eventuais barreiras estruturais existentes no DNA, como lesões causadas por mutações no DNA-molde, ou mesmo para a substituição de uma DNA-polimerase III defeituosa, sem que haja a necessidade de desfazer todo o replissomo de uma forquilha de replicação.

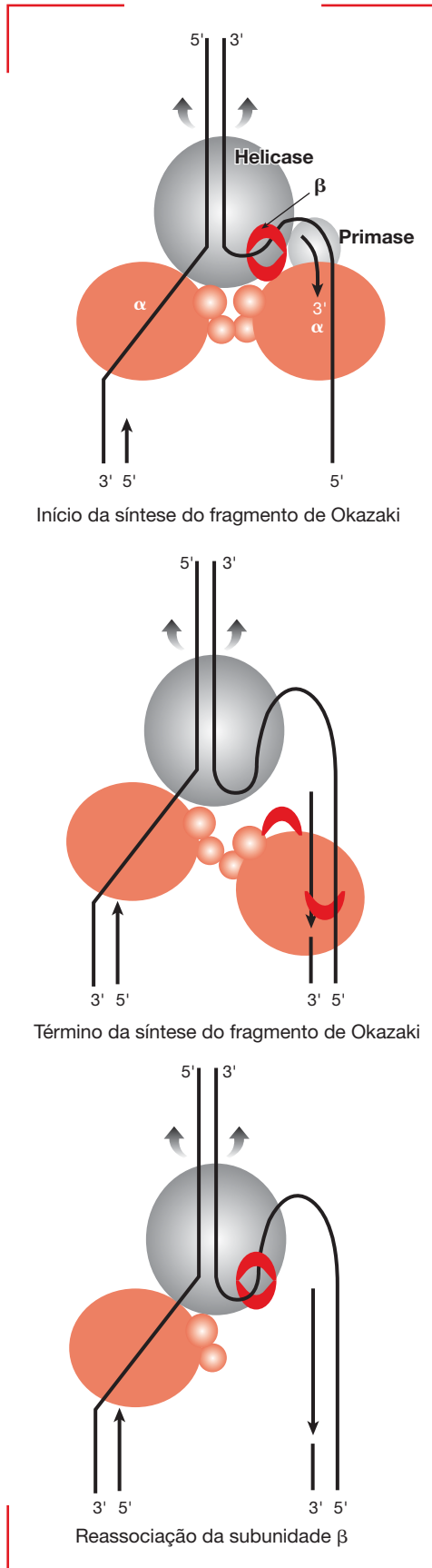
6. Término da replicação

Os cromossomos de organismos procarióticos geralmente apresentam estrutura circular fechada sem extremidades livres. Esse tipo de estrutura circular também pode ser encontrada em plasmídeos, bacteriófagos, DNA de cloroplastos e de mitocôndrias. A replicação de uma molécula circular não apresenta problemas de término, pois ele ocorre quando as duas forquilhas de replicação se encontram (replicação bidirecional), ou quando a forquilha atinge novamente a região de origem de replicação (replicação unidirecional).

O término da replicação de cromossomos lineares, como os das células eucarióticas, possivelmente ocorra pelo encontro das forquilhas de replicação adjacentes. Assim, a fusão das bolhas de replicação dispensaria o uso de sequências específicas para a regulação dos processos de replicação. No entanto, são encontrados sítios de pausa de replicação em eucariotos, com função mais especializada para alguns genes, como, por exemplo, a região onde estão localizados os genes codificadores de rRNA. Porém, esses sítios de pausa não estão relacionados ao controle da fusão das forquilhas de replicação.

6.1 Genomas circulares

Na replicação do cromossomo circular de *E. coli*, as duas forquilhas de replicação se encontram na região terminal

**Figura 6.11**

Modelo de replicação do DNA em *E. coli*. Modelo proposto demonstrando a participação da subunidade β da DNA-polimerase III (β em ■) na síntese da fita descontínua do DNA. Além das subunidades β , somente as subunidades catalíticas (α) da DNA-polimerase III estão representadas. As setas representam o movimento da forquilha de replicação e o sentido da síntese das fitas ($5' \rightarrow 3'$) está indicado por setas (em ■).

denominada sítio *Ter*, localizada diametralmente oposta ao *oriC*. Existem duas regiões gerais de terminação (T1 e T2), cada uma com várias sequências específicas, localizadas de cada lado do ponto central de encontro das forquilhas. Cada sítio *Ter* seria específico para uma direção do movimento da forquilha, fazendo com que cada forquilha passe pela outra para finalizar a replicação (Figura 6.12A). O término da replicação necessita do produto do gene *tus*, que codifica uma proteína que reconhece e liga-se à sequência consenso presente no sítio *Ter*, impedindo, com isso, o movimento da forquilha de replicação. A proteína Tus liga-se aos sítios *Ter* na forma de monômero, bloqueando a atividade de helicase da DnaB de uma maneira chamada de “orientação-específica” (Figura 6.12B). Diversas teorias existem para explicar esta parada da replicação orientação-específica. Existem evidências indicando que, quando DnaB se aproxima do complexo Tus*Ter*, com a “orientação-específica” correta, seu movimento fica bloqueado e a helicase não consegue desfazer a ligação da Tus ao sítio *Ter*.

Diferentemente das proteínas envolvidas no reconhecimento das origens da replicação, as que estão presentes no término não apresentam conservação. As diferenças no término da replicação entre *E. coli* e *B. subtilis* são exemplos bem conhecidos, demonstrando diferenças tanto nas sequências *Ter* como nas proteínas terminadoras (Tus em *E. coli* e RTP em *B. subtilis*). Essas duas proteínas terminadoras da replicação apresentam apenas similaridade de função biológica, sendo totalmente diferentes tanto nas suas sequências de aminoácidos como nas suas estruturas tridimensionais.

6.2 Genomas lineares

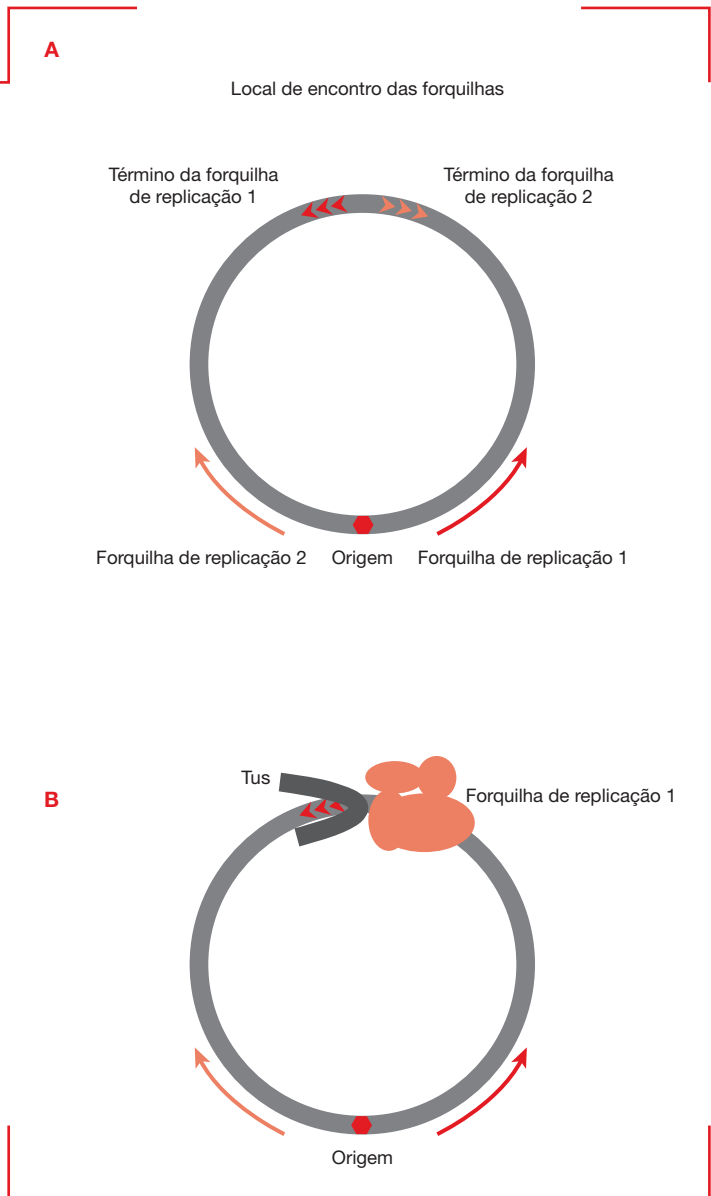
Como discutido na Seção 2.3, as DNA-polimerases, para a síntese das novas fitas de DNA, necessitam de um iniciador e somente incorporam os nucleotídeos no sentido $5' \rightarrow 3'$. Portanto, na síntese da fita descontínua, a remoção do iniciador localizado na extremidade $5'$ -P da cadeia resultaria na perda de parte do DNA a cada geração celular, o que acabaria por ocasionar a perda da informação genética (Figura 6.13A).

Diversos organismos podem apresentar genomas formados por moléculas de DNAs lineares, como ocorre com as extremidades dos cromossomos de organismos

Figura 6.12

Término da replicação de genomas circulares.

(A) Demonstração do sentido do movimento das forquilhas de replicação (setas em ■ e ■). Os sítios de término (*Ter*) estão marcados (triângulos em ■ e ■). (B) Representação da atividade da proteína Tus bloqueando o movimento do replissomo 1 (em ■), responsável pela forquilha de replicação 1. A proteína Tus liga-se às sequências terminadoras *ter* (triângulos ■), impedindo a passagem do replissomo 1.



eucarióticos e em genomas de vírus tanto de eucariotos como de procariotos. Contudo, existem alguns exemplos de mecanismos utilizados por estes organismos com o objetivo de evitar a perda da informação genética como resultado do término da replicação de moléculas de DNAs lineares.

6.2.1 Telômeros e telomerase

Os **telômeros** (extremidade dos cromossomos de eucariotos) estão relacionados à estabilidade genômica. Telômeros são compostos de sequências repetidas ricas em “G”, formando uma estrutura tridimensional, o que evita que as extremidades sejam reconhecidas como quebras de DNA de fita dupla. Durante os eventos de replicação, em muitas células, os telômeros diminuem seu tamanho conduzindo a **senescência (morte) celular**.

No entanto, em células com uma alta atividade replicativa (como as células germinativas) os telômeros são alongados pela atividade de uma enzima específica, a **telomerase**. Em células humanas, a telomerase é uma ribonucleoproteína composta por uma unidade catalítica com atividade de **transcriptase reversa** (denominada de hTERT) e uma molécula de RNA (denominada de hRT). A molécula de RNA serve de molde para a atividade de polimerização da transcriptase reversa (síntese de DNA no sentido 5'→3', utilizando um molde de RNA) (Figura 6.13B).

O mecanismo de ação da telomerase consiste em alongar uma das fitas de DNA por meio da adição de nucleotídeos na extremidade 3'-OH livre. A enzima adiciona as bases complementares à molécula de RNA presente na sua estrutura em direção à extremidade distal do cro-

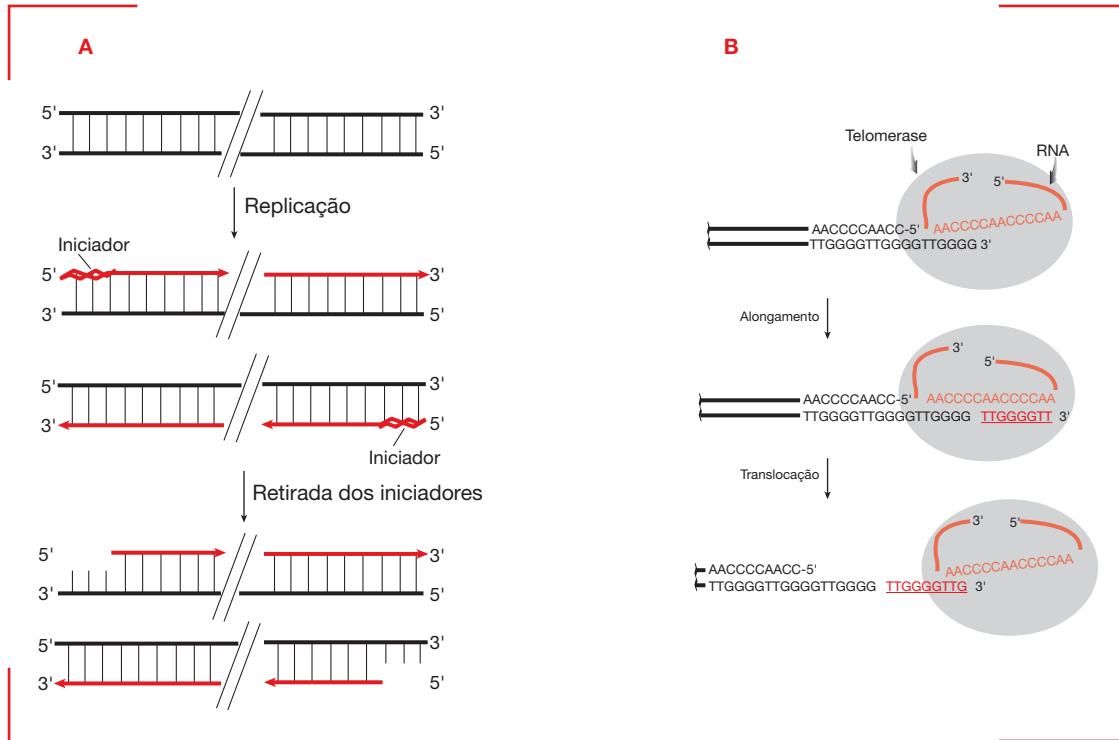


Figura 6.13

Término da replicação de genomas lineares. (A) Esquema do aparecimento de regiões de fita simples nas extremidades de DNAs lineares recém-replicados, após a remoção dos iniciadores. As novas fitas sintetizadas estão representadas como setas em ■, para indicar o sentido da síntese (5'→3'). **(B)** Modelo de replicação dos telômeros em organismos eucarióticos. Na figura estão representadas somente as extremidades dos cromossomos contendo as sequências repetidas (■). A telomerase (elipse em ■), contendo a molécula de RNA (■) que serve de molde, realiza a síntese da fita no sentido 5'→3'. Os nucleotídeos adicionados pela telomerase na molécula de DNA estão representados em ■ e sublinhados. Após a síntese dos primeiros nucleotídeos, a telomerase se desloca para a extremidade, mantendo o pareamento RNA-DNA, e reinicia a síntese.

mosso, ao atingir o final da fita-molde a telomerase se desloca para a extremidade recém-sintetizada, pareando com os últimos resíduos e iniciando novamente a síntese. Assim, várias unidades repetidas de TTGGGG são adicionadas em uma das fitas do DNA. A síntese da fita complementar destas regiões teloméricas pode ocorrer por meio da maquinaria normal de replicação da fita descontínua.

Apesar do mecanismo de manutenção dos telômeros pela atividade da telomerase ser bastante comum em vários organismos eucarióticos, ele não pode ser considerado um sistema absoluto. Em *Drosophila melanogaster*, por exemplo, não foi encontrada nenhuma telomerase, sendo a manutenção dos telômeros realizada pela adição de retroelementos (ver Capítulo 9) nas extremidades dos cromossomos.

6.2.2 Sistemas iniciadores alternativos

Genomas virais desenvolveram diferentes mecanismos para solucionar o problema de perda da parte do seu DNA ao término da replicação. Muitos DNAs virais apresentam estruturas secundárias nas suas extremidades,

circularizam seu DNA ou utilizam sistemas iniciadores alternativos, baseados em uma proteína ou um tRNA, para resolver este problema.

Alguns genomas de DNA lineares de diversos vírus apresentam uma proteína ligada covalentemente na extremidade 5'-P do seu DNA. Os exemplos melhor caracterizados são os do adenovírus e do bacteriófago $\phi 29$. Esses vírus replicam seus respectivos genomas a partir de uma das extremidades, por meio de um mecanismo denominado **deslocamento de fita**. Na extremidade 5'-P destes DNAs encontra-se uma proteína (terminal) ligada covalentemente ao DNA por meio de uma ligação fosfodiéster com uma serina. Essa proteína apresenta duas funções: é utilizada como iniciador, devido à presença de um nucleotídeo citidina para o início da replicação (**ver Material Complementar Online, Figura 1**); e interage com a DNA-polimerase. Desta forma, após a replicação, são produzidas novas fitas de DNA ligadas na extremidade 5' com a proteína terminal.

Outros vírus utilizam como iniciador para o início da replicação um tRNA, como no caso de alguns retrovírus.

Estes vírus infectam células eucarióticas e o genoma de alguns deles é composto por duas moléculas de RNA fita simples, possuindo como característica fundamental para sua multiplicação a integração, como DNA fita dupla, no cromossomo das células hospedeiras. As partículas virais contêm, além do material genético (RNA), um tRNA, que funciona como iniciador, e várias proteínas codificadas pelo vírus, como a transcriptase reversa. Essa enzima realiza a síntese de uma molécula de DNA fita dupla, a partir do RNA viral, etapa necessária para o processo de integração do genoma viral a um cromossomo da célula hospedeira.

As moléculas de DNA, contendo **sequências palindrômicas** nas extremidades, permitem a ocorrência de pareamentos intracadeia, gerando estruturas em grampo (*hairpins*). O início da replicação ocorre pelo pareamento na extremidade 3'-OH, que serve de iniciador para a replicação. Após a síntese da fita pela DNA-polimerase, uma nuclease específica cliva dentro da sequência palindrômica, permitindo que a síntese do DNA prossiga na direção 5'→3'.

Alguns fagos, como os bacteriófagos λ e T7, possuem genomas com regiões terminais contendo uma sequência repetida de nucleotídeos. O bacteriófago λ utiliza essas regiões terminais para tornar o seu DNA circular dentro da célula bacteriana e, assim, replicar utilizando a maquinaria da célula hospedeira. Estas extremidades podem também realizar ligações entre as unidades de DNA formando longas cadeias, chamadas de **concatâmeros**, contendo múltiplos genomas associados. Esses concatâmeros são gerados por um mecanismo alternativo de replicação, denominado **círculo rolante**. Assim, estes fagos não necessitam iniciadores para evitar perdas de seus genomas a cada ciclo de replicação.

7. Sistemas de replicação por círculo rolante

O mecanismo de círculo rolante é utilizado por alguns genomas para gerar cópias de seu DNA. Esse mecanismo pode ser encontrado durante diversos processos: de replicação de vírus que possuem DNA de fita dupla ou fita simples, de viroides que possuem RNA circular, de pequenos plasmídeos, ou até mesmo nos mecanismos de transferência de DNA durante o processo de conjugação bacteriana.

Entre os bacteriófagos de genoma constituídos por DNA de fita simples circular podemos encontrar aqueles com capsídeos icosaédricos, como ϕ X174, ou filamentosos, como M13. O ciclo de replicação desses bacteriófagos pode ser dividido em três estágios: transformação do genoma viral da forma de fita simples em uma forma replicativa (RF) de fita dupla; replicação do RF para aumentar seu número de cópias por célula; e finalização da replica-

ção pelo mecanismo de círculo rolante, gerando DNA fita simples para ser empacotado no capsídeo viral.

A produção de RF, a partir do genoma viral fita simples, depende apenas da maquinaria da célula hospedeira. Basicamente, ocorre um reconhecimento de uma região do genoma, pela primase ou pela RNA-polimerase celular, para a síntese de um iniciador, utilizado pela DNA-polimerase bacteriana para a síntese da fita complementar.

Os mecanismos envolvidos na replicação do RF e na geração de fita simples mais conhecidos são os utilizados pelo bacteriófago ϕ X174. A “proteína A”, codificada pelo fago, apresenta função indispensável para esse processo, sendo necessária tanto para o início da replicação da RF, como para o seu término. Essa proteína liga-se a uma região do DNA da RF circular, induzindo a quebra de uma ligação fosfodiéster de uma das fitas e permanecendo ligada na extremidade 5'-P. A seguir, a helicase da célula hospedeira separa as duas fitas, que são recobertas pelas proteínas SSB. A extremidade 3'-OH serve de iniciador para a síntese da nova fita e a DNA-polimerase utiliza a fita complementar intacta como molde. A proteína A ligada à extremidade 5'-P direciona a formação das moléculas de fita simples circulares. Consequentemente, as moléculas de fita simples circulares podem ser duplicadas (gerando mais RF e reiniciando o processo) ou participar da formação da partícula viral infectiva (**ver Material Complementar Online, Figura 2**).

Outros fagos, como o bacteriófago λ , apresentam o genoma composto por uma molécula de DNA de fita dupla linear, de 48 kb, possuindo regiões terminais com 12 nucleotídeos de fita simples complementares, denominadas regiões ou sítios coesivos (sítios *cos*). Após a infecção de células de *E. coli*, o DNA injetado circulariza-se devido ao pareamento dos sítios *cos*, originando uma molécula de DNA de fita dupla circular (**ver Material Complementar Online, Figura 3**).

Para realizar sua replicação, o bacteriófago λ , assim como outros, necessita das enzimas da célula hospedeira. Entretanto, o genoma de λ codifica duas proteínas envolvidas na sua replicação: as proteínas O e P, responsáveis pelo reconhecimento da origem da replicação do fago. A proteína O teria função equivalente a da DnaA no reconhecimento e na ligação à origem da replicação (*OriC*) do cromossomo bacteriano, induzindo a abertura das fitas em uma região rica em AT (ver Seção 4). A proteína P auxilia no processo, por interagir com a helicase DnaB. A necessidade das proteínas O e P para a replicação condiciona a transcrição da região do genoma, que codifica essas proteínas, a ser o primeiro processo que ocorre após a infecção.

Durante os primeiros 5 a 15 min após a infecção da célula pelo bacteriófago λ e após a transcrição das proteínas O e P, a replicação procede de forma semelhante àquela descrita para o cromossomo de *E. coli* (ver Seção

4), sem ocorrer, no entanto, reinício na origem da replicação antes do final do ciclo iniciado. Após 15 min, a replicação do bacteriófago λ passa a ocorrer pelo mecanismo de círculo rolante, para originar as moléculas ou fita dupla lineares. Isto acontece, porque formas monoméricas do DNA de λ não podem ser empacotadas no capsídeo viral, sendo necessária a formação de múltiplos de DNA (concatâmeros) para a produção de partículas virais infectivas. A replicação por círculo rolante permite então a produção dos concatâmeros lineares.

Mecanismos replicativos de círculo rolante também são utilizados por elementos genéticos móveis presentes em células eucarióticas. Viroides são dependentes de suas células hospedeiras, pois não codificam no seu genoma nenhuma proteína. Alguns viroides, que apresentam seu

genoma composto por uma única molécula de RNA circular, replicam no interior do núcleo ou de cloroplastos utilizando um mecanismo de círculo rolante.

Alem de vírus, plasmídeos também utilizam mecanismos de círculo rolante para sua replicação. Alguns plasmídeos pequenos (menores do que 15 kb) replicam via um mecanismo de círculo rolante similar ao descrito para os fagos. O processo de círculo rolante plasmidial inicia pela ligação de uma proteína específica (denominada Rep), codificada pelo plasmídeo em uma origem da replicação de fita dupla (**ver Material Complementar Online 6.2**). Diferentes plasmídeos apresentam este tipo de mecanismo, existindo variações tanto nas proteínas Rep como nas origens da replicação, o que permite a classificação dos plasmídeos em pelo menos cinco grupos.



Resumo

Todas as células replicam seu DNA de maneira semi-conservativa, considerando a complementaridade de bases e o antiparalelismo das fitas duplas de DNA. A replicação do genoma se origina sempre em um local específico, denominado origem da replicação, e o número de origens da replicação varia entre os diferentes organismos.

As enzimas que realizam a replicação do DNA, as DNA-polimerases, sintetizam as novas fitas de DNA incorporando os nucleotídeos na extremidade 3'-OH livre. Portanto, uma das fitas novas é sintetizada continuamente no sentido 5'→3' e a outra fita deve ser sintetizada descontinuamente, na forma de fragmentos, para manter a incorporação no sentido 5'→3'. Além da DNA-polimerase, diversas proteínas participam do processo de replicação, formando um complexo denominado replissomo. Na forquilha de replicação estão presentes enzimas que rompem pontes de hidrogênio

entre as bases, separam as fitas (helicase) e sintetizam sequências iniciadoras (primase), fornecendo segmentos nucleotídicos com extremidades 3'-OH livres (iniciadores) para o alongamento por parte da DNA polimerase.

O reconhecimento da origem de replicação por meio de proteínas específicas é uma etapa fundamental para o processo de replicação. Em bactérias, como *E. coli*, a proteína DnaA é responsável pelo reconhecimento da origem da replicação, e os sistemas de metilação e o sequestro da origem são responsáveis pelos mecanismos de controle da ativação da origem da replicação. Em genomas circulares, o término da replicação ocorre quando duas forquilhas de replicação se encontram, permitindo a síntese completa das moléculas de DNA. Em genomas lineares existem alguns mecanismos para permitir a síntese das extremidades das moléculas, evitando perdas de material genético.



Leituras recomendadas

Bates D. The bacterial replisome: back on track? *Mol Microbiol.* 2008;69(6):1341-8.

Duggin IG, Wake RG, Bell SD, Hill TM. The replication fork trap and termination of chromosome replication. *Mol Microbiol.* 2008;70(6):1323-33.

Duncker BP, Chesnokov IN, McConkey BJ. The origin recognition complex protein family. *Genome Biol.* 2009;10(3):214.

Groth A. Replicating chromatin: a tale of histones. *Biochem Cell Biol.* 2009;87(1):51-63.

Klinge S, Núñez-Ramírez R, Llorca O, Pellegrini L. 3D architecture of DNA Pol alpha reveals the functional core of multi-subunit replicative polymerases. *EMBO J.* 2009;28(13):1978-87.

Lutzmann M, Méchali M. How to load a replicative helicase onto chromatin: a more and more complex matter during evolution. *Cell Cycle.* 2009;8(9):1309-13.

Reyes-Lamothe R, Sherratt DJ, Leake MC. Stoichiometry and architecture of active DNA replication machinery in *Escherichia coli*. *Science.* 2010;328(5977):498-501.

Sakakibara N, Kelman LM, Kelman Z. How is the archaeal MCM helicase assembled at the origin? Possible mechanisms. *Biochem Soc Trans.* 2009;37(Pt 1):7-11.

Simonetta KR, Kazmirski SL, Goedken ER, Cantor AJ, Kelch BA, McNally R, et al. The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. *Cell.* 2009;137(4):659-71.

Tahirov TH, Makarova KS, Rogozin IB, Pavlov YI, Koonin EV. Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. *Biol Direct.* 2009;4(1):11.

Waldminghaus T, Skarstad K. The Escherichia coli SeqA protein. *Plasmid.* 2009;61(3):141-50.

Luciane M. P. Passaglia

Mutação e Reparação do DNA

1. Mutações 134

- 1.1 Natureza das mutações 134
- 1.2 Classificação das mutações 134
 - 1.2.1 Classificação por localização 134
 - 1.2.2 Classificação por tipo de alteração molecular 135
 - 1.2.3 Classificação por efeitos fenotípicos 135
- 1.3 Mutações espontâneas 136
 - 1.3.1 Erros durante a replicação do DNA 136
 - 1.3.1.1 Deslizes de replicação 136
 - 1.3.1.2 Modificações tautoméricas 136
 - 1.3.2 Lesões hidrolíticas: desaminação e depurinação 137
 - 1.3.3 Lesões oxidativas 140
- 1.4 Frequência de mutações 140
- 1.5 Mutações induzidas 140
 - 1.5.1 Mutações induzidas quimicamente 140
 - 1.5.1.1 Análogos de bases 140
 - 1.5.1.2 Ácido nitroso, corantes de acridina e agentes alquilantes 141
 - 1.5.2 Mutações induzidas por radiações 142
 - 1.5.2.1 Radiações ionizantes 142
 - 1.5.2.2 Radiação ultravioleta 142

- 1.6 Mutações no homem 144
 - 1.6.1 Doença das células falciformes 144
 - 1.6.2 Grupos sanguíneos ABO 145
 - 1.6.3 Distrofias musculares 145
- 1.7 Aplicações práticas das mutações 145

2. Reparação do DNA 146

- 2.1 Revisão de leitura e reparação de mal pareamentos 146
- 2.2 Reversão direta de lesão 147
 - 2.2.1 Fotorreativação enzimática 147
 - 2.2.2 Reparação enzimática de bases alquiladas 148
- 2.3 Reparação por excisão 148
 - 2.3.1 Excisão de base 148
 - 2.3.2 Excisão de nucleotídeos 149
- 2.4 Sistema de reparação pós-replicação 153
 - 2.4.1 Reparação por recombinação 153
 - 2.4.2 Sistema SOS 154
 - 2.4.3 Reparação sujeita a erro: síntese translesão 155
- 2.5 Reparação de quebras duplas no DNA 156
 - 2.5.1 Por recombinação homóloga 156
 - 2.5.2 União de extremidades não homólogas 158
 - 2.5.3 União de extremidades mediada por micro-homologias 158
 - 2.5.4 Anelamento de fitas simples 158

1. Mutações

As moléculas no DNA de um organismo não são estáticas. Com frequência, as bases nitrogenadas de seus nucleotídeos estão expostas aos agentes, naturais ou artificiais, que provocam modificações em sua estrutura ou composição química. Todo organismo que exibe uma forma diferente da de seus ascendentes, resultado da presença de uma mutação, é denominado **mutante**. Utilizado em um sentido amplo, o termo **mutação** refere-se a qualquer modificação súbita e hereditária no conjunto gênico de um organismo, não explicável pela recombinação da variabilidade genética preexistente. Essas alterações incluem mudanças, no número de cromossomos (euploidia e aneuploidia), na estrutura cromossômica (aberrações cromossômicas) e alterações nos genes individuais. Neste capítulo, o termo mutação será utilizado em um sentido mais restrito, referindo-se apenas às alterações detectadas em nível de genes individuais.

A perpetuação do material genético, de geração para geração, depende da manutenção das taxas de mutação em níveis mínimos. Taxas de mutação elevadas na linhagem germinativa levariam à extinção da espécie, ao passo que taxas de mutação elevadas nas linhagens somáticas provocariam a morte do indivíduo. Dessa forma, as células dependem do funcionamento correto de milhares de genes, sendo que cada um deles pode ser danificado por mutação em qualquer um dos sítios da sequência codificadora da proteína ou do RNA funcional, ou, ainda, das regiões reguladoras que promovem sua expressão.

Em geral, os organismos portadores de mutação, em um determinado gene, apresentam problemas na sua sobrevivência e acabam sendo eliminados por seleção natural. No entanto, nem toda mutação resulta em uma consequência deletéria para o seu portador. Na verdade, a mutação é a base de toda a variabilidade genética, constituindo a matéria-prima para a evolução. A recombinação rearranja essa variabilidade em combinações novas e a seleção natural (ou artificial) preserva as combinações melhor adaptadas às condições ambientais existentes (ou desejadas). Sem a mutação, todos os genes existiriam apenas em uma forma e os organismos não seriam capazes de evoluir e de se adaptar às condições ambientais.

1.1 Natureza das mutações

Todos os seres vivos sofrem mutações como resultado de funções celulares normais ou interações aleatórias com o ambiente. Essas mutações são denominadas **espontâneas**, e a sua frequência de ocorrência é característica para cada espécie, constituindo o chamado nível basal (*background*). A ocorrência de mutações pode aumentar pelo tratamento com determinados compostos. Esses compostos são denominados **agentes mutagênicos** e as modificações que eles causam, **mutações induzidas**. Muitos agentes mutagênicos atuam diretamente no DNA, devido a sua capacidade de atuar como base nitrogenada ou de incorporar-se à cadeia polinucleotídica. O efeito de

um agente mutagênico é mensurado pelo grau em que ele aumenta a frequência de mutação acima do nível basal.

Qualquer base nitrogenada do DNA pode ser mutada. Uma **mutação pontual** envolve modificação em um único par de bases (**substituição**) ou em poucos pares de bases (**adição/inserção** ou **deleção**). A mutação pontual pode ser o resultado do mau funcionamento do sistema celular que replica ou repara o DNA, inserindo uma base incorreta na cadeia polinucleotídica, que está sendo sintetizada, ou de uma interferência química sobre uma das bases do DNA.

1.2 Classificação das mutações

As mutações devem causar alguma modificação detectável, para que a sua presença seja reconhecida. As modificações podem ser muito pequenas, identificadas apenas por técnicas genéticas e bioquímicas especiais; grandes, ao ponto de afetar a morfologia da célula que as contém; ou ainda, podem ser letais. A extensão na qual uma mutação altera as características de um organismo depende do local onde essa mutação ocorreu e o grau em que ela afetou o funcionamento do produto gênico.

Devido à grande variedade de tipos e efeitos das mutações, os geneticistas classificam as mutações de diversas maneiras. Essas classificações não são mutuamente exclusivas, podendo uma mesma mutação ser classificada por mais de uma maneira.

1.2.1 Classificação por localização

As mutações podem ser classificadas conforme o tipo celular ou a localização cromossômica em que ocorrem, podendo surgir em qualquer estágio do ciclo celular. **Mutações somáticas** acontecem em qualquer célula do organismo, exceto as células germinativas, e serão perpetuadas apenas nas células descendentes da célula original na qual a mutação ocorreu, podendo não afetar o organismo inteiro. **Mutações germinativas** são aquelas que ocorrem nos gametas e serão transmitidas à descendência. **Mutações autossômicas** são aquelas que ocorrem em genes localizados nos cromossomos autossômicos e **mutações ligadas ao X** são as que afetam genes localizados no cromossomo X.

Quando uma mutação autossômica **recessiva** ocorrer em uma célula somática de um organismo diploide é pouco provável que ela resulte em um fenótipo detectável. A expressão da maioria dessas mutações provavelmente será mascarada pelo alelo de tipo selvagem presente na célula. As mutações somáticas terão um impacto fenotípico apenas se forem **dominantes** ou, em indivíduos de sexo masculino, se elas estiverem ligadas ao X, uma vez que essas mutações podem ser expressas de imediato. De forma semelhante, o impacto das mutações somáticas ou ligadas ao X será mais aparente quando elas ocorrerem no início do desenvolvimento do organismo, em que um pequeno número de células indiferenciadas se multiplica para dar origem aos diferentes tecidos e órgãos. Mu-

tações dominantes, que ocorrem em células de tecidos adultos, frequentemente são mascaradas pelas centenas de milhares de células não mutantes do mesmo tecido, que desempenham a função celular normal.

Mutações germinativas têm o potencial de serem expressas em todas as células de um descendente. Uma mutação autossômica dominante herdada será fenotipicamente detectável na primeira geração. Mutações recessivas ligadas ao X, que surgirem em um gameta de uma fêmea **homogamética**, poderão ser expressas no descendente macho **hemizigótico**, caso ele receba o cromossomo X afetado. Devido à **heterozigosidade**, a ocorrência de uma mutação autossômica recessiva em gametas, tanto de machos como de fêmeas (mesmo que resulte em um alelo letal), pode passar despercebida por muitas gerações, até que o alelo mutante tenha se espalhado na população. Em geral, o novo alelo se tornará evidente apenas quando cruzamentos aleatórios juntarem duas cópias suas na condição de **homozigose**.

1.2.2 Classificação por tipo de alteração molecular

Frequentemente as mutações são classificadas em termos das alterações de nucleotídeos que constituem a mutação. Como já mencionado, a substituição de um par de bases por outro em uma molécula de DNA é denominada mutação pontual, ou substituição de base. A troca de um nucleotídeo dentro da região codificadora de um gene pode resultar na criação de um novo códon, que corresponderá a um aminoácido diferente no produto proteico (ver Capítulo 12). Essa mutação é denominada **mutação com sentido trocado** (de *missense mutation*). Um segundo resultado possível é que o códon original seja substituído por um códon de parada, resultando na terminação prematura da proteína traduzida. Essa mutação é denominada **mutação sem sentido** (de *nonsense mutation*). Caso a mutação pontual altere o códon, mas essa alteração não resulte na troca do aminoácido correspondente na proteína (devido à degeneração do código genético), ela será denominada **mutação silenciosa**.

Dois outros termos também são utilizados com frequência para descrever uma substituição de bases. A substituição de uma purina por outra purina ou de uma pirimidina por outra pirimidina é chamada de **transição** e a substituição de uma purina por uma pirimidina, ou vice-versa, é chamada de **transversão**. São possíveis quatro diferentes transições e oito diferentes transversões (Figura 7.1).

Um terceiro tipo de mutação pontual envolve a adição ou a deleção de um ou de alguns pares de bases e são referidas coletivamente como **mutações que modificam a fase de leitura** (de *frameshift mutations*), pois alteram a fase de leitura de todas as trincas de pares de bases no gene depois do sítio mutado. Uma mutação que modifica a fase de leitura ocorrerá quando um número qualquer de bases for adicionado ou deletado, com exceção dos múltiplos de três, o que iria restabelecer a fase

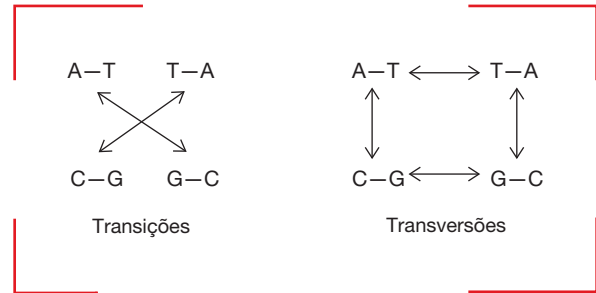


Figura 7.1

Diagramas mostrando as possíveis substituições de bases no DNA, incluindo quatro transições (diagrama à esquerda) e oito transversões (diagrama à direita).

de leitura original. É possível que um dos muitos códons alterados seja um UAA, UAG ou UGA, que são códons de parada (ver Capítulo 12). Quando um desses códons é encontrado durante a tradução, a síntese do polipeptídeo é interrompida naquele ponto. O resultado de uma mutação, que modifica a fase de leitura, pode ter consequências graves, principalmente se ela ocorrer nas porções iniciais da sequência codificadora.

1.2.3 Classificação por efeitos fenotípicos

Dependendo do seu tipo e localização, as mutações podem ter uma ampla faixa de efeitos fenotípicos, desde nenhum até letais. Uma **mutação de perda de função** é aquela que reduz ou elimina a função do produto gênico. Qualquer tipo de mutação, desde uma mutação pontual até a deleção total do gene, pode levar à perda de função. Mutações que resultam na perda total da função são denominadas **mutações nulas**. Uma mutação de perda de função com efeito dominante pode resultar na presença de um produto proteico defeituoso, que interage com o produto gênico normal ou inibe sua ação, estando, também, presente no mesmo organismo. Uma **mutação de ganho de função**, por outro lado, resulta em um produto gênico com função aumentada ou inteiramente nova. Isso pode ser devido a uma troca na sequência de aminoácidos da proteína, resultando em uma nova atividade, ou o resultado de uma mutação na região reguladora do gene, elevando seu nível de expressão ou a síntese de seu produto em momentos ou locais celulares inapropriados.

Entretanto, como os genomas eucarióticos consistem, principalmente, de regiões não codificadoras (ver Capítulo 5), a maioria das mutações têm alta probabilidade de ocorrer nas grandes porções do genoma, desprovidas de genes. Essas mutações são consideradas **mutações neutras**, caso elas não afetem produtos gênicos ou a expressão dos genes.

As mutações observadas com mais facilidade são as que afetam uma característica morfológica. Essas mutações são denominadas **mutações visíveis** e são reconhecidas pela sua capacidade de alterar um fenótipo

normal ou de tipo selvagem. Exemplos clássicos dessas mutações são todas as características observadas por Mendel em seus experimentos com ervilhas-de-jardim, como a forma da semente (lisa ou rugosa), a cor da vagem (verde ou amarela), a altura da planta (alta ou baixa), entre outras.

Algumas mutações exibem efeitos bioquímicos ou nutricionais. Em bactérias e fungos, uma típica **mutação nutricional** resulta na perda da capacidade de sintetizar um aminoácido ou uma vitamina. Nos seres humanos, a doença das células falciformes e a hemofilia são exemplos de doenças resultantes de **mutações bioquímicas**. Embora essas mutações nem sempre afetem características morfológicas, elas podem ter efeito no bem estar e na sobrevivência do indivíduo afetado.

Existe, também, uma categoria de mutação que afeta os padrões de comportamento de um organismo, por exemplo, o comportamento de acasalamento ou o ritmo circadiano de um animal. O efeito primário dessas **mutações comportamentais** é difícil de analisar.

Outros tipos de mutações incluem ainda: **mutações reguladoras**, que afetam a regulação da expressão de um gene, ativando ou reprimindo a sua expressão de forma imprópria, e as **mutações letais**, que podem interromper um processo essencial à sobrevivência do organismo. Dentro da categoria de mutações letais, a maioria das mutações utilizadas na análise genética de processos biológicos são as **mutações letais condicionais**. Essas mutações são letais em um determinado ambiente, nas denominadas **condições restritivas**, mas não letais em um outro ambiente, com **condições permissivas**.

As três maiores classes de mutantes com mutações letais condicionais são: (1) **mutantes auxotróficos**, incapazes de sintetizar um metabólito essencial (um aminoácido, uma purina, uma pirimidina, etc.) e que são sintetizados novamente pelos indivíduos de tipo selvagem da espécie. Esses mutantes se reproduzem quando o metabólito é fornecido pelo meio (condição permissiva) e não se reproduzem quando o metabólito está ausente (condição restritiva); (2) **mutantes sensíveis à temperatura**, que sobrevivem em uma determinada temperatura, mas não em outra. A sensibilidade à temperatura resulta de um aumento na sensibilidade ao calor ou ao frio do produto gênico mutado (por exemplo, uma enzima que é ativa em temperaturas baixas, mas é parcial ou totalmente inativa em temperaturas elevadas); e (3) **mutantes sensíveis ao supressor**, que sobrevivem quando um segundo fator genético, um **supressor**, está presente, mas não são viáveis na ausência dele. O gene supressor pode corrigir ou compensar um defeito no fenótipo causado pela mutação sensível ao supressor, ou pode restituir o produto gênico alterado pela mutação original.

1.3 Mutações espontâneas

Alterações espontâneas nas sequências de nucleotídeos dos genes não parecem ter uma causa conhecida. Ne-

nhum agente específico está associado com a ocorrência dessas mutações e, geralmente, assume-se que elas surjam como resultado de processos químicos ou biológicos normais do organismo, os quais alteram a estrutura das bases nitrogenadas.

1.3.1 Erros durante a replicação do DNA

Como visto no Capítulo 6, o processo de replicação do DNA não é perfeito. Algumas vezes, a DNA-polimerase insere nucleotídeos incorretos durante a síntese de uma fita de DNA. Embora as DNA-polimerases possam corrigir esses erros de replicação, utilizando a sua capacidade inerente de correção de erro no sentido 3'→5', nucleotídeos incorporados, de forma errônea, podem persistir após a replicação. Esses erros podem causar mutações caso não sejam detectados e corrigidos pelos mecanismos de reparação do DNA. Erros de replicação, devido a mal pareamentos, originam mutações pontuais. O fato de que as bases do DNA podem assumir várias formas, conhecidas como **tautômeros**, aumenta a chance de ocorrência de mal pareamentos durante a replicação do DNA.

1.3.1.1 Deslizes de replicação

Além das mutações pontuais, a replicação do DNA pode levar à introdução de pequenas inserções ou deleções de nucleotídeos, ocasionando os chamados **deslizes de replicação**. Essas alterações podem ocorrer quando uma fita do DNA-molde forma uma alça e é deslocada durante a replicação, ou quando a DNA-polimerase desliza ou para em um local durante a replicação. Caso uma alça ocorra na fita-molde, a DNA-polimerase pode passar direto pelos nucleotídeos que formam a alça e uma pequena deleção será introduzida na fita recém-sintetizada. Caso a DNA-polimerase introduza, de forma repetida, nucleotídeos que não estão presentes na fita-molde, a inserção de um ou mais nucleotídeos ocorrerá, criando uma alça não pareada na fita recém-sintetizada. Essas inserções e deleções podem provocar mutações que alteram a fase de leitura ou adições e deleções de aminoácidos no produto gênico.

Deslizes de replicação podem ocorrer em qualquer local do DNA, mas são mais frequentes em regiões que contêm sequências repetidas. Sequências repetidas são sítios quentes (ver adiante) para mutações no DNA e, em alguns casos, elas contribuem para doenças hereditárias. A hipermutabilidade de sequências repetidas em regiões não codificadoras do genoma é a base para os métodos atuais de análise forense do DNA.

1.3.1.2 Modificações tautoméricas

Um mecanismo para a ocorrência de mutações espontâneas foi sugerido por Watson e Crick, em 1953, no seu trabalho sobre a hélice dupla do DNA. Eles notaram que alguns átomos de hidrogênio podem mover-se de uma posição para a outra nas purinas e pirimidinas, por exemplo, de um grupamento amínico para um anel nitrogenado. Essas flutuações químicas são denominadas **modificações tautoméricas**. Embora as modificações tautoméricas sejam raras, elas podem ser de considerá-

vel importância no metabolismo do DNA, uma vez que alteram o pareamento de bases normal, em que a adenina pareia com a timina e a guanina com a citosina. Poucas vezes, porém, as formas **cetônicas** mais estáveis da timina e da guanina, e as formas **amínicas** da adenina e da citosina podem sofrer modificações tautoméricas para, respectivamente, as formas **enólicas** e **imínicas**, menos estáveis (**Figura 7.2**). Mesmo que essas formas existam por curtos períodos, se as bases estiverem na sua forma rara, no momento em que a cadeia estiver sendo replicada e forem incorporadas à cadeia de DNA nascente, pode ocorrer uma mutação. Nos seus estados menos estáveis, imínico e enólico, a adenina pode parear com a citosina e a timina, com a guanina (**Figura 7.3A**). Por exemplo, se o tautômero imínico da adenina parear com uma citosina será formado o par A^{*}C (o asterisco denota a forma tautomérica), incorporando, assim, uma citosina na cadeia de DNA no lugar reservado a uma timina. No próximo ciclo de replicação, a adenina provavelmente voltará a sua forma amínica estável e irá parear com uma timina, e a cito-

sina, incorporada erroneamente em uma das fitas da cadeia, será mantida e irá parear com uma guanina. Assim, uma das cadeias-filha da molécula de DNA irá conter um par GC no lugar do par de bases normal AT (**Figura 7.3B**).

As mutações resultantes de modificações tautoméricas nas bases do DNA envolvem a substituição de um par de bases por outro e são o tipo mais comum de mutação pontual, criando transições ou transversões.

1.3.2 Lesões hidrolíticas: desaminação e depurinação

As mutações espontâneas no DNA surgem não apenas por erros durante a sua replicação, mas também, por lesões provocadas por agentes mutagênicos ambientais. Apesar da própria estrutura de hélice dupla do DNA depender de um ambiente aquoso, o DNA pode sofrer lesões pela ação da água (**Figura 7.4A**). O tipo mais frequente e importante de lesão hidrolítica é a **desaminação** da base citosina. Sob condições fisiológicas normais, a cito-

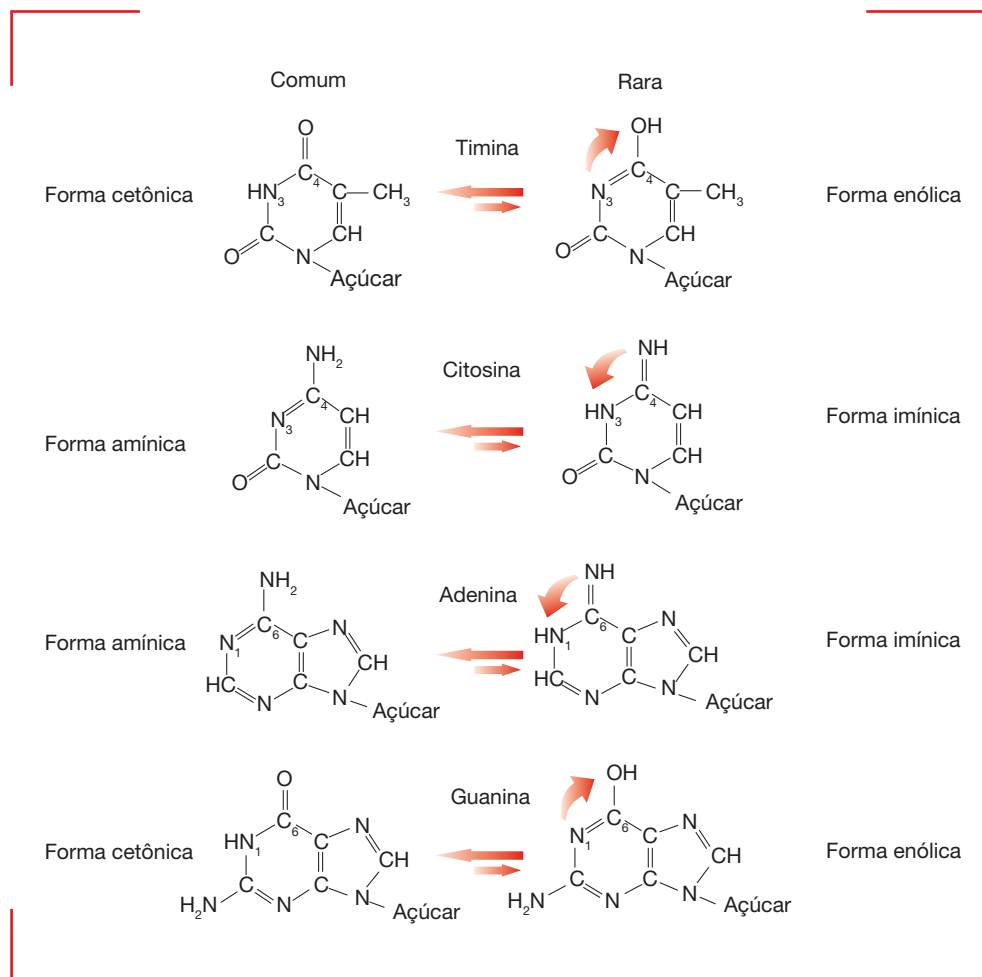


Figura 7.2

Formas tautoméricas das quatro bases comuns no DNA. A troca de átomos de hidrogênio entre as posições de número 3 e 4 das pirimidinas e entre as posições de número 1 e 6 das purinas, mostradas pelas setas, modifica o pareamento normal das bases.

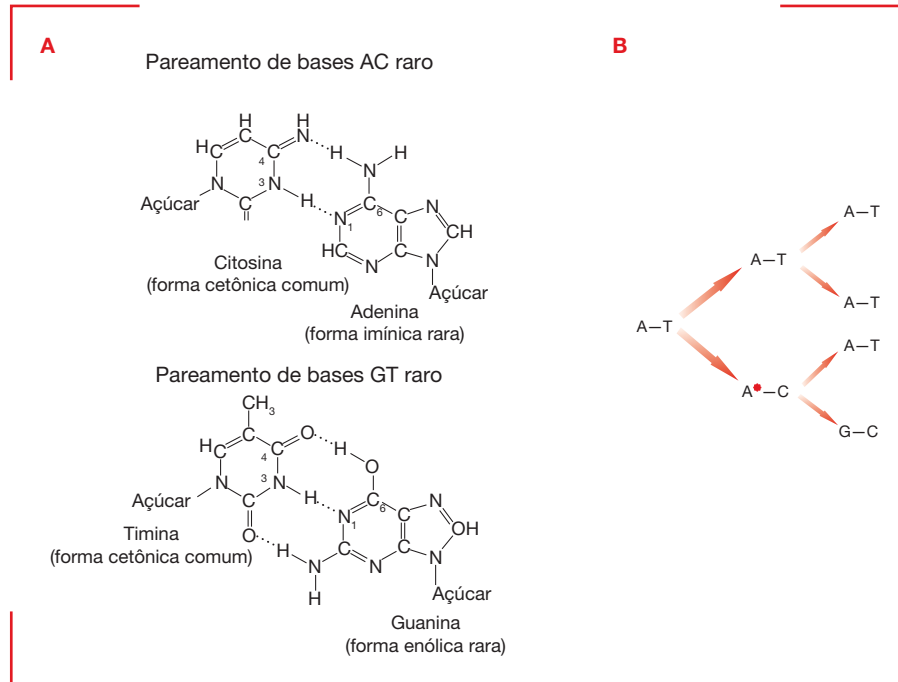


Figura 7.3

Modificações tautoméricas e pareamento das bases. (A) Exemplos de pareamento incorreto entre bases (AC e GT), que podem resultar da ocorrência de uma das formas tautoméricas raras das purinas (no exemplo, a adenina na sua forma imínica rara e a guanina na sua forma enólica rara, no momento da replicação). **(B)** O pareamento do tautômero raro da adenina (A*) com a citosina leva à formação do par de bases GC na geração seguinte.

sina sofre desaminação espontânea, perdendo seu grupamento amínico e originando uracila, que não é uma base natural do DNA. A uracila realiza pareamento com a adenina, introduzindo-a na fita oposta após a replicação, em vez de guanina, que deveria ter sido pareada com a citosina original. Adenina e guanina também estão sujeitas à desaminação espontânea. A desaminação converte adenina em hipoxantina, que faz pontes de hidrogênio com a citosina, em vez de com a timina; a guanina é convertida em xantina, que continua a parear com a citosina, embora por apenas duas pontes de hidrogênio.

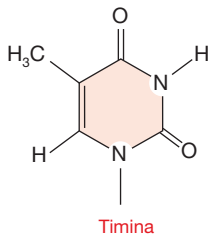
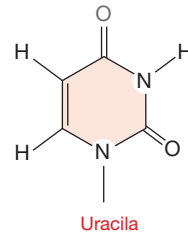
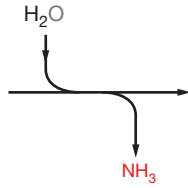
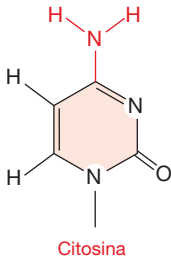
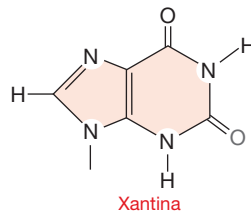
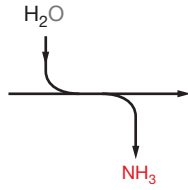
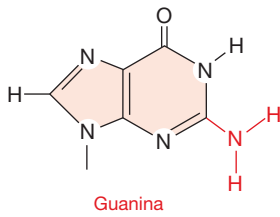
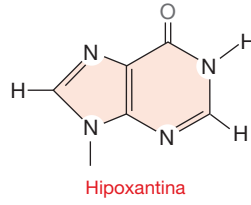
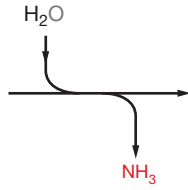
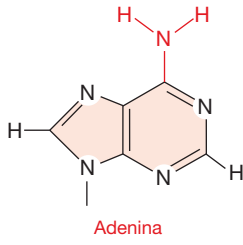
O dano gerado a partir da formação de uma base de ocorrência natural por desaminação é ilustrado por meio do problema causado pela presença de **5-metil-citosina** (Figura 7.4B). O DNA dos vertebrados frequentemente contém 5-metil-citosina no lugar de citosina, como resultado da ação de metil-transferases. Essa base modificada desempenha um papel importante na regulação da transcrição de muitos genes. Sítios contendo 5-metil-citosina constituem sítios quentes para mutações pontuais espontâneas. A existência destes sítios quentes decorre do fato de que a 5-metil-citosina sofre desaminação espontânea em uma frequência apreciável; e a substituição do grupamento amínico por um

grupamento cetônico converte a 5-metil-citosina em timina. Como essa base é um dos constituintes normais do DNA, o sistema de reparação existente para remover bases incomuns no DNA não detecta a alteração, resultando em uma mutação do tipo transição. A conversão cria um par anômalo GT, cuja separação, na replicação subsequente, produz um par GC de tipo selvagem e um par AT mutante.

O DNA também sofre depurinação pela hidrólise espontânea da ligação N-glicosídica. **Depurinação** é a perda de uma das bases nitrogenadas em uma cadeia de DNA de fita dupla intacta. Com frequência, a base perdida é uma purina – seja a adenina ou a guanina. As bases podem ser perdidas quando a ligação glicosídica entre os átomos de carbono 1' da desoxirribose e o da posição número 9 do anel purínico é rompida, criando um **sítio apurínico** em uma das fitas do DNA (**Figura 7.5**). Estima-se que centenas dessas lesões espontâneas sejam formadas diariamente no DNA de células humanas em cultura. Caso os sítios apurínicos não sejam reparados, não haverá uma base naquela posição para atuar como molde durante a replicação do DNA, fazendo com que a DNA-polimerase introduza um nucleotídeo de forma aleatória naquele sítio.

A

BASES NATURAIS



SEM DESAMINAÇÃO

B

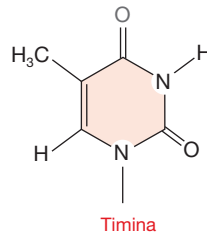
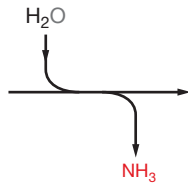
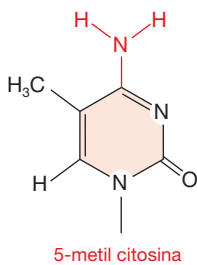


Figura 7.4

Alterações das bases por desaminação. (A)

Resultado da desaminação espontânea das bases comuns, adenina, guanina e citosina, originando as bases incomuns, hipoxantina, xantina e uracila, no DNA. A timina, por não possuir um grupamento amínico, não sofre desaminação. (B) A desaminação de 5-metil-citosina gera timina, que é uma base comum do DNA.

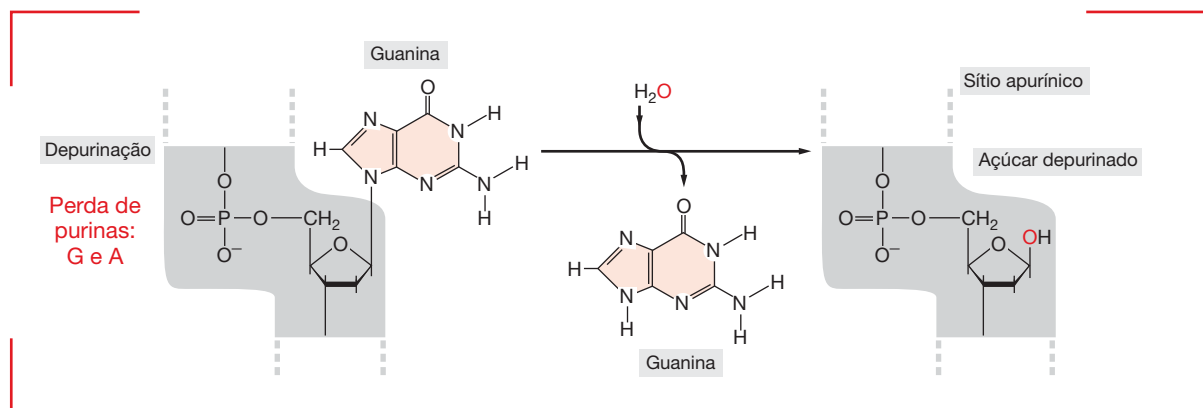


Figura 7.5

Resultado da depurinação. A base guanina é perdida, criando um sítio apurínico na cadeia de DNA.

1.3.3 Lesões oxidativas

O DNA também pode sofrer lesões em decorrência de produtos secundários dos processos celulares normais. Esses produtos secundários incluem espécies reativas de oxigênio (oxidantes eletrofilicos), geradas durante a respiração aeróbia normal. Por exemplo, radicais superóxidos (O_2^-), hidroxílicos (OH) e peróxido de hidrogênio (H_2O_2) são criados durante o metabolismo celular, sendo ameaças constantes à integridade do DNA. Esses **oxidantes reativos**, que também são gerados pela exposição à radiação de alta energia, podem produzir mais de 100 diferentes tipos de modificações às bases, que podem causar mal pareamentos durante a replicação.

1.4 Frequência de mutações

Muitas generalizações podem ser feitas em relação às taxas de mutações espontâneas nos organismos. A taxa de mutação espontânea é bastante baixa para todos os organismos, mas ela pode variar consideravelmente nos diferentes organismos, com valores que vão desde cerca de 10^{-4} por par de bases por geração para determinados vírus de RNA, de 10^{-6} a 10^{-8} para a maioria das bactérias, até 10^{-9} para células humanas. Mesmo dentro de uma mesma espécie, a taxa de mutação espontânea pode variar de gene para gene.

Em geral, a frequência de mutações cresce com o aumento do tamanho populacional e o decréscimo do conteúdo de informação do genoma, o que resulta em uma taxa mutacional muito estável, aproximadamente 1/300 mutações não neutras por genoma por geração. Entretanto, esse tema se torna mais complicado pelo fato das taxas de mutações variarem por todo o genoma. Estudos iniciais no repressor Lac de *E. coli* revelaram uma variação significativa na taxa de mutação por toda a extensão do gene, ao passo que estudos genômicos recentes sobre mutações silenciosas em seres humanos revelaram a presença de **sítios quentes** (*hot spots*) e **sítios frios** (*cold spots*) que cobrem centenas de quilobases. A denominação de sítio quente foi

atribuída para os sítios nos quais ocorre um número de mutações maior do que o esperado para uma distribuição aleatória – por exemplo, de 10 a até 100 vezes a mesma mutação no mesmo sítio. Em oposição, sítios frios são locais do genoma onde a frequência de mutações é menor do que a esperada para uma ocorrência aleatória. Em alguns genomas pequenos de bactérias, alguns sítios quentes mutacionais foram relacionados a sequências específicas de DNA, como repetições invertidas ou em tandem.

As razões para a variação nas frequências de mutações nos genomas não são claras, entretanto, a variação entre organismos pode refletir as eficiências relativas de seus mecanismos de revisão de leitura e de reparação do DNA.

1.5 Mutações induzidas

Todas as células estão expostas a um número enorme de agentes mutagênicos, com o potencial de danificar o DNA e provocar mutações. Alguns desses agentes, como algumas toxinas de fungos, raios cósmicos e luz ultravioleta são componentes naturais do ambiente e outros, incluindo alguns poluentes industriais, raios X médicos e produtos químicos (dentre esses a fumaça do cigarro) podem ser considerados como contribuições não naturais, provenientes do mundo moderno. Nesta seção, será analisada a forma como alguns desses agentes, naturais ou não, causam mutações. O **Material Complementar Online 7.1** descreve uma técnica rápida e de baixo custo, o Teste de Ames, que é utilizada para detectar a potencialidade mutagênica de um determinado composto químico.

1.5.1 Mutações induzidas quimicamente

1.5.1.1 Análogos de bases

Análogos de bases são substâncias mutagênicas que possuem estrutura suficientemente similar às bases normais para serem metabolizadas e incorporadas ao DNA durante a replicação. Contudo, elas são diferentes o su-

ficiente para aumentar a frequência de pareamento incorreto e, portanto, de mutação. As duas bases análogas mais utilizadas são a **5-bromouracila** e **2-aminopurina**. A 5-bromouracila (BrdU) é uma análoga da timina que, em geral, pareia com a adenina. No entanto, a proporção da forma tautomérica enólica da BrdU é maior do que a da timina, provavelmente pelo fato de o átomo de bromo ser mais eletronegativo do que o grupamento metílico do carbono 5. A forma enólica da BrdU pareia com a guanina, causando a transição $AT \leftrightarrow GC$ (Figura 7.6), e a 2-aminopurina, em sua maioria, pareia com timina. Ao contrário da adenina, o tautômero raro da 2-aminopurina pode formar uma única ponte de hidrogênio com a citosina e, assim, a 2-aminopurina pode produzir a transição $AT \leftrightarrow GC$.

1.5.1.2 Ácido nitroso, corantes de acridina e agentes alquilantes

Existem mutagênicos que atuam por meio de modificações químicas das bases do DNA. Por exemplo, o **ácido nitroso** reage com as bases que possuem grupamentos aminícos. Como visto anteriormente, a adenina é desaminada oxidativamente para a **hipoxantina**, a citosina

para a **uracila** e a guanina para a **xantina** (ver Figura 7.4). Devido às alterações de pareamento provocadas, o ácido nitroso causa transições $AT \leftrightarrow GC$. A **hidroxilamina** (NH_2OH) é um mutagênico específico, reagindo, quase que exclusivamente, com a citosina para gerar um derivado que pareia com a adenina, em vez de com a guanina. A hidroxilamina produz uma transição unidirecional de CG para AT .

Um tipo diferente de mutação é produzido por moléculas de corantes aromáticos como as **acridinas**. Estes compostos intercalam-se no DNA, isto é, eles introduzem-se entre pares de bases adjacentes da hélice dupla e, em consequência, causam a inserção ou a deleção de um ou mais pares de bases. O efeito dessas mutações é a alteração da fase de leitura na tradução, a menos que um número múltiplo de três seja inserido ou deletado (ver Capítulo 12). Foi a análise desses mutantes que revelou que o código genético era decifrado em trinca de nucleotídeos.

Um outro tipo de dano resulta da **alquilação**, transferência de grupamentos metílico ou etílico para os sítios

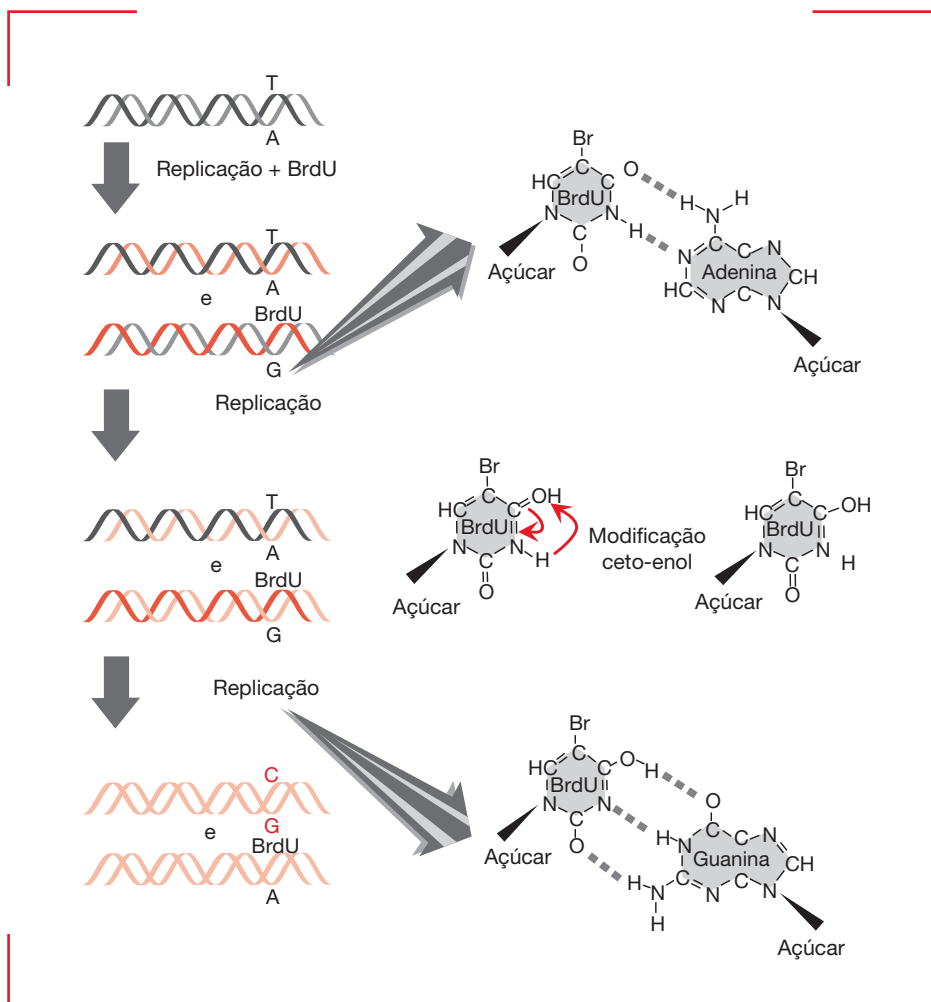


Figura 7.6

A base nitrogenada 5-bromouracila (BrdU), análoga da timina, ocasionalmente pareia com a guanina em vez de com a adenina. A presença do átomo de bromo (Br) no carbono 5 aumenta a proporção do tautômero raro, formado pelo deslocamento de um próton do nitrogênio 3 para o átomo de oxigênio do carbono 4, causando a transição $AT \leftrightarrow GC$.

reativos das bases e dos fosfatos da cadeia de DNA. Alquilantes químicos incluem **nitrosaminas** e um agente mutagênico de laboratório muito potente, a **metil-nitrosoguanidina**. Um dos sítios de alquilação mais vulneráveis é a base guanina, sujeita à metilação do oxigênio ligado ao seu carbono 6. O produto dessa metilação, a **O⁶-metil-guanina**, com frequência, parecia incorretamente com a timina, trocando GC por AT quando o DNA é replicado (**Figura 7.7**).

1.5.2 Mutações induzidas por radiações

A porção do espectro eletromagnético, contendo comprimentos de onda (λ) menores e de maior energia que a luz visível ($\lambda \cong 0,1 \mu\text{m}$), pode ser subdividida em **radiações ionizantes** (raios X, raios gama e raios cósmicos) e **radiações não ionizantes** (luz ultravioleta). As radiações ionizantes, como os raios X ($\lambda = 0,1 - 1 \text{ nm}$), têm muita energia e, portanto, são úteis para o diagnóstico médico, já que podem penetrar nos tecidos vivos. No processo de penetração, estes raios de muita energia colidem com os átomos da matéria e causam a liberação de elétrons, deixando radicais livres (íons) positivamente carregados. Estes íons, por sua vez, colidem com outras moléculas, provocando a liberação de outros elétrons. O resultado final é a formação de uma “trilha” de íons ao longo do caminho de cada raio de alta energia por meio da matéria ou tecido vivo. Esse processo de ionização é induzido por máquinas que produzem raios X, prótons e nêutrons, assim como pela radiação α , β e γ liberada pelos isótopos radioativos dos elementos (por exemplo, ^{32}P , ^{35}S , ^{60}Co , ^{137}Cs , etc.). Os raios ultravioleta (UV), que têm pouca energia, penetram apenas nas camadas de células superficiais de vegetais e animais superiores e não induzem ionizações.

Os raios UV dissipam sua energia para os átomos que encontram, elevando os elétrons de orbitais exteriores para os níveis de alta energia, um estado denominado **excitação**. As moléculas contendo átomos na forma iônica ou no estado excitado são quimicamente mais reativas do que as que contêm átomos no seu estado normal. A reatividade elevada dos átomos presentes nas moléculas de DNA é a base dos efeitos mutagênicos da luz UV e da radiação ionizante.

1.5.2.1 Radiações ionizantes

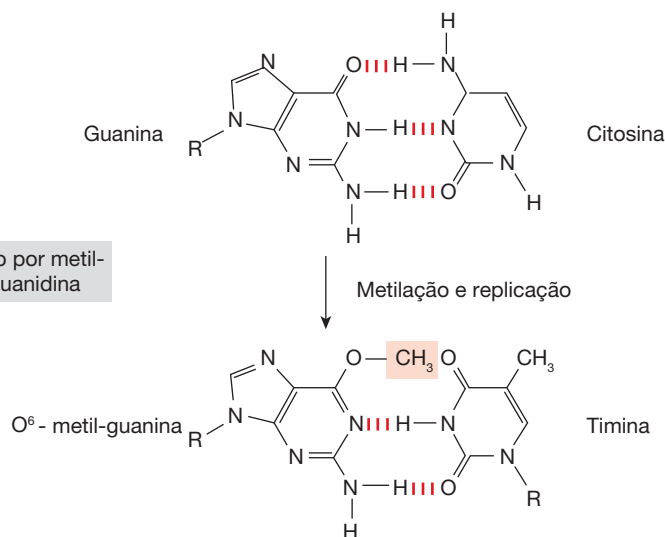
Os raios X e a maioria das outras formas de radiação ionizante são quantificados em unidades **roentgen** (r), em que uma unidade r equivale a radiação ionizante que produz uma unidade eletrostática de carga em um cm^3 de volume. A dose de irradiação em unidades r não envolve uma extensão de tempo. A mesma dose pode ser obtida por uma baixa intensidade de irradiação por um longo período de tempo ou, uma alta intensidade de irradiação por um curto período. Esse aspecto é muito importante porque, em geral, a frequência de mutações pontuais é diretamente proporcional à dose de irradiação. A **Figura 7.8** mostra um gráfico da porcentagem de mutações induzidas recessivas letais ligadas ao cromossomo X, contra a dose de raios X administrada. Existe uma relação linear entre a dose de raios X e a indução de mutações; para cada duplicação da dose, o dobro de mutações é induzido. Uma vez que a linha se aproxima do zero no eixo das abcissas, o gráfico sugere que mesmo uma dose pequena de radiação seja mutagênica.

1.5.2.2 Radiação ultravioleta

Os raios UV não possuem energia suficiente para induzir ionizações, entretanto, são absorvidos por certas

Figura 7.7

A reação da guanina com um agente alquilante cria O⁶-metil-guanina, que parece com a timina, trocando GC por AT quando o DNA é replicado.



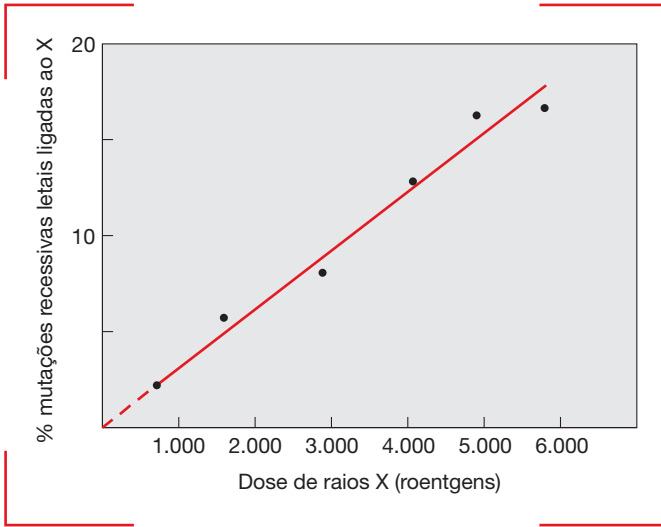


Figura 7.8

Relação entre a frequência de mutações recessivas letais ligadas ao cromossomo X induzidas e a dose de irradiação ionizante aplicada em *D. melanogaster*. Se extrapolado, o gráfico cruza os eixos no ponto zero, como demonstrado pela linha pontilhada.

substâncias, como purinas e pirimidinas que, por sua vez, ficam mais reativas ou no estado excitado. Nos organismos multicelulares, os raios UV atingem apenas a camada de células superficiais. Nos seres unicelulares, no entanto, os raios UV são um potente agente mutagênico. A absorção máxima de UV pelo DNA ocorre no comprimento de onda de 254 nm. O comprimento de onda de 254 nm também é o que possui a capacidade mutagênica máxima, sugerindo que o processo de mutação induzida por UV é mediado pela sua absorção por purinas e pirimidinas. Estudos *in vitro* mostram que as pirimidinas, principalmente a timina, absorvem ativamente a luz UV com comprimento de onda de 254 nm e, como resultado, se tornam muito reativas. Os dois

principais produtos da absorção de UV pelas pirimidinas são os **hidratos de pirimidinas** e os **dímeros de pirimidinas** (Figura 7.9).

A relação entre a dose de luz UV e a frequência de mutação é muito variável, dependendo do tipo de mutação, do organismo e das condições vigentes. Uma série de evidências indica que, provavelmente, a dimerização de timinas é o principal efeito mutagênico da luz UV. Os dímeros de timinas parecem causar mutações de duas formas indiretas: os dímeros alteram a hélice dupla de DNA, interferindo na replicação precisa da molécula; e erros ocasionais ocorrem durante os processos de reparação do DNA danificado.

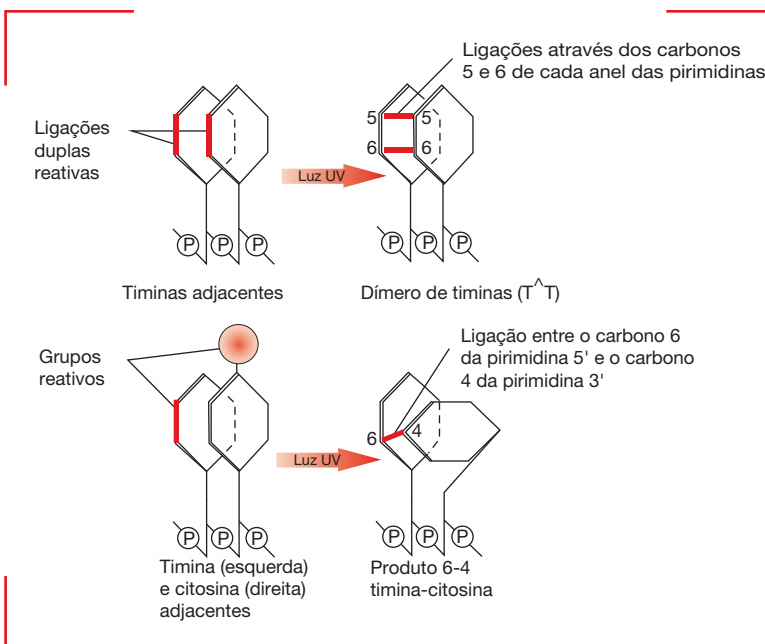


Figura 7.9

Dois importantes fotoprodutos da radiação UV: dímero de timinas e 6-4 timina-citosina.

1.6 Mutações no homem

1.6.1 Doença das células falciformes

Um exemplo bem conhecido e bastante estudado de mutação na espécie humana envolve a molécula de hemoglobina. A hemoglobina é uma macromolécula transportadora de oxigênio, presente nos glóbulos vermelhos do sangue. A hemoglobina é uma proteína complexa, constituída por quatro cadeias da proteína globina combinadas com o pigmento férrico heme. Dois tipos principais de hemoglobina estão presentes no homem em diferentes estágios de seu desenvolvimento: a hemoglobina fetal (HbF) e a hemoglobina do adulto (HbA). A HbA contém duas cadeias α idênticas, compostas por 141 aminoácidos cada uma, e duas cadeias β , também idênticas, compostas, cada uma, por 146 aminoácidos. Na HbF, duas cadeias γ estão presentes no lugar das cadeias β e possuem igualmente 146 aminoácidos. Devido às similaridades nas sequências de aminoácidos, acredita-se que todas as cadeias de hemoglobina e, portanto, os genes correspondentes, tenham evoluído de um ancestral comum (ver Capítulo 5).

Foram identificadas muitas variantes das hemoglobinas nas populações humanas, sendo a maioria de HbA. Algumas variantes apresentam sérios efeitos fenotípicos para os seus portadores, ao passo que outras apresentam efeitos menores.

A **hemoglobina da célula falciforme** (HbS) é uma das variantes de hemoglobinas. Os indivíduos homocigotos para o alelo da HbS (HbS β /HbS β , β indica que a HbS é uma variante da cadeia β) desenvolvem uma anemia hemolítica severa, chamada de **doença das células falciformes** (conhecida como anemia falciforme).

As moléculas de HbS precipitam na ausência de oxigênio, formando agregados cristaloides que distorcem a morfologia dos glóbulos vermelhos do sangue, alongando-se e ficando com a forma de foice. Estas células em forma de foice (falciformes) obstruem os pequenos vasos sanguíneos e impedem o transporte de oxigênio para os diversos tecidos. Na maioria dos casos, a doença das células falciformes provoca a morte durante a infância.

Quando as sequências de aminoácidos das cadeias β de HbA e HbS foram determinadas e comparadas, verificou-se que a cadeia β da HbS diferenciava-se em apenas um aminoácido (o sexto aminoácido da extremidade amínica) em relação à cadeia β da HbA. Na cadeia β da HbA este aminoácido é o ácido glutâmico (com carga negativa) e, na cadeia β da HbS ele é uma valina (sem carga). As cadeias α da HbA e HbS são idênticas, portanto, uma modificação em apenas um único aminoácido pode ter graves efeitos no fenótipo.

Atualmente, as sequências nucleotídicas dos genes Hb $^{\alpha}\beta$ e Hb $^{\delta}\beta$, de muitos indivíduos afetados, são conhecidas e, em sua maioria, a mutação no gene Hb $^{\alpha}\beta$, que dá origem ao gene Hb $^{\delta}\beta$, foi caracterizada como sendo a substituição de uma timina por uma adenina, trocando um par de bases AT por TA, uma transversão. Essa substituição de nucleotídeos provoca uma mutação com sentido trocado, em que o códon correspondente ao ácido glutâmico é substituído pelo códon correspondente a um resíduo de valina. No momento, são conhecidas cerca de 75 hemoglobinas variantes com aminoácidos trocados na cadeia β . Todas, exceto uma, diferem da cadeia β normal (HbA β) pela substituição de um único aminoácido. Alguns exemplos estão ilustrados na **Figura 7.10**.

Figura 7.10

Exemplos de variantes da cadeia β da hemoglobina que possuem a sequência de aminoácidos determinada. Em todos os casos, a sequência da cadeia β da hemoglobina A difere em apenas um aminoácido, como está indicado. Muitas das substituições envolvem mudanças de carga, o que facilita a detecção das hemoglobinas mutadas e com cargas alteradas por meio de eletroforese. Essas substituições provavelmente alteram mais a estrutura e a função da proteína, do que as que não envolvem mudança de carga.

	Posição do aminoácido													
	Val 1	His ⁺ 2	Glu ⁻ 6	Glu ⁻ 7	Gly 16	Gly 24	Glu ⁻ 26	Gly 56	His ⁺ 63	Lys ⁺ 95	Asn 108	Pro 124	Leu 141	
Hemoglobina A														
Hb variante														
Tokuchi		Tyr												
Hb S			Val											
Hb C			Lys ⁺											
Hb G				Gly										
J Baltimore					Asp ⁻									
Savannah						Val								
Hb E							Lys ⁺							
J Bangkok								Asp ⁻						
Zurich									Asp ⁺					
M Saskatoon										Tyr				
N Baltimore											Glu ⁻			
Yoshizuka												Asp ⁻		
Khartoum													Asp ⁺	
Olmsted													Asp ⁺	

1.6.2 Grupos sanguíneos ABO

O sistema de grupo sanguíneo ABO, descoberto por Karl Landsteiner, no começo do século XX, é ainda considerado o mais importante sistema de grupos sanguíneos na medicina clínica transfusional. Esse sistema se baseia em uma série de determinantes antigênicos encontrados nos eritrócitos e outras células, principalmente as células epiteliais. O locus ABO está localizado no braço longo do cromossomo 9. A heterogeneidade fenotípica do sistema sanguíneo ABO é devida à diferença estrutural do gene da glicosil-transferase, enzima responsável pela transferência dos resíduos específicos de açúcar, $\alpha 1 \rightarrow 3$ -N-acetilgalactosamina transferase ou $\alpha 1 \rightarrow 3$ -N-galactosil transferase, para a substância H, convertendo-a nos antígenos A ou B respectivamente.

Três alelos do gene *I* codificam a enzima glicosil-transferase. A substância H será modificada para um dos antígenos, A ou B, como resultado da ação dessa enzima, dependendo se a enzima foi codificada pelo alelo I^A ou I^B . A incapacidade de modificar a substância H resulta da presença do alelo nulo I^O .

Quando as sequências de DNA dos alelos I^A e I^B foram comparadas, quatro substituições de nucleotídeos, únicas e consistentes, foram encontradas. As trocas resultantes na sequência de aminoácido do produto do gene da glicosil-transferase originaram produtos gênicos com funções alteradas e, em consequência, a diferentes modificações na substância H.

Indivíduos homocigotos para o alelo I^O possuem grupo sanguíneo de tipo O, não apresentam atividade de glicosil-transferase e, por isso, não conseguem modificar a substância H. A análise da sequência de DNA do alelo I^O mostrou uma alteração que é única, comparada com as sequências dos outros alelos: a deleção de um só nucleotídeo no início da sequência codificadora, causando uma mutação que modifica a fase de leitura. Um RNA mensageiro completo é transcrito, mas, no momento de ser traduzido, a fase de leitura muda no local da deleção e continua em uma outra fase por cerca de 100 nucleotídeos antes de encontrar um códon de terminação. Nesse ponto, a cadeia polipeptídica da glicosil-transferase termina prematuramente, resultando em um produto que não é funcional.

Esses resultados fornecem uma explicação molecular direta sobre o sistema ABO e a base para a biossíntese dos antígenos correspondentes. A base molecular para os fenótipos antigênicos é relacionada a mutações de substituição de bases e de modificação de fase de leitura dentro do gene que codifica a enzima glicosil-transferase.

1.6.3 Distrofias musculares

Distrofias musculares são doenças genéticas caracterizadas por fraqueza e degeneração muscular progressiva. Existem muitos tipos de distrofias musculares que diferem em suas severidades, época de surgimento e causas genéticas. Duas formas relacionadas de distrofia mus-

cular – a **distrofia muscular de Duchenne** (DMD) e a **distrofia muscular de Becker** (BMD, de *Becker muscular dystrophy*) – são condições recessivas ligadas ao X. A DMD é a mais severa das duas doenças, com uma degeneração muscular progressiva rápida e o envolvimento do coração e dos pulmões. Meninos com DMD perdem a capacidade de caminhar por volta dos 12 anos de idade e podem morrer perto dos 20 anos. A incidência de 1 em 5.000 nascimentos de meninos vivos faz da DMD uma das doenças hereditárias, que encurtam o tempo de vida, mais comuns. Ao contrário, a BMD não envolve o coração e os pulmões, progredindo de forma lenta, a partir da adolescência até a idade dos 50 anos ou mais.

O gene responsável pela DMD e pela BMD, o *dystrophin*, é muito grande, consistindo em cerca de 2,5 milhões de pares de bases. Nos indivíduos normais (não afetados) a transcrição e o processamento subsequente do transcrito primário do *dystrophin* resultam em um mRNA, contendo cerca de 14.000 pb. Ele é traduzido na proteína **distrofina**, que contém 3.685 aminoácidos.

Recentemente, as bases moleculares das mutações que causam a DMD ou a BMD foram determinadas. Com poucas exceções, as mutações que causam a DMD alteram a fase de leitura do gene *dystrophin*, ao passo que as mutações causadoras de BMD, normalmente não. Estudos mostraram que cerca de dois terços das mutações no gene *dystrophin*, que levam à DMD ou à BMD, são deleções e inserções, sendo apenas um terço mutações pontuais. A maioria das mutações que causam a DMD leva à terminação prematura da tradução. Isto, por sua vez, leva à degradação do transcrito impropriamente traduzido do gene *dystrophin* e a quase que completa ausência da proteína distrofina. Ao contrário, a maioria das mutações que causa a BMD altera as sequências internas do transcrito do gene *dystrophin* e da proteína, mas não alteram a fase de leitura da tradução. Como resultado, uma proteína distrofina modificada, mas de alguma forma funcional, é produzida, evitando as consequências severas da DMD.

1.7 Aplicações práticas das mutações

Apesar de a maioria das mutações tornarem o organismo menos adaptado e serem, portanto, desvantajosas, a possibilidade de desenvolver novas características (desejáveis) por meio de mutações induzidas tem intrigado muitos pesquisadores. Existem relatos de mutantes induzidos de cevada, trigo, aveia, soja, tomate e árvores frutíferas, capazes de melhorar com rapidez as linhagens cultivadas. Foram obtidos, por exemplo, mutantes de cevada mais produtivos, com resistência à ferrugem, palha rija, maior quantidade de proteína e sementes sem casca.

As mutações também têm sido utilizadas para elucidar as vias pelas quais os processos biológicos ocorrem. A sequência de passos de uma rota metabólica pode ser determinada por meio do isolamento e do estudo das mutações nos genes que codificam as enzimas envolvidas.

Como uma mutação apropriada elimina a atividade de um único polipeptídeo, a mutação é uma ferramenta de grande utilidade, com a qual se pode dissecar os processos biológicos. Um exemplo desse tipo de aplicação das mutações ocorre quando se quer determinar a função de um gene recentemente identificado. A informação sobre a sua função pode, em princípio, ser obtida pela geração de um organismo mutante, no qual o funcionamento daquele gene esteja ausente.

2. Reparação do DNA

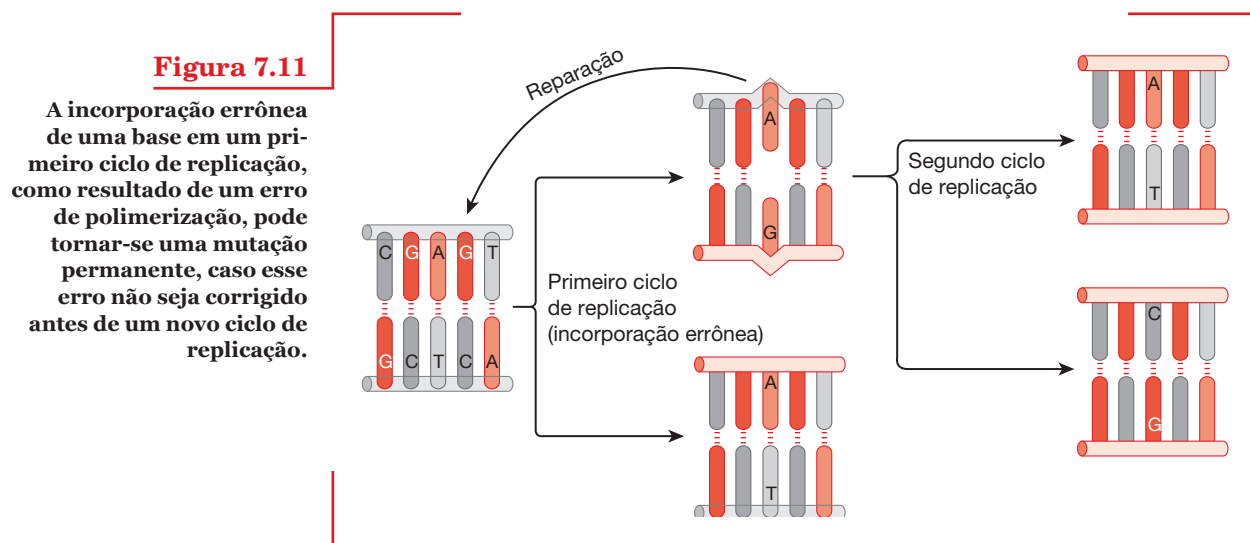
Um mutante sobrevive quando a alteração genética sofrida não é prejudicial ou, em alguns casos raros, é benéfica. A maioria das mutações é desvantajosa e as células não conseguiriam sobreviver se não possuísem mecanismos enzimáticos para reverter os efeitos dos processos mutagênicos, naturais ou artificiais. Como visto anteriormente, as bases do DNA podem ser alteradas ou perdidas, as ligações fosfodiéster da cadeia podem ser quebradas e as fitas livres podem ser cruzadas e covalentemente ligadas. Essas lesões são produzidas por radiações ionizantes, luz ultravioleta e uma variedade de compostos químicos. Muitos dos danos sofridos pelo DNA podem ser reparados, pois a informação genética é preservada em ambas as fitas da hélice dupla, de forma que a informação perdida em uma fita pode ser recuperada a partir da fita complementar.

Os mecanismos existentes e conhecidos de reparação de DNA lesado são provavelmente comuns a todos os organismos, sendo que uma célula pode ter vários sistemas de reparação capazes de atuar ao mesmo tempo. Esses sistemas são, no entanto, melhor compreendidos e estudados em *E. coli*. Nesta seção, discutiremos os mecanismos de reparação do DNA conhecidos até o momento.

2.1 Revisão de leitura e reparação de mal pareamentos

Alguns dos tipos mais comuns de mutações surgem durante a replicação do DNA, quando um nucleotídeo incorreto é inserido pela DNA-polimerase (**Figura 7.11**). A enzima de bactérias (DNA-polimerase III) comete um erro em aproximadamente 100.000 inserções, originando uma taxa de erro na ordem de 10^{-5} . Porém, a própria enzima revisa a sua síntese, por meio da revisão de cada etapa, detectando 99% dos erros. Caso um nucleotídeo tenha sido erroneamente incorporado durante a polimerização, a enzima tem o potencial de reconhecer esse erro e “reverter” a sua direção. Ela utiliza, então, sua atividade de exonuclease $3' \rightarrow 5'$, removendo o nucleotídeo incorreto e substituindo-o pelo correto (Capítulo 6). Essa ação aumenta a precisão da replicação em uma centena de vezes, deixando apenas um nucleotídeo a cada 10^7 polimerizados, com uma taxa final de erro de 10^{-7} .

Para lidar com os erros que permanecem após a revisão da polimerização, *E. coli* possui o **sistema de reparação de mal pareamentos** (de *mismatch repair*, **Figura 7.12**), que confere precisão ao processo de replicação. Esse sistema faz com que a ordem final de ocorrência de um erro seja a cada 10^{10} ou 10^{11} nucleotídeos (nt). O sistema de reparação de mal pareamentos consiste de várias proteínas, codificadas pelos genes *mut*. Esse sistema percorre o DNA recentemente sintetizado a procura de pares de bases mal pareadas e remove os segmentos de fita simples contendo nucleotídeos incorretos, em um processo que envolve pelo menos quatro etapas. Isto permite que a DNA-polimerase insira a base correta na lacuna formada. A dificuldade que surge é a de distinguir qual das bases de um par erroneamente formado é a incorreta, porque ambas são componentes naturais do DNA. Se a remoção de uma das bases fosse ao acaso, haveria a probabilidade de 50% de a base correta ser removida e a mutação poderia ser perpetuada, ao invés de corrigida. Existe,



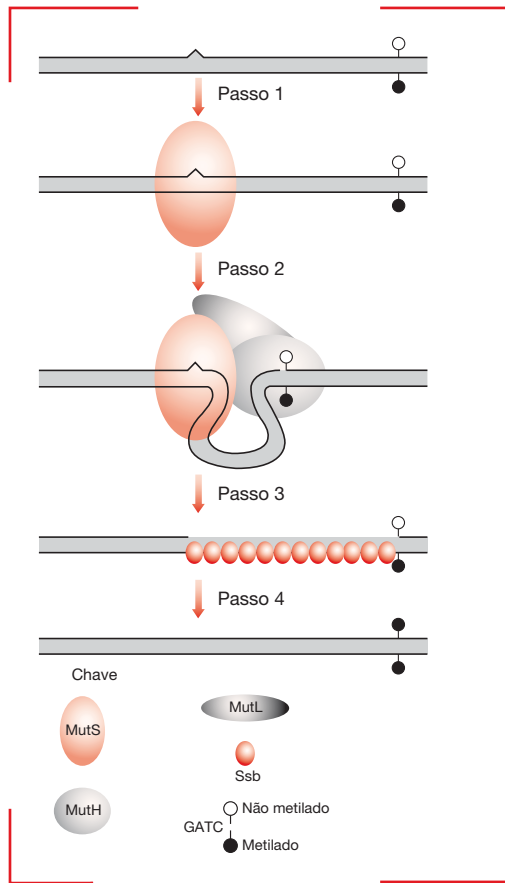


Figura 7.12

Representação esquemática do processo de reparação de bases mal pareadas em *E. coli*. Passo 1: ligação de MutS ao par mal pareado. Passo 2: MutS recruta MutH e MutL. A montagem do complexo torna a atividade endonucleásica latente de MutH funcional, que inicia o processo de excisão realizando uma clivagem na fita recém-sintetizada, que possui a sequência GATC não metilada. A degradação da fita contendo o nucleotídeo incorretamente incorporado inicia nesse corte e continua através e além das bases mal pareadas. Passo 3: a região de fita simples exposta é protegida da degradação pela proteína SSB. Passo 4: a lacuna é preenchida pela ação da DNA-polimerase III e a fita é unida pela DNA-ligase.

porém, um sinal específico, controlado temporalmente, que direciona o sistema de reparação de mal pareamentos, exclusivamente, para a fita recém-sintetizada. Este sinal implica no reconhecimento de sequências GATC próximas ao erro. Contudo, se uma metilase, codificada pelo gene *dam*, modificar primeiro a adenina da sequência GATC para N⁶-metil-adenina, esse sistema de reparação não irá mais atuar e a excisão não irá ocorrer. Muitas se-

quências GATC na célula modificam-se após sua síntese, mas somente após um pequeno espaço de tempo – alguns segundos ou minutos. Este tempo é suficiente para que o sistema de reparação ligue-se à fita recém-sintetizada, ao invés de ligar-se à fita-molde. O sistema de reparação detecta não somente pares únicos de bases mal pareadas, mas também adições e deleções, o que reduz a incidência de mutações por modificação na fase de leitura, além daquelas envolvendo substituição de bases.

As células eucarióticas apresentam proteínas ortólogas à MutS (chamadas MSH) e MutL (chamadas MLH e PMS) para corrigir mal pareamentos. Os eucariotos apresentam diversas proteínas semelhantes à MutS, com especificidades diferentes, por exemplo, uma delas é específica para mal pareamentos simples, ao passo que outras reconhecem pequenas inserções ou deleções, resultantes de deslizes durante a replicação do DNA.

Ainda que as células eucarióticas apresentem sistemas de reparação de mal pareamentos, elas não têm a endonuclease MutH, nem o artifício engenhoso de *E. coli* de utilizar a hemimetilação para marcar a fita parental. Existem evidências de que descontinuidades na nova fita, que está sendo sintetizada, podem estar implicadas no processo discriminatório. Antes da etapa de ligação, os fragmentos de Okazaki (ver Capítulo 6) estão separados do DNA já sintetizado por uma quebra, que pode ser considerada equivalente à quebra gerada por MutH na fita recém-sintetizada em *E. coli*. Resultados recentes indicam que os homólogos humanos de MutS (MSH) interagem com um componente do PCNA (ver Capítulo 6) e poderiam, portanto, ser recrutados para o sítio de síntese descontínua do DNA na fita tardia. A interação com componentes do PCNA também recrutaria as proteínas de reparação de mal pareamentos para a extremidade 3' (crescente) da fita sintetizada continuamente.

2.2 Reversão direta de lesão

2.2.1 Fotorreativação enzimática

Um dos mecanismos de reparação melhor estudado é o de remoção dos dímeros de pirimidinas, que, como visto, são formados pela exposição do DNA à luz UV. Resíduos de pirimidinas adjacentes podem tornar-se covalentemente ligados sob estas condições. Este dímero não se encaixa na hélice dupla e, assim, a replicação e a expressão gênica são bloqueadas até que o dímero seja removido. Este tipo de lesão pode ser reparado de diferentes formas. A forma mais direta e menos frequente envolve enzimas que reverter a modificação química que originou o dano.

Dímeros de pirimidinas são alvos da enzima **fotorreparação**, que se liga ao dímero e catalisa uma segunda reação fotoquímica, na presença da luz visível, desfazendo o anel ciclobutânico formado pela luz UV e refazendo as bases pirimídicas individuais. Este processo é chamado de **fotorreativação** e envolve duas etapas. Inicialmente, a enzima reconhece e liga-se especificamente ao dímero no escuro e, depois, a absorção de luz fornece

energia para converter o dímero de ciclobutila em monômeros de pirimidina, então, a enzima dissocia-se do DNA (**Figura 7.13**). Este sistema é comum na natureza e parece ser muito importante em vegetais. Em *E. coli* ele depende da ação do produto de um único gene (*phr*), que codifica a fotoliase.

2.2.2 Reparação enzimática de bases alquiladas

Em *E. coli*, bases alquiladas são removidas pela enzima **O⁶-metil-guanina-metil-transferase**, codificada pelo gene *ada*, que reconhece O⁶-metil-guanina na fita dupla de DNA e remove o grupamento metílico, transferindo-o para um resíduo de cisteína da enzima (**Figura 7.14**). A mesma enzima também pode remover grupamentos metílicos dos fosfatos que, possivelmente, interrompem a cadeia de DNA. Uma característica incomum desta reação é o fato de não existir meios de recuperar a enzima metilada, como acontece com a maioria das enzimas que são recicladas e ficam novamente funcionais ao final do processo de catálise. No caso da O⁶-metil-guanina-metil-transferase, para cada grupamento metilíco removido uma molécula de enzima é utilizada.

2.3 Reparação por excisão

A remoção de uma base defeituosa ou não habitual pode ser feita pela clivagem da ligação base-açúcar (excisão da base) ou, mais comumente (exceto para a uracila), pela liberação dos nucleotídeos por incisões endonucleolíticas nos dois lados da lesão, seguida pelo preenchimento da região pela ação da DNA-polimerase I e posterior ligação.

2.3.1 Excisão de base

O sistema de **reparação por excisão de base** (BER, de *base excision repair*) corrige lesões causadas nas bases nitrogenadas pela hidrólise espontânea ou por agentes que as alteram quimicamente, provocando, por exemplo, a oxidação. Esse processo também lida com uma variedade de outras lesões, incluindo bases desaminadas e quebras de fita simples (SSBs, de *single-strand breaks*).

As reações envolvidas no sistema de BER estão representadas na **Figura 7.15**. A primeira etapa do sistema BER em *E. coli* envolve o reconhecimento de uma base quimicamente alterada por **DNA-glicosilases**, que são específicas para diferentes tipos de lesões no DNA. Por exemplo, a citosina do DNA é desaminada, de forma espontânea, em uma frequência perceptível para uracila. A desaminação da citosina é um evento mutagênico potencial, porque a uracila parece com a adenina e, assim, uma das moléculas-filha conterá um par de bases AU no lugar do par original GC. Esta mutação é evitada, pois o sistema de reparação por excisão de base reconhece uracila como sendo uma base estranha ao DNA.

A primeira ação da **uracila-DNA-glicosilase**, no sistema BER, é hidrolisar a ligação glicosídica entre a uracila e as moléculas de desoxirribose. Neste estágio, a cadeia de DNA está intacta, mas sua base é perdida. Esta fenda é chamada de **sítio AP**, porque se torna **apurínico** (perde A ou G) ou **apirimidínico** (perde C ou T). A enzima **AP endonuclease** reconhece este defeito e cliva a cadeia de DNA na posição 5' do sítio AP, deixando um 3'-OH; e a exonuclease cliva a posição 3' do sítio AP, deixando um fosfato 5'. A lacuna resultante é preenchida pela DNA-polimerase I, que insere a citosina, de acordo com a orientação fornecida pela fita complementar não danificada contendo uma guanina. Finalmente, a integridade da fita corrigida é restaurada pela DNA-ligase. Além desse exemplo existem outras glicosilases que reconhecem e removem hipoxantina, 3-metil-adenina e purinas com anel imidazol aberto por radiação ionizante ou pH elevado.

Sistemas de reparação por excisão de base também existem em eucariotos, desde leveduras até seres humanos. Em seres humanos, a quebra de fita simples, resultante da ação da AP endonuclease, é reconhecida pela enzima PARP (poli-ADP-ribose-polimerase) que, uma vez ativada, sinaliza o dano para as demais proteínas envolvidas na reparação da quebra. Evidências experimentais mostraram que tanto células de camundongo como humanas, defectivas em atividade de BER, são hipersensíveis aos efeitos letais dos raios gama e agentes oxidantes.

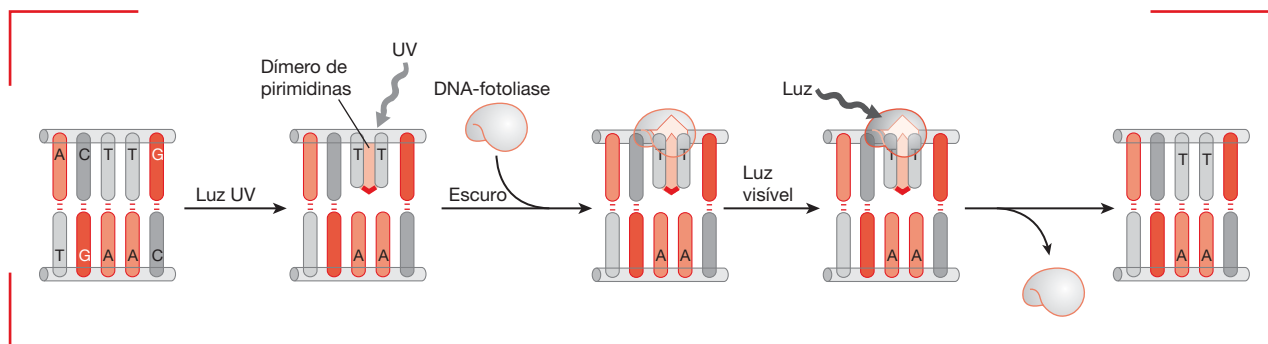
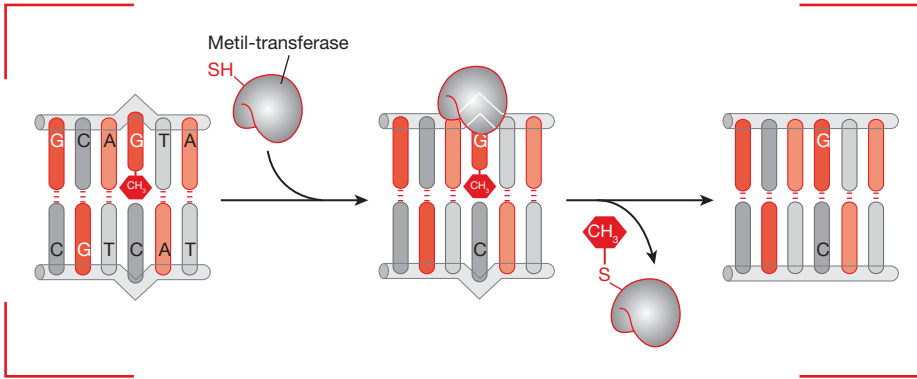
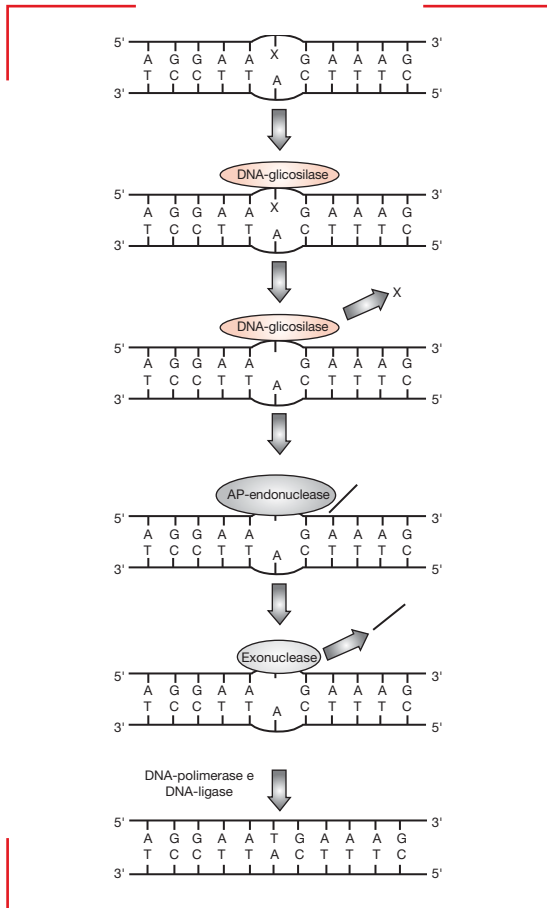


Figura 7.13

Mecanismo de reparação do DNA lesionado por luz UV, realizado pela enzima fotoliase na presença da luz visível (reparação por fotorreativação enzimática).

**Figura 7.14**

A metil-transferase catalisa a transferência do grupo metil da O⁶-metil-guanina para um resíduo de cisteína da enzima, restaurando a base guanina normal no DNA e provocando a inativação da enzima.

**Figura 7.15**

Sistema de reparação por excisão de base (BER). A enzima DNA-glicosilase reconhece uma base estranha no DNA e hidrolisa a ligação glicosídica. A base, então, é perdida, formando o sítio AP (apurínico ou apirimidínico), onde a enzima AP endonuclease reconhece o defeito e cliva a cadeia de DNA na posição 5' do sítio AP, deixando um 3'-OH. A exonuclease cliva a posição 3' do sítio AP, deixando um fosfato 5'. A lacuna resultante é preenchida pela DNA-polimerase I, tendo como molde a fita não danificada. A DNA-ligase restaura a integridade da fita.

2.3.2 Excisão de nucleotídeos

Um dos mecanismos de reparação mais importantes e gerais envolve as etapas conhecidas coletivamente sob o nome de **reparação por excisão de nucleotídeos (NER, de nucleotide excision repair)**. A **Figura 7.16** resume os componentes centrais envolvidos no processo de NER (em *E. coli*) e os seus respectivos papéis no mecanismo. As principais propriedades das proteínas-chave UvrA-D e dos dois fatores auxiliares, Phr e Mfd, são apresentadas na **Tabela 7.1**. Em todos os organismos em que já foi estudado, o processo de NER consiste de cinco etapas: reconhecimento da lesão; incisão da fita lesada em ambos os lados da lesão e a alguma distância dela; excisão do segmento (oligonucleotídeo) contendo a lesão; síntese de um novo segmento de DNA, utilizando a fita não danificada como molde; e, por fim, ligação. A reação é, em princípio, livre de erro. *E. coli* é o organismo em que todo o mecanismo é conhecido em detalhes. No entanto, um processo equivalente ao de NER também já foi encontrado em leveduras e em eucariotos mais complexos.

Em *E. coli*, a reação de NER inicia quando duas moléculas de UvrA dimerizam na presença de ATP e se complexam com uma molécula de UvrB. Este heterotrímero liga-se ao DNA e sua fraca atividade de DNA-helicase 5' → 3' permite sua movimentação ao longo de uma das fitas da hélice à procura de distorções na cadeia. Quando o complexo Uvr A₂B encontra uma lesão, a translocação é interrompida e UvrB fica “enganchada” no DNA, induzindo uma conformação específica da cadeia e formando uma região parcialmente desnaturada. As moléculas de UvrA se dissociam (essa dissociação requer ATP), permitindo que UvrC ligue-se à UvrB. O complexo UvrBC promove, então, duas incisões na fita danificada: a primeira, localizada à distância de 8 nt da extremidade 5' da lesão e a segunda, à distância de 4 a 5 nt da extremidade 3' da lesão. Essas incisões também requerem a utilização de ATP. As ações conjuntas da UvrD (que também é uma helicase) e da DNA-polimerase I promovem a excisão do segmento de 12 a 13 pb contendo a lesão, a liberação de UvrB e UvrC e a síntese de uma cópia não danificada da sequência que foi removida, tendo como molde a fita complementar intacta. Por fim, a DNA-ligase une as extremidades clivadas na cadeia reparada.

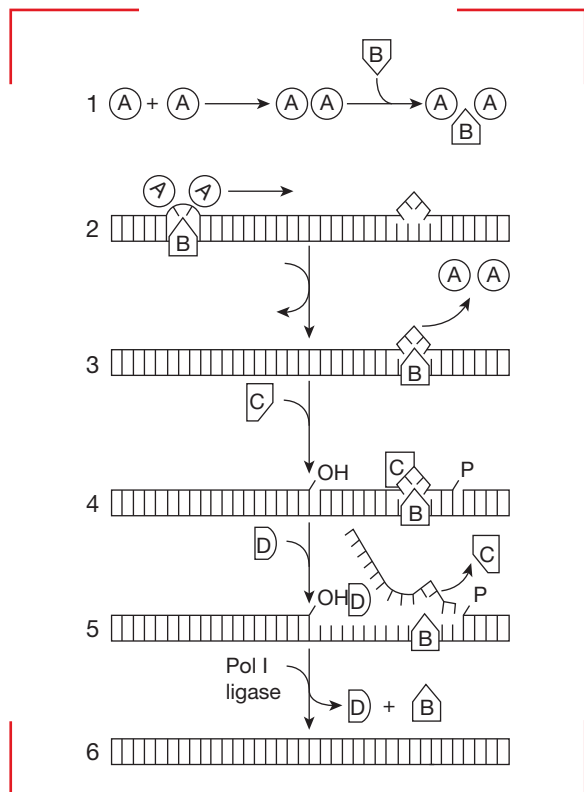


Figura 7.16

Modelo para o mecanismo de reparação por excisão de nucleotídeos em *E. coli*.

A, B, C e D representam as moléculas de UvrA, UvrB, UvrC e UvrD, respectivamente; Pol I, DNA-polimerase I; ligase, DNA-ligase. As fitas estão representadas por linhas e os estágios sequenciais do processo de NER estão numerados.

A extensão média do fragmento de DNA removido é de 12 nt, razão pela qual esse modelo recebe a denominação de **reparação de regiões curtas**. A enzima envolvida na síntese do novo segmento é, na maioria das vezes, a DNA-polimerase I, embora as DNAs-polimerases II e III possam substituí-la. Em caso de grandes quantidades de lesões, a reparação de regiões curtas contribui com 99% dos eventos de reparação por excisão. O 1% restante envolve a substituição de fitas de DNA com, em média, 1.500 nt de extensão, mas podendo chegar a tamanhos de 9.000 nt.

Esse modelo de reparação também requer a ação dos genes *uvr* e envolve a DNA-polimerase I, tendo sido denominado **reparação de regiões longas**. A diferença entre esses dois modelos de reparação é que a reparação de regiões curtas é uma função constitutiva da célula bacteriana, ao passo que a reparação de regiões longas deve ser induzida por lesões no DNA. O processo de reparação de regiões longas atua, provavelmente, sobre lesões localizadas em regiões próximas das forquilha de replicação. As diferenças entre esses dois modelos de reparação, por exemplo, quanto às proteínas envolvidas, ainda não foram caracterizadas.

O passo mais importante deste processo, a detecção da lesão pela discriminação entre estruturas de DNA normais e anormais, é realizado pela ação cooperativa das proteínas UvrA e UvrB. O complexo UvrA₂B é capaz de reconhecer uma grande variedade de estruturas danificadas (dímeros de pirimidinas, adutos químicos, ligações cruzadas, etc.), que estão presentes no DNA em quantidades muito baixas (no máximo 1 a cada 10⁶ nt). No entanto, nem todas estas lesões são identificadas com a mesma eficiência e algumas são substrato para mais de um sistema de reparação.

Tabela 7.1 Propriedades das principais proteínas do sistema de NER de *E. coli*^a

Proteína	Nº de aminoácidos	Nº de moléculas por célula ^b	Propriedades da proteína
UvrA	940	25 (250)	Liga-se ao DNA de fita dupla irradiado por UV. Forma um heterotrímero com UvrB.
UvrB	673	500 (2.000)	Liga-se à UvrA ₂ , exercendo atividade de helicase 5'→3' do complexo UvrA ₂ B. Faz a incisão na região 3'.
UvrC	610	10	Liga-se ao DNA de fita simples e dupla. Atividade de endonuclease no complexo com UvrB. Faz a incisão na região 5'.
UvrD (Helicase II)	720	3.000 (4.500)	DNA-helicase 3'→5' ou RNA-DNA-helicase. Liga-se aos nucleotídeos e ao DNA.
Phr	471	10-20	Fotorreativa aos dímeros de pirimidina. Estimula a reparação por UvrABC. Liga-se à FADH ₂ , pterina e DNA de fita de dupla irradiado por UV.
Mfd	1.148	?	Confere especificidade para a fita transcrita na reparação <i>in vitro</i> por UvrABC. Desloca a RNA-polimerase. Interage com UvrA.

^a A DNA-polimerase I e a DNA-ligase não estão incluídas.

^b O número de moléculas por célula após a indução SOS é mostrado entre parênteses.

O esquema mostrado na Figura 7.16 representa a reação central do processo de NER em *E. coli*, que pode ser realizada em um sistema *in vitro*, contendo uma cadeia de DNA danificada como substrato e as proteínas mencionadas anteriormente. Entretanto, na célula, fatores adicionais estão envolvidos na modulação do sistema para permitir sua ação sobre a cromatina, no caso de DNA eucariótico. Um exemplo interessante é o **sistema de reparação de genes ativos** ou **sistema de reparação fita-seletivo (TC-NER, de transcription-coupled excision repair)**, um subcaminho do processo de NER primeiramente identificado em mamíferos.

Em *E. coli*, foi demonstrado que, tanto *in vitro* como *in vivo*, o sistema de NER remove lesões que bloqueiam a transcrição da fita contendo um gene ativo, com mais rapidez do que da fita que não é transcrita ou do restante do genoma da célula. Esse sistema pode distinguir entre as duas fitas de um gene, que sítios mais críticos do genoma devem ser reparados primeiro, embora contenham o mesmo tipo de lesão encontrada em locais menos críticos do DNA. Este sofisticado processo requer ao menos um fator adicional: o fator de ligação reparação-transcrição, produto do gene *mfd* (ver Tabela 7.1).

In vitro, a proteína Mfd é capaz de dissociar uma molécula de RNA-polimerase parada em uma lesão e ligar-se à subunidade de UvrA do complexo de reparação. Alguns componentes auxiliares deste processo foram identificados, entre eles a enzima de fotorreativação, a fotoliase. Essa enzima, além de monomerizar dímeros de pirimidinas por meio de uma fotorreversão direta, parece facilitar a reparação destas lesões pelo sistema de NER. Provavelmente, outros fatores acessórios serão ainda descobertos.

O processo de NER em *E. coli* está sob o controle do sistema SOS, o que permite uma expressão constitutiva basal de um grande número de genes do sistema de NER, em condições normais, e uma forte indução deste sistema quando a bactéria sofre várias lesões no seu DNA.

Embora o mecanismo molecular básico do processo de NER de *E. coli* seja considerado um modelo válido para eucariotos, o sistema eucariótico é bem mais complexo. Em mamíferos, em que a quantidade de DNA presente no núcleo é muito maior do que em *E. coli* ou mesmo em leveduras, a reparação preferencial de sequências transcritas TC-NER também ocorre. Aparentemente, uma maior prioridade é dada no sentido de resolver, com rapidez, um bloqueio da transcrição de um processo vital causado por uma lesão no molde. O reconhecimento inicial e a sinalização da lesão do DNA neste subcaminho do sistema de NER são realizados pela RNA-polimerase II, ao passo que a reparação em regiões do genoma não contendo genes (**sistema de reparação global do genoma, GG-NER, de global genome excision repair**) é mais lenta e incompleta. As lesões que ocorrem nestas regiões são detectadas por complexos de reparação que exploram o DNA. A procura por anormalidades conformacionais, feita pela RNA-polimerase II, pode ser mais rigorosa, explicando por que algumas lesões, como díme-

ros de pirimidinas, são removidas com mais eficiência de fitas transcritas do que do restante do genoma.

As consequências de uma deficiência no sistema de reparação por excisão tornam-se evidentes devido ao aparecimento de sintomas clínicos que originam diversas doenças nos seres humanos, como a xeroderma pigmentosa (XP), a síndrome de Cockayne (SC) e uma forma especial de tricotiodistrofia (a PIBIDS). Pacientes que sofrem de uma dessas doenças autossômicas recessivas são muito sensíveis à luz solar (UV). Portadores de XP mostram pigmentação anormal e numerosas lesões pré-malignas quando expostos à luz do sol, bem como uma predisposição ao câncer de pele e uma acelerada neurodegeneração. Uma disfunção neurológica é também uma característica marcante na SC, embora, neste caso, a desmielinização dos neurônios seja diferente. Todos os pacientes com SC e alguns portadores de XP apresentam crescimento retardado. Uma diferença marcante entre esses dois tipos de doenças é que indivíduos com SC não apresentam uma elevada incidência de câncer de pele.

PIBIDS é a sigla de uma complexa síndrome que compreende fotossensibilidade, ictiose (escamação da pele), cabelo quebradiço, inteligência diminuída, decréscimo de fertilidade e estatura reduzida. A PIBIDS assemelha-se à SC em muitos aspectos, como seus efeitos sobre o crescimento, sistema nervoso, desenvolvimento sexual, fotossensibilidade e sua baixa associação com o câncer. Cabelo quebradiço, devido a uma redução do conteúdo de cisteínas da matriz proteica, é uma das características mais marcantes da tricotiodistrofia, doença que também afeta pacientes não fotossensíveis.

Em muitos portadores de XP e PIBIDS, a reparação de lesões provocadas por UV é reduzida ou completamente ausente. O defeito do sistema de NER em pacientes com SC clássica é limitado à via da reparação preferencial de genes ativos; o sistema menos eficiente, o GG-NER que repara o genoma total, ainda é funcional. Análises de complementação mostraram que todas as três doenças envolvendo o sistema de NER são geneticamente heterogêneas, pois fusões de células de diferentes pacientes podem complementar cada defeito de reparação. Sete grupos de complementação XP excisão-deficientes (designados XP-A a XP-G) foram catalogados até o momento (**Tabela 7.2**). Um oitavo grupo, chamado de XP-variante, tem um sistema de NER normal, mas apresenta deficiência no caminho de reparação pós-replicação. Pacientes que têm SC se enquadram nos grupos SC-A ou SC-B, ao passo que um número pequeno de indivíduos, com manifestações combinadas de XP e SC, são classificados nos grupos XP-B, D e G. Pelo menos dois grupos de complementação podem ser distinguidos para PIBIDS. A deficiência no sistema de reparação é mais severa nos grupos XP-A, B e G e afeta tanto a reparação preferencial como a do genoma total.

Na última década, todos os genes que codificam os fatores-chave para o sistema de NER em seres humanos foram clonados, e a reação central de “clivagem e ligação” foi reconstituída *in vitro* a partir dos compo-

Tabela 7.2 Propriedades dos grupos de complementação de XP, CS e PIBIDS

Grupos de complementação	Características clínicas		Características de reparação		
	Câncer de pele	Anomalias neurológicas	Frequência de ocorrência	Sensibilidade à UV	Observações
XP-A	+	++	Alta	+++	
XP-B	+/-	+++/+	Muito rara	++	XP/SC Combinadas.
XP-C	+	-	Alta	+	Deficiente no sistema de reparação global do genoma.
XP-D	+	+/ -	Intermediária	++	Pacientes com PIBIDS e com XP-SC.
XP-E	+/-	-	Rara	+/-	
XP-F	+/-	-	Rara/intermediária	+	Reparação lenta mas prolongada.
XP-G	+/-	+++/+	Rara	++	SC
XP-V	+	-	Alta	+	Deficiente na reparação pós-replicação.
SC-A normal	-	++	Rara	+	Deficiente na reparação fita-seletiva.
SC-B normal	-	++	Alta	+	Deficiente na reparação fita-seletiva.
TTD1 BR	-	+	Muito rara	+	PIBIDS

OBS: o número de sinais “+” indica a intensidade da característica, o sinal “+/-” indica que a característica apresenta intensidade moderada e o sinal “-” indica que a doença não apresenta a característica.

nentes purificados. A **Tabela 7.3** mostra alguns dos genes humanos já isolados e seus correspondentes em levedura. Recentemente, a proteína XPC foi identificada como um sensor de lesões no DNA e como um fator de recrutamento do sistema de reparação global do genoma, não participando, portanto, do TC-NER. O complexo fator de transcrição geral, TFIIH, que contém as helicases XPB e XPD, promove a separação das fitas no local da lesão. XPA verifica a lesão em uma conforma-

ção do DNA aberta, e é crucial na montagem do restante da maquinaria de reparação. A proteína de replicação A (RPA) estabiliza o complexo de DNA aberto e está envolvida no posicionamento das endonucleases XPG e XPF-ERCC1, responsáveis pelas incisões no DNA ao redor da lesão. Após a remoção da sequência contendo a lesão, em geral um oligonucleotídeo de 24-32 bases, os fatores de replicação gerais preenchem o espaço remanescente e refazem a ligação.

Tabela 7.3 Genes do sistema de NER humanos conhecidos

Gene (proteína)	Local no cromossomo	Tamanho (kb) da proteína	Nº de aminoácidos em <i>S. cerevisiae</i>	Gene homólogo	Propriedades da proteína
<i>xpac</i> (XPA)	9q34	25	273	<i>rad14</i>	Afinidade por DNA de fita simples ou com lesões.
<i>xpcc</i> (XPC)			940	<i>rad4</i>	Sensor da lesão e fator de recrutamento do sistema de reparação. Restrito ao sistema GG-NER.
<i>xpbc/ercc3</i> (XPB)	2q21	45	782	<i>rad25/ssl2</i>	Helicase 3' → 5' subunidade de TFIIH.
<i>xpdc/ercc2</i> (XPD)	19q13	20	760	<i>rad3</i>	Helicase 5' → 3' subunidade de TFIIH.
<i>xpfc/ercc1</i> (XPF-ERCC1)	19q13	15-17	297	<i>rad10</i>	Catalisa a incisão 5'.
<i>xpgc/ercc5</i> (XPG)	13q32-	32	1186	<i>rad2</i>	Catalisa a incisão 3' e estabiliza o complexo totalmente aberto.
<i>rpac</i> (RPA)			616	<i>rfa1</i>	Estabiliza o complexo de DNA aberto e posiciona as nucleases.

A partir do conhecimento das propriedades de alguns destes genes e da analogia com o mecanismo conhecido em *E. coli*, o mecanismo proposto para o sistema de NER em mamíferos é o seguinte:

1. as proteínas XPB, XPD e XPA estão envolvidas na detecção da lesão, processo em que as helicases percorrem a hélice e XPA reconhece a lesão;
2. após encontrar a lesão, o complexo desnatura o DNA do local do dano. Esta conformação específica é reconhecida pelas endonucleases XPF-ERCC1 e XPG, que fazem a dupla incisão na cadeia, gerando um oligonucleotídeo de 24-32 bases contendo a lesão;
3. as DNAs-polimerases δ e ϵ e seus cofatores PCNA e RF-C estão envolvidos na remoção do oligonucleotídeo contendo o dano, na liberação do complexo de incisão e na ressíntese do DNA.
4. por fim, a DNA-ligase I realiza a ligação da extremidade 5' da região recém-sintetizada à sequência original.

A **Figura 7.17** mostra as etapas do sistema de reparação por excisão de nucleotídeos (NER) em mamíferos

e as proteínas envolvidas em cada subcaminho (GG-NER e TC-NER). Como pode ser notado, os processos diferenciam-se em relação às proteínas envolvidas no reconhecimento da lesão: HHR23B e XPC no GG-NER e a RNA-polimerase II, junto das proteínas CSB e CSA, no TC-NER. Uma vez reconhecida a lesão, as proteínas que atuam na sua remoção e na correção da fita danificada são comuns a ambos os processos. **A Tabela 1 do Material Complementar Online** deste capítulo resume algumas das principais doenças humanas relacionadas com algum tipo de deficiência dos mecanismos de reparação.

2.4 Sistema de reparação pós-replicação

2.4.1 Reparação por recombinação

No processo de reparação por excisão, a fita complementar não danificada é utilizada como molde na substituição do fragmento lesado, porém, algumas vezes, este molde não está disponível. Isso ocorre, por exemplo, quando, durante o processo de replicação do DNA, uma lesão do tipo dímero de pirimidinas impede o sítio danificado de atuar como molde. A DNA-polimerase, então, é forçada a “saltar”, passando pela lesão. A DNA-polimerase para

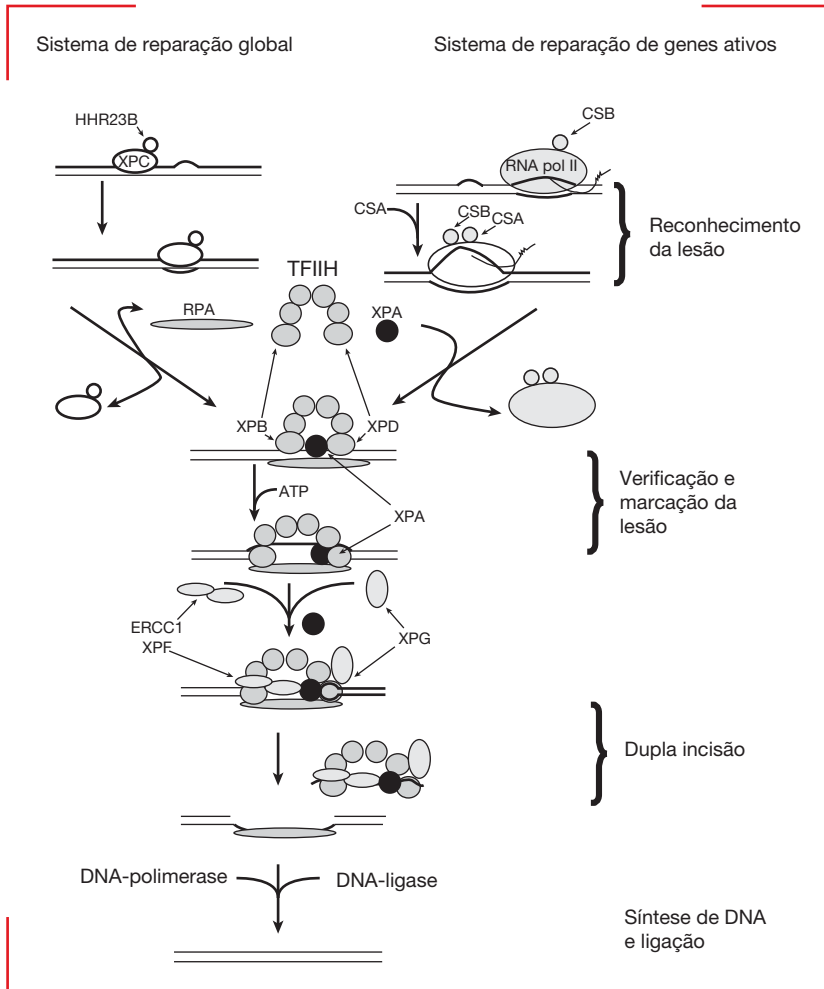


Figura 7.17

Etapas do sistema de reparação por excisão de nucleotídeos (NER) em mamíferos e as proteínas envolvidas em cada subcaminho: GG-NER e TC-NER. Detalhes da atividade de cada proteína são mostrados na Tabela 7.3.

a síntese da cadeia-filha e a replicação somente reinicia a alguma distância, além da lesão. Uma lacuna, com um tamanho considerável, é deixada na fita recém-sintetizada. Neste caso, as duas fitas-filha resultantes serão diferentes: uma terá a fita parental contendo a lesão pareada com uma fita recém-sintetizada, que contém uma lacuna de tamanho considerável; e a outra fita duplicada terá a fita parental não danificada, que terá sido copiada, originando uma fita complementar normal. Além disso, se ambas as bases de um par complementar forem alteradas, nenhuma poderá atuar como molde para a outra; ou, ainda, se a hélice dupla for quebrada nas duas fitas e se um determinado segmento do DNA for perdido, não haverá a possibilidade de que uma reparação direta e precisa ocorra. Em todos estes casos, toda a informação do sítio danificado é perdida e apenas pode ser recuperada pela utilização de uma outra molécula idêntica de DNA. Um modelo simplificado deste sistema de reparação, denominado **reparação por recombinação**, é apresentado na **Figura 7.18**.

Uma fonte apropriada de DNA, para servir como molde na reparação por recombinação na maioria das vezes pode ser encontrada em qualquer célula. Nas bactérias, devido a sua rápida multiplicação, as células quase sempre possuem uma ou mais cópias extras de seu cromossomo. Já as células de organismos superiores geralmente são diploides e, portanto, possuem a mesma sequência em cada um dos cromossomos de um par de homólogos.

A principal rota do processo de reparação por recombinação em *E. coli* é responsabilidade dos genes *rec*, cujas atividades no sistema de recombinação serão discutidas no Capítulo 8. Em linhagens de *E. coli*, deficientes no sistema de reparação por excisão, mutações no gene *recA* abolem todos os processos remanescentes para a reparação e recuperação de lesões. Tentativas de replicação do DNA em células *uvr⁻ rec⁻* produziram fragmentos cujos tamanhos corresponderam às distâncias esperadas entre dímeros de pirimidinas. Esse resultado sugeriu que os dímeros constituem um obstáculo letal à replicação, na ausência das funções de RecA. Isso explica por que um mutante duplo não pode tolerar mais do que um ou dois dímeros em seu genoma (em comparação com a capacidade de uma bactéria de tipo selvagem, que pode tolerar aproximadamente 50).

2.4.2 Sistema SOS

Uma vez que a célula pode regular a expressão dos genes conforme a necessidade de seus produtos, muitas enzimas envolvidas na reparação do DNA são induzidas pelo próprio defeito na molécula do ácido nucleico. O grupo mais importante destas enzimas consiste nos produtos dos **genes SOS** (**Tabela 7.4**), induzidos por um defeito que é severo o bastante para impedir a síntese de DNA, ao invés de apenas trocar as propriedades de pareamento de bases. Os dímeros de pirimidinas constituem um exemplo deste tipo de dano. Quando a forquilha de replicação encontra um dímero, a síntese para e só é reiniciada a al-

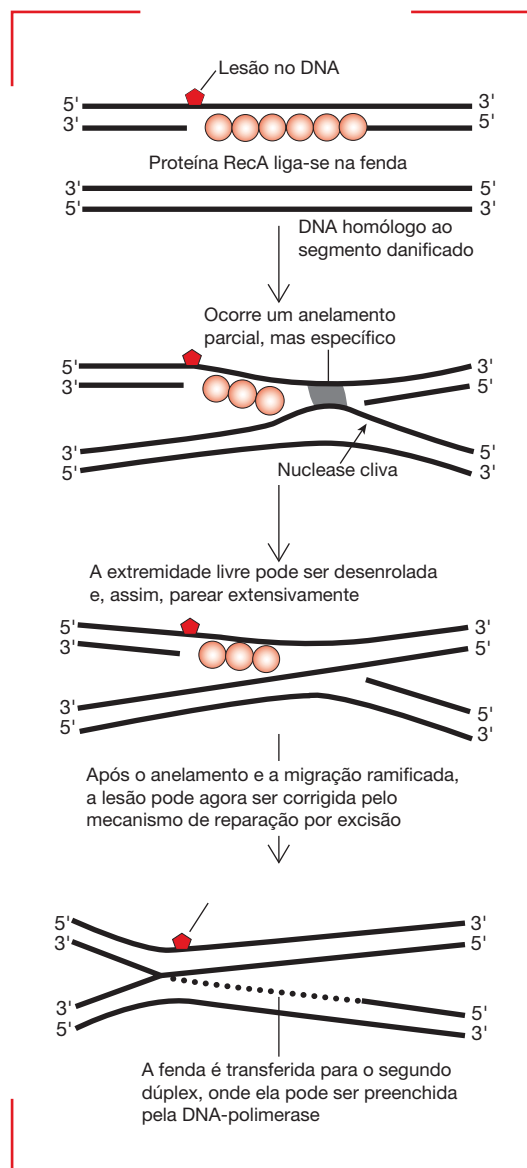


Figura 7.18

Exemplo de ação do mecanismo de reparação por recombinação pós-replicação, em que a fita complementar à região contendo a lesão não está disponível para servir como molde, sendo necessária a utilização do cromossomo homólogo.

guma distância além do dímero, ficando uma lacuna no DNA, na qual a enzima de recombinação RecA se liga. Antes de iniciar a troca da fita de DNA danificada (e, assim, a reparação por recombinação), a ligação da proteína RecA ao DNA de fita simples ativa, nessa proteína, uma atividade enzimática independente da recombinação, que é a estimulação da atividade autoproteolítica do repressor dos genes SOS (proteína LexA). Desta maneira, a proteína RecA promove a reparação do DNA de ambas as formas, por recombinação e pela indução de cerca de

Tabela 7.4 Genes SOS induzíveis por lesões do DNA

Nome do gene	Papel na reparação do DNA
Genes com funções conhecidas	
<i>uvrA</i>	Codificam endonucleases de excisão
<i>uvrB</i>	
<i>uvrC</i>	
<i>urvD</i>	Codifica a helicase II
<i>dinB</i>	DNA-polimerase IV, síntese translesão
<i>umuD</i>	DNA-polimerase V, síntese translesão
<i>umuC</i>	
<i>sulA</i>	Codifica uma proteína que inibe a divisão celular, permitindo que a reparação do DNA aconteça
Genes envolvidos no metabolismo do DNA e cujo papel específico no sistema de reparação não é bem conhecido	
<i>ssb</i>	Codifica proteína SSB
<i>himA</i>	Codifica uma subunidade do fator de interação ao hospedeiro, envolvido na recombinação sítio-específica
<i>recN</i>	Envolvido na reparação por recombinação
Genes com funções desconhecidas	
<i>dinA</i>	
<i>dinD</i>	
<i>dinF</i>	

15 genes SOS. Todos esses genes são expressos em um nível significativo, mesmo sem indução, mas seu nível de expressão aumenta muito nas células em que o DNA é lesado, uma vez que o repressor do sistema (LexA) foi removido. Outros genes SOS codificam proteínas envolvidas na síntese e na reparação do DNA ou com funções ainda não identificadas.

A proteína LexA reprime a expressão de seus genes-alvo por meio da sua ligação a uma sequência de 20 pb, chamada de **box SOS**, que inclui uma sequência consensual com oito posições conservadas. Como em outros operadores, os **boxes SOS** sobrepõem-se com os respectivos promotores. No locus *lexA*, sujeito a autorrepressão, existem dois **boxes SOS** adjacentes.

RecA e LexA são alvos recíprocos no circuito SOS, RecA aciona a clivagem de LexA, que, por sua vez, reprime *recA*. A resposta SOS, portanto, ocasiona a amplificação tanto da proteína RecA como do repressor LexA, embora pareça, esse resultado não é contraditório. O aumento na expressão de RecA é necessário devido ao seu papel nos processos de reparação por recombinação. Sob indução, o nível de RecA aumenta em até cerca de 50 vezes o seu nível basal de ~1.200 moléculas/célula. O alto nível em células induzidas significa a existência de RecA suficiente para garantir que todas as proteínas LexA serão clivadas. Isso deveria impedir LexA de restabelecer a

repressão de seus genes-alvo, porém, a principal importância desse circuito para a célula é a sua capacidade de retornar rapidamente à normalidade. Quando o sinal de indução é removido, a proteína RecA perde a sua capacidade de desestabilizar LexA. Nesse momento, o gene *lexA* está sendo expresso em um alto nível, e, na ausência de RecA ativada, a proteína LexA acumula-se com rapidez na sua forma intacta, inativando os genes SOS. Isso explica porque a resposta SOS é facilmente reversível.

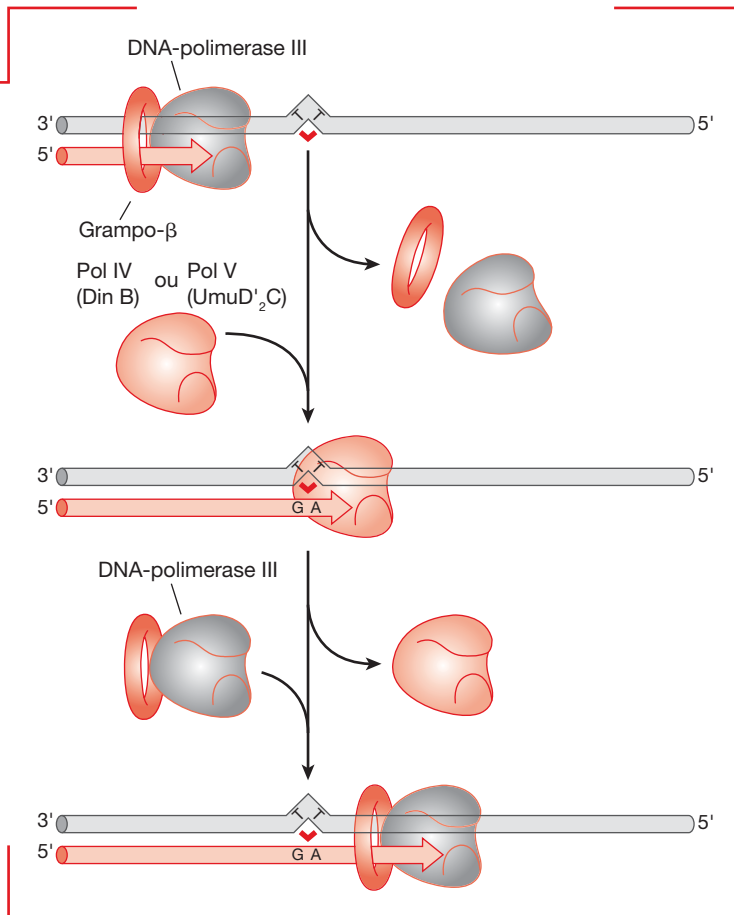
2.4.3 Reparação sujeita a erro: síntese translesão

Existem ocasiões em que o dano no DNA é extremo, ao ponto de não haver maneira de os mecanismos celulares de reparação corrigirem de forma precisa a molécula. Por exemplo, quando uma DNA-polimerase está replicando o DNA e encontra uma lesão, como um dímero de pirimidina ou um sítio apurínico que não foi corrigido, o **sistema de reparação sujeito a erro** (*error-prone*), em que qualquer uma das quatro bases é inserida no local lesado, a fim de garantir a continuidade do processo de replicação. Devido ao fato de não existir a informação (molde) precisa, o próprio mecanismo de reparação acaba sendo o causador de uma mutação, pois a chance de introduzir uma base incorreta na cadeia é grande. De qualquer forma, para a célula, é preferível incorporar uma base incorreta e permitir que a replicação continue do que não replicar mais. Esse mecanismo é conhecido como **síntese translesão** (ou síntese *by pass*) (**Figura 7.19**), que, embora seja altamente sujeito a erro, poupa a célula de um destino pior, causado por um cromossomo replicado de forma incorreta.

A síntese translesão é catalisada por uma classe especializada de DNA-polimerases, que sintetizam DNA diretamente sobre o sítio do dano. Nesse processo, em *E. coli*, o núcleo da DNA-polimerase III perdido na forquilha de replicação colapsada é substituído por uma das duas **DNA-polimerases de desvio** (*by pass DNA-polymerases*), a DNA-polimerase IV e a DNA-polimerase V, cujas sínteses são induzidas pela resposta SOS. A **DNA-polimerase IV** é o produto do gene *dinB* e a **DNA-polimerase V** é constituída pelos produtos dos genes *umuC* e *umuD*. Ambas as enzimas pertencem a uma família distinta de DNA-polimerases, encontrada em diversos organismos, conhecida como **DNA-polimerases da família Y** (**ver Material Complementar Online 7.2**). Uma característica importante dessas DNA-polimerases é que, embora dependam de um molde, elas podem incorporar nucleotídeos independentemente do pareamento de bases. Isso explica como as enzimas conseguem sintetizar DNA por sobre a lesão na fita-molde, inserindo nucleotídeos de uma maneira que não é dirigida pelo pareamento de bases. Entretanto, a incorporação do nucleotídeo pode não ser aleatória – algumas DNA-polimerases translesão incorporam nucleotídeos específicos. Por exemplo, um membro da família Y de DNA-polimerases translesão, presente em seres humanos, insere corretamente dois resíduos de adenina em oposição a um dímero de timinas.

Figura 7.19

Sistema de reparação translesão. Ao encontrar uma lesão no DNA, a DNA-polimerase III dissocia-se do DNA e é substituída por uma DNA-polimerase translesão, que continua a síntese por sobre o dímero de pirimidinas no molde. Após a inserção aleatória de nucleotídeos, a DNA-polimerase translesão é substituída pela DNA-polimerase III.



2.5 Reparação de quebras duplas no DNA

Até o momento, foram apresentados sistemas de reparação que lidam com lesões ou erros em uma das fitas do DNA. Será considerado, agora, o que acontece quando ambas as fitas de uma mesma molécula de DNA são clivadas. Essas quebras ocorrem nas células em diversas circunstâncias e se constituem em uma das principais ameaças à integridade do genoma. Se não forem reparadas, as quebras podem causar a morte celular e, em organismos multicelulares, o câncer.

Uma grande variedade de agentes exógenos, incluindo a radiação ionizante e um número considerável de drogas anticancerígenas (como, por exemplo, a bleomicina) causam **quebras duplas no DNA (DSB, de DNA double-strand break)**, bem como agentes endógenos, como os radicais livres – o produto secundário do metabolismo oxidativo. Além disso, em mamíferos, essas quebras ocorrem como etapas intermediárias dos processos recombinacionais, que geram a enorme variedade de anticorpos e proteínas receptoras de células-T durante o desenvolvimento de células linfocitárias. Em razão da redundância dos sistemas de **reparação de quebras duplas no DNA (DSBR, de DNA double-strand break repair)**, as quebras duplas podem ser reparadas de diferentes maneiras, de forma que a

perda de um sistema de reparação (seja por causas genéticas ou intervenção química) não impede a correção da lesão.

2.5.1 Por recombinação homóloga

Um dos sistemas envolvidos na reparação de quebras duplas é o sistema de **reparação de quebras duplas por recombinação homóloga**. O início desse processo envolve a atividade de uma enzima, que reconheça a quebra dupla na cadeia de DNA e promova a digestão das fitas contendo as extremidades 5'-P da hélice de DNA quebrada, criando, assim, extremidades 3'-OH sobressalentes (**Figura 7.20**). Uma das extremidades 3'-OH sobressalente procura uma região de complementaridade de sequência na cromátide irmã e, então, invade a molécula de DNA homóloga, alinhando as sequências complementares. Com as sequências complementares alinhadas, ocorre a síntese de DNA a partir da extremidade 3'-OH sobressalente, utilizando a fita de DNA não danificada como molde. A interação das duas cromátides irmãs é necessária, pois, uma vez que ambas as fitas da mesma hélice estão quebradas, não existe uma fita de DNA parental não danificada disponível para ser utilizada como molde durante a reparação. Após a síntese, a molécula heterodúplex resultante é re-

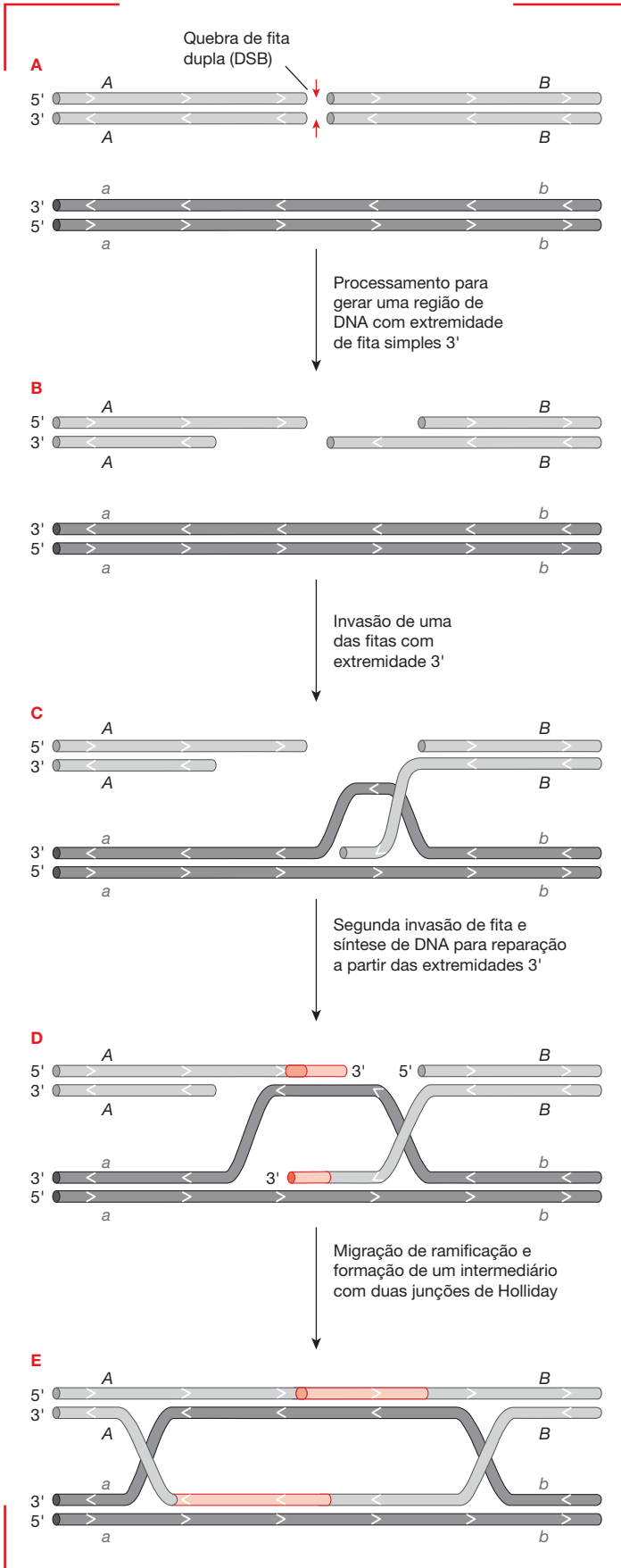


Figura 7.20

Etapas do sistema de reparação de quebras duplas por recombinação homóloga.

solvida e as duas cromátides são separadas. Por se tratar de um tópico que envolve a maquinaria enzimática da recombinação homóloga, esse processo é discutido com maiores detalhes no Capítulo 8.

O processo de recombinação homóloga ocorre durante o final da fase S ou no início da fase G2 do ciclo celular, após a replicação do DNA, quando as cromátides irmãs estão disponíveis para serem usadas como moldes na reparação. Uma vez que um molde não danificado é utilizado durante a síntese de reparação, o sistema de reparação de quebras duplas por recombinação homóloga é um processo livre de erro.

2.5.2 União de extremidades não homólogas

Em organismos multicelulares, o principal mecanismo de reparação de quebras duplas é chamado de **união de extremidades não homólogas** (NHEJ, de *nonhomologous end joining*) e consiste na ligação de extremidades não coesivas de uma mesma molécula de DNA. Esse tipo de união não requer homologia entre as duas moléculas que irão recombinar e é bem diferente do processo de recombinação homóloga. O mesmo complexo enzimático realiza o processo tanto na NHEJ como na recombinação dos genes de resposta imune.

Na NHEJ, as extremidades quebradas de uma DSB devem ser alinhadas, suas extremidades com projeções de fita simples devem ser aparadas e/ou preenchidas e as fitas devem ser ligadas. O núcleo da maquinaria enzimática da NHEJ, em eucariotos, consiste em uma proteína-quinase DNA-dependente (DNA-PK). Essa proteína apresenta dois componentes, uma subunidade catalítica (DNA-PK_{cs}), com aproximadamente 465 kDa, e a proteína de ligação à extremidade do DNA, Ku (um heterodímero de subunidades homólogas de 70 e 83 kD, Ku70 e Ku80). DNA-PK_{cs}, em associação com a proteína Ku ligada ao DNA, realiza a função catalítica de DNA-PK. Além desta proteína, fazem parte do processo a DNA-ligase IV e a proteína acessória Xrcc4. A Ku é uma proteína nuclear abundante, que liga-se a DSBs independentemente da presença de extremidades cegas ou de projeções de fita simples e, portanto, parece ser o sensor primário de DSBs na célula.

Foi demonstrado que complexos Ku-DNA formam dímeros, a fim de alinharem para ligação os membros de uma DSB, ambos com as extremidades aparadas e com projeções de fita simples curtas (1 a 4 pb) e complementares, como está demonstrado na **Figura 7.21**. As extremidades do DNA ficam expostas ao longo de uma das faces de cada um dos complexos Ku-DNA, o que, presumivelmente, as torna acessíveis às DNA-polimerases que irão preencher as lacunas, e às nucleases que irão aparar as extremidades sobressalentes e inapropriadas, preparando-as para a ligação pela DNA-ligase IV em complexo com Xrcc4. A remoção de nucleotídeos das extremidades (*nucleotide trimming*), que gera mutações, parece ser realizada de modo dependente de ATP pelo complexo Mre11, que consiste em duas subunidades da nuclease Mre11 e duas subunidades da ATPa-se Rad50. A Ku aca-

ba sendo liberada do DNA religado, talvez por clivagem proteolítica.

A NHEJ mediada por Ku é um processo ineficiente, possibilitando a sobrevivência de apenas uma a cada mil células de levedura nas quais uma quebra cromossômica foi introduzida e resulta na formação de deleções e inserções de um a quatro nucleotídeos nos produtos reparados.

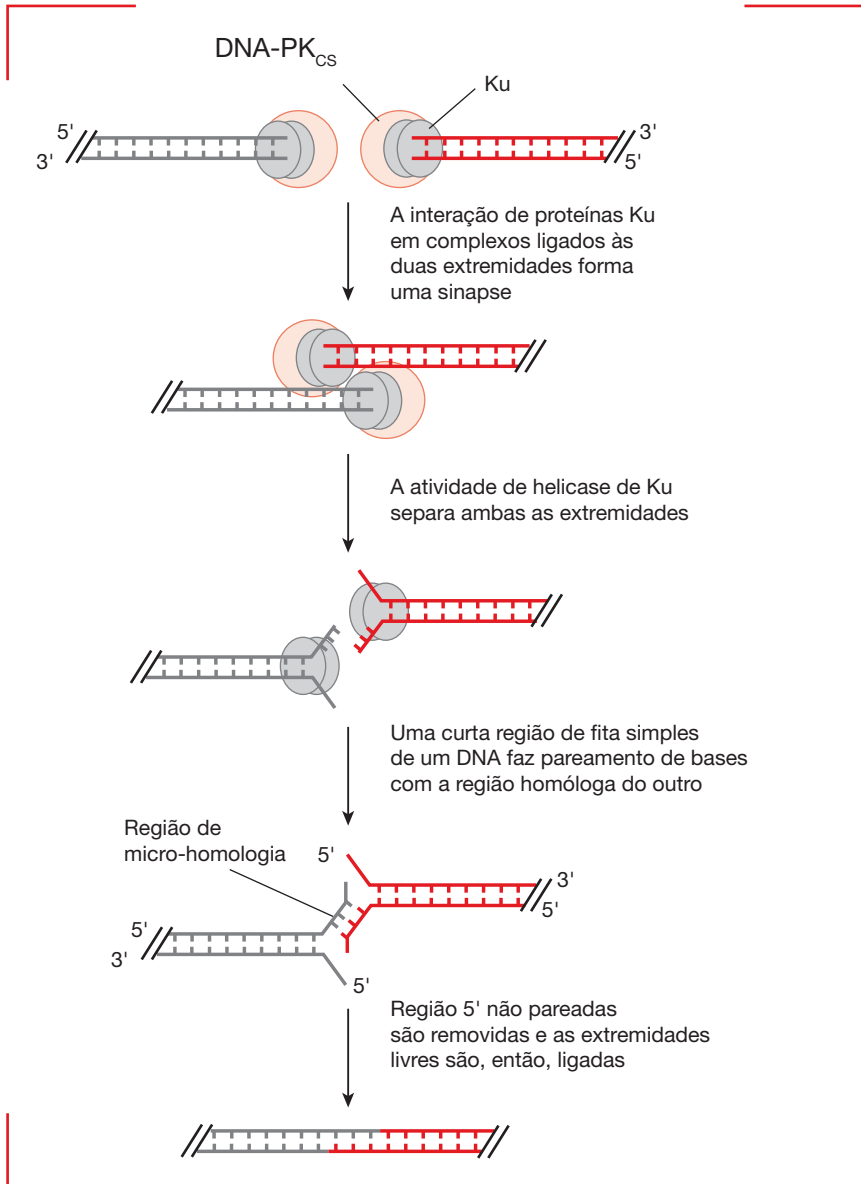
2.5.3 União de extremidades mediada por micro-homologias

Estudos recentes indicaram a existência de um mecanismo de união de extremidades de uma cadeia de DNA quebrada distinto da NHEJ e igualmente sujeito a erro. Esse mecanismo, denominado **união de extremidades mediada por micro-homologias** (MMEJ, de *microhomologymediated end joining*), distingue-se da NHEJ devido à necessidade da existência de micro-homologias de 5 a 25 pb entre as extremidades das cadeias que serão reunidas para a ligação entre elas ocorrer. Dessa forma, deleções maiores do que àquelas presentes na NHEJ são encontradas flanqueando a quebra original. A MMEJ também pode estar associada com anormalidades cromossômicas, como deleções, translocações, inversões e outros rearranjos cromossômicos complexos. A **Figura 7.22** apresenta as etapas que ocorrem durante a reparação de DSB pelo mecanismo de MMEJ e a **Tabela 7.5** contém as proteínas, conhecidas até o momento, envolvidas no processo de MMEJ, em outros processos de reparação de DSBs, ou na regulação de sua atividade.

A existência de sistemas de reparação de DSBs independentes de recombinação homóloga, que utilizam micro-homologias entre cinco ou mais bases durante a religação, foi demonstrada há mais de 20 anos. Em relação à MMEJ, muitas ambiguidades ainda permanecem sobre seu mecanismo de ação e finalidade. Por exemplo, por muito tempo, a MMEJ foi considerada um mecanismo de segurança para reparar DSBs, quando a NHEJ e outros mecanismos deixavam de agir. Essa premissa contribuiu com a ideia de que a MMEJ pudesse ser irrelevante sob condições fisiológicas normais. Entretanto, já foi demonstrado que a MMEJ em mamíferos é muito importante, pelo menos em determinados contextos, durante o desenvolvimento do sistema imune. Na falta de componentes da NHEJ, como a DNA-ligase IV ou Xrcc4, a recombinação de troca de classe, que ocorre nos linfócitos B, e a recombinação V(D)J, que gera a diversidade dos receptores de células T, ocorrem por meio da ligação das extremidades das cadeias de DNA pela utilização de micro-homologias extensas.

2.5.4 Anelamento de fitas simples

O mecanismo de união de DSBs, denominado de **anelamento de fitas simples** (SSA, de *single-strand annealing*), é o preferencialmente utilizado na reparação de uma DSB que ocorre entre sequências repetidas e orientadas na mesma direção. As extremidades de ambos os

**Figura 7.21****Etapas do processo de união das extremidades não homólogas em células de mamíferos.**

Nesse processo, o heterodímero Ku liga-se às extremidades livres do DNA, juntamente com a subunidade catalítica da proteína quinase DNA-dependente (DNA-PK_{CS}), ativando sua atividade quinásica. Após a formação de uma sinapse, na qual as extremidades quebradas se sobrepõem, Ku desenrola as extremidades, revelando micro-homologias aleatórias nas sequências das duas cadeias de DNA. As extremidades de fita simples 5'-P não pareadas são removidas por nucleases e as duas moléculas de DNA de fita dupla, ligadas pela ação conjunta da DNA-ligase IV e Xrcc4. Como resultado, a DSB é reparada, mas alguns pares de bases no local onde ocorreu a quebra são perdidos.

lados da quebra são processadas por uma exonuclease 5' → 3', que expõem regiões de fita simples com extremidades 3' complementares, situadas nas repetições diretas e que podem se anelar. Isso resulta na deleção da sequência que não era repetida, mas existia entre as sequências repetidas (normalmente >30 nt são perdidos). As extremidades sobressalentes são removidas, ocorre síntese de DNA para o preenchimento de lacunas e a ligação.

Propriedades fundamentais do SSA foram documentadas em *Saccharomyces cerevisiae*, pela criação de DSBs *in vivo*. As análises de SSA tanto em plasmídeos como no cromossomo da levedura confirmaram que, quando uma DSB está situada dentro de uma das repetições, ela pode ser reparada por SSA ou por conversão gênica (ver Capítulo 8). O aumento da distância entre as duas repetições diminui a eficiência de SSA na competi-

ção com o evento de conversão gênica, consistente com a necessidade de um tempo maior para que a degradação 5' → 3' exponha as homologias complementares. A eficiência do SSA é influenciada pelo comprimento das sequências flangeadoras e, também, pelo grau de identidade das sequências.

Em relação às proteínas envolvidas no SSA, muitas são compartilhadas com outros mecanismos de correção de DSBs (ver Tabela 7.5). Entretanto, estudos de diversos grupos demonstraram que a ausência da proteína Rad52, uma proteína de anelamento de fitas, praticamente elimina o SSA quando as sequências homólogas apresentam entre 1 a 2 kb de comprimento. A **Figura 7.23** mostra uma comparação entre os três mecanismos de reparação de DSBs, como elucidado para *S. cerevisiae*, apresentados neste capítulo.

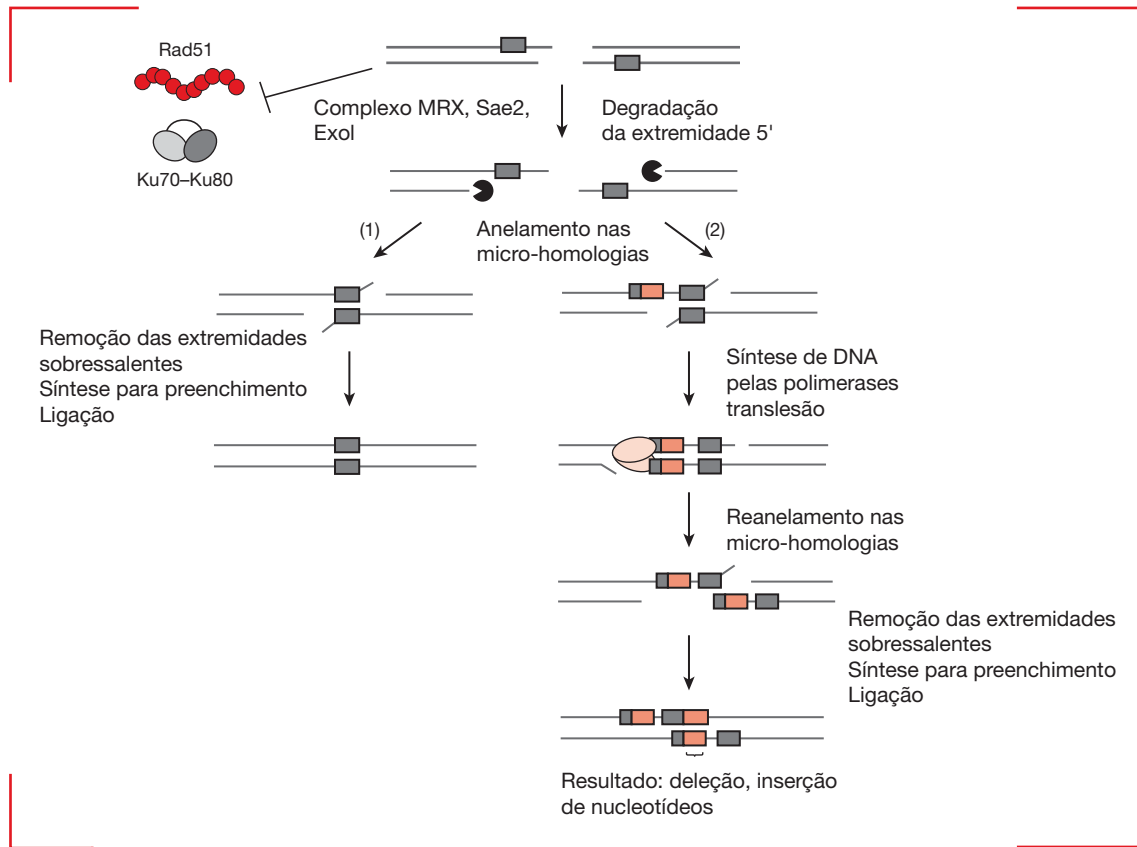


Figura 7.22

Modelo para o MMJE e mecanismos alternativos de reparação de DSBs. Durante os estágios iniciais do MMJE, Ku70-Ku80 e Rad51, que inibem o MMJE, são impedidas de ligar-se ao DNA ou removidas dele. Isso possibilita a degradação de 30-50 nucleotídeos pelo complexo MRX, Sae2 e Exo1 (indicadas pelos círculos parciais em ■), que revela micro-homologias entre as sequências (quadros em ■). Essas micro-homologias se anelam de forma transitória e dinâmica. (1) Nos casos em que o anelamento é estável, a reparação é completada por remoção das extremidades sobressalentes, síntese de DNA para o preenchimento de lacunas e ligação, resultando em deleções em relação à sequência original. (2) Alternativamente, uma ou mais DNA-polimerases translesão (■) podem estender as sequências aneladas (representadas por quadros em ■ e ■) pela síntese sujeita a erro. A dissociação das micro-homologias iniciais e o realinhamento de outras sequências micro-homólogas, seguido de remoção das sequências sobressalentes, síntese de DNA para o preenchimento de lacunas e ligação, finaliza a reparação, resultando em eventos de deleção, seguidos de inserção de nucleotídeos no local da quebra. Muitas variações e interações de (2) podem ocorrer, resultando em junções complexas, com deleções e inserções.

Tabela 7.5 Proteínas envolvidas no processo de MMEJ e em outros processos de reparação de DSBs, ou na regulação de sua atividade^a

Proteína	Espécie(s) estudada(s)	Função proposta no MMJE	Outro(s) sistema(s) de reparação do(s) qual(is) participa
Mre11-Rad50-Xrs2 ^b	<i>S. cerevisiae</i>	Degradação para expor as micro-homologias	NHEJ, HR
Exo1	<i>S. pombe</i> <i>S. cerevisiae</i>	Degradação usando atividade exonucleásica 5' → 3'	HR, SSA, MMR
Sae2 ^c	<i>S. cerevisiae</i>	Promoção da degradação	HR, SSA
Tel1 ^d	<i>S. cerevisiae</i>	Fosforilação de proteínas importantes para a degradação	HR
Srs2	<i>S. cerevisiae</i>	Remoção de Rad51 do DNA de fita simples	HR, SSA
Rad1-Rad10 ^e	<i>S. cerevisiae</i>	Remoção das extremidades 3' sobressalentes	SSA, NER
PolIV	<i>S. cerevisiae</i> <i>S. pombe</i>	Síntese de pequenos fragmentos de DNA	NHEJ
Pol η, Pol ζ	<i>S. cerevisiae</i>	Síntese de DNA	TLS
Pol32 (subunidade de Pol δ)	<i>S. cerevisiae</i>	Síntese para preencher lacunas de fitas simples	HR
DNA-ligase IV	<i>S. cerevisiae</i>	Ligação	NHEJ
DNA-ligasas I, III	homem e roedores	Ligação	HR, BER, NER
PARP-1	homem e roedores	Sinapse das extremidades	BER
Histona H1	homem	Pode estimular ou inibir o MMEJ em diferentes estudos	N/A

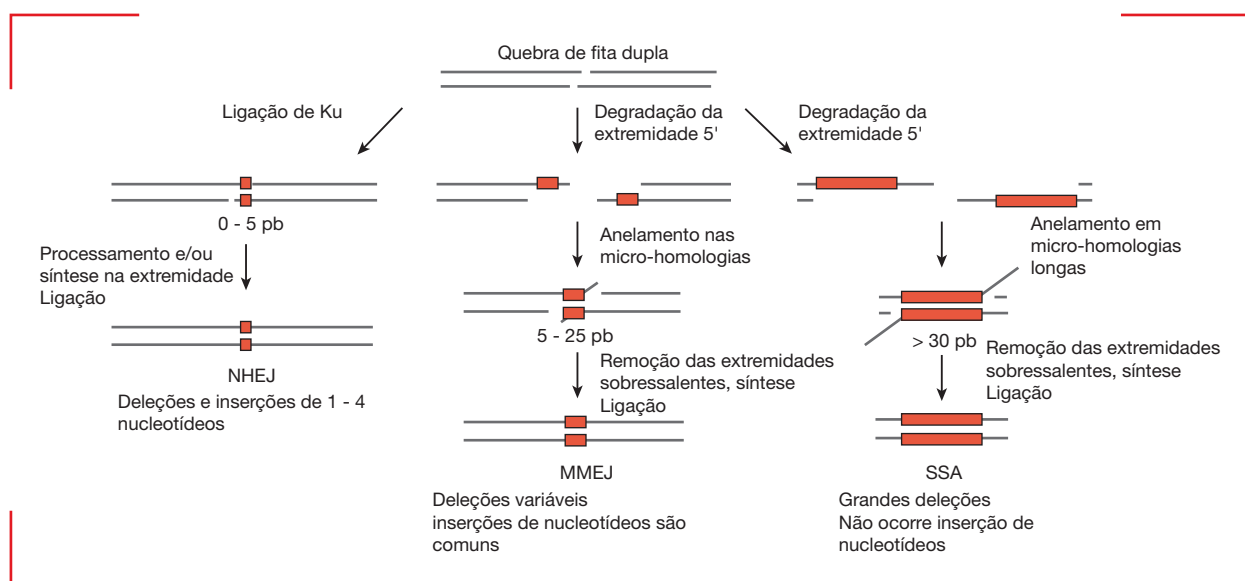
^aAbreviações: BER: reparação por excisão de base; HR: recombinação homóloga; MMR: reparação de mal pareamentos (*mismatch repair*); NER: reparação por excisão de nucleotídeos; SSA: anelamento de fitas simples; TLS: síntese translesão; N/A: não avaliado.

^bNBS1 em mamíferos.

^cCtIP em mamíferos.

^dATM em mamíferos.

^eXPF-ERCC1 em mamíferos.

**Figura 7.23**

Comparação entre os sistemas de NHJE, MMEJ e SSA em *S. cerevisiae*. Durante o NHJE, o heterodímero Ku70-Ku80 liga-se à DSB, impedindo a degradação das extremidades do DNA. A reparação prossegue por meio do anelamento de regiões curtas de micro-homologia (quadros ■), preenchimento pela Pol IV e ligação, utilizando a DNA-ligase IV, que resulta em produtos com pequenas deleções e inserções. Tanto o MMEJ como o SSA necessitam da degradação das extremidades e do desenrolamento das fitas para revelar sequências homólogas, embora a extensão de homologia necessária para o MMJE (5-25 pb) seja mais curta do que para o SSA. SSA e MMJE também necessitam da remoção das extremidades sobressalentes, antes da síntese de DNA, para o preenchimento das lacunas e a ligação. Nos produtos do MMJE encontram-se nucleotídeos inseridos, mas eles nunca são observados nos produtos de SSA. Embora os três processos compartilhem várias proteínas (ver Tabela 7.5), em geral eles são bastante distintos entre si.



Resumo

Todo organismo que exibe uma forma diferente da de seus ascendentes, resultado da presença de uma mutação, é denominado mutante. Qualquer base do DNA pode ser mutada. Uma mutação pontual envolve modificação em um único par de bases (substituição) ou em poucos pares de bases (adição ou deleção). Devido à grande variedade de tipos e efeitos das mutações, elas são classificadas de diversas maneiras, podendo uma mesma mutação se enquadrar em mais de uma classificação. Mutações espontâneas podem surgir devido a erros durante a replicação do DNA, pelas lesões provocadas por agentes ambientais, como a hidrólise pela água, ou por lesões oxidativas. A taxa de mutação espontânea é muito baixa para todos os organismos, mas varia consideravelmente entre diferentes organismos e, em uma mesma espécie, a taxa de mutação espontânea varia de gene para gene.

A ocorrência de mutações pode ser aumentada pelo tratamento com determinados compostos, denominados agentes mutagênicos, causadores de mutações induzidas. Muitos mutagênicos atuam diretamente no DNA devido a sua capacidade de atuar como uma determinada base (análogos de base), de incorporar-se à cadeia polinucleotídica (corantes de acridina), ou de alterar quimicamente as bases (ácido nitroso e agentes alquilantes). Raios X, gama, cósmicos e luz ultravioleta também são mutagênicos. Os principais produtos da luz UV, dímeros e hidratos de pirimidinas, são lesões graves no DNA, para as quais existem vários mecanismos de reparação.

Muitos dos danos sofridos pelo DNA podem ser reparados, porque a informação genética é preservada em ambas as fitas da hélice dupla. Erros que ocorrem durante a replicação do DNA são corrigidos pela capacidade de revisão de leitura da própria DNA-polimerase e pelo sistema de reparação de mal pareamentos. Dímeros de pirimidinas e outros danos provocados pela luz UV são reparados por fotorreativação enzimática, reparação por excisão de nucleotídeos, ou por recombinação. Bases modificadas são corrigidas pelo sistema de excisão de bases. Bases alquiladas são reconhecidas por metil-transferases específicas. O sistema de reparação por excisão de nucleotídeos pode ser dividido em dois subcaminhos: o de reparação global do genoma e o de genes ativos, que tem a preferência de reparação.

Várias doenças humanas estão associadas aos defeitos nos sistemas de reparação, como a xeroderma pigmentosa, a síndrome de Cockayne e a PIBIDS. Quando o número de lesões é muito grande, *E. coli* utiliza o sistema SOS, no qual uma série de genes envolvidos em diversos mecanismos de reparação têm a sua expressão induzida. Entre esses genes estão os que codificam as DNA-polimerases translesão, fazendo uma reparação sujeita a erro. Quebras duplas em uma mesma molécula de DNA podem ser corrigidas pela reparação por recombinação, união de extremidades não homólogas, união de extremidades mediada por micro-homologias e por anelamento de fitas simples. Todos esses processos, exceto o por recombinação, provocam mutações no sítio onde ocorreu a quebra.



Leituras recomendadas

Klug WS, Cummings MR, Spencer CA, Palladino MA. Concepts of genetics. 9th ed. San Francisco: Pearson Benjamin Cummings; 2009.

Lainé JP, Egly JM. When transcription and repair meet: a complex system. Trends Genetic. 2006;22(8):430-436.

Lewin B. Genes VII. 7. ed. Porto Alegre: Artmed; 2001.

Lodish H, Baltimore D, Berk A, Zipursky L, Matsudaira PT. Biologia celular e molecular. 5. ed. Porto Alegre: Artmed; 2005.

Madhusudan S, Hickson ID. DNA repair inhibition: a selective tumor targeting strategy. Trends Genetic. 2005;11(11):503-511.

Mcvey M, Lee SE. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genetic. 2008;24(11):529-538.

Rando OJ, Verstrepen KJ. Timescales of genetic and epigenetic inheritance. Cell. 2007;128(4):655-68.

Schumacher B, Garinis GA, Hoeijmakers, HJ. Age to survive: DNA damage and aging. Trends Genetics. 2008;24(2):77-85.

Voet D, Voet JG. Bioquímica. 3. ed. Porto Alegre: Artmed; 2006.

Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick, R. Biologia molecular do gene. 5. ed. Porto Alegre: Artmed; 2006.

Luciane M. P. Passaglia

Recombinação Genética

1. Recombinação homóloga 164

- 1.1 Modelo de Holliday para recombinação homóloga 165
- 1.2 Mecanismos de recombinação e suas associações com a reparação do DNA 166
 - 1.2.1 Reparação de forquilhas de replicação colapsadas 167
 - 1.2.2 Reparação de quebras duplas pela recombinação homóloga 167
 - 1.2.3 Reparação para o preenchimento de lacunas por recombinação 170
- 1.3 Conversão gênica 170
- 1.4 Enzimas que atuam na recombinação homóloga em procariotos 170
 - 1.4.1 Proteína RecA 171
 - 1.4.2 Proteínas RecBCD 172
 - 1.4.3 Proteínas RecFOR 173
 - 1.4.4 Outras proteínas que participam da recombinação homóloga procariótica 173

1.5 Recombinação homóloga em eucariotos 175

- 1.5.1 A segregação correta dos cromossomos durante a meiose depende da recombinação homóloga 175
- 1.5.2 Enzimas de recombinação homóloga em eucariotos 176

2. Recombinação sítio-específica 179

- 2.1 Mecanismos de recombinação sítio-específica 180
 - 2.1.1 Troca de fitas 181
- 2.2 Integração e excisão do bacteriófago λ 181
- 2.3 Regulação da expressão gênica por meio da recombinação sítio-específica 182

No capítulo anterior, foram discutidos os mecanismos pelos quais as sequências do DNA de uma célula são mantidas de geração a geração com poucas alterações. Embora essa estabilidade genética seja decisiva para a sobrevivência em curto prazo, em períodos de tempo mais longos, a sobrevivência dos organismos pode depender da variação genética, pela qual eles podem se adaptar às mudanças ambientais. Uma importante propriedade do DNA das células é a sua capacidade de sofrer rearranjos, que podem ocasionar desde novas combinações entre os genes presentes em qualquer genoma individual até alterações qualitativas e quantitativas na expressão desses genes. Esses rearranjos do DNA são realizados pela **recombinação genética**. Duas amplas classes de recombinação são conhecidas como: recombinação homóloga, ou geral, e recombinação sítio-específica.

Na **recombinação homóloga**, a troca genética envolve sequências de DNA homólogas, em geral, localizadas nas duas cópias de um mesmo cromossomo. Um dos exemplos mais importantes desse tipo de troca entre cromossomos homólogos (denominado *crossing-over*) acontece na meiose. O *crossing-over* ocorre entre cromossomos altamente relacionados nos estágios iniciais de desenvolvimento de óvulos e espermatozoides e permite, por exemplo, que diferentes versões (alelos) do mesmo gene sejam testadas em novas combinações aumentando a chance de pelo menos alguns membros de uma população fundadora sobreviverem em um ambiente que está se modificando. A vantagem deste tipo de “mistura gênica” é tão proveitosa que cruzamentos e reagrupamentos de genes pela recombinação homóloga são também encontrados em bactérias. A recombinação homóloga também atua como um importante mecanismo para a reparação de quebra da fita dupla (DSB, de *DNA double-strand break*), que ocorre nos cromossomos em qualquer etapa do ciclo celular (Seção 2.5 do Capítulo 6). Se estas quebras não forem reparadas, ou se forem reparadas de forma incorreta, elas podem promover eventos mutagênicos, tais como perdas, deleções, duplicações ou translocações cromossômicas. Estes eventos podem levar a processos carcinogênicos em organismos eucarióticos, ou à morte de organismos unicelulares, como leveduras e bactérias.

Na **recombinação sítio-específica**, não é necessária a homologia extensa do DNA, neste caso as trocas ocorridas são curtas. Sequências específicas de nucleotídeos são reconhecidas por uma enzima de recombinação sítio-específica. Esta enzima altera a posição relativa das sequências de nucleotídeos nos genomas. Em alguns casos, estas trocas são programadas e organizadas, como quando um vírus integrado ao cromossomo bacteriano é induzido a deixar o hospedeiro devido a um estresse da bactéria; em outros casos, elas são aleatórias, quando, por exemplo, a sequência de DNA de um transposon é inserida aleatoriamente em um sítio qualquer do cromossomo (ver Capítulo 9).

Muito do que se sabe sobre a bioquímica da recombinação genética provem dos estudos de organismos mais simples, principalmente de *Escherichia coli* e de vírus.

1. Recombinação homóloga

A recombinação homóloga pode ocorrer em qualquer lugar ao longo de duas moléculas complementares de DNA, e o seu principal resultado é que essas moléculas de DNA se sobrepõem e trocam partes (**crossing-over = sobre-cruzamento** – evento de sobreposição cruzada que viabiliza a troca de fitas). As hélices duplas das moléculas de DNA homólogas se quebram, e as duas extremidades quebradas se unem com suas parceiras opostas para formar novamente duas hélices intactas, cada uma composta por partes das duas moléculas de DNA iniciais.

O sítio de troca pode ocorrer em qualquer lugar da sequência de nucleotídeos homóloga das duas moléculas de DNA envolvidas (embora alguns tipos de sequências possam ser favorecidos em relação a outros). No sítio de troca uma fita de uma das moléculas de DNA faz um pareamento de bases com uma das fitas da outra molécula, criando a junção “alternativa” (*staggered joint*), geralmente chamada de **junção heterodúplex**, entre as duas diferentes hélices duplas (**Figura 8.1**); a região de heterodúplex pode ter centenas de pares de bases de comprimento. Não há alteração nas sequências de nucleotídeos no sítio de troca, e a quebra e os eventos de religação ocorrem de forma muito precisa, não havendo perda, ganho ou alteração de nenhum nucleotídeo.

A frequência de recombinação não é constante ao longo de todo o genoma, sendo influenciada por efeitos tanto gerais como locais. A frequência total pode ser diferente nos oócitos e nos espermatozoides; em mulheres, a frequência de recombinação é o dobro da que ocorre

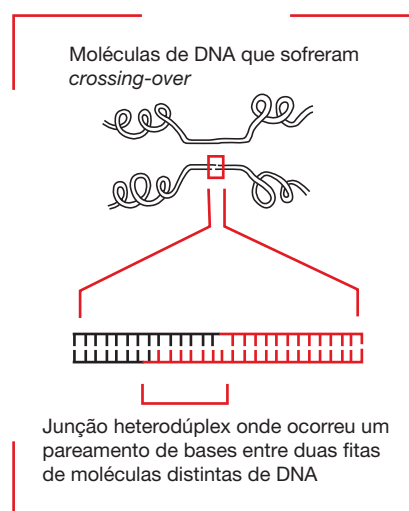


Figura 8.1

Uma junção heterodúplex une duas moléculas de DNA no sítio de *crossing-over*. Normalmente, essas junções possuem centenas de pares de bases de extensão.

em homens. Dentro de um mesmo genoma, a frequência de *crossing-over* depende da estrutura do cromossomo e da distância física entre dois eventos de troca, por exemplo: o *crossing-over* é suprimido nas proximidades das regiões condensadas e inativas da heterocromatina e a ocorrência de um *crossing-over* inibe a ocorrência de outro nas suas proximidades.

1.1 Modelo de Holliday para recombinação homóloga

A ação de conectar duas moléculas de DNA de fita dupla é o ponto principal do processo de recombinação. A **Figura 8.2** ilustra um processo que inicia com quebras em pontos determinados das fitas homólogas de dois dúplices de DNA pareados. Essas quebras permitem o movimento das extremidades livres, criadas pelas clivagens. Cada fita deixa a sua parceira e pareia com a fita complementar do outro dúplice. Essa etapa é chamada de **invasão de fita**. A troca recíproca cria uma conexão entre os dois dúplices, que formam, então, uma **molécula unida** (*joint molecule*). O ponto no qual uma fita de DNA cruza de um dúplice para outro é chamado de **junção recombinante** (*recombinant joint*). Assim, é formada a região de heterodúplex, na qual cada dúplice

possui uma porção que consiste em uma fita de cada molécula de DNA parental.

Uma característica importante de uma junção recombinante é a sua capacidade de mover-se ao longo do dúplice. Esta mobilidade é chamada de **migração de ramificação**. O ponto de ramificação pode migrar em ambas as direções à medida que uma fita é deslocada pela outra. Quando a recombinação envolve duas moléculas de DNA de fita dupla, manipulações topológicas podem ser necessárias. Ambos os dúplices devem estar livres para rotar, ou deve ocorrer um alívio equivalente da repressão topológica. Quando a molécula unida da Figura 8.2 rota um dúplice em relação ao outro, é possível visualizar isso em um plano como uma **estrutura de Holliday** (em homenagem ao seu descobridor Robin Holliday). Esse processo está ilustrado na **Figura 8.3**.

A molécula unida, formada pela troca de fitas, deve ser resolvida em dois dúplices separados. A resolução necessita um par de clivagens adicionais. O resultado da reação depende de qual par de fitas é cortado, como pode ser visto nas Figuras 8.2 e 8.3. Se os cortes forem feitos no par de fitas que não foram originalmente clivadas (o par de fitas que não iniciou a troca), todas

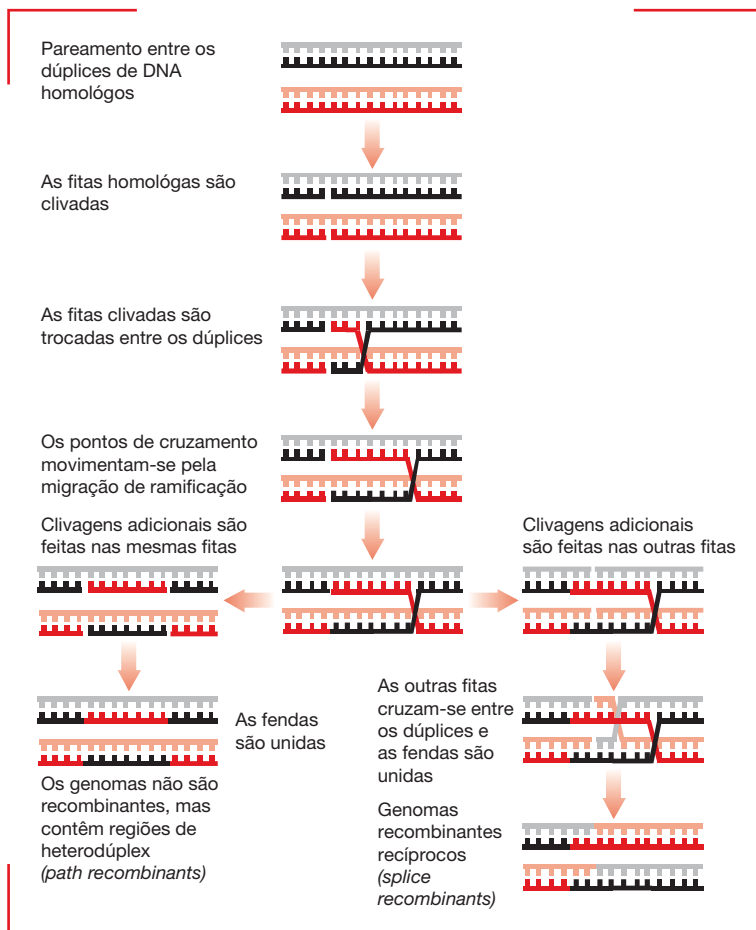
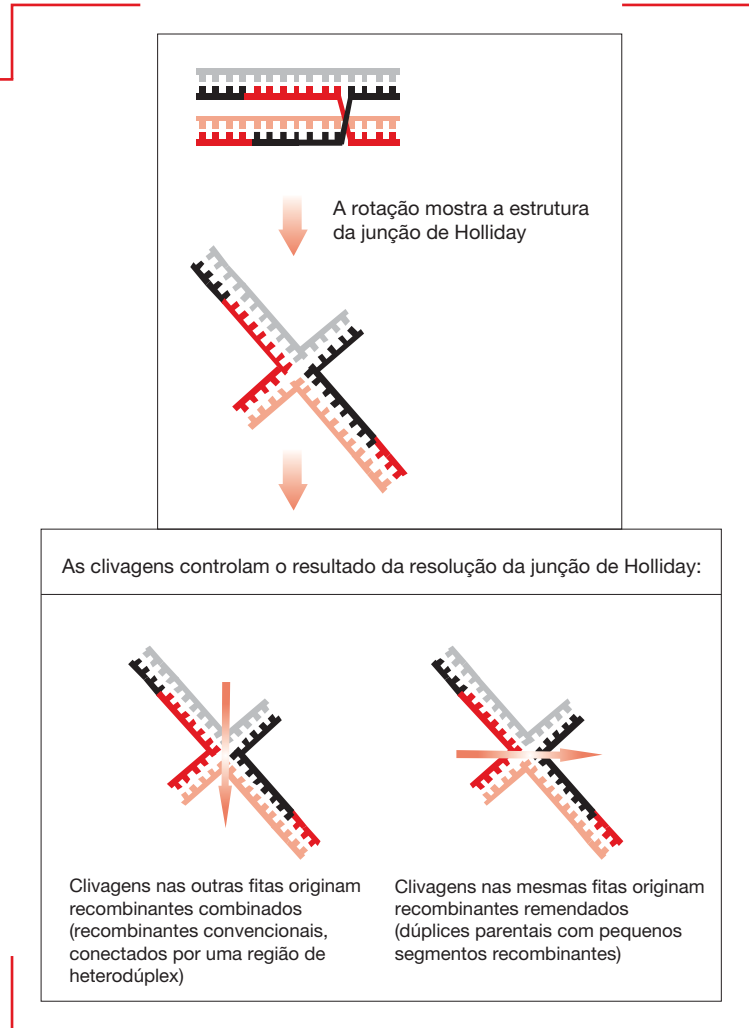


Figura 8.2

A recombinação entre dois dúplices de DNA pareados pode resultar na troca recíproca de fitas simples, na migração de ramificação ou em clivagens.

Figura 8.3

A resolução de uma junção de Holliday pode gerar dúplexes parentais ou recombinantes, dependendo de quais fitas forem cortadas. Ambos os produtos, no entanto, apresentarão uma região de heterodúplex.



as quatro fitas originais são clivadas. Isso resulta em moléculas de DNA **recombinantes combinadas** (ou **recombinantes combinados**, de *splice recombinants*). O dúplex de um DNA parental é covalentemente ligado ao dúplex do outro DNA parental por meio de uma região heterodúplex. Se as mesmas duas fitas envolvidas nas clivagens originais forem clivadas novamente, as outras duas fitas permanecerão intactas. As clivagens liberam os dúplexes parentais, os quais permanecem intactos, salvo que cada um possuirá um vestígio do evento de recombinação na forma de uma região de heterodúplex. Esses recombinantes são chamados de **recombinantes “remendados”** (*patch recombinants*).

Estas alternativas para a resolução da molécula unida estabelecem o princípio de que uma troca de fitas entre dúplexes de DNA sempre cria uma região de heterodúplex, mas a troca pode ser, ou não, acompanhada pela recombinação das regiões adjacentes (*crossing-over* bem-sucedido X *crossing-over* malsucedido).

1.2 Mecanismos de recombinação e suas associações com a reparação do DNA

A recombinação genética entre sequências de DNA de cromossomos homólogos permite o resgate de cromossomos quebrados ou replicados de forma incompleta em todos os organismos. Existem três tipos de reações de recombinação que podem mediar a reparação de cromossomos danificados: reparação de forquilhas de replicação colapsadas, reparação de quebras duplas e reparação para o preenchimento de lacunas por recombinação. As proteínas centrais na recombinação homóloga são relacionadas nos três domínios da vida: RecA em bactérias, RadA em arqueas e Rad51 em eucariotos. A princípio, serão considerados apenas os mecanismos, antes de discutirmos as enzimas que atuam especificamente neles. Em *E. coli* e organismos superiores, a inativação de proteínas que atuam na recombinação provoca redução na viabilidade celular e, com uma elevada taxa de lesões no DNA, as células são levadas à morte.

1.2.1 Reparação de forquilhas de replicação colapsadas

Forquilhas de replicação colapsadas podem se originar como resultado de dificuldades que surgem durante a replicação do DNA, por exemplo, quando a maquinaria de replicação encontra uma lesão e a forquilha de replicação é obrigada a parar o seu avanço, tornando-se alvo para endonucleases. Na ausência de uma reparação por recombinação para corrigir a forquilha quebrada, o cromossomo danificado será degradado.

No processo de reparação da forquilha (**Figura 8.4A**), uma das fitas do cromossomo quebrado é digerida para dar origem a uma extremidade 3'-OH de fita simples. Essa extremidade invade o dúplex homólogo, promovendo o deslocamento de uma das suas duas fitas e a formação de um intermediário ramificado, conhecido como alça-D (*D-loop*, de *displacement*). A extensão da região de heterodúplex entre a fita invasora e a sua parceira com o concomitante deslocamento da outra fita promove a migração de ramificação, criando uma junção de Holliday. A clivagem (ou resolução) apropriada das fitas das moléculas envolvidas, seja na região da alça-D ou na junção de Holliday, e a ligação das fitas restauram a estrutura da forquilha. Após a finalização do processo de reparação da forquilha de replicação, o replissomo deve ser restaurado para reiniciar a replicação.

1.2.2 Reparação de quebras duplas pela recombinação homóloga

Quebras nas duas fitas do DNA são lesões comuns que ocorrem em todas as células. Devido ao fato da acumulação dessas quebras ser prejudicial para as células, foram desenvolvidos múltiplos mecanismos para a sua reparação em diferentes organismos. Entre esses mecanismos, a recombinação homóloga parece ser: a única maneira de reparação desse tipo de lesão em *E. coli*, o mecanismo predominante em alguns eucariotos, incluindo a levedura, e um mecanismo alternativo nos mamíferos, incluindo o homem. Como detalhado no Capítulo 6, a ligação simples das extremidades do DNA (por meio de diferentes processos) é a principal alternativa de reparação em mamíferos e pode resultar na perda de material genético e em modificações cromossômicas.

O mecanismo de reparação de quebras duplas por meio de recombinação foi bastante conservado durante a evolução, bem como suas proteínas mais importantes, sendo necessário para as funções normais da célula. Além disso, esse tipo de lesão pode servir como início de um processo de recombinação entre cromossomos homólogos. A Figura 8.4B mostra como a recombinação pode atuar na reparação de quebras nas duas fitas de uma mesma molécula de DNA.

Após a formação de uma DSB, uma exonuclease degradada, de forma sequencial, a molécula de DNA quebrada, para gerar regiões de DNA de fita simples. Esse processamento cria extensões de fita simples nas moléculas

de DNA quebradas, e essas extensões terminam em extremidades 3'-OH. Em alguns casos, ambas as fitas em uma DSB são processadas, ao passo que em outros apenas a fita que termina com a extremidade 5'-P é degradada. As extremidades de fita simples geradas por esse processamento invadem, então, o dúplex de DNA homólogo não quebrado. Em cada caso, a fita invasora realiza um pareamento de bases com a sua fita complementar na outra molécula de DNA. Uma vez que as fitas invasoras terminam em uma extremidade 3'-OH, elas podem servir como iniciadores para a síntese de DNA. O alongamento dessas extremidades do DNA, utilizando a fita complementar do dúplex homólogo como molde, serve para restaurar as regiões de DNA que foram degradadas durante o processamento das fitas no sítio de quebra (Figura 8.4B).

Caso os dois cópias de DNA originais não apresentem sequências idênticas próximas ao sítio da quebra (por exemplo, apresentando alterações em alguns poucos pares de bases), a informação de sequência pode ser perdida durante o processo de recombinação via sistema de reparação de DSB. No evento de recombinação mostrado na Figura 8.4B, a informação de sequência perdida da molécula de DNA na parte superior, como resultado do processamento do DNA, é substituída pela sequência presente no dúplex da parte inferior, como resultado da síntese de DNA. Essa etapa, não recíproca na reparação de DSB, origina, algumas vezes, um evento genético denominado **conversão gênica**.

As duas junções de Holliday encontradas nos intermediários da recombinação gerados pelo modelo exposto na Figura 8.4B deslocam-se via migração de ramificação e, no final, são resolvidas para terminar a recombinação. Novamente, as fitas que foram clivadas durante a resolução dessas junções de Holliday determinam se os produtos das moléculas de DNA conterão ou não genes reagrupados nas regiões que flanqueiam o sítio de recombinação (ou seja, resultando em *crossing-over*). As diferentes maneiras para resolver um intermediário de recombinação contendo duas junções de Holliday são explicadas no **Quadro** da página 169.

Uma situação alternativa para a reparação de quebra da fita dupla, conhecida como **anelamento de fita dependente de síntese (SDSA, de *synthesis-dependent strand annealing*)**, ocorre com bastante frequência para a reparação de quebras provocadas durante a mitose, que, frequentemente, não estão associadas com *crossing-overs*. O mecanismo de SDSA é semelhante ao modelo de reparação de DSB, nas etapas iniciais do processamento das extremidades da DSB e na invasão de um cromossomo homólogo, mas, ao invés de capturar a segunda extremidade da DSB em um intermediário de recombinação, a fita invasora é deslocada após uma etapa de síntese e volta a reanelar com a região de fita simples da outra extremidade da DSB (Figura 8.4C). Isso resulta no término do processo de reparação da quebra sem a necessidade de clivagem de nenhuma estrutura ramificada e, portanto, não gera produtos recombinantes.

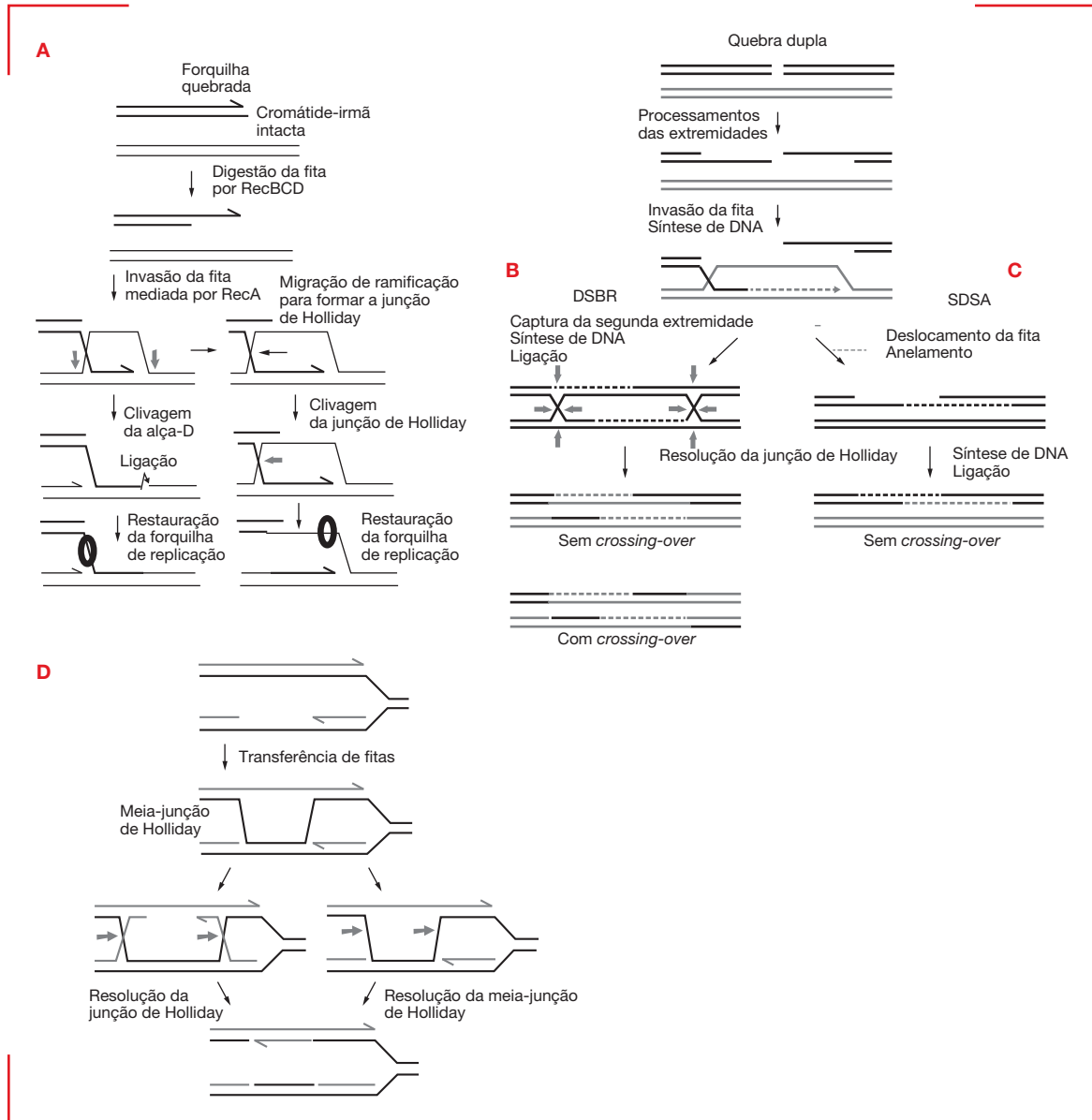


Figura 8.4

Reações de recombinação para a reparação de cromossomos danificados. (A) Reparação de forquilhas de replicação colapsadas: o cromossomo quebrado é digerido por RecBCD para criar uma região de DNA de fita simples 3'-OH. RecA é colocada sobre esse DNA de fita simples e catalisa a invasão de fita na região de homologia da cromátide-irmã intacta, gerando a alça-D. A migração de ramificação estende a região de heterodúplex e gera uma junção de Holliday. As estruturas ramificadas são resolvidas por clivagem da alça-D ou da junção de Holliday, formando novas forquilhas de replicação. **(B)** Reparação de quebras duplas no DNA (DSBR): nesse processo, a reparação é iniciada pela digestão da DSB para fornecer extremidades de DNA de fita simples 3'-OH. A invasão das fitas de uma molécula de DNA homóloga por essas extremidades 3'-OH é seguida pela síntese de DNA a partir das extremidades invasoras. A captura da segunda extremidade da quebra forma um intermediário com duas junções de Holliday, que podem ser resolvidas em produtos com ou sem *crossing-overs*. **(C)** Alternativamente, uma DSB pode ser processada pelo mecanismo de SDSA, pelo deslocamento da fita invasora e o anelamento da região de fita simples que foi estendida à região de fita simples da outra extremidade da quebra, seguido por preenchimento das lacunas e ligação. Os produtos reparados por SDSA nunca geram *crossing-overs*. **(D)** Reparação para o preenchimento de lacunas por recombinação: a lacuna gerada durante a replicação é preenchida por recombinação da região correspondente da cromátide-irmã. O recrutamento das extremidades que flanqueiam a junção produz junções de Holliday que podem ser resolvidas por clivagem.



Resolução do intermediário de recombinação com duas junções de Holliday

A maneira como as junções de Holliday presentes em um intermediário de recombinação são clivadas (resolvidas) representa um profundo impacto estrutural nos produtos das moléculas de DNA. Os produtos podem ter o DNA que flanqueia o sítio de recombinação reorganizado (nos produtos combinados/com sobre cruzamento), ou não (nos produtos remendados/sem sobre cruzamento), dependendo de como a resolução é alcançada. Uma vez que os intermediários gerados pelo sistema de reparação de DSB contêm duas junções de Holliday, pode ser difícil visualizar quais produtos serão gerados por todas as combinações possíveis dos eventos de clivagem da junção de Holliday. Existe um padrão simples que determina se produtos com ou sem sobre cruzamento serão gerados.

Para explicar as diferentes maneiras, como esses intermediários podem ser resolvidos, considere as duas junções (marcadas por x e y) na **Figura Q.1**. Para cada junção, existem dois possíveis sítios de

clivagem (marcados por sítio 1 e sítio 2). Existe uma regra simples que determina se a resolução resultará ou não em produtos com sobre cruzamento *versus* produtos sem sobre cruzamento. Caso as junções forem clivadas da mesma maneira, ou seja, em ambos os lados do sítio 1 ou do sítio 2, então, produtos sem sobre cruzamento serão gerados. Um exemplo desse tipo de produto é mostrado no painel b da figura. Essas moléculas são geradas quando ambas as junções de Holliday forem clivadas no sítio 2. Note que os alelos marcadores *A/B* e *a/b* ainda estão nas mesmas moléculas de DNA, da mesma forma como eles se encontravam nos cromossomos parentais. A clivagem de ambas as junções no sítio 1 também gera produtos sem sobre cruzamento.

Em contraste, quando as duas junções são clivadas utilizando sítios diferentes, então, os produtos com sobre cruzamento são gerados. Um exemplo desse tipo de resolução é mostrado no painel c da Figura Q.1, onde a junção x foi clivada no sítio 1, e a junção y foi clivada no sítio 2. Note que, nesse caso, o gene *A* está ligado ao gene *b*, e o gene *a* está ligado ao gene *B*; assim, ocorreu o reagrupamento dos genes flanqueadores. A clivagem da junção x no sítio 2 e da junção y no sítio 1 também gerou produtos com sobre cruzamento.

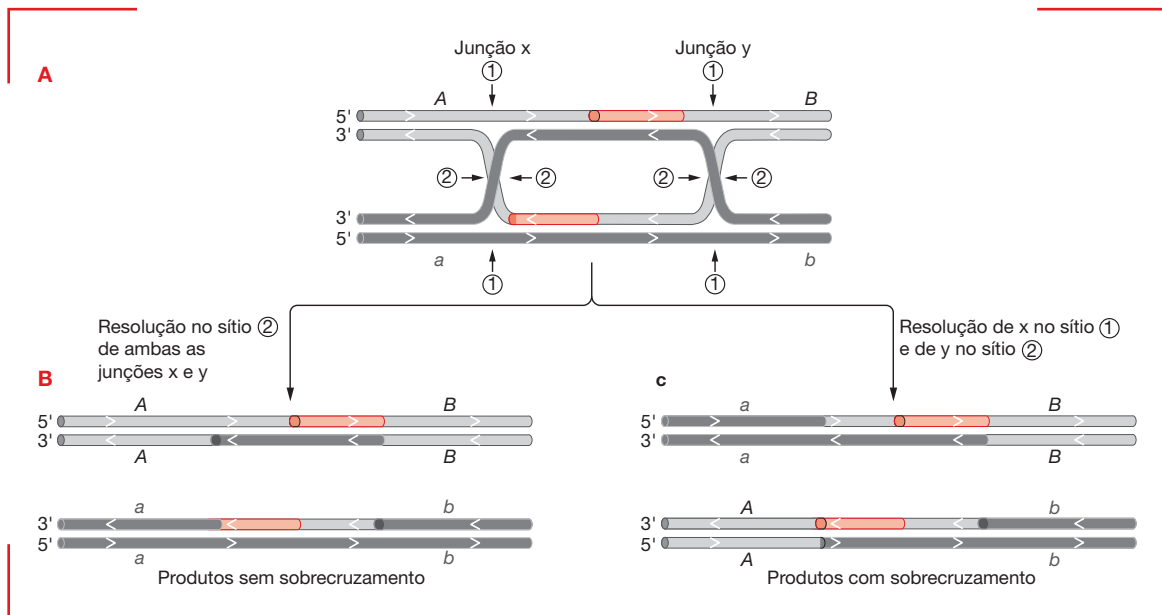


Figura Q.1

As duas maneiras possíveis de resolver um intermediário a partir do sistema de reparação de DSB. As moléculas de DNA parentais são semelhantes àsquelas na Figura 8.4.

1.2.3 Reparação para o preenchimento de lacunas por recombinação

No terceiro tipo de reparação do DNA associada com a recombinação homóloga, lacunas de DNA de fita simples podem ser preenchidas pelas reações de trocas de fitas entre cromossomos homólogos ou cromátides-irmãs (Figura 8.4D). Uma lacuna de fita simples em uma cromátide-irmã pode ser produzida por uma replicação incompleta, na qual a síntese foi bloqueada por uma lesão no DNA-molde ou na fita nascente, por uma estrutura de DNA secundária ou pela ligação muito forte de proteínas. A Figura 8.4D mostra um modelo para esse processo. A região de fita simples da lacuna parecia com a sua região complementar que pertence a uma molécula de DNA intacta. Essa invasão de fita pode ser auxiliada pelo deslocamento de uma ou ambas as fitas que flanqueiam a lacuna. Como essa sinapse não envolve uma extremidade livre, esse pareamento apresenta problemas topológicos, que devem ser resolvidos por ação de topoisomerases. A síntese de DNA iniciada a partir das extremidades substitui a fita transferida, originando uma molécula ramificada constituída de uma ou mais fitas cruzadas. A clivagem dessas fitas restaura as duas moléculas de DNA de fita dupla e a lacuna é preenchida. Como no modelo de reparação de DSB, a clivagem nessas junções também pode gerar produtos recombinantes ou não.

1.3 Conversão gênica

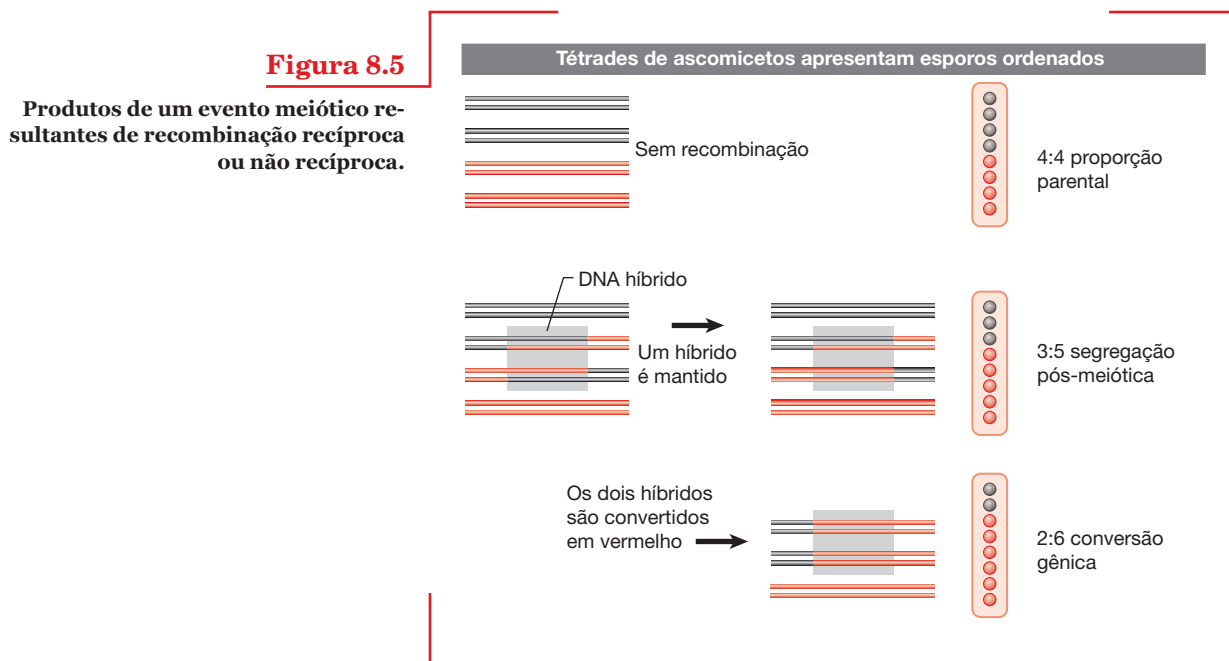
Investigações iniciais de *crossing-over* entre diferentes genes indicavam que parecia ser obrigatória a ocorrência de recombinantes recíprocos. Entretanto, foram descobertas exceções quando surgiram recombinantes entre sítios próximos do mesmo gene, nos quais o comporta-

mento não recíproco era observado com frequência. Este fenômeno, chamado de **conversão gênica**, foi bem estudado em leveduras e *Neurospora*, em que todos os produtos de um único evento meiótico podem ser analisados. Nesses organismos, ao contrário da segregação normalmente esperada (2:2) dos alelos envolvidos, foram observados casos de segregação 3:1, nos quais um alelo é perdido e uma cópia extra do outro aparece.

Como ilustrado na **Figura 8.5**, o modelo de segregação final dos genes localizados ao redor do sítio de cruzamento pode ser afetado por enzimas de reparação, que reconhecem as distorções do heterodúplex na hélice dupla e removem aleatoriamente um dos pares de bases mal-pareados. Dependendo de qual base for removida, tanto a proporção 2:2 como a de 3:1 podem ser encontradas. Se nenhuma troca for feita na sequência, as fitas irão se separar na replicação subsequente, cada uma dando origem a um dúplex que irá perpetuar a sua informação. Esse evento é chamado de **segregação pós-meiótica**, pois reflete a separação das fitas de DNA após a meiose. O outro tipo de efeito, que resulta na conversão gênica, causa a troca de fitas entre as moléculas de DNA e a subsequente transferência não recíproca da informação de uma cromátide para a outra. Como os heterodúplex são curtos em relação à distância entre muitos genes recombinantes, o processo de conversão gênica não é muito comum, prevalecendo o aparecimento de recombinantes recíprocos na maioria dos casos.

1.4 Enzimas que atuam na recombinação homóloga em procariontos

Entre o que se sabe sobre as propriedades físicas do DNA, não há razões para se esperar que duas moléculas que se quebram, passem a se unir simplesmente porque uma



delas possui uma quebra ou uma região de fita simples. Como demonstrado para bacteriófagos, bactérias e leveduras, e válido para todos os organismos em geral, existem enzimas que catalisam os passos bioquímicos da recombinação.

Assim, para que a recombinação aconteça, os DNAs que servirão de substrato para o processo deverão ser processados por nucleases e helicases, que produzirão regiões de DNA de fita simples. Essas regiões serão cobertas pela proteína central na troca de fitas, a proteína RecA. *In vivo*, RecA é ativamente colocada no fragmento de DNA de fita simples por pelo menos dois mecanismos distintos, envolvendo os complexos RecBCD ou RecFOR. O filamento de DNA que fica recoberto por RecA, além de proteger o DNA de eventuais degradações, potencializa os eventos subsequentes de sinapse e troca de fitas. Várias outras proteínas modulam a estabilidade do filamento pré-sináptico recoberto por RecA. Proteínas estruturais específicas também devem coordenar ou manter unidas as extremidades do DNA.

1.4.1 Proteína RecA

Muito do conhecimento sobre o mecanismo de recombinação homóloga é devido aos estudos realizados *in vitro* com a proteína RecA de *E. coli*. RecA possui um papel central na recombinação, atuando como uma proteína estrutural e em reações catalíticas. Estudos bioquímicos demonstraram que a RecA é capaz de parear duas moléculas de DNA homólogas e catalisar as reações de trocas de fitas, levando à formação do heterodúplex de DNA.

A RecA é uma proteína, de 37,8 kDa, que polimeriza sobre o DNA para formar uma estrutura similar a um colar de pérolas, em que ocorrem as reações de pareamento. As propriedades enzimáticas da RecA mostram por que as regiões de DNA de fita simples ou quebradas são necessárias para iniciar a recombinação. Nenhuma reação ocorre se duas moléculas homólogas, mas intactas, são misturadas com a proteína RecA. Portanto, RecA reconhece especificamente o DNA de fita simples e anela este segmento a uma sequência complementar de um dúplice homólogo, substituindo, de forma simultânea, a fita original complementar pela nova fita (**Figura 8.6**). A RecA liga-se primeiro ao DNA de fita simples a uma razão fixa de cerca de um polipeptídeo para cada três a cinco nucleotídeos, formando um filamento DNA-proteína. Se ATP está presente, a proteína RecA do filamento pode, então, desenrolar sucessivamente diferentes partes do dúplice, na tentativa de anelar o filamento de fita simples a uma região desenrolada. Quando uma região complementar da sequência da molécula de fita dupla é encontrada, inicia-se o processo de pareamento. Uma vez que uma pequena porção é pareada de forma correta, a energia do ATP direciona a reação de pareamento até o final, movendo o conjunto na direção 5'→3' em relação à molécula de fita simples. A proteína RecA é deslocada quando a nova molécula de DNA híbrida é formada.

Bacteriófagos, como T7 e λ , utilizam um sistema enzimático mais simplificado para promoverem a sua re-

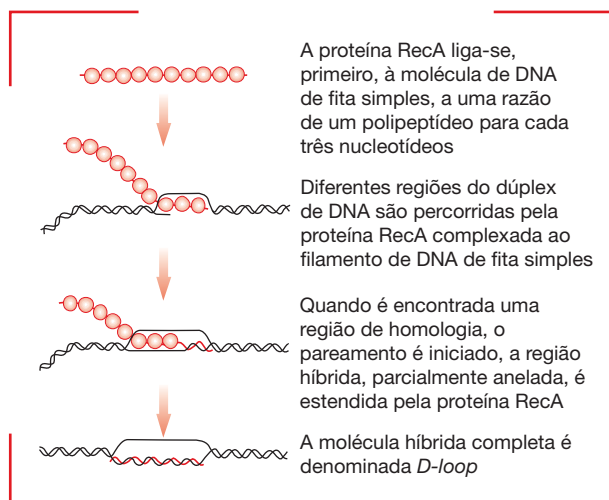


Figura 8.6

Atividade da proteína RecA, mostrada *in vitro*, em uma reação utilizando fragmentos de DNA de fita simples e de fita dupla homólogos.

combinação. Nucleases são utilizadas para criar regiões de fita simples na molécula de DNA, e proteínas, que se ligam a DNA, estendem esta região de fita simples facilitando o pareamento por meio de colisões aleatórias. É provável que esta estratégia seja bem-sucedida devido à alta concentração de genomas virais em uma célula infectada, o que aumenta a probabilidade das sequências de fitas simples complementares parearem ao acaso. Organismos superiores, por outro lado, necessitam de enzimas sofisticadas (similares à RecA) para direcionar a busca de regiões homólogas devido à abundância de sequências. No entanto, os mecanismos de pareamento e trocas de fitas em eucariotos podem diferir bastante daqueles catalisados pela RecA.

A proteína RecA de *E. coli* possui um papel central na regulação da recombinação genética, na reparação do DNA e na mutagênese induzida por UV. Como já visto, a RecA tem uma função-chave na resposta SOS (ver Capítulo 6), catalisando a digestão do repressor LexA e da proteína UmuD (DNA-polimerase IV). Além dessas funções, RecA catalisa reações *in vitro*, que são consistentes e atuam direto no processo de recombinação.

Estudos iniciais demonstraram que a proteína RecA é uma ATPase DNA-dependente. A hidrólise de ATP é necessária para a formação do filamento nucleoproteico, porque os oligonucleotídeos menores que 50 pb são cofatores fracos. O consumo de ATP é uniforme ao longo de toda a extensão do filamento. No entanto, embora ocorram ao mesmo tempo, as reações de troca de fitas e hidrólise de ATP são independentes. A reação de troca de fitas é isoenergética (para cada par de bases quebrado, outro par é formado) e ocorre mesmo na ausência de ATP. Dessa forma, a necessidade de RecA ligar-se a ATP, e não hidrolisá-lo, pode ser explicada da seguinte maneira: no

processo de ligação de ATP, RecA sofre uma transição alostérica para um estado de alta afinidade de ligação ao DNA, o filamento é formado sobre o DNA, e as moléculas de DNA realizam as reações de pareamento. O processo de hidrólise de ATP resulta na produção de moléculas de ADP ligadas ao filamento, induzindo a transição para um estado de baixa afinidade pelo DNA. A formação desse estado de baixa afinidade resulta na liberação do heterodúplex de DNA.

Embora a proteína RecA transfira, inicialmente, apenas uma fita para o interior de um dúplice de DNA, a segunda troca ocorre como consequência da primeira. A proteína RecA coordena o anelamento entre a extremidade de fita simples de uma molécula e a região complementar de outra. A junção se difunde envolvendo toda a região do dúplice, desde o início do pareamento, movimentando-se para a direita e desenrolando a hélice dupla intacta, como ilustrado na **Figura 8.7**. Simultaneamente, a segunda hélice híbrida é formada, pelo preenchimento da região que antes era de fita simples, utilizando como molde a cadeia intacta, vinda da outra molécula de DNA. Assim, a proteína RecA pode atuar guiando a formação de uma hélice com quatro fitas, composta pelas duas moléculas de DNA, acelerando o proces-

so de troca de fitas e permitindo que a junção se mova ao longo de toda a região do dúplice (Figura 8.7).

1.4.2 Proteínas RecBCD

A RecA não é a única proteína envolvida no processo de recombinação, embora seja a mais importante. O complexo formado pelas proteínas RecB, RecC e RecD, com 300 kDa, também executa diversas atividades. Ele possui uma potente atividade nucleásica, que degrada DNA, uma atividade de helicase, que pode desenrolar um dúplice de DNA na presença da proteína SSB (proteína de ligação ao DNA de fita simples, ver Capítulo 5), além de possuir atividade de ATPase. Todas essas atividades combinadas expõem regiões de DNA de fita simples com uma extremidade 3' livre, e é possível que a RecA se ligue a esse segmento de DNA para iniciar as reações de pareamento. Uma importante característica do complexo RecBCD é que ele inicia o processo de desenrolamento e degradação do DNA somente em uma molécula contendo uma extremidade livre. Uma vez ligado à extremidade da cadeia, a energia do ATP é utilizada para percorrer a molécula, desenrolando o DNA durante a sua passagem e degradando a fita que possui a extremidade 3'-OH livre (**Figura 8.8**).

A terceira atividade do complexo RecBCD, que acontece enquanto o DNA está sendo desenrolado, é decisiva para a ocorrência da recombinação. Experimentos genéticos demonstraram que o complexo RecBCD promove a recombinação com uma maior frequência em cadeias de DNA que contenham o chamado sítio *chi*, com sequência 5'-GCTGGTGG-3'. Quando um DNA, que está sendo desenrolado, possui a sequência *chi*, uma atividade endonucleásica específica de RecBCD cliva uma das fitas do DNA em uma região próxima ao sítio *chi* (Figura 8.8). O reconhecimento do sítio *chi* faz com que a subunidade RecD se dissocie do complexo ou se torne inativa, fazendo o complexo perder a sua atividade endonucleásica, mas manter suas funções de helicase. Como consequência, é deixada para trás uma região de DNA de fita simples, sobre a qual RecA pode ligar-se e iniciar a troca de fitas com uma sequência homóloga.

O DNA de *E. coli* possui centenas de sítios *chi*, cerca de um a cada 5-10 kb e, portanto, oferece muitas oportunidades para a ação do complexo RecBCD. Estas sequências, em geral não estão disponíveis para o complexo em uma célula normal, em que o DNA não apresenta extremidades livres, mas são particularmente importantes na conjugação bacteriana. Neste processo, uma extremidade do DNA é introduzida pela linhagem doadora (macho) em uma linhagem receptora (fêmea). É possível que RecBCD se ligue na extremidade do DNA introduzido, movendo-se ao longo da molécula e clivando-a nos sítios *chi*, promovendo, dessa forma, a recombinação com o DNA da célula receptora.

Para RecBCD iniciar a sua atividade nucleásica é necessário que a extremidade do DNA esteja emparelhada (cega). A atividade nucleásica de RecBCD é inibida por regiões de fita simples 3'-OH ou 5'-P de tamanhos tão pequenos quanto 25 nucleotídeos. Portanto, durante o

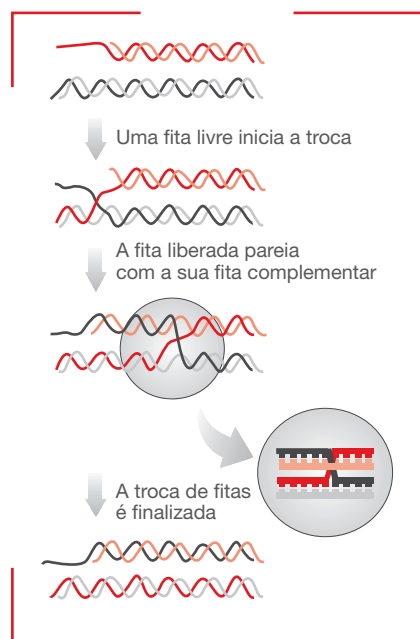


Figura 8.7

A troca de fitas, realizada pela proteína RecA entre um dúplice de DNA, que possui uma região de fita simples e um dúplice intacto, origina uma molécula unida com a mesma estrutura observada em uma etapa intermediária da recombinação.

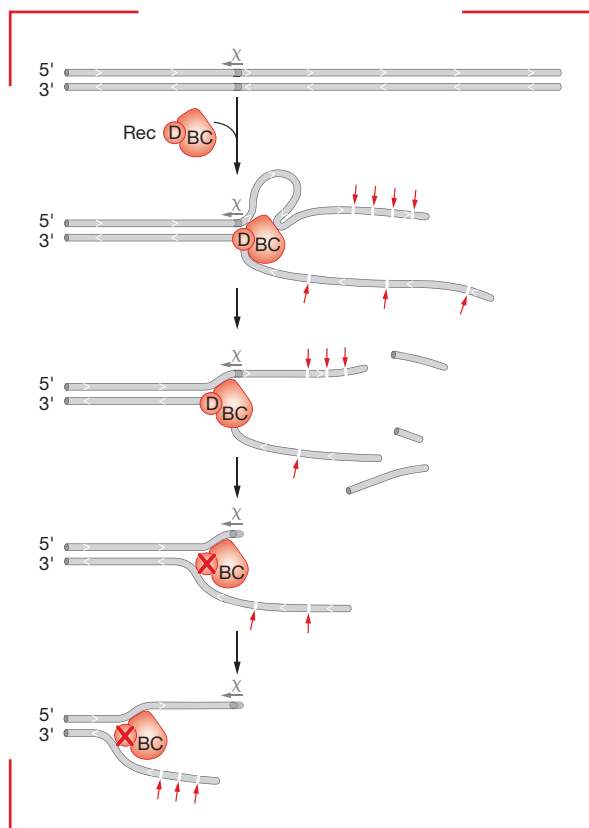


Figura 8.8

A nucleasa RecBCD aproxima-se de uma sequência *chi*, degradando o DNA enquanto avança; em uma região próxima ao sítio *chi*, ela realiza uma clivagem endonucleolítica, perde RecD, mas continua com a sua atividade de helicase, gerando uma região de DNA de fita simples, na qual RecA poderá ligar-se e iniciar um processo de recombinação homóloga.

processamento de quebras cromossômicas de ocorrência natural por RecBCD, existe a necessidade de uma etapa de processamento das extremidades para que se tornem emparelhadas (ou *end-blunting*), pela ação de DNA-exonucleases de fita simples. O efeito na atividade nucleásica de RecBCD sobre DNAs de fagos linearizados sugerem que as DNA-exonucleases de fita simples ExoI, RecJ e SbcCD desempenham o papel de potencializar substratos de DNA para RecBCD.

O reconhecimento do sítio *chi*, além de atenuar a atividade nucleásica de RecBCD, reduzir sua atividade de helicase pela metade e fazer com que a fita com a extremidade 3'-OH seja preservada (a fita com a extremidade 5'-P continua a ser degradada), também estimula o complexo RecBCD a colocar ativamente a proteína RecA no fragmento de fita simples com extremidade 3'-OH emergente. Essa atividade ocorre pelo domínio C-terminal da subunidade RecB, que forma complexos estáveis com RecA. Devido a essa colocação ativa e imediata de RecA

no filamento de DNA de fita simples que está sendo formado, RecBCD é capaz de impedir que a proteína SSB se ligue nesse DNA. Isso explica por que o sistema RecBCD não é inibido durante a superprodução de SSB.

1.4.3 Proteínas RecFOR

Um mecanismo alternativo para a colocação de RecA em um DNA de fita simples é fornecido pelas proteínas RecFOR. Essas proteínas fornecem uma solução para RecA ter acesso a um DNA de fita simples que esteja coberto por proteínas SSB. O gene *recF* foi descoberto como sendo um dos componentes de um processo de recombinação inicialmente denominado “sistema RecF”, que funcionaria independentemente da nucleasa RecBCD. Análises subsequentes identificaram outros dois genes, *recO* e *recR*, com efeitos bastante similares, definindo, assim, o agora denominado “sistema RecFOR”. Análises genéticas sugerem que RecF, RecO e RecR atuem em conjunto para promover a ação de RecA.

A caracterização *in vitro* das proteínas RecFOR confirmou que elas agem no controle da formação do filamento DNA-RecA, direcionando RecA para lacunas de DNA de fita simples. Ao contrário de RecBCD, as proteínas RecFOR não atuam em quantidades estequiométricas. Ao invés disso, RecF, RecO e RecR atuam de forma independente ou aos pares. As proteínas RecO e RecR agem como “mediadoras”, possibilitando que RecA substitua SSB em um DNA de fita simples. A proteína RecO pode se ligar diretamente em um DNA de fita simples coberto por SSB, bem como em um DNA de fita simples livre de proteínas. O par RecOR promove a dissociação de SSB e a substituição por RecA em uma região que apresente uma lacuna de fita simples em um DNA de fita dupla. O papel de RecF parece ser o de direcionar e restringir a formação do filamento de RecA na região da lacuna de fita simples. A proteína RecF liga-se, de preferência, às junções DNA de fita simples/DNA de fita dupla e, em conjunto com RecR, impede a colocação de RecA em regiões de DNA de fita dupla. O fenótipo de mutantes para RecF pode ser parcialmente suprimido pela superprodução de RecOR, suportando a ideia de que RecOR exerce a função de mediador de RecA e que RecF direcionaria RecA para a localização correta.

1.4.4 Outras proteínas que participam da recombinação homóloga procarionótica

Um dos pontos mais decisivos da recombinação é a resolução da junção de Holliday, o que determinará a ocorrência ou não de uma recombinação recíproca (*crossing-over* bem-sucedido X *crossing-over* malsucedido). A migração de ramificação, a partir do ponto de troca, determina a extensão da região de DNA híbrida (região de heterodúplex), independentemente da formação de recombinantes bem ou malsucedidos. As proteínas envolvidas na estabilização e na resolução das junções de Holliday foram identificadas em *E. coli* como produtos dos genes *ruvA* e *ruvB*. As proteínas RuvA e RuvB aumentam a frequência de formação das estruturas de heterodúpli-

ces, e RuvA reconhece a estrutura da junção de Holliday. A proteína RuvA liga-se às quatro fitas de DNA no ponto de cruzamento e forma dois tetrâmeros, que posicionam o DNA no meio deles, como em um “sanduíche”. RuvB é uma ATPase que funciona como um hexâmero e possui atividade de helicase, agindo como o motor da migração de ramificação. Os anéis hexaméricos de RuvB se ligam ao redor de cada dúplex de DNA, a montante do ponto de cruzamento. Um diagrama do funcionamento do complexo é apresentado na **Figura 8.9**. O complexo RuvAB pode acelerar a migração de ramificação para até 10 a 20 pb/segundo e deslocar RecA do DNA durante a sua ação.

O terceiro gene *ruv*, o *ruvC*, codifica uma endonuclease que reconhece as junções de Holliday e pode clivá-las *in vitro*, para resolver estruturas intermediárias da recombinação. Uma sequência consensual de quatro nucleotídeos funciona como um “sítio quente” para promover a resolução da junção de Holliday mediada por RuvC. O tetranucleotídeo (ATTG) é assimétrico e, assim, pode direcionar a resolução em relação ao par de fitas que será clivado. Isso determina se o resultado será a formação de recombinantes remendados (sem uma recombinação completa ou malsucedidos) ou a formação de recombinantes combinados (com recombinação entre as regiões adjacentes ou recombinantes bem-sucedidos).

Outra proteína que também atua na resolução de intermediários de recombinação é a helicase RecG. Essa enzima pode se ligar a junções de Holliday, alças-D e outras estruturas ramificadas, a partir das quais promove migração de ramificação dependente de ATP. Essa helicase pode migrar não apenas junções de Holliday, mas também, formar, *in vitro*, junções de Holliday a partir de regressão da forquilha, um mecanismo que auxilia na restauração da replicação em forquilhas de replicação bloqueadas. RecG liga-se com grande facilidade às estruturas de alça-D, e sua atividade de helicase é direcional sobre esse tipo de substrato. Ao contrário de RuvAB, a atividade de helicase de RecG atua para desfazer as estru-

turas de alças-D e, *in vitro*, pode reverter a troca de fitas mediada por RecA.

Os complexos RuvABC e RecG contribuem para a recombinação *in vivo*. Embora as atividades catalíticas de RuvABC para promover a migração de ramificação e a clivagem possam atuar para estabilizar e resolver junções de heterodúplexes, tanto *in vivo* como *in vitro*, pouco se sabe sobre a contribuição da atividade de helicase da RecG para a recombinação, uma vez que sua direcionalidade preferencial dissolve essas junções. Uma possibilidade é que RecG promova a etapa de dissolução do processo de reparação por SDSA, no qual, após a síntese de DNA no intermediário de recombinação, a sinapse é revertida para permitir o anelamento dos cromossomos quebrados.

A troca de fitas fornece extremidades de DNA 3'-OH que podem ser alongadas pela síntese de DNA. Essa síntese pode estabilizar moléculas unidas (Figura 8.4B) e restaurar a informação perdida a partir das quebras. A síntese de DNA que segue a recombinação pode envolver a ação de uma única DNA-polimerase, ou estar associada com o restabelecimento de uma nova forquilha de replicação completa, com a síntese coordenada das fitas líder e tardia.

A coordenação das últimas etapas da recombinação com a replicação provavelmente está sujeita à regulação, dependendo da natureza do intermediário de recombinação formado. Dessa forma, faz sentido supor que em eventos de recombinação de uma só quebra, como no caso da reparação de forquilhas de replicação colapsadas (Figura 8.4A), seja obrigatório o estabelecimento de novas forquilhas para permitir uma eficiente retomada da replicação após o colapso da forquilha inicial. Entretanto, eventos de reparação de quebras da dupla fita (Figura 8.4B e 8.4C) e recombinação para o preenchimento de lacunas de fita simples (Figura 8.4D) podem necessitar apenas uma limitada extensão de síntese de DNA nos intermediários de recombinação para restaurar a informação perdida. O estabelecimento de forquilhas de replicação, nesses casos, pode ser desnecessário. Os fatores e os mecanismos que devem atuar nessa discriminação ainda não são completamente entendidos.

E. coli possui cinco DNA-polimerases (Capítulo 5). Durante a replicação do cromossomo, a DNA-polimerase III sintetiza grande parte do DNA com a DNA-polimerase I envolvida na maturação dos fragmentos de Okazaki. As DNA-polimerases II, IV e V são induzidas pela resposta SOS aos danos de DNA e funcionam nas reações de síntese translesão. O envolvimento dessas últimas polimerases de reparação nos eventos de recombinação ainda não é bem definido, possivelmente devido aos seus papéis redundantes. As propriedades da DNA-polimerase II tornam essa enzima bem adequada para a síntese em lacunas de DNA, uma vez que ela parece estar presente em intermediários de recombinação. A DNA-polimerase II apresenta uma alta fidelidade, devido a sua atividade exonucleásica 3'→5' e é especificamente estimulada pelo DNA de fita simples coberto por SSB. Existem evidências, também, que a DNA-polimerase IV (ou proteína DinB) seja requerida para determinados eventos de recombinação iniciados por pa-

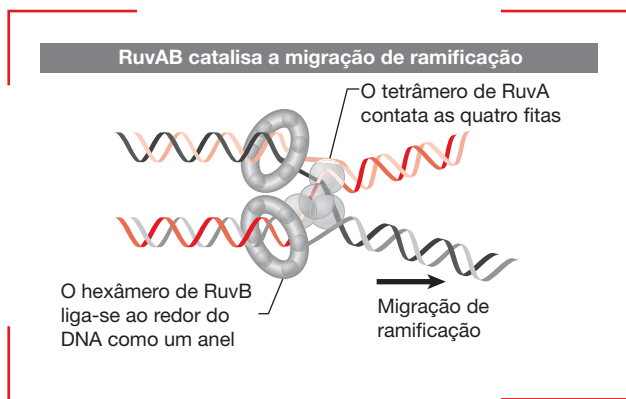


Figura 8.9

As proteínas RuvAB formam um complexo assimétrico que promove a migração de ramificação de uma junção de Holliday, que será, posteriormente, clivada pela ação de RuvC.

radas na replicação e que a polimerização dependente de DinB acompanha a reparação de quebras de fita dupla.

O restabelecimento da forquilha de replicação necessita de fatores específicos para recolocar a helicase da forquilha de replicação, DnaB, nas estruturas de DNA reparadas. Uma vez que DnaB tenha sido colocada, acredita-se que o replissomo composto por DNA-polimerase III e DNA-primase se reuna automaticamente, por meio de interações com DnaB.

1.5 Recombinação homóloga em eucariotos

1.5.1 A segregação correta dos cromossomos durante a meiose depende da recombinação homóloga

A recombinação homóloga é fundamental para a meiose. Durante a meiose, a recombinação homóloga é necessária para o correto pareamento dos cromossomos e, assim, para a manutenção da integridade do genoma. Cada par

de cromossomos, independentemente de seus tamanhos, deve apresentar pelo menos um evento de *crossing-over* (o *crossing-over* obrigatório), que é essencial para a segregação dos cromossomos homólogos na primeira divisão meiótica. Essa recombinação também reembaralha os genes entre os cromossomos parentais, garantindo a variação nos conjuntos de genes que são passados para a próxima geração.

A **Figura 8.10** compara os progressos visíveis dos cromossomos pela meiose e pelas interações moleculares que estão envolvidas na troca de material entre os dúplexes de DNA. A meiose inicia com uma prófase prolongada, cujos cinco estágios estão resumidos na figura.

O início da meiose é marcado pelo ponto no qual os cromossomos individuais tornam-se visíveis. Cada um desses cromossomos foi previamente replicado e consiste de duas cromátides-irmãs, cada uma delas contendo um DNA de fita dupla. Os cromossomos homólogos aproximam-se um do outro e começam a parear entre si em uma ou mais regiões, formando os **bivalentes**. O

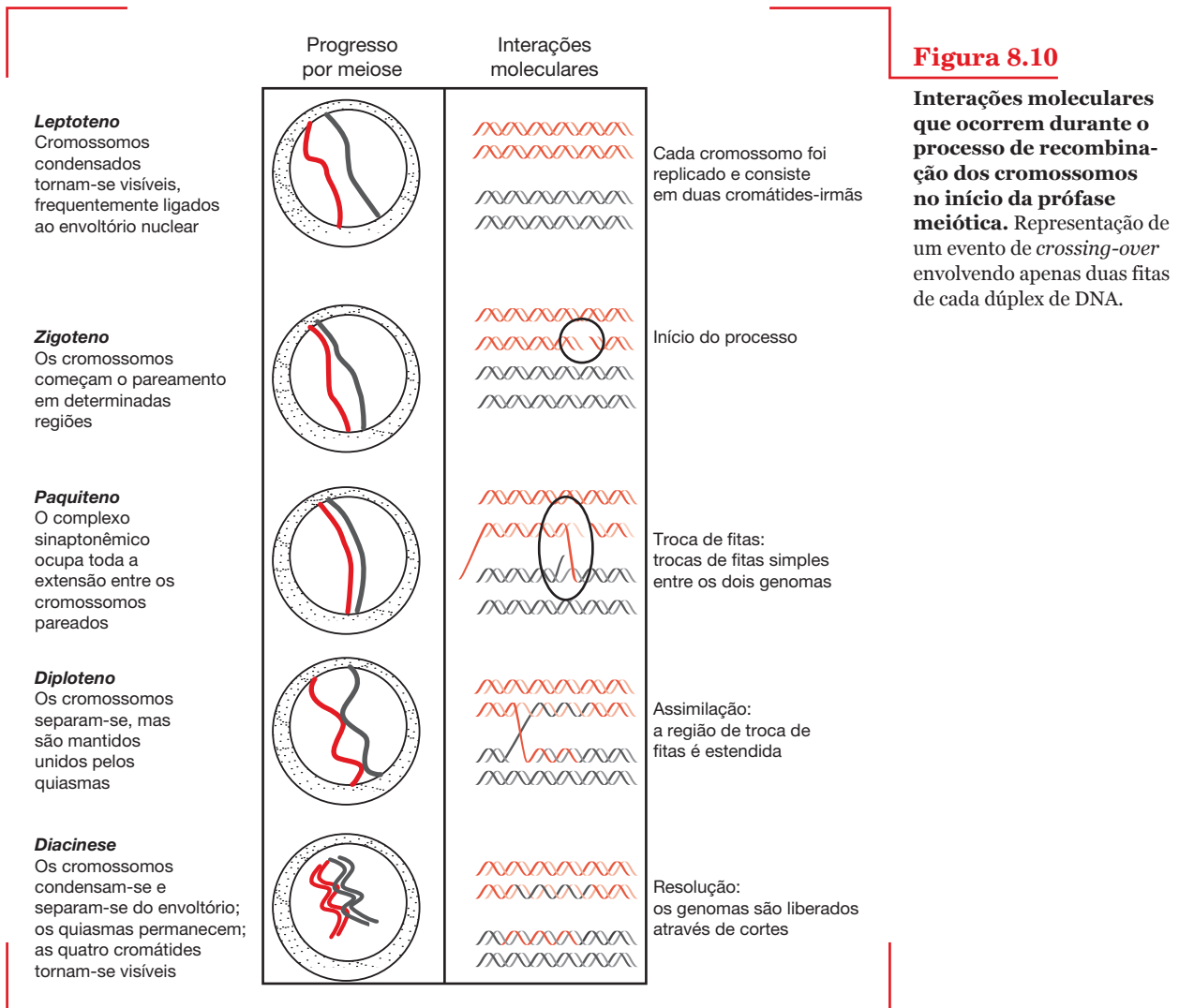


Figura 8.10

Interações moleculares que ocorrem durante o processo de recombinação dos cromossomos no início da prófase meiótica. Representação de um evento de *crossing-over* envolvendo apenas duas fitas de cada dúplex de DNA.

pareamento se estende até que cada cromossomo esteja unido com o seu homólogo em toda a sua extensão. Esse processo é chamado de **sinapse** ou **pareamento cromossômico**. Quando o processo é completado, os cromossomos estão lateralmente associados na forma de um **complexo sinaptonêmico**, que possui uma estrutura característica em cada espécie, embora exista uma ampla variação de detalhes entre as espécies.

A recombinação entre os cromossomos envolve a troca física de partes, geralmente representada por um evento de quebra seguido por um de religação, na qual duas cromátides não irmãs (cada uma contendo um dúplex de DNA) são quebradas e, depois, ligadas uma à outra. Quando os cromossomos começam a se separar eles podem ser mantidos unidos em sítios discretos, os **quiasmas**. O número e a distribuição dos quiasmas comparam-se às características do *crossing-over* genético. Análises tradicionais sustentam que um quiasma representa um evento de *crossing-over*. O quiasma permanece visível quando os cromossomos se condensam e todas as quatro cromátides tornam-se evidentes.

Em termos moleculares, cada uma das cromátides-irmãs de cada um dos cromossomos homólogos contém um dúplex de DNA e, assim, cada bivalente contém quatro moléculas de DNA de fita dupla. A recombinação necessita um mecanismo que permita a um dúplex de uma cromátide-irmã interagir com o outro dúplex da cromátide-irmã do outro cromossomo homólogo. O mecanismo utilizado provém da maneira pela qual os ácidos nucleicos reconhecem um ao outro, com base nas suas sequências de nucleotídeos, ou seja, a complementaridade entre as fitas simples.

Acredita-se que qualquer evento que resulte em quebras na cadeia de DNA estimule o início de um processo de *crossing-over*. Por isso, radiação ultravioleta, raios X e reagentes químicos que criam quebras na fita dupla ou regiões de fita simples em uma molécula de DNA aumentam a frequência de *crossing-over* na célula. Em bactérias, mutações nos genes que codificam DNA-ligase e DNA-polimerase I, que originam quebras e regiões fita simples porque impedem a reconstituição do DNA durante os processos de replicação ou reparação, possuem efeito semelhante. Do mesmo modo, enzimas celulares induzem ativamente a recombinação na meiose pelo fato de fazerem clivagens ocasionais ao longo de um par de cromossomos homólogos.

1.5.2 Enzimas de recombinação homóloga em eucariotos

Muitos dos genes que atuam nos processos de recombinação homóloga em eucariotos são denominados de genes *Rad*, uma vez que eles foram primeiramente isolados em buscas por mutantes com sensibilidade elevada à irradiação por raios X. Raios X causam quebras nas duas fitas do DNA e, dessa forma, mutantes *rad*, sensíveis a raios X, também tornam-se deficientes nos processos de recombinação mitóticos ou meióticos. DSBs que ocor-

rem na meiose e em outros processos naturais como a troca de tipos acasalantes na levedura *S. cerevisiae*, são fortes indutores de recombinação homóloga. A recombinação também acontece em resposta a outras lesões não programadas no DNA. Como visto anteriormente, determinados tipos de lesões impõem um forte impedimento para a maquinaria de replicação do DNA, e a recombinação de uma molécula de DNA danificada com a sua cromátide-irmã restabelece a forquilha de replicação. A recombinação meiótica ocorre com uma frequência de 100 a 1.000 vezes maior do que a recombinação mitótica e, em geral, envolve cromossomos homólogos, resultando em cromossomos com *crossing-overs*. Esses *crossing-overs* são fundamentais para a segregação apropriada dos cromossomos na primeira divisão meiótica. A recombinação homóloga que ocorre durante a mitose difere daquela da meiose, pelo fato de que poucos eventos estão associados com *crossing-overs*, e a formação de *crossing-overs* é suprimida na mitose por DNA-helicases especializadas.

Significantes progressos foram alcançados nos últimos anos relacionados às proteínas que catalisam a recombinação homóloga em eucariotos e nas definições de seus mecanismos de ação. Conceitualmente, a recombinação homóloga pode ser dividida em três estágios: pré-sináptico, sináptico e pós-sináptico. As proteínas e estruturas envolvidas em cada um desses estágios serão descritas a seguir (**Figura 8.11**).

Nas células mitóticas, quebras no DNA de fita dupla são produzidas por agentes exógenos, tais como radiação ou tratamento químico, ou por fontes endógenas, tais como topoisomerasas e quebras na fita-molde. Durante a replicação essas quebras são convertidas em quebras nas duas fitas. Na meiose, as quebras na fita dupla são induzidas pela clivagem dependente de Spo11, cuja ação endonucleásica é a responsável pelo início dos processos de troca de fitas e *crossing-over* em alguns organismos. Assim, no **estágio pré-sináptico**, o DNA danificado é processado para formar uma região de fita simples estendida, que é coberta pela proteína de ligação ao DNA de fita simples eucariótica (RPA, de *replication protein A*). Para DSBs que ocorrem no DNA da levedura *Saccharomyces cerevisiae*, essa etapa é bastante complexa e envolve quatro nucleases: o complexo MRX (formado pelas proteínas Mre11-Rad50-Xrs2 [em humanos esse complexo é formado pelas proteínas Mer11-Rad50-Nbs1 e é chamado de MRN]), ExoI, Dna2 e Sae2 (CtIP em humanos) e pela proteína Sgs1 (Blm em humanos) com atividade de helicase.

Acredita-se que a proteína Rad50 dos complexos MRX/MRN mantenha unidas as extremidades das duplas fitas quebradas, por meio de dímeros que se conectam a essas extremidades em uma estrutura com a forma de um gancho, que se torna ativa na presença de íons de zinco. Rad50 e Mre11 são relacionadas às proteínas bacterianas SbcC e SbcD, que apresentam atividades de exonuclease em DNA de fita dupla e de endonucleases em DNA de fita simples.

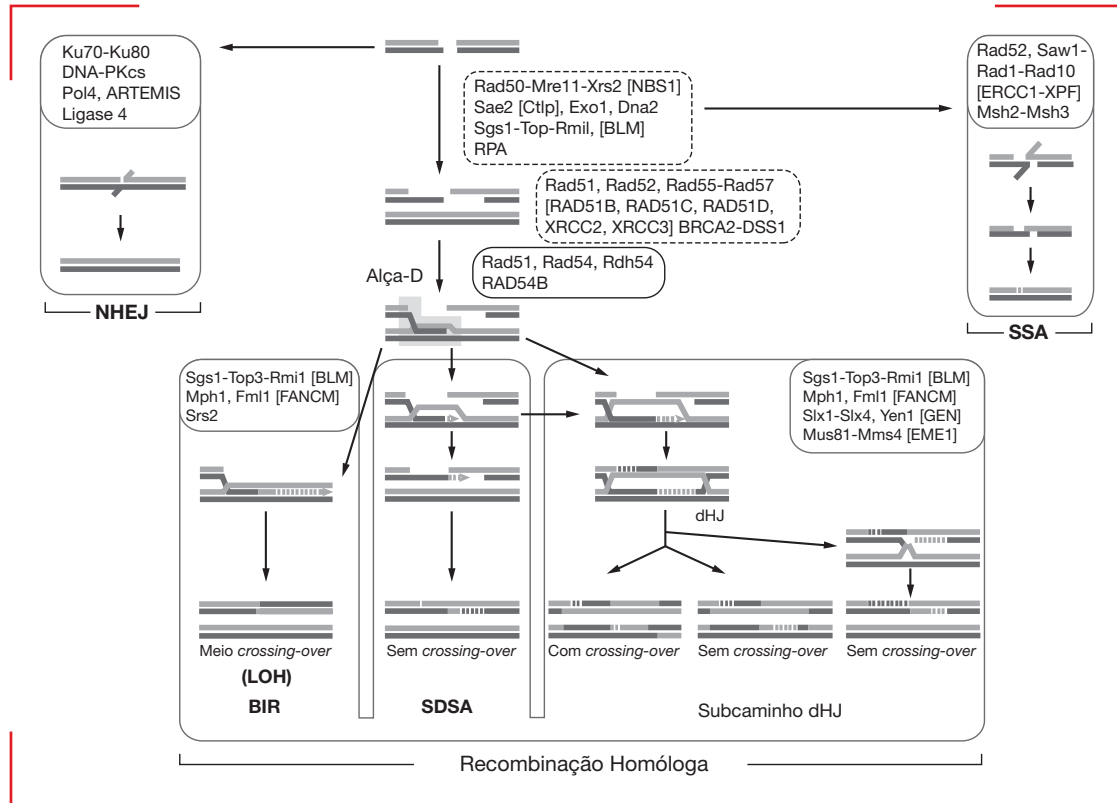


Figura 8.11

Caminhos para a reparação de quebras duplas. Os nomes das proteínas se referem às da levedura *S. cerevisiae*. Quando esses nomes forem diferentes em humanos, eles são mostrados entre colchetes. Linhas quebradas indicam síntese de DNA e regiões de heterodúplex que, após a reparação de mal pareamentos, podem levar à conversão gênica. Abreviaturas: BIR, replicação induzida por quebra; dHJ, junções de Holliday duplas; NHEJ, união de extremidades não homólogas; LOH, perda de heterozigotidade; SDSA, anelamento de fita dependente de síntese; SSA, anelamento de fita simples.

Após MRX/MRN e Sae2/CtIP terem preparado as extremidades da quebra dupla e removido qualquer proteína que estivesse nelas associadas ou adutos que poderiam inibir a degradação de uma das fitas, as extremidades são digeridas por nucleases que agem em conjunto com DNA helicases, que desenrolam o dúplice de DNA para expor extremidades de DNA de fita simples. As exonucleases ExoI e Dna2 e as helicases Sgs1 (em leveduras) e Blm (em humanos) foram identificadas como fatores decisivos para essa etapa.

Após as extremidades das fitas duplas terem sido processadas para gerar extremidades de fita simples 3'-OH, a ligação de RPA nesse DNA elimina estruturas secundárias. Com o auxílio de proteínas mediadoras, Rad51 substitui RPA, formando o nucleofilamento DNA de fita simples-Rad51. Rad51 é a homóloga eucariótica da RecA bacteriana, com 30% de identidade entre as sequências. Rad51 forma um nucleofilamento helicoidal voltado para a direita em um processo dependente de ATP, com seis moléculas de Rad51 e 18 nucleotídeos do DNA de fita simples por volta da hélice. Rad51 é necessária para todos os processos de recombinação homóloga, com exceção dos mecanismos de anelamento de fita

simples e união de extremidades não homólogas (discutidos no Capítulo 6). Nos eventos de recombinação que ocorrem durante a meiose, outra proteína com funções semelhantes à Rad51 foi identificada. Essa proteína é denominada Dmc1 e, de modo geral, atua de forma análoga à Rad51 para a reparação de quebras duplas nos cromossomos.

Três diferentes classes de proteínas mediadoras foram descritas, mas seus mecanismos de ação e a forma como eles se relacionam entre si ainda são pouco entendidos. O primeiro grupo é formado pelos parálogos de Rad51 e compreendem quatro proteínas em dois complexos separados em *S. cerevisiae* (Rad55-Rad57, Shu1-Psy3) e cinco em mamíferos (Rad51B, Rad51C, Rad51D, XRCC2 e XRCC3). Essas proteínas compartilham a porção central de RecA com Rad51 (com identidade de 20 a 30% entre suas sequências), mas não são capazes de formar filamentos extensos sobre o DNA e são incapazes de realizar a ampla faixa de reações catalisadas por Rad51. O segundo grupo é caracterizado pela proteína Rad52 de *S. cerevisiae*, que executa dois papéis independentes, uma função de mediadora e uma segunda função tardia no anelamento de RPA em um DNA de fita simples. A tercei-

ra classe de proteínas mediadoras, aparentemente ausente em leveduras, é caracterizada pela proteína BRCA2, a proteína supressora de câncer de mama e ovário humana. BRCA2 humana apresenta motivos de ligação ao DNA de fita simples (dobras-OB), um motivo de ligação ao DNA de fita dupla (domínio de torre) e vários sítios de ligação à Rad51, sugerindo que ela direcione a nucleação do filamento de Rad51 às junções de DNA de fita dupla nas extremidades processadas.

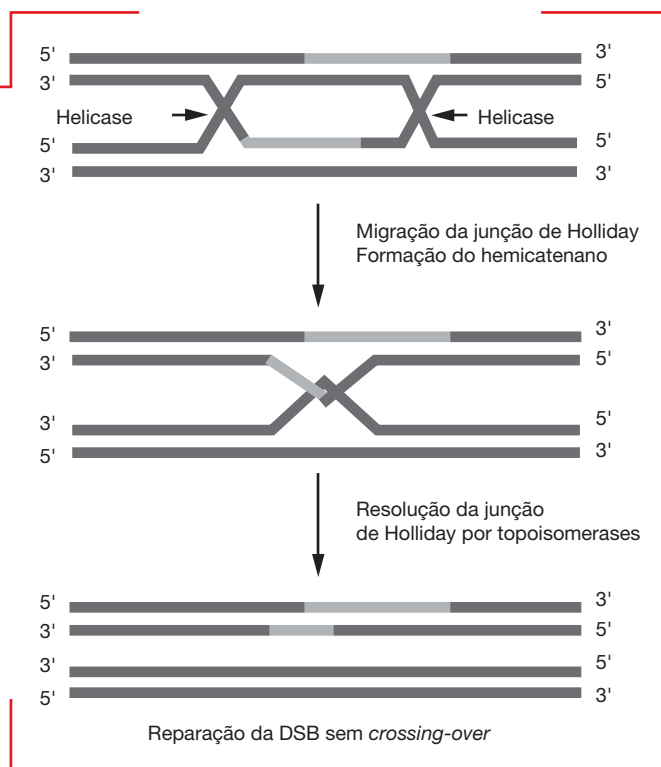
Durante o **estágio sináptico**, o filamento de Rad51 (ou de Dmc1) realiza a procura por regiões de homologia e a invasão da fita de DNA homóloga, gerando a alça-D, dentro da qual a extremidade 3'-OH da fita invasora serve como iniciador para a síntese de DNA (Figura 8.11). A proteína motora Rad54 é necessária para a estabilização do filamento de Rad51 e pela estimulação da formação da alça-D por Rad51. Ela também é a responsável em promover a transição da etapa de invasão de fita para a de síntese de DNA, uma vez que ela dissocia Rad51 do heterodúplex de DNA.

Finalmente, no **estágio pós-sináptico**, os três subcaminhos da recombinação homóloga são distinguidos (Figura 8.11), cada um com suas atividades enzimáticas específicas, muitas das quais ainda são parcialmente definidas. Como mostrado na Figura 8.11, a alça-D representa o ponto de ramificação para os múltiplos subcaminhos da recombinação homóloga (BIR, SDSA, dHJ). Na ausência de uma segunda extremidade, a alça-D pode se tornar uma forquilha de replicação completa, em um processo denominado replicação induzida por quebra (BIR, de

break-induced replication). Embora esse processo restaure a integridade do cromossomo, ele pode levar à perda de heteroziguidade (LOH, de *loss-of-heterozygosity*) de toda a informação genética que esteja após a quebra dupla. Na presença de uma segunda extremidade, o caminho predominante para a reparação da DSB em células somáticas parece ser o processo de SDSA, no qual a alça-D é revertida, levando ao anelamento da região recém-sintetizada com a fita simples da segunda extremidade. Esse caminho evita *crossing-overs* de forma inerente, o que reduz as probabilidades de rearranjos genômicos. Ao passo que a geração de *crossing-overs* pela formação de junções de Holliday duplas (dHJ, de *double Holliday junction*) é o objetivo da recombinação meiótica, a reparação por recombinação em células somáticas raramente está associada aos *crossing-overs*. A formação de junções de Holliday duplas envolve a captura da segunda extremidade, um processo que é ativamente bloqueado pela proteína Rad51 *in vitro*, sugerindo um mecanismo inerente em direção ao SDSA. O intermediário de dHJ pode ser resolvido por endonucleases de uma maneira descrita para a proteína bacteriana RuvC em produtos com ou sem *crossing-overs*, mas os mecanismos exatos e as proteínas envolvidas ainda são alvo de discussões. Alternativamente, dHJs podem ser dissolvidas por um complexo mecanismo que envolve a proteína motora de DNA da família RecQ (Sgs1 em *S. cerevisiae* e Blm em humanos), topoisomerase 3 e cofatores. As duas junções migram uma em direção à outra, criando uma estrutura de um hemicatenano (Figura 8.12), onde as fitas de DNA estão enroladas uma em volta da outra. Essa estrutura é, então,

Figura 8.12

A dissolução de junções de Holliday duplas pelas ações de uma DNA helicase e de uma topoisomerase. As duas junções de Holliday são empurradas uma em direção à outra pela migração de ramificação, usando a atividade de DNA-helicase. A estrutura resultante é um hemicatenano, em que as fitas simples de diferentes hélices de DNA estão enroladas uma ao redor da outra. Essa estrutura é clivada por uma DNA topoisomerase, que desenrola e libera as duas moléculas de DNA, formando produtos sem *crossing-over*.



resolvida pela ação de uma DNA topoisomerase associada: Topo3, no caso de Sgs1 e hTOPOIII α , no caso de Blm.

Enquanto o complexo helicase-topoisomerase pode resolver junções de Holliday como produtos sem *crossing-over* em células mitóticas, a resolvase que atua nas junções de Holliday meióticas e resulta em produtos com *crossing-over* não foi, ainda, completamente identificada. Atividades de endonuclease adicionais presentes nos complexos Mus81/Mms4 em leveduras e Mus81/Eme1 em células de mamíferos podem clivar estruturas semelhantes a junções de Holliday quebradas e estruturas de DNA ramificadas *in vitro*. A relação dessas atividades com a formação de *crossing-overs* meióticos, entretanto, ainda não está definida. Recentemente, homólogos de resolvases eucarióticos foram identificados em humanos e leveduras. As proteínas Gen1 em humanos e Yen1 em levedura são capazes de resolver estruturas de Holliday *in vitro*, embora, *in vivo*, seus papéis na recombinação homóloga ainda tenham que ser demonstrados.

Como visto até o momento, a reparação de DSBs é um processo que deve ser bem regulado e pode ser executado por diferentes caminhos, dependendo da fase do ciclo celular. Assim, a recombinação homóloga geralmente está restrita às fases S (de síntese) e G2, quando o DNA já foi replicado e as cromátides-irmãs estão disponíveis para atuarem como moldes na reparação das quebras duplas. Mecanismos como o de união de extremidades não homólogas (NHEJ, Capítulo 6), ao contrário, operam através de todo o ciclo celular, porém tendo mais importância na fase G1. A escolha entre estes caminhos é governada por proteínas quinases dependentes de ciclinas (CDKs, de *cyclin-dependent protein kinases*), e várias linhas de evidências sugerem que o processamento das quebras duplas, um evento necessário para a ocorrência de recombinação homóloga e inibitório para o NHEJ, seja uma etapa reguladora crucial nessa decisão.

2. Recombinação sítio-específica

Em razão de suas propriedades, o *crossing-over* preserva a ordem das sequências de DNA em cromossomos homólogos. Em casos excepcionais, no entanto, as células também utilizam um elaborado processo recombinacional de regulação, que tem como consequência o rearranjo de sequências por meio da recombinação direta entre sítios especiais. Segmentos de DNA podem ser movidos pela **recombinação sítio-específica**, resultando, frequentemente, na expressão de diferentes genes ou grupos de genes.

A recombinação sítio-específica (**Figura 8.13**) não envolve uma extensa homologia entre as sequências de DNA como ocorre no *crossing-over*. Ela necessita apenas que as sequências de ligação sejam localizadas por enzimas especializadas, que catalisam a quebra e a reunião das moléculas. Esse tipo de recombinação é iniciado por

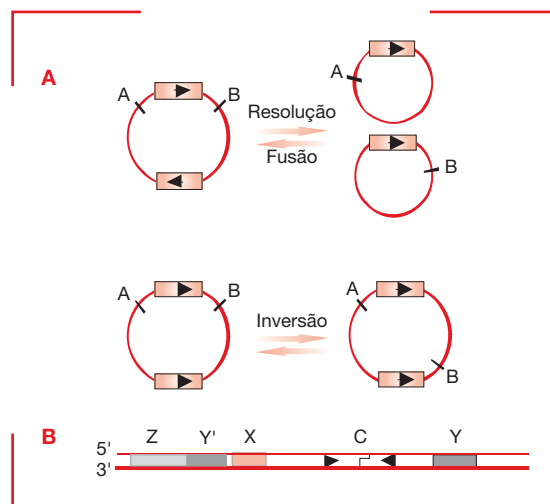


Figura 8.13

Recombinação sítio-específica: o DNA é quebrado em dois sítios e as extremidades são trocadas, sem que ocorram eventos de síntese ou degradação de DNA. (A) Resultado de um evento de recombinação sítio-específica entre moléculas circulares de DNA. As setas indicam as orientações dos sítios recombinantes. **(B)** Representação esquemática de um sítio de recombinação. A região de clivagem e religação está localizada no interior de uma sequência curta com simetria bilateral, C, na qual a recombinase liga-se. Próximos a esta sequência podem existir outros sítios de ligação para a recombinase ou proteínas acessórias (X, Y, Y', Z). Os sítios verdadeiros são todas variações desta situação.

processos reguladores que tornam as enzimas corretas disponíveis, ao contrário dos eventos de quebra aleatórios, nos quais proteínas, como a RecA, podem ligar-se a qualquer segmento de DNA e iniciar um *crossing-over*.

Na recombinação sítio-específica conservativa, uma molécula de DNA é clivada em ambas as fitas em dois locais específicos e as extremidades são religadas às suas novas parceiras por uma série de eventos catalíticos precisos, que não envolvem qualquer síntese ou degradação de DNA ou hidrólise de ligações fosfodiéster.

As funções da recombinação sítio-específica *in vivo* incluem a integração e a excisão de fagos, a resolução de intermediários na transposição, as modificações na expressão gênica, a regulação do número de cópias de plasmídeos e a monomerização de cromossomos e plasmídeos bacterianos multiméricos (ver Figura 8.13A). Os sistemas de recombinação sítio-específica melhor caracterizados são os de fagos, bactérias e leveduras. Nestes organismos, os substratos normais das reações são moléculas de DNA circulares. O evento recombinacional em um DNA circular pode apresentar um de três possíveis resultados: um único círculo pode ser resolvido em dois; ou dois círculos

podem ser fusionados em um; ou um segmento de DNA interno pode ser invertido em relação à sequência do restante da molécula (ver Figura 8.13B). Exemplos bem conhecidos destes resultados, discutidos no final deste capítulo, são a integração e a excisão do bacteriófago λ e a inversão do segmento Hin em *Salmonella*.

2.1 Mecanismos de recombinação sítio-específica

Na recombinação sítio-específica, proteínas que se ligam ao DNA exercem um papel fundamental na manutenção da união entre as fitas que irão sofrer um processo de recombinação. Os sítios necessários para a recombinação em geral são pequenos (com poucas dezenas de pares de bases) e contêm sequências nucleotídicas, reconhecidas especificamente por uma ou mais proteínas envolvidas na reação. Os dois sítios de recombinação não necessitam possuir a mesma sequência ou estrutura molecular, embora em muitos sistemas eles sejam idênticos.

O primeiro passo na recombinação sítio-específica é a ligação de proteínas a alguns ou a todos os elementos de reconhecimento de um ou ambos os sítios que irão recombinar. Depois disso, esses sítios fazem uma **sinapse** (ou troca de fitas), o contato sináptico pode ser entre proteínas já ligadas a cada sítio individual ou entre proteínas ligadas a um sítio e ao DNA clivado do outro sítio. Não é necessária homologia entre as partes recombinantes nesta etapa. A ligação e a sinapse são processos que envolvem múltiplos eventos. A ligação das subunidades proteicas pode ser cooperativa e criar uma elaborada estrutura de DNA-proteínas que está apta a buscar e a contatar o sí-

tio parceiro de uma maneira específica. Em alguns casos, a formação da sinapse depende da topologia do substrato de DNA e das orientações relativas dos sítios. Isso permite a obtenção de uma seletividade na recombinação entre os pares de sítios situados na mesma molécula de DNA, que apresentam repetições diretas ou invertidas.

A complexidade dos sistemas de recombinação sítio-específica varia muito, desde os que necessitam de uma única sequência curta (aproximadamente 30 pb) de reconhecimento e uma só proteína até aqueles que, como λ , envolvem sítios complexos (com um total de 270 pb) e quatro proteínas diferentes. Cada sistema tem uma proteína (**recombinase**) que reconhece as sequências do DNA. Ela cliva e religa as fitas, estando envolvida na catálise da troca de fitas. Com frequência, a região codificadora desta recombinase está intimamente relacionada com seus sítios de reconhecimento e, algumas vezes, com um elemento reforçador (*enhancer*) recombinacional. As recombinases examinadas *in vitro* podem ser classificadas em duas famílias com base nas suas sequências de aminoácidos. Uma família, denominada tirosino-recombinase, é exemplificada pela integrase do bacteriófago λ , e a outra, denominada serino-recombinase, pela resolvase do transposon Tn3 (**Tabela 8.1**). Enzimas de ambos os tipos são utilizadas para catalisar inversões de DNA e resolução de transposons. Quando adicionados à recombinase, muitos sistemas necessitam de proteínas acessórias, que possuem outras importantes funções celulares, por exemplo: IHF, FIS, HU e ArgR. Na recombinação sítio-específica, essas proteínas estão envolvidas na curvatura do DNA, na montagem do complexo DNA-proteína e no controle ou na seletividade das reações.

Tabela 8.1 Características das famílias de recombinases e resumo dos mecanismos de recombinação sítio-específica

	Família das tirosino-recombinases	Família das serino-recombinases
Tamanho da subunidade da recombinase	Variável, 20 a 65 kDa	Todas com ~ 20 kDa
Conservação de sequência	Baixa	Alta
Nº de exemplos conhecidos	~ 30	~ 30
Exemplos	Int de λ , Cre do bacteriófago P1, XerC de <i>E. coli</i>	Resolvases: Tn3, γ S, Tn21, Tn552; invertases: Gin, Hin e Cin.
Organização do sítio de recombinação	Variável: algumas com apenas o sítio de <i>crossing-over</i> , outras com muitos sítios de ligação acessórios	Conservado: resolvases possuem três subsítios (todos ligam resolvases); invertases possuem o sítio de <i>crossing-over</i> com um reforçador em <i>cis</i>
Tamanho do sítio de recombinação	~ 30-240 pb	Resolvases: 100-130 pb Invertases: 30 pb; necessitam de reforçador (60 pb)
Sítio de <i>crossing-over</i>	30 pb com simetria imperfeita	30 pb com simetria imperfeita
Sequência de sobreposição	6-8 pb	2 pb
Proteínas acessórias	Frequentemente necessárias	Resolvases não necessitam; invertases necessitam da FIS
Ligação covalente DNA-proteína	3' fosfodiéster com Tyr	5' fosfodiéster com Ser
Subunidades necessárias para a troca de fitas	4	4
Intermediários	Junções de Holliday	Substrato de quatro fitas clivado
Mecanismo de troca de fitas	Isomerização de Holliday	Rotação simples

2.1.1 Troca de fitas

A clivagem do DNA e a religação ocorrem no interior de um pequeno sítio de *crossing-over* (de aproximadamente 30 pb), com simetria bilateral imperfeita e que serve para unir as duas subunidades da recombinase. Em alguns casos, o sítio de *crossing-over* é flanqueado por outras sequências de ligação para proteínas, como a própria recombinase ou outras proteínas acessórias (ver Figura 8.13B). A família das serino-recombinases promove clivagens próximas ao centro do sítio de *crossing-over*. As fitas são cortadas, de forma assimétrica, a 2 pb de distância, deixando extremidades 5' recessivas (Figura 8.14A). A ligação DNA-proteína ocorre por meio de uma ligação fosfodiéster entre a extremidade 5' do DNA e um resíduo conservado de serina, próximo ao aminoterminal da recombinase. O grupamento hidroxílico da serina é o agente nucleofílico que ataca a cadeia de DNA. A estrutura química das junções após a troca de fitas (ver Figura 8.14A) é idêntica à estrutura do substrato inicial, de forma que as etapas de troca de fitas são energeticamente neutras. A observação de que cada extremidade 5' é ligada covalentemente a uma molécula de proteína fortalece a ideia de que a troca de fitas é catalisada por quatro subunidades de recombinase, com duas subunidades ligadas a cada sítio de *crossing-over*.

A troca de fitas na família das tirosino-recombinases é um pouco diferente. As fitas de cima e as de baixo são clivadas em locais fixos no interior do sítio de *crossing-over*, mas espaçadas em 6 a 8 pb, de forma que a clivagem em ambas as fitas pode gerar extremidades 3' recessivas (Figura 8.14B). A ligação DNA-proteína também é

fosfodiéster, mas entre a extremidade 3' do DNA no ponto de clivagem e um grupamento hidroxílico de um resíduo de tirosina conservado. Outra diferença é a ocorrência da formação de um intermediário na recombinação, no qual um par de fitas foi trocado e o outro permanece não recombinante, semelhante à junção de Holliday da recombinação homóloga. A reação catalisada pela tirosino-recombinase é, por isso, subdivisível em dois estágios, cada um deles envolvendo a clivagem e a religação das duas fitas do DNA (ver Figura 8.14B).

2.2 Integração e excisão do bacteriófago λ

A recombinação sítio-específica foi descoberta a partir de estudos genéticos com o bacteriófago λ . O DNA de λ integra-se, via recombinação, a um sítio específico do cromossomo de *E. coli*, onde replica de maneira passiva, como se fosse parte do DNA da célula hospedeira – um estágio denominado **profago**. As duas principais características da integração de λ são: o fato de que a troca é recíproca, preservando todo o DNA preexistente, e que ela ocorre em uma sequência específica de nucleotídeos, no interior de uma região curta de homologia entre o DNA do fago e o da bactéria.

O DNA do fago codifica uma enzima – **integrase de λ** – que direciona a inserção do DNA do fago no cromossomo de *E. coli* pela recombinação entre sítios específicos de ambos os DNAs, gerando uma molécula circular a partir de duas (Figura 8.15). A integrase é bastante sintetizada durante o período de infecção, de forma que, a integração deve ocorrer em praticamente todas as células infectadas. Uma reação de integração pode ser realizada

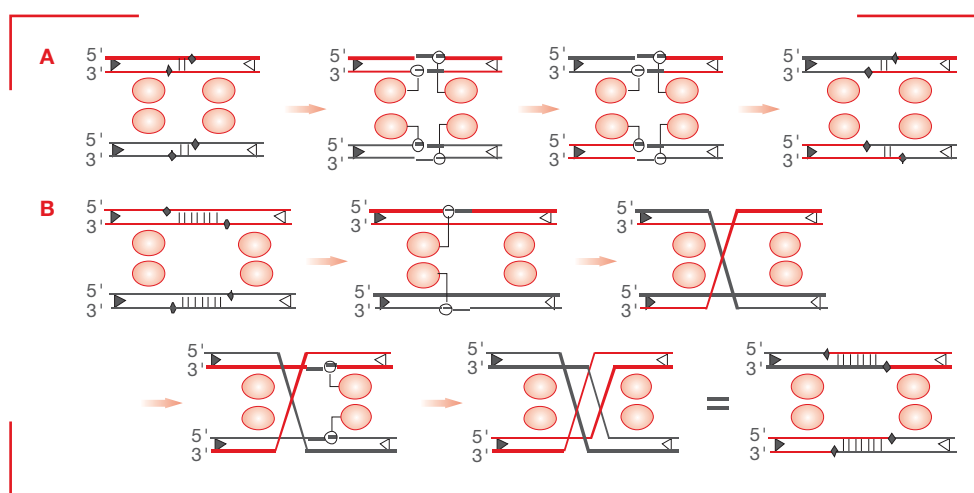


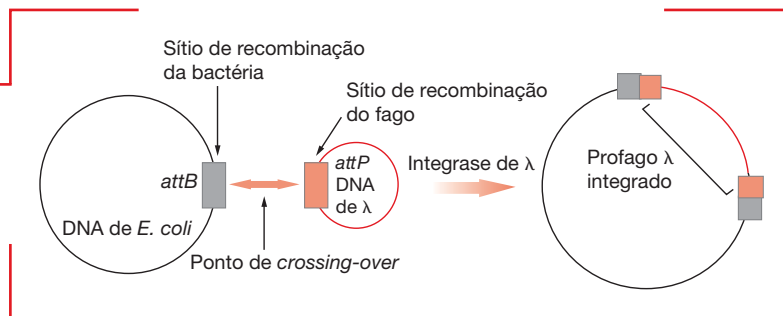
Figura 8.14

Eventos que ocorrem nos sítios de *crossing-over* durante a recombinação.

(A) Enzimas do tipo serino-recombinase: as subunidades proteicas estão representadas por elipses, as extremidades dos sítios de *crossing-over* por pontas de setas invertidas em \blacksquare e \square ; os dois pares de bases da região de sobreposição por linhas verticais; e os fosfatos, que são atacados pela recombinase, por losangos em \blacklozenge . As fitas de DNA são mostradas como linhas de cor \blacksquare ou \square , como forma de diferenciar os dois substratos do evento de *crossing-over*. (B) Enzimas do tipo tirosino-recombinase: a região de sobreposição (linhas verticais) varia de 6 a 8.

Figura 8.15

Integração do DNA do bacteriófago λ no cromossomo de *E. coli* por meio da recombinação sítio-específica.



com facilidade *in vitro* preparando-se uma mistura contendo: integrase purificada; uma proteína acessória de *E. coli*, chamada de IHF (o fator de integração ao hospedeiro); íons magnésio e DNA com sítios específicos do fago e da bactéria (chamados de *attP* e *attB*, respectivamente).

Todas as etapas da reação de integração estão associadas: as quatro fitas são clivadas, trocadas e religadas, sem que qualquer estrutura intermediária estável seja observada. O ponto-chave para a integração do profago é um evento de recombinação recíproca que ocorre entre o sítio de ligação especial do DNA de λ, na forma circular, (sítio *attP*) e o sítio correspondente, mas não idêntico, no cromossomo circular da bactéria (sítio *attB*). O resultado é a inserção do cromossomo do fago no genoma bacteriano.

Cada sítio *att* consiste em três componentes: uma sequência de 15 pb, comum a todos os sítios *att* e simbolizada por **o** (local onde o evento de recombinação ocorre); e um par de sequências flanqueadoras distintas, B e B' no cromossomo bacteriano, e P e P' no cromossomo do fago. Representando-se o sítio *attB* por BoB' e o *attP* por PoP', após o evento de recombinação, a integração irá gerar dois novos sítios de ligação híbridos, conhecidos por *attL* (ou BoP') e *attR* (ou PoB'). A integração é independente da proteína RecA, podendo ocorrer em hospedeiros RecA⁻, mas absolutamente dependente da ação da integrase e de IHF, uma proteína pequena e básica de *E. coli*, que se liga a sítios preferenciais e provoca curvatura no DNA. A reação total pode ser representada da seguinte forma:



Quando o fago passa para o estágio induzido, a integração é revertida (processo chamado de excisão) e os DNAs deles e da bactéria são reconstituídos. O fago provoca a excisão pela expressão de uma segunda proteína, uma **excisionase**, a qual permite que a integrase catalise a recombinação entre os sítios de ligação híbridos do fago, por meio de uma ação conjunta integrase-excisionase. A reação de excisão pode ser representada como:



2.3 Regulação da expressão gênica por meio da recombinação sítio-específica

A recombinação entre os sítios de uma única molécula de DNA pode ter duas consequências possíveis, dependendo de como os sítios envolvidos estão orientados. A recombinação pode remover segmentos intermediários ou invertê-los. As células, em alguns momentos, utilizam a inversão recombinacional para escolher entre dois arranjos alternativos de DNA que permitem a diferentes proteínas ou grupos de proteínas serem expressos.

Um exemplo bem estudado envolve a expressão alternada de duas proteínas flagelares da bactéria patogênica *Salmonella*, chamadas de H1 e H2. Em um determinado período, uma célula expressa uma dessas proteínas, mas nunca as duas simultaneamente. O promotor do gene *H2* é encontrado sobre um segmento de DNA, com 970 pb, adjacente ao gene *H2* e pode ser invertido por recombinação. Esse mecanismo é apresentado em detalhes no Capítulo 13.



Resumo

O DNA das células é capaz de sofrer rearranjos, que podem ocasionar desde novas combinações entre os genes presentes em qualquer genoma individual até alterações qualitativas e quantitativas na expressão desses genes. Esses rearranjos são realizados pela recombinação genética, na qual duas amplas classes são reconhecidas: a recombinação homóloga ou geral e a recombinação sítio-específica. A recombinação homóloga pode ocorrer em qualquer lugar ao longo de duas moléculas complementares de DNA e o seu principal resultado é que essas moléculas de DNA se sobrepõem e trocam partes (*crossing-over*). O sítio de troca pode ocorrer em qualquer lugar da sequência de nucleotídeos homóloga das duas moléculas de DNA envolvidas. Nesse local, uma fita de uma das moléculas de DNA faz pareamento de bases com uma das fitas da outra molécula, criando a junção heterodúplex entre as duas diferentes hélices duplas, que pode ter centenas de pares de bases de comprimento. Não há alteração nas sequências de nucleotídeos no sítio de troca; a quebra e os eventos de religação ocorrem de forma tão precisa que não há perda, ganho ou alteração de nenhum nucleotídeo. O modelo de Holliday demonstra como ocorre a recombinação entre dois cromossomos homólogos e como ela é resolvida. A recombinação genética entre sequências de DNA de cromossomos homólogos permite o resgate de cromossomos quebrados ou incompletamente replicados em todos os organismos. Existem três tipos de reações de recombinação que podem mediar a reparação de cromossomos danificados: reparação de forquilhas de replicação colapsadas, reparação de quebras duplas e reparação para o preenchimento de lacunas por recombinação. Em *E. coli* e organismos superiores, a inativação de proteínas que atuam na recombinação provoca redução na viabilidade celular e, com uma elevada taxa de lesões no DNA, as células são levadas à morte. Entre as diversas enzimas participantes da re-

combinação homóloga, RecA de *E. coli* e seus homólogos correspondentes em outros organismos, Rad51 e Dmc1, têm a função de encontrar a molécula de DNA correspondente para iniciar um evento de recombinação. Outras proteínas, como RecBCD, RecFOR e RuvABC, também participam da resolução dos intermediários da recombinação. Em eucariotos, a recombinação homóloga também é fundamental para o correto pareamento dos cromossomos durante a meiose e, assim como em bactérias, envolve uma série de enzimas especializadas.

Na recombinação sítio-específica, não é necessária uma homologia extensa do DNA. Neste caso, as trocas ocorridas são curtas. Sequências específicas de nucleotídeos (sobre uma ou ambas as moléculas de DNA participantes) são reconhecidas por uma enzima de recombinação sítio-específica, chamadas de recombinases. As funções da recombinação sítio-específica *in vivo* incluem: a integração e excisão de fagos, a resolução de intermediários na transposição, as modificações na expressão gênica, a regulação do número de cópias de plasmídeos, o *splicing* de genes em nível de DNA e a monomerização de cromossomos e plasmídeos bacterianos multiméricos. As recombinases examinadas *in vitro* podem ser classificadas em duas famílias, com base nas suas sequências de aminoácidos. Uma família, denominada tirosino-recombinase, é exemplificada pela integração do bacteriófago λ , e a outra, denominada serino-recombinase, pela resolvase do transposon Tn3. Em adição à recombinase, muitos sistemas necessitam de proteínas acessórias, que possuem outras importantes funções celulares, por exemplo: IHF, FIS, HU e ArgR. Na recombinação sítio-específica, estas proteínas estão envolvidas na curvatura do DNA, na montagem do complexo DNA-proteína e no controle ou na seletividade das reações.



Leituras recomendadas

Glover DM. Gene cloning: the mechanics of DNA manipulations. New York: Chapman and Hall; 1984.

Kuzminov A, Stahl FW. Double-strand end repair via the RecBC pathway in *Escherichia coli* primes DNA replication. *Gene Dev.* 1999;13(3):345-56.

Lewin B. Genes IX. 9. ed. Porto Alegre: Artmed; 2009.

Lodish H, Baltimore D, Berk A, Zipursky L, Matsudaira PT. *Biologia celular e molecular*. 5. ed. Porto Alegre: Artmed; 2005.

Mézard C, Vignard J, Drouaud J, Mercier R. The road to crossovers: plants have their say. *Trends Genet.* 2007;23(2):91-9.

Motamedi MR, Szigety SK, Rosenberg SM. Double-strand-break repair recombination in *Escherichia coli*: physical evidence for a DNA replication mechanism *in vivo*. *Gene Dev.* 1999;13(21):2889-903.

Ptashne MA. A genetic switch: phage lambda and higher organisms. 2nd ed. Cambridge: Cell Press; 1992.

Sambrook J, Russel DW. Molecular cloning: a laboratory manual. 3. ed. New York: Cold Spring Harbor Laboratory; 2001.

Schuermann D, Molinier J, Fritsch O, Hohn B. The dual nature of homologous recombination in plants. Trends Genet. 2005;21(3):172-81.

Stark WM, Boocock MR, Sherratt DJ. Catalysis by site-specific recombinases. Trends Genet. 1992;8(12):432-8.

Van Gool A, Hajibagheri NMA, Stasiak A, West SC. Assembly of the Escherichia coli RuvABC resolvosome

directs the orientation of Holliday junction resolution. Gene Dev. 1999;(14)13:1861-870.

Van de Putte P, Goosen N. DNA inversions in phages and bacteria. Trends Genet. 1992;8(12):457-62.

Voet D, Voet JG. Bioquímica. 3. ed. Porto Alegre: Artmed; 2006.

Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R, et al. Biología molecular do gene. 5. ed. Porto Alegre: Artmed; 2006.

Elgion Loreto | Henrique Bunselmeyer Ferreira

Elementos Genéticos Móveis

1. Plasmídeos 186

- 1.1 Estrutura e classificação dos plasmídeos 186
- 1.2 Mobilidade plasmidial em bactérias 188

2. Bacteriófagos 191

- 2.1 Estrutura e classificação dos bacteriófagos 191
- 2.2 Mobilização de material genético bacteriano mediada por bacteriófagos 191

3. Elementos de transposição 193

- 3.1 Características gerais e mecanismos de transposição 193
- 3.2 Elementos de DNA ou transposons 196
- 3.3 Retrotransposons 197
- 3.4 Consequências biológicas da transposição 201

O início do século XXI foi marcado pelo sequenciamento completo de mais de mil genomas (ver Capítulos 4 e 5), e parte substancial dessas sequências determinadas é formada por segmentos de DNA que podem movimentar-se dentro dos genomas ou mesmo entre diferentes genomas. Em razão disso, esses segmentos são denominados **elementos genéticos móveis**. O conjunto dos elementos genéticos móveis de um genoma é chamado de **mobiloma**.

Três são os exemplos de elementos genéticos móveis: os **plasmídeos**, os **bacteriófagos** e os **elementos transponíveis**. Muitas vezes esses elementos estão integrados aos cromossomos, embora possam, ao longo do tempo, mudar de localização dentro deles. Porém, outras vezes, eles podem estar fisicamente dissociados dos cromossomos e replicarem-se de forma independente. Quando estes elementos podem ser mantidos ou replicados de maneira autônoma dos cromossomos, são chamados de **episômos**, mesmo que às vezes possam ser encontrados na forma integrada.

Alguns elementos móveis, como certos elementos transponíveis, não codificam qualquer proteína e não se manifestam fenotipicamente nos organismos que os possuem. Por isso, são, às vezes, incluídos na categoria de DNA “sem função”. Outros elementos genéticos móveis codificam enzimas necessárias a sua mobilização e, muitas vezes, contêm genes que podem conferir características especiais aos organismos que os possuem. Por exemplo, genes que conferem resistência a antibióticos, codificam toxinas ou são responsáveis por algumas funções celulares, como a construção de telômeros, podem estar presentes em certos elementos genéticos móveis.

A seguir, será apresentado de forma detalhada os plasmídeos, os fagos e os elementos transponíveis. Além dos aspectos estruturais de cada tipo de elemento genético móvel, serão enfatizados os mecanismos de mobilidade e as suas consequências para as células ou organismos hospedeiros.

1. Plasmídeos

1.1 Estrutura e classificação dos plasmídeos

Os plasmídeos são elementos genéticos extracromossômicos com capacidade de replicação autônoma (**Figura 9.1**). Ainda que a replicação dos plasmídeos seja independente da replicação dos cromossomos, ela é dependente das mesmas enzimas que replicam o genoma da célula hospedeira. Plasmídeos são encontrados principalmente em procariotos (bactérias e arqueas), mas também em alguns eucariotos, como leveduras. A grande maioria dos plasmídeos é formada por moléculas de DNA de fita dupla circulares. Entretanto, alguns plasmídeos ocorrem na forma de DNA de fita dupla lineares, sendo encontrados em procariotos, que também possuem cromossomos lineares (ver Capítulo 4).

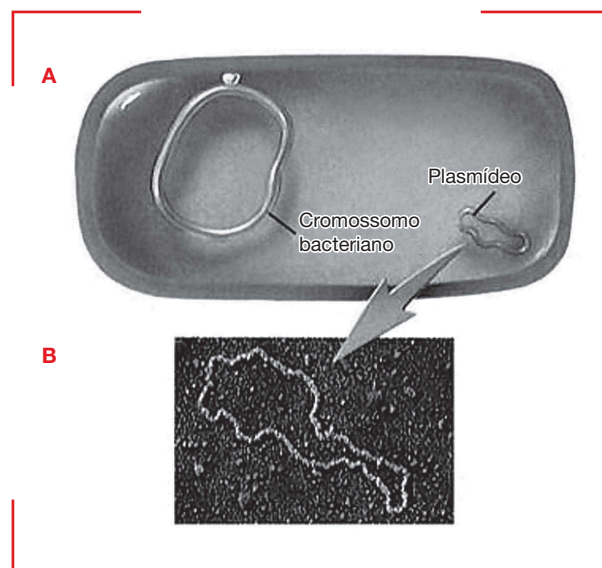


Figura 9.1

Material genético de uma bactéria. (A) Geralmente o cromossomo bacteriano é circular. Além dele, as bactérias podem possuir material genético extracromossômico, os plasmídeos, também circulares. **(B)** Micrografia eletrônica de um plasmídeo.

Quanto ao tamanho, os plasmídeos podem ser bem variados. Os menores, de 1 kb, têm apenas um ou poucos genes. Os maiores podem atingir 1.000 kb e contêm muitos genes. O elenco de genes presentes em um plasmídeo pode conferir funções e determinar fenótipos importantes para as células hospedeiras (**Tabela 9.1**). Por exemplo, além de genes que codificam proteínas envolvidas na própria replicação, os plasmídeos podem conter genes relacionados à virulência, codificadores de toxinas ou envolvidos na determinação de resistência a substâncias nocivas exógenas, como antibióticos e outros xenobióticos.

Os plasmídeos estão entre as principais ferramentas da biologia molecular, e muitos plasmídeos artificiais vêm sendo construídos com os mais variados propósitos. Por exemplo, os plasmídeos artificiais podem ser utilizados como vetores de clonagem molecular, para a transferência de genes de uma célula ou organismo para outro, ou para a expressão heteróloga de um gene (isto é, em um organismo diferente daquele de origem do gene) (ver Capítulo 16).

Para a autorreplicação de forma autônoma, os plasmídeos precisam de uma origem de replicação, *ori*, constituída por sequências específicas, reconhecidas por proteínas envolvidas na replicação do DNA. As origens de replicação plasmidiais são, portanto, funcionalmente análogas à *oriC* (cromossômica) bacteriana, por exemplo (ver Capítulo 6).

O **número de cópias** de um plasmídeo encontrado por célula é determinado pela sua origem de replicação

Tabela 9.1 Classificação funcional dos plasmídeos

Classe	Gene(s) funcionalmente relevantes	Função	Exemplos
Plasmídeos de fertilidade ou plasmídeos conjugativos	Genes <i>tra</i>	Promover a transferência do plasmídeo de uma célula doadora para uma receptora	Plasmídeo F de <i>Escherichia coli</i>
Plasmídeos de resistência	Genes para enzimas envolvidas no metabolismo de antibióticos	Conferem resistência à antibióticos ou agentes antimicrobianos	Plasmídeo R100, que confere resistência a sulfonamidas, estreptomicina/espectromicina, ácido fusídico, cloranfenicol, tetraciclina e mercúrio em espécies dos gêneros <i>Escherichia</i> e <i>Klebsiella</i>
Plasmídeos produtores de bacteriocinas	Genes que codificam bacteriocinas como as colicinas	Inibem a multiplicação ou matam bactérias de espécies relacionadas ou de outras linhagens da mesma espécie	Plasmídeos Col como o ColE1 de <i>E. coli</i>
Plasmídeos degradativos	Genes que metabolizam xenobióticos	Permitem degradar substâncias estranhas à célula	Plasmídeo pCAR1 de <i>Pseudomonas putida</i> , que confere à bactéria a capacidade de degradar carbazol
Plasmídeos de virulência	Genes de virulência (codificam proteínas responsáveis pela adesão da bactéria à mucosa intestinal e invasão do epitélio)	Transforma as bactérias em patogênicas	Plasmídeos codificadores de enterotoxinas de <i>E. coli</i>

(*ori*), pois é ela que regula a replicação. Essa regulação é necessária para que o plasmídeo não sature a célula com um número excessivo de cópias, no caso de uma replicação muito acelerada, e nem seja perdido ao longo de uma linhagem celular em decorrência de um processo de replicação muito lento, em comparação à velocidade do ciclo celular da espécie hospedeira. Alguns plasmídeos, como o pIJ101 de *Streptomyces coelicolor*, replicam um grande número de vezes por ciclo celular, de modo a gerar centenas de cópias por célula, sendo, por isso, denominados **plasmídeos multicópia**. Porém, outros plasmídeos, como o plasmídeo F de *E. coli*, replicam apenas uma ou poucas vezes durante o ciclo celular da bactéria, o que resulta em células-filha recebendo apenas uma ou poucas cópias a cada divisão. Quando um plasmídeo tem a sua replicação regulada de forma a manter em média apenas uma cópia por célula, ele é chamado de **plasmídeo de cópia única**. Muitas vezes, são também denominados plasmídeos de cópia única aqueles presentes em um número menor que 10 cópias por célula, embora seja mais correto tratá-los como **plasmídeos de baixo número de cópias** (incluindo os plasmídeos de cópia única).

Os mecanismos reguladores da replicação, utilizados por plasmídeos multicópia, são bastante diferentes daqueles utilizados por plasmídeos de baixo número de cópias. Os plasmídeos multicópia necessitam de apenas um mecanismo que iniba o início da replicação plasmidial quando o número de cópias por célula atingir um determinado nível, sendo, por isso, chamados de **plasmídeos relaxados**. Já os plasmídeos de baixo número de cópias, denominados **plasmídeos estringentes**,

necessitam de um mecanismo regulador mais restritivo, que permita apenas um ou poucos eventos de replicação a cada ciclo celular.

Além da regulação do número de cópias, muitos plasmídeos têm sistemas que regulam a sua segregação durante a divisão celular, denominados **sistemas de partição** (Figura 9.2). O sistema de partição baseia-se em sequências nucleotídicas específicas, presentes no plasmídeo e chamadas de **sítios *par***, que interagem com a membrana celular da bactéria. À medida que a célula cresce, as cópias do plasmídeo são afastadas umas das outras, distribuindo-se em números aproximadamente iguais nas duas células-filha geradas durante a divisão celular (Figura 9.2A). Na ausência de um sistema de partição, a segregação das cópias de um determinado plasmídeo a cada divisão celular ocorre ao acaso, o que pode levar à eventual eliminação em uma linhagem celular (Figura 9.2B).

Na natureza encontram-se bactérias portadoras de inúmeros plasmídeos distintos, como a bactéria *Borrelia burgdorferi* B31, uma das espiroquetas causadoras da doença de Lyme em seres humanos, que possui pelo menos 17 plasmídeos (7 ou 8 circulares e 10 lineares). Entretanto, alguns plasmídeos são **incompatíveis**, ou seja, incapazes de manterem-se juntos em uma mesma linhagem celular. Geralmente, esses plasmídeos possuem a mesma *ori* e, por isso, têm similaridades nos mecanismos de regulação do número de cópias (Figura 9.3). Os plasmídeos podem ser classificados pela sua capacidade de manterem-se ou não juntos em uma linhagem celular,

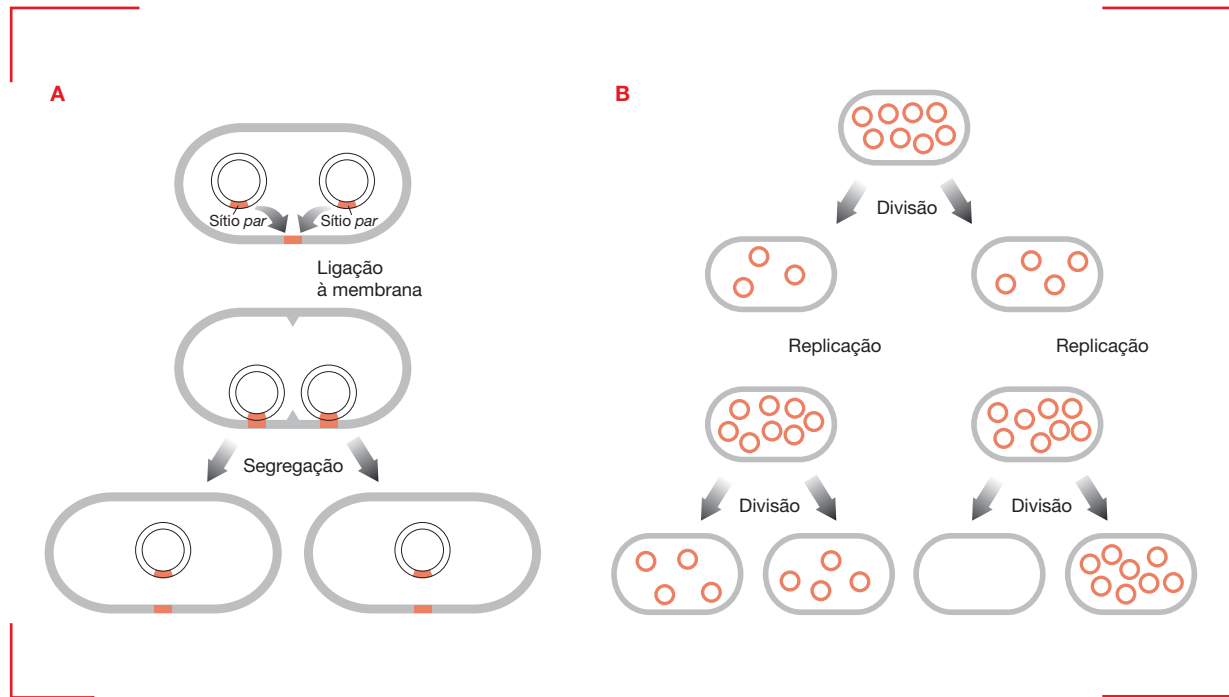


Figura 9.2

Segregação plasmidial. Na segregação regulada por sistemas de partição (A), os sítios *par* interagem com a membrana celular da bactéria e, à medida que a célula cresce, as cópias do plasmídeo são afastadas umas das outras e segregadas em números aproximadamente iguais nas duas células-filha geradas durante a divisão celular. Na ausência de um sistema de partição (B), a segregação das cópias de um plasmídeo (círculos em ■), a cada divisão celular, ocorre ao acaso, o que pode levar à eventual eliminação desse plasmídeo em uma linhagem celular.

formando diversos **grupos de incompatibilidade** ou **grupos Inc**.

1.2 Mobilidade plasmidial em bactérias

Dois importantes fenômenos envolvendo a mobilização de plasmídeos em bactérias são a conjugação e a transformação bacteriana. A **conjugação** bacteriana é a transferência de material genético através de um contato direto célula-célula. Este processo depende da formação de longos apêndices na superfície da célula bacteriana, chamados de **fímbrias** ou de **pilus sexuais**. Esses apêndices são formados, principalmente, por subunidades de uma glicoproteína, a **pilina**. Embora as fímbrias sexuais sejam essenciais para a formação de pares associados de células, elas não são necessárias para a subsequente transferência do DNA. A transferência depende da formação de poros e de uma ponte citoplasmática (**tubo** ou **ponte de conjugação**) entre as células associadas.

Na **conjugação** ocorre a transferência de um plasmídeo, denominado **conjugativo**, de uma bactéria para outra. Existe um número bastante grande de plasmídeos conjugativos, que se caracterizam pelos respectivos repertórios gênicos e determinam o **sistema de transferência** ao qual pertence uma bactéria. Em geral, esses plasmídeos são grandes, tendo em torno de 100 kb, e contêm os genes de transferência **genes tra**, em torno de

30, responsáveis pela síntese da fímbria sexual, pelo reconhecimento entre células conjugativas, pela transferência do DNA e pela regulação do processo de conjugação. Além disso, os plasmídeos conjugativos podem conter seqüências de inserção e transposons.

O **plasmídeo F** (de fertilidade) foi o primeiro elemento genético extracromossômico identificado em *E. coli*, e consiste em uma molécula de DNA circular de 94,5 kb. Ele é constituído por três grupos de genes: um grupo de vários genes envolvidos na sua replicação vegetativa; um grupo de pelo menos 28 genes coordenadamente regulados, que estão envolvidos no processo de conjugação, constituindo a chamada região *tra*; e um grupo de quatro seqüências de inserção transponíveis, responsáveis pela integração de F no cromossomo da bactéria hospedeira. Os genes *tra* do plasmídeo F estão organizados em 3 óperons – *traM*, *traJ* e *traYZ* – que, juntos, ocupam 33 kb do plasmídeo. Quatorze genes (*traA-G*) são necessários para o estabelecimento e o equilíbrio do contato célula-célula e cinco genes (*traMYDIZ*) são necessários para a transferência do DNA. A regulação da transcrição destes óperons é função dos produtos dos genes *finO*, *finP* e *traJ*.

As bactérias que possuem o plasmídeo F, são chamadas de F⁺ (doadoras), e as que não o possuem são F⁻ (receptoras). Quando células F⁺ entram em contato com células F⁻ ocorre uma interação entre elas, mediada por uma fímbria sexual. A extremidade da fímbria sexual es-

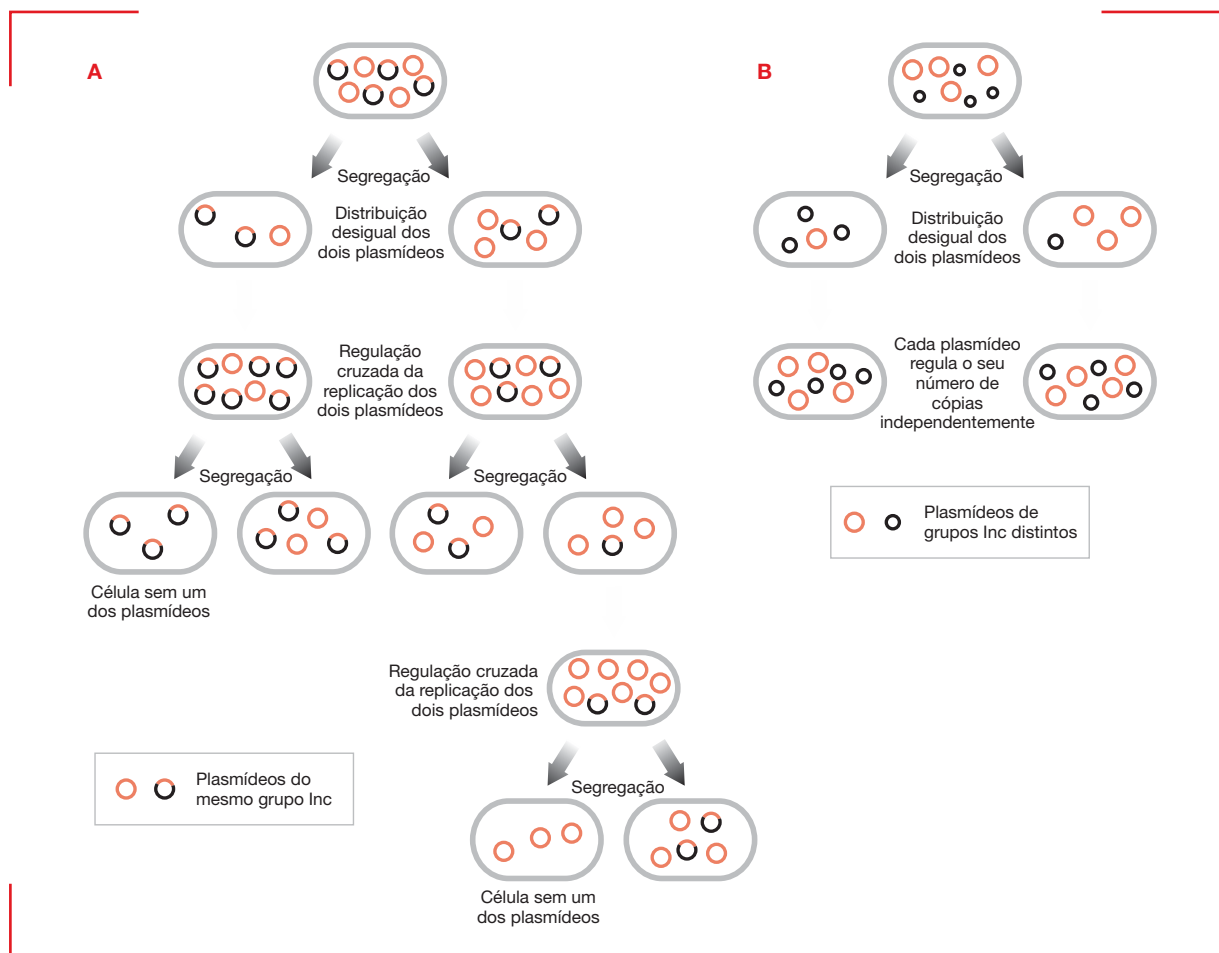


Figura 9.3

Incompatibilidade de plasmídeos. Cada plasmídeo possui um sistema de controle da sua replicação que determina o seu número de cópias na célula hospedeira e define o seu grupo **Inc**. Quando dois plasmídeos pertencentes ao mesmo grupo **Inc** residem em uma mesma célula hospedeira (**A**), o sistema de replicação comum limita o número de cópias de ambos, de modo que a soma das cópias de cada tipo de plasmídeo não pode ultrapassar o número total de cópias permitido para aquele grupo **Inc**. Na figura, o número total de cópias permitido é de 8; assim, a replicação e/ou a segregação desigual em ciclos de divisão celular sucessivos leva à perda de um dos plasmídeos e à fixação do outro na linhagem. Quando dois plasmídeos pertencentes a grupos **Inc** distintos residem em uma mesma célula hospedeira (**B**), os sistemas de controle da replicação não relacionados controlam de forma independente o número de cópias de cada um deles. Em uma situação como esta, os dois plasmídeos podem coexistir em uma mesma linhagem celular.

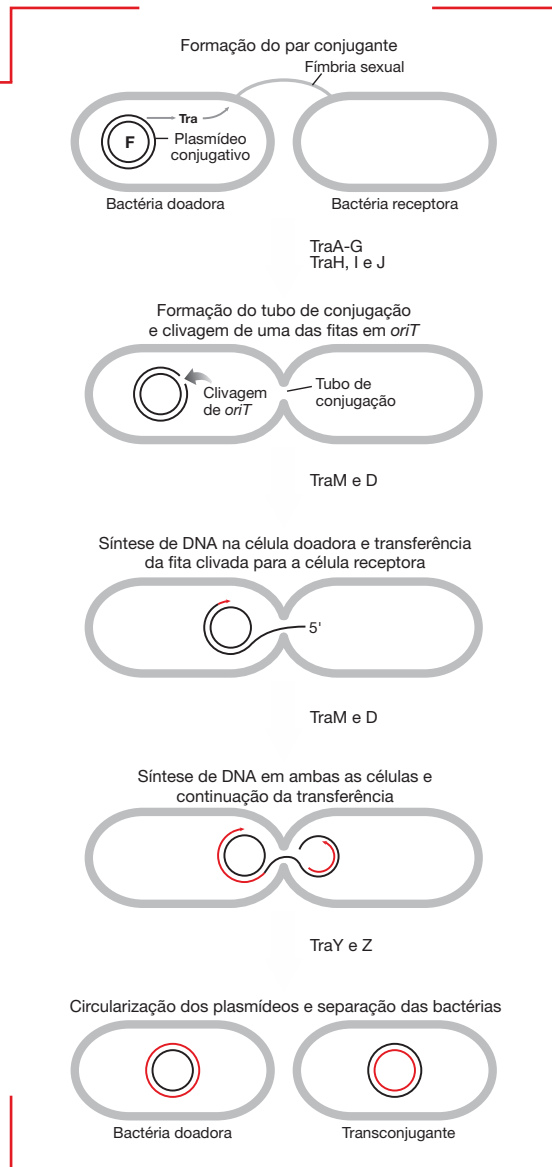
tabelece uma conexão estável entre os dois tipos de célula, aparentemente pela interação com um sítio-específico na superfície da célula F^- . Após a união, ocorre a contração da fímbria sexual, aproximando ainda mais as duas células.

A transferência do DNA (**Figura 9.4A**) inicia em um sítio-específico do plasmídeo, denominado **origem de transferência** (*oriT*), e prossegue por um mecanismo de replicação do tipo círculo-rolante, similar ao que ocorre com o bacteriófago λ (ver Capítulo 6). Este processo começa com a formação de um complexo nucleoprotéico (**relaxossomo**) pela ligação dos produtos dos genes *traH*, *traI* e *traJ* à *oriT*. O produto de *traI* é uma helicasa, que cliva uma das fitas de DNA de F na sequência da *oriT* e a desenrola a partir da sua extremidade 5'. Essa

fita passa através do tubo de conjugação para a célula F^- , ao mesmo tempo em que uma fita complementar àquela que permanece na célula doadora é sintetizada para substituí-la. Uma fita complementar àquela que foi transferida para a célula receptora é também sintetizada, de maneira descontínua, utilizando a própria maquinaria de replicação da célula hospedeira. Por fim, um mecanismo envolvendo os produtos dos genes *traY* e *traZ* circulariza a molécula de DNA de F, até então linear, na célula receptora. O produto do evento de conjugação completo é a transferência do plasmídeo F para uma célula F^- , com a retenção de uma cópia idêntica na célula doadora. Uma vez recebido o plasmídeo F, a célula receptora passa agora a ser F^+ , transformando-se também em doadora.

Figura 9.4

Representação esquemática dos processos de conjugação e mobilização do plasmídeo F integrado ao cromossomo bacteriano. Transferência de um plasmídeo F por conjugação. A célula doadora produz uma fímbria sexual que estabelece um contato estável com a célula receptora. A transferência do DNA inicia com a clivagem de uma das fitas do DNA plasmidial em *oriT*. A fita clivada é desenrolada a partir da sua extremidade 5' e passa para a célula receptora, ao mesmo tempo em que uma fita complementar àquela que fica na célula doadora é sintetizada para substituí-la. A fita transferida, por sua vez, serve de molde para a síntese de uma fita complementar a ela na célula receptora. Ao final do processo, uma cópia do plasmídeo permanece na célula doadora, enquanto uma nova cópia é gerada na célula receptora, que passa a ser chamada de transconjugante. Na figura, as fitas originais do plasmídeo conjugativo estão representadas em ■, e as novas fitas sintetizadas durante o processo de transferência estão representadas em ■. Os produtos dos principais genes *tra* envolvidos em cada etapa do processo estão indicados.



Os plasmídeos conjugativos podem inserir-se no cromossomo bacteriano ou em outros plasmídeos. Quando esses plasmídeos integrados transferem-se para células receptoras, transferem junto parte do cromossomo bacteriano (**ver Material Complementar Online 9.1**). Na célula receptora, este DNA irá se recombinar com o DNA do cromossomo, tendo como consequência a recombinação genética entre as duas linhagens (doadora e receptora). Como visto anteriormente, a transferência ocorre com a passagem de um fita de DNA da célula doadora para a receptora, começando sempre pela *oriT*. Assim, os genes do cromossomo bacteriano, que estão mais próximos do epissomo F, serão transferidos primeiro. Em 1950, Wollman e Jacob demonstraram que era possível fazer um mapeamento cromossômico de bactérias, interrompendo a conjugação em tempos diferentes, agitando-as em um liquidificador. Pela frequência de transferên-

cia dos diferentes genes marcadores, pode-se determinar a ordem dos genes no cromossomo bacteriano.

A **transformação** genética é outro importante método de transferência de plasmídeos entre as bactérias. Na transformação, pequenos segmentos de DNA extracelular (exógeno) são captados por uma bactéria viva, podendo promover mudança genética na célula receptora. Porém, neste caso, não se faz necessário uma bactéria doadora. Moléculas de DNA purificadas são capazes de ser interiorizadas. A passagem ao interior da célula de moléculas de DNA é, em geral, um processo ativo, que requer energia e moléculas específicas para este transporte.

As células capazes de fazer este transporte são chamadas de **competentes**. Algumas bactérias são naturalmente competentes, ainda que outras precisem ser induzidas a captarem o DNA exógeno.

2. Bacteriófagos

2.1 Estrutura e classificação dos bacteriófagos

Bacteriófagos, também chamados simplesmente de **fagos**, são vírus de bactérias. Como os demais vírus, de arqueas e de eucariotos, os fagos não possuem uma maquinaria metabólica própria e, por isso, são incapazes de se multiplicarem sem parasitar células de um hospedeiro sensível. Os fagos estão entre os seres mais diversos e amplamente distribuídos da biosfera. Alguns possuem genoma de RNA, como os bacteriófagos R17, MS2, f2 e Q, ao passo que outros têm genoma de DNA de fita simples (ϕ X174, M13) ou de DNA linear de fita dupla (T4, T5, T7, λ , P1). O tamanho do genoma dos fagos varia muito. O bacteriófago ϕ X174, por exemplo, tem um genoma de 5,4 kb, com apenas 11 genes, sendo 6 deles total ou parcialmente sobrepostos; já o bacteriófago T4 tem um genoma de 168,9 kb, contendo mais de 270 genes.

Em geral, os fagos se propagam em uma estrutura composta por um capsídeo proteico envolvendo o genoma viral. Quanto ao capsídeo, os fagos são muito variados, sendo alguns bastante simples, como o bacteriófago M13, cujo capsídeo é formado por 5 diferentes proteínas e tem a forma de um bastão (**Figura 9.5A**). Outros, como o bacteriófago T4, possuem um capsídeo complexo, formado por um total de 24 diferentes proteínas, que compõem-se com uma cabeça icosaédrica, além de um colar, de uma cauda e de fibras (**Figura 9.5B**). Devido à grande diversidade, os fagos são classificados em 13 diferentes famílias, refletindo a variação na organização dos seus genomas, capsídeos e diferenças relacionadas aos hospedeiros.

A distinção entre fagos e plasmídeos nem sempre é muito exata. Plasmídeos são considerados replicons não letais (pelo menos para as suas células hospedeiras), ao passo que fagos são replicons potencialmente letais. Entretanto, bacteriófagos como P1 e λ dv podem ser mantidos na forma de plasmídeos.

O “ciclo de vida” dos fagos podem ser de dois tipos: os do tipo **virulento**, que se multiplicam dentro da célula hospedeira e originam vários fagos que são liberados após a lise celular; e os do tipo **temperado**, que podem integrar-se ao genoma bacteriano, formando uma linhagem lisogênica, ou entrar em ciclo lítico como os fagos virulentos.

Na **infecção** de uma bactéria por um fago ocorrem três etapas principais: a colisão, a adsorção e a injeção do ácido nucleico. A ligação do fago na superfície da célula não somente ativa a liberação do ácido nucleico, como também pode causar alterações na membrana, facilitando sua penetração. No caso dos bacteriófagos P1, T4, T2 e ϕ X174, um lipopolissacarídeo da membrana bacteriana parece ser a molécula receptora. Já o bacteriófago λ requer uma proteína do hospedeiro

codificada pelo gene *lamB*, ao passo que os bacteriófagos M13 e R17 necessitam de uma fímbria sexual para a adsorção.

Após a entrada do genoma viral em um hospedeiro suscetível, o fago pode entrar em um ciclo lítico ou lisogênico, dependendo da sua natureza (se virulenta ou temperada) e do estado fisiológico da célula hospedeira. No **ciclo lítico** (**Figura 9.6A**), o genoma do fago é expresso em uma sequência de eventos reguladores que requer fatores codificados tanto pelo fago como pela bactéria. Logo após a infecção, a própria maquinaria de transcrição da bactéria hospedeira determina a expressão dos denominados **genes iniciais** do fago. Os genes iniciais codificam proteínas reguladoras que determinam a transcrição dos **genes intermediários**, cujos produtos são, na maioria, proteínas envolvidas no processo de replicação do genoma do fago. Genes intermediários também produzem algumas proteínas reguladoras, que determinam a expressão dos **genes tardios**. Os produtos dos genes tardios são proteínas do capsídeo do fago e fatores envolvidos na formação das partículas virais maduras e na lise da célula hospedeira. Ao final do ciclo, as novas partículas virais formadas são liberadas e podem infectar novas células hospedeiras suscetíveis.

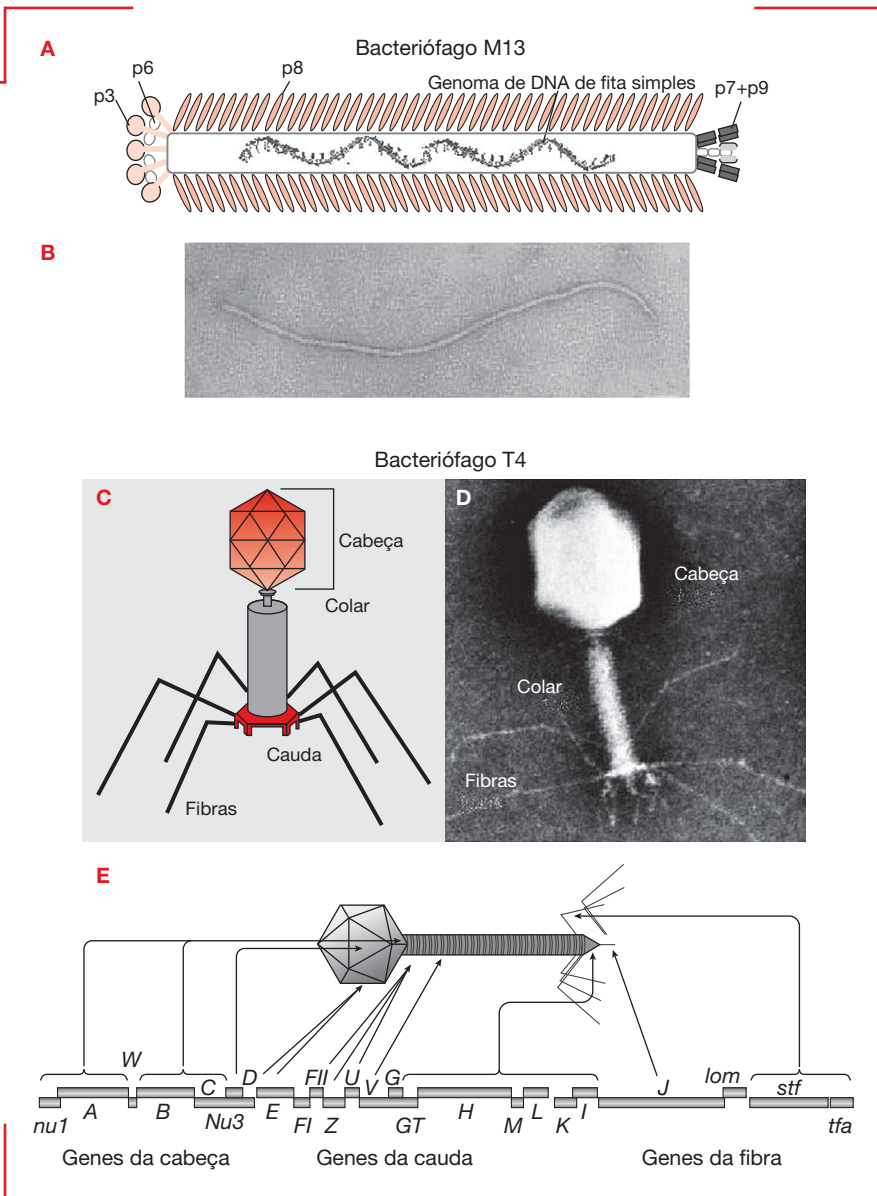
E um **ciclo lisogênico** (**Figura 9.6B**), fagos temperados são capazes de perpetuar o seu material genético sem provocar a morte da célula hospedeira. Nesse processo, o genoma do fago se integra, por recombinação sítio-específica, no genoma bacteriano (ver Capítulo 8). O processo é reversível, de forma que o genoma do fago pode ser eventualmente liberado, tornando possível sua entrada em um ciclo lítico. A bactéria que carrega uma cópia do genoma viral integrada no seu cromossomo é chamada de **lisogênica**, e o genoma do fago, que mantém a capacidade potencial de lise da célula hospedeira, é denominado **profago**. Bactérias lisogênicas possuem **imunidade** contra a infecção adicional por outros fagos do mesmo tipo, pois o profago possui um gene determinando a produção de uma proteína repressora que, além de manter seu circuito regulador lisogênico, reprime o ciclo lítico de qualquer outro fago possuindo no seu genoma os sítios reguladores onde essa proteína atua (a chamada **região de imunidade** ou *imm*). Por exemplo, o bacteriófago λ , quando integrado no genoma de *E. coli*, expressa uma proteína repressora (produto do gene *cI*) que confere imunidade à célula hospedeira contra a infecção por outro bacteriófago λ ou por qualquer outro lambdoídeo, como ϕ 80, 21 ou 434. Isso significa que os bacteriófagos λ , ϕ 80, 21 e 434 possuem a mesma região de imunidade nos seus genomas, embora as demais regiões sejam diferentes.

2.2 Mobilização de material genético bacteriano mediada por bacteriófagos

Certos fagos têm a capacidade de atuar como vetores naturais para a transferência de pequenos segmentos de

Figura 9.5

Organização do capsídeo de dois diferentes fagos. (A) Esquema da estrutura do bacteriófago filamentoso M13, em que o capsídeo é formado principalmente pela proteína p8, contendo também outras 4 proteínas (p3, p6, p7 e p9) nas extremidades. O genoma é composto por uma molécula de DNA de fita simples. **(B)** Micrografia eletrônica de M13. **(C)** Organização do capsídeo do bacteriófago T4. A cabeça icosaédrica, o colar, a cauda e as fibras são formadas por várias dezenas de proteínas. **(D)** Micrografia eletrônica de T4. **(E)** Os 24 genes que codificam as proteínas no capsídeo (*nu1*, *A*, *W*, *B*, etc.).

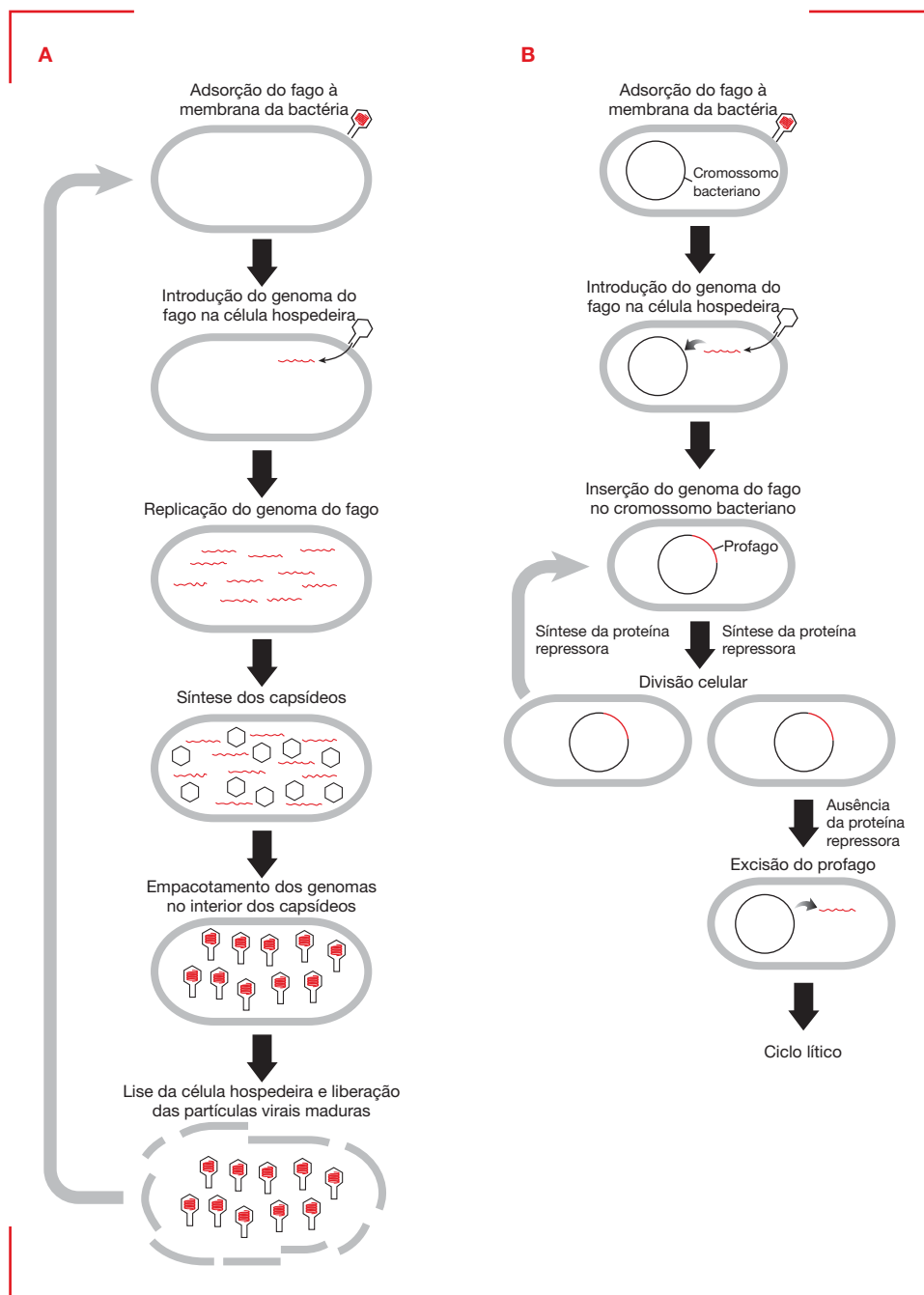


DNA do cromossomo bacteriano de uma célula para outra, em um processo denominado **transdução**. A transdução pode ser de dois tipos: generalizada, quando envolve a mobilização de qualquer região do cromossomo; e especializada, quando envolve uma região específica do cromossomo.

Na **transdução generalizada** (que ocorre, por exemplo, com o bacteriófago P1), durante a multiplicação do fago, eventualmente (em uma frequência de 3×10^{-3}), uma pequena quantidade de DNA do cromossomo bacteriano é empacotada no lugar do genoma viral (**Figura 9.7A**). Esse empacotamento incorreto ocorre ao acaso, de modo que qualquer região do cromossomo da célula hospedeira pode estar envolvida. O capsídeo do fago atua, então, como vetor na transferência do segmento de DNA cromossômico que é, em parte ou totalmente,

integrado ao genoma da célula receptora por recombinação homóloga.

A **transdução especializada** envolve fagos que se integram em sítios-específicos do genoma bacteriano, como o bacteriófago λ . A integração do genoma de λ ao cromossomo da célula hospedeira ocorre por recombinação sítio-específica, na região entre os óperons *bio* e *gal* do genoma de *E. coli*, no chamado sítio *att*, em um processo reversível. Assim, na transdução especializada mediada por λ (**Figura 9.7B**), quando fagos lisogênicos têm sua excisão induzida, esse processo ocasionalmente (uma vez a cada 10^5) ocorre de forma anormal. Quando isso acontece, é produzido um fago com genoma defeutivo, com parte do seu cromossomo substituído por um segmento de DNA da bactéria, que só pode incluir seqüências adjacentes ao sítio *att*.

**Figura 9.6****Representações esquemáticas dos ciclos lítico e lisogênico de um fago.**

(A) Ciclo lítico: um fago adere a uma bactéria hospedeira e injeta o seu genoma (no caso, uma molécula de DNA linear de fita dupla) na célula. Após a infecção, a maquinaria de transcrição/tradução da bactéria expressa genes do fago que determinam a replicação do genoma viral, a síntese das proteínas do capsídeo e a formação de partículas virais maduras. Ao final do ciclo, as novas partículas virais formadas são liberadas e podem infectar novas células hospedeiras suscetíveis. Na figura, o DNA viral está representado em ■, e o cromossomo da célula hospedeira não é mostrado.

(B) Ciclo lisogênico: um fago temperado, ao infectar uma bactéria, pode integrar o seu genoma (em ■) no cromossomo da célula hospedeira por recombinação sítio-específica. O genoma integrado pode ser mantido nesta forma, chamada de profago, ao longo da linhagem bacteriana, que é chamada de lisogênica. A manutenção do ciclo lisogênico depende da produção de uma proteína repressora, que mantém o circuito regulador lisogênico e reprime o ciclo lítico (ver Capítulo 13). Na ausência da proteína repressora ocorre a excisão do profago e a entrada no ciclo lítico.

3. Elementos de transposição

3.1 Características gerais e mecanismos de transposição

Elementos de transponíveis ou **transposons** são segmentos de DNA que possuem a propriedade de transposição, ou seja, são capazes de se mover de um ponto a outro de um cromossomo, mudar de cromossomo e, às vezes, “saltar” do genoma de uma espécie para o de outra. Esses elementos foram descobertos a partir de experimen-

tos genéticos realizados na década de 1940 (**ver Material Complementar Online 9.1**) e, desde então, a sua representatividade e importância funcional vem sendo cada vez mais evidenciada em diversas espécies. Além disso, a partir da aquisição de conhecimentos sobre os mecanismos e consequências dos eventos de mobilização dos transposons nos genomas, alguns desses elementos tornaram-se importantes ferramentas experimentais para a biologia molecular (**ver Material Complementar Online 9.2**).

Os elementos transponíveis estão presentes em quase todos os genomas já estudados, tanto de procariontos

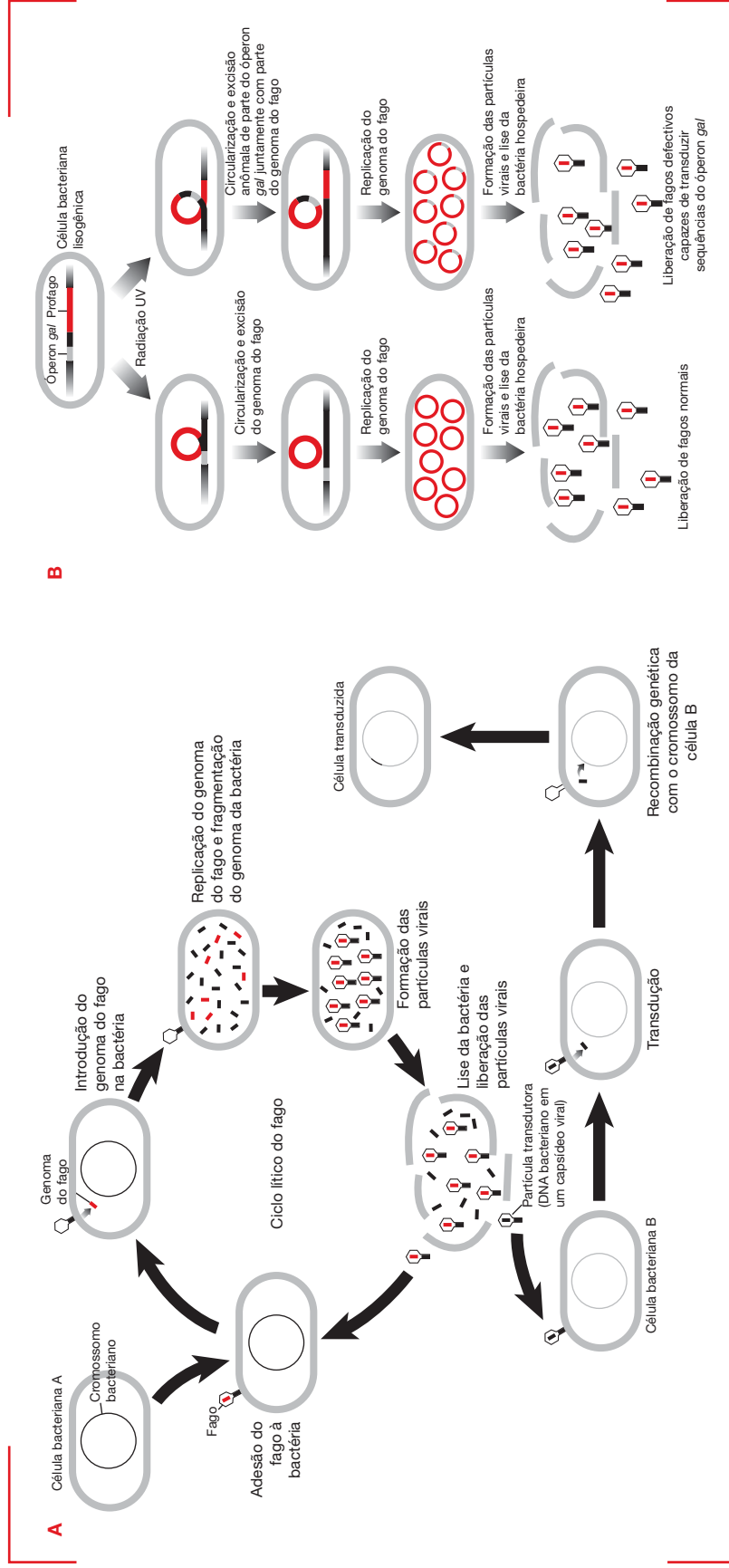


Figura 9.7

Representação esquemática de eventos de transdução. (A) Na transdução generalizada, um fago infecta uma bactéria e entra em ciclo lítico, seu genoma é replicado inúmeras vezes e o cromossomo da célula hospedeira fragmentado. Durante o processo de formação das partículas virais, um fragmento de DNA da bactéria pode ser acidentalmente empacotado em um capsídeo viral, formando uma partícula transdutora. A partícula transdutora, ao infectar uma outra célula bacteriana, transfere para ela o fragmento de DNA da primeira bactéria, executando a transdução. Às vezes, o fragmento transferido pode ser incorporado ao genoma da célula receptora por recombinação genética, produzindo uma célula transduzida estável. (B) Na transdução especializada mediada pelo bacteriófago λ , o fago tem, no ciclo lisogênico, seu genoma integrado ao sítio-específico do cromossomo da células hospedeira, adjacente ao operon *gal*. A excisão do profago depende de indução, que desencadeia o ciclo lítico. Em um evento de excisão normal (à esquerda), o genoma do fago é precisamente removido do cromossomo bacteriano e, depois de replicado, origina partículas virais normais. Um evento de excisão anormal (à direita) gera um genoma defeitivo, no qual parte do genoma do fago é substituída por sequências do operon *gal*. Este genoma defeitivo, depois de replicado, gera partículas virais transdutoras capazes de transferir as sequências do operon *gal* para outra bactéria (no evento de transdução).

como de eucariotos. Em muitos genomas, principalmente nos de espécies eucarióticas, uma parte significativa é constituída por elementos transponíveis (ver Capítulo 5). Por exemplo, 45% do genoma humano é formado por elementos transponíveis e, em algumas plantas e anfíbios, esta fração é ainda maior, podendo ser de 70 a 90%. Em geral, há uma correlação entre o tamanho do genoma e a proporção de elementos transponíveis (**Tabela 9.2**).

Poucos são os exemplos de genomas em que não foram detectados elementos transponíveis. Entre os eucariotos, elementos transponíveis não foram encontrados apenas em alguns organismos unicelulares, geralmente patógenos intracelulares e que têm genomas reduzidos, como espécies dos gêneros *Plasmodium*, *Babesia* e *Cryptosporidium*. Em procarionotos há ocorrência significativa de elementos transponíveis (ver Capítulo 4), e os exemplos de genomas sem elementos transponíveis também são de bactérias parasitas intracelulares e com genomas reduzidos, como algumas espécies dos gêneros *Mycoplasma*, *Buchnera* e *Blochmannia*. Entretanto, esta questão não pode ser generalizada, pois algumas espécies dos gêneros *Mycoplasma* e *Wolbachia* (parasitas intracelulares) chegam a ter 13% de seus genomas compostos por elementos transponíveis.

Além de abundantes em muitos genomas, os elementos transponíveis são também muito diversos, podendo variar quanto aos mecanismos de transposição que empregam e aos genes que possuem (até mesmo não possuir gene algum). No que diz respeito à organização estrutural, os elementos transponíveis podem também variar quanto à presença de sequências repetidas, a ordem dos genes, entre outras características.

Em relação aos **mecanismos de transposição**, existem dois tipos principais, diferentes quanto à molécula, de RNA ou DNA, que será intermediária na transposição. Alguns elementos transponíveis movem-se por um intermediário de RNA (**Figura 9.8A**). Nestes casos, os elementos transponíveis são transcritos e o RNA serve de molde para transcrição reversa pela enzima **transcriptase reversa**, ou seja, uma cópia de DNA é sintetizada a partir do RNA do elemento transponível. O DNA proveniente é inserido no genoma, resultando em uma nova cópia do elemento transponível. Esse processo de transposição é chamado de **“cópia e colagem”** ou **retrotransposição**, e os elementos transponíveis que o

utilizam são denominados **retrotransposons** ou **elementos transponíveis de Classe I**.

Outros elementos usam um intermediário de DNA para transposição, sendo chamados de **transposons** ou **elementos transponíveis de Classe II**. Esses elementos empregam uma enzima chamada de **transposase**, que, em alguns casos, pode retirar o elemento do seu sítio e transpô-lo para um novo sítio (Figura 9.8B). Esse processo é chamado de **“corte e colagem”**. Em outros casos, a transposase pode manter o elemento em seu sítio original e transpor uma cópia dele para um novo sítio.

O mecanismo de transposição utilizado pelos retrotransposons (cópia e colagem) sempre resulta no aumento do número de cópias, pois o elemento que origina a nova cópia nunca sai de seu sítio original e, por isso, é chamado de **transposição replicativa**. O mecanismo de corte e colagem utilizado pelos transposons de DNA, pode manter o número de cópias dos elementos transponíveis no genoma, pois, no processo chamado de **transposição conservativa**, o transposon sai de seu sítio de origem e vai ocupar um novo. Entretanto, cabe salientar que os transposons de DNA são capazes de fazer tanto a transposição conservativa como a replicativa e, no segundo caso, o número de cópias de elementos transponíveis aumenta no genoma.

Em ambas as classes de elementos transponíveis, retrotransposons e transposons, os elementos podem ser ou não autônomos. Os elementos transponíveis, denominados **autônomos**, possuem os genes que codificam os produtos requeridos para transposição (transcriptases reversas ou transposases, por exemplo) (**Figura 9.9A**), e podem promover sua própria mobilização. Em contrapartida, os **não autônomos** (Figura 9.9B) não codificam proteínas para transposição, mas são capazes de se transpor, pois retêm as sequências necessárias para a transposição (ver Seções 3.3 e 3.4). Eles dependem, porém, da atividade de uma transcriptase reversa ou de uma transposase suprida por um elemento autônomo. Há ainda nos genomas **elementos inativos**, representados pelos elementos que possuem mutações nas sequências necessárias à mobilização e que, por isso, não são mais mobilizáveis.

As classes de elementos transponíveis I e II são ainda subdivididas em subclasses, superfamílias, famílias e

Tabela 9.2 Tamanho de genoma e fração ocupada por elementos transponíveis em diferentes espécies

Organismo	Tamanho do genoma (Mb)	Fração de elementos transponíveis (%)
<i>Rana esculenta</i> (rã verde)	6.850	77
<i>Homo sapiens</i> (ser humano)	3.038	45
<i>Drosophila melanogaster</i> (mosca-das-frutas)	150	15
<i>Caenorhabditis elegans</i> (nematódeo)	108	12
<i>Saccharomyces cerevisiae</i> (levedura)	12	3
<i>E. coli</i> (bactéria)	4,6	0,3

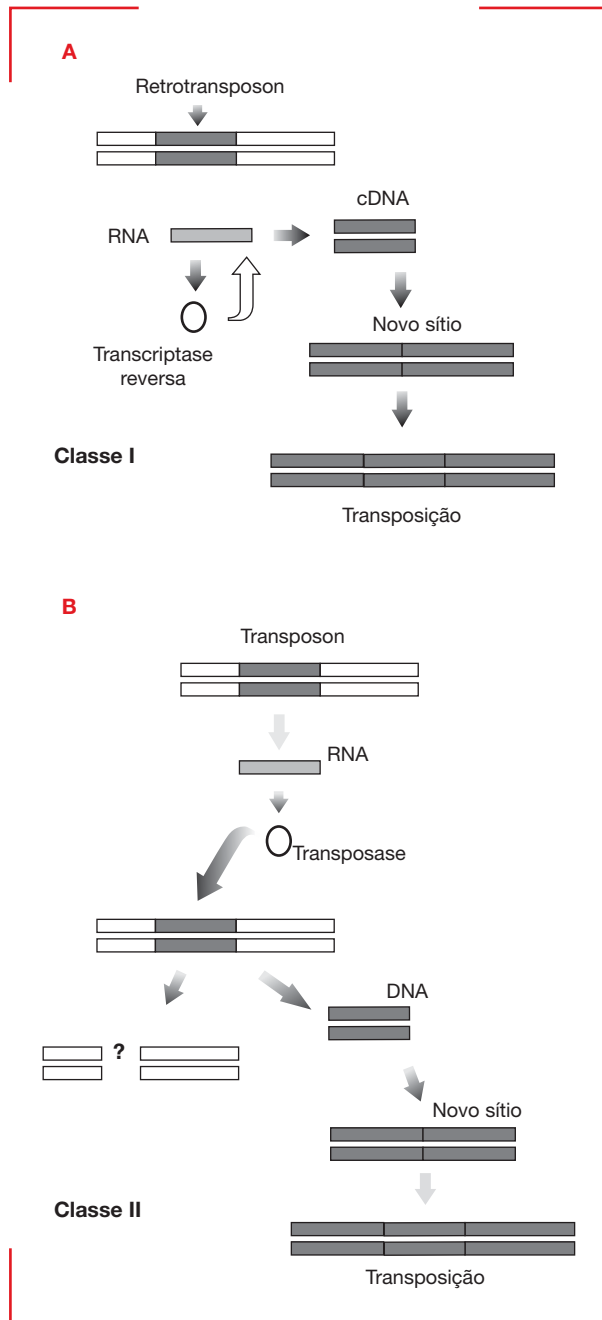


Figura 9.8

Comparação entre os mecanismos de transposição dos retrotransposons (classe I) e dos transposons de DNA (classe II).

(A) Os retrotransposons são transcritos e seu RNA traduzido. Algumas proteínas são produzidas, entre elas a enzima transcriptase reversa, que utiliza o RNA como molde para fazer uma cópia de DNA complementar ao RNA (cDNA). O cDNA será transposto a um novo sítio por ação de enzimas como a integrase, que também é codificada pelo retrotransposon. (B) Os transposons de DNA são transcritos e seu RNA traduzido para uma enzima, a transposase. Essa enzima pode excisar o transposon de seu antigo sítio e o transpor para um novo.

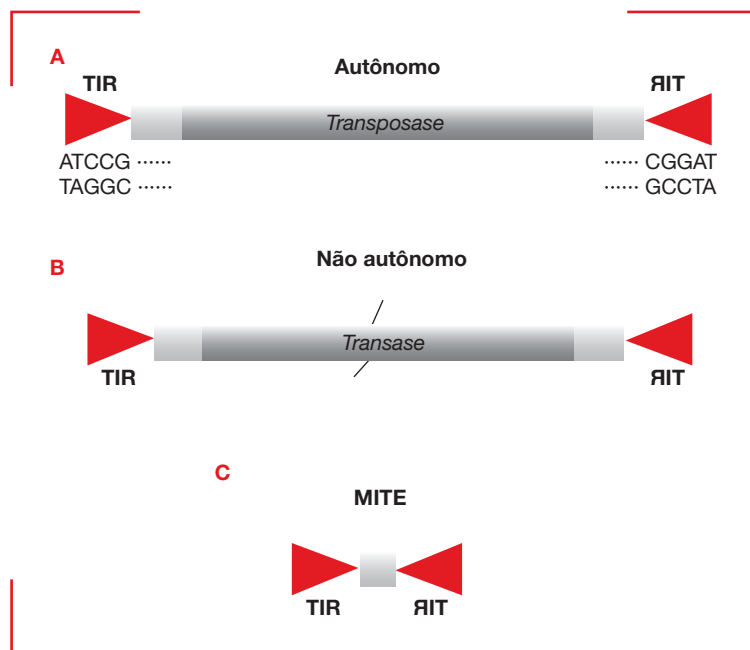
subfamílias. Estas classificações se baseiam em características estruturais e moleculares, como os genes que os elementos possuem, a ordem desses genes, a similaridade de sequências ou a presença de certos domínios nas proteínas codificadas. Esta classificação se faz necessária porque há uma diversidade muito grande de elementos transponíveis, principalmente em genomas maiores, como os de metazoários e plantas. Por exemplo, mesmo em um genoma não muito grande, como o da mosca *D. melanogaster*, encontramos aproximadamente 100 diferentes famílias de elementos transponíveis. A representatividade dos diferentes tipos de elementos transponíveis nos genomas eucarióticos é discutida no Capítulo 5.

3.2 Elementos de DNA ou transposons

Os elementos que se transpõem por um intermediário de DNA utilizam a enzima transposase para este processo (corte e colagem). Essa enzima reconhece determinadas sequências do DNA, presentes em ambas as extremidades do elemento, mas invertidas uma em relação à outra. Estas sequências, denominadas **TIR** (de *terminal inverted repeats* = repetições invertidas terminais), são essenciais para a mobilização do transposon (ver Figura 9.9A). Em geral, essas repetições são curtas (entre 10 e 40 pb), entretanto, em alguns elementos transponíveis, como no *Merlin*, encontrado no nematódeo *Caenorhabditis briggsae*, e nos elementos *Bari* e *Paris* de *Drosophila* spp., as TIRs são bem maiores, tendo em torno de 250 pb. Os elementos de DNA não autônomos típicos, com o gene da transposase “truncado” (ver Figura 9.9B), possuem TIRs como as dos elementos íntegros dos quais são derivados.

Há ainda uma classe de elementos não autônomos com estrutura mais diferenciada. Esses elementos, chamados **MITEs** (de *miniature inverted-repeat transposable elements* = elementos transponíveis miniatura com repetições invertidas), geralmente possuem uma ou poucas centenas de pares de bases (Figura 9.9C), TIRs conservadas e a sequência interna pode ter similaridades com parte da sequência de elementos da família com a qual são relacionados, ou ter, ainda, sequências sem qualquer similaridade com outros elementos. Alguns genomas, como os de plantas e de mosquitos, podem conter milhares de cópias de MITEs. Por exemplo, no mosquito *Aedes aegypti*, 16% do genoma é composto por MITEs e milhares dessas sequências estão presentes nos genomas do arroz e do milho.

No processo da transposição, a transposase atua como um dímero, sendo que cada monômero da enzima se liga a uma TIR. Quando o dímero da transposase liga-se às TIRs, a parte da molécula de DNA contendo o transposon forma uma alça (Figura 9.10). Quando a enzima “corta” o elemento transponível do sítio doador, o cromossomo “doador” sofre uma quebra, que pode ser reparada pelo sistema de reparação, que liga as fitas quebradas ou utiliza o cromossomo homólogo para fazer a reparação por recombinação. Se o mecanismo de reparação usar o cromossomo homólogo e nele existir uma

**Figura 9.9**

Estrutura esquemática de transposons de DNA. (A) Elemento autônomo com a presença das terminações repetidas e invertidas (TIR) nas suas extremidades e o gene da transposase. (B) Elemento não autônomo com o gene da transposase “truncado”. (C) Estrutura de um MITE (*miniature inverted-repeat transposable element*).

cópia do elemento transponível, o resultado será a manutenção desse elemento na posição original, constituindo um evento de transposição replicativa.

A transposase, carregando o transposon “recortado” do sítio doador, inserirá o elemento em um sítio-alvo. As transposases não são seletivas em relação à sequência do sítio-alvo e esta inserção ocorre ao acaso. A transposase cliva o sítio-alvo de forma a gerar extremidades de fita simples complementares (extremidades coesivas; ver Capítulo 16) e, em consequência disso, a inserção do elemento transponível seguida pela ação do mecanismo de reparação (Capítulo 7), provoca a duplicação da sequência do sítio-alvo nos dois lados do elemento transponível (**Figura 9.11**). Esta duplicação é chamada de **TSD** (de *target sequence duplication* = duplicação do sítio-alvo).

Nem todos os elementos de DNA empregam a transposase como enzima para mobilização. Alguns elementos utilizam um processo de replicação do tipo “círculo rolante” (ver Capítulo 6), usando uma helicase, que separa as fitas do DNA, e uma integrase, que abre a fita dupla de DNA no sítio-alvo para a inserção de uma nova cópia do elemento transponível. Estes elementos são chamados de **hélitrons** e **Maverick** e ambos são amplamente distribuídos em plantas, metazoários e fungos.

3.3 Retrotransposons

Os **retrotransposons**, elementos que transpõem através de um intermediário de RNA, empregam a enzima transcriptase reversa para fazer uma cópia de DNA a partir de um transcrito de RNA do elemento. Entretanto nos elementos autônomos, o gene dessa enzima não é o único presente, pois outras proteínas são necessárias para a realização da retrotransposição, entre elas a

RNase H, que degrada o híbrido DNA-RNA formado após a transcrição reversa e a integrase.

Os retrotransposons podem ser divididos em dois grupos, o dos **retrotransposons com LTR**, que possuem longas repetições terminais (**LTR**, de *long terminal repeats*), e o dos **retrotransposons sem LTR** (**Figura 9.12**). As LTRs são sequências repetidas diretas, com 200 a 500 pb, que estão envolvidas no processo de transposição e regulação da expressão de genes do elemento transponível.

Os retrotransposons com LTR (Figura 9.12A e B) podem ser abundantes nos genomas de animais, por exemplo, 10% do genoma do mosquito *Aedes aegypti* e aproximadamente 5% do genoma de *D. melanogaster* é formado por retrotransposons com LTR. Porém, é nas plantas, principalmente em gramíneas, que esses elementos são muito abundantes. Nas gramíneas, os elementos das duas principais superfamílias de retrotransposons com LTR, *gypsy* e *copia*, podem estar representados por milhões de cópias, perfazendo 60% ou mais do genoma.

Alguns retrotransposons com LTRs possuem um gene para a formação de capsídeo (gene GAG) e, por isso, são capazes de produzir **partículas similares a vírus (VLPs)**, de *virus-like particles* sem capacidade infecciosa. Outros retrotransposons com LTRs, além do gene GAG, possuem o gene ENV, que codifica as proteínas necessárias para a formação do envelope viral. Os elementos que possuem ambos os genes (GAG e ENV) são capazes de produzir partículas virais infecciosas. Em virtude da presença desses genes, os retrotransposons com LTR apresentam tamanhos maiores do que os dos transposons de DNA. O tamanho médio dos retrotransposons com LTR é de 5 a 7 kb, mas alguns podem chegar a 25 kb.

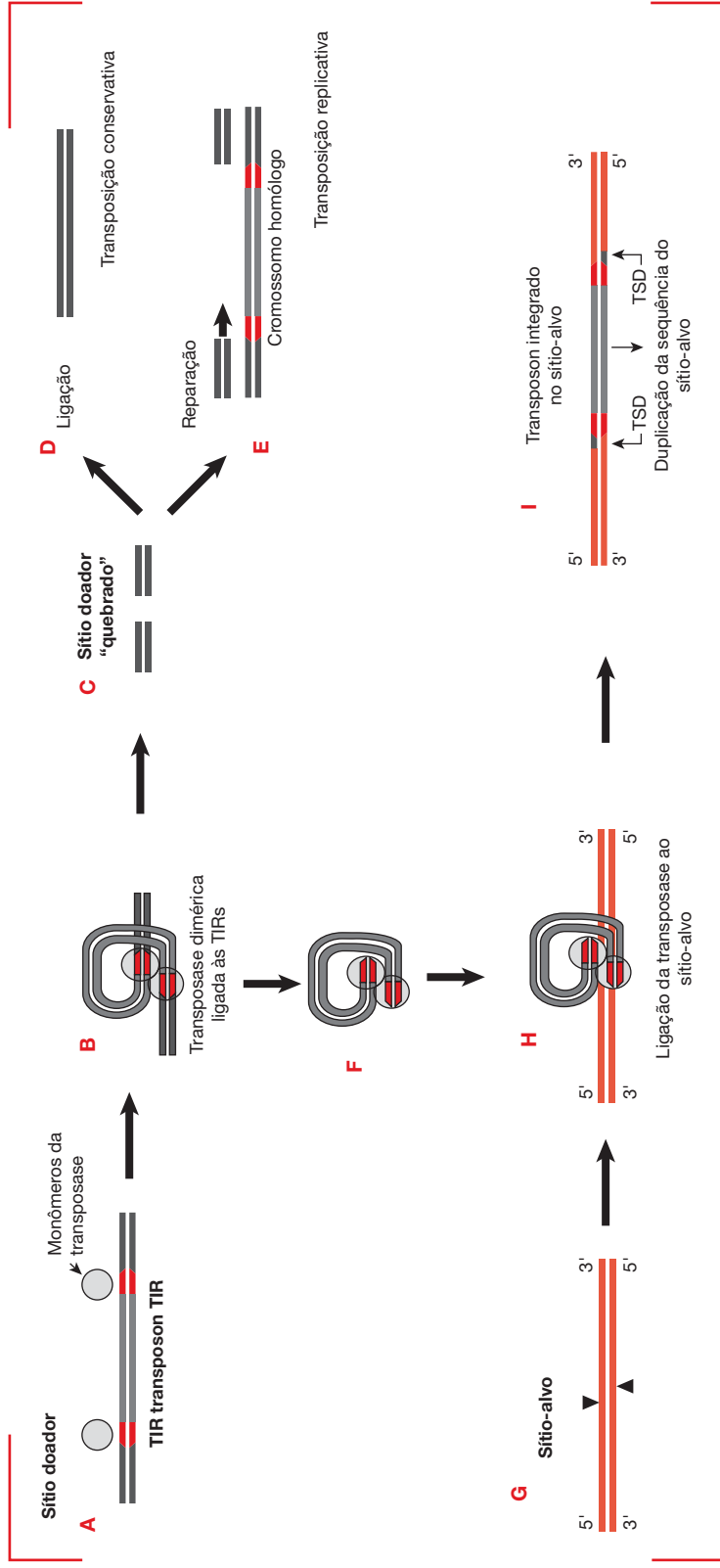


Figura 9.10

Processo de transposição mediado pela transposase (corte e colagem). (A) A enzima transposase liga-se às terminações repetidas e invertidas (TIRs). (B) O dímero da enzima ligado ao DNA do transposon forma uma "alça". (C) Ao ser cortado pela transposase, o sítio doador se quebra e é reparado pelo sistema de reparação celular, que poderá ligar as duas fitas de DNA (D), ou usar o cromossomo homólogo para o reparação (E). A transposase clivará o sítio-alvo de modo a gerar extremidades coesivas de fita simples (G) e inserirá o transposon neste sítio (H). Como a clivagem do DNA gera extremidades coesivas de fita simples, haverá a duplicação do sítio-alvo em ambos os lados do transposon (TSDs) (I).

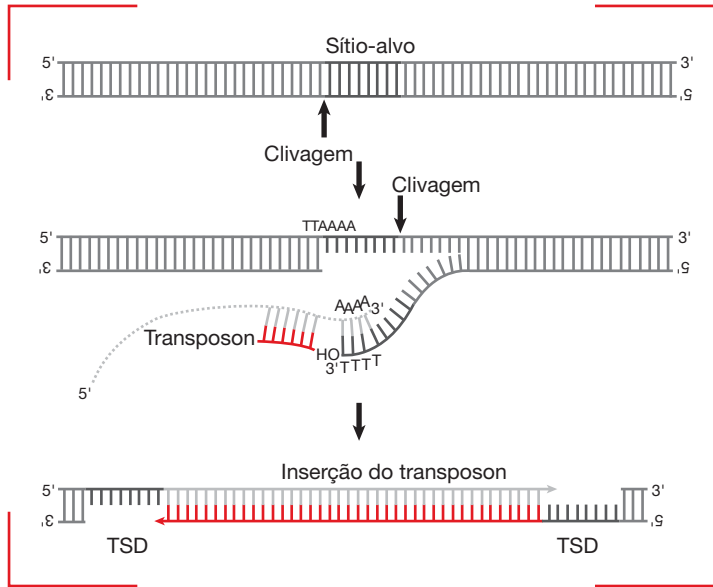


Figura 9.11

Esquema mostrando a duplicação do sítio-alvo. As enzimas clivam o DNA no sítio-alvo de forma a gerar extremidades coesivas de fita simples, que flanqueiam o transposon após a sua inserção. Quando o sistema de reparação preenche as lacunas de fita simples, há a duplicação desta sequência (TSD).

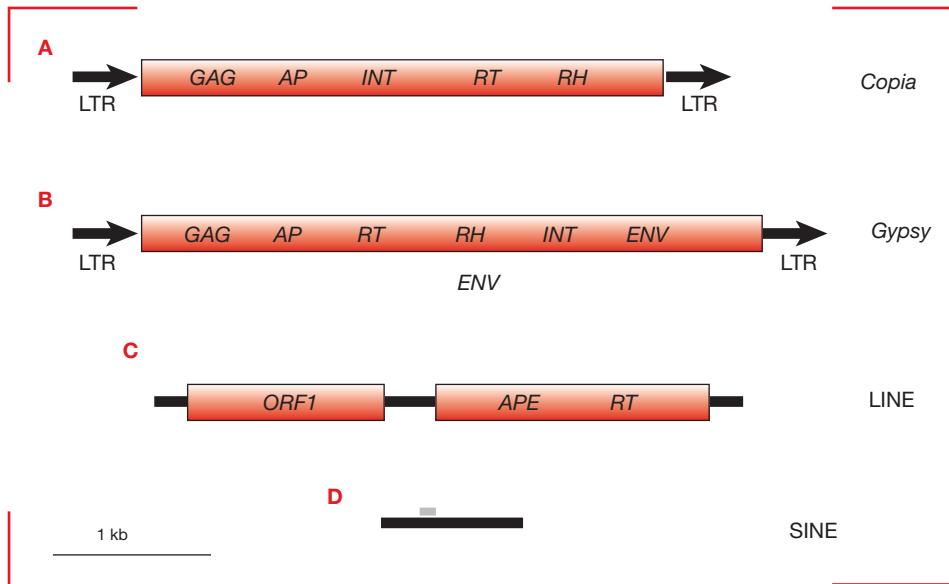


Figura 9.12

Organização estrutural de retrotransposons com LTR, representados pelos elementos *Copia* (A) e *Gypsy* (B), e de retrotransposons sem LTR, representados pelos LINEs (C) e SINEs (D). Em (A) e (B), estão representados os genes GAG, para as proteínas do capsídeo; AP, que codifica uma proteinase; INT, da integrase; RT, da transcriptase reversa; RH, da RNase H; e ENV, para as proteínas do envelope viral (este gene é facultativo em elementos da família *Gypsy*). Em (C), ORF1 é a sequência codificadora de uma proteína de ligação a RNA de função desconhecida e APE/RT codifica uma enzima bifuncional com atividade de endonuclease e transcriptase reversa, encontradas em LINEs; os SINEs (D) não possuem genes e a barra menor, em ■, representa sequências dos elementos as quais se ligam iniciadores para a retrotranscrição e que são características para cada família de SINE.

Os **retrotransposons sem LTR** ou **retroposons** (Figura 9.12C e D), por sua vez, não possuem genes para a formação de capsídeo e, portanto, não formam VLPs ou partículas virais. Como exemplos desses elementos transponíveis podem-se citar os **LINEs** (de *long interspersed nuclear elements* = elementos nucleares longos dispersos), que, quando completos, têm de 6 a 7 kb e possuem dois genes, um codificador de uma proteína que se liga ao RNA, e outro codificador de uma proteína que tem atividades de transcriptase reversa e de endonuclease. O genoma humano contém aproximadamente meio milhão de cópias de elementos LINEs e estima-se que de 50 a 100 dessas cópias ainda sejam capazes de ser mobilizadas.

Outros retrotransposons sem LTR são os **SINEs** (de *short interspersed nuclear elements* = elementos nucleares curtos dispersos). Os SINEs são curtos (com menos de 500 pb), não possuem qualquer gene e dependem dos LINEs para a sua transposição. No genoma humano, os SINEs também são muito abundantes. Só o elemento *Alu*,

o SINE mais abundante em nosso genoma, está representado por mais de um milhão de cópias.

As relações evolutivas entre os retrotransposons e os transposons são inferidas a partir da presença de sequências funcionais comuns nesses elementos (**Figura 9.13**). Os elementos de DNA teriam sido a fonte do gene da integrase para os retrotransposons. Em uma forma intermediária, os retrotransposons teriam os genes da transcriptase reversa e da RNase H, além do gene das proteínas de capsídeo. A partir desse ancestral, um ramo daria origem aos LINEs e outro “adquirido” as LTRs, originando os retrotransposons com LTR.

A maior parte dos retrotransposons com LTR não possui o gene ENV, sendo chamados de **retrovírus endógenos**, mas alguns retrotransposons com LTR incorporaram na sua sequência o gene ENV, originando os retrotransposons com ENV e os retrovírus. Por isso, a distinção entre retrotransposons com LTR e retrovírus não é fácil, pois parece haver um *continuum* estrutural e funcional entre essas entidades. Por exemplo, retro-

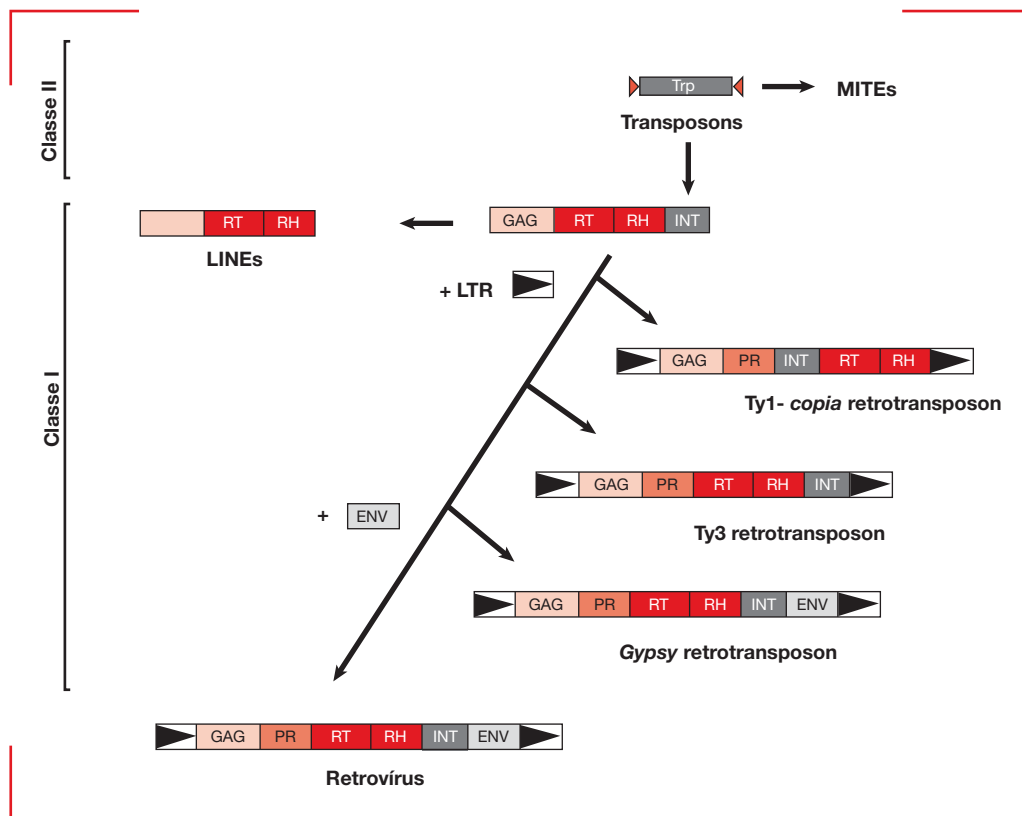


Figura 9.13

Relações evolutivas sugeridas para os elementos de transposição. Os transposons de DNA seriam a fonte da integrase dos retrotransposons. Um ramo dos retrotransposons teria perdido a integrase e originado os LINEs, ao passo que outro adquiriu LTRs e originou os atuais grupos de retrotransposons com LTR. Em um segundo momento, esses elementos receberam o gene ENV, originando os elementos relacionados a *Gypsy* e os retrovírus. Os genes representados na figura são os codificadores de transposase (*Trp*); proteínas de capsídeo (*GAG*); proteinase (*AP*); integrase (*INT*); transcriptase reversa (*RT*); RNase H (*H*); e proteínas de envelope viral (*ENV*).

transposons com LTR como o *gypsy*, de *D. melanogaster*, que possuem o gene ENV, apesar de se propagarem como retrotransposons típicos, também são capazes de produzir partículas virais infectivas e de se propagarem como retrovírus.

3.4 Consequências biológicas da transposição

A transposição dos elementos transponíveis pode gerar mutações, quebras e rearranjos cromossômicos, assim como alterações epigenéticas (mudanças reversíveis e herdáveis no genoma) que não alteram a sequência de nucleotídeos do DNA. As mutações geradas pela inserção de um elemento transponível podem ter características diferentes daquelas provocadas por agentes químicos ou físicos. Isto porque, a inserção de um elemento transponível em uma região codificadora ou reguladora de um gene pode fazer com que ele não seja mais expresso, ou que a proteína por ele codificada não seja mais funcional. Outras vezes, o resultado da mutação causada por um elemento transponível é a formação de um novo padrão de expressão gênica. Esse processo acontece porque os elementos transponíveis possuem suas próprias regiões reguladoras transcricionais, que respondem aos fatores de transcrição produzidos pelos genes reguladores do organismo hospedeiro (ver Capítulo 14). Assim, quando um elemento transponível se insere próximo à região codificadora de um gene, pode levá-la a ter um novo padrão de expressão, pois esta região passa a ser expressa a partir dos sítios reguladores do elemento transponível.

Alterações no funcionamento gênico, causadas pela mobilização de elementos de transposição, podem ser consideradas responsáveis por aproximadamente 1% das doenças humanas. Foram relatados casos de hemofilia e distrofia muscular de Duchene, assim como tumores e câncer de mama causados pela mobilização de elementos LINES e SINES. Casos de teratocarcinoma, leucemia, esquizofrenia e diabetes já foram associados a inserções de retrovírus endógenos humanos (**HERV**, de *human endogenous retrovirus*), que são retrotransposons.

Os elementos transponíveis, por apresentarem muitas cópias no genoma, podem se tornar sítios para pareamento e recombinação entre diferentes regiões cromossômicas. Este fenômeno, chamado de **recombinação ectópica**, gera quebras e rearranjos cromossômicos.

A taxa de mutações causadas pela mobilização dos elementos transponíveis é bastante variada entre os organismos. Em espécies do gênero *Drosophila*, por exemplo, estima-se que de 50 a 80% das mutações espontâneas são decorrentes de inserções de elementos transponíveis. Já em seres humanos, esta estimativa é de apenas 0,1 a 1%, apesar do genoma humano ter uma fração maior de elementos transponíveis (da ordem de 45%) do que o genoma de *Drosophila* (da ordem de 15%). Isso reflete o fato de que, em seres humanos, esses elementos são, na maioria, inativos em decorrência da ação de mecanismos celu-

lares de regulação e silenciamento gênico mais eficientes do que o observado em *Drosophila*.

Para minimizar efeitos deletérios associados à transposição, os organismos desenvolveram, no decorrer da evolução, vários mecanismos para a regulação ou silenciamento dos elementos de transposição. Os principais são a **metilação do DNA**, a **modificação de histonas** e a **interferência por RNA**. A **metilação** dos resíduos de citosina, em genomas eucarióticos, é frequentemente associada a sequências repetitivas, incluindo elementos transponíveis. A metilação dessas sequências reflete a ação de um sistema de defesa genômica, que pode silenciar os elementos transponíveis para minimização dos efeitos deletérios causados pela sua mobilização.

Nos eucariotos, em que o DNA está complexado com histonas (ver Capítulo 13), a presença de elementos de transposição resulta, muitas vezes, em alterações na cromatina nas regiões cromossômicas onde eles se encontram. Estas alterações decorrem da **modificação de histonas**, que são modificações químicas reversíveis nas proteínas capazes de fazer a cromatina adotar formas mais compactas (inclusive de heterocromatina). Isso reduz a acessibilidade do DNA à maquinaria de transcrição e, onde a transcrição é dificultada ou impedida, há o silenciamento dos elementos transponíveis presentes.

Tem sido demonstrado, em diferentes organismos, o envolvimento da **interferência por RNA (RNAi, de RNA interference)**, descrita no Capítulo 14) na proteção contra proliferação de elementos transponíveis. Esse importante mecanismo de silenciamento gênico é mediado por RNA de fita dupla (dsRNA, de *double-stranded RNA*). Neste processo, de uma até várias cópias do elemento transponível são transcritas a partir de ambas as fitas (senso e antissenso), formando um dsRNA. Esse dsRNA será alvo do mecanismo de RNAi, promovendo a degradação de todos os mRNAs produzidos por elementos transponíveis com sequências similares às do elemento que produziu o dsRNA e, assim, impedindo a produção das enzimas necessárias a transposição.

Se para os indivíduos a mobilização dos elementos transponíveis pode trazer efeitos deletérios, como mutações somáticas e germinativas, para a evolução estes elementos são considerados como fatores fundamentais. Os elementos transponíveis são considerados importantes para a geração de variabilidade genética, que é a matéria-prima para a evolução dos genomas (ver Capítulos 4 e 5). Há vários exemplos de transposons que foram cooptados pelo genoma hospedeiro para o desempenho de novas funções, esse fenômeno é chamado de **domesticação gênica**.

O gene *sincitina*, fundamental para o desenvolvimento normal da placenta em seres humanos e ratos, corresponde a uma parte de um retrovírus endógeno. Outros exemplos são os genes *RAG1* e *RAG2*, codificadores das recombinases essenciais para a recombinação das cadeias V(D)J no sistema imune, que evoluíram a partir de

transposons de DNA. Os telômeros dos cromossomos de *D. melanogaster* são formados pela mobilização de elementos transponíveis (TART e Het-A). Acredita-se que, no decorrer do processo evolutivo, os mecanismos de metilação, modificação de histonas e RNA; (já discutidos) foram cooptados para controle da expressão de outros genes, responsáveis por vários aspectos do fenótipo dos organismos, criando, assim, novas redes regulatórias e novos fenótipos importantes para a evolução.

Uma característica marcante, não só dos elementos transponíveis como também dos elementos genéticos móveis em geral, é a capacidade de passar do genoma de uma espécie para o de outra. Este fenômeno é chamado de **transferência horizontal**. A transferência de informação genética entre espécies é também uma importante fonte de variabilidade, que pode contribuir no processo evolutivo (ver Capítulo 4).



Resumo

Elementos genéticos móveis são segmentos de DNA que podem movimentar-se em um genoma ou entre diferentes genomas. O conjunto dos elementos genéticos móveis de um genoma é chamado de mobiloma. Existem três tipos de elementos genéticos móveis: os plasmídeos, os fagos e os elementos transponíveis. Plasmídeos são elementos genéticos extra-cromossômicos, com capacidade de replicação autônoma, encontrados principalmente em procaríotos e podendo variar de 1 a 1.000 kb. Alguns plasmídeos estão envolvidos no fenômeno da conjugação bacteriana, que é a transferência de material genético através de um contato direto célula-célula.

Bacteriófagos ou fagos são vírus de procaríotos. Alguns fagos possuem genoma de RNA, ao passo que outros têm genoma de DNA de fita simples ou de DNA linear de fita dupla. Alguns fagos realizam ciclo lítico, ou seja, entram em uma célula, replicam-se e a matam, liberando as partículas virais. Outros, ainda, podem fazer, além do ciclo lítico, o lisogênico, no qual se integram ao cromossomo bacteriano. Certos fagos têm a capacidade de atuar como vetores naturais para a transferência de pequenos segmentos de DNA do cromossomo bacteriano de uma célula para outra, em um processo denominado transdução.

Elementos de transposição ou transposons são segmentos de DNA que possuem a propriedade de mover-se nos genomas. Alguns elementos de transposição, chamados de retrotransposons, se transpõem utilizando um intermediário de RNA. Os retrotransposons e alguns elementos de DNA transpõem-se através de mecanismos de “cópia e colagem”, denominado transposição replicativa, ao passo que outros, chamados de transposons, se transpõem por um mecanismo chamado “corte e colagem”, que é uma transposição não replicativa. Uma das principais consequências biológicas da transposição é a indução de mutações que podem ser germinativas ou somáticas, associando os transposons ao surgimento de doenças genéticas e câncer. Porém os elementos de transposição são também fonte importante de variabilidade genética, sendo esta a matéria-prima para a evolução.

Muitos elementos de transposição foram recrutados para executar funções celulares, em um processo chamado de domesticação gênica. Para minimizar os efeitos deletérios da transposição, os genomas desenvolveram mecanismos de silenciamento dos elementos de transposição. Os principais mecanismos são a metilação do DNA, a modificação de histonas e o de interferência por RNA (RNAi).



Leituras recomendadas

ACLAME. Description [Internet]. ACLAME; 2008 [capturado 28 ago. 2011]. Disponível em: <http://aclame.ulb.ac.be/Classification/description.html>.

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Biologia molecular da célula. 5. ed. Porto Alegre: Artmed; 2010.

Biémont C, Vieira C. Genetics: junk DNA as an evolutionary force. *Nature*. 2006;443(7111):521-24.

Capy P, Bazin C, Higuier D, Langin T. Dynamics and evolution of transposable elements. Austin: Landes Bioscience; 1998.

Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington: ASM Press; 2002.

IS Finder Database [Internet]. [capturado 28 ago. 2011]. Disponível em: <http://www-is.biotoul.fr/>.

Lankenau DH, Volff JN, editors. Transposons and the dynamic genome. Berlin: Springer; 2009.

Venner S, Feschotte C, Biéumont C. Dynamics of transposable elements: towards a community ecology of the genome. *Trends Genet.* 2009;25(7):317-23.

Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R. *Biologia molecular do gene*. 5. ed. Porto Alegre: Artmed; 2006.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet.* 2007;8(12): 973-82.

WikiPoson.org [Internet]. Hudson: Bioinformatics Organization; c2007 [capturado 23 ago. 2011]. Disponível em: <http://www.wikiposon.org>.

Augusto Schrank

Transcrição

- 1. Ciclo da transcrição 206**
- 2. Proteínas envolvidas 207**
 - 2.1 RNA-polimerases 207
 - 2.1.1 RNAPs de procariotos: bactérias e arqueas 208
 - 2.1.2 RNAPs de eucariotos 209
 - 2.2 Outras proteínas 210
 - 2.2.1 Topoisomerasas 210
 - 2.2.2 Fatores de transcrição 210
- 3. Promotores 211**
 - 3.1 Procariotos 211
 - 3.2 Eucariotos 213
- 4. Início da transcrição 214**
 - 4.1 Procariotos 214
 - 4.2 Eucariotos 217
 - 4.2.1 Início da transcrição por RNAPII 218
 - 4.2.2 Início da transcrição por RNAPI e RNAPIII 221
- 5. Alongamento da cadeia de RNA 222**
 - 5.1 Alongamento em procariotos 222
 - 5.2 Alongamento em eucariotos 224
- 6. Terminação da transcrição 226**
 - 6.1 Terminação em procariotos 226
 - 6.2 Terminação em eucariotos 227
- 7. Transcrição na cromatina 228**
- 8. Fidelidade e correção de erros 230**
- 9. Fábricas de transcrição 230**

A transcrição é o processo pelo qual são sintetizados todos os RNAs celulares. Esse processo conecta a informação presente na sequência de bases no genoma (genótipo) com as características funcionais da célula (fenótipo). Os RNAs sintetizados na transcrição produzem os RNAs funcionais e as proteínas e, portanto, a transcrição tem uma posição central na expressão gênica, sendo o seu primeiro passo. A transcrição também é um dos principais processos em que ocorre a **regulação da expressão** gênica, definindo quais genes ou sequências serão transcritas, quando esta expressão ocorre e quais as quantidades de produtos serão produzidas. A falha na transcrição de um determinado mRNA provoca a ausência da proteína por ele codificada e poderá causar alterações na célula.

A transcrição é mediada em todas as células por **RNA-polimerases** dependentes de DNA (RNAP), cuja estrutura e função são conservadas desde bactérias até seres humanos. Transcrição e síntese de RNA nem sempre são sinônimos. No caso dos vírus que possuem moléculas de RNA de fita dupla, como genoma, para que haja a sua replicação deve ocorrer a síntese de RNA (replicação), que não é transcrição, pois não está relacionada à expressão dos genes do vírus. Embora a transcrição seja muitas vezes referida como expressão dos genes, formalmente a expressão apenas ocorre quando a proteína madura ou o RNA funcional forem sintetizados. As alterações que acontecem nos RNAs após a transcrição são denominadas **alterações pós-transcricionais** e as ocorridas após a tradução das proteínas são denominadas **pós-traducionais**. Essas alterações são, em geral, necessárias para que ocorra a expressão dos genes.

Devido a sua importância, a transcrição é um dos processos mais estudados e também, no qual houve mais avanços na área da biologia molecular. Neste capítulo serão abordados aspectos básicos desse processo, mostrando um paralelo entre os sistemas menos complexos (procariotos) e mais complexos (eucariotos).

1. Ciclo da transcrição

A transcrição é um processo cíclico de síntese de RNA dividido em três fases: o **início**, em que sequências específicas do DNA (**promotores**) sinalizam o local de formação do complexo de transcrição para iniciar a cópia das sequências do DNA em RNA; a fase de **alongamento da cadeia**, na qual a molécula do RNA é sintetizada; e a **terminação** da transcrição, o processo em que a síntese do RNA é terminada em resposta a sinais específicos de terminação da transcrição (**terminadores**). Cada ciclo de transcrição ocorre várias vezes na mesma região do DNA (gene) e antes que um complexo atinja a região de terminação, outros complexos de transcrição estão iniciando a síntese de RNA. Portanto, embora normalmente seja mostrado um complexo de transcrição em um determinado gene, na realidade, a transcrição é dinâmica e várias cópias de um determinado RNA são sintetizadas concomitantemente a partir do gene. Em geral, os componen-

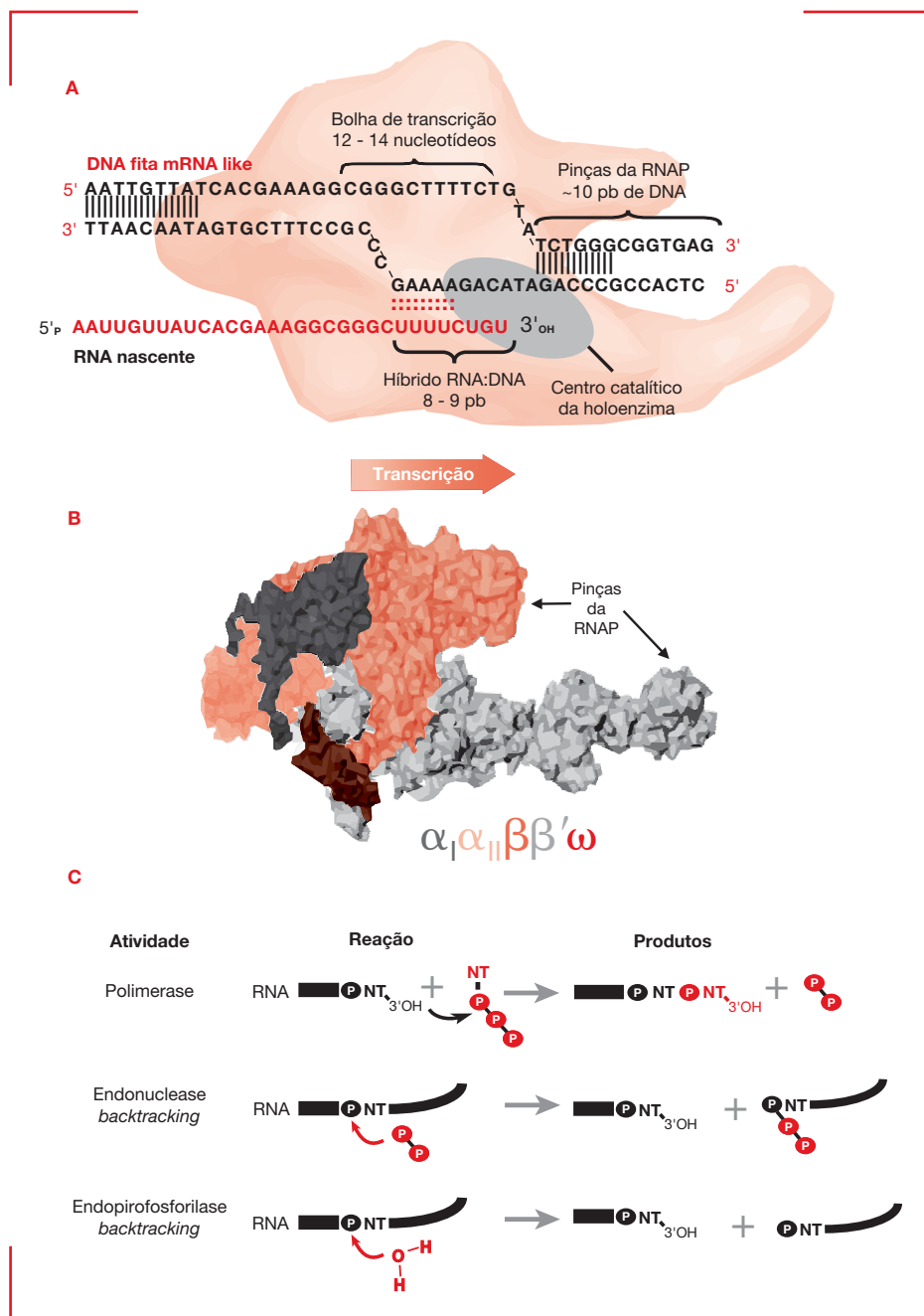
tes do complexo de transcrição são dissociados, em parte ou em sua totalidade, ao final da transcrição e reiniciam um novo ciclo de transcrição.

A transcrição tem muitas características comuns à replicação do DNA (ver Capítulo 6), como a reação básica, que é bastante semelhante, pois trata da incorporação de nucleotídeos à macromolécula que está sendo sintetizada. Na transcrição, os ribonucleotídeos (ATP, UTP, CTP e GTP) são adicionados sequencialmente à cadeia crescente de RNA. A reação depende, ainda, de um molde cuja sequência de bases é copiada por complementaridade. A molécula molde, neste caso, é uma das fitas do DNA hélice dupla. A reação também é **processiva** e a síntese ocorre na direção 5' para 3'. Da mesma forma que na replicação do DNA, na transcrição, a hélice dupla do DNA deve ser **desnaturada** (separada) por ruptura das pontes de hidrogênio para que a sequência de nucleotídeos da **fita-molde** possa ser exposta e copiada. Embora com consequências diferentes da replicação, na transcrição, a fidelidade da cópia deve ser alta e ocorrer em grande velocidade.

Sobre o DNA, é necessário compreender, de início, a sua nomenclatura e o fato de que apenas uma das suas fitas é utilizada como molde. (**Figura 10.1A**). Como referido no Capítulo 2, apenas uma das fitas de DNA é representada e escrita na orientação 5' para 3'. Essa fita é denominada fita similar ao RNA (*RNA-like*) e para obter a sequência da molécula de RNA que será sintetizada, respeitando seu início e sua terminação, basta substituir as **Ts** por **Us**. Por isso, também podemos chamar essa fita de **fita codificante** ou codificadora. A fita de RNA sintetizada é, portanto, copiada da outra fita de DNA, denominada **fita-molde**, pois a síntese ocorre por complementaridade (ver Figura 10.1A). Existe alguma confusão na literatura em relação a estas definições. Consideramos aqui que a fita codificadora é aquela que contém a sequência direta do RNA a ser transcrito e que, no caso dos mRNAs, contém os códons.

A fita dupla é desnaturada na região do DNA onde ocorre a transcrição, e o complexo RNAP mantém esta estrutura estável, ao passo que o RNA sintetizado permanece parcialmente pareado com a fita-molde do DNA e a parte já sintetizada deixa o complexo de transcrição. A essa estrutura atribui-se a denominação **bolha de transcrição**. O complexo movimenta-se ao longo da molécula de DNA para expor e copiar novas regiões do gene.

A transcrição começa pela **iniciação abortiva**, que é um estado estático formado por um **complexo aberto** no promotor, onde a RNAP ainda está ligada ao sítio de início da transcrição. A fase produtiva da transcrição é a fase de alongamento da cadeia de RNA. As RNAPs adicionam processivamente na extremidade 3' um a um os ribonucleotídeos complementares ao molde de DNA. O complexo de alongamento é dissociado no final da síntese em resposta a sinais específicos de terminação. A RNAP livre do complexo pode, com auxílio de fatores de iniciação, ligar-se novamente a um promotor e reiniciar o ciclo de transcrição.

**Figura 10.1****RNA-polimerase, funções básicas e estrutura. (A)**

Esquema de uma bolha de transcrição na fase de alongamento. A direção de síntese do RNA e os contatos principais da bacRNAP com os ácidos nucleicos estão indicados, e a RNAP está representada em **(B)** Representação esquemática do modelo tridimensional da bacRNAP, enzima principal com suas subunidades. Dados gerados pela cristalografia por raios X. **(C)** Algumas das principais atividades catalíticas das RNA-polimerases. A atividade de polimerase é responsável pela síntese de RNA. As duas outras atividades (*backtracking*) são importantes na correção de erros e na reativação de complexos de transcrição estagnados. NT (nucleotídeos) com seus grupamentos fosfato(P).

A seguir serão descritos os complexos de transcrição, as proteínas que as compõem e as suas características estruturais.

2. Proteínas envolvidas

2.1 RNA-polimerases

Todas as RNAPs celulares são complexos proteicos com subunidades múltiplas e partilham um complexo básico comum entre todos os domínios da vida. Embora existam

diferenças no número de subunidades, existem muitas semelhanças na estrutura e na sequência de aminoácidos das proteínas que compõem as RNAPs, sendo aceito que todas tiveram um ancestral comum (**Tabela 10.1**). Existe também uma grande semelhança na conformação espacial das RNAPs.

Os **bacteriófagos** (o modelo é o bacteriófago T7, um vírus da bactéria *Escherichia coli*) apresentam uma única RNAP, capaz de desempenhar todas as funções básicas das demais RNAPs celulares. Os genomas das mitocôndrias e dos cloroplastos são transcritos por RNAPs semelhantes à RNAP dos fagos. Nas bactérias, nas arque-

Tabela 10.1 RNA-polimerases: composição e similaridades nos três domínios da vida*

Bactéria	Arquea	Eucariotos			Função
		RNAPII	RNAPI	RNAPIII	
β'	A'	Rpb1	A190	C160	Core catalítico
	A''		Rpb3	A135	
β	B'	Rpb2		AC40	
α	B''		Rpb10	Rpb10	
	D	Rpb11	AC19	Ac19	
α	N	Rpb12	Rpb12	Rpb12	
ω	L	Rpb5	Rpb5	Rpb5	Funções auxiliares especializadas
	P	Rpb6	Rpb6	Rpb6	
F	H	Rpb4	A14	C17	
	E	Rpb7	A43	C25	
TFS/X	K	Rpb9	A12	C11	
		Rpb8	Rpb8	Rpb8	
		A49			
		A34.5			
			C82		
			C53		
			C37		
			C31		

*As subunidades relacionadas e com similaridade de sequência ou estrutura são representadas pela mesma cor ou marca.

as e nos eucariotos as RNAPs são compostas por complexos proteicos, contendo de 6 a até 17 proteínas. A Tabela 10.1 apresenta uma comparação entre as proteínas componentes, as homologias e as funções dos complexos de RNAPs nos diferentes domínios da vida.

Para a fluência da leitura serão adotadas as seguintes abreviaturas: **bacRNAP**, para polimerases bacterianas, **arqRNAP**, para as polimerases de arqueas e **euRNAP** para as polimerases de eucariotos. Esses complexos proteicos são capazes de realizar a **transcrição basal**, ou seja, sintetizar RNA a partir do promotor de um DNA-molde. Em geral, os complexos de transcrição na célula apresentam outras proteínas que não fazem parte do complexo básico, mas participam da transcrição.

Independentemente da sua complexidade, as RNAPs têm múltiplas atividades: (1) reconhecem e se ligam às sequências específicas de DNA; (2) separam a hélice dupla do DNA, expondo a sequência de nucleotídeos a ser copiada; (3) mantêm as fitas de DNA separadas na região de síntese do RNA; (4) mantêm estável o híbrido DNA-RNA na região de síntese; (5) renaturam o DNA na região posterior à da síntese; e (6) sozinhas ou com o auxílio de proteínas específicas, terminam a síntese do RNA.

A reação catalisada pelas RNAPs é mecanisticamente idêntica à reação catalisada pelas DNA polimerases. A

reação ocorre entre o grupamento 3'-OH de um ribonucleotídeo e o grupamento fosfato do carbono 5' do ribonucleosídeo trifosfatado a ser incorporado (Figura 10.1C). A reação de adição, portanto, ocorre no sentido 5'→3' e a fita de DNA molde (copiada) é aquela de sentido 3'→5'. O ribonucleotídeo, para ser incorporado, deve ter como característica a complementariedade ao desoxirribonucleotídeo do DNA-molde. As RNAPs não necessitam de um **iniciador** para começar a síntese do RNA e isso contrasta com as DNA-polimerases. Convém ressaltar que a síntese do iniciador de RNA, utilizado para iniciar cada ciclo de replicação do DNA, pode ser realizada pela RNAP (DNA primase). Além da atividade de polimerase as RNAPs catalisam outras reações importantes (Figura 10.1C), como a clivagem do RNA sintetizado (**backtracking**) que permite a correção de erros e a recuperação de complexos de transcrição pausados, como será visto adiante.

2.1.1 RNAPs de procariotos: bactérias e arqueas

Entre as RNAPs com subunidades múltiplas, a bacteriana é a mais simples, sendo composta por cinco proteínas: duas subunidades que formam o sítio ativo do complexo (subunidades beta, β e beta', β'), um homodímero alfa (α_I, α_{II}) e uma subunidade omega (ω), que estabiliza o

complexo (Tabela 10.1). Essas cinco proteínas formam o complexo $\alpha_2\beta\beta'\omega$ que é suficiente para sintetizar RNA a partir de um molde de DNA na presença dos ribonucleotídeos (ATP, UTP, CTP e GTP) e em condições apropriadas de íons e pH. O complexo $\alpha_2\beta\beta'\omega$ é definido como o complexo central (**core**) das bacRNAPs e nos modelos cristalográficos apresenta o formato de “pinça de caranguejo” (Figura 10.1B), também observado nas arqRNAPs e nas euRNAPII.

Este complexo central da bacRNAP não apresenta uma característica fundamental que é a capacidade de reconhecer o promotor e iniciar a transcrição. Para o início da síntese de RNA, em bactérias, é necessário que outra proteína (fator sigma, σ), seja adicionada ao complexo, formando $\alpha_2\beta\beta'\omega\sigma$. Esse complexo plenamente funcional é denominado **holoenzima**, tem cerca de 400 kDa e foi muito estudado do ponto de vista estrutural nas bactérias *E. coli* e *Thermus thermophilus*. A maior parte da massa molecular do complexo $\alpha_2\beta\beta'\omega\sigma$ é formada pelas subunidades β e β' . As subunidades α , apesar de idênticas, possuem localização diferente no complexo bacRNAP e, por isso, são diferenciadas em α_I e α_{II} . O dímero de subunidades α forma a estrutura básica para as subunidades catalíticas β e β' serem montadas e a subunidade ω auxilia a ligação da subunidade β' ao subcomplexo $\alpha_I - \alpha_{II} - \beta$.

Nos **complexos de iniciação** da transcrição, aminoácidos específicos do fator σ estão ligados ao DNA, conferindo ao complexo RNAP a capacidade de reconhecer e ligar-se ao promotor. Em *E. coli* e outras bactérias, fatores σ diferentes são responsáveis pela ligação da RNAP aos diversos grupos de promotores com sequências características. A maioria dos genes de *E. coli* têm promotores que são reconhecidos pelo fator σ^{70} . O número 70 refere-se à massa molecular da proteína σ (70 kDa). Outros genes, por exemplo, aqueles envolvidos na resposta ao choque térmico, utilizam o fator σ^{32} . Assim, a célula tem fatores σ específicos para respostas metabólicas rápidas. Quando a bactéria é transferida para um meio de temperatura elevada, por exemplo, deve sintetizar as proteínas de choque térmico que a protegem do aumento da temperatura com rapidez. Várias bactérias têm fatores σ alternativos e, portanto formam diferentes RNAPs holoenzimas. Em *E. coli* existem sete fatores σ e cada um deles é específico para um tipo de promotor. Em muitas bactérias existem diversos fatores σ , podendo chegar a mais de 60 fatores σ diferentes, como é o caso de *Streptomyces coelicolor*.

Algumas das funções de cada proteína no complexo bacRNAP holoenzima são conhecidas. As subunidades α atuam na montagem do complexo multiproteico, ligam proteínas regulatórias e apresentam os **domínios** carbóxi-terminal (α -CTD, de *carboxy terminal domain*) e aminoterminal (α -NTD, de *aminoterminal domain*). A região **CTD** participa do complexo reconhecendo e ligando-se a sequências de DNA na região anterior (*upstream* ou **a montante**) do promotor. A região **NTD** das subunidades α ligam-se a outras subunidades do complexo bacRNAP. A subunidade β' possui a função de ligar-se ao

DNA-molde e parece não ter especificidade por sequência. A subunidade β contém a principal atividade catalítica da bacRNAP, sendo responsável pela reação de síntese do RNA. As subunidades β e β' formam as “pinças” da estrutura da bacRNAP. Durante algum tempo, a subunidade ω não era considerada como integrante necessária do complexo bacRNAP, entretanto, atualmente é reconhecida como fundamental para restaurar a atividade do complexo bacRNAP quando desnaturado. A subunidade ω parece ser fundamental para a montagem do complexo bacRNAP *in vivo* e, possui a função de manter a estrutura da subunidade catalítica β . Com exceção da subunidade σ , todos os demais componentes da bacRNAP tem homologia com componentes das RNAPs de arqueas e eucariotos. As funções de alguns componentes flexíveis do complexo RNAP estão apresentadas no **Material Complementar Online 10.1**.

As arqueas possuem a arqRNAP organizada de forma semelhante às RNAPs de eucariotos (euRNAPs), em especial a euRNAPII (Tabela 10.1). Nas arqueas, assim como nas bactérias, existe um único complexo RNAP que sintetiza todo o RNA celular, sendo o complexo principal semelhante à bacRNAP, mas com subunidades adicionais presentes. Há dois tipos principais de arqRNAP: um, composto por 11 subunidades (subunidades BA'A"DE'FLHNKP, em *Pyrococcus*) e outro, por 12 subunidades (subunidades A'B'B"A"DE'FLHNPK, em arqueas metanogênicas). As subunidades das arqRNAPs apresentam homologia com as subunidades das bacRNAPs: A e A' com β' , B com β e D e L com α . A estrutura geral das arqRNAPs também tem a forma de “pinças de caranguejo”.

2.1.2 RNAPs de eucariotos

Em eucariotos existem três classes de euRNAPs nucleares: euRNAPI, euRNAPII e euRNAPIII, que sintetizam diferentes **classes de RNA**. Em plantas (angiospermas), existem mais duas classes de RNAPs: a RNAPIV e a RNAPV. Além das euRNAPs nucleares, as organelas, as mitocôndrias e os plastídeos, possuem RNAPs que transcrevem genes dos **genomas das organelas** (Tabela 10.2). As RNAPs das organelas são compostas por apenas uma proteína e se assemelham as RNAPs de fagos. Nesta seção serão analisadas as três euRNAPs nucleares.

Cada euRNAP é responsável pela transcrição de classes específicas de RNA. A RNAPI, localizada no nucléolo, transcreve os genes para o precursor dos rRNAs, que quando processado produz as subunidades 28S, 5.8S e 18S, que irão formar os ribossomos (Capítulo 11). A RNAPIII transcreve os genes dos tRNAs do rRNA 5S e de uma variedade de RNAs denominados **RNAs pequenos** (snRNAs, snoRNAs) que, em geral, se complexam com proteínas formando as ribonucleoproteínas. Os RNAs pequenos estão envolvidos em uma variedade de processos celulares fundamentais, como no processamento de precursores de mRNAs e no direcionamento de proteínas para o retículo endoplasmático (ver a descrição de algumas classes de RNAs no Capítulo 2).

Tabela 10.2 RNA-polimerases em eucariotos

Polimerase	Localização	Moléculas/célula	Produtos	Atividade
RNAP I	Nucléolo	40.000	35-45S pré-rRNA (28S, 18S e 5S)	50-70%
RNAP II	Nucleoplasma	40.000	hnRNA e snRNAs (U1, U2, U4, U5)	20-40%
RNAP III	Nucleoplasma	20.000	5S rRNA tRNA, snRNA U6, 7S RNA, outros snRNAs	10%
RNAPIV	Plantas (angiospermas)	–	–	–
RNAPV	Plantas (angiospermas)	–	–	–
RNAP de mitocôndria	Mitocôndria	–	RNAs Mitocondriais	–
RNAP Cloroplasto	Cloroplasto	–	RNAs do Cloroplasto	–

As euRNAPs são mais complexas em relação às arqRNAPs e às bacRNAPs (Tabela 10.1). Nas euRNAPs existem proteínas homólogas a todas as 11 subunidades das arqueas, com exceção da Rpo13, e cada uma das euRNAPs tem um homólogo do **fator geral de transcrição** TFS/M (TF, de *transcription factor*, descritos a seguir) das arqueas. Nas euRNAPs esses fatores homólogos são subunidades permanentes do complexo. A euRNAPII possui 12 subunidades, sendo 11 homólogas às subunidades de arqRNAP (Rbp1 a 8 e Rbp10 a 12) e uma (Rbp9) homóloga ao TFS/M de arqueas. Cada uma das 12 subunidades da euRNAPII compartilha homologia com subunidades da euRNAPI e euRNAPIII, ou é codificada por um **gene parálogo** na euRNAPI ou na euRNAPIII. As euRNAPI (com mais duas subunidades, totalizando 14) e euRNAPIII (com mais cinco subunidades, 17 no total) não apresentam homologia com nenhuma outra RNAP. As euRNAPs também apresentam uma conformação geral em forma de “pinças de caranguejo”.

A euRNAPII é a mais estudada em termos de estrutura. Apresenta uma característica importante que a diferencia das euRNAPI e euRNAPIII, a extremidade C-terminal da sua subunidade maior (Rpb1). Essa região, chamada de domínio carbóxi-terminal, CTD, é fundamental para a funcionalidade da euRNAPII. A região CTD é composta por repetições de sete aminoácidos (tirosina-serina-prolina-treonina-serina-prolina-serina) que estão ligadas de forma flexível com o restante da subunidade. O número de repetições deste heptapeptídeo na CTD varia entre os organismos, por exemplo, a RNAPII de leveduras apresenta 26 ou mais repetições e a de seres humanos apresenta 52 repetições. Essas repetições são fundamentais para a funcionalidade das euRNAPII, e as mutações que reduzem o número de repetições abaixo de um determinado limite, são letais. Os aminoácidos serina e treonina da CTD podem ser fosforilados para iniciar a fase de alongamento da transcrição por euRNAPII. As prolinas podem sofrer **isomerização** e alterar a conformação da CTD.

As euRNAPs I e III contêm subunidades que não apresentam homologia com as subunidades de euRNAPII. Entretanto, duas subunidades das euRNAPs I e III apresentam relação estrutural e funcional com o fator de transcrição TFIIF, que participa da transcrição por euRNAPII. É possível que essas subunidades repre-

sentem TFs que foram acoplados de modo definitivo aos complexos euRNAPs I e III durante a evolução.

2.2 Outras proteínas

2.2.1 Topoisomerases

O contexto de **superenrolamento** (Capítulo 2) é de fundamental importância em todas as reações que ocorrem no DNA, isso não é diferente para a transcrição. Tanto nos sistemas procarióticos como nos eucarióticos a transcrição é iniciada com mais facilidade e eficiência quando o molde está superenrolado. *In vivo* a eficiência dos promotores e, em alguns casos, o controle da expressão é exercido por superenrolamento. Como visto anteriormente (Capítulo 2), o superenrolamento altera a geometria do pareamento entre as fitas do DNA e pode expor as bases nitrogenadas. Como no caso dos promotores, o complexo de transcrição deve reconhecer sequências específicas e desnaturar a hélice dupla, assim, pode-se compreender que as alterações provocadas no superenrolamento estão diretamente relacionadas a esses processos. Portanto, as **topoisomerases** (enzimas que atuam no controle do superenrolamento) estão também diretamente relacionadas à transcrição.

No momento da abertura da hélice dupla e durante seu deslocamento no DNA as RNAPs enrolam e desenrolam o DNA-molde, gerando o superenrolamento positivo, a jusante da região de síntese e o superenrolamento negativo, quando as fitas de DNA são renaturadas a montante. Portanto, as diferentes topoisomerases, tanto em procariotos como em eucariotos, participam da transcrição estabilizando as estruturas e refazendo o superenrolamento original após a transcrição. A reconstituição do superenrolamento original é fundamental, pois após a transcrição o DNA deve retornar à sua condição estrutural original.

2.2.2 Fatores de transcrição

O termo “fator de transcrição” (TF, de *transcription factor*) é utilizado na literatura em vários contextos e em algumas situações com definições diferentes. Alguns autores consideram todas as moléculas que atuam sobre RNA, DNA ou RNAPs em todos os estágios da transcrição, como fatores de transcrição. Esses autores consideram que os fatores de transcrição possam ser:

proteínas, peptídeos, RNAs não codificadores, aminoácidos e vitaminas, entre outros. Essa é uma definição ampla de qualquer molécula que possa regular, ativar, inativar ou modular a expressão gênica.

Outros autores definem os fatores de transcrição, como proteínas que reconhecem e ligam-se à sequências de DNA para ativar ou inibir a expressão de algum gene em particular. Neste caso, a definição é muito restrita e não considera interações entre as proteínas na regulação da expressão gênica. As divergências de definição se deve às diferenças de complexidade entre os modelos procarionto e eucarioto. Neste capítulo, será adotada, para os fatores de transcrição, uma definição em relação à etapa específica da expressão gênica considerada. Na transcrição os TFs são proteínas que se ligam a sequências específicas do DNA e que sozinhas, ou por interação com outras proteínas, promovem ou bloqueiam a ligação das RNAPs nas regiões promotoras dos genes.

Fatores basais de transcrição são aquelas proteínas necessárias para iniciar a transcrição e que não fazem parte das RNAPs (complexo de proteínas com atividade na síntese de RNA). Na ausência dos fatores basais o complexo RNAP não pode iniciar a transcrição a partir de um promotor. Nessa definição enquadra-se o fator sigma nas bacRNAPs, por exemplo. Em relação às euRNAPs, na categoria de TF enquadram-se todas as proteínas que, para cada classe de gene, são essenciais para a formação dos **complexos de pré-iniciação** (PICs, de *preinitiation complex*). São exemplos, os fatores basais de transcrição SL1 para euRNAPI, TFIID para euRNAPII e TFIIB para euRNAPIII. O termo fatores gerais de transcrição é inadequado, pois alguns fatores de transcrição basais estão ausentes em determinados PICs. A **proteína TBP** (de *TATA-binding protein*), que participa da transcrição das três classes de genes em eucariotos, é chamada de fator universal de transcrição.

Existem proteínas que interagem com os complexos de RNAP nas etapas de alongamento e de terminação da transcrição que, em alguns casos, são denominados fatores de transcrição, embora o mais adequado fosse chamá-los de fatores de alongamento da transcrição (por exemplo, a proteína NusA) e fatores de terminação (por exemplo, a proteína Rho, ρ). Há ainda fatores de transcrição proteicos, denominados fatores de transcrição específicos, que atuam em genes ou grupos de genes específicos ou são funcionais apenas em tecidos específicos. Existem outras proteínas envolvidas na regulação da transcrição como coativadores, proteínas de remodelamento da cromatina, acetilases e deacetilases de histonas, quinases e metilases, que não são classificadas como fatores de transcrição.

Grande parte dos fatores de transcrição conhecidos apresentam um dos seguintes motivos estruturais para ligação ao DNA: hélice-volta-hélice (*helix-turn-helix*) / homeodomínio; dedo de zinco (*zinc finger*); zíper de leucina (*leucine zipper*); hélice-alça-hélice (*helix-loop-helix*); e hélice “alada” (*winged-helix*). As características gerais desses motivos são apresentadas no Capítulo 14 e no **Material Complementar Online 10.2**.

3. Promotores

3.1 Procariotos

As sequências específicas na molécula de DNA, que determinam o local de formação dos complexos e iniciam a transcrição, são denominadas **promotores** (Figura 10.2). O primeiro nucleotídeo da sequência do DNA, copiado no RNA, é denominado **sítio de início da transcrição** e demarcado como +1 (IT) na molécula de DNA. Os desoxirribonucleotídeos localizados antes do sítio de início da transcrição (direção 5' ou a montante do sítio +1) recebem sinal negativo e números crescentes. Os desoxirribonucleotídeos, após o sítio +1 em direção ao terminador (direção 3' ou a jusante do sítio +1), recebem números crescentes positivos (Figura 10.2A).

Existem diferenças básicas e fundamentais na definição de promotor em procariotos e eucariotos. Nos procariotos o promotor é o conjunto de sequências do DNA onde o complexo RNAP se liga para iniciar a transcrição, ao passo que em eucariotos o promotor é o conjunto de sequências do DNA onde os fatores de transcrição gerais (TF) se ligam para posicionar a RNAP para o início da transcrição. Em eucariotos a RNAP não se liga diretamente as sequências de DNA do promotor, isso só ocorre após a ligação de alguns TFs ao promotor.

Nos promotores procarióticos a RNAP holoenzima ($\alpha_1\alpha_{II}\beta\beta'\omega\sigma$) reconhece sequências de DNA específicas e liga-se para o início da transcrição. Embora a RNAP central ($\alpha_1\alpha_{II}\beta\beta'\omega$) possa ligar-se aos moldes de fita dupla de DNA e sintetizar RNA, somente após a formação do complexo RNAP holoenzima, com a participação do fator σ , a iniciação ocorre a partir dos promotores. As sequências dos promotores demarcam o local exato onde, no genoma, estão as regiões que são transcritas. A princípio a localização dos promotores de *E. coli* foi realizada com base em experimentos de mutação e indicavam que as sequências de DNA dos promotores eram mapeadas na extremidade 5' dos genes analisados (a montante do códon de início da tradução). Assim, os primeiros genes começaram a ter suas sequências de nucleotídeos determinadas e comparadas entre si.

Uma hipótese óbvia era a de que deveria haver muita semelhança entre as sequências dos promotores dos diferentes genes em *E. coli*, pois existe apenas uma RNAP nessa bactéria, e este complexo enzimático é capaz de reconhecer, em princípio, todos os promotores dessa bactéria. A partir disso, as sequências de nucleotídeos dos genes começaram a ser comparadas, principalmente a região 5' a montante do sítio de início da síntese de proteínas, na tentativa de encontrar sequências comuns (sequências consenso) que pudessem ser reconhecidas pelo complexo RNAP, como promotores. Com essas análises e, atualmente, com o sequenciamento completo do genoma desta e de outras bactérias, chegou-se às seguintes regras para promotores de *E. coli*, aplicáveis, com adequações, às bactérias em geral:

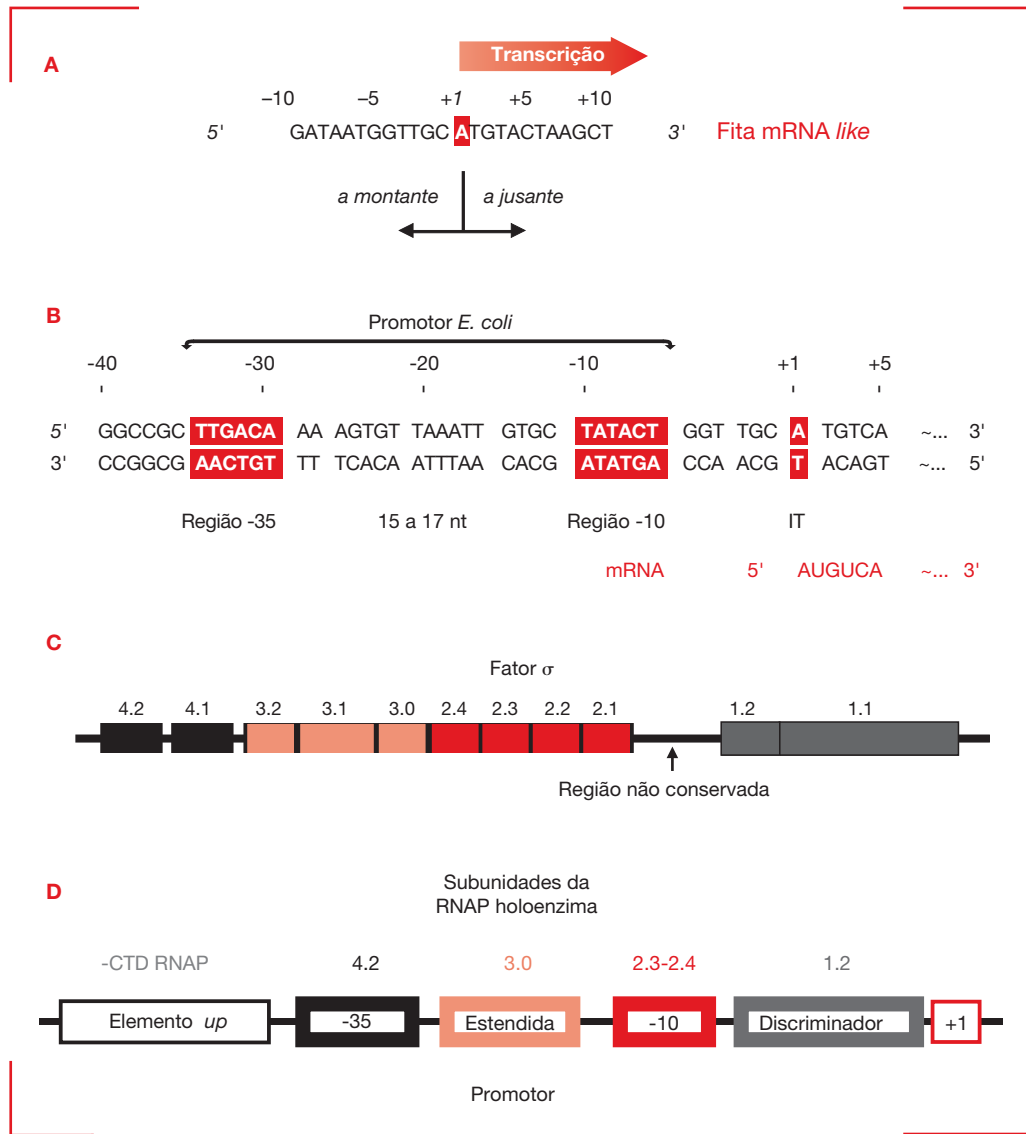


Figura 10.2

Estrutura geral de promotores de bactérias. (A) Nomenclatura e convenções utilizadas na transcrição. (B) Elementos que compõem um promotor genérico de *E. coli*. (C) Domínios do fator σ de *E. coli*. Representação gráfica da sequência de aminoácidos de fatores σ . A escala de tamanho é aproximada e representa os aminoácidos. As cores representam as quatro regiões conservadas e os números as subregiões. (D) Elementos de um promotor genérico de *E. coli* e interações com as subunidades da bacRNAP no início da transcrição. As regiões de σ que interagem com sequências específicas e conservadas dos diferentes tipos de promotor, estão indicadas α -CTD: extremidade carboxi-terminal da subunidade α da bacRNAP. Estendida: sequência a montante do elemento -10 presente em alguns promotores de *E. coli*. Discriminador: presente entre o elemento -10 e o sítio de início da transcrição +1.

- o primeiro nucleotídeo a ser transcrito (+1, IT) é uma purina (A ou G, em mais de 90% dos genes de *E. coli*), normalmente no centro da sequência
 $-1 +1 +2$
 $5'-C \mathbf{A} T-3'$
- duas sequências são conservadas na maioria dos genes comparados: uma localizada na região -10 e outra na região -35
- as regiões -10 e -35 são separadas por 15 a 17 nucleotídeos, o que corresponde a uma volta da hélice do DNA.

A região -10 recebe, também, as denominações **TATA box** ou *Pribnow box*. As regiões -10 e -35 definem os promotores de *E. coli* para a holoenzima RNAP, quando o fator σ é σ^{70} . Os milhares de promotores nos genomas de cada bactéria têm eficiências diferentes. As sequências contidas nos dois hexâmeros principais do promotor (-10 e -35 , ou equivalentes em outras bactérias) são os determinantes primários da eficiência dos promotores. Alterações nessas sequências, em alguns promotores, afetam a interação com duas das subunidades da RNAP holoenzima, a subunidade α e o fator σ . Essas alterações, ainda, geram a grande variabilidade na eficiência dos diferentes promotores, associados a um mesmo fator sigma.

Existem também, dois outros modelos de promotores em *E. coli*. Um deles, em que além dos elementos -10 e -35 ocorre uma sequência localizada a montante de -35 denominada elemento *up*, esses promotores são característicos de genes de rRNAs. No outro modelo não existe o elemento -35 e ocorre uma extensão a montante do elemento -10 , denominada estendida, composta por uma sequência que faz contatos com a RNAP e compensa a ausência do elemento -35 . Este tipo de promotor está presente nos genes que participam da utilização de galactose. Em alguns promotores pode ocorrer ainda uma sequência rica em GC entre os elementos -10 e o sítio $+1$, nos promotores de genes de rRNA e tRNA, que é denominada de discriminador e interage com o domínio 1.2 fator σ (Figura 10.2.C).

Além disso, a atividade dos promotores pode ser alterada por proteínas ativadoras e repressoras, que se ligam a sequências específicas na região dos promotores (Capítulo 13). Os promotores de alguns genes de *E. coli*, como *lac* e *ara*, diferem um pouco da sequência consenso, principalmente na região -10 . Esses genes são pouco transcritos (promotores fracos). Outros promotores têm sequências muito parecidas com a consenso, e os genes são transcritos com grande eficiência (promotores fortes), como é o caso do óperon *rrn*, que codifica RNA ribossômico em *E. coli*. Assim, a eficiência dos promotores pode variar em diversas ordens de magnitude. Por exemplo, em um mesmo espaço de tempo, um determinado promotor pode produzir apenas uma cópia de mRNA por célula e outro pode gerar milhares de cópias. Algumas das diferenças na utilização e na eficiência dos promotores estão relacionadas às RNAP holoenzimas associadas a diferentes fatores σ em decorrência do estado fisiológico da célula.

Outra variável é que o número de complexos principais de RNAP na célula é limitado, o que resulta em uma competição entre os milhares de promotores no genoma para a ligação com a RNAP holoenzima. Existem diferentes afinidades entre os vários fatores σ e a RNAP principal. Além disso, existe uma variação muito grande na quantidade dos diversos fatores σ , que é decorrente de mudanças nas condições nutricionais e ambientais as quais a célula está submetida. O balanço entre os fatores σ , a disponibilidade de RNAP principal para a formação da RNAP holoenzima e a disponibili-

dade de promotores são peças importantes no **controle da expressão** dos genes (Capítulo 13).

3.2 Eucariotos

Nos organismos eucarióticos a definição de sequências comuns aos diferentes promotores é mais complexa devido a maior diversidade de genes; a grande complexidade dos sistemas, que atribui uma maior complexidade no processo de controle da expressão (tecidos, órgãos etc.); a diversidade dos organismos eucarióticos (mamíferos, plantas, insetos, fungos, por exemplo); e a existência de pelo menos três euRNAPs, que atuam na transcrição de classes diferentes de RNA. Assim, a definição dos promotores depende de um acúmulo de informações nos diferentes sistemas e de análises gerais de genomas completamente sequenciados.

Em eucariotos as euRNAPs não se ligam diretamente à sequência do promotor, que primeiro é reconhecida pelos fatores de transcrição (TFs) que dirigem a ligação das euRNAPs para o início da transcrição. Em geral, os promotores de eucariotos têm uma sequência principal capaz de produzir a transcrição em nível basal, que é denominada promotor principal (*core promoter*). Os promotores principais “dirigem” o início da transcrição.

As análises de genomas de mamíferos mostraram dois tipos de sítios de iniciação de transcrição e levaram a uma redefinição da estrutura dos promotores, propondo dois tipos de promotores principais para as euRNAPII: o promotor focalizado, no qual a transcrição inicia em um único nucleotídeo ou em uma região bastante restrita com grupos de vários nucleotídeos; e o promotor disperso, no qual a iniciação da transcrição ocorre em múltiplos sítios em uma região ampla de 50 a 100 nucleotídeos. Os promotores focalizados são mais antigos do ponto de vista evolutivo e ocorrem em todos os organismos, sendo o modo predominante ou único, de início de transcrição em organismos menos complexos. Em vertebrados, mais de 70% dos genes apresentam promotor do tipo disperso e são encontrados nas “ilhas CpG”.

Estudos de análise ampla sobre promotores em genomas propuseram que os promotores do tipo focalizado ocorrem em genes regulados, ao passo que os promotores do tipo disperso ocorrem em genes constitutivos. Essa diferença pode ter como base a maior facilidade em regular a atividade de promotores com início de transcrição único. Da mesma forma, variações dos níveis de expressão em genes constitutivos seriam reduzidas pela existência de sítios múltiplos de início de transcrição. Alguns promotores possuem as características dos dois tipos, focalizado e disperso, sendo compostos por vários sítios $+1$ dispersos, mas apresentando um, em particular, forte (mais frequente).

A análise dos genes transcritos pela euRNAPII mostrou que existem sequências atuando na regulação da expressão gênica, localizadas tanto na região 5' a montante do sítio de início da transcrição, como em outras

regiões do gene. As porções dos genes envolvidas na regulação transcricional são constituídas por duas regiões distintas, o promotor principal e as regiões regulatórias (**enhancer**, silenciador, ativador). O promotor principal foi definido como a sequência mínima capaz de iniciar a transcrição correta pela euRNAPII. Na maioria dos casos, o promotor principal é constituído por uma região que contém o sítio de início da transcrição (sítio +1 ou IT) e se estende a montante ou a jusante por cerca de 35 nucleotídeos. Em eucariotos também são utilizadas as siglas TSS (de *transcription start site*) e Inr (de *initiator*), que é a sequência consenso contendo o sítio +1 e as sequências flanqueadoras (**Figura 10.3**).

Um promotor principal contém a sequência mínima necessária para permitir a transcrição por RNAPII. Para efeito da formação dos PICs deve-se considerar as duas categorias de promotores principais, a com TATA *box* e a chamada de TATA *less*. Embora o elemento TATA *box* seja o mais conhecido, os promotores TATA *less* representam a maioria dos promotores principais dos genes da classe II e apresentam muita diversidade de organização.

Atualmente, com a disponibilidade de genomas sequenciados em sua totalidade, mostrou-se que existem pelo menos nove diferentes tipos de elementos reguladores nos promotores principais. Esses promotores podem conter diversos elementos, como TATA *box*, BRE, Inr, TEM, DPE e XCPE1, que especificam padrões de transcrição e respostas a *enhancers* diferentes (Figura 10.3). É importante salientar a grande diversidade dos promotores principais dos genes da classe II, e que muitos desses promotores não apresentam nenhum desses elementos, sendo, mesmo assim, funcionais.

Em geral, os **elementos canônicos** são necessários para a funcionalidade do promotor, incluindo a montagem do PIC e a sua regulação. Elementos diferentes podem ocorrer simultaneamente no mesmo promotor, mas determinadas combinações ocorrem com mais frequência do que outras. Entretanto, alguns promotores não apresentam nenhum elemento identificável.

Alguns genes cujo nível de expressão é baixo, como os genes que codificam enzimas do metabolismo intermediário, não apresentam TATA *box* nem Inr, mas uma região rica em GC, em um segmento de 100-200 pb loca-

lizado a montante do sítio de início de transcrição. Esses genes pertencem a um grupo que sempre é expresso (expressão constitutiva), pelo fato de serem constantemente requisitados pelas células (**genes housekeeping**).

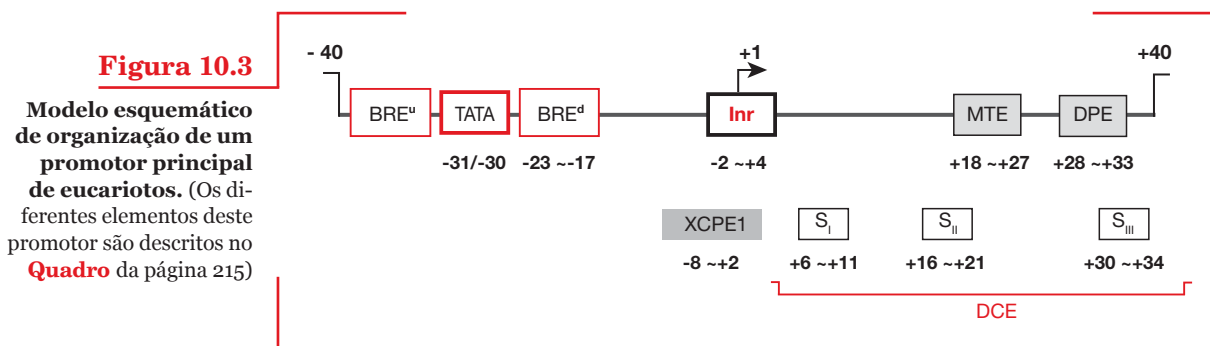
Existem ainda os elementos reguladores localizados próximos aos promotores (promotores proximais), situados a uma distância de 50 pb a algumas poucas centenas de pares de bases a montante do sítio de início de transcrição, e os elementos localizados a uma distância razoável (milhares de pares de bases), tanto a montante como a jusante do sítio de início de transcrição, denominados promotores distais ou reforçadores (*enhancers*).

4. Início da transcrição

4.1 Procariotos

A primeira etapa na transcrição em bactérias é o reconhecimento e a ligação do complexo RNAP holoenzima ao promotor. A energia necessária para esse processo não é obtida pela hidrólise de ATP, mas sim, provém dos rearranjos e interações entre as proteínas da RNAP e as sequências de DNA do promotor que geram energia livre. O complexo formado entre a RNAP e o promotor no momento da ligação é denominado **complexo binário fechado** (CBF), pois é composto por RNAP e DNA (hélice dupla). Em seguida, o complexo RNAP se alonga e cobre uma região de DNA desde a posição -55 até a posição +1 (TSS), que permanece em sua conformação de fita dupla com o CBF.

A partir deste ponto o complexo passa por alterações conformacionais importantes na RNAP e no promotor (denominadas isomerização) e a RNAP pode cobrir desde a região -35 até a região +12. Embora possam ocorrer algumas rupturas nas pontes de hidrogênio na região -10, o DNA ainda permanece na conformação de hélice dupla. Após novas alterações conformacionais as pontes de hidrogênio no DNA são rompidas entre -11 e +3 e alterações na hélice dupla ocorrem até a região +20. Quando as fitas de DNA são separadas é formado um **complexo** denominado **binário aberto** (CBA). Esses rearranjos sucessivos, que resultam na abertura das fitas do DNA, posicionam o **centro catalítico** da RNAP no sítio de início da trans-





Modelo esquemático de organização de um promotor principal de eucariotos

O **iniciador (Inr)** contém o sítio de início de transcrição TSS (+1), sendo o elemento mais comum encontrado em promotores principais. É definido pela sequência consenso YYANWYY, em seres humanos e TCAKTY, em *Drosophila*. Está posicionado entre -2 a +5 e é independente do TATA box. O TATA box e os elementos Inr são os únicos que sozinhos podem recrutar o PIC para iniciar a transcrição. Esse elemento é reconhecido por TFIID. O elemento TATA box apresenta a sequência consenso TATAWAWR e está posicionado entre -31 (ou -30) a -24 em relação ao Inr (+1). O TATA box foi o primeiro elemento dos promotores principais a ser descoberto, e é um dos sítios de ligação dos fatores de transcrição mais estudados. A sequência consenso TATAA liga-se ao fator geral de transcrição TBP (de *TATA binding protein*), que faz parte do PIC e participa da transcrição pelas três euRNAPs. TBP e o elemento TATA box são conservados desde Arquea até seres humanos. TATA box está associado a promotores tecido-específicos fortes e, com frequência, ocorrem conjuntamente com sequências semelhantes ao elemento iniciador (*Inr-like*). TATA box está presente em apenas 10 a 15% dos promotores principais de mamíferos. Outros dois elementos, BREu e BREd, atuam em conjunto com o TATA box e tanto aumentam como diminuem os níveis de transcrição basal. Os **elementos BRE** (de *TFIIB recognition element* = elemento que reconhece TFIIB) são sequências de ligação de TFIIB, que podem ligar-se a montante (BREu) ou a jusante (BREd) do TATA box. BREu apresenta a sequência consenso SSRCGCC e está posicionado em -38 a -32 e BREd apresenta a sequência consenso RTDKKKK. Dependendo das sequências do promotor, BREu e BREd podem aumentar, ou diminuir o nível de transcrição. O **elemento DPE** (de *downstream promoter element* = elemento promotor a jusante) é um elemento promotor a jusante, importante para a atividade basal (mínima) de transcrição. Esse elemento tem como sequência consenso RGWYVT ou RGWYV, e é conservado de *Drosophila* até seres humanos, estando localizado nas posições +28 a +33 ou +32,

em promotores sem TATA box. Em geral, DPE atua cooperativamente com elementos Inr e acredita-se que tenha função semelhante a TATA box no direcionamento da localização do PIC próximo ao TSS. O espaçamento entre DPE e o elemento Inr é decisivo para a transcrição máxima. O **elemento MTE** (de *motif ten element* = elemento motivo dez) apresenta a sequência consenso CSARCSSAACGS e está posicionado em +18 a +27. Semelhante a DPE, o MTE funciona cooperativamente com Inr e são elementos de reconhecimento para TFIID. MTE pode atuar independentemente de TATA e DPE, entretanto, existe sinergia entre MTE e DPE, assim como entre TEM e o TATA box. Foi construído um super promotor principal contendo TATA, Inr, TEM e DPE, mais potente que o promotor principal, tanto *in vitro* como em cultura de células, mostrando que os níveis de expressão podem ser modulados pelo promotor principal. O **elemento DCE** (de *downstream core element* = elemento cor a jusante) apresenta três subelementos: SI, CTTC (+6 a +11); SII, CTGT (+16 a +21); e SIII, AGC (+30 a +34). O DCE frequentemente ocorre com o TATA box e parece ser distinto de DPE. O **elemento XCPE1** (de *X core promoter element 1* = elemento promotor principal X 1) tem a sequência consenso DSGYGGGRASM e está posicionado em -8 a +2. Esse elemento aparece em cerca de 1% dos promotores principais de genes humanos, sendo a maioria sem TATA box, e atua em conjunto com ativadores sequência-específicos. O **elemento XCPE2** (de *X core promoter element 2* = elemento promotor núcleo X 2) apresenta a sequência consenso VCYCRTTRCMY e está posicionado em -9 ~+2. Descoberto no gene X do vírus da hepatite B, mostrou-se também presente em regiões promotoras de genes humanos e parece possibilitar o início da transcrição a partir de um dos múltiplos TSSs em promotores sem TATA. As **ilhas CpG** correspondem a segmentos genômicos, nos quais os dinucleotídeos CG estão super-representados. Tem-se mostrado que cerca de 70% dos promotores de genes humanos estão associados à ilhas CpG. As ilhas CpG são, com frequência, associadas a promotores de genes de expressão ubíqua com promotores principais do tipo disperso. Os nucleotídeos degenerados são designados de acordo com os códigos da *International Union of Pure and Applied Chemistry* (IUPAC). **D= A, T, ou G; K= G ou T; S= C ou G; Y= C ou T; R= A ou G; M= C ou A; W= T ou A; V= A, C, ou G.**

crição (+1). Quando a síntese de RNA é iniciada, o **complexo** formado chama-se **terciário aberto** (CTA), pois agora uma terceira molécula, o RNA, está presente.

No complexo RNAP holoenzima ($\alpha_1\alpha_{II}\beta\beta'\omega\sigma$) ocorre a formação de dois canais moleculares. Entre as subunidades β e β' existe uma estrutura em α -hélice que forma uma ponte entre essas subunidades. Esta ponte está localizada na região onde fica a “articulação” das “pinças”, formadas pelas subunidades β e β' . Nessa região são formados dois canais: em um deles, o canal maior, o DNA a jusante do promotor é alojado, sendo também o local onde será aberta a bolha de transcrição que abrange entre 12 a 14 nucleotídeos; o outro canal, que é menor e denominado canal secundário, é o local por onde os ribonucleosídeos trifostatados (NTPs), que serão incorporados ao RNA nascente, entram no complexo, próximo ao centro catalítico da RNAP que contém íon Mg^{+2} . Esse canal secundário tem 12 Å de tamanho e comporta exatamente a passagem dos ribonucleotídeos até o centro catalítico da RNAP, onde ocorre o pareamento com o DNA-molde e a sua incorporação ao RNA nascente.

O fator σ no complexo RNAP holoenzima está orientado com seu domínio C-terminal na região a montante do promotor e o domínio N-terminal a jusante. O fator σ^{70} de *E. coli* pode ser dividido em quatro domínios (σ_1 , σ_2 , σ_3 e σ_4 , que são ainda divididos em subregiões) que estão relacionados a sua função e localização durante a etapa de reconhecimento e ligação da holoenzima RNAP ao promotor (Figura 10.2C e D). Dessa forma, o fator σ assume uma conformação estendida em uma das faces da RNAP. O domínio α -CTD da subunidade α interage com a sequência do DNA na região -35 e a montante do promotor. O local exato onde α -CTD se liga ao DNA pode alterar a eficiência do promotor.

Uma alteração conformacional importante ocorre quando as fitas de DNA são abertas e a RNAP desenrola uma volta da hélice dupla na região +1, formando o CTA. Nesse momento as “pinças” da RNAP (subunidades β e β') se fecham sobre a estrutura do DNA que é seguida pelo início da incorporação dos ribonucleotídeos pelo centro catalítico.

O centro catalítico da RNAP é formado por resíduos de aspartato presentes em uma região muito conservada da subunidade β' . Esses resíduos coordenam os dois íons Mg^{+2} e estão envoltos por vários domínios das subunidades β e β' . A reação de adição de cada NTP envolve as seguintes etapas: entrada do ribonucleotídeo pelo canal secundário e sua ligação ao complexo; incorporação do ribonucleotídeo na cadeia nascente na extremidade 3' do RNA, pela formação de uma ligação fosfodiéster com a liberação de pirofosfato; e translocação do complexo de transcrição para a liberação do sítio de inserção para a entrada do próximo ribonucleotídeo a ser incorporado.

A chegada do ribonucleotídeo ao centro catalítico passa por várias etapas de interação com domínios da RNAP, responsáveis por reconhecer o ribonucleotídeo correto a ser adicionado na cadeia do RNA em relação à

fita-molde de DNA. Esse reconhecimento ocorre a partir de uma rede de interações que englobam todas as características químicas dos NTPs: as bases, o açúcar e os fosfatos. Essas interações monitoram o pareamento correto do NTP na base do DNA-molde.

Na etapa de formação do complexo de iniciação, a RNAP inicia um **ciclo abortivo de transcrição**, em que ocorre a síntese e a liberação de pequenos RNAs. Quando ocorre a síntese de RNAs com tamanho entre 9 e 11 nucleotídeos, o complexo RNAP escapa da região do promotor, formando o complexo de alongamento e dando início a próxima etapa da transcrição, que é o alongamento da cadeia de RNA que resultará na síntese do RNA.

O mecanismo molecular, pelo qual o centro catalítico da RNAP é translocado da iniciação abortiva para o complexo de alongamento e consequentemente escapa do promotor, foi elucidado em 2006. Até então muitos modelos foram propostos e testados, mas dois dados considerados contraditórios dificultavam a interpretação do mecanismo: (1) durante a iniciação abortiva o centro catalítico da RNAP se move sobre o DNA-molde para sintetizar os RNAs pequenos e (2) experimentos de localização da RNAP no promotor sugeriam que a parte anterior do complexo RNAP permanecia ligada a sequência do promotor e que, portanto, a RNAP não se movia em relação ao DNA-molde.

Experimentos realizados utilizando uma única molécula de RNAP com DNA-molde mostraram que o complexo RNAP não se move em relação ao molde de DNA durante a iniciação abortiva. O que ocorre é que a RNAP, mantendo-se fixa na região -35 e na região +12, se “encolhe”, puxando o DNA que está na região do sítio +1 para trás (no sentido do promotor). Esta ação, causa um estresse na estrutura da hélice dupla que, para manter a estabilidade, desenrola a hélice com ruptura das pontes de hidrogênio. Isso ocorre sucessivamente a cada par de bases cuja ponte de hidrogênio é rompida, sendo a hélice dupla desenrolada. As fitas de DNA abertas formam alças para se acomodar no complexo RNAP. Esse processo é repetido inúmeras vezes na iniciação abortiva liberando pequenos RNAs de ~8 nucleotídeos.

Quando uma das moléculas de RNA atinge entre 9 a 11 nucleotídeos em tamanho, o que significa um “encolhimento” de 7 a 9 pares de bases do DNA-molde, a RNAP altera a sua conformação, escapa do promotor e inicia a fase de alongamento da cadeia de RNA. Esse mecanismo de encolher e puxar (*scrunching*) é o atualmente aceito, e em 2009 foi detectada pela primeira vez em células a existência de produtos de RNA derivados de iniciação abortiva *in vivo*. Essa estrutura formada no sítio de início da transcrição é reconhecida por proteínas regulatórias que podem adicionar mais um nível de regulação ao início da transcrição.

Embora tenha existido controvérsia durante algum tempo, atualmente sabe-se que o fator σ é desligado do complexo RNAP holoenzima quando o complexo deixa o promotor e entra na fase de alongamento da transcrição.

Na maioria dos promotores de *E. coli* a interação entre o fator σ e a RNAP ocorre em etapas, assim como a sua dissociação.

O início da transcrição nas arqueas é semelhante ao que será visto para os eucariotos. Existe semelhança estrutural e evolucionária entre a arqRNAP e a euRNAPII. As arqueas também possuem um fator de transcrição TF (homólogo de TBP), um TFB (homólogo de TFIIB) e um TFE (homólogo de TFIIE). O reconhecimento e a ligação de arqTBP aos promotores ricos em AT promove a formação do complexo de pré-iniciação e recruta outros TFs ao promotor. Após a ligação de arqTBP e TFB ao promotor, a arqRNAP é recrutada. Um domínio de TFB interage com o centro catalítico da arqRNAP e estabiliza a fita-molde para o início da incorporação dos ribonucleotídeos para a síntese do RNA. TFE, ao ser recrutado para o complexo, promove a formação do complexo aberto no promotor e atua na sua estabilização, possibilitando o início da transcrição.

4.2 Eucariotos

O início da transcrição em eucariotos envolve mais etapas e um número maior de fatores, quando comparado com os procariotos. Como visto, em bactérias, uma única proteína (o fator σ) é necessária para a iniciação correta da transcrição pela bacRNAP. Nas arqueas existe apenas uma RNAP e somente dois TFs são necessários para o iní-

cio da transcrição, por essa razão, servem como um modelo simplificado para o estudo do início da transcrição.

Em eucariotos, o início da transcrição ocorre pela ligação dos fatores de transcrição (TFs) às sequências de DNA na região do promotor. O conjunto de TFs e a euRNAP nuclear (I, II ou III) formam o PIC, que é o complexo mínimo capaz de iniciar a transcrição *in vitro*. Essa transcrição basal é mínima e não regulada, pois para isso, outros TFs devem ser recrutados aos PICs. O número de proteínas que participam da transcrição *in vivo* é muito grande e variável, pois depende do promotor envolvido, do organismo e do tecido em questão. Essa descoberta dependeu da caracterização de inúmeros TFs e proteínas e somente foi demonstrada após a clonagem e a expressão dos TFs basais e da montagem de sistemas *in vitro*, onde, após a adição dos TFs minimamente necessários, o RNA é sintetizado por promotores também mínimos.

O início da transcrição promotor-específico por RNAPII, por exemplo, depende de cinco fatores basais de transcrição, compreendendo cerca de 30 polipeptídeos e formando com a RNAPII um complexo de cerca de 2.000 KDa, que deve ser montado a cada ciclo de transcrição. Cada uma das três RNAPs nucleares necessita de TFs característicos para transcrever os genes de classe I, II e III, forma PICs característicos e reconhece promotores específicos (**Figura 10.4**).

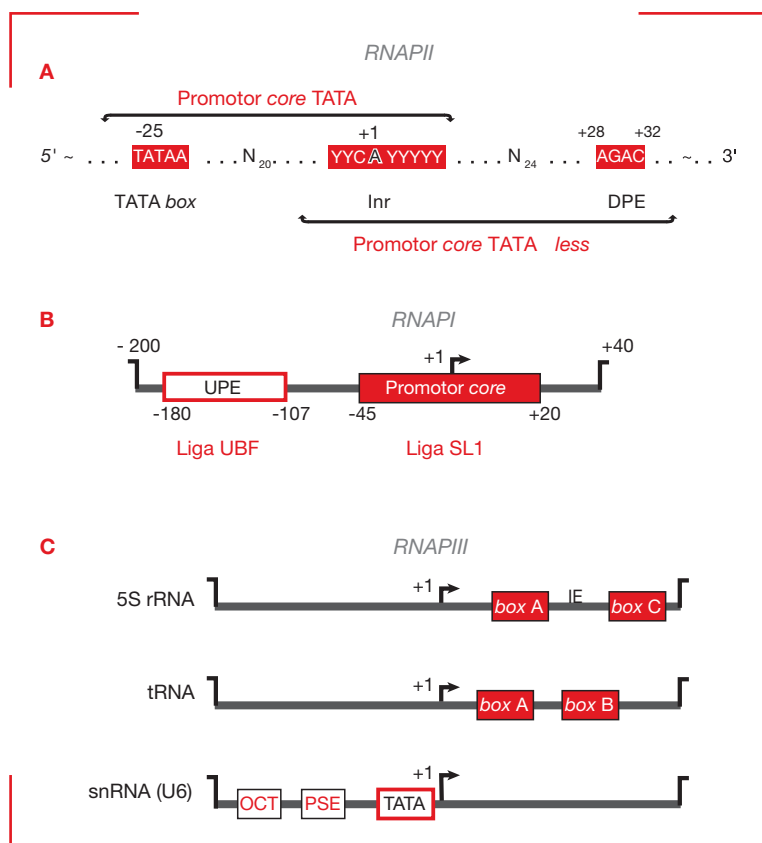


Figura 10.4

Estrutura geral de promotores das euRNAPs nucleares. (A) Esquema geral de promotores TATA e TATA less para euRNAPII. **(B)** Esquema geral de promotores para euRNAPI e os elementos de ligação dos complexos SL1 e UBS. **(C)** Esquema geral de promotores para euRNAPIII. Existem três classes de promotores para essa RNAP. Os promotores para 5S rRNA e tRNA são “internos”.

Existe uma ordem hierárquica de ligação dos TFs para a montagem dos PICs funcionais. Uma vez formados, esses complexos podem iniciar a transcrição mínima ou basal dos genes. Os fatores de transcrição eucarióticos podem ser classificados como TFs basais, necessários para a maioria dos genes de cada classe; TF universal, que é a proteína TBP, necessária para todas as classes de eRNAP nucleares; e TFs específicos. A transcrição mediada por RNAPII, conta ainda, com a participação do mediador, um complexo multiproteico muito grande que pode ser considerado um TF basal e transmite sinais de fatores regulatórios, que interagem com regiões regulatórias de genes transcritos ao complexo RNAPII. As proteínas componentes do **mediador** (20 a 30) formam um complexo bastante grande, de cerca de um milhão de daltons e está organizado em uma estrutura modular formando uma cabeça, uma porção intermediária e uma cauda.

4.2.1 Início da transcrição por RNAPII

A RNAPII transcreve todos os genes que produzem mRNA, além dos genes que codificam vários snRNAs (Tabela 10.2). Como visto anteriormente, os promotores de eucariotos apresentam grande variabilidade e o reconhecimento das sequências do promotor é o passo inicial para a montagem dos PICs.

Os TFs basais reconhecem e se ligam aos promotores em uma ordem hierárquica fundamental: promotor + TFIID; TFIIA; TFIIB; TFIIF + RNAPII; TFIIE e TFIIH. Esse é o complexo básico para a formação dos PICs nos genes de classe II em eucariotos (**Figura 10.5**).

O primeiro reconhecimento do promotor é realizado por TFIID. Esse TF foi descrito originalmente como sendo capaz de reconhecer a sequência canônica TATA *box*. Entretanto, como se sabe, TFIID reconhece todos os promotores principais de genes da classe II, mesmo aqueles promotores classificados como **TATA less**. TFIID é um complexo multiproteico formado por TBP ligado a 13 ou 14 fatores associados a TBP (**TAFs**, de *TBP-associated factors*).

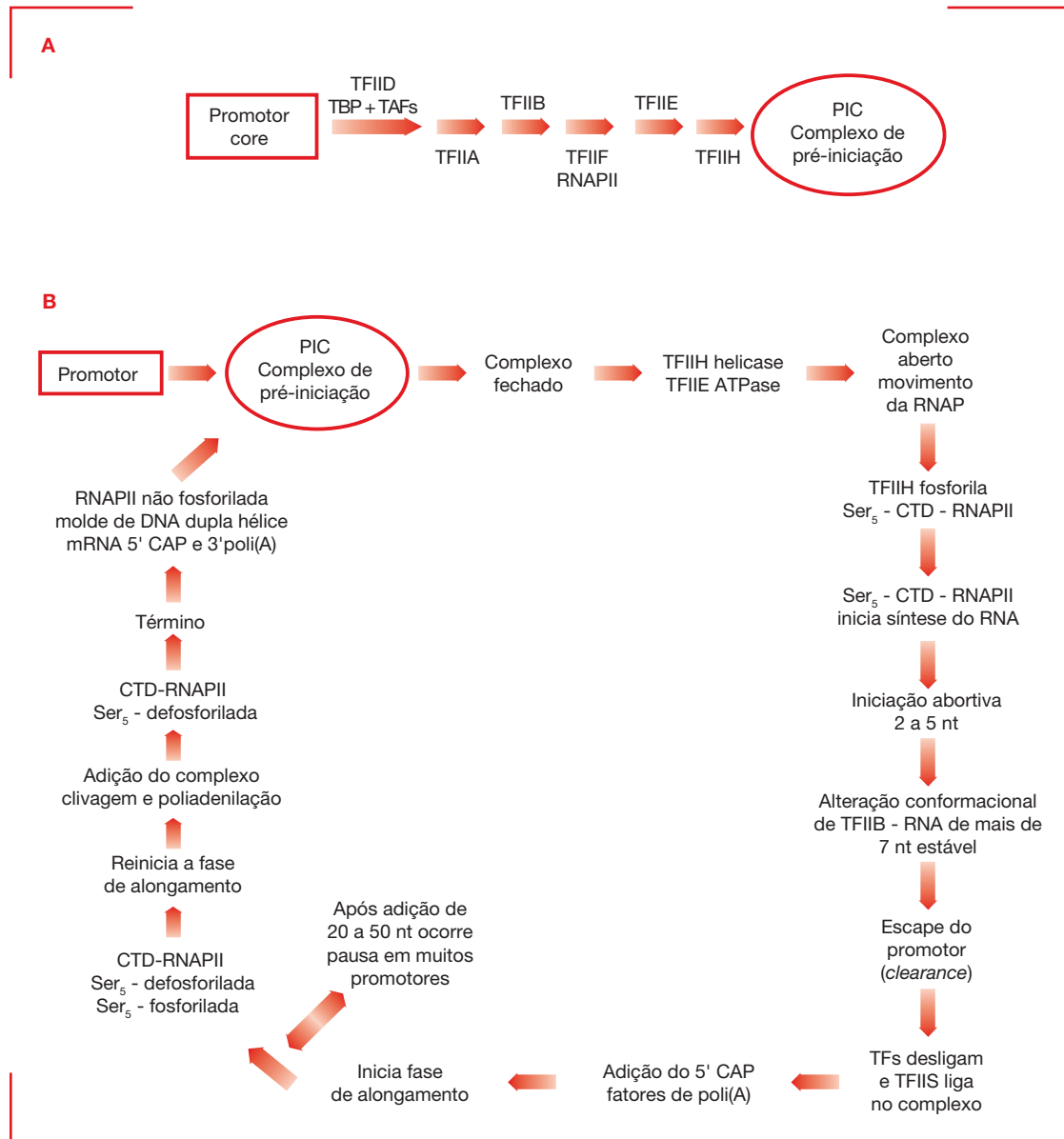
O componente principal de TFIID é a proteína TBP de 37 kDa, fundamental para o início da transcrição pelas eRNAPs nucleares. Cada classe de genes TBP forma um tipo específico de complexo com os TAFs. Para RNAPII o complexo formado é TFIID, para RNAPI o complexo é SL1 e para a RNAPIII o complexo é TFIIB. Nos genes da classe II, que contêm TATA *box*, TBP liga-se a sequência consenso TATA(A/T)A(A/T)(A/G). TBP liga-se ao DNA na sua cavidade menor, fato que é incomum, pois a cavidade maior é o sítio mais comum de ligação de proteínas que interagem com sequências de DNA. A hélice dupla fica acomodada na parte interna de TBP (sua porção C-terminal) que é muito conservada entre os organismos e apresenta domínios de ligação ao DNA. Nessa porção existem entre 32 a 39 repetições do aminoácido glutamina, responsáveis pela ligação ao DNA. A porção N-terminal, por sua vez, fica exposta, sendo muito variável e responsável pela interação com outras proteínas.

A estrutura cristalográfica de TBP e suas interações com o DNA foram estudadas de forma detalhada. TBP possui uma conformação de “sela” e se liga a uma das faces do DNA na região da cavidade menor, nas posições -37 a -25. A exposição das sequências, cerca de 8 pb, dessa região do promotor é aumentada, pois TBP força uma dobra na estrutura da hélice dupla de cerca de 80° que desenrola cerca de 1/3 de volta da hélice. Embora não ocorra a ruptura das pontes de hidrogênio da hélice dupla, essa alteração conformacional expõe as sequências flangeadoras ao sítio de ligação de TBP, de forma que outros TAFs ou TFs possam se ligar ao complexo. Após essa alteração do DNA, nas regiões flangeadoras do sítio de ligação de TBP, outros TFs e TFIID podem interagir com uma região mais extensa do promotor, que vai da posição -45 a -10. No TFIID livre, a proteína acessória TAF1 liga TBP na região da “sela”, impedindo que TBP possa se ligar ao DNA. A região N-terminal de TAF1 “imita” estruturalmente a cavidade menor do DNA, precisando ser deslocada para que TFIID possa se ligar ao promotor. TFIIA altera a conformação do TAF1 e promove uma maior interação entre TFIID e o promotor.

Outras proteínas que apresentam sequências relacionadas a TBP e que fazem parte do complexo TFIID em determinados tipos celulares, são denominadas fatores relacionados a TBP (TRFs, de *TBP related factors*). A **Figura 10.6** ilustra um complexo TFIID. TBP é responsável pela ligação ao TATA *box* e os outros TAFs componentes de TFIID auxiliam na ligação ao promotor. Por exemplo, os TAFs 1 e 2 ligam-se ao iniciador (Inr) e os TAFs 6 e 9 aos elementos a jusante dos promotores principais. Em geral, TFIID apresenta 800 kDa de tamanho e nem todos os complexos têm a mesma composição.

Em sequência à ligação de TFIID ao promotor, associa-se ao complexo os fatores TFIIA e TFIIB. Estudos de estrutura mostraram que TFIIA liga-se à sequência situada a montante do sítio de ligação de TFIID e TFIIB liga-se a jusante. Portanto, TFIIA e TFIIB “fixam” o DNA nos dois lados de TBP/DNA. Esses dois TFs promovem a formação do PIC, estabilizando a interação TBP/DNA e recrutando a RNAPII e outros TFs. A partir disso, a interação entre TFIID e o promotor é estabilizada e a sua região de ligação ao DNA estende-se a montante. TFIIA desempenha, ainda, mais uma função importante que é a de liberar a repressão exercida pelo TAF1 em TBP, permitindo a interação mais definitiva entre TBP e o promotor.

TFIIB estende a interação de TFIID de -10 até +10 e interage com o elemento BRE a montante. TFIIB possui também a função de conferir polaridade ao promotor, ou seja, posiciona a RNAPII quando ela entra no PIC e auxilia na seleção da fita-molde. Essa função de “posicionamento” do complexo é muito importante e característica de alguns TFs e define o sentido da transcrição. TFIIB auxilia no posicionamento e no alinhamento de TFIIE, TFIIF e TFIIH no PIC. TFIIB tem, ainda, mais uma função importante na passagem da fase de início para a de alongamento da transcrição.

**Figura 10.5**

Sequência de eventos na transcrição mediada por euRNAPII. (A) Etapa de início da transcrição, montagem do PIC. **(B)** Etapas de alongamento e terminação da transcrição.

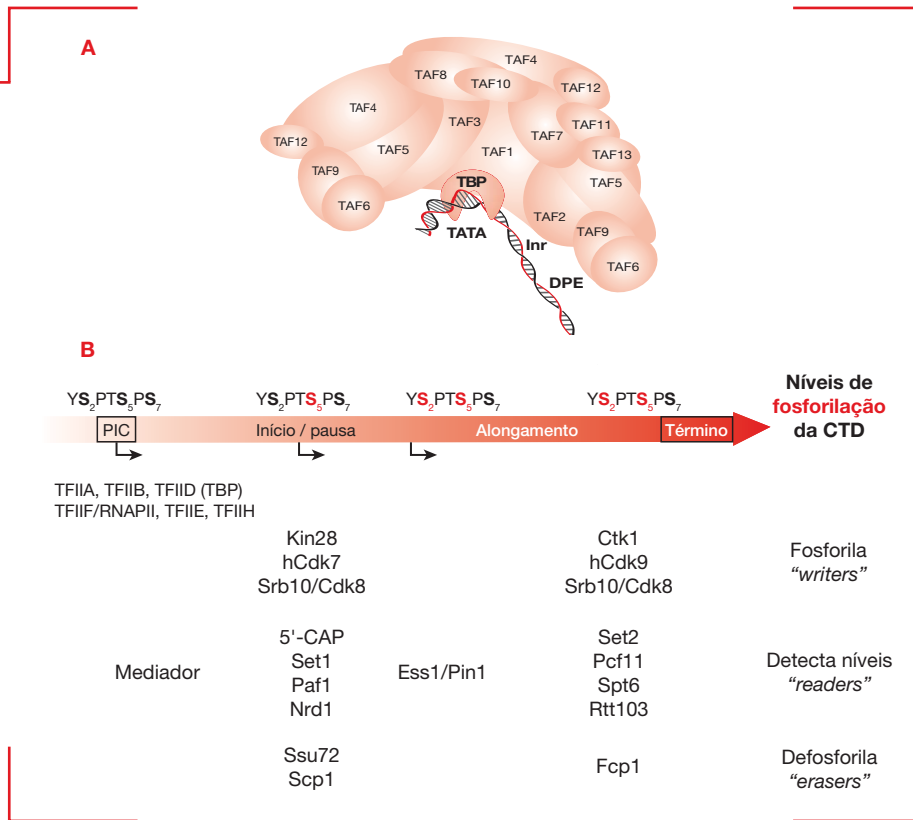
No complexo promotor+TFIID+TFIIA+TFIID ocorre a ligação de TFIIF que, provavelmente, já está ligado com a RNAPII. Há evidências de que ocorra primeiro a formação de um complexo TFIIF/RNAPII que depois é recrutado ao PIC. TFIIF é um complexo com quatro proteínas que forma duas subunidades. A subunidade maior apresenta uma atividade de helicase (dependente de ATP) e auxilia na desnaturação da hélice dupla na região do promotor. A subunidade menor apresenta homologia com regiões do fator σ bacteriano e é responsável pela forte ligação com a RNAPII. Quando TFIIF/RNAPII chega ao PIC outras interações ocorrem entre o TFIIF e o DNA e entre a RNAPII e o TFIID

e o TFIIB. A RNAPII estende a região de ligação até a região +25.

Neste ponto, uma região muito maior do DNA, na região do promotor, está interagindo com TFs e RNAPII, estendendo a sequência reconhecida por TBP no início da formação do complexo. A montagem do PIC é finalizada pela entrada dos dois últimos TFs basais: TFIIE e TFIH. TFIIE é um heterotetrâmero que além da interação com a RNAPII interage ainda com TFIIF, TFIIB e com o promotor e auxilia no recrutamento de TFIH. A sua interação com a RNAPII ocorre na região do centro ativo e com o promotor na região +10, onde ocorrerá a abertura da hélice dupla para o início da transcrição. TFIIE tem

Figura 10.6

TFIID e CTD-RNAPII. (A) Representação esquemática do complexo multiproteico TFIID, formado pelo TBP associado a 13 ou 14 TAFs. No desenho do DNA estão representados o TATA *box* interagindo com TBP, o Inr interagindo com TAF2 e o DPE interagindo com TAF9 e 6. **(B)** O domínio C-terminal da subunidade maior da euRNAPII (CTD) é fundamental para as diferentes fases da transcrição. São mostradas as alterações de fosforilação das serinas das repetições do CTD durante a transcrição. A figura mostra exemplos de proteínas que fosforilam, defosforilam e detectam os níveis de fosforilação do CTD.



afinidade por DNA de fita simples e estimula as atividades enzimáticas de TFIIH, facilitando a formação de um complexo de iniciação competente com a RNAPII para o início da transcrição.

O complexo TFIIH é o último TF basal necessário para a formação do PIC *in vitro* (Figura 10.5A). O TFIIH é composto por nove proteínas e tem três atividades enzimáticas fundamentais para o início da transcrição: a atividade ATPase DNA dependente; a atividade de helicase ATP-dependente; e a atividade de quinase, que fosforilará a CTD da subunidade maior da RNAPII (Rpb1). A sua atividade de helicase é importante para a separação das fitas no promotor e, portanto, para a formação do complexo aberto de transcrição. A atividade de abertura do promotor é realizada em associação com o TFIIE e é importante na transição da fase de início para o alongamento da transcrição. A afinidade de TFIIE com a fita simples do DNA é consistente com essa função. A necessidade de TFIIE e TFIIH depende da topologia do DNA (superenrolamento) e da sequência do promotor. Em modelos experimentais, em que o promotor é previamente desnaturado, a transcrição pode ocorrer na ausência de TFIIE e TFIIH.

TFIIH é um dos TFs mais versáteis, pois participa da formação do PIC, juntamente com TFIIB e TFIIE; participa da transição entre as fases de início e de alongamento da transcrição; e participa da reparação de DNA danificado (Capítulo 7).

Mecanicamente e de forma simplificada, podemos considerar as seguintes etapas na formação do PIC: (1) TBP (TFIID) reconhece o promotor e provoca uma dobra no DNA em torno da RNAPII e do domínio C-terminal de TFIIB; (2) o domínio N-terminal de TFIIB posiciona o DNA na região do TSS no centro ativo da RNAPII; (3) TFIIE entra no complexo e recruta TFIIH que induz tensão negativa na hélice dupla do DNA, forçando a sua abertura; (4) essa abertura permite a ligação de TFIIF na fita não molde; (5) a fita-molde é posicionada no centro catalítico da RNAPII e sob controle de uma alça de TFIIB ocorre a iniciação abortiva, sintetizando oligômeros de 6 nucleotídeos. A **Tabela 10.3** sumariza os componentes e sua ordem de entrada para a formação do PIC.

O complexo definido como PIC e permite a transcrição *in vitro* por RNAPII. Como visto na Seção 3.2, além dos TFs basais, o **mediador** também participa do processo. O mediador é atraído ao PIC por proteínas ativadoras que se ligam a promotores ou *enhancers* e se associam a TFs basais e a RNAPII. O mediador dissocia-se da RNAPII após o início da transcrição, permanecendo ligado ao promotor e mantendo uma estrutura pré-formada que permite, mais facilmente, o reinício da transcrição naquele gene. Além de atuar como um TF basal, o mediador pode funcionar como coativador ou correpressor da transcrição. O mediador sofre grandes alterações conformacionais causadas por interações entre a

Tabela 10.3 Componentes mínimos para a transcrição basal por RNAPII

Componente (por ordem de entrada no PIC)	Composição de proteínas (número de proteínas)	Função no PIC
TFIID	TBP/TAFs (1-14) (quinze)	TF que se liga ao promotor principal
TFIIB	p35 (α), p19 (β), p12 (γ) (três) p33 (uma)	Seleção do TSS, estabiliza TBP-TATA, recruta o complexo RNAPII/TFIIF, posiciona a RNAPII e confere direcionalidade, posiciona TEFIIE, F e H no PIC
TFIIF	PAP30, RAP74 (duas)	Liga a RNAPII e facilita a sua entrada no PIC, recruta TFIIE e TFIIH, atua com TFIIB e a RNAPII na seleção do TSS, facilita o escape com complexo RNAPII do promotor e o início da transcrição, aumenta a eficiência do alongamento
RNAPII	RPB1 –12 (doze)	Participa do início, do alongamento e do término da transcrição, recruta enzimas do <i>capping</i> , acopla a transcrição e o <i>splicing</i> , recruta fatores de processamento na extremidade 3', participa da fosforilação da CTD, glicosilação e ubiquitinação
TEFIIE	p56 (α), p34 (β) (duas)	Recruta TFIIH, interage com a RNAPII para o início da transcrição (<i>clearance</i> , liberação do promotor)
TFIIH	P89/XPB, p80/XPD, p62, p52, p44, p40/ CDK7, p38/ciclina H, p34, p32/MAT1, e p8/ TFB5 (dezesesseis)	Atividade ATPase para a iniciação e liberação do promotor, atividade helicase para abertura da dsDNA no promotor, acoplamento entre transcrição e o sistema de reparação, atividade de quinase para fosforilação da CTD da RNAPII

RNAPII e as proteínas regulatórias. Da mesma forma, o mediador pode expor sítios crípticos de ligação da RNAPII modulando a formação dos PICs. Essas alterações na conformação da RNAPII por ativadores e repressores regulam a transcrição e explicam o mecanismo regulatório do mediador.

Com apenas algumas exceções, as proteínas que compõem os mediadores apresentam grande variabilidade estrutural e de sequência, embora a sua organização espacial seja semelhante e mantenha a estrutura cabeça, região mediana e cauda. Essas diferenças entre os mediadores podem oferecer alvos flexíveis para diversas proteínas regulatórias com diferentes estruturas, representar regiões ligantes variáveis entre domínios globulares e facilitar a montagem e a desmontagem dos complexos em resposta a sinais regulatórios.

A sequência de eventos para a formação do PIC nos promotores **TATA less** é bastante semelhante. A diferença é que TFIID, na ausência de TATA *box*, reconhece e liga-se a Inr. Esse processo é realizado por um ou mais TAFs e Inr posiciona TFIID, ao passo que outros TAFs reconhecem o elemento DPE. A função de TBP, neste caso, é interagir com a RNAPII e a posicionar no TSS. Nos promotores TATA *less* esse posicionamento do TSS é menos eficiente, resultando em um início de transcrição difuso. Isso não significa perda na eficiência do início da transcrição, apenas que os transcritos terão alguma variabilidade no seu nucleotídeo +1. É importante lembrar que a maioria dos genes de vertebrados tem seu início da transcrição difuso e mesmo assim eficiente.

4.2.2 Início da transcrição por RNAPI e RNAPIII

A RNAPI é especializada na síntese do pré-rRNA que irá produzir os três rRNAs principais (28S, 18S e 5S), representando 80% de todo o RNA da célula. Essa síntese ocorre no nucléolo e os precursores são processados até seus produtos finais funcionais, antes de atingirem o citoplasma para fazerem parte dos ribossomos (Capítulo 11). A RNAPI é composta por 14 subunidades, sendo similar a RNAPII em termos de estrutura. Entretanto, a sua subunidade maior não apresenta a CTD, sendo essa a razão da não existência das estruturas 5'-CAP e cauda de poli(A) nestes rRNAs.

A RNAPI deve ter alta eficiência e os promotores dos genes para rRNA devem ser altamente especializados e eficientes. A razão para isso é que a célula deve produzir a cada ciclo uma quantidade suficiente de rRNA para compor os milhões de ribossomos (1 a 2 milhões por geração), que participarão da síntese de todas as proteínas da célula. Assim, a RNAPI deve sintetizar cerca de dez milhões de cópias dos pré-rRNAs. Uma das estratégias desenvolvidas para responder a essa demanda foi o aumento do número de cópias desses genes, podendo variar entre cerca de 100 a 5.000. Essas cópias são organizadas **em tandem** no cromossomo e cada cópia (unidade de transcrição) é separada por regiões não transcritas, denominadas espaçadores não transcritos (Capítulo 5).

Os promotores dos genes de rRNA são iguais em todas as cópias e altamente especializados para garantir e-

eficiência máxima na transcrição. Esse processo sofreu tamanha pressão evolutiva que os promotores para RNAPI diferem em cada organismo analisado. As diferenças entre os promotores contrastam com o alto grau de conservação da sequência de nucleotídeos da região codificadora, que tem função estrutural. Os rRNAs estão entre as sequências mais conservadas filogeneticamente. A maior parte dos promotores para RNAPI são compostos por dois domínios de atividade: o promotor principal, que tem sequências ricas em GC e está localizado entre -45 e $+20$, a partir do TSS (localizado em uma região rica em AT); e o elemento promotor a montante, UPE (de *upstream promoter element*, também denominado UCE), rico em GC e localizado entre -180 e -107 (Figura 10.4.B). Essa organização é conservada nos diferentes eucariotos, embora as sequências específicas sejam variáveis.

Para a transcrição dos genes de rRNA, a RNAPI necessita dos fatores SL1 e UBF1. O fator SL1 (ou TIF-1B ou Rib1) é um complexo formado por quatro proteínas, sendo uma delas a proteína TBP. Como TBP não se liga às sequências ricas em GC do promotor, outras proteínas no complexo devem fazer este reconhecimento. TBP participa interagindo com a RNAPI. SL1 em conjunto com a RNAPI pode transcrever, mas em níveis baixos. A provável função de SL1 é o posicionamento da RNAPI no TSS.

O fator UBF é uma proteína de 97 kDa e liga-se a regiões do DNA ricas em GC do elemento UPE, atuando na forma de dímero. As duas funções atribuídas a UBF são estimular a liberação da RNAPI do promotor, iniciando a fase de alongamento, e estimular a ligação de SL1 e sua interação com a RNAPI. Depois de se ligar à região do promotor, UBF força o DNA a formar uma alça (360°) possibilitando que o promotor principal fique próximo ao UPE, permitindo, assim, que UBF estimule a ligação de SL1 ao promotor. A distância entre UPE e o promotor principal é importante para a formação dessa alça de DNA mediada pelo dímero de UBF. Essa distância pode ser alterada introduzindo fragmentos de DNA que tenham o tamanho de uma volta da hélice, sem alterar a eficiência da iniciação. Se, no entanto, meia volta for adicionada a eficiência diminui. Quando os dois fatores (SL1 e UBF) estão especificamente ligados ao DNA, a RNAPI liga-se ao promotor central e inicia a transcrição.

A polimerase RNAPIII catalisa a síntese de rRNA 5S, tRNAs, snRNA, U6, e RNA 7S. Esses RNAs representam cerca de 10% do RNA total de uma célula. Os promotores de RNAPIII (Figura 10.4.C) pertencem a duas classes distintas: a dos genes de snRNAs, em que os promotores são similares aos até aqui analisados e ocorrem a montante do TSS; e a dos promotores dos genes de rRNA 5S e dos tRNAs, que são “internos” e localizam-se a jusante do TSS, ou seja, na própria região que é transcrita.

Os promotores dos genes de tRNAs apresentam o elemento A (*box A*), que ocorre entre $+8$ e $+19$, com o consenso T(A/G)GC-AG(T/C)-GG, e o elemento B (*box B*), que ocorre entre $+52$ e $+62$, com o consenso GGTTCCGA-TCC. A distância entre A e B varia, em diferentes organismos, entre 31 a 93 pb.

Os três fatores de transcrição que participam do reconhecimento dos promotores internos são TFIIA, uma **proteína dedo de zinco**; TFIIB, composto por TBP e mais duas proteínas; e TFIIC, um complexo de seis proteínas. Nos promotores de genes de tRNAs, TFIIC liga-se nos elementos A e B. Nos genes de RNA 5S, antes que TFIIC possa ser ligado, o fator TFIIA deve ligar-se ao elemento A. TFIIC posiciona a ligação de TFIIB (contém TBP) no TSS que recruta a ligação da RNAPIII. TFIIA e TFIIC são fatores de montagem do complexo, cuja função é auxiliar a ligação de TFIIB, que é o fator de posicionamento. Após a ligação de TFIIB, os complexos TFIIA e TFIIC podem ser desligados do promotor, pois TFIIB é suficiente para que a RNAPIII identifique e ligue-se ao TSS. TFIIB além do TBP é formado pela proteína Brf, relacionada ao TFIIB e pela proteína B". A proteína B" é desnecessária se o DNA estiver desnaturado e, portanto, deve participar da abertura das fitas de DNA e parece exercer uma função no início da formação da bolha de transcrição, semelhante ao fator sigma nas bacRNAPs.

Nos promotores dos genes de snRNAs o elemento TATA pode iniciar a transcrição, mas a eficiência é aumentada pelos *enhancers* PSE e OCT e a montagem do complexo assemelha-se a que ocorre nos promotores de classe I e II. É possível, ainda, que o elemento TATA seja o responsável por conferir especificidade por RNAPIII ou RNAPII nos promotores de snRNAs, pois alguns genes de snRNA são transcritos pela RNAPIII e outros genes de snRNA são transcritos pela RNAPII.

A estratégia de formação de PICs nos genes de classe III demonstra algumas características que diferenciam a transcrição em eucariotos: a RNAP nunca se liga diretamente ao promotor ou na ausência de outros TFs; alguns TFs ligados primeiro podem ser removidos depois que outros TFs são ligados, sem prejuízo da iniciação (denominados fatores de montagem); e alguns TAFs, em geral associados a TBP, têm a função de posicionar a RNAP no TSS.

5. Alongamento da cadeia de RNA

5.1 Alongamento em procaríotos

O complexo de alongamento foi detalhado por análises estruturais de modelos cristalográficos, principalmente das bactérias *E. coli* e *Thermus thermophilus*. Esses estudos de estrutura associados aqueles realizados com levaduras mostraram que embora existam particularidades, os modelos desenvolvidos podem ser aplicados a todos os organismos.

Durante a etapa de alongamento, o complexo RNAP deve desempenhar várias atividades: (1) separar (desnaturar) a hélice dupla do DNA na região a jusante do complexo de iniciação, expondo a sequência de nucleotídeos a ser copiada; (2) manter as fitas de DNA separadas na

região de síntese, acomodando a fita não molde; (3) catalisar a reação de incorporação dos ribonucleotídeos; (4) permitir o tráfego dos ribonucleotídeos ao centro catalítico e a saída dos pirofosfatos resultantes da reação de incorporação e dos ribonucleotídeos erroneamente incorporados; (5) manter estável o híbrido DNA-RNA na região de síntese; (6) separar o RNA sintetizado do híbrido e permitir a sua saída do complexo de transcrição; (7) renaturar o DNA na região a montante da bolha de replicação; e (8) responder a alterações ocasionadas pela ligação de proteínas regulatórias ao complexo e detectar estruturas secundárias e terciárias no DNA e no RNA.

Os estudos de estrutura mostraram que o complexo de alongamento se organiza em canais ou cavidades onde algumas atividades são realizadas. A **Figura 10.7** mostra de forma esquemática algumas dessas regiões do complexo de transcrição. Cada uma dessas regiões é o resultado das interações entre as diferentes subunidades da RNAP.

Três regiões do complexo são responsáveis por ligar-se a ácidos nucleicos (Figura 10.7A):

1. Sítio de ligação a dsDNA a frente (jusante) da bolha de transcrição, que interage com cerca de 9 pb da hélice dupla imediatamente a jusante do centro catalítico. Esse sítio é formado pelas “pinças” da subunidade β' .
2. Sítio de ligação ao híbrido DNA-RNA com cerca de 7 a 9 pb de tamanho na região do DNA desnaturado que apresenta cerca de 12 a 15 pb de extensão. Esse sítio é formado por interações entre as subunidades β e β' .
3. Sítio de ligação ao RNA nascente distante cerca de 8 a 14 nucleotídeos do centro catalítico. Esse sítio é formado pela extremidade N-terminal da subunidade β' .

As três regiões interagem alostericamente com o centro catalítico e, portanto, influenciam a catálise, dependendo das alterações conformacionais que ocorrem nos ácidos nucleicos com os quais estão em contato. Assim, alterações conformacionais no DNA ou no RNA afetam a velocidade da adição dos ribonucleotídeos. Além dessas regiões o complexo deve acomodar a fita de DNA não molde.

O complexo RNAP contém ainda o centro catalítico, onde a reação de incorporação dos NTPs ocorre, e o canal (referido como canal secundário) que permite a entrada dos NTPs e a saída do pirofosfato resultante da reação e dos NTPs errados. Esse canal é bastante profundo e conecta diretamente o meio externo ao centro catalítico.

O centro catalítico da RNAP é formado por três resíduos de aspartato em uma região muito conservada da subunidade β' . Esses resíduos coordenam os dois íons Mg^{+2} e estão envolvidos por vários domínios das subunidades β e β' . A reação de adição de cada NTP envolve três etapas: a entrada do ribonucleotídeo pelo canal secundário e sua ligação ao complexo; a incorporação do ribonucleotídeo na cadeia nascente da extremidade livre 3' do

RNA, pela formação de uma ligação fosfodiéster com a liberação de pirofosfato; e a translocação do complexo de transcrição para liberar o sítio de inserção para a entrada do próximo ribonucleotídeo a ser incorporado. A chegada do ribonucleotídeo até o centro catalítico passa por várias etapas de interação com domínios da RNAP, que são responsáveis por reconhecer o ribonucleotídeo correto a ser adicionado na cadeia do RNA em relação a fita-molde de DNA. Esse reconhecimento ocorre a partir de uma rede de interações que englobam todas as características químicas dos NTPs: as bases, o açúcar e os fosfatos. As interações monitoram o pareamento correto do NTP na base do DNA-molde.

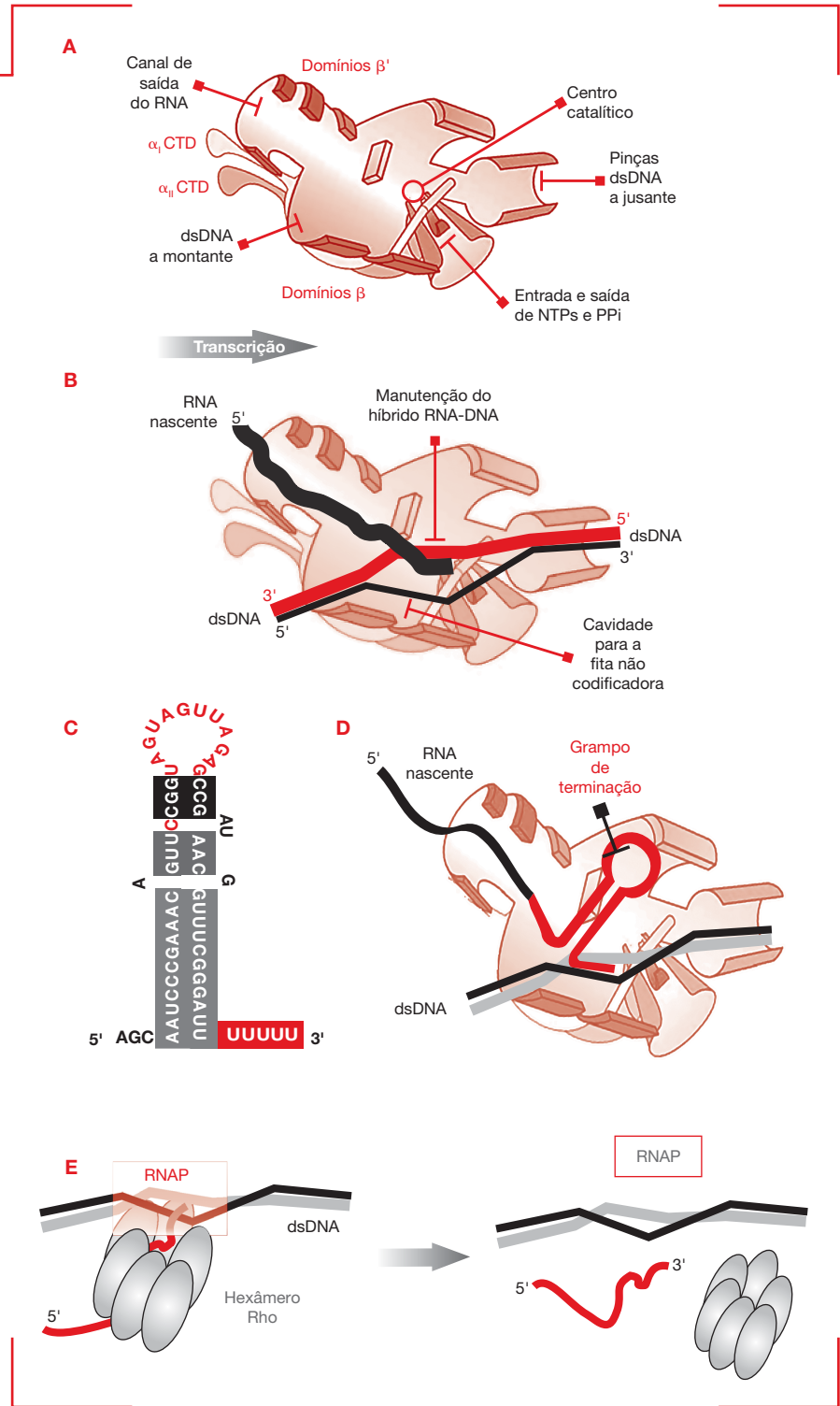
A região a jusante do complexo é formada pelas “pinças” da RNAP que interagem com 13 pb dos DNAs hélice dupla. No sentido a montante, as fitas são abertas e a fita não molde é acomodada, evitando sua renaturação. Na região a montante do centro catalítico, um canal mantém estável o híbrido DNA-RNA. No centro catalítico existe um canal permitindo a entrada dos ribonucleotídeos, a saída dos produtos da reação de incorporação (pirofosfato) e de ribonucleotídeos que foram incorretamente incorporados ou que não são incorporados ao molde. Outro canal estabiliza o RNA sintetizado, de forma a ser separado do DNA-molde, e direciona o RNA para fora do complexo de transcrição. Em um outro canal, as fitas do DNA são renaturadas para formar novamente a hélice dupla (Figura 10.7B).

A velocidade de transcrição não é constante ao longo de toda a fase de alongamento. Em geral, a velocidade de transcrição pela RNAP de *E. coli* é de 50 a 100 nucleotídeos por segundo, mas é alterada dependendo da sequência de nucleotídeos da região do DNA, que está sendo transcrita, da estrutura do RNA, que está sendo sintetizado e da ligação de proteínas ao complexo de transcrição. Sequências ricas em GC tendem a reduzir a velocidade, provavelmente porque esses pareamentos são mais difíceis de desnaturar e a etapa de abertura das fitas é mais demorada do que em outras regiões. Em algumas regiões, a transcrição para rapidamente e, depois, é retomada, sem que o complexo seja desfeito. Essas paradas são denominadas **pausas** e podem durar dezenas de segundos.

A processividade (tempo que uma enzima permanece em contato com o substrato) da RNAP é elevada e, mesmo em pausas prolongadas, o complexo de transcrição não é dissociado antes do sinal de terminação da transcrição. Outro fator que propicia as pausas é o acoplamento, em procaríotos, dos processos de transcrição e tradução (Capítulo 13). A proteína NusA, que se liga ao complexo após o desligamento de σ , amplifica o sinal nos sítios de pausa, principalmente quando **sequências repetidas invertidas** ou palíndromos estão presentes. Nesse caso, o RNA sintetizado forma um **grampo** por pareamento interno das sequências complementares, o qual parece ser estabilizado pela proteína NusA. Essas pausas podem durar de 10 segundos a 30 minutos. Em alguns casos, a pausa é utilizada como mecanismo de controle da expressão gênica (Capítulo 13).

Figura 10.7

Domínios funcionais das RNAPs. (A) Domínios funcionais das RNAPs. Desenho ilustrativo a partir da estrutura tridimensional do complexo. (B) Posicionamento dos ácidos nucleicos em relação aos domínios da RNAP. (C) Grupo de terminação de *E. coli*. (D) Formação do grupo de terminação no complexo RNAP. (E) Terminação Rho dependente. Modelo alostérico.



5.2 Alongamento em eucariotos

A transcrição nos eucariotos, mesmo no modelo mais simples de levedura, ocorre em um contexto estrutural muito diferente do que ocorre em procariotos. A estrutura da cromatina impõe restrições espaciais à acessi-

bilidade das sequências contidas nas diferentes regiões regulatórias e os nucleossomos impõem uma barreira ao alongamento da transcrição (Capítulo 3). Além disso, a diversidade das três euRNAPs é um fator adicional de complexidade nos eucariotos. Ainda assim, os estudos estruturais nos complexos de alongamento da transcrição

de bactérias, arqueas e eucariotos têm evidenciado muitas semelhanças, mostrando grande conservação estrutural e mecânica em seus fundamentos. Portanto, muito do que foi descrito, para o alongamento da transcrição em bactérias, tem analogia com os modelos eucarióticos, em sua essência. O processo em arqueas ocupa uma posição intermediária de complexidade e apresenta algumas características do modelo eucarioto.

Assim como ocorre com o fator σ em bactérias, os TFs, que participam da formação do PIC nos eucariotos, se dissociam do complexo na transição entre as fases de início e de alongamento. Entretanto, outros fatores são recrutados e participam do complexo de alongamento. A transcrição é regulada por alterações na cromatina e o RNA deve ser, no caso dos mRNAs, concomitantemente processado (adição de 5'-CAP, remoção dos íntrons e poliadenilação) e, por isso, muitos fatores e complexos participam desse processo. Os modelos estruturais eucarióticos mais estudados e detalhados são os da euRNAPII. As etapas de associação da euRNAPII em um complexo fechado, com o DNA no promotor, a formação do complexo aberto e o “escape” (*clearance*) do promotor para o início da fase de alongamento da transcrição são análogos ao da bacR-NAP. Entretanto, existem algumas diferenças em termos de mecanismo: nos eucariotos o complexo aberto depende da atividade de helicase do complexo TFIIB; ocorre fosforilação da RNAPII (domínio C-terminal, CTD), e o movimento da euRNAPII sobre o molde de DNA para o escape do promotor é dependente da hidrólise de ATP.

Como mostrado na Figura 10.5B, após a formação do PIC os fatores basais TFIIE e TFIIH exercem suas atividades de helicase e de hidrólise de ATP, que provocam a abertura das fitas de DNA e proveem energia para o complexo RNAPII se movimentar para o escape do promotor. Após a incorporação dos primeiros ribonucleotídeos na cadeia de RNA o complexo deve ser deslocado, liberando o promotor e iniciando o alongamento. Os TFs e o mediador que estavam complexados no promotor, deixam, na sua grande maioria, o complexo e não acompanham o movimento da RNAPII sobre o molde de DNA. Na iniciação abortiva de eucariotos, RNAs de apenas 2 a 3 nt são sintetizados e rapidamente metabolizados até que a fase de alongamento seja iniciada. Uma etapa adicional ocorre em muitos genes de eucariotos, a pausa na transcrição após a formação de RNAs de 20 a 50 nt de tamanho.

Recentemente foi demonstrado que TFIIB tem participação direta na transição entre o início e o alongamento da transcrição. TFIIB estabiliza o complexo até que o híbrido RNA-DNA apresente um tamanho suficiente para iniciar o alongamento. A estrutura formada pelo complexo TFIIB-RNAPII é similar aquela formada entre bacRNAP e o fator σ . TFIIB tem regiões com similaridade ao fator σ e apresenta uma alça (denominada *finger B*) que interage com o centro catalítico da RNAPII. Portanto, TFIIB participa tanto do início da transcrição, ligando-se a sequências regulatórias no promotor, como das al-

terações conformacionais, que determinam o escape da RNAPII do promotor. Mostrou-se, também, que TFIIB é responsável pelo desligamento da maioria dos TFs na passagem do PIC para o alongamento da transcrição. Um TF que se liga ao complexo para o alongamento é o TFIIS, homólogo do fator bacteriano GreA, cuja atividade aumenta a eficiência da transição entre o início e o alongamento. Esses TFs estão associados diretamente ao centro ativo da RNAP, tem atividade de clivagem de RNA e provavelmente atuam auxiliando a remoção das cadeias de RNA não produtivas.

Outra atividade fundamental na liberação do promotor e na passagem para a etapa de alongamento é a fosforilação da cauda CTD da subunidade maior da RNAPII (Figura 10.6B). O domínio CTD coordena vários processos durante a transcrição, sendo responsável pelo recrutamento de proteínas envolvidas nas modificações das histonas, na fase de alongamento, na terminação e no processamento do RNA. A fosforilação de CTD também promove o desligamento de TFs e do mediador do complexo de iniciação.

O padrão de fosforilação de CTD é dinâmico e se altera durante o ciclo da transcrição, sendo controlado por enzimas que promovem a sua fosforilação ou desfosforilação. A Figura 10.6 mostra algumas dessas alterações. CTD contém a repetição $YS_2PTS_5PS_7$ (26 repetições em *Saccharomyces* e 52 em seres humanos) e pode ser fosforilada em cinco posições. Entretanto, apenas a fosforilação das serinas S_2 e S_5 estão bem caracterizadas. A fosforilação da S_7 foi descrita na transcrição de snoRNA em células humanas, mas a sua ocorrência na transcrição de outros genes ainda não está caracterizada.

Muitas proteínas reconhecem e ligam-se a CTD com padrões específicos de fosforilação. Durante a formação do PIC, a RNAPII que entra no complexo não tem a sua CTD fosforilada. Durante o início da transcrição CTD é fosforilada na S_5 (pela quinase Kin28 de TFIIH). Essa marca de fosforilação atrai alguns fatores, incluindo a enzima (guanosil transferase) que fará a adição do 5'-CAP ao mRNA (a formação do 7-metil-guanosil-CAP, protege o mRNA da degradação e o marca para o transporte através da membrana nuclear); um complexo de verificação da qualidade dos mRNAs (Nrd1-Nab3 que detecta mRNAs defeituosos e portanto não funcionais); e o complexo PAF (conecta a CTD fosforilada a uma rede de modificação de histonas com funções regulatórias).

Com a continuidade do alongamento da cadeia de RNA ocorre uma alteração no padrão de fosforilação de CTD da S_5 para a S_2 . Portanto, a marca de fosforilação S_5 é “apagada” por uma fosfatase (Ssu72), que apresenta outras atividades importantes durante a transcrição: está associada a TFIIB no PIC; é necessária na transição S_5 para S_2 ; e faz parte do complexo de clivagem e adição da cauda de poli(A). As prolinas presentes nas repetições da CTD também têm um papel importante no controle das alterações no padrão de fosforilação, pois elas podem sofrer isomerização e nas repetições da CTD permitem que este domínio controle suas próprias alterações con-

formacionais. A fosforilação da S_2 dispara uma onda de comunicação entre a CTD e a cromatina local.

Todas as fases da transcrição (início, pausa, alongamento e terminação) são reguladas pela fosforilação ou defosforilação de S_5 e S_2 e por isomerização das prolinas nas repetições contidas na CTD. Esses processos controlam a entrada e a saída dos complexos de modificação da cromatina e os fatores de pausa e de terminação, permitindo que a RNAPII navegue, de forma eficiente, pelos nucleossomos produzindo mRNAs completos com precisão em seu início e terminação. Além disso, a CTD participa direta ou indiretamente do processamento pós-transcricional do mRNA, da ligação de fatores que farão o transporte para o citoplasma através da membrana nuclear e, de modificações na cromatina como a metilação de histonas (ver Capítulos 3 e 14).

Uma característica inesperada no início da fase de alongamento da transcrição é, que além da formação da extremidade 5'-CAP, alguns fatores responsáveis pela clivagem do RNA para a adição da extremidade 3' poli(A) também ligam-se ao complexo. Embora esses fatores sejam utilizados somente na fase de terminação da transcrição a sua ligação ocorre no início da síntese do RNA.

Muitos genes são encontrados pausados na transcrição, ou seja, o PIC foi formado, ocorreu o início da transcrição até a formação dos RNAs de 20 a 50 nt, mas o complexo entra em pausa e os RNAs sintetizados são degradados. Para que esses complexos sejam ativados para a transcrição devem ocorrer alterações no padrão de fosforilação de CTD.

6. Terminação da transcrição

Todas as fases da transcrição estão sob controle estrito. Portanto, a terminação é determinada ativamente. Após seu início e alongamento, o complexo RNAP deve reconhecer o terminador, cessar a incorporação de ribonucleotídeos, dissociar o complexo de transcrição liberando os fatores, a RNAP e o DNA renaturado em sua conformação original e liberar o RNA sintetizado. Durante o alongamento, o controle da velocidade de transcrição é exercido por pausas, nas quais a transcrição é interrompida durante algum tempo podendo ser retomada. A diferença entre os mecanismos de pausa e de terminação é que na pausa não ocorre a dissociação do complexo de transcrição.

As pausas podem ocorrer por dois mecanismos, um que depende de sinais oriundos da formação de grampos de RNA e, outro que independe desse tipo de sinal. A terminação é normalmente determinada por uma pausa, que é seguida pela dissociação do complexo de transcrição. Em menor ou maior escala, dependendo da sua classe, o RNA sintetizado é processado. O processamento dificulta o estudo da terminação da transcrição, pois o RNA produzido não contém mais suas sequências originais na extremidade 3'.

6.1 Terminação em procariotos

Nas bactérias os sinais de terminação não dependem diretamente do reconhecimento de sequências no DNA, mas sim das sequências já transcritas no RNA. Em bactérias existem dois mecanismos gerais de terminação da transcrição, a terminação intrínseca e a terminação dependente do complexo ρ (*proteína Rho*). Sequências no DNA determinam o fim da transcrição, provocando a desestabilização do complexo de transcrição, e o RNA sintetizado participa ativamente da terminação de sua síntese, formando estruturas secundárias ou ligando a outros fatores.

A terminação intrínseca, também denominada independente da proteína ρ ou da terminação por formação de grampo no RNA (Figura 10.7C), ocorre na terminação da transcrição da maioria dos genes de bactérias. Neste tipo de terminação uma sequência definida, localizada na região 3' do gene, sinaliza o final da síntese do RNA. Um terminador intrínseco é reconhecido na sequência do DNA na porção 3' terminal dos genes por conter dois componentes essenciais, uma sequência invertida repetida rica em GC e uma sequência de adeninas no DNA-molde (uracilas no RNA, cerca de 7 Us).

A sequência invertida repetida, localizada a montante do último nucleotídeo a ser transcrito, permite que o RNA nascente forme um pareamento interno a partir das bases complementares na sequência, denominado **grampo de terminação**. O grampo é estabilizado pela região da alça e pelo pareamento GC. Quando a RNAP transcreve essa sequência ocorre a formação do grampo na molécula de RNA nascente (Figura 10.7D). Essa estrutura impede espacialmente o movimento da RNAP sobre o molde de DNA e provoca uma longa pausa no alongamento da cadeia, pois a reação de incorporação de ribonucleotídeos cessa.

Essa pausa, associada a uma região de pareamento AU, entre a fita-molde de DNA e o RNA, que se segue imediatamente após a sequência do grampo, desestabiliza o híbrido DNA-RNA no centro ativo da RNAP. O pareamento AU é um dos pareamentos mais fracos formados entre DNA e RNA e, portanto, somado a pausa da síntese do RNA ocasiona a desestabilização do complexo de transcrição RNAP-DNA-RNA. O próprio pareamento AU provoca a pausa na transcrição e permite que o grampo no RNA seja formado. As “pinças” da RNAP, que a mantinham firmemente associada ao DNA, são abertas e a RNAP perde seus contatos com o DNA, desestabilizando o complexo e liberando o RNA, o DNA e a RNAP livre. Logo após as fitas do DNA são renaturadas com o auxílio da proteína NusA.

Um outro tipo de terminação, dependente de Rho, ocorre pela ação de um complexo hexamérico da proteína Rho (Figura 10.7D). Nesses terminadores, estruturas evidentes, como as encontradas nos terminadores intrínsecos, não estão presentes na sequência do DNA. As sequências que participam dessa terminação estão espalhadas ao longo de cerca de 100 pb antes do último nu-

cleotídeo a ser incorporado no RNA. Essas sequências sinalizam as pausas na transcrição e, portanto, nas regiões 3' desses genes, a transcrição está pausada ou ocorrendo em baixas velocidades. Na ausência da proteína Rho essas sequências são inativas para a terminação e, em geral, é letal para a célula.

A proteína Rho se liga formando um anel em torno do RNA, necessitando para isso uma região de 60 a 80 nt. O complexo Rho apresenta uma atividade de translocase (e de helicase) que permite ao complexo se deslocar na molécula de RNA na direção de 5' para 3'. A energia necessária para esse deslocamento é fornecida pela hidrólise de ATP e pela atividade ATPase de Rho. O domínio com atividade ATPase está próximo de um domínio de ligação a RNA na extremidade N-terminal e existe, ainda, outro domínio de ligação a RNA na extremidade C-terminal da proteína Rho. Provavelmente esta conformação confere ao complexo Rho a direcionalidade de deslocamento no RNA. Esse complexo é responsável pela terminação da transcrição em muitas unidades de transcrição em bactérias.

Foram propostos três possíveis modelos para explicar como o complexo Rho provoca a terminação da transcrição. O primeiro modelo propõe a colisão do complexo RhoRNA nascente com o complexo de transcrição. O complexo Rho, na colisão, desloca (empurra) a RNAP do híbrido RNA-DNA, provocando a parada da transcrição. Ocorre, então, o desligamento da RNAP do complexo, resultando na terminação da transcrição. Como a transcrição e a tradução são acopladas em procariotos, o mRNA nascente somente estará livre de ribossomos na região 3' após o sinal de terminação da tradução. Experimentalmente, se códons de parada da síntese de proteínas são introduzidos em um determinado gene a montante do local original, os ribossomos se desligam prematuramente do mRNA e RNAs mais curtos são sintetizados. Isso demonstra que a ausência dos ribossomos permite a ligação de Rho ao RNA nascente e a colisão com o complexo de transcrição com a consequente terminação. O segundo modelo propõe que o complexo Rho internaliza (puxa) o RNA causando o deslocamento do híbrido RNA-DNA, o que determina a parada da transcrição e o desligamento da RNAP do complexo, resultando no término da transcrição.

Esses dois modelos têm como base a ligação da proteína Rho no RNA nascente do complexo de transcrição. Entretanto, recentemente, foi demonstrado que o complexo Rho não pode ligar-se ao RNA nascente. Devido à ligação dos ribossomos, não existe um tamanho mínimo de sequência do RNA disponível para a ligação do complexo Rho. Além disso, foi demonstrado que o complexo Rho liga-se a RNAP, estabelecendo contato com parte da subunidade β' e provocando alterações conformacionais na RNAP. Essa ligação pode ocorrer com a RNAP livre, na ausência de RNA e de DNA, mas acontece quando a RNAP está associada ao DNA no complexo aberto de transcrição. O complexo Rho, portanto, já estaria ligado a RNAP durante o alongamento da transcrição.

Foi proposto, então, um terceiro modelo, que atende melhor às evidências experimentais, denominado modelo “alostérico” para a terminação Rho dependente. Nesse modelo, o complexo Rho, ligado a RNAP, sinaliza uma parada na transcrição pelo rearranjo alostérico do centro catalítico da RNAP, provavelmente induzido por sequências na região de terminação. Essa interação da proteína Rho no canal principal da RNAP desestabiliza o complexo de transcrição estagnado, provocando a sua dissociação e o término da transcrição. Portanto, ao invés de “empurrar” a RNAP ou “puxar” o RNA, Rho invade o canal principal da RNAP provocando alterações de conformação e, com sua atividade helicase, separa o híbrido RNA-DNA. Como a organização estrutural e os mecanismos de atividade básicos são conservados nas RNAPs, é possível que este mecanismo de terminação seja de ampla utilização.

Um terceiro tipo de terminação tem sido considerado, embora não seja responsável pela terminação no final dos genes. Muitos complexos de transcrição podem ficar paralisados por mutações no DNA. O sistema de terminação com base no complexo Mfd promove a terminação da transcrição nesses complexos. O fator de acoplamento entre a transcrição e o sistema de reparação de DNA é uma translocase, e sua atividade dissocia complexos de transcrição por se ligar a RNAP e ao DNA que emerge a montante do complexo de transcrição. Mfd reconhece qualquer complexo de transcrição que esteja paralisado e não é sítio-específico. Além de promover a terminação da transcrição, Mfd recruta o sistema de reparação para corrigir as lesões do DNA naquele gene.

Nas arqueas, pouco ainda é conhecido sobre os mecanismos de terminação da transcrição. A terminação é comprovadamente mediada por sequências ricas em timinas na região 3' dos genes. Entretanto, os genomas de arqueas que foram analisados não revelaram nenhuma sequência homóloga à proteína Rho e, também não foram encontradas estruturas que se assemelhem aos grampos de terminação bacterianos. Todavia, as arqRNAPs apresentam atividade de terminação intrínseca em resposta a sequências presentes no DNA-molde, mas não necessitam da formação de grampos de RNA como nas bacRNAPs. É possível que mecanismos específicos atuem a terminação da transcrição em arqueas, pois apesar da transcrição em seus mecanismos e componentes possuir semelhanças com os eucariotos, sua transcrição e tradução são acopladas como nas bactérias.

6.2 Terminação em eucariotos

O conhecimento sobre os mecanismos envolvidos na terminação da transcrição nos eucariotos é incompleto e menos detalhado que os da iniciação. A principal dificuldade é que os RNAs sintetizados sofrem modificações significativas, denominadas pós-transcricionais, também na sua extremidade 3' antes que sejam funcionais (Capítulo 11).

Geralmente, antes da terminação da transcrição, o RNA nascente é clivado em sua porção 3', portanto, as

sequências que demarcam a terminação da transcrição não estão presentes no RNA sintetizado. Na maioria dos casos, regiões de simetria invertida, como as que ocorrem em procarionotos, não são encontradas. Da mesma forma, proteínas relacionadas a Rho não foram encontradas em eucariotos.

A terminação da transcrição da RNAPIII é definida por sequências ricas em T no DNA. Esses oligômeros de T desestabilizam o complexo de transcrição e a RNAPIII se desliga do DNA-molde. Os transcritos sintetizados pela RNAPI são produtos de alto peso molecular e processados com rapidez, o que gera moléculas funcionais menores (Capítulo 11). A terminação da transcrição ocorre em uma região de cerca de 1.500 nt após a extremidade 3' dos RNAs processados. Sequências específicas ocorrem no DNA após o final da região correspondente ao RNA maduro. Essas sequências localizam-se em dois sítios, denominados T1 e T2. Algumas proteínas se ligam a estes sítios, onde também ocorre a clivagem do RNA que está sendo transcrito. Essa clivagem acontece de 15 a 50 nt após a extremidade 3' do RNA maduro e é mediada por uma endonuclease. A clivagem no RNA se assemelha àquela mediada pelo complexo de poliadenilação dos genes transcritos por RNAPII.

A terminação da transcrição dos RNAs sintetizados pela RNAPII sucede, de forma difusa, em uma região de aproximadamente 1.500 nt da extremidade 3' dos genes após a extremidade 3' dos RNAs processados. A clivagem do transcrito pelo complexo de poliadenilação determina o final da transcrição. Aparentemente as alterações que ocorrem durante o processamento da extremidade 3' dos mRNAs são mais relevantes, em termos de controle da expressão, do que o final da transcrição.

Existem dois modelos propostos para o mecanismo de terminação da transcrição dos genes da classe II. Esses modelos não são mutuamente exclusivos, podendo ocorrer em diferentes genes, e ambos dependem da clivagem do transcrito pelo complexo de poliadenilação. Um dos modelos baseia-se na existência de uma exonuclease 5' → 3', que é capaz de se ligar à extremidade 5' do RNA e promove a sua degradação em direção à extremidade 3'. Nesse modelo, denominado “torpedo”, a exonuclease se liga à extremidade 5' do RNA que está sendo transcrito, após a sua clivagem pelo complexo de poliadenilação. Então, a exonuclease migra em direção ao complexo de transcrição (lembrando o movimento do complexo Rho), degradando o RNA em direção à extremidade 3'. Se a velocidade desse processo for maior que a da síntese de RNA pela RNAPII, a degradação atingiria o complexo e a transcrição seria terminada. Existe interação entre a exonuclease e o domínio CTD da RNAPII, o que permite supor a ocorrência de alterações na RNAPII mediadas pelo domínio CTD. Quando a RNAPII se aproxima do final da transcrição a fosforilação de S₂ predomina sobre S₅ no domínio CTD. Essa condição promove o recrutamento de componentes do complexo de poliadenilação e da exonuclease, o que explicaria a natureza difusa da terminação da transcrição por RNAPII, pois o ponto de

contato entre a exonuclease e o complexo RNAPII dependeria de possíveis pausas na transcrição que diminuiriam a sua velocidade. Sequências localizadas na extremidade 3' do gene poderiam determinar as pausas ocasionando a terminação difusa da transcrição. O outro modelo é denominado “alostérico” e propõe que a clivagem do RNA nascente no sítio de poliadenilação desencadearia alterações conformacionais na RNAPII e na cromatina local, levando a desestabilização do complexo de transcrição e sua terminação.

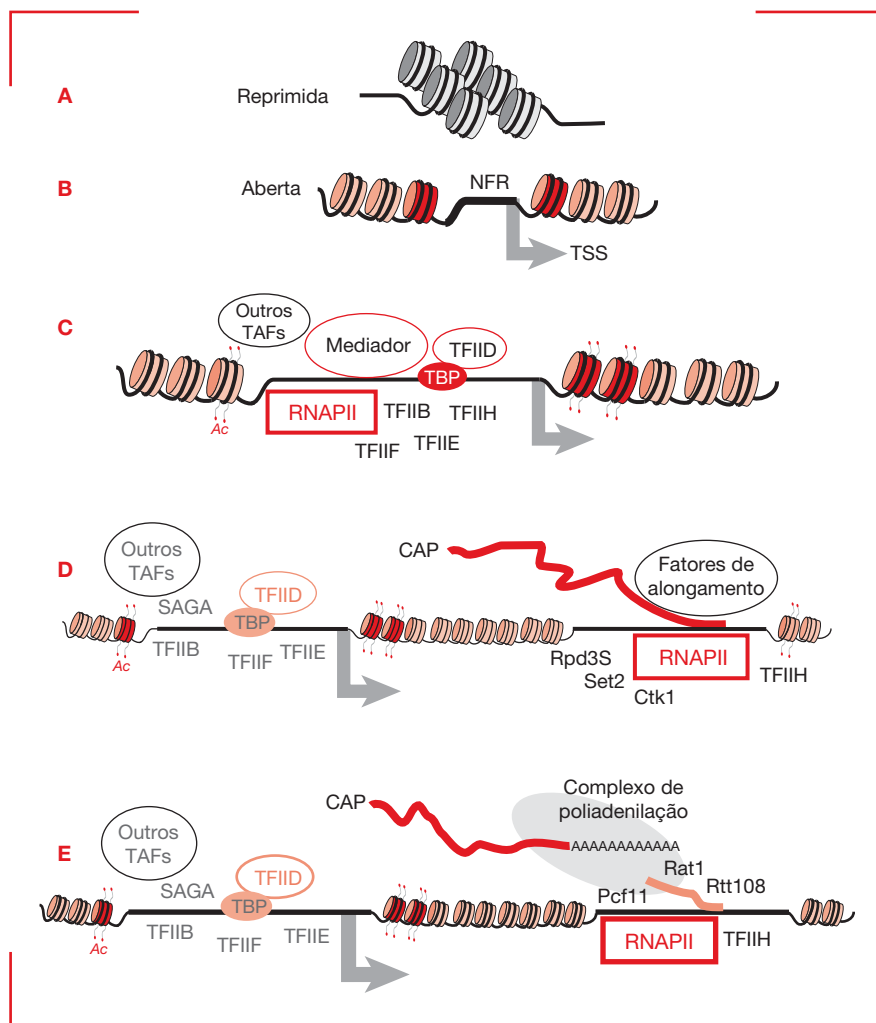
7. Transcrição na cromatina

Nos eucariotos os genes estão organizados no núcleo na forma de **cromatina** (Capítulo 3). Essa estrutura, fundamental para a estabilidade e a compactação do DNA, representa uma grande barreira para processos básicos celulares, principalmente a transcrição. A complexidade e a compactação da cromatina são obstáculos para que elementos (sequências ou motivos) sejam reconhecidos para o início, o alongamento e a terminação da transcrição. A cromatina pode assumir uma estrutura totalmente **reprimida**, pela sua compactação em fibras de 30 nanômetros. O estado reprimido da cromatina está associado à **metilação de histonas** e à ligação concomitante das proteínas de **heterocromatina**, que reconhecem histonas metiladas. Ainda não é conhecido se essas marcas de repressão, e suas proteínas cognatas são apenas um indicador da repressão ou se realmente promovem e mantêm a cromatina reprimida. Alguns **RNAs não codificadores** podem participar no estabelecimento de regiões de cromatina reprimida. A cromatina também pode ser inativada para a transcrição pela metilação do DNA.

Em seu estado denominado aberto, a cromatina apresenta uma estrutura mais favorável à transcrição (**Figura 10.8A e B**). Mesmo assim, os nucleossomos continuam a representar uma barreira para a transcrição. Esse bloqueio pelos nucleossomos foi demonstrado *in vitro*, pela inativação da transcrição na presença de nucleossomos e, por sua ativação *in vivo* quando da depleção das histonas. Esses dados foram comprovados por análises do padrão geral da transcrição e pela descrição dos muitos mecanismos de **remodelamento da cromatina** e suas relações com a expressão gênica.

A presença de nucleossomos no promotor dos genes é um impeditivo para o início da transcrição, pois bloqueia a ligação dos TFs. Entretanto, em descoberta recente, constatou-se que os nucleossomos não estão estocasticamente dispersos nos genomas. Demonstrou-se que a existência de **regiões livres de nucleossomos** (NFR, de *nucleosome-free region*) é uma característica comum dos promotores (Figura 10.8B). O início da transcrição é, portanto, regulado pelo posicionamento dos nucleossomos a distâncias específicas do TSS.

As regiões NFR dos promotores são flanqueadas por nucleossomos (ditos -1 e +1), que apresentam histonas

**Figura 10.8**

Cromatina e transcrição. (A) Cromatina reprimida (fechada). (B) Cromatina aberta. NFR são regiões livres de nucleossomos no promotor. TSS sítio de início da transcrição. (C) Formação do PIC no NFR. (D) Alongamento da transcrição. Note que a região do promotor permanece “comprometida” com TFs e outros complexos. (E) Término da transcrição.

alteradas e assim marcam os promotores. Esse padrão de organização é encontrado quando as posições dos nucleossomos dos genomas de leveduras, moscas e seres humanos estão alinhadas em relação ao TSS. As regiões NFR e o posicionamento uniforme dos nucleossomos em relação ao TSS possuem conservação evolutiva. Entretanto, não se conhece, com exatidão, os mecanismos que determinam a existência das NFR. Existem fortes evidências de que esse posicionamento seja controlado pela estrutura local do DNA e pela interação de proteínas com sequências específicas.

A ativação da transcrição na cromatina envolve o deslocamento dos nucleossomos de suas posições. Esses deslocamentos são rigorosamente controlados pelos denominados complexos de remodelamento da cromatina. Esses complexos pertencem a quatro famílias, de acordo com a sua sequência: SWI/SNF, SWRI/INO80, ISWI e CHD (ver Capítulo 14). Esses complexos são inespecíficos quanto a sequência do DNA e, portanto, são recrutados para as regiões específicas por interação com outras proteínas, que reconhecem as sequências específicas.

A organização da cromatina é alterada durante a transcrição, pois quando a transcrição é ativada, os nucleossomos ficam muito mobilizáveis e adotam localizações alternativas. Porém, a estrutura dos nucleossomos é preservada durante a transcrição. Com poucas exceções, em genes altamente transcritos não existem nucleossomos na região transcrita. A abertura das fitas de DNA e o movimento da RNAP na direção da transcrição deslocam os nucleossomos, que são imediatamente restaurados após a passagem do complexo.

Os complexos de remodelamento da cromatina utilizam a energia do ATP para alterar a estrutura, a posição e a composição dos nucleossomos. As voltas do DNA em torno dos nucleossomos, por ação dos remodeladores, formam alças translocadas e provocam o simples deslocamento ou a dissociação dos octâmeros de histona. Os nucleossomos adjacentes aos promotores são dinâmicos, sugerindo que os complexos de remodelamento estejam presentes e ativos em muitos promotores, mesmo naqueles não engajados na transcrição. Os nucleossomos -1 e $+1$ controlam a acessibilidade aos promotores e provavelmente são alvos dos complexos de remodelamento (Figura 10.8).

8. Fidelidade e correção de erros

A transcrição é um processo muito preciso, tanto em procariotos como em eucariotos, tendo uma estimativa de erro de um a cada cem mil ribonucleotídeos incorporados (10^{-5}). Para atingir esse nível de precisão, a maior estabilidade quando o ribonucleotídeo correto é pareado com o molde de DNA não é suficiente, necessitando também um mecanismo, descrito recentemente, para correção de erros.

Na subunidade β' das bacRNAP e nas correlatas em eucariotos existe um motivo, próximo ao centro catalítico das RNAPs, responsável pela captura do ribonucleotídeo correto complementar ao molde de DNA. Quando o ribonucleotídeo correto está ligado a esse motivo, a estabilidade do complexo aumenta dez vezes. A entrada de um ribonucleotídeo incorreto causa um desalinhamento, reduzindo a velocidade de incorporação. A entrada do ribonucleotídeo correto também é verificada no canal secundário da RNAP. Além disso, existem pelo menos dois mecanismos de correção de erros (ver Figura 10.1C): o primeiro se baseia na atividade de clivagem do transcrito da RNAP, estimulada por fatores de alongamento da transcrição em bactérias (GreA e GreB) e em eucariotos (TFIIS); e o segundo mecanismo de correção se baseia na atividade de exonuclease intrínseca da RNAP.

O primeiro mecanismo é conhecido como monitoramento (*backtracking*). A extremidade 3' terminal do RNA torna-se um substrato para a reação de clivagem, que é estimulada por fatores de clivagem. Essa reação depende da estabilidade do pareamento do RNA com o DNA-molde. Quanto mais instável o pareamento, maior é a atividade de clivagem, portanto a incorporação de um ribonucleotídeo errado estimula a correção pelo sistema de monitoramento.

O segundo mecanismo de correção de erros na transcrição parece ocorrer por atividade intrínseca de exonuclease da RNAP e resulta na remoção de um dinucleotídeo da extremidade do RNA. Embora sua participação exata

na fidelidade da transcrição ainda não seja totalmente conhecida, esse mecanismo já foi demonstrado *in vivo* em eucariotos (TFIIS e a subunidade Rpb9 da euRNAPII).

9. Fábricas de transcrição

Foi demonstrado, em vários modelos, que os organismos concentram ou agrupam moléculas para promover a sua interação, facilitando muitas das reações necessárias. Em particular, a existência de compartimentos especializados (organelas) para determinados processos demonstra essa tendência evolutiva. Esse tipo de macro-organização pode ser estendido a vários processos celulares, incluindo a transcrição. Essa organização hierárquica e espacial pode explicar como muitas das proteínas e complexos são atraídos nas diferentes fases da transcrição e podem permanecer ativando ou reprimindo determinados genes com maior eficiência.

A transcrição dos rRNAs pela RNAPI no nucléolo apresenta esse tipo de macro-organização. Para a transcrição por RNAPIII, o agrupamento espacial dos tRNAs também foi descrito. No caso da RNAPII, a tendência à organização de maior complexidade tem sido evidenciada na formação dos PICs (TFs, RNAP, mediador, remodeladores da cromatina) e na formação de complexos durante a fase de alongamento da transcrição, com a incorporação de complexos de processamento do RNA (adição do 5'-CAP, complexo de poliadenilação, spliceossomos).

A partir de evidências experimentais recentes, um modelo de organização hierárquica maior foi proposto para a transcrição. Esse modelo, denominado “fábricas de transcrição”, propõe que a transcrição ocorra em pontos focais no núcleo, formando estruturas ancoradas. Essas estruturas conteriam todos os componentes necessários para a transcrição e seu controle e as regiões do cromossomo a serem transcritas seriam translocadas para as “fábricas”. Existem ainda muitos questionamentos sobre esse modelo, mas provavelmente ele será detalhado nos próximos anos.



Resumo

A transcrição é o processo pelo qual todos os RNAs celulares são sintetizados, sendo a etapa inicial da expressão gênica e onde ocorre grande parte da regulação da expressão. As enzimas que catalisam esse processo são as RNA polimerases (RNAP), que apresentam complexidade crescente durante a evolução. Nos fagos, a RNAP é composta por uma proteína, ao passo que nas bactérias o complexo ativo contém seis subunidades ($\alpha_2\beta\beta'\omega\sigma$) e nas arqueas podem ser onze ou doze subunidades. Nesses sistemas há apenas uma RNAP que sintetiza todas as clas-

ses de RNA. Nos eucariotos existem três classes de euRNAPs nucleares especializadas: a euRNAPI que sintetiza rRNAs (35-45S pré-rRNA processado nos rRNAs 28S, 18S e 5,8S), a euRNAPII que transcreve todos os genes que codificam proteínas (mRNAs) e alguns RNAs pequenos nucleares (snRNAs), e a RNAPIII que sintetiza rRNA 5S, tRNAs e snRNAs. Os eucariotos contêm ainda RNAPs de organelas e as angiospermas possuem mais duas RNAPs. A reação catalisada é a adição de ribonucleotídeos trifosfatados no sentido 5'→3', utilizando uma das fitas do DNA

como molde, sem necessitar de iniciadores para começar a síntese.

A transcrição é dividida nas fases de início, alongamento e terminação. O início da transcrição ocorre em sequências definidas nos genomas, denominadas promotores, que devem ser adequadamente reconhecidos pelo complexo de transcrição. As bacRNAPs reconhecem os promotores e iniciam a transcrição, quando o fator σ faz parte do complexo principal (holoenzima). Por sua vez, as euRNAPs não reconhecem os promotores com facilidade, necessitando proteínas acessórias, os fatores de transcrição (TF). Alguns TFs reconhecem e ligam-se aos promotores e recrutam a euRNAP para montar os complexos de pré-iniciação (PIC), que podem iniciar a transcrição. Os promotores são mais simples em procariotos e muito complexos em eucariotos.

O início da transcrição ocorre pela desnaturação da hélice dupla do DNA, que expõe a sequência molde e o primeiro ribonucleotídeo é incorporado (+1, IT ou TSS). Estabilizado o complexo de iniciação, é iniciada a fase de alongamento da cadeia em que o fator σ , nas bactérias, e a maioria dos TFs, nos eucariotos, é desligada do complexo e a síntese prossegue até o término da transcrição. O desenrolamento do DNA gera supertorção no DNA a jusante que é neutralizada por topoisomerases, também atuando a montante refazendo o superenrolamento original. Nos genes da classe II (mRNAs eucarióticos), nesta fase, ocorre a adição do 5'-CAP e alguns fatores de poliadenilação são ligados ao complexo.

A transcrição sofre várias pausas durante a fase de alongamento e nos eucariotos é concomitante com a excisão dos íntrons pelos spliceossomos. As RNAPs têm capacidade de corrigir erros durante a transcrição e removem o RNA já sintetizado em complexos estagnados antes de reiniciar a transcrição. Esses mecanismos de correção de erro são denominados *back-tracking* e, como na polimerização, são atividades intrínsecas das RNAPs.

A terminação da transcrição nas bactérias é sinalizada por sequências no RNA sintetizado, reconhecíveis na sequência do DNA (terminadores intrínsecos), ou ocorre por atividade da proteína Rho que atua no complexo de transcrição, provocando a sua dissociação. Nos eucariotos a terminação da transcrição varia de acordo com a classe de euRNAP, mas em geral ocorre por pausas longas e sequências ricas em Ts. A terminação da transcrição em eucariotos tende a ser dispersa, em uma região com mais de mil nucleotídeos do terminal 3' dos genes. Na fase de terminação da transcrição o complexo é dissociado, liberando o RNA e a RNAP, e o DNA é renaturado.

A cromatina representa uma barreira para a transcrição em eucariotos, e os nucleossomos são repositionados durante o alongamento da transcrição. Nos promotores existem regiões livres de nucleossomos (NFR), o que facilita seu reconhecimento pelos TFs. Vários complexos participam do remodelamento da cromatina para a transcrição. A transcrição pode ocorrer em regiões de macro-organização denominadas “fábricas de transcrição”.



Leituras recomendadas

Carninci P. RNA dust: where are the genes? *DNA Res.* 2010;17(2):51-9.

Goldman SR, Ebright RH, Nickels BE. Direct detection of abortive RNA transcripts in vivo. *Science.* 2009;324(5929):927-8.

Kornberg RD. The molecular basis of eucaryotic transcription. *Cell Death Differ.* 2007;14(12):1989-97.

Revyakin A, Liu C, Ebright RH, Strick TR. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science.* 2006;314(5802):1139-43.

Roberts JW, Shankar S, Filter JJ. RNA polymerase elongation factors. *Annu Rev Microbiol.* 2008;62:211-33.

Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet.* 2007;8(6):424-36.

Sutherland H, Bickmore WA. Transcription factories: gene expression in unions? *Nat Rev Genet.* 2009;10(7):457-66

Sydow JF, Cramer P. RNA polymerase fidelity and transcriptional proofreading. *Curr Opin Struct Biol.* 2009;19(6):732-9.

Thomas MC, Chiang CM. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol.* 2006;41(3):105-78.

Venters BJ, Pugh BF. How eukaryotic genes are transcribed. *Crit Rev Biochem Mol Biol.* 2009;44(2-3):117-41.

Wade JT, Struhl K. The transition from transcriptional initiation to elongation. *Curr Opin Genet Dev.* 2008;18(2):130-6.

Charley Christian Staats

Mecanismos de Processamento e Maturação de RNA

1. Mecanismos de processamento de precursores de mRNAs 234

- 1.1 Adição de 5'-*cap* em hnRNAs 234
- 1.2 Poliadenilação em hnRNAs 236
- 1.3 *Splicing* em hnRNAs 237
- 1.4 *Splicing* alternativo 240
- 1.5 *Trans-splicing* 243

2. Mecanismos de processamento de rRNA 244

- 2.1 Processamento e maturação de transcritos primários de rRNAs em procariotos 244
- 2.2 Processamento e maturação de transcritos primários de rRNAs em eucariotos 245

3. Mecanismos de processamento de tRNA 246

- 3.1 Geração de tRNAs clivados e adição de 3' CCA 246

- 3.2 Modificações de bases nos tRNAs 246
- 3.3 *Splicing* em tRNAs 247

4. Mecanismos de *auto-splicing* 248

- 4.1 *Auto-splicing* associado a íntrons do grupo I 248
- 4.2 *Auto-splicing* associado a íntrons do grupo II 249
- 4.3 *Auto-splicing* associado a íntrons do grupo III 250
- 4.4 Mobilidade de íntrons 250

5. Mecanismos de edição de RNA 250

6. Degradação de RNA 252

O processo de transcrição, especialmente em eucariotos, geralmente não produz moléculas de RNA completamente funcionais. Os transcritos primários, também denominados pré-RNAs ou precursores de RNAs, são os produtos diretos da transcrição e, com frequência, não estão associados a nenhuma atividade biológica. Antes de realizar as funções específicas de cada RNA, as moléculas de pré-RNA devem passar por uma série de reações que tendem a alterar a sua sequência primária. Estas alterações são conjuntamente denominadas **processamento de RNA (ou maturação do RNA)**, compreendendo reações como a adição, a remoção ou até mesmo a modificação de nucleotídeos (**Figura 11.1**).

As três principais classes de RNA (tRNAs, rRNAs e mRNAs) são processadas para resultar em RNAs ativos. Tanto em organismos procarióticos quanto em eucarióticos, a maturação de rRNAs e tRNAs envolve diversas reações de clivagem para liberação das moléculas de RNA funcional e também modificações de bases. Os mRNAs de procariotos sofrem poucas alterações, muito mais complexas nos eucariotos. Especificamente em eucariotos, os hnRNAs (do inglês, *heterogeneous nuclear*), moléculas de RNAs nucleares precursoras dos mRNAs, são alvo de muitas modificações para sua maturação. Essas modificações ocorrem nas extremidades 5' e 3' e no interior dos hnRNAs. Conjuntamente à transcrição dos hnRNAs, ocorre a adição de nucleotídeos nas extremidades 5' e 3', denominadas respectivamente de adição de 5'-cap e poliadenilação. Outra modificação fundamental para a maturação de hnRNAs é o *splicing*, modificação caracterizada pela retirada ou excisão das sequências intervenientes (íntrons) presentes nas moléculas de hnRNA. O *splicing* depende de proteínas e de outras moléculas de RNA que

não podem ser enquadradas nas três classes acima descritas (tRNAs, rRNAs e mRNAs). Estas moléculas catalíticas de RNA, também denominadas de **ribozimas**, têm corroborado teorias evolutivas de que organismos ancestrais baseavam o fluxo da informação apenas em RNA.

1. Mecanismos de processamento de precursores de mRNAs

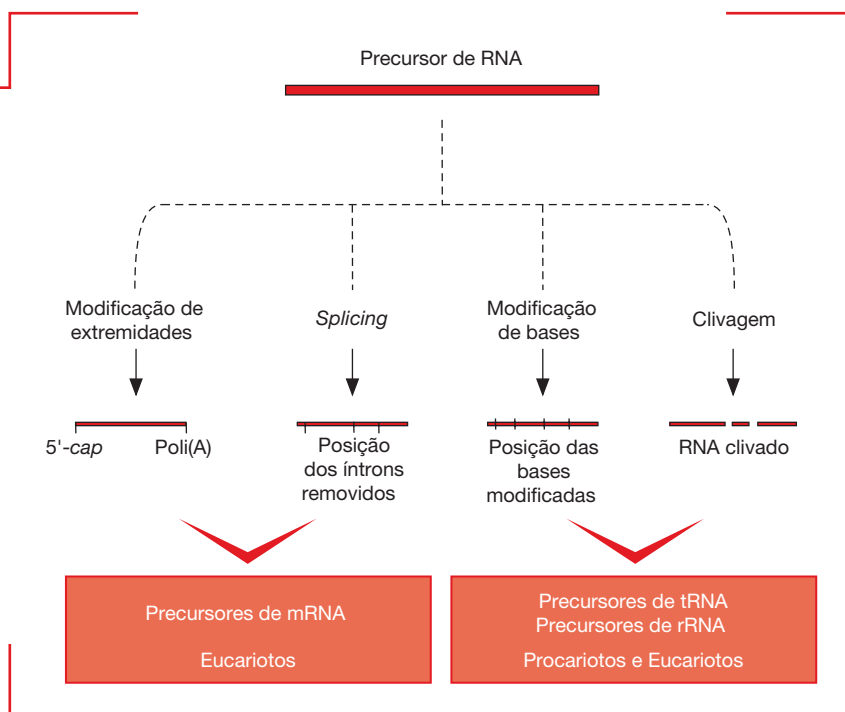
Existem distinções muito claras quando os eventos que governam o destino de mRNAs são comparados em organismos procarióticos e eucarióticos. Em procariotos, logo após a transcrição, os mRNAs já são reconhecidos por ribossomos, o que leva imediatamente à síntese proteica. Por outro lado, em organismos eucarióticos, os hnRNAs devem passar por uma série de processos antes que possam se associar aos ribossomos para servirem como molde para a síntese de proteínas. Estas alterações incluem: (i) formação do 5'-cap, que é a adição de um nucleotídeo na extremidade 5' do RNA, podendo ocorrer metilação de alguns ribonucleotídeos; (ii) a formação da cauda de poli(A) pela adição de uma cauda de adenilatos na porção 3' do hnRNA; e (iii) a remoção dos íntrons e o transporte para o citoplasma (**Figura 11.2**).

1.1 Adição de 5'-cap em hnRNAs

No início da transcrição, as moléculas de hnRNA sendo transcritas são caracterizadas pela presença de um nucleotídeo trifosfatado (geralmente pppG ou pppA). Uma

Figura 11.1

Modificações comumente encontradas no processo de maturação de RNAs.



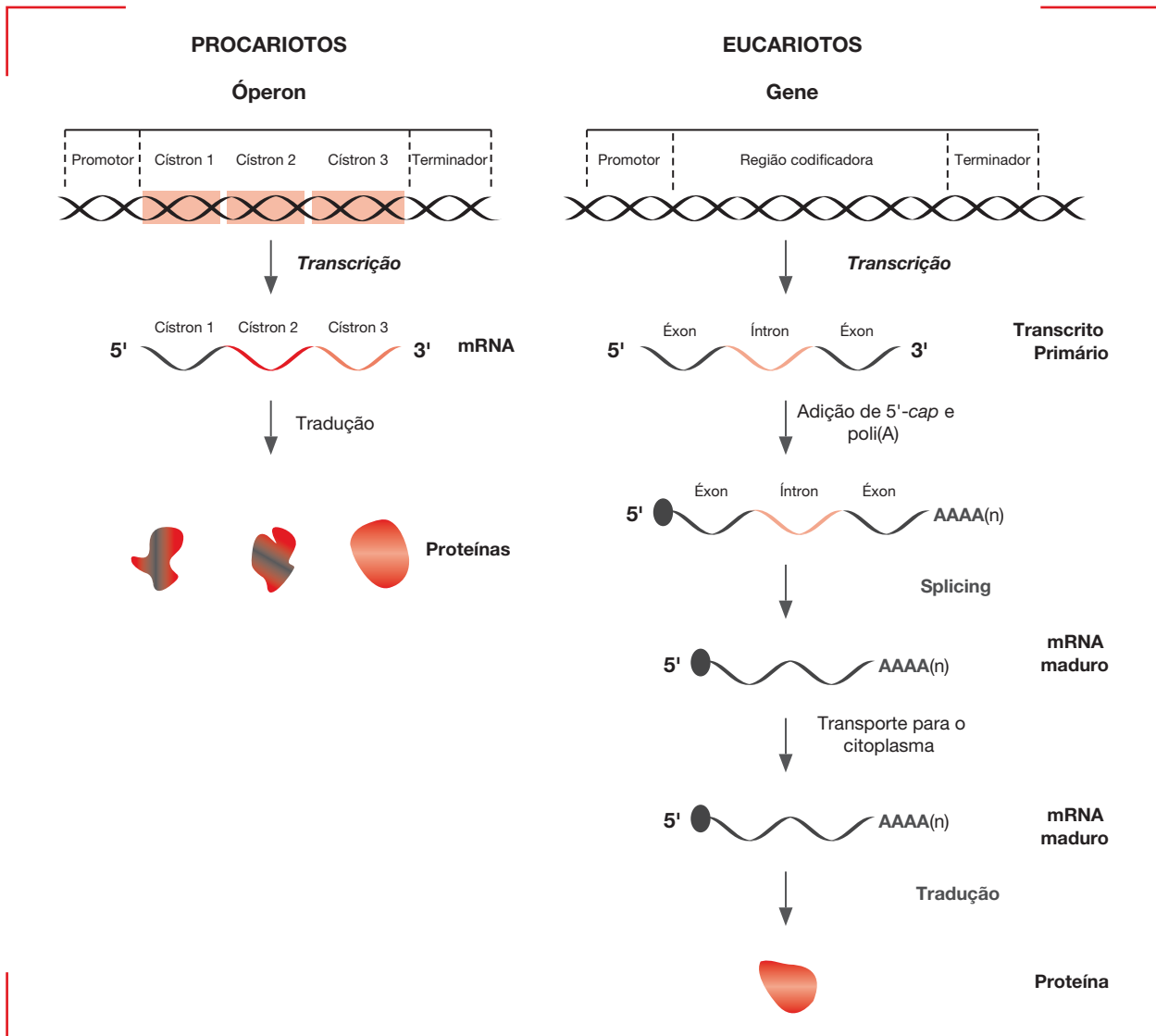


Figura 11.2

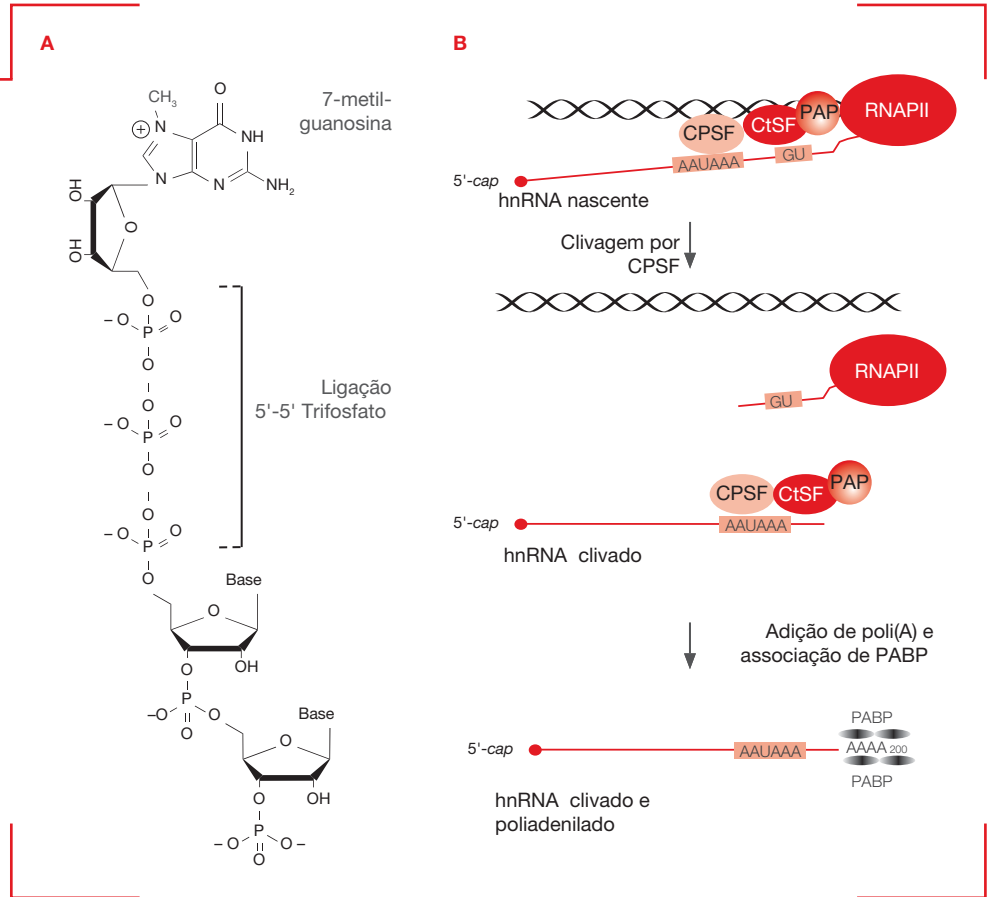
Processo de expressão de proteínas em procariotos e eucariotos. O painel à esquerda esquematiza um óperon característico de procariotos, a delimitação de cada cístron, a síntese de um RNA policístrônico e a produção de proteínas de acordo com o número de cístrons presentes neste RNA. Este processo acontece de maneira acoplada, sendo o RNA produzido rapidamente reconhecido pelos ribossomos, propiciando a síntese proteica. Em eucariotos, esquematizado no painel à direita, é mostrado um esquema generalizado de um gene de eucariotos e de seu precursor de mRNA, caracterizado pela presença de íntrons e éxons. Diversas etapas de modificações devem ser realizadas no precursor do mRNA para que seja alcançada a sua maturação e consequente exportação para o citoplasma, onde geralmente se dá início a síntese proteica.

das estratégias desenvolvidas pelas células para que hnRNAs não sejam rapidamente degradados por exonucleases ocorre pela proteção da extremidade 5' do hnRNA por adição de um nucleotídeo utilizando uma ligação pouco comum. A adição de uma molécula de 7-metil-guanosina (**5'-cap**) na extremidade 5' do hnRNA é a primeira etapa de modificação pós-transcricional que ocorre em hnRNAs, sendo extremamente importante para a expressão gênica, pois além de determinar as taxas adequadas de transcrição do gene, participa da associação com os ribossomos, o que determina a sua tradução (ver Capítulo 12).

A formação do 5'-cap (**Figura 11.3A**) é realizada em duas etapas. Conjuntamente com a transcrição ocorre a formação de uma ligação 5' – 5' entre os fosfatos do primeiro nucleotídeo transcrito e de um guanidilato (7-metil-guanosina). A enzima responsável pela catálise desta reação é a guanilil-transferase, sendo necessárias diversas atividades enzimáticas associadas para a formação dos intermediários. O grupo fosfato γ presente na extremidade 5' do RNA nascente é removido pela atividade de uma fosfo-hidrolase. Em seguida, a enzima guanilil-transferase realiza a hidrólise da ligação fosfodiéster entre os fosfa-

Figura 11.3

Modificações por adição nas extremidades de mRNA. (A) A adição de 5'-cap (painel à esquerda), uma molécula de 7 metil-guanosina, à extremidade 5' do precursor de mRNA se dá por uma ligação pouco comum 5'-5' trifosfato e se dá logo no início da transcrição. (B) A adição da cauda de poli(A) à extremidade 3' do precursor de mRNA está associada ao término da transcrição. Esta se dá em parte devido à existência de sequências canônicas reconhecidas por um complexo que cliva o precursor de RNA ainda sendo transcrito (CPSF e CtsF). A adição da cauda de poli(A) se dá devido à atividade de uma polimerase (PAP) e à sua estabilização por proteínas ligadoras de poli(A) –PABP.



tos α e β do GTP (ribonucleosídeo trifosfatado), liberando pirofosfato, e também catalisa a ligação entre o GMP (ribonucleosídeo monofosfatado) resultante e o RNA nascente cujo fosfato γ tenha sido previamente hidrolisado. A segunda etapa consiste na metilação desta estrutura. Muitas etapas de metilação podem acontecer no 5'-cap, mas a predominante e diretamente envolvida com a maturação de hnRNAs se refere à metilação na posição 7 da base nitrogenada adicionada. Esta reação é catalisada por uma enzima denominada guanina-7-metiltransferase. Estas duas enzimas (guanina-7-metiltransferase e guanilil-transferase) estão associadas ao complexo da RNA-polimerase II (RNAPII – ver Capítulo 10), pela associação com o seu domínio carboxiterminal (CTD). Assim, após a pausa na transcrição que ocorre aproximadamente 30 nucleotídeos a jusante do sítio de início, a RNAPII recruta as enzimas responsáveis pela adição do 5'-cap à extremidade do RNA nascente. Desta forma, conjuntamente à transcrição, esta modificação pós-transcricional confere proteção contra a atividade de 5' – 3' exonucleases, que poderiam levar à degradação do transcrito nascente. Adicionalmente à proteção da degradação mediada por nucleases, o 5'-cap tem função importante na etapa de alongamento da transcrição, no *splicing*, na poliadenilação, no transporte de mRNAs para o citoplasma, na estabilidade do mRNA no citoplasma e na tradução.

1.2 Poliadenilação em hnRNAs

Grande parte dos mRNAs eucarióticos possui em sua extremidade 3' uma sequência com cerca de duas centenas de adenilatos, denominada de **cauda de poli(A)**. Diferentemente de outros RNAs, este grande trecho de RNA não é transcrito pela RNAPII, mas é uma modificação pós-transcricional dependente da atividade da enzima **poli-A polimerase** e outras enzimas, e tem relação direta com o término da transcrição. A adição da cauda de poli(A) é um evento necessário para formação de mRNAs maduros, uma vez que está associado a fatores como a estabilidade, já que previne a degradação do mRNA por 3' – 5' exonucleases, assim como no transporte para o citoplasma. Apesar da cauda de poli(A) ser geralmente associada a hnRNAs eucariotos, os quais possuem uma importante atividade contra a degradação, transcritos poliadenilados também podem ser encontrados em organismos procarióticos. Entretanto, ainda não são plenamente conhecidas as implicações deste tipo de modificação na funcionalidade de mRNAs em procariotos. Nos organismos eucarióticos, uma exceção importante são os genes de histonas de alguns organismos, cujos mRNAs, em geral, não apresentam a cauda de poli(A).

A adição da cauda de poli(A) é uma modificação pós-transcricional com relação direta com o término da

transcrição, sendo realizada em diversos passos por um grande complexo proteico associado ao domínio carboxi-terminal da RNAPII. A região de término da transcrição é caracterizada pela presença de algumas sequências em *cis* que, uma vez transcritas, sinalizam para a clivagem e a poliadenilação. O sítio onde o hnRNA nascente será clivado e poliadenilado é caracterizado por duas sequências conservadas: 5' AAUAAA 3', localizada a uma distância de 10 a 30 nucleotídeos a montante do sítio de clivagem, e uma sequência rica em Gs e Us, cuja localização varia de 20 a 40 nucleotídeos a jusante do sítio de clivagem. A RNAPII transcreve além do ponto onde ocorre a clivagem e a poliadenilação e essas sequências presentes na extremidade 3' do hnRNA nascente então sinalizam para o complexo proteico, associado à RNAPII, realizar a clivagem e a poliadenilação. A região 5' AAUAAA 3' é reconhecida pelo complexo CPSF (fator de clivagem e poliadenilação – do inglês, *cleavage and polyadenylation specific factor*), e a região rica em GU é reconhecida pelo complexo CstF (fator estimulatório de clivagem – do inglês, *cleavage stimulation factor*) (Figura 11.3B). Estes dois complexos têm a capacidade de interagir entre si e estimular a afinidade da ligação nas suas respectivas sequências. No complexo CPSF, a subunidade de 73 kDa possui atividade de endonuclease, sendo a responsável pela hidrólise da ligação fosfodiéster do ponto de clivagem. Em seguida, a extremidade 3' gerada pela clivagem é alvo da poli(A) polimerase, a qual é responsável pela síntese da cauda de poli(A). Esta modificação ocorre em duas etapas. A primeira é caracterizada pela atividade propriamente dita da poli(A) polimerase, que é dependente do reconhecimento da sequência *cis* 5' AAUAAA 3' para adição da cauda de poli(A). A segunda é realizada por um conjunto de proteínas nucleares denominadas PAPB (proteína ligadora de poli(A) – do inglês, *poly(A) binding protein*), que se associam à cauda de poli(A), permitindo a extensão, por parte da poli(A) polimerase, até o comprimento de aproximadamente 200 resíduos (Figura 11.3B). A restrição deste tamanho se dá pela PAPB, por mecanismos ainda não completamente compreendidos.

1.3 *Splicing* em hnRNAs

A primeira noção da existência de íntrons foi originada por meio da análise do sequenciamento de genes eucarióticos, realizada por Phillip Sharp e Richard Roberts, no ano de 1977. Estes dados diferiam substancialmente daqueles obtidos por sequenciamento de genes procarióticos, caracterizados pela colinearidade entre a sequência gênica e seu produto proteico (ver Capítulos 4 e 5). Neste contexto, foi possível definir o conceito de íntrons, sequências intervenientes presentes nos precursores de mRNA que devem ser excisadas para união dos éxons e posterior síntese da proteína de maneira correta. Genes eucarióticos, em especial de organismos multicelulares, são ricos em íntrons. Por exemplo, o gene da distrofina de humanos possui 2.400 kb e 78 íntrons, os quais constituem 98% da sequência codificante deste gene. Com a disponibilidade de dados de sequenciamento de genes, genomas e de transcritos maduros, foi possível definir,

por análises bioinformáticas, que a grande maioria dos íntrons presentes em hnRNAs pode ser enquadrada em dois grupos distintos, de acordo com a sequência que determina o seu início e o seu fim: os íntrons 5' GU..AG 3', denominados íntrons U2, e os íntrons 5' AU..AC 3', denominados íntrons U12 (Figura 11.4A). Apesar de muito parecidos quanto ao mecanismo geral na sua excisão, estes dois tipos de íntrons requerem mecanismos com particularidades próprias, devendo ser removidos do hnRNA, com consequente união dos éxons de uma maneira ordenada e correta em um processo denominado ***splicing*** (Figura 11.4B). Desta forma, o mRNA maduro pode ser exportado para o citoplasma e então traduzido e/ou degradado. A localização dos íntrons em uma sequência gênica pode variar, sendo possível estar presente nas regiões 5' e 3' não traduzidas (5'- e 3'-UTR, do inglês, *untranslated region*, respectivamente), assim como na região codificadora, de modo muito comum. Quando presente na região codificadora, um íntron pode ou não interromper um códon. Assim, se a inserção estiver entre dois códons, este íntron é denominado de íntron de fase 0. No caso de interromper os dois primeiros ou dois últimos nucleotídeos de um códon, eles serão denominados íntrons de fase 1 ou íntrons de fase 2, respectivamente.

O *splicing* em íntrons 5' GU..AG 3', com distribuição ubíqua em eucariotos, não depende apenas das sequências que o delimitam. Outros motivos conservados são reconhecidos pela maquinaria para que os mesmos sejam excisados. Os íntrons 5' GU..AG 3' são caracterizados pela sequência 5' AG-GURAGU 3' (o traço separa a sequência do éxon a montante e o íntron) como sítio de *splicing* 5'. Já a sequência característica do sítio de *splicing* 3' é 5' YAG-G 3' (o traço separa a sequência do íntron e o éxon a jusante). Há ainda o sítio de ramificação, caracterizado pela sequência 5' CURACU 3' (o nucleotídeo sublinhado representa o adenilato do sítio de ramificação). Adicionalmente, em eucariotos superiores, pode existir uma região rica em pirimidinas (Py) justaposta ao sítio de *splicing* 3'. Os íntrons 5' AU..AC 3', que não existem em alguns organismos como *Drosophila melanogaster* e *Saccharomyces cerevisiae*, possuem como sítio de *splicing* 5' a sequência 5' N-AUAUCCUG 3' (o traço separa a sequência do éxon a montante e o íntron), como sítio de *splicing* 3' a sequência 5' PyAC-N 3' (o traço separa a sequência do íntron e o éxon a jusante), e como sítio de ramificação a sequência 5' CCUUPuAC 3' (o nucleotídeo sublinhado representa o adenilato do sítio de ramificação) – Figura 11.4A. A denominação do tipo de íntron U2 ou U12 provém do mecanismo de remoção de íntrons mediada por um grande complexo ribonucleoproteico denominado ***spliceossomo***.

Em organismos eucarióticos, as reações de excisão de íntrons são realizadas pelos ***spliceossomos***, cuja massa molecular pode atingir aproximadamente 12 milhões de Daltons. Existem basicamente 5 snRNAs (RNAs nucleares pequenos, do inglês, *small nuclear RNAs* – ver Capítulo 2) que, associados a proteínas específicas, originam os ***snRNPs*** (pequenos complexos ribonucleoproteicos nucleares, do inglês, *small nuclear RNPs*). De acordo com

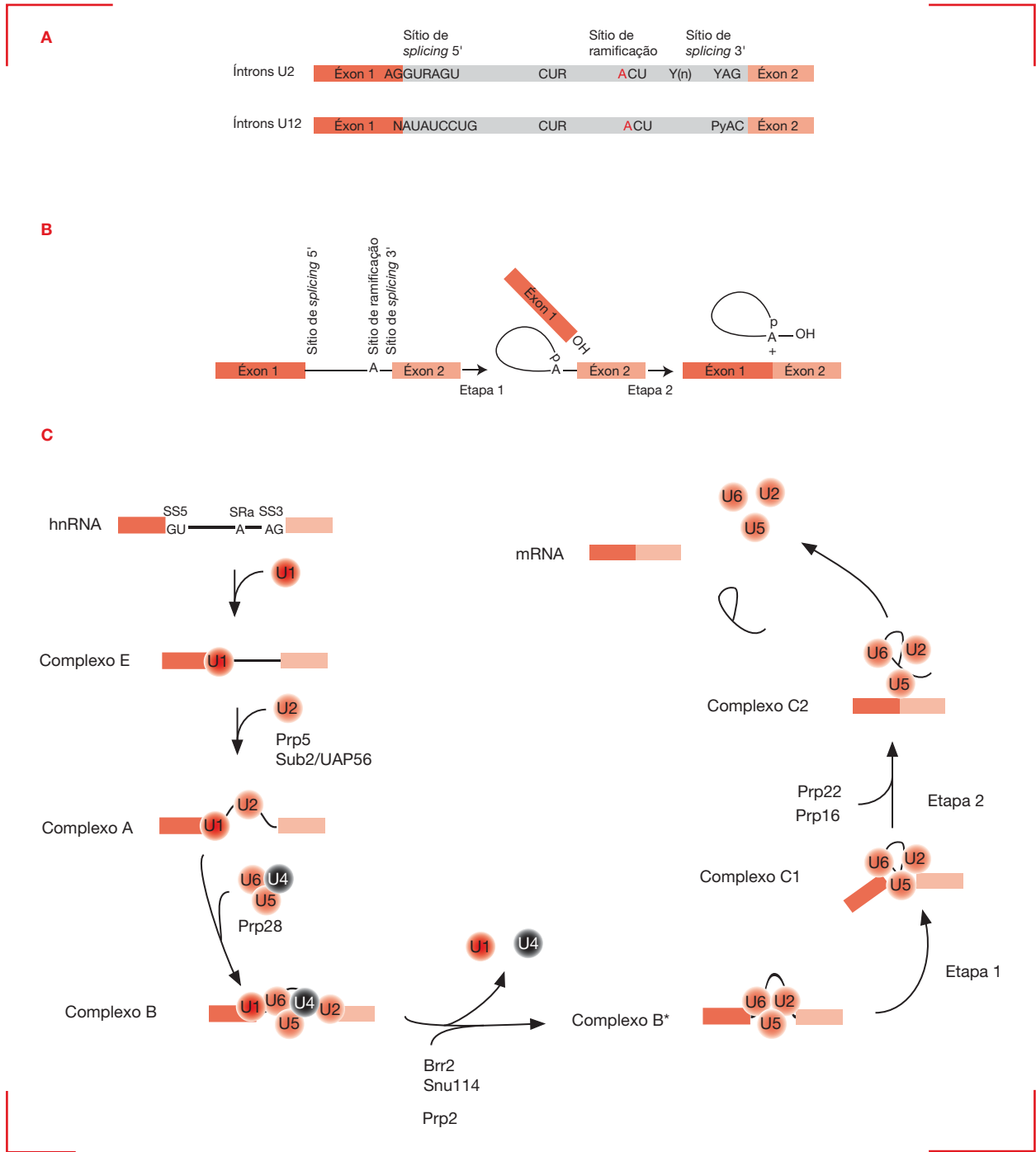


Figura 11.4

Remoção de íntrons e formação do spliceossomo. (A) A delimitação de íntrons e éxons se dá por seqüências conservadas nas extremidades de cada íntron, as quais são diferentes de acordo com o spliceossomo que processa estes tipos de íntrons, sendo classificados em íntrons U2 e U12. Além das extremidades dos íntrons, outros elementos podem estar presentes, como trechos ricos em pirimidinas (Y). (B) Esquema geral do processo de *splicing*, o qual é caracterizado por duas reações de transesterificação. A primeira reação (etapa 1) se dá pelo ataque nucleofílico no sítio de *splicing* 5' pela adenosina presente no sítio de ramificação. A segunda reação se dá pelo ataque nucleofílico no sítio de *splicing* 3' pela hidroxila do último nucleotídeo do éxon a montante, com conseqüente união dos éxons e liberação do íntron. (C) A formação do spliceossomo se dá de maneira orquestrada e dependente de pareamento de bases entre o precursor de mRNA e snRNAs (presentes nos snRNPs) ou entre diferentes snRNAs. Diferentes snRNPs (U1, U2, U4, U5 e U6) e proteínas se associam ao sítio de *splicing* 5' (SS5) e 3' (SS3), assim como ao sítio de ramificação (SRa), para promover a excisão do íntron e a união dos dois éxons.

os snRNAs que carregam, os snRNPs são denominados U1, U2, U5, U4, U6, U11 e U12. A composição dos spliceossomos formados varia em função do tipo de íntron a ser excisado: íntrons do tipo 5' GU..AG 3' estão associados a spliceossomos contendo os snRNPs U1, U2, U5, U4 e U6, ao passo que íntrons do tipo 5' AU..AC 3' estão associados a spliceossomos contendo os snRNPs U11, U12, U5 e as variações U4atac e U6atac. Estes grandes complexos são estruturas dinâmicas, que somente figuram como spliceossomos após o agrupamento ordenado de uma série de proteínas, snRNPs e outras moléculas em um hnRNA a ser processado.

O processo geral de *splicing* de hnRNAs engloba apenas duas reações de transesterificação. A primeira reação é caracterizada pela clivagem do sítio de *splicing* 5' (sítio doador) por intermédio de uma reação de substituição nucleofílica por parte da hidroxila ligada ao carbono 2' de um adenilato localizado na sequência do íntron, mais especificamente conhecido como sítio de ramificação. Como resultado, ocorre a hidrólise da ligação fosfodiéster entre os nucleotídeos do sítio de *splicing* 5' e concomitante formação de uma ligação fosfodiéster pouco comum, 5' – 2' entre o primeiro guanidilato do íntron e o adenilato do sítio de ramificação, lembrando a estrutura de um laço (do termo em inglês *lariat*) (Figura 11.4B). A segunda reação de transesterificação culmina na hidrólise da ligação fosfodiéster do sítio de *splicing* 3' (também conhecido como sítio aceptor) por parte da hidroxila ligada ao carbono 3' do éxon a montante, gerado na primeira reação de transesterificação. Com esta reação, os dois éxons são unidos e ocorre a liberação do íntron na forma de um laço, que é posteriormente degradado.

O spliceossomo é montado sobre um íntron de maneira conjunta à transcrição, envolvendo diversas etapas com o recrutamento de diversos snRNPs e proteínas acessórias. Assim que o primeiro íntron de um gene é transcrito, o snRNP U1 reconhece e se associa ao sítio de *splicing* 5'. Isto ocorre pela hibridização entre parte do snRNA presente no snRNP U1 que é complementar ao início da sequência do íntron. Este pareamento é estabilizado pela associação do complexo U2AF/Mud2 à região rica em pirimidinas e da proteína BBP (proteína ligadora de sítio de ramificação, do inglês, *branchpoint binding protein*) ao sítio de ramificação. Este primeiro complexo, cuja formação é independente de ATP, é denominado complexo E (inicial, do inglês, *early*). Adicionalmente, especialmente em eucariotos superiores, há a associação de proteínas da família SR, caracterizadas por um alto conteúdo de Serina e Arginina e por motivos de ligação ao RNA, que participam não apenas desta etapa da montagem do spliceossomo, mas também de várias etapas do *splicing*. Este conjunto de moléculas forma um complexo com grande força de interação, já que as proteínas SR são capazes de se associar tanto com o snRNP U1 quanto com U2AF (Figura 11.4C).

Uma vez montado o complexo E, há a associação ordenada de diversas outras proteínas, assim como de alguns snRNPs. O snRNP U2 se associa ao spliceossomo

em formação por meio da hibridização de parte de seu snRNA com a região equivalente ao sítio de ramificação. Esta etapa, dependente de ATP, leva à formação do complexo A. Com a associação de um heterotrímero contendo os snRNPs U4, U5 e U6, o complexo passa a ser denominado B, ou spliceossomo, pois todas as moléculas necessárias para a realização do *splicing* estão presentes. Entretanto, para que a excisão dos íntrons de fato ocorra, faz-se necessário a dissociação do snRNP U1, permitindo que os demais componentes, em especial o snRNP U6, fiquem intimamente próximos ao sítio de *splicing* 5'. O spliceossomo passa a ter atividade catalítica com a liberação dos snRNP U4, o que permite que o snRNP U6 possa de fato se associar e interagir com o sítio de *splicing* 5'. Evidências apontam para um papel regulador da atividade do snRNP U6 por parte do snRNP U4. Por isto, o complexo passa a ser denominado C1, também denominado de spliceossomo ativado, ocorrendo então a primeira reação de transesterificação, com a clivagem do sítio de *splicing* 5' e a formação do laço. O complexo C2 é caracterizado pela catálise da segunda reação de transesterificação, com a clivagem do sítio de *splicing* 3' e consequente união dos éxons. Assim, há liberação do íntron na forma de laço e do mRNA processado (Figura 11.4C).

Os mecanismos acima apresentados referem-se ao *splicing* de íntrons 5' GU..AG 3' (tipo U2). Para os hnRNAs que possuem íntrons 5' AU..AC 3', o processo é muito parecido, mas diferentes snRNPs participam da excisão do íntron. Quando os spliceossomos associados a cada tipo de íntron são comparados, o único snRNP em comum é o U5. A função dos snRNPs U1 e U2 são desenvolvidas por U11/U12-snRNP, e as demais funções são desempenhadas por snRNP U4atac/U6atac.

O spliceossomo deve reconhecer de forma exata os sítios de *splicing* 5' e 3' associados a um determinado íntron. Caso contrário, regiões que devem ser incluídas como éxons podem ser simplesmente não processadas e consequentemente ser retiradas da sequência final do mRNA maduro, ou ainda uma determinada sequência pode ser reconhecida erroneamente como sítio de *splicing* 3'. Tais processos aberrantes de *splicing* são denominados *exon skipping* e sítio de *splicing* críptico. Como resultado destas possíveis falhas de reconhecimento dos spliceossomos, moléculas de mRNA maduras muito diferentes seriam geradas, levando à tradução de proteínas defeituosas. Neste contexto, o spliceossomo depende das proteínas SR também para a adequada seleção do sítio de *splicing* a ser processado para excisão de um determinado íntron da sequência do hnRNA. Isto se dá em parte pela associação de proteínas SR com elementos *cis* denominados **reforçadores de *splicing* exônicos** (ESEs – do inglês, *exonic splicing enhancers*), indicando que a montagem do spliceossomo não depende apenas da delimitação do íntron pelo reconhecimento de seus limites, mas também pela influência das sequências presentes nos éxons adjacentes. Existem ainda outros elementos com função reguladora do *splicing*, como os **reforçadores de *splicing* intrônicos** (ISEs – do inglês, *intronic splicing enhancers*), os **silenciadores de *splicing***

mRNAs maduros oriundos de eventos de *splicing* alternativo a partir do transcrito primário deste gene. Como resultado, possivelmente 500 proteínas diferentes podem ser produzidas a partir do mesmo gene. A importância do *splicing* alternativo é tão grande que resultados recentes apontam que em mais de 95% dos genes de mamíferos possa ocorrer *splicing* alternativo, afetando diretamente diversos processos regulatórios, incluindo a modificação de cromatina (ver Capítulo 3) e a transdução de sinal.

Diversas estratégias para a geração da variabilidade de mRNAs já foram descritas. Podem ocorrer a retenção de íntrons, a seleção alternativa de sítios de *splicing* 5' e/ou 3', a inclusão ou exclusão de éxons, a utilização distinta de promotores associados ao *splicing* alternativo, e ainda a seleção diferencial de sítios de poliadenilação associados ao *splicing* alternativo (Figura 11.6A). Desta forma, pode-se classificar os éxons como éxons constitutivos, presentes em todos os mRNAs maduros produzidos a partir do mesmo gene, ou éxons facultativos, os quais podem ser removidos da sequência do hnRNA durante o processo de *splicing*. Com base nestas variações, diferentes proteínas podem ser sintetizadas, podendo desempenhar funções semelhantes, mas com localização espaço-temporal diferencial. Por exemplo, o gene de mamíferos *CaMKII δ* , que codifica para uma quinase dependente de cálcio, possui em seu hnRNA três éxons (14, 15 e 16) que estão sujeitos a *splicing* alternativo. A exclusão do éxon 14 origina uma proteína com direcionamento para a membrana, ao passo que a exclusão mútua dos éxons 15 e 16 ou dos éxons 14, 15 e 16 gera proteínas endereçadas para o núcleo ou para o citoplasma, respectivamente. Com a disponibilidade cada vez maior de sequências genômicas e transcriptômicas, uma grande quantidade de exemplos de *splicing* alternativo e a função de seus produtos gerados têm sido descritos. Entretanto, um dos exemplos mais bem caracterizados e ilustrativos reside no *splicing* alternativo como mecanismo de determinação de sexo na mosca doméstica *Drosophila melanogaster* (Figura 11.6B). Esta determinação segue uma cascata de eventos dependentes de *splicing* alternativo e tem como ponto inicial a expressão do gene *sxl* (*sex lethal*). A razão entre cromossomos X e autossomos em embriões determina o *splicing* alternativo de transcritos do gene *sxl*. Em machos, todos os éxons são incorporados, levando à síntese de uma proteína truncada em decorrência de um sítio de parada da tradução prematuro presente no éxon 3 do gene *sxl*. Nas fêmeas, o éxon 3 é removido, levando à formação de uma proteína funcional SXL. Esta proteína regula o *splicing* alternativo do segundo hnRNA da cascata, produzido a partir do gene *tra*. Em machos, o *splicing* de transcritos primários de *tra* acontece normalmente, o que abole completamente a expressão da proteína TRA em machos. Em fêmeas, entretanto, a ligação de SXL ao sítio de *splicing* 3' do íntron a montante do segundo éxon do gene *tra* leva à formação de um sítio de *splicing* 3' críptico interno no segundo éxon. Como resultado, é produzida a proteína TRA exclusivamente em fêmeas. Um terceiro e último nível de controle ocorre no *splicing* alternativo dos transcritos primários do gene

dsx, cujos produtos são os principais determinantes de sexo em *Drosophila*. A proteína TRA, produzida apenas em fêmeas, se associa conjuntamente a proteínas SR e a um ESE presente no éxon 4 deste transcrito. Desta forma, o mRNA maduro produzido em fêmeas possui o éxon 3 e o éxon 4, ao passo que, em machos, na ausência da proteína TRA, este éxon é removido no *splicing*, levando à formação de um mRNA maduro contendo os éxons 3 e 5. Assim, são produzidas duas variantes da mesma proteína DSX, mas que são sexo-específicas.

Os mecanismos que regulam o *splicing* alternativo dependem de várias sequências *cis* associadas aos íntrons e aos éxons adjacentes. Os sítios de *splicing* podem ser fracos ou fortes, dependendo do quão conservados são em relação à sequência consenso, o que determina a avidéz na interação com os componentes do spliceossomo. Em geral, os sítios de *splicing* fortes determinam a ocorrência de *splicing* constitutivo, enquanto a excisão de íntrons contendo sítios de *splicing* fracos depende de fatores adicionais. Estes fatores adicionais reconhecem sequências *cis* presentes tanto nos éxons (ESSs e ESEs) quanto nos íntrons (ISSs e ISEs). Os fatores que reconhecem estas sequências *cis* geralmente pertencem à grande família de proteínas SR e a outros hnRNPs, assim como fatores tecido-específicos, como NOVA e FOX. Por meio do reconhecimento destes fatores pode ocorrer a ativação ou a inativação da participação de um determinado sítio de *splicing*, como visto acima para as proteínas SXL na determinação de sexo em *D. melanogaster*. Evidências recentes apontam para um conceito global para definição do *splicing* alternativo: se o *splicing* ocorrer unicamente após o término da transcrição, o grau de inclusão de um éxon alternativo será dependente da força dos seus sítios de *splicing* 5' e 3', da presença de diferentes ESEs e ESSs no éxon alternativo, da presença de diferentes ISEs e ISSs nos íntrons adjacentes, de possíveis estruturas secundárias que ocorrem nas proximidades do éxon que podem promover algum impedimento para acesso do spliceossomo, da concentração nuclear dos fatores regulatórios, assim como da atividade destes fatores regulatórios, que pode ser modulada por modificações pós-traducionais. Entretanto, considerando que o *splicing* alternativo pode estar associado à transcrição, especialmente pela interação de proteínas SR com o domínio CTD da RNAPII, sabe-se que o cenário é mais complexo e que fatores que regulam a transcrição possuem um papel importante no *splicing* alternativo.

Como um grande número de fatores são necessários para a montagem correta do spliceossomo, um controle adequado deve existir para garantir a excisão precisa dos íntrons e éxons alternativos. De forma contrária, mRNAs maduros aberrantes podem ser gerados, podendo resultar em anomalias celulares e possíveis doenças associadas. Por exemplo, algumas doenças humanas, como algumas formas de atrofia muscular, retinite pigmentosa, síndrome de Hutchinson-Gilford e alguns tipos de câncer, têm sua origem em mutações nas sequências *cis* que governam o *splicing* alternativo. Além das mutações somáticas, a edição de RNA (ver Seção Mecanismos de

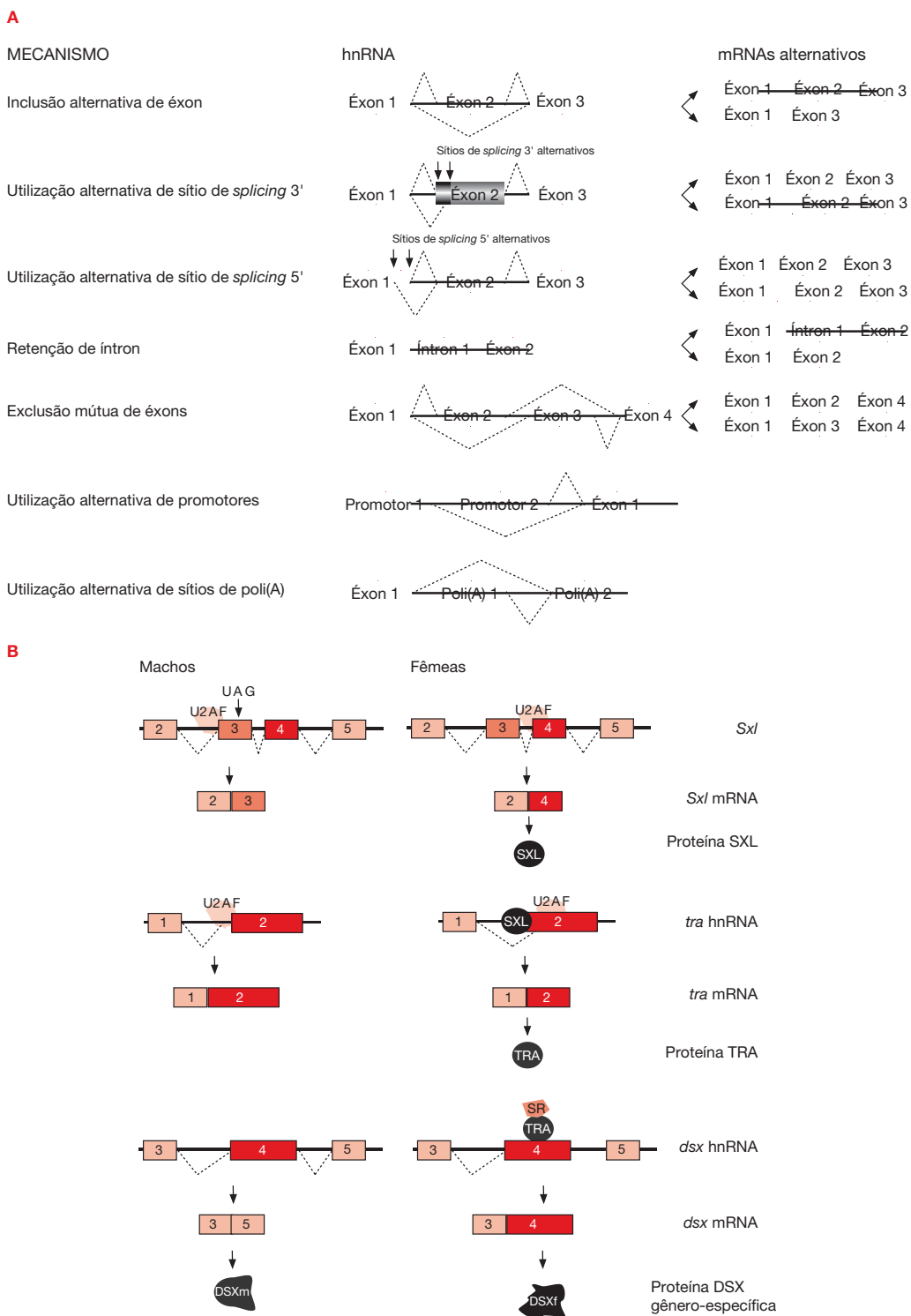


Figura 11.6

Splicing alternativo como mecanismo de geração de diversidade estrutural e funcional de proteínas. (A) Diversos mecanismos de *splicing* alternativo em precursores de mRNA e os seus respectivos produtos. (B) Processo de programação gênica para determinação de sexo em *D.melanogaster*. O *splicing* alternativo é responsável pela formação de proteínas funcionais SXL (exclusão alternativa de éxon) ou TRA (utilização alternativa de sítio de *splicing* 5'), que atuam em uma cascata que determina a formação, por *splicing* alternativo, de uma proteína gênero-específica DSX, cuja atividade está diretamente associada à expressão de genes gênero-específicos.

Edição de RNA) pode provocar mudanças nestas sequências *cis*, seguido de falha no reconhecimento destas sequências por parte dos fatores de *splicing*. Um dos tipos mais comuns da edição de RNA em mamíferos é a troca da base adenosina por uma inosina, a qual é reconhecida pelo spliceossomo como uma guanina. Como consequência, sítios de *splicing* podem ser eliminados ou gerados, alterando completamente a sequência do mRNA maduro.

1.5 Trans-splicing

Os mecanismos de *splicing* alternativo por definição são intramoleculares e resultam na excisão de sequências de

íntrons da mesma molécula de hnRNA; portanto, ocorrem em *cis*. O *splicing* alternativo é, então, *cis-splicing*. Em alguns organismos, especialmente nematódeos e tripanossomatídeos, ocorre um fenômeno de *splicing* no qual éxons provenientes de diferentes moléculas de hnRNA podem estar presentes em um mRNA maduro. Este fenômeno é denominado *trans-splicing* (Figura 11.7A), e existem evidências obtidas por análises de sequências de que possa ocorrer também em humanos.

Em tripanossomatídeos, os genes codificadores de proteínas são organizados como longas unidades de transcrição policistrônica, que podem abrigar até 100

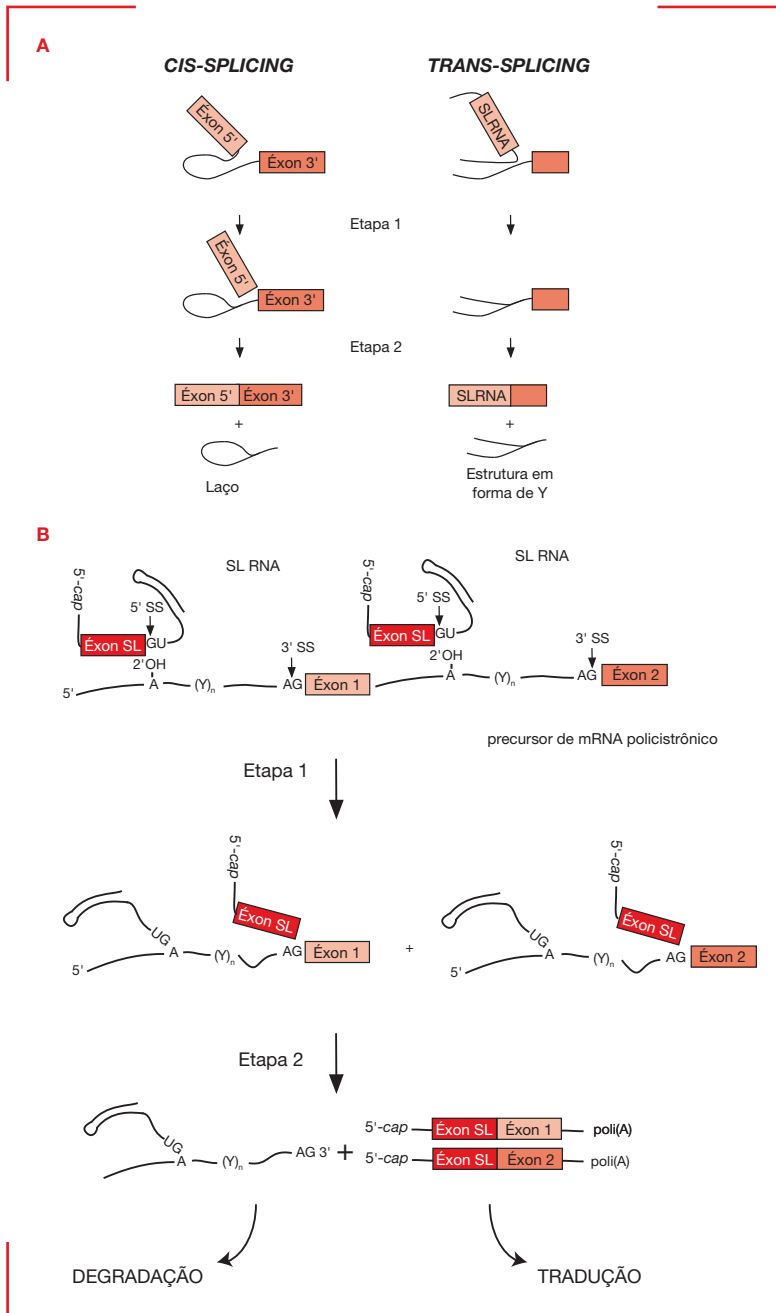


Figura 11.7

Trans-splicing em tripanossomatídeos. (A) Esquematização comparativa do processo de *cis-splicing* e de *trans-splicing*. (B) Processo de *trans-splicing* em um mRNA policistrônico, com reconhecimento do sítio de *splicing* 5' (5' SS), presente no SL RNA, e do sítio de *splicing* 3' (3' SS), presente no RNA policistrônico, pelo spliceossomo, e consequente união dos éxons, levando à formação de mRNAs maduros.

diferentes genes. Os hnRNAs oriundos da transcrição destas unidades contendo vários genes são processados por mecanismos de *splicing* e poliadenilação, liberando os mRNAs maduros. Entretanto, os mecanismos de *splicing* que processam estes longos transcritos policistrônicos são bastante diferentes dos mecanismos de *cis-splicing*. A geração dos mRNAs maduros depende do *trans-splicing*, pelo qual ocorre a união de um mini-éxon de 39 nucleotídeos proveniente de um transcrito denominado sequência-líder de *splicing* (SL, do inglês, *spliced leader*), não presente na região 5' dos genes individuais. Como consequência, cada mRNA maduro terá os mesmos 39 nucleotídeos iniciais. A adição do SL ao início de cada cístron para formação de um mRNA maduro é importante, pois providencia o 5'-*cap* no seu início, fundamental para exportação para o citoplasma e consequente tradução. Devido a alta demanda para suprir o *trans-splicing* nos hnRNAs dos mais de 9.000 genes destes organismos, seus genomas contêm mais de 200 cópias dos genes SL. Quando o SL e o hnRNA são processados por uma reação de *trans-splicing*, a região 3' do SL e a região 5' de cada cístron compõem as porções 5' e 3' de um íntron. Após a reação de *trans-splicing*, uma ligação 5' – 2' é formada entre o sítio de *splicing* 5', presente no SL, e o sítio de ramificação, presente no hnRNA, acarretando na formação de uma estrutura em forma de Y, lembrando do laço formado nas reações de *cis-splicing* (Figura 11.7B).

No nematódeo *Caenorhabditis elegans*, cerca de 70% dos genes são processados por *trans-splicing*, podendo estes eventos ser classificados em dois grupos distintos. Uma classe destes genes é caracterizada por hnRNAs monocistrônicos processados tanto por *cis-splicing*, para remoção dos introns internos, quanto por *trans-splicing*, que fornece os 22 nucleotídeos do SL característico deste organismo. A segunda classe de genes é caracterizada por hnRNAs policistrônicos cuja maturação ocorre por *trans-splicing*, como ocorre em tripanossomatídeos. Existem dois tipos de SLs em *C. elegans* que participam diferencialmente no *trans-splicing* destas duas classes de genes. SL1 é associado apenas ao *trans-splicing* de hnRNAs monocistrônicos contendo uma região 5' característica não traduzida, denominada de *outtron*. SL2 participa do *trans-splicing* de hnRNAs policistrônicos.

2. Mecanismos de processamento de rRNA

Os ribossomos são complexos ribonucleoproteicos responsáveis pelo processo de síntese proteica (ver Capítulo 12). Para que sejam funcionais, a sua montagem ordenada depende da inclusão das moléculas de rRNA maduras corretamente processadas. Este processamento pós-transcricional de rRNA ocorre em todos os organismos, mas existem diferenças entre organismos procarionóticos e eucarióticos. A ordem lógica deste processamento é baseada na excisão das unidades funcionais a partir de um transcrito primário de alta massa molecular. O rRNA re-

presenta cerca de 70% de todo RNA celular e, portanto, os organismos desenvolveram mecanismos muito eficientes para a sua síntese e seu processamento. No momento da divisão celular, esta quantidade de rRNA deve quase duplicar para compor as novas células. Nos eucariotos, por exemplo, uma região especial do núcleo da célula procede a síntese de rRNA, o nucléolo.

2.1 Processamento e maturação de transcritos primários de rRNAs em procarionotos

Genomas bacterianos possuem uma diversidade de cópias de óperons responsáveis pela síntese de rRNA, podendo variar de 1 a 15, dependendo do tamanho do genoma (ver Capítulo 4). Estes óperons são caracterizados pela presença dos genes 16S, 23S e 5S, nesta ordem, podendo estar presentes também alguns genes de tRNAs e sequências espaçadoras que separam os genes (Figura 11.8A). Para estes RNAs, é utilizada a unidade S (de Svedberg), que representa o tamanho da molécula de acordo com a sua medida da velocidade de sedimentação. Pela atividade da RNAP nestes óperons, um transcrito primário (denominado 30S) é produzido, e subsequentemente é alvo de diversas enzimas que têm a função de retirar os espaçadores e originar os rRNAs maduros. Este processo é dinâmico e é iniciado mesmo antes da transcrição do óperon ter terminado.

O processo de maturação dos rRNAs em procarionotos é bastante conservado, tendo sido elucidado primeiramente em *E. coli*. Nesta bactéria, existem sete óperons de rRNA que apresentam a estrutura básica 16S-23S-5S no transcrito primário. A primeira endonuclease a clivar o transcrito primário do rRNA é **RNase III**, que separa os precursores de rRNA e de tRNA. Durante a transcrição, o conjunto de nucleotídeos que flanqueiam os sítios correspondentes aos rRNAs 16S e 23S maduros formam uma fita dupla de RNA por pareamento de sequências complementares, gerando os sítios de reconhecimento de RNase III (Figura 11.8A). Os produtos desta reação são os precursores do rRNA 16S (17S rRNA), do rRNA 23S e do rRNA 5S (9S rRNA), assim como os precursores de alguns tRNAs, caso estejam presentes no RNA precursor, dependendo do organismo. A maturação final do rRNA 16S depende ainda das atividades das enzimas RNase E, RNase G e de uma RNase ainda não completamente caracterizada. Para os rRNAs 23S e 5S, a maturação final depende das atividades das enzimas RNase III e RNase E, respectivamente (Figura 11.8A).

As moléculas de rRNA podem ainda ser submetidas a uma série de modificações químicas, algumas das quais são fundamentais para a correta montagem e funcionalidade do ribossomo. Tanto a base nitrogenada quanto a ribose dos ribonucleotídeos podem ser alvo de mais de 80 modificações diferentes, cada uma catalisada por enzimas próprias. Por exemplo, o 16S rRNA maduro possui 11 modificações, sendo 10 delas metilações, ao passo que o rRNA 23S possui 25 modificações, sendo 14 metilações e 9 isomerizações de uridina para pseudouridina.

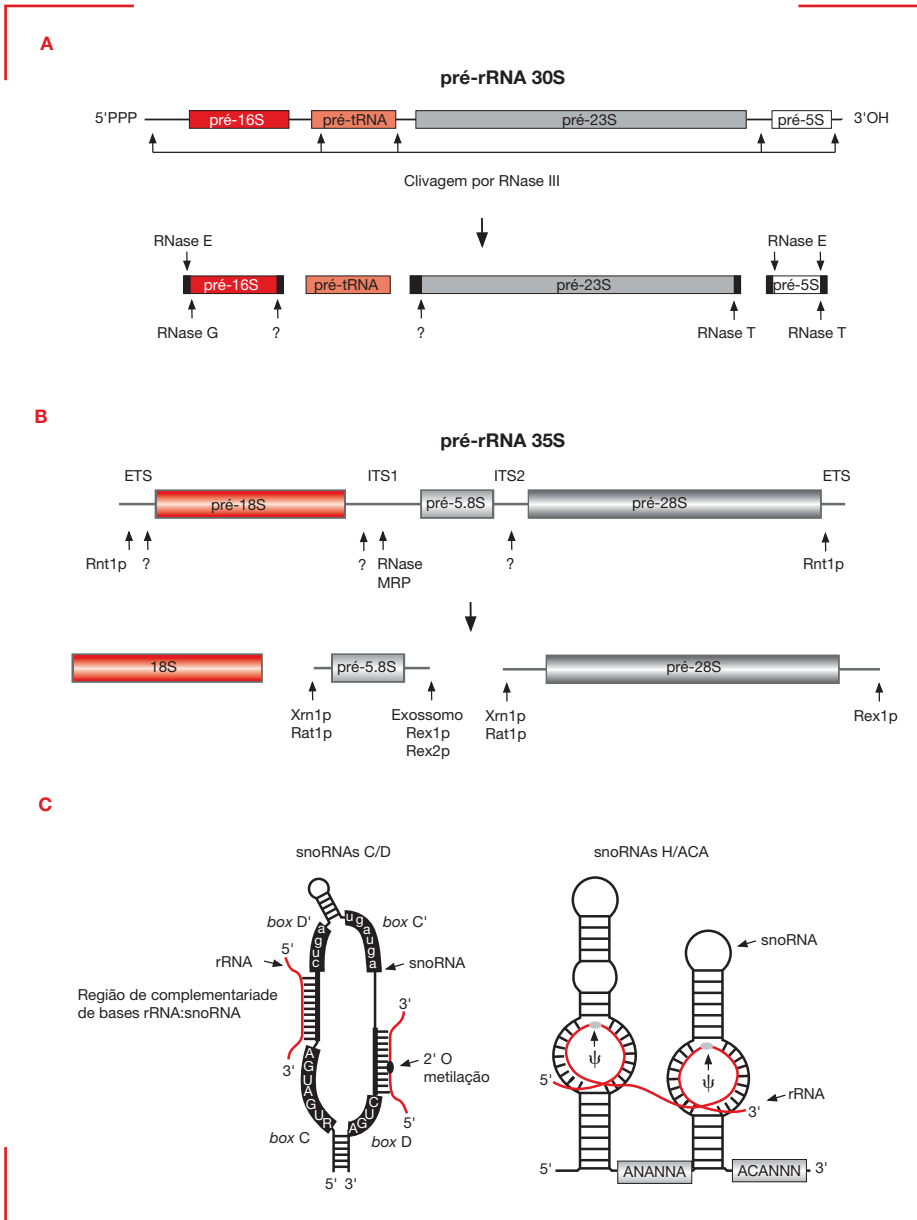


Figura 11.8

Maturação de rRNAs. (A) Esquemática representando o transcrito primário de um *cluster* rRNA procarionótico, o qual contém as três moléculas de rRNA (23S, 16S e 5S), assim como um tRNA. A atividade de diferentes endonucleases são responsáveis pela maturação destas moléculas. (B) Esquemática representando o transcrito primário de um *cluster* rRNA eucariótico, o qual contém os espaçadores transcritos externos (ETS – do inglês, *external transcribed spacers*) e internos (ITS – do inglês, *internal transcribed spacers*), assim como os precursores de três moléculas de rRNA (28S, 18S e 5.8S). As setas indicam o local de atuação de diversas endonucleases e exonucleases responsáveis pela maturação destes rRNAs. (C) snoRNAs das classes *box C/D* e *H/ACA*, os quais são responsáveis por modificações de bases nas moléculas de rRNA. Por meio de pareamento de bases, os snoRNAs C/D reconhecem sequências específicas nos rRNAs e participam da metilação de bases específicas, ao passo que snoRNAs H/ACA participam da pseudouridilação de bases específicas.

2.2 Processamento e maturação de transcritos primários de rRNAs em eucariotos

Em eucariotos, as moléculas de rRNAs são também produzidas a partir de grandes moléculas precursoras, as quais devem ser submetidas a diversas etapas de clivagem e modificações para serem totalmente funcionais. Existem quatro tipos de moléculas de rRNAs em organismos eucarióticos, sendo três produzidas a partir de um precursor maior sintetizado pela RNA-polimerase I – RNAPI (28S, 18S e 5.8S). Outra polimerase, RNA-polimerase III – RNAPIII, transcreve rRNA 5S, que está isolado dos outros genes de rRNA. Os genes dos precursores de rRNA em eucariotos estão alocados nos genomas em *tandem* (distribuídos sequencialmente), e o número de cópias

pode variar de algumas dezenas a milhares, dependendo da complexidade do organismo (ver Capítulo 5).

Os genes de rRNAs 28S, 18S e 5.8S estão agrupados e geram um transcrito precursor, denominado 45S em organismos multicelulares e 35S na levedura *S. cerevisiae*. Os rRNAs maduros são liberados a partir dos precursores por atividade enzimática de uma diversidade de enzimas nucleolíticas, tanto endonucleases quanto exonucleases, removendo os espaçadores transcritos externos (ETS, do inglês, *external transcribed spacers*) e os espaçadores transcritos internos (ITS, do inglês, *internal transcribed spacers*). De uma maneira geral, este processo é bastante conservado em eucariotos, sendo que a geração de extremidades 5' está associada a atividades de endonucleases, enquanto as extremidades 3' dos rRNAs são originadas por endonucleases, seguido da atividade de exonucleases

(Figura 11.8B). Muitos fatores de processamento estão associados, os quais reconhecem fatores *cis* presentes tanto nos ETSS quanto nos ITSS. O processamento de rRNA que ocorre no nucléolo tem como alvo um grande complexo ribonucleoproteico, o qual está diretamente envolvido com a biogênese de ribossomos.

Assim como ocorre em procariotos, além da excisão de nucleotídeos, as moléculas de rRNA passam por modificações antes de sua completa maturação. Estas modificações dependem de pequenas moléculas nucleolares de RNA denominadas **snoRNAs** (RNAs nucleolares pequenos, do inglês, *small nucleolar RNAs*). Existem centenas de snoRNAs diferentes em eucariotos sendo classificados de acordo com sequências conservadas e sua estrutura secundária. Os snoRNAs do grupo C/D são requeridos para metilar riboses na posição 2' dos rRNAs. Isto se deve em parte a uma região de complementariedade de uma sequência próxima ao *box D* com as sequências dos rRNAs 28S e 18S (Figura 11.8C). Os snoRNAs do grupo H/ACA promovem a conversão de uridina a pseudouridina, que estão presentes entre 50 a 100 resíduos por rRNA, dependendo do organismo. Os snoRNAs do grupo H/ACA atuam em conjunto com uma diversidade de proteínas e a atividade enzimática é provida pela pseudouridina sintase. Este complexo reconhece a região a ser modificada por pareamento com o snoRNA, que resulta na conversão de uridina a pseudouridina (Figura 11.8C).

3. Mecanismos de processamento de tRNA

A maturação de tRNAs, moléculas que transportam os aminoácidos na síntese proteica, depende da participação de uma série de atividades enzimáticas. Em procariotos e eucariotos, as moléculas de tRNA transcritas não são funcionais. Diversas modificações pós-transcricionais são necessárias para a sua completa maturação, que incluem a remoção de íntrons, a remoção das sequências adjacentes às extremidades 5' e 3', a adição da sequência CCA à extremidade 3', a modificação de bases e, em eucariotos, a exportação para o citoplasma.

3.1 Geração de tRNAs clivados e adição de 3' CCA

Genes de tRNAs estão organizados em procariotos, em óperons ou agrupados com genes de rRNA. Em eucariotos, os genes de tRNAs geralmente estão dispostos em unidades transcricionais isoladas, podendo estar repetidas diversas vezes no genoma. A transcrição sintetiza precursores que subsequentemente têm suas extremidades 5' e 3' removidas. As extremidades 5' dos tRNAs maduros são gerados pela atividade da **RNase P**. É uma ribozima composta por uma molécula de RNA e um número variável de proteínas, dependendo do organismo (uma proteína em bactérias, quatro proteínas distintas em arqueas e nove proteínas distintas em eu-

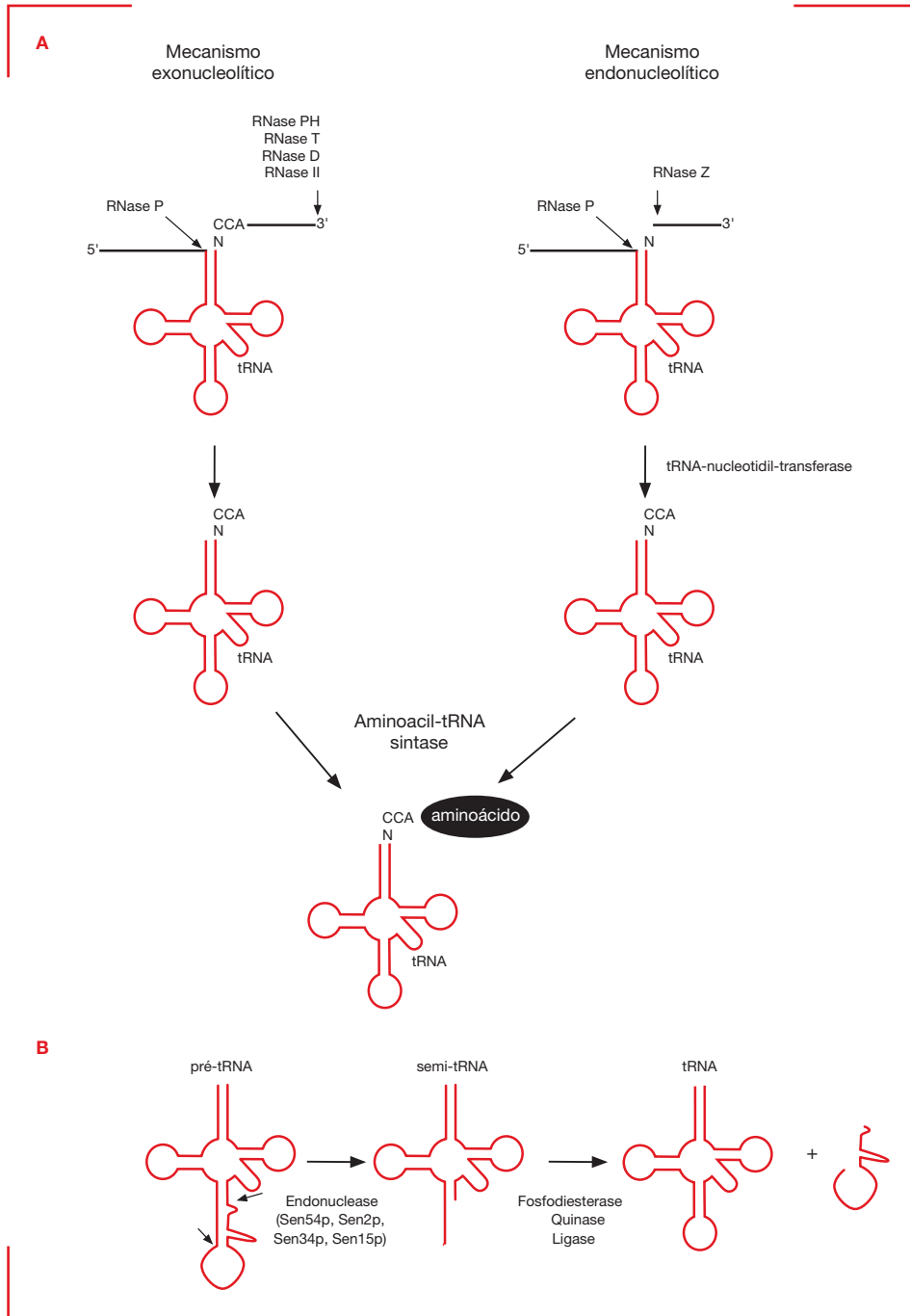
cariotos). O RNA que compõe a RNase P é conservado em todos os organismos e tem atividade endonucleásica. Esta ribozima reconhece o formato em L característico do tRNA, hidrolisando especificamente a ligação fosfodiéster do tRNA e liberando a sequência adjacente 5' (Figura 11.9A).

Em procariotos, são conhecidos dois grandes conjuntos de atividades responsáveis pela maturação da extremidade 3' dos tRNAs. A maturação exonucleolítica de tRNAs depende da atividade de enzimas como RNase PH e RNase T, que reconhecem a extremidade 3' do tRNA e hidrolisam as ligações fosfodiéster de maneira exonucleolítica. A maturação endonucleolítica de tRNAs ocorre em uma única reação, catalisada pela enzima ubíqua **RNase Z**. Em eucariotos, a atividade principal de maturação dos tRNAs é devida à atividade endonucleolítica da RNase Z. Esta enzima reconhece a ligação fosfodiéster associada ao nucleotídeo discriminador, ou seja, o nucleotídeo não pareado localizado no final do braço acceptor do tRNA (Figura 11.9A), liberando o tRNA para os passos seguintes de maturação.

A remoção da sequência 3' do precursor do tRNA é uma etapa fundamental para a sua completa maturação, pois fornece a estrutura adequada para a ligação do aminoácido de acordo com o anticódon daquele tRNA. Esta ligação é catalisada por um grande conjunto de enzimas denominadas aminoacil-tRNA-sintetases, discutidas em detalhe no Capítulo 12. Para a atividade correta destas enzimas, cada tRNA deve exibir no seu braço acceptor a sequência CCA não pareada. A maioria dos tRNAs de procariotos apresenta esta sequência no gene. Os tRNAs de eucariotos e aqueles de procariotos gerados pelo processamento por RNase Z devem ser modificados pela adição da sequência CCA na extremidade 3'. Esta adição é catalisada pela enzima tRNA-nucleotidil-transferase, uma RNA-polimerase que adiciona sequencialmente os ribonucleotídeos C, C e A à extremidade 3' do tRNA sem a necessidade de molde. As enzimas tRNA-nucleotidil-transferase podem ser classificadas em dois grupos de acordo com a conservação de aminoácidos no sítio catalítico, sendo as do grupo I encontradas em arqueas e as do grupo II, presentes em bactérias e eucariotos.

3.2 Modificações de bases nos tRNAs

As modificações pós-transcricionais de bases que ocorrem no tRNA definem de forma drástica a sua funcionalidade. Existem cerca de 90 modificações que podem ocorrer nas bases de tRNAs, tanto de procariotos quanto de eucariotos. Os quatro nucleotídeos (A, U, C e G) podem ser modificados tanto na base nitrogenada quanto na ribose. Além disto, estas modificações podem ser relativamente simples, como uma metilação na ribose ou na base uridina, ou intensa reformulação gerando estruturas complexas, como, por exemplo, a adição de um anel pentenil por meio de uma ligação N-C à 7-metilguanossina para a formação de uma base Q (como a queosina), ou a adição de um anel à estrutura da guanossina para a formação de uma base Y (como a wyosina).

**Figura 11.9**

Maturação de tRNAs. (A) Esquema comparativo entre os mecanismos exonucleolítico e endonucleolítico de clivagem de precursores de tRNAs. Ambos dependem da atividade de nucleases que reconhecem a estrutura com o dobramento adequado e geram os precursores clivados. Para os precursores de tRNA oriundos do mecanismo endonucleolítico, há necessidade de adição de CCA à extremidade 3', logo após o nucleotídeo discriminador (N). Esta reação é catalisada pela enzima tRNA-nucleotidil-transferase. **(B)** O *splicing* de íntrons em tRNAs é catalisado por uma série de endonucleases que reconhecem o íntron e o removem. Diferentes enzimas então promovem a ligação das duas moléculas de semi-tRNA para formação do tRNA funcional.

As modificações são catalisadas por enzimas ou conjunto de enzimas e apresentam variação muito grande em relação à sua especificidade. Algumas têm atividade específica sobre um determinado nucleotídeo de um determinado tRNA, ao passo que outras apresentam alvos em diferentes posições de diferentes moléculas de tRNA. Como resultado destas modificações, os tRNAs apresentam maior estabilidade, e o acesso e o reconhecimento por parte de proteínas e outros RNAs da maquinaria de tradução podem ser eficientemente modulados.

3.3 *Splicing* em tRNAs

A descoberta de íntrons em tRNAs ocorreu por meio de análises de seqüências da levedura *S. cerevisiae* realizadas nos anos 1980. Desde então, foram descritos íntrons em arqueas, plantas, mamíferos e bactérias. Em arqueas, aproximadamente 70% dos precursores de tRNA possuem íntrons cujo tamanho varia de 16 a 44 nucleotídeos. Íntrons de tRNA em eucariotos são relativamente pequenos (12 a 104 nucleotídeos) e estão em uma abundância menor em relação a arqueas, sendo descritos em 20% dos

tRNAs de levedura e aproximadamente 6% dos tRNAs de mamíferos. Os mecanismos que governam a remoção destes íntrons de tRNAs diferem significativamente daqueles associados ao *splicing* de íntrons via spliceossomos, sendo que não há conservação de sequências. Contudo, a excisão de íntrons de tRNAs em bactérias e algumas organelas ocorre por um mecanismo de *auto-splicing* (ver Seção *Auto-splicing*). A remoção de íntrons de tRNAs em eucariotos depende de mecanismos associados a diferentes atividades enzimáticas.

Os mecanismos de remoção de íntrons de tRNAs foram mais bem caracterizados na levedura *S. cerevisiae*. Os íntrons de tRNAs diferem significativamente dos íntrons presentes nos hnRNAs, o que impossibilita a excisão de íntrons de tRNAs pela atividade de spliceossomos. Os íntrons de tRNAs não apresentam conservação na sua sequência e apresentam na sua porção 5' uma sequência complementar ao anticódon. A formação da estrutura secundária característica do precursor do tRNA (Figura 11.9B) propicia o reconhecimento da sequência primária independente, mas estrutura terciária dependente, por uma endonuclease. Esta enzima, que em *S. cerevisiae* é um tetrâmero formado por duas subunidades catalíticas e duas subunidades estruturais, reconhece e cliva as extremidades 5' e 3' do íntron. Como resultado, são geradas duas moléculas de semi-tRNA pareadas e o íntron é liberado. Entretanto, a extremidade 3' do tRNA gerada por esta hidrólise é caracterizada por uma ligação 2' – 3' fosfato cíclico, e a extremidade 5' do íntron apresenta uma hidroxila. Uma enzima multifuncional e dependente de ATP catalisa uma série de reações que têm como produto final a união dos éxons. A atividade de fosfodiesterase converte a ligação 2' – 3' fosfato cíclico em 2' fosfato; a atividade de quinase fosforila a extremidade 5' OH do éxon a jusante; a atividade de ligase catalisa a formação da ligação fosfodiéster entre os dois éxons, cuja proximidade é garantida pela estrutura terciária da molécula, e finalmente a atividade de fosfatase remove o fosfato 2' (Figura 11.9B). Este mecanismo é bastante conservado, apesar de pequenas diferenças serem observadas em diferentes modelos. Por exemplo, o reconhecimento do íntron em eucariotos ocorre pela proximidade com dois braços do tRNA, entretanto, em arqueas, as estruturas secundárias do íntron do tipo *bulge-helix-bulge* determinam a associação da maquinaria para remoção do íntron.

4. Mecanismos de *auto-splicing*

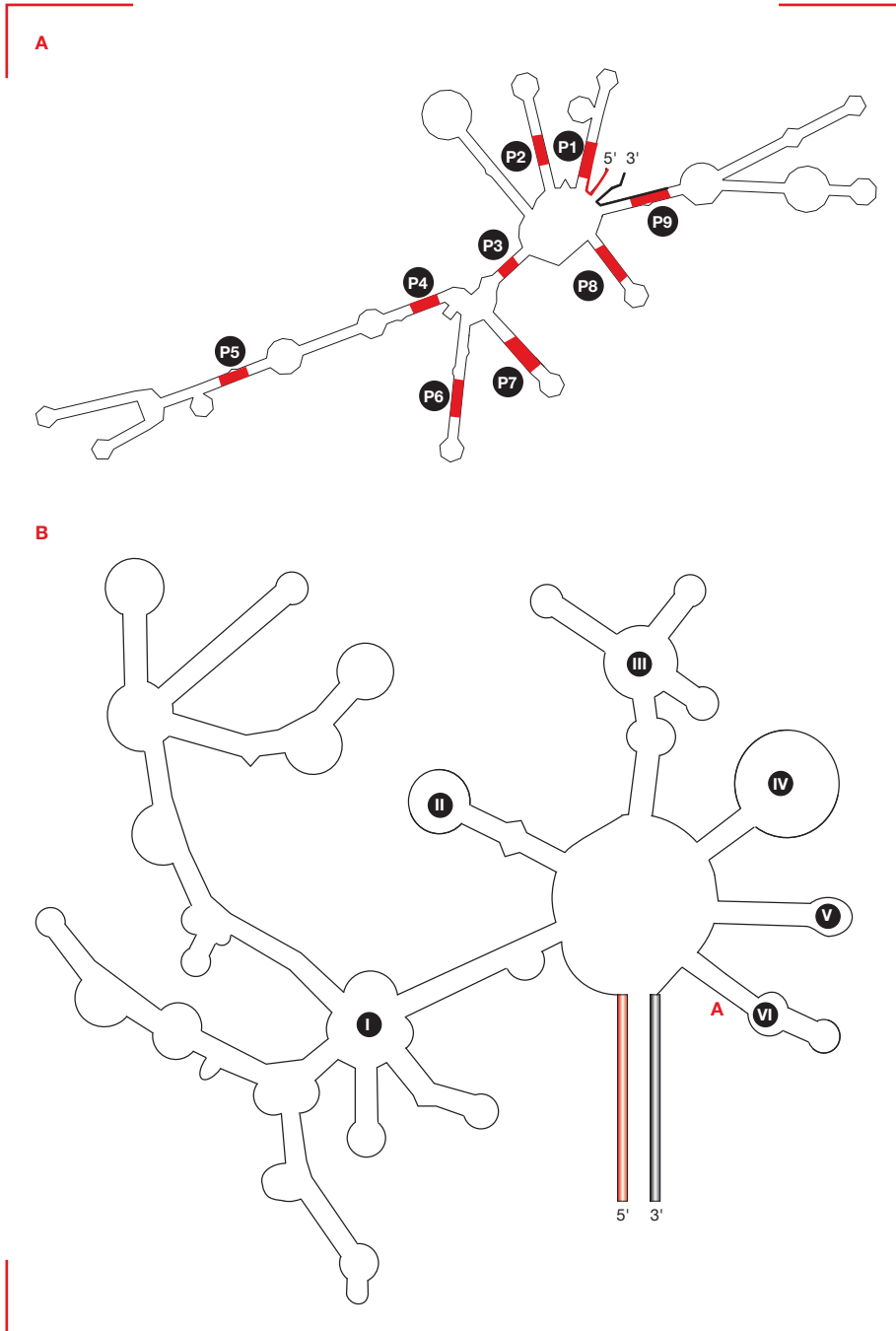
Além dos íntrons GU-AG e AU-AC (presentes em hnRNAs e removidos pelo spliceossomo), de tRNAs (removidos por maquinaria específica), são descritos ainda os *twintrons* e os íntrons dos grupos I, II e III. Os *twintrons* representam um arranjo especial de íntrons caracterizados pela presença de dois tipos distintos de íntron ocupando a mesma posição em um precursor de RNA. O grande diferencial dos íntrons dos grupos I, II e III é a sua capacidade de autoexcisão dos precursores de RNA e são denominados de ribozimas (RNAs com atividade

catalítica). O mecanismo associado à remoção destes íntrons depende da formação de estruturas secundárias que aproximam as extremidades 5' e 3' do íntron, levando à hidrólise das ligações fosfodiéster entre o íntron e os éxons a montante e a jusante. A hidrólise é seguida da união dos dois éxons por meio de uma ligação covalente. Devido às características gerais de distribuição e mecanismo de *splicing*, estes íntrons são classificados em três grupos distintos: íntrons do grupo I, do grupo II e do grupo III, que serão detalhados a seguir.

4.1 *Auto-splicing* associado a íntrons do grupo I

Íntrons do grupo I são encontrados em bactérias, bacteriófagos, vírus, e em genomas organelares e nucleares de eucariotos inferiores, como algas, fungos e líquens, sendo originalmente descritos em genes de rRNA do protozoário de vida livre *Tetrahymena thermophila*, em meados de 1980. Estes íntrons interrompem genes de rRNA, mRNA e tRNA em genomas bacterianos, assim como em genomas mitocondriais e cloroplastidiais de eucariotos inferiores. Entretanto, estes íntrons estão presentes apenas em genes de rRNA em genomas nucleares de eucariotos inferiores. Os íntrons do grupo I são relativamente pequenos (de 250 a 500 nucleotídeos) e apresentam uma estrutura secundária conservada, que determina seu mecanismo de excisão (Figura 11.10A). Após a transcrição, os íntrons presentes nos precursores de RNAs adotam uma estrutura baseada em elementos pareados (P1 a P9), que atuam diretamente no *auto-splicing* usando guanossina, GMP ou GTP como cofator. Os elementos P4, P5 e P6 formam um domínio estrutural, ao passo que a hélice composta pelos elementos P3, P7 e P9 forma o sítio catalítico, contendo o sítio para ligação de guanossina (GBS, do inglês, *guanosine binding sequence*). P1 e P10 são complementares ao final do éxon a montante e ao início do éxon a jusante, respectivamente, sendo denominados sequência guia interna (IGS, do inglês, *internal guide sequence*).

No primeiro passo do *auto-splicing* deste grupo de íntrons, ocorre um ataque nucleofílico, por parte da hidroxila 3' de uma guanossina exógena associada ao GBS, ao sítio de *splicing* 5', presente no elemento P1. Após este passo, a guanossina exógena permanece covalentemente ligada à extremidade 5' do íntron por uma ligação 3' – 5' fosfodiéster, e o sítio GBS é então ocupado por uma guanossina conservada que está presente no final do íntron, o que determina o sítio de *splicing* 3'. Em seguida, o grupamento hidroxila presente na extremidade 3' livre do éxon a jusante realiza um ataque nucleofílico ao sítio de *splicing* 3', resultando na ligação dos dois éxons e a liberação do íntron, o qual pode estar em uma forma circular ou linear. A autoexcisão de íntrons do grupo I também passa por duas reações de transesterificação, assim como ocorre na remoção dependente de spliceossomo. Com base na estrutura do elemento P7 e nas sequências adjacentes, existem pelo menos 13 subgrupos de íntrons do grupo I. Segundo o banco de dados RFAM, a existência destes íntrons ocorre em mais de 30.000 espécies.

**Figura 11.10**

Estrutura secundária de íntrons capazes de realizar *auto-splicing*. **(A)** Íntrons do grupo I apresentam nove pareamentos conservados (P1 a P9), sendo que os éxons a montante (vermelho) e a jusante (preto) ao íntron estão presentes em P1 e P9, respectivamente. **(B)** Íntrons do grupo II apresentam seis hélices conservadas (I a VI), sendo que os éxons a montante (vermelho) e a jusante (cinza) ao íntron são aproximados por esta estrutura secundária. A adenina responsável pela primeira etapa do *splicing* é representada em vermelho na hélice VI.

4.2 *Auto-splicing* associado a íntrons do grupo II

Íntrons do grupo II são caracterizados por RNAs catalíticos ou retroelementos móveis que catalisam o seu *auto-splicing* por meio da formação de um intermediário em forma de laço, semelhante ao mecanismo de remoção de íntrons de hnRNAs por spliceossomos. Adicionalmente, estes íntrons têm a capacidade de serem transferidos para sítios ectópicos (aleatórios) no genoma por meio de retrotransposição, o que se deve, em parte, à atividade de um complexo ribonucleoproteico que contém uma pro-

teína codificada pelo próprio íntron do grupo II (IEP, do inglês, *intron encoded protein*). Desta forma, íntrons do grupo II excisados são importantes mediadores da diversidade genômica. Em termos evolutivos, acredita-se que estes íntrons são os ancestrais do próprio spliceossomo, assim como de retrotransposons com LTR (ver Capítulo 9). Devido a esta capacidade de disseminação, os íntrons do grupo II têm sido encontrados em procarionotos e nos genomas de organelas de fungos, plantas e protistas. Mais de 25% dos genomas de bactérias sequenciadas apresentam íntrons do grupo II, geralmente pouco numerosos e apresentando-se como retroelementos ativos. Íntrons do

grupo II presentes em genomas de organelas, por sua vez, não se apresentam plenamente funcionais. Entretanto, estes íntrons são raros em arqueas.

O mecanismo associado ao *auto-splicing* em íntrons do grupo II foi consolidado a partir de ensaios utilizando um íntron da bactéria *Oceanobacillus iheyensis*. Íntrons do grupo II apresentam uma estrutura secundária conservada caracterizada por seis domínios (DI – DVI) e um anel central (Figura 11.10B). O domínio V contém os resíduos catalíticos, e o domínio VI contém uma adenosina protudente que tem papel análogo ao sítio de ramificação em íntrons spliceossômicos. Inserida no domínio IV está a região codificante do IEP. A excisão do íntron se dá por duas reações de transesterificação e inicia com o ataque nucleofílico da hidroxila 2' da adenosina protudente, presente no domínio VI, ao sítio de *splicing* 5', com consequente formação de intermediário em forma de laço associado ao éxon a jusante. A segunda reação ocorre pelo ataque nucleofílico da hidroxila 3' do éxon a montante ao sítio de *splicing* 3', levando à ligação dos éxons e à liberação do íntron na forma de um laço. Outro mecanismo, no qual uma molécula de água atua na hidrólise do sítio de *splicing* 5', foi descrito e denominado mecanismo hidrolítico, o que contrasta com o mecanismo de ramificação.

4.3 *Auto-splicing* associado a íntrons do grupo III

Os íntrons do grupo III assemelham-se aos do grupo II, apresentando domínios característicos e uma adenosina protudente que atua como ponte de ramificação no processo de *splicing*. Uma característica que os diferencia é sua distribuição, pois foram encontrados somente em genoma cloroplastidial de espécies de *Euglena*, outra é o seu diminuto tamanho em relação aos íntrons do grupo II, com tamanho médio de 99 nucleotídeos. O mecanismo de *auto-splicing* é também muito semelhante àquele dos íntrons do grupo II.

4.4 Mobilidade de íntrons

Íntrons dos grupos I e II também são conhecidos por sua capacidade de inserção em diferentes locais do genoma em um processo denominado mobilidade de íntrons. Os íntrons do grupo I utilizam um mecanismo de inserção mediado por DNA, enquanto nos íntrons do grupo II o mecanismo de inserção é baseado em RNA. Isto é possível devido ao fato de que estes íntrons possuem regiões codificantes de genes que atuam diretamente nesse processo.

O mecanismo que desenvolve a mobilidade de íntrons tendo como substrato o DNA depende das enzimas denominadas **endonucleases de homing**. Estas enzimas, codificadas por genes geralmente presentes em íntrons do grupo I, são consideradas meganucleases, visto que o seu sítio de reconhecimento é consideravelmente maior (14 a 44 pb) que aqueles comuns de enzimas de restrição (4 a 6 pb). Devido a este tamanho, esses sítios são extremamente raros nos genomas. Existem cinco famílias de endonucleases de *homing*, de acordo com a

conservação de sequências. As famílias LAGLIDADG e GIY-YIG são as mais numerosas e apresentam ampla distribuição, estando presentes tanto em procariotos quanto em eucariotos. Estas endonucleases reconhecem seu sítio no genoma, hidrolisam uma ligação fosfodiéster e permitem a inserção da sequência do íntron por meio de recombinação (conversão gênica – ver Capítulo 8). A origem do íntron duplicado ocorre por um mecanismo de transposição replicativa (ver Capítulo 9). Como a inserção do íntron altera o sítio de reconhecimento das endonucleases de *homing*, eventos posteriores de inserção não ocorrem. Muitas proteínas com atividade do tipo endonucleases de *homing* são **inteínas**. Trata-se de domínios presentes em proteínas que são capazes de realizar a sua autoexcisão em um processo denominado *splicing* proteico. Cerca de 350 inteínas já foram descritas em diversos organismos, segundo o banco de dados Inbase.

A mobilidade de íntrons tendo como substrato o RNA é particularmente associada a íntrons do grupo II. Esses mecanismos de mobilidade e as suas características próprias sugerem que estes íntrons foram os ancestrais evolutivos dos íntrons spliceossômicos, do spliceossomo e de retrotransposons em eucariotos. Estes íntrons possuem em um dos domínios característicos uma sequência que codifica para uma enzima multifuncional (IEP) que tem as atividades de maturase, transcriptase reversa e endonuclease. Após o *auto-splicing*, o íntron liberado em forma de laço associa-se à enzima IEP, formando um complexo ribonucleoproteico. Este complexo, por um conjunto de atividades que incluem a endonuclease, promove a inserção do íntron em sítios distintos do genoma que possuam identidade ao sítio de *splicing* (clivagem da sequência alvo). Há também atividade de transcriptase reversa (síntese de DNA a partir de molde de RNA), e o mecanismo é denominado *retrohoming*. Este processo também é dependente de reparo de DNA para que ocorra de modo adequado a ligação deste retroelemento no respectivo sítio.

Análises genômicas revelaram a existência de íntrons tanto do grupo I quanto do grupo II em uma diversidade de organismos, mas especificamente com ampla distribuição em genomas de organelas. Muita importância tem sido atribuída à mobilidade de íntrons, pois pode ser utilizada como ferramenta na manipulação gênica, com aplicação em terapia gênica, pela inserção de regiões codificantes tanto por meio de endonucleases *homing* construídas para o reconhecimento de uma sequência específica quanto pela modificação dos nucleotídeos que flanqueiam íntrons do grupo II, determinantes do sítio de integração dos mesmos no genoma por *retrohoming*.

5. Mecanismos de edição de RNA

No ano de 1986, o termo edição de RNA foi cunhado após análises de sequenciamento de mRNAs do gene *coxII* provenientes de mitocôndrias do protozoário *Trypanosoma brucei*. Uma diferença significativa era encontrada

nas sequências destes mRNAs, pois continham grandes segmentos não codificados no gene *coxII*. Estas diferenças são devidas a modificações pós-transcricionais dependentes da inserção na sequência do mRNA de resíduos de uridilato em posições específicas. Posteriormente, foi demonstrado que a edição de RNA ocorre em uma ampla gama de RNAs, como mRNAs, tRNAs, rRNAs e até microRNAs (ver Capítulo 14). A edição de RNA tem ampla distribuição taxonômica, desde fungos até mamíferos e apesar de ser majoritariamente encontrada em organismos eucarióticos, alguns exemplos de edição de RNA já foram descritos em arqueas. É importante salientar que a edição de RNA é um processo diferente daquelas alterações pós-transcricionais como o *splicing*, a adição de poli-A e de modificação do 5'-cap vistas anteriormente.

A edição de RNA tem por finalidade tornar o precursor de RNA funcional. Por exemplo, na maior parte dos casos, deve existir a adição ou a remoção de nucleotídeos nos precursores de mRNAs para gerar uma fase de leitura aberta, ou seja, uma sequência delimitada por um códon de início e um códon de término de tradução (ver Capítulo 12). O mRNA funcional poderá então codificar para a proteína correta. Desta forma, pode-se considerar a edição de RNA como um sistema de reparo de RNA, pois somente depois desta edição o mRNA terá a sequência correta. Os mecanismos de edição de RNA podem ser classificados em dois grandes grupos: inserção/remoção ou substituição.

Os mecanismos de edição por inserção ou deleção têm sido descritos apenas em protozoários da ordem Kinetoplastida, que incluem as espécies do gênero *Trypanosoma* e *Leishmania*, alguns dos quais são importantes patógenos humanos. Estas reações são caracterizadas pela inserção ou remoção de resíduos de uridilato no precursor de mRNA. Em alguns tripanossomatídeos, foram descritas intensas modificações em 12 diferentes precursores de mRNA mitocondriais, com a adição de mais de 3.000 resíduos ou a remoção de cerca de 300 resíduos de uridilato em sítios específicos. Este processo sítio-específico depende de RNAs pequenos não codificantes denominados RNAs guia (gRNA). A edição ocorre pela formação de um híbrido entre o precursor de mRNA e o gRNA que atua como um “quasi-molde”. Esta interação depende de regiões conservadas que propiciam a formação de uma âncora. No caso da inserção de uridilatos, ocorre uma clivagem logo após esta âncora, na primeira base não pareada. A atividade de adição de resíduos de uridilato é catalisada pela enzima Terminal-uridiltransferase (TUTase). Na remoção de uridilatos, a atividade da 3' exonuclease (exoUase) processa a região não pareada. O mecanismo de determinação do número de resíduos de uridilato a ser inserido ou removido depende do tamanho das regiões complementares entre o gRNA e o precursor de mRNA. Por fim, os fragmentos oriundos da edição do precursor do mRNA são unidos pela atividade de uma RNA-ligase. Este conjunto de reações ocorre no sentido 3' – 5' do precursor de mRNA e a maturação completa geralmente depende de ciclos múltiplos com diferentes gRNAs (Figura 11.11). Uma série de proteínas acessó-

rias participa deste processo, que compõe um complexo de grande massa molecular denominado editossomo.

A edição de RNA baseada na modificação da sequência do precursor de mRNA foi descrita pela primeira vez

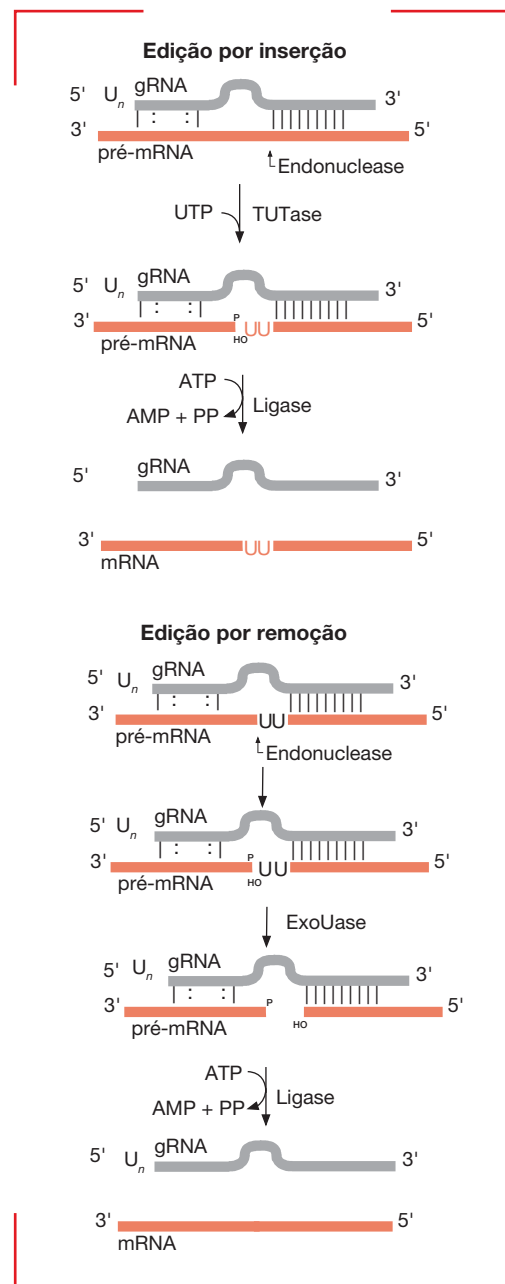


Figura 11.11

Mecanismos de edição de RNA em tripanossomatídeos. Uma vez associados por pareamento A-U, G-C ou G-U, o RNA guia (gRNA) recruta endonucleases que clivam o mRNA. A inserção de uridilatos (U) é realizada pela enzima TUTase, ao passo que sua remoção se dá pela ExoUase. As extremidades então são ligadas (Ligase), seguindo para tradução.

em oócitos de *Xenopus* no ano de 1987, baseado no fato de que moléculas de RNA antissenso injetadas não eram capazes de formar RNA fita dupla com o mRNA sintetizado e teoricamente capaz de hibridizar com o RNA injetado. Esta incapacidade de hibridização ocorre porque o mRNA sintetizado pela célula sofre edição com modificações que convertem resíduos específicos de adenosina para inosina (A-I), por desaminação hidrolítica. Esta reação é catalisada por uma enzima denominada ADAR (do inglês, *adenosine deaminase acting on RNA*). A edição A-I ocorre em regiões de fita dupla de RNA, sendo catalisada por ADARs, e ocorre em ampla gama de organismos, desde anêmonas até humanos. A inosina preferencialmente parecia com citidina, sendo funcionalmente equivalente à guanosina. Desta forma, a edição A-I pode afetar a funcionalidade de um RNA de diferentes maneiras. Quando considerados hnRNAs, esta edição pode provocar alteração da capacidade codificante de um mRNA, pela geração ou eliminação de sítios de *splicing* assim como de início ou término da tradução. A edição pode ainda produzir alterações na sequência para um códon não sinônimo, o que acarreta em uma proteína diferente. Outro possível resultado da edição é alterar a capacidade de silenciamento gênico por miRNAs ou siRNAs (ver Capítulo 14). O controle adequado da edição A-I é fundamental, sendo que diversas doenças humanas estão associadas a falhas desta maquinaria.

A edição C-U, desaminação da citosina, é bastante comum em RNAs mitocondriais de plantas, tendo sido originalmente descrito em tripanossomatídeos. É conhecida a sua ocorrência em mRNAs e tRNAs, tanto organelares quanto nucleares, de mamíferos, plantas e protozoários. Entretanto, um caso bastante característico da edição C-U refere-se à edição C-U que converte o códon CAA (glutamina) ao UAA (término da tradução) no ponto mediano do mRNA que codifica para a apolipoproteína B, um componente da lipoproteína de baixa densidade (LDL, do inglês, *low-density lipoprotein*). A geração do códon de UAA resulta na terminação prematura da tradução deste mRNA em células do intestino, produzindo a proteína truncada apoB48, ao passo que a proteína completa (apoB) é produzida no fígado. A edição C-U no mRNA apoB é mediada pelo complexo APOBEC (do inglês, *apoB mRNA editing complex*), o qual possui a atividade catalítica de citidina desaminase em uma subunidade (APOBEC-1) e um cofator ACF (do inglês, *APOBEC-1 complementation factor*), que associa a APOBEC-1 e o substrato na forma de RNA fita simples. As proteínas do tipo APOBEC possuem distribuição filogenética restrita, mas há um grupo de proteínas do qual as APOBECs fazem parte, denominado CDARs (do inglês, *citidine deaminase acting on RNA*), com maior distribuição entre eucariotos.

6. Degradação de RNA

A síntese de proteínas é um dos processos de maior consumo de energia para a célula, por isso o controle da ex-

pressão é tão fundamental (ver Capítulos 13 e 14). Um dos mecanismos possíveis de controle é pela degradação do mRNA sintetizado, abolindo, assim, a sua tradução. Portanto, pela degradação do mRNA, a célula pode controlar eficientemente quais proteínas deverão ser sintetizadas. Além disso, a degradação do RNA fornece os ribonucleotídeos necessários para a síntese de outros RNAs essenciais na célula. Portanto, a degradação do mRNA é um processo muito dinâmico e controlado. É comum utilizar os termos decaimento ou vida média do mRNA para referir a estabilidade dos mRNA à degradação.

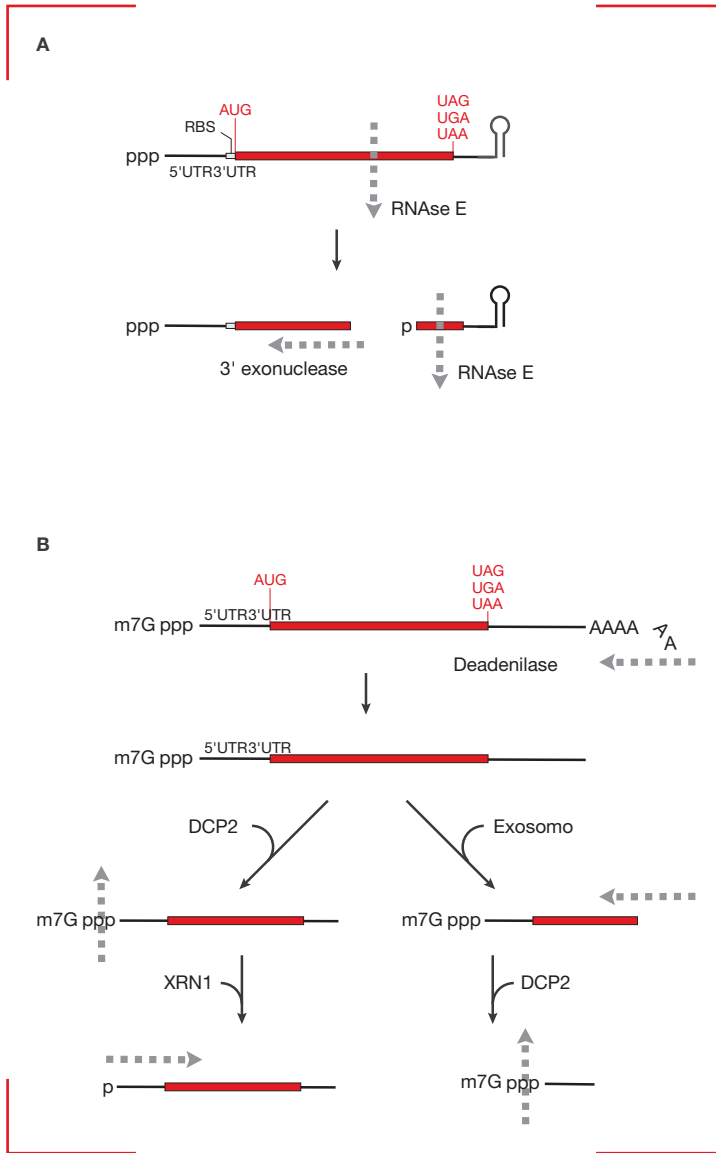
Devido às muitas diferenças entre procariotos e eucariotos, a expectativa inicial era de que existissem muitas diferenças nos sistemas de degradação de RNA nestes organismos. Entretanto, estudos mais recentes têm mostrado que existem muitas semelhanças. Serão discutidos aqui, de maneira geral, os mecanismos básicos destes dois modelos.

Em bactérias, os dois fosfatos da extremidade 5' do mRNA são removidos, deixando uma extremidade monofosfatada. Este mRNA é o substrato para endonucleases (RNase E) que clivam internamente a molécula, expondo extremidades 3'OH livres (**Figura 11.12A**). Estes fragmentos gerados são o substrato para uma série de 3'-exonucleases (PNPase, RNase II, RNase R) que degradam o RNA no sentido 3' – 5'. Pequenos fragmentos de RNA restantes são degradados por oligoribonucleases da mesma forma.

Todas essas 3'-exonucleases atuam em RNA de fita simples. Nos mRNAs bacterianos que apresentam o grampo de terminação (terminação intrínseca, Capítulo 10), além da presença do pareamento interno do grampo existe uma sequência muito curta após o grampo para a atividade das 3'-exonucleases. O mecanismo que contorna essa barreira é a adição de uma cauda de poliadeninas (de 10 a 40) no terminal 3' do mRNA. Essa adição é promovida por poli(A) polimerases (PAP) e resulta em moléculas que podem então serem degradadas pelas 3'-exonucleases.

É fundamental ressaltar o contexto em que os mRNAs devem ser degradados em procariotos, pois a transcrição e a tradução estão acopladas e ocorrem concomitantemente em procariotos. Naturalmente, os mRNAs estarão protegidos por ribossomos que iniciam a síntese de proteínas imediatamente após o mRNA nascente deixar o complexo de transcrição. Mesmo assim, a vida média dos mRNAs bacterianos é extremamente curta a ponto de mesmo antes de uma proteína estar completamente sintetizada pelos ribossomos o mRNA que a codificou já estar degradado. Portanto, esse é um processo muito dinâmico e vários complexos de ribossomos iniciam a tradução do RNA nascente imediatamente, pois assim que a endonuclease RNase E consegue ligar-se ao RNA a sua degradação é iniciada.

O modelo clássico para a degradação dos mRNAs eucarióticos (**Figura 11.12B**) envolve várias enzimas que degradam os diferentes componentes do mRNA, como

**Figura 11.12**

Degradação de mRNA. (A) Os mecanismos de degradação de mRNAs em procariotos iniciam-se pela atividade endonucleolítica da RNase E, que atua em regiões do mRNA não protegidas pelos ribossomos. Em seguida, diversas exonucleases degradam os subprodutos da hidrólise dos mRNAs pela RNase E. (B) Em eucariotos, a degradação inicia-se com a remoção da cauda poli(A) pela atividade de uma adenilase. Em seguida, a degradação ocorre de maneira exonucleolítica, tanto 5'-3', catalisado pela enzima XRN1, quanto 3'-5', catalisado pela enzima DCP2.

3' e 5' exonucleases e endonucleases. Basicamente, as duas modificações nas extremidades do mRNA – 5'-cap e poli(A) – conferem estabilidade e resistência à degradação. Assim, a sua remoção é importante para a degradação do mRNA que ocorre por ação de adenilases, as quais removem as adeninas da cauda poli(A) e enzimas que removem o 5'-cap. O principal processo de degradação de mRNAs, caracterizado em metazoários, começa com a diminuição da cauda poli(A), seguido pela remoção do 5'-cap pela atividade de Dcp2 e consequente degradação do mRNA de maneira exonucleásica, tanto pela atividade de Xrn1 (exonuclease 5'-3') quanto pelo complexo **exossomo** (exonucleases 3'-5'). O exossomo é um complexo de degradação de RNA, com atividade na extremidade 3', composto por muitas subunidades e que atua tanto no citoplasma quanto no núcleo e é ativo sempre que a proteína PABP não está complexada com o RNA. Entretanto, existe outro mecanismo importante que não depende da

deadenilação prévia à degradação, sendo dependente de endonucleases que podem hidrolisar o mRNA tanto na região codificante quanto na região 3'UTR, seguido da ação de exonucleases tanto na direção 5' – 3' quanto 3'-5'. Este fenômeno é diferente da degradação mediada por miRNAs e siRNAs, que envolvem a atividade da enzima argonauta (Capítulo 14), sendo pouco conhecido sobre as enzimas que participam da degradação de mRNAs iniciada por endonucleases.

Existem diversos mecanismos adicionais de degradação de mRNA, como, por exemplo, o RNAi, importante mecanismo de regulação gênica pós-transcricional. Este mecanismo é dependente de um grande complexo denominado RISC e leva à degradação sequencia-específica de determinados mRNAs. Este processo é melhor discutido no Capítulo 14 (ver Seção Regulação pós-transcricional).



Resumo

Nos diversos domínios da vida, quase todos os RNAs recém-sintetizados não são funcionais. As três principais classes de RNAs (mRNA, rRNA e tRNA) são produzidas como precursores que devem passar por uma série de modificações para que sejam funcionais. Este conjunto de modificações é denominado processamento de RNA. Em eucariotos, os precursores de mRNA passam por três modificações principais: a adição do 5'-*cap*, a adição de uma cauda de poli(A) na extremidade 3', e a excisão de íntrons ou *splicing*. As moléculas de tRNA funcionais são originadas pela retirada de seqüências nas extremidades 5' e 3', com eventual adição de seqüências, para en-

tão permitir a ligação de um aminoácido na sua extremidade 3' de acordo com o anticódon daquele tRNA. Já as moléculas de rRNA maduras são oriundas de eventos de clivagem e remoção de extremidades, assim como modificação de bases. Existem outros mecanismos de modificação de RNAs, como a edição, caracterizada pela adição ou remoção específica de bases na molécula de precursor de mRNA, ou ainda a troca de nucleotídeos. As moléculas de mRNA podem, após ser traduzidas nos ribossomos, avançar para um passo de degradação, que segue duas vias independentes baseadas na atividade de endonucleases e exonucleases.



Leituras recomendadas

Chen HC, Cheng SC. Functional roles of protein splicing factors. *Biosci Rep*. 2012;32(Pt 4):345-59.

De Conti L, Baralle M, Buratti E. Exon and intron definition in pre-mRNA splicing. *Wires RNA*. 2013;4(1):49-60.

El Yacoubi B, Bailly M, de Crécy-Lagard V. Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annu Rev Genet*. 2012;46:69-95.

Gray MW. Evolutionary origin of RNA editing. *Biochemistry*. 2012;51(26):5235-42.

Kaczanowska M, Rydén-Aulin M. Ribosome biogenesis and the translation process in *Escherichia coli*. *Microbiol Mol Biol Rev*. 2007;71(3):477-94.

Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ. Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat Rev Mol Cell Bio*. 2013;14(5):153-65.

Lambowitz AM, Zimmerly S. Group II Introns: mobile ribozymes that invade DNA. *Cold Spring Harb Perspect Biol*. 2011;3(8):a003616.

Mullineux ST, Lafontaine DL. Mapping the cleavage sites on mammalian pre-rRNAs: Where do we stand? *Biochimie*. 2012;94(7):1521-32.

Preußner C, Jaé N, Bindereif A. mRNA splicing in trypanosomes. *Int J Med Microbiol*. 2012;302(4-5):221-4.

Raghavan R, Minnick MF. Group I introns and inteins: disparate origins but convergent parasitic strategies. *J Bacteriol*. 2009;191(20):6193-202.

Schoenberg DR, Maquat LE. Re-capping the message. *Trends Biochem Sci*. 2009;34(9):435-42.

Tian B, Graber JH. Signals for pre-mRNA cleavage and polyadenylation. *Wires RNA*. 2012;3(3):385-96.

Turunen JJ, Niemelä EH, Verma B, Frilander MJ. The significant other: splicing by the minor spliceosome. *Wires RNA*. 2013;4(1):61-76.

Irene Silveira Schrank | Marilene Henning Vainstein

Código Genético e Síntese de Proteínas

- 1. Código genético 256**
- 2. RNAs transportadores 259**
 - 2.1 Estrutura dos tRNAs 259
 - 2.2 Ativação dos tRNAs e as aminoacil-tRNA-sintetases 261
 - 2.3 tRNAs supressores 262
- 3. Ribossomos 263**
 - 3.1 Composição e estrutura 263
 - 3.2 Estrutura e dinâmica dos sítios ativos do ribossomo 265
- 4. Síntese de proteínas 266**
 - 4.1 Início 266
 - 4.1.1 Procariotos 266
 - 4.1.2 Eucariotos 266
 - Etapa 1: Formação do complexo de pré-iniciação 43S 266
 - Etapa 2: Ligação do complexo 43S ao mRNA 267
 - Etapa 3: Formação do complexo 48S 267
 - Etapa 4: Formação do complexo iniciador 80S 268
 - 4.2 Alongamento da cadeia polipeptídica 269
 - 4.2.1 Decodificação e atividade de peptidil-transferase 270
 - 4.2.2 Translocação 271
 - 4.3 Término da síntese de proteínas 273
 - 4.4 Saída da cadeia peptídica do ribossomo 273

Durante os processos evolutivos, todos os organismos desenvolveram a habilidade de traduzir eficiente e precisamente a informação genética do DNA. Esta precisão é importante, pois os genomas evoluíram em sua organização gênica estabelecendo para cada gene a codificação de um produto altamente específico.

Na síntese de proteínas (ou tradução), as interações entre os **códons** (uma sequência de três nucleotídeos correspondente a um determinado aminoácido) das moléculas de RNA mensageiro (mRNA) com os **anticódons** (uma sequência de três nucleotídeos presente no tRNA complementar e que interage com o códon por pareamento de bases) dos RNAs transportadores (tRNA) governa a formação da sequência de aminoácidos da proteína nascente. Além da interação mRNA-tRNA, a presença dos ribossomos é decisiva para a precisão e a interpretação do código genético. Os tRNAs, além de interagirem com os mRNAs por meio do pareamento códon-anticódon, carregam os aminoácidos específicos, sendo denominados **aminoacil-tRNAs**. O pareamento entre os códons dos mRNAs e os anticódons dos aminoacil-tRNAs é favorecido pela maquinaria dos ribossomos. Os ribossomos são considerados como “minúsculas fábricas”, que compreende todas as atividades necessárias para a síntese de proteínas.

Os ribossomos são formados por duas subunidades e apresentam um tamanho muito maior que as moléculas de aminoacil-tRNAs. A **Figura 12.1** mostra um esquema que relaciona o tamanho de um ribossomo com o das moléculas de RNA. Portanto, um ribossomo pode conter, durante a síntese de proteínas, uma região de 35 bases do mRNA e pelo menos duas moléculas de tRNA.

1. Código genético

Uma das características fundamentais encontrada em todos os organismos vivos está relacionada à utilização,

com pequenas variações, do mesmo código genético. A descoberta do código genético, em 1961, por Nirenberg e Matthaei, revelou que a informação genética está armazenada no DNA na forma de trinca (trinucleotídeos). A análise geral do código genético demonstra três principais aspectos: (1) as quatro bases nitrogenadas presentes no RNA se classificam em (A e G) no grupo das purinas e (C e U) no grupo das pirimidinas; (2) cada códon é representado por uma trinca de bases e, portanto, existem 64 possíveis códons (4^3); (3) são codificados, pelo código genético, os 20 aminoácidos presentes nas proteínas. Organismos que apresentam aminoácidos adicionais, como selenocisteína ou pirrolisina, desenvolvem mecanismos especiais de codificação.

A tabela do código genético padrão apresenta uma estrutura organizada, principalmente no que diz respeito ao posicionamento dos códons. Essa estrutura do código genético pode estar relacionada com o aperfeiçoamento seletivo, ao longo da evolução, buscando a robustez do processo de tradução e minimizando os erros na síntese de proteínas. Estudos sobre a origem e a evolução do código genético padrão demonstram que os códons usados na tabela permitem uma tradução muito mais precisa, se comparados com a maioria dos códons gerados aleatoriamente.

Na tabela do código genético padrão, aminoácidos similares são codificados por alguns códons, diferindo em uma única substituição de base. Considerando que existem 64 códons e somente 20 aminoácidos, pode-se concluir que a maioria dos aminoácidos é codificada por mais de um códon. Por isso, a tabela do código genético é muito redundante, e apenas os aminoácidos metionina e triptofano são codificados por uma única trinca de bases. A descoberta do código genético padrão ocorreu em 1966 e sua primeira base é demonstrada na **Tabela 12.1**. Na tabela, os códons aparecem com as bases presentes nas moléculas de mRNA, que são utilizados para a leitura do código genético durante a tradução. A análise da tabela demonstra que, dos 64 códons, somente 61 codificam os 20 ami-

Figura 12.1

Componentes fundamentais para a síntese de proteínas e sítios do ribossomo. A figura demonstra a relação de tamanho entre um ribossomo completo, as moléculas de tRNA e uma região de 35 bases do mRNA. Os valores em angstrom (Å) servem de referência para comparação de tamanho entre estas moléculas. Também estão representados os sítios aminoacil (A), peptidil (P) e o sítio de saída do tRNA (E) no ribossomo. As setas horizontais representam a direção do movimento dos tRNAs, do mRNA e do próprio ribossomo durante a tradução.

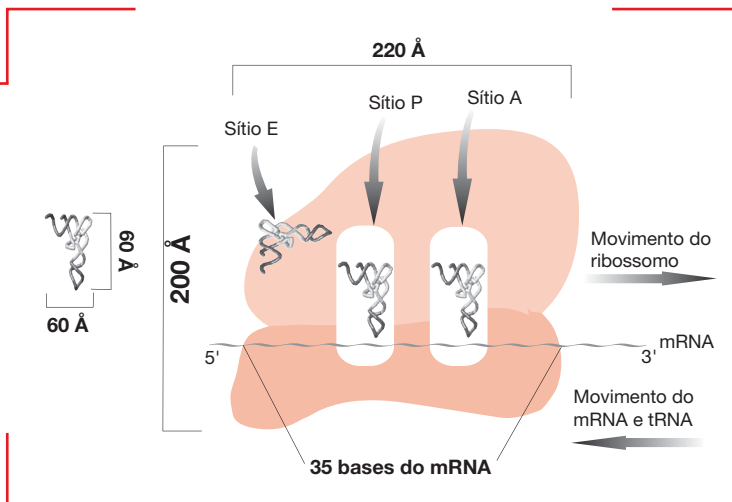


Tabela 12.1 Tabela do código genético

A

		Segunda base do código genético							
		U		C		A		G	
Primeira base do código genético	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
		UUC		UCC		UAC		UGC	
		UUA	Leu	UCA		UAA	Parada	UGA	Parada
		UUG		UCG		UAG		UGG	Trp
	C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
		CUC		CCC		CAC		CGC	
		CUA		CCA		CAA	Gln	CGA	
		CUG		CCG		CAG	CGG		
	A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
		AUC		ACC		AAC		AGC	
		AUA		ACA		AAA	Lys	AGA	Arg
		AUG	Met	ACG		AAG	AGG		
	G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
		GUC		GCC		GAC		GGC	
		GUA		GCA		GAA	Gln	GGA	
		GUG		GCG		GAG	GGG		

	Definições únicas
	Terceira base irrelevante
	Terceira base purina
	Terceira base pirimidina

B

		Segunda base do código genético							
		U		C		A		A	
Primeira base do código genético	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
		UUC		UCC		UAC		UGC	
		UUA	Leu	UCA		UAA	Parada	UGA	Parada
		UUG		UCG		UAG		UGG	Trp
	C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
		CUC		CCC		CAC		CGC	
		CUA		CCA		CAA	Gln	CGA	
		CUG		CCG		CAG	CGG		
	A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
		AUC		ACC		AAC		AGC	
		AUA		ACA		AAA	Lys	AGA	Arg
		AUG	Met	ACG		AAG	AGG		
	G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
		GUC		GCC		GAC		GGC	
		GUA		GCA		GAA	Gln	GGA	
		GUG		GCG		GAG	GGG		

Grau de hidrofobicidade: Os blocos contendo os códons sinônimos estão marcados conforme a escala de polaridade que define a hidrofobicidade dos aminoácidos. Tonalidade mais escura é relativa ao maior grau de hidrofobicidade.

noácidos, sendo os três códons restantes (**UAA, UAG e UGA**) responsáveis pelo término da síntese de proteínas, recebendo a denominação **códons de terminação**.

Os códons que representam o mesmo aminoácido são chamados de sinônimos e tendem a apresentar uma sequência similar. Normalmente, a base na terceira posição do códon é irrelevante, pois códons com variação na terceira posição codificam tal aminoácido. Esta característica define o código genético como **degenerado**. Essa degeneração do código genético pode ser observada, na tabela, para os códons dos aminoácidos leucina (leu), prolina (pro), serina (ser), alanina (ala), valina (val), glicina (gly), treonina (thr) e arginina (arg). Para outros aminoácidos, o código genético é parcialmente degenerado, pois a variação da base na terceira posição ocorre apenas dentro dos grupos das purinas ou pirimidinas. Para os aminoácidos fenilalanina (phe), histidina (his), cisteína (cys), asparagina (asn), aspartato (asp), tirosina (tyr) e serina (ser), a substituição da terceira base ocorre entre as pirimidinas; e para os aminoácidos leucina (leu), glutamina (gln), glutamato (glu), lisina (lys) e arginina (arg), ocorre entre as purinas. Esta redundância do código genético pode minimizar o efeito deletério das mutações, pois a variação da base na terceira posição, em geral, não afeta o tipo do aminoácido presente no peptídeo.

Na Tabela 12.1A, os códons para os aminoácidos, contendo as substituições na terceira posição entre purinas e pirimidinas, estão marcados. São ressaltados também, aqueles códons contendo definições únicas, como AUG (metionina), UGG (triptofano), e os códons AUU, AUC e AUA para isoleucina. A organização do código genético reflete também as similaridades na estrutura físico-química dos aminoácidos, como, por exemplo, todos os códons contendo a base U na segunda posição codificam aminoácidos hidrofóbicos (Tabela 12.1B). A distribuição dos códons correlacionada à característica de hidrofobicidade dos aminoácidos demonstra que os códons que codificam o mesmo aminoácido estão organizados em blocos, com exceção da serina. Essa organização em blocos, relacionada ao grau de hidrofobicidade, pode ser evidenciada mesmo para os aminoácidos leucina e arginina, que são codificados por seis códons.

O código genético padrão é utilizado pela maioria dos organismos, existindo, no entanto, algumas variações esporádicas. Estas alterações ocorrem mais frequentemente em genomas mitocondriais, nos quais foram encontradas substituições de aminoácidos. Como exemplo podemos citar o códon AUA, que codifica metionina, e o AAA, que codifica asparagina. No entanto, nos genomas de diferentes organismos, a maioria das alterações no código genético afeta os códons de terminação. Assim, por exemplo, são encontradas alterações em que os códons UAA e UAG passam a codificar glutamina, e o códon UGA passa a codificar triptofano ou cisteína. Podem ocorrer também alterações em que os códons de alguns aminoácidos geram códons de terminação, como os códons da arginina CGG e AGA. Tais alterações parecem ocorrer de maneira independente e só em algumas rotas específicas durante a evolução.

A característica principal do código genético, que correlaciona uma trinca de bases para um aminoácido, define que para cada sequência de mRNA existem três possíveis modos de síntese da proteína, denominadas **fases de leitura**. Na **Figura 12.2** está representada uma sequência de mRNA com os peptídeos originados das fases de leitura dependendo da trinca utilizada para o início da síntese. Neste exemplo, caso a tradução utilize a trinca a partir da primeira base (CAU), será formado um peptídeo contendo apenas histidina; se o início ocorrer a partir da segunda base da trinca (AUC), será produzido o peptídeo contendo isoleucina e, por fim, na terceira fase de leitura, será formado o peptídeo composto exclusivamente de serina. As fases de leitura são fundamentais para a síntese de proteínas, e qualquer alteração na sequência do mRNA terá consequência direta na sequência de aminoácidos da proteína formada. Uma inserção ou deleção de bases pode alterar a sequência de aminoácidos na proteína. Uma fase de leitura que contenha somente trincas que codificam aminoácidos é denominada sequência aberta de leitura (ORF, de *open reading frame*). Portanto, uma região do DNA que codifica uma proteína contém uma ORF, onde a primeira trinca é um códon AUG (metionina) seguido de várias outras trincas e finalizando com um dos códons de terminação.

Atualmente, uma pergunta fundamental refere-se a “como e por que” ocorreu a evolução do código genético. Esta questão vem causando muitas discussões, com várias teorias. Existem, no momento, três teorias principais para explicar a origem e a evolução do código genético. A primeira considera que a definição dos códons para seus aminoácidos está relacionada à afinidade físico-química entre o aminoácido e o anticódon. A segunda teoria, a da

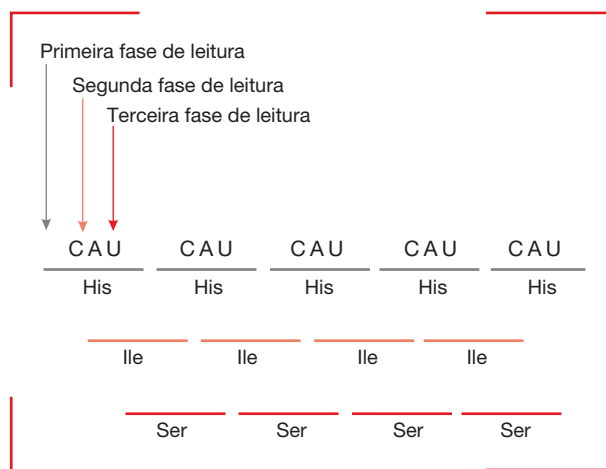


Figura 12.2

Fases de leitura. A figura representa uma sequência de nucleotídeos de uma molécula de mRNA, com três possíveis fases de leitura durante a síntese de proteínas. No exemplo apresentado, a primeira fase de leitura produz um peptídeo contendo apenas histidina (His), na segunda, o peptídeo contém apenas isoleucina (Ile), e na terceira fase, o peptídeo contém só serina (Ser).

coevolução, sugere que a estrutura do códon coevoluiu com as rotas de biossíntese dos aminoácidos. A última teoria, denominada “minimizar o erro”, indica que o fator principal na evolução está relacionado com a seleção de códons que diminuem o erro das mutações durante a tradução. Além dessas três teorias, existe uma outra hipótese, em que o código genético foi estabelecido porque todas as formas de vida são originadas de um ancestral comum. Todas essas teorias não se excluem, e a origem e a evolução do código genético podem ser uma combinação de uma ou mais destas hipóteses.

2. RNAs transportadores

Uma característica essencial de todas as células é a tradução da informação genética na forma de quatro letras (A, T ou U, C e G), que estão presente no DNA e no mRNA, no alfabeto de 20 letras que corresponde aos aminoácidos constituintes das proteínas. A fidelidade da tradução

do mRNA durante a síntese proteica é determinada por dois fatores: a disponibilidade do aminoacil-tRNA contendo o aminoácido correspondente e a seleção precisa desse aminoacil-tRNA no ribossomo.

Os tRNAs são pequenas moléculas polinucleotídicas de fita simples contendo de 74 a 95 bases apresentando uma estrutura muito conservada. A descoberta dos tRNAs foi muito importante para o entendimento da transferência da informação genética, pois o tRNA contém a sequência de bases (anticódon) complementar ao códon do mRNA e carrega o aminoácido específico.

2.1 Estrutura dos tRNAs

Várias moléculas de tRNAs foram isoladas a partir de organismos procarióticos e eucarióticos, todas apresentando uma estrutura secundária semelhante. Essa estrutura, denominada **folha de trevo**, é mantida por meio de pareamento entre sequências complementares na molécula de tRNA (**Figura 12.3**). A partir de um gru-

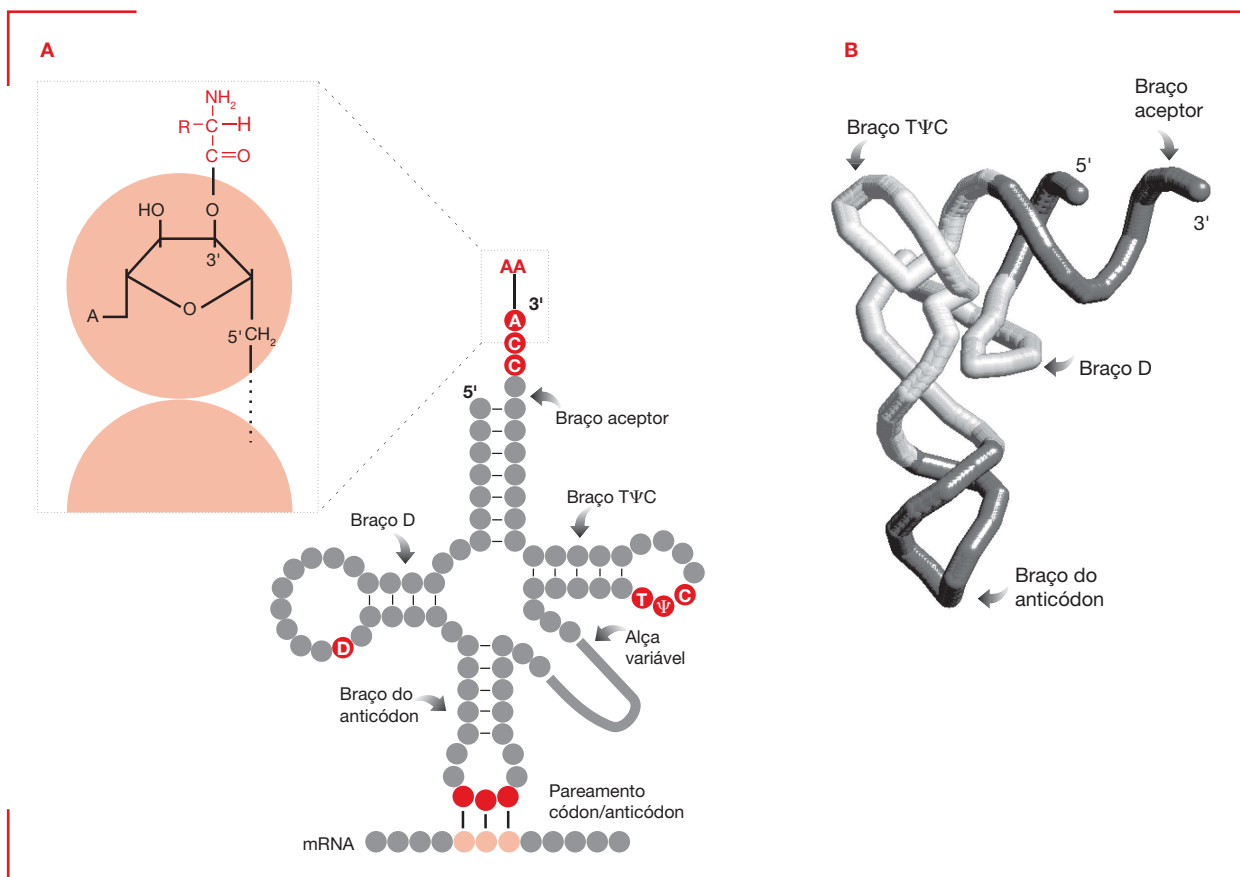


Figura 12.3

Estrutura dos tRNAs. (A) Representação dos pareamentos intramoleculares e denominação dos braços e alças dos tRNAs. As letras representam as bases mais relevantes na estruturados tRNAs. Também é mostrado o pareamento com uma molécula de mRNA (pareamento códon-anticódon) e a ligação do aminoácido (AA) na extremidade 3' do tRNA. O quadro contendo a ampliação das duas últimas bases da extremidade 3' (CA) ressalta a ligação da região carboxi-terminal do aminoácido com o grupamento hidroxílico da posição 3' da ribose. (B) Representação da estrutura terciária dos tRNAs.

pamento fosfato na extremidade 5', a forma de folha-de-trevo apresenta braços conservados contendo regiões pareadas (hastes) e regiões de fita simples (alças), caracterizadas como:

1. Braço acceptor: uma haste com 7 pares de bases (pb) que consiste em uma região pareada entre as extremidades 5' e 3' do tRNA, contendo, na região 3', uma sequência de fita simples conservada, **CCA**, com o grupamento 3'-OH livre. A ligação com o aminoácido ocorre pelo grupamento hidroxílico livre do carbono 2' ou 3' da ribose da última base (adenina) (ver Figura 12.3).
2. Braço D: uma haste com 3 a 4 pb que termina em uma alça de 5 a 7 nucleotídeos, cuja denominação também ocorre em função da presença de bases modificadas (diidouridina; base representada pela letra D).
3. Braço do anticódon: uma haste terminando em um laço que contém a trinca do anticódon no centro da sequência, correspondendo à alça desse braço.
4. Braço T_ψC: uma haste que contém, em geral, cinco pareamentos e, na alça, apresenta a base não usual pseudouridina (Ψ). Essa região do tRNA também pode ser denominada braço T.
5. Braço variável: o braço que apresenta maiores variações entre os tRNAs. Existem tRNAs que possuem, nesse braço, uma sequência de 3-5 bases, ao passo que um outro grupo de tRNAs apresenta esse mesmo braço longo, podendo ter entre 3 e 21 nucleotídeos, com uma haste de até 7 pb. Os tRNAs com braço variável pequeno representam, aproximadamente, 75% dos tRNAs presentes na célula.

A presença de **bases não usuais**, também denominadas bases modificadas, em uma proporção de até mais de 25%, é uma característica dos tRNAs. A definição para bases não usuais considera qualquer anel purínico ou pirimídico que não represente as quatro bases tradicionais A, C, G e U. Essas bases são geradas por meio de modificações em qualquer uma das quatro bases já incorporadas no RNA. Em geral, as modificações ocorrem nas bases já existentes na molécula pelas alterações ou adições realizadas por enzimas específicas que modificam tRNAs. Essas modificações conferem aos tRNAs uma maior versatilidade em estrutura, podem estar relacionadas a diferentes funções e estarem localizadas ao longo da molécula dos tRNAs. Nenhuma dessas modificações é essencial para a manutenção da integridade estrutural do tRNA ou para sua ligação adequada ao ribossomo. No entanto, a posição na estrutura do tRNA, onde estão algumas das bases modificadas, é conservada entre os diferentes tRNAs.

Os tRNAs são denominados em função do aminoácido que representam, existindo pelo menos um tRNA para cada aminoácido. Por exemplo, o tRNA contendo o anticódon da metionina é denominado tRNA^{Met}. Como discutido, quando um tRNA está ligado ao seu aminoácido, denomina-se aminoacil-tRNA (por exemplo, Met-

-tRNA). Após a ligação do aminoácido ao seu tRNA correspondente, a participação do aminoacil-tRNA, durante a síntese da proteína, é governada exclusivamente pelo pareamento códon-anticódon.

Todos os tRNAs dos diferentes organismos apresentam uma estrutura terciária conservada contendo, no entanto, pequenas variações individuais. Essa estrutura em forma de L é a forma funcional dos tRNAs. Nessa estrutura, permanecem os pareamentos entre as bases, característicos da estrutura secundária, surgindo uma estrutura tridimensional com duas hélices duplas em ângulo específico (Figura 12.3B). A haste da região acceptora e a haste T_ψC representam uma hélice dupla contínua, e as hastes D e do anticódon formam a segunda hélice dupla. A região entre essas hélices duplas é aquela que forma a volta da forma L, contendo as alças D e T_ψC. Na estrutura terciária, a região 3' e a região do anticódon estão posicionadas nas duas extremidades do L com uma distância máxima adequada para sua participação na síntese de proteínas. A estrutura dos tRNAs se caracteriza e é mantida pela presença de bases modificadas, incluindo associações por pontes de hidrogênio envolvendo bases não pareadas, com os grupamentos fosfato ou com os grupamentos 2'-OH dos resíduos da ribose, na estrutura secundária. Essas interações terciárias contribuem para a estrutura compacta dos tRNAs.

Os tRNAs apresentam propriedades que são muito importantes: capacidade de representar apenas um aminoácido ligando-se covalentemente a ele; e capacidade de conter uma sequência de trinucleotídeo e o anticódon, que é complementar ao códon do mRNA e representa o aminoácido. Tais propriedades garantem a fidelidade durante a síntese de proteínas, pois será adicionado o aminoácido correto a partir da interação códon-anticódon.

Devido à própria conformação da molécula de tRNA o pareamento com a terceira base do códon é menos rígido. Dessa maneira, ocorre o que se chama de pareamento oscilante (*wobble base-pairing*). O pareamento oscilante permitido é aquele que mantém a distância usual entre as riboses. Assim, não é possível o pareamento entre duas purinas ou duas pirimidinas (distâncias muito grandes ou muito pequenas, respectivamente).

Os pareamentos possíveis, segundo o conceito de oscilação, são mostrados na **Figura 12.4A**. A presença de bases não usuais no anticódon influencia também na capacidade de pareamento tRNA-mRNA. A base não usual inosina, que é uma modificação da base adenosina, pode estar presente na primeira posição do anticódon. Esses anticódons, que possuem a base inosina (I) ocupando a primeira posição (que pareia com a terceira posição do códon), podem parear com três diferentes códons (Figura 12.4B).

A presença de inosina na primeira posição do anticódon é fundamental para o processo de diferenciação de alguns códons. Devido ao pareamento oscilante uma adenosina ou guanossina, na terceira posição do códon, poderiam ser reconhecidas por um tRNA contendo uma

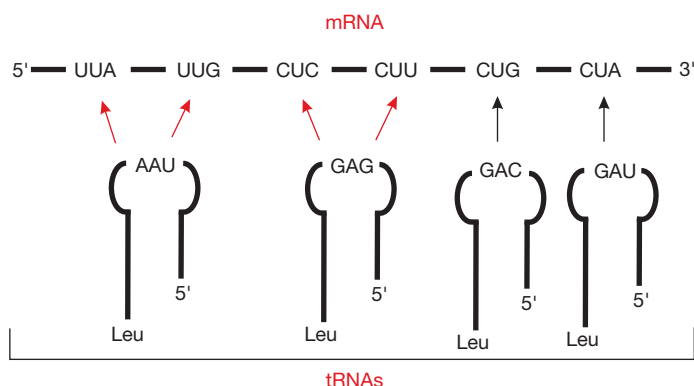
A

Pareamento oscilante entre o tRNA e mRNA	
Base na primeira posição do anticódon	Base na terceira posição do códon
U	A ou G
G	C ou U
I	C ou U ou A

Figura 12.4

Pareamento oscilante. (A) Apresenta os pareamentos oscilantes possíveis de serem realizados entre o mRNA e os diferentes tRNAs. (B) Mostra um exemplo, utilizando os códons para o aminoácido leucina (Leu). As setas indicam que o anticódon pode parer com o códon.

B



uracila. Analisando a tabela do código genético, pode-se observar que o códon para metionina (AUG) difere somente na terceira posição de um dos códons da isoleucina (AUA). Neste caso, poderia ocorrer erro durante a síntese de proteínas, pois haveria pareamento cruzado entre os tRNAs dos dois aminoácidos, devido ao pareamento oscilante da base U do anticódon com as bases A ou G do códon (ver Figura 12.4A). A presença de inosina na primeira posição do anticódon para o aminoácido isoleucina apresenta uma solução para este problema, pois elimina a possibilidade de pareamento com o códon da metionina.

A existência de pareamento oscilante explica a variação do número de tRNAs para os diferentes códons, pois não existem 61 diferentes tRNAs. Para códons em que a degeneração é total, existe a presença de pelo menos dois tRNAs carregando o mesmo aminoácido. Quando existe mais de um tRNA para o mesmo aminoácido, sua denominação é $tRNA_1$ e $tRNA_2$, sendo definidos como **isoaceptores**. Portanto, para o aminoácido prolina, por exemplo, os tRNAs isoaceptores seriam classificados como $tRNA_1^{Pro}$ e $tRNA_2^{Pro}$.

2.2 Ativação dos tRNAs e as aminoacil-tRNA-sintetases

O processo de ligação do aminoácido com o tRNA é realizado por um grupo de enzimas denominadas aminoacil-tRNA sintetases (aaRSs). Devem existir pelo menos 20 diferentes aaRSs, cada uma reconhecendo um aminoáci-

do, e os tRNAs contendo o anticódon específico. O aminoácido e os tRNAs reconhecidos por uma aaRSs recebem a denominação de cognato. A aaRSs que reconhece o aminoácido prolina é denominada proilil-tRNA^{Pro}-sintetase, tendo a prolina como o aminoácido cognato, e os $tRNA_1^{Pro}$ e $tRNA_2^{Pro}$ seriam seus tRNAs cognatos.

As aaRSs realizam a ligação precisa do aminoácido correto ao tRNA cognato em uma reação enzimática dividida em duas etapas. Primeiro, ocorre a ativação do aminoácido (aa) por meio da ligação ao ATP, formando um intermediário aminoaciladenilato (aa-AMP) e liberando pirofosfato (PPi). Após a ativação, o aminoácido é transferido para a extremidade 3' do tRNA correspondente, para ser utilizado posteriormente na síntese de proteínas. As etapas dessa reação estão esquematicamente representadas a seguir.

1. $aa + ATP \rightarrow aa-AMP + PPi$
2. $aa-AMP + tRNA \rightarrow aa-tRNA + AMP$

Os erros durante a síntese de proteínas são minimizados de duas maneiras: pela ligação preferencial do aminoácido cognato e pela correção de erro na ligação dos aminoácidos similares. Por meio deste mecanismo de correção de erro, a frequência de incorporação errônea diminui bastante e contribui para o controle de qualidade da tradução. Essas enzimas possuem especificidade variável, existindo aaRSs capazes de ativar somente um tipo de aminoácido, e outras que podem ativar, também, ami-

noácidos relacionados. No entanto, como as aaRSs realizam a correção do erro, não transferem o aminoácido ativado não cognato para o tRNA. Um exemplo a ser citado, ocorre com a isoleucil-tRNA-sintetase, capaz de ativar tanto isoleucina como valina (**Figura 12.5**). Na presença de tRNA para isoleucina ($tRNA^{Ile}$), ocorre apenas a formação de ile-tRNA, ao passo que a ligação val-AMP é hidrolisada originando val + AMP + $tRNA^{Ile}$.

As aminoacil-tRNA-sintetases são divididas em duas classes dependendo da sua estrutura e atividade. As aaRSs foram classificadas em classe I e II, considerando características como a posição da aminoacilação na extremidade 3' da ribose do tRNA e a topologia e orientação da ligação do tRNA no sítio ativo da aaRSs. A ligação específica entre um aminoácido e seu tRNA, realizada pela aaRSs, sugere um modelo geral de ligação da enzima com o tRNA cognato. Existem três áreas da molécula do tRNA que podem realizar o contato com a aaRSs: a haste acceptora (que deve estar envolvida na interação, pois é o local de ligação com o aminoácido), a haste D e a haste do anticódon. Portanto, a liberação do tRNA não carregado com o aminoácido, pelo mecanismo de correção de erro realizado pelas aaRSs, pode ser considerada como essencial para o desenvolvimento dos diferentes organismos. Informações adicionais são apresentadas no **Material Complementar Online 12.2**.

2.3 tRNAs supressores

As moléculas de tRNA supressores se caracterizam por conterem um anticódon alterado que realiza a leitura do códon introduzindo modificações no peptídeo. Uma variedade desses tRNAs já foi identificada, tanto em bactérias como em organismos eucarióticos, sendo uti-

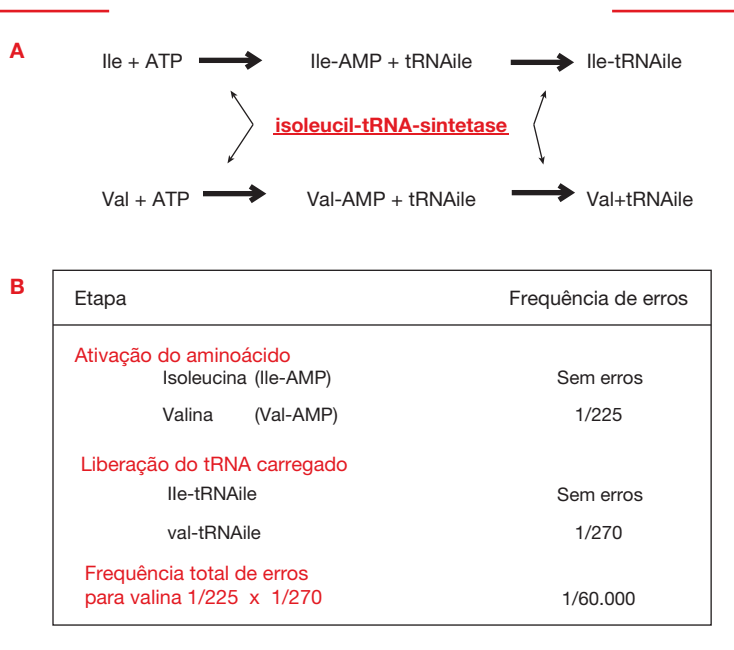
lizados em diferentes estudos genéticos e bioquímicos. Normalmente, durante o processo de tradução, o término da síntese do peptídeo ocorre quando um dos códons de terminação é apresentado no sítio “A” do ribossomo (ver Seção 4.3). No entanto, existem diferentes processos que podem alterar o término da tradução, como a alteração da fase de leitura e a presença de tRNA supressor. A presença de tRNAs supressores para os três tipos de códons de terminação foi demonstrada em organismos procarióticos. Esses tRNAs supressores são capazes de realizar a leitura de um códon de terminação como um códon para um aminoácido, alterando a sequência da proteína.

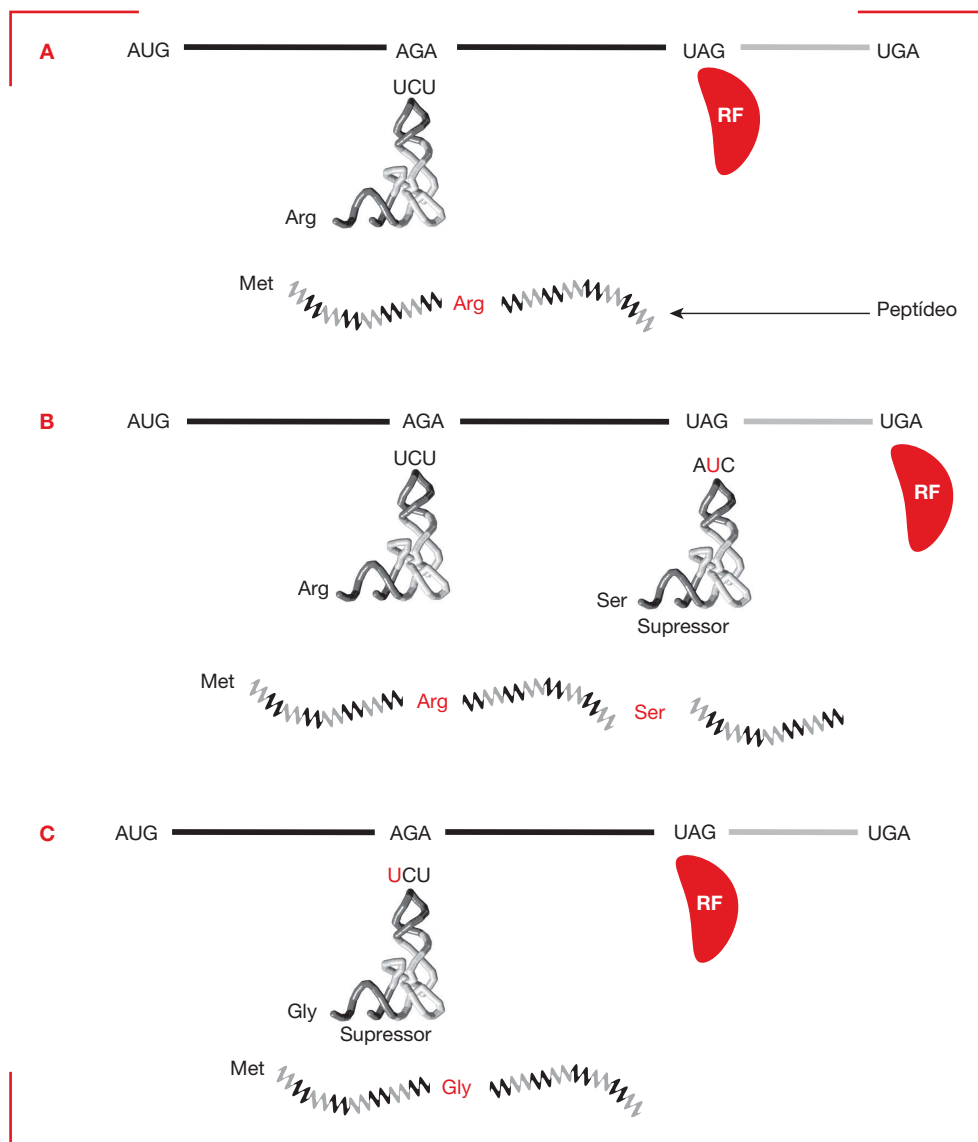
Na **Figura 12.6**, é apresentado um exemplo utilizando dois tipos de tRNAs supressores. Nesse exemplo, o peptídeo inicia na metionina e termina no primeiro códon de terminação (UAG). Células que contêm um tRNA supressor com uma alteração no anticódon podem competir com o processo de terminação, introduzindo um aminoácido adicional (no exemplo seria a serina, **B**) na proteína e, conseqüentemente, alongando a síntese da cadeia peptídica até o próximo códon de terminação (UGA). Os tRNAs supressores também podem alterar o aminoácido na cadeia peptídica, como mostrado na Figura 12.6 **C**. Esta Figura demonstra a presença do tRNA supressor carregando uma glicina com o anticódon mutado capaz de parear com o códon da arginina. Neste tipo de célula, o peptídeo sintetizado pode conter uma substituição de aminoácidos (Arg para Gly).

Os processos de supressão são bastante ineficientes, pois estão sempre competindo com os tRNAs regulares da síntese de proteínas. No entanto, a síntese de um peptídeo, mesmo em pequena escala, originado por supressão, pode ser, em algum momento, o suficiente para garantir a viabilidade do organismo.

Figura 12.5

Mecanismo de controle de erros durante a ligação do aminoácido ao tRNA. (A) Representa a reação de ligação do aminoácido ao tRNA, e **(B)**, a frequência em que podem ocorrer erros. Isoleucina (Ile), valina (Val).



**Figura 12.6**

Exemplos de supressão realizada por tRNAs mutados. **(A)** Representa uma região codificante onde estão apresentados o códon iniciador AUG, um códon AGA para o aminoácido arginina e dois códon de terminação. O fator RF reconhece o primeiro códon de terminação finalizando a síntese do peptídeo. **(B)** Demonstra a tradução da mesma região, mas na presença de um tRNA para a serina, que sofreu uma mutação no anticódon (CGA para CUA). **(C)** Está representada a presença de outro tRNA supressor, onde ocorreu uma mutação no anticódon do tRNA para glicina (CCU para UCU). O símbolo Z representa qualquer aminoácido.

3. Ribossomos

Os ribossomos são formados por duas subunidades contendo moléculas de RNA associadas com proteínas, sendo responsáveis pela decodificação da informação genética, convertendo-a em aminoácidos componentes da proteína. Eles representam próximo a 30% da massa celular, existindo, respectivamente, um número de 10^5 e 10^6 ribossomos em bactérias e em células de mamíferos. Em geral, em células de metabolismo ativo, a maioria dos ribossomos está realizando a síntese de proteínas.

Podemos correlacionar a função geral dos ribossomos com aquela encontrada nas RNA ou DNA-polimerases, pois apresentam uma atividade enzimática processiva que inicialmente reconhece os elementos do mRNA, determinando onde deve iniciar a síntese

da proteína e, logo após, no estágio de alongamento da cadeia peptídica, perdem o contato com os elementos iniciadores.

3.1 Composição e estrutura

Os ribossomos são grandes complexos ribonucleoproteicos, em que as duas subunidades apresentam funções especializadas. A subunidade menor é responsável pela decodificação do mRNA, e a maior contém a atividade de peptidil-transferase essencial para a formação da cadeia peptídica (ver Seção 4.2). A união das duas subunidades gera os sítios ativos de ligação às três moléculas de tRNA necessárias para a síntese de proteínas. Nos ribossomos, também existe a formação de um túnel para a saída da cadeia da proteína nascente. O processo de tradução se encontra, física e funcionalmente, acoplado aos mecanismos de enrolamento e direcionamento das proteínas

sintetizadas. Existem evidências demonstrando que os ribossomos promovem a formação da estrutura secundária das cadeias nascentes de proteínas.

A **Figura 12.7A** apresenta a composição dos ribossomos de organismos procarióticos e eucarióticos, demonstrando uma estrutura organizacional semelhante. Em todos os organismos, as subunidades dos ribossomos se encontram separadas, unindo-se para formar a estrutura funcional. Na síntese de proteínas, os ribossomos são muito rápidos e eficientes podendo produzir continuamente uma proteína a uma velocidade de 20 ligações peptídicas por segundo. A composição dos ribossomos, contendo diferentes proteínas (proteínas ribossômicas ou r-proteínas) e moléculas de rRNA (RNA ribossômico) na proporção de aproximadamente 1:2, tem sido mantida durante a evolução.

Em procariotos, a subunidade menor, denominada 30S (definido pelo seu coeficiente de sedimentação), contém uma molécula de rRNA (16SrRNA) e entre 20-21 r-proteínas, ao passo que a subunidade maior (50S) apresenta duas moléculas de rRNA (23SrRNA e 5SrRNA) e um total de 30-35 proteínas ribossômicas diferentes. Ribossomos encontrados em organismos eucarióticos são estru-

turalmente maiores (80S) que os de procarióticos (70S). Ribossomos de eucariotos se diferenciam também no tamanho das subunidades (40S e 60S), no tamanho dos rRNAs (18S, 28S, 5S e 5,8S) e na quantidade de r-proteínas que formam as subunidades (ver Figura 12.7A).

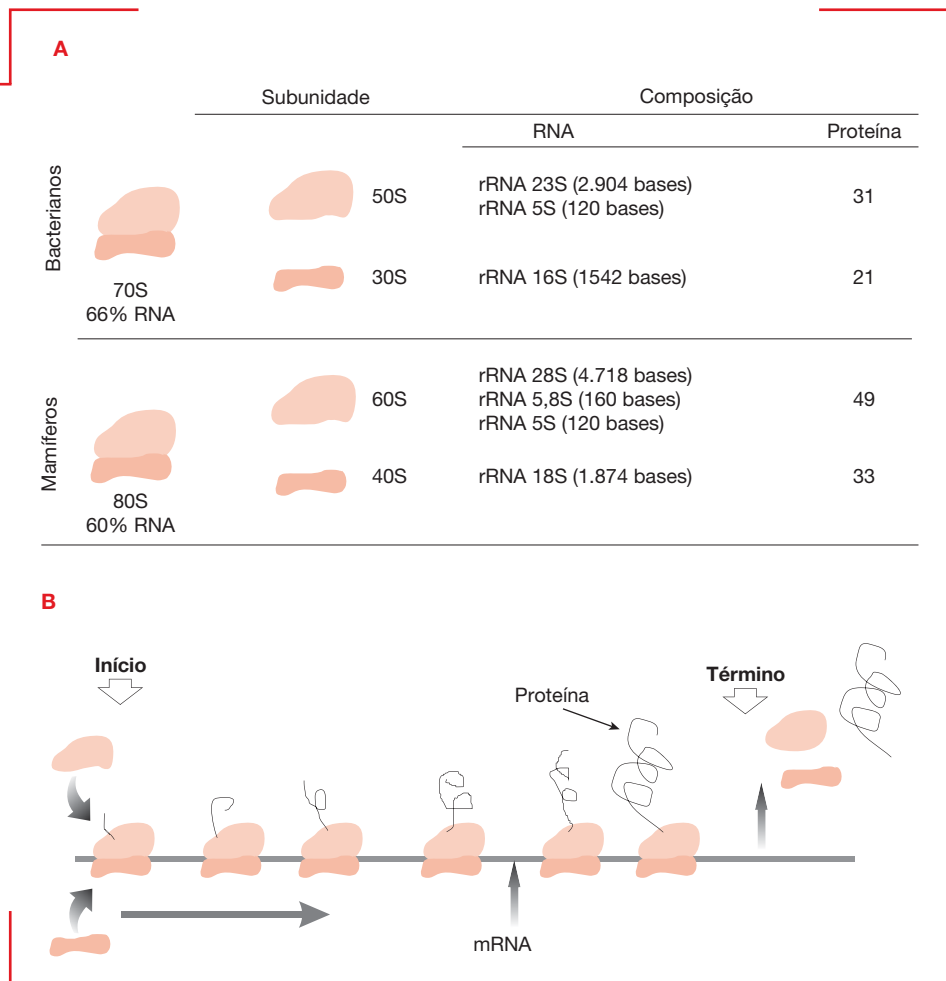
O ribossomo completo da bactéria *Escherichia coli* apresenta uma estrutura assimétrica, sendo composto por uma região basal e por outra contendo a cabeça ou protuberância.

Essa estrutura completa do ribossomo 70S deve ser mantida por associações discretas entre determinadas regiões das duas subunidades. Essa mesma estrutura assimétrica foi também demonstrada em ribossomos de organismos eucarióticos.

Atualmente, a maioria dos mecanismos relacionados às funções dos ribossomos está bem elucidada, devido aos trabalhos de cristalografia dos ribossomos e determinação da sua estrutura. A região de interface entre as duas subunidades do ribossomo é composta principalmente pelas moléculas de rRNA, incluindo a região dos sítios ativos. Portanto, os ribossomos estão sendo considerados como ribozimas (moléculas de RNA contendo

Figura 12.7

Composição dos ribossomos. (A) Os ribossomos completos estão identificados pelos seus tamanhos 70S e 80S, assim como a subunidade maior (50S e 60S) e a subunidade menor (30S e 40S). A porcentagem indicada abaixo do número 70S ou 80S representa a composição em quantidade de RNA. A coluna referente a composição de proteínas indica o número de proteínas que compõem cada subunidade, e a coluna do RNA indica o tipo e o tamanho (entre parênteses) dos RNAs presentes na subunidade. **(B)** Estão representados vários ribossomos realizando simultaneamente a síntese de proteínas. A seta horizontal representa a direção do movimento dos ribossomos durante a síntese do peptídeo.



atividade enzimática), pois nos seus sítios ativos não são encontradas as r-proteínas, isto é, são constituídos essencialmente pelas moléculas de RNA ribossômicas. As proteínas ribossômicas de cada subunidade estão embebidas dentro do complexo conformacional dos rRNA, sendo essenciais para a manutenção da estrutura dinâmica, que é responsável por diversas atividades dos ribossomos (decodificação precisa, formação da ligação peptídica e eficiente polimerização).

Durante a síntese de proteínas, um grande número de ribossomos atua simultaneamente realizando a síntese proteica, e desta maneira, aumentando a eficiência da tradução. Esses ribossomos recebem a denominação de polissomos, que realizam a adição de cada aminoácido na cadeia nascente enquanto se deslocam na molécula de mRNA molde (Figura 12.7B). A localização e o arranjo dos polissomos no mRNA maximiza a distância entre as cadeias nascentes de proteínas, prevenindo interações desfavoráveis.

3.2 Estrutura e dinâmica dos sítios ativos do ribossomo

Cada subunidade dos ribossomos possui três sítios ativos para ligação das moléculas de tRNA. O **sítio A** liga preferencialmente aminoacil-tRNA que será incorporado na cadeia polipeptídica; o **sítio P** liga tanto Met-tRNA (tRNA iniciador) como posiciona o peptidil-tRNA; e o **sítio E** é ocupado com os tRNAs desacetilados antes da sua dissociação do ribossomo (ver Figura 12.1). Portanto, com exceção do tRNA iniciador, todos os aminoacil-tRNAs se associam ao ribossomo, inicialmente pelo sítio A, depois ocupam o sítio P e realizam sua dissociação pelo sítio E.

Quando os tRNAs se encontram ocupando os sítios A e P, o anticódon dos tRNAs está pareado com o códon do mRNA, que está associado à subunidade menor, e o braço acceptor contendo o aminoácido ou o peptídeo, que está localizado na subunidade maior. Portanto, como os sítios dos ribossomos estão localizados nas duas subunidades, deve ocorrer uma ação cooperativa das subunidades durante a síntese de proteínas. Resumidamente, podemos relacionar a subunidade menor com as funções de apresentar o caminho que deve ser realizado (mRNA), conter a região decodificadora e o mecanismo para controlar a fidelidade da tradução. Na subunidade maior, estariam presentes o centro catalítico (ligação peptídica), a atividade de polimerização dos aminoácidos e o túnel de saída da proteína.

Durante a síntese de proteína, a transformação da informação genética em proteínas depende exclusivamente do pareamento códon-anticódon entre mRNA e tRNA. No entanto, os ribossomos controlam a precisão da leitura do códon através das várias regiões de sua estrutura.

Em organismos procarióticos, para o início da síntese de proteínas, é necessária a ligação entre o mRNA e o ribossomo. O tamanho dessas sequências do mRNA recobertas pelos ribossomos pode variar de 35 a 40 nu-

cleotídeos. Essa região do mRNA apresenta duas características principais: (1) deve conter o códon de iniciação AUG (raramente GUG ou UUG); (2) deve possuir a sequência de nucleotídeos, denominada sítio de ligação dos ribossomos (RBS) ou sequências S/D (Shine-Dalgarno), complementar, de forma parcial, a uma região 3' do rRNA16S. A sequência apresentada a seguir representa um consenso obtido após a análise de um grande número de genes de procariotos que localiza-se a aproximadamente 7 nucleotídeos do códon AUG.

```
5'xxxAGGAGGxxxxxxxAUG 3' – região do mRNA
3'xxxUCCUCCxxxxxx 5' – região do rRNA16S
Met – região aminoterminal da proteína
```

Apesar de existir certa similaridade entre as regiões 3' dos rRNAs de organismos procarióticos e eucarióticos, não existe evidência que demonstre a ligação do mRNA eucariótico com o rRNA 18S. O modelo proposto para o início da síntese de proteínas, em eucariotos, baseia-se no reconhecimento de CAP em 5' pela subunidade 40S (ver Seção 4.1.2). Posteriormente, essa subunidade desloca-se ao longo do mRNA, até encontrar o primeiro códon de iniciação AUG.

No início da síntese de proteínas, o códon iniciador é, na maioria das vezes, um AUG, mas em bactérias podem ser utilizados GUG ou, mais raramente, UUG. O códon AUG codifica metionina, sendo que dois tRNAs podem carregar esse aminoácido. Em grande parte dos organismos procarióticos e das mitocôndrias, o tRNA iniciador (tRNA_i) está ligado a uma metionina contendo seu radical amino formilado, gerando a molécula de N-formil-metionil-tRNA (tRNA^{Met}). Esse tRNA é utilizado apenas como iniciador, reconhecendo os códons AUG, GUG e, raramente, UUG. A outra molécula de tRNA carregando uma metionina (tRNA^{Met}) reconhece os códons AUG internos, nunca participando do complexo de início da tradução. Por isso, a leitura dos códons AUG e GUG depende de sua localização no mRNA. Quando AUG está localizado no início, é lido como uma formil-metionina; quando está inserido na região codificadora, é lido como metionina. O mesmo fundamento é utilizado para o códon GUG que, quando no princípio da síntese, é utilizado como iniciador é lido como formil-metionina, ao passo que, internamente, na região codificadora, corresponde a uma valina. Em organismos eucarióticos, também existem dois tRNAs carregando metionina e somente um deles é utilizado como iniciador.

Os tRNAs iniciadores apresentam duas características que os diferenciam dos demais tRNAs, permitindo sua participação na formação do complexo de início da síntese de proteínas: somente o tRNA_i apresenta a capacidade de ligação ao sítio P parcial, presente na subunidade menor do ribossomo (30S ou 40S); e o tRNA_i é o único dos tRNAs capaz de se ligar ao fator de iniciação IF2 ou eIF2 (ver Seção 4.1). Essas duas características podem ser dependentes da estrutura diferenciada determinada para

o tRNA iniciador. Um exemplo são os tRNA iniciadores dos procariotos, que apresentam diferenças em relação aos outros tRNAs, o braço acceptor contém cinco bases não pareadas, ao contrário dos outros tRNAs (quatro bases não pareadas), e o braço do anticódon apresenta três pares GC na região da haste.

4. Síntese de proteínas

O processo de síntese de proteínas pode ser dividido em três etapas: início, alongamento da cadeia peptídica e término. Os elementos já discutidos, mRNA, tRNA e ribossomos, são essenciais e participam nas três etapas. No entanto, tanto em organismos procarióticos como eucarióticos, outros elementos, como proteínas, são necessários durante todas as etapas da síntese proteica. Essas proteínas, denominadas fatores, recebem uma identificação ou nomenclatura relacionada à fase da síntese de proteínas na qual participam. Assim, fatores do início da síntese proteica são identificados como IF, do alongamento como EF e do término como RF. Essa nomenclatura é utilizada também em organismos eucarióticos, sendo que nos fatores equivalentes em eucarióticos, foi adicionada a letra “e” antes da sigla, ficando eIF, eEF e eRF.

Os mecanismos utilizados na síntese de proteínas ou tradução são bastante conservados entre os organismos procarióticos e eucarióticos, existindo, no entanto, algumas diferenças significativas. Será discutido, a seguir, cada uma das etapas, considerando os mecanismos comuns entre os organismos e especificando as diferenças importantes existentes.

4.1 Início

Para o início da síntese de proteínas são necessários o mRNA, o tRNA iniciador (tRNA_i) e a subunidade menor do ribossomo. Na formação do complexo de início da tradução, somente o tRNA iniciador apresenta a característica de ligação ao sítio P parcial na subunidade menor do ribossomo (**Figura 12.8**). Além destes elementos, incluem-se também a participação dos fatores de início da tradução (IF ou eIF). Nessa etapa, em todos os organismos, as subunidades do ribossomo devem estar dissociadas para que haja a formação do complexo de iniciação. O fator IF₃, ou eIF₃, interage com a subunidade menor do ribossomo, impedindo a sua ligação com a subunidade maior, mantendo, dessa forma, o ribossomo dissociado. Um segundo fator iniciador, o IF₂ ou eIF₂, apresenta a função de se ligar ao tRNA iniciador, permitindo a interação deste complexo (IF₂-tRNA_i ou eIF₂-tRNA_i) com a subunidade menor do ribossomo. A **Figura 12.8** mostra um esquema comparativo da atividade destes fatores no início da síntese de proteínas de organismos procarióticos e eucarióticos. As características dos fatores envolvidos do início da síntese de proteínas são apresentados no **Material Complementar Online 12.1**.

4.1.1 Procariotos

Em bactérias, o início da síntese de proteínas ocorre por meio da formação de vários complexos intermediários entre o ribossomo, o tRNA iniciador e o mRNA. Esses complexos diferem em composição e estrutura, sendo formados em etapas sucessivas do processo de início da tradução. Se compararmos os domínios Eukarya, Archaea e Bacteria, o início da tradução, nesse último, contém o menor número de elementos atuando em *trans*, como fatores de tradução.

O processo de tradução inicia com a subunidade 30S do ribossomo dissociado da subunidade maior pela interação com o fator IF₃. Os outros dois fatores (IF₁ e IF₂) posteriormente se ligam a este complexo para favorecer o posicionamento do tRNA iniciador e do mRNA (ver **Figuras 12.8** e **12.9**). Portanto, os três fatores de início da tradução atuam cineticamente, controlando o processo que garante a precisão e a fidelidade da montagem do complexo de iniciação. Outra característica específica de alguns organismos procarióticos é a presença de formil-metionil-tRNA, necessário para as interações de alta afinidade com o fator IF₂, e a presença de sequências RBS no mRNA para o pareamento com o 16SrRNA.

Por meio de estudos de cristalografia, a formação dos três principais complexos durante a tradução pode ser considerada da seguinte forma: (1) complexo pré-30S_{IC} contendo a subunidade 30S, com os três fatores (IF₁, IF₂ e IF₃), associado ao mRNA; (2) complexo 30S_{IC} onde ocorre uma alteração conformacional do complexo, permitindo o pareamento do códon AUG com o anticódon do tRNA no sítio P; (3) complexo de início da tradução 70S onde ocorre a ligação da subunidade maior com ativação do mecanismo de hidrólise de GTP associado a IF₂. Com a ligação da subunidade 50S, os fatores se dissociam do complexo (ver **Figura 12.9**).

4.1.2 Eucariotos

Em organismos eucarióticos, o início da síntese de proteínas também pode ser considerado como contendo três principais complexos intermediários. Os processos que conduzem a formação do ribossomo 80S consistem em vários estágios interligados, sendo mediados pelos fatores de início da síntese de proteínas (**Figura 12.10**).

Etapa 1: Formação do complexo de pré-iniciação 43S

Neste estágio, ocorre a seleção do tRNA iniciador pelo fator eIF₂ e a formação do complexo terciário eIF₂-GTP-Met-tRNA_i (ver **Figura 12.8**). Os fatores de iniciação eIF₁, eIF_{1A} e eIF₃ se ligam de forma cooperativa à subunidade 40S, promovendo ou facilitando a ligação do complexo terciário eIF₂-GTP-Met-tRNA_i. Esses fatores são também essenciais para a formação do complexo 43S/ mRNA e para o reconhecimento do códon iniciador.

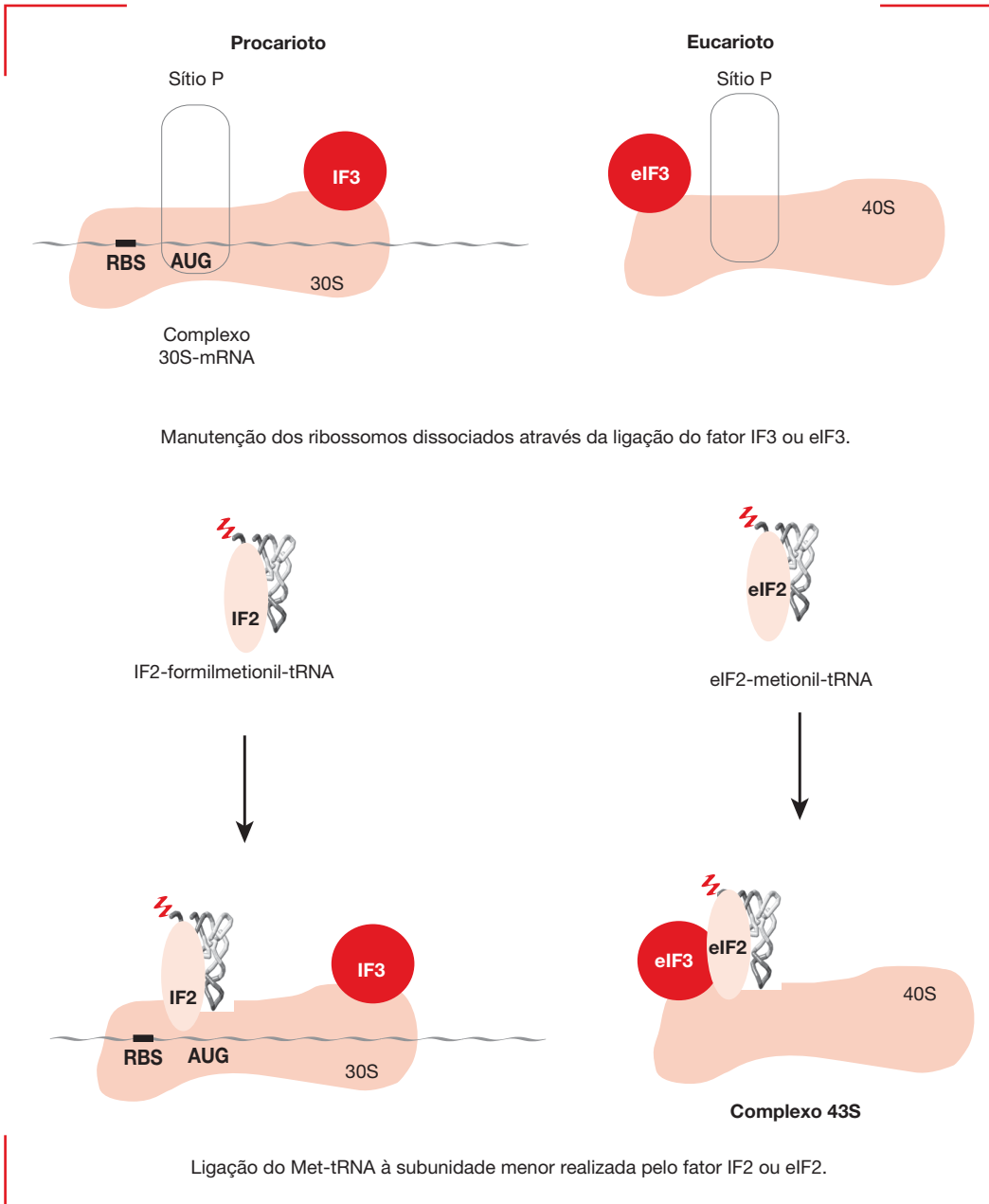


Figura 12.8

Início da síntese de proteínas. A figura demonstra as etapas conservadas na formação do complexo de início da síntese de proteínas em organismos procarióticos e eucarióticos. RBS representa o sítio de ligação do ribossomo. Fatores de início da síntese de proteínas em procariotos IF2 e IF3 e de eucariotos eIF2 e eIF3 estão em . O símbolo ⚡ se refere ao aminoácido ligado ao tRNA iniciador.

A dissociação do ribossomo também é favorecida pela atividade do fator eIF6. Esse fator apresenta uma atividade antiassociativa ligando-se à subunidade maior (60S) do ribossomo e prevenindo sua interação com a subunidade 40S. No nucléolo, eIF6 é parte componente das partículas pré-ribossômicas, sendo necessário para a biogênese da subunidade 60S.

Etapa 2: Ligação do complexo 43S ao mRNA

Neste estágio, ocorre a ligação do complexo 43S ao mRNA pelo reconhecimento da região 5'-CAP. Um dos fatores do complexo eIF4F, o fator eIF4E, reconhece a extremidade 7-metilguanocap dos mRNAs ligando-se e facilitando a ligação do complexo ribossômico 43S. Essa

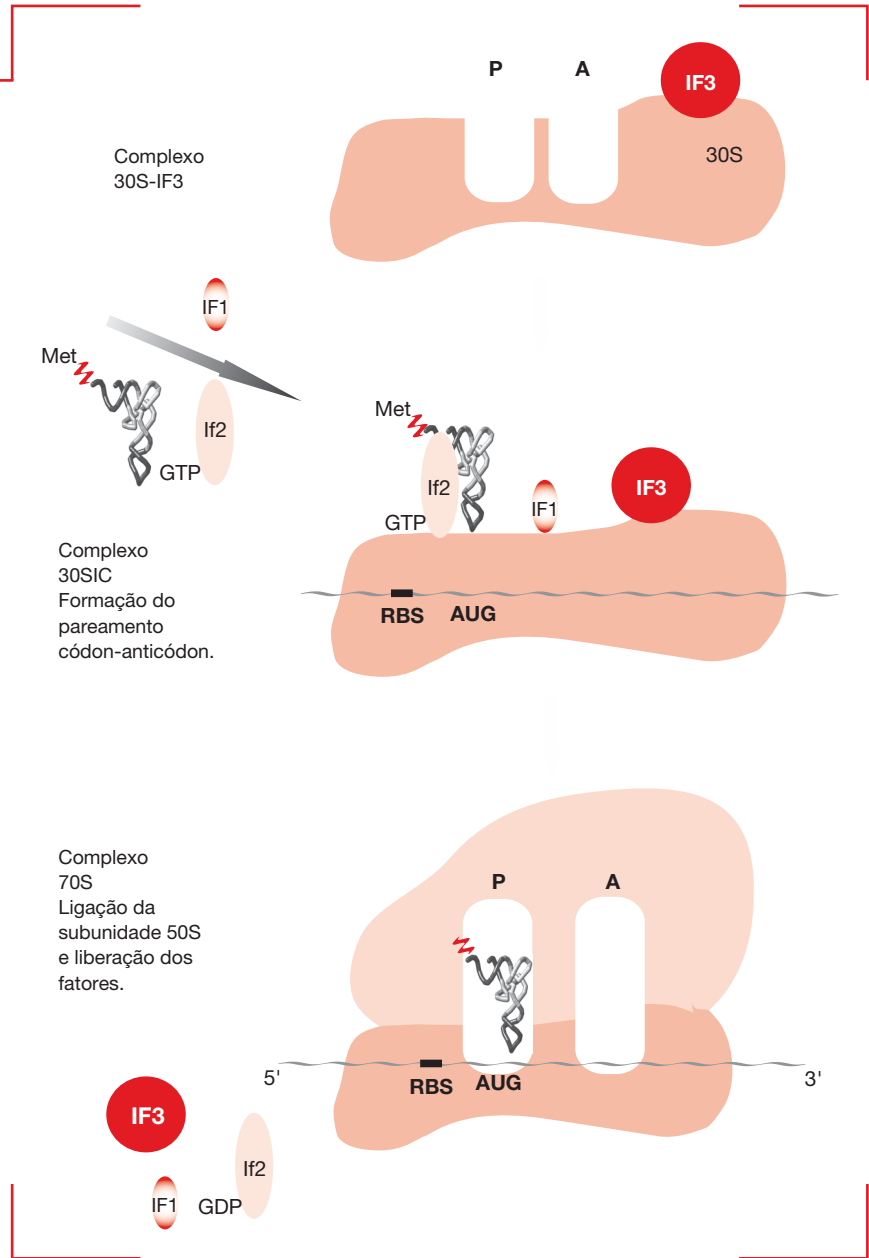
ligação é estimulada pela proteína PABP (*polyA-binding protein*) que se encontra ligada à extremidade 3' poli(A) dos mRNAs. Portanto, a ligação do complexo 43S ao mRNA conduz a circularização desse mRNA propiciando a ligação da região de CAP e poli(A) pela interação do fator eIF4G (outro fator do complexo eIF4F) com PABP. Neste estágio, o complexo 43S encontra-se competente para realizar a varredura na região 5' do mRNA na presença de eIF1 e eIF1A.

Etapa 3: Formação do complexo 48S

O complexo 48S se forma quando o 43S encontra o códon iniciador dentro de um contexto favorável estabelecendo as interações códon-anticódon. Nessa etapa, o complexo

Figura 12.9

Formação do complexo de início da tradução em procariotos. Esquema simplificado da formação do complexo iniciador da tradução. As letras “A” e “P” representam respectivamente os sítios aminoacil e peptidil do ribossomo.



43S realiza a varredura do mRNA na busca do primeiro códon AUG, considerado o códon iniciador. Em células de mamíferos, esse códon iniciador necessita estar em um contexto adequado, denominado sequência Kozak (GCC(A/G)CCAUGG), para ser selecionado eficientemente. As bases mais importantes desta sequência consenso são: uma purina na posição -3 e um G na posição +4 (relativos ao A do AUG).

A seleção do AUG é a primeira etapa irreversível conduzindo para o início do processo de tradução. No momento do pareamento códon-anticódon ocorre, de maneira irreversível, a hidrólise do GTP associado a eIF2. Essa hidrólise é promovida pelo fator iniciador eIF5 por sua atividade de GTPase. Após a hidrólise, o

fator GDP-eIF2 possivelmente se dissocia do complexo, posicionando o Met-tRNA_i no sítio peptidil da subunidade 40S.

Etapa 4: Formação do complexo iniciador 80S

A dissociação dos primeiros fatores do complexo ocorre durante a hidrólise de GTP-eIF2, em que são liberados eIF1 e eIF1A. A ligação da subunidade 60S ao complexo deve ocorrer simultaneamente com a liberação de GDP-eIF2 (estimulada pelo fator eIF5B); e no momento da ligação da subunidade maior, em que ocorre a dissociação dos outros fatores de início da síntese de proteínas. Na formação do ribossomo 80S, o Met-tRNA_i se encontra posicionado, por meio do pareamento códon-anticódon, no sítio peptidil, dessa forma,

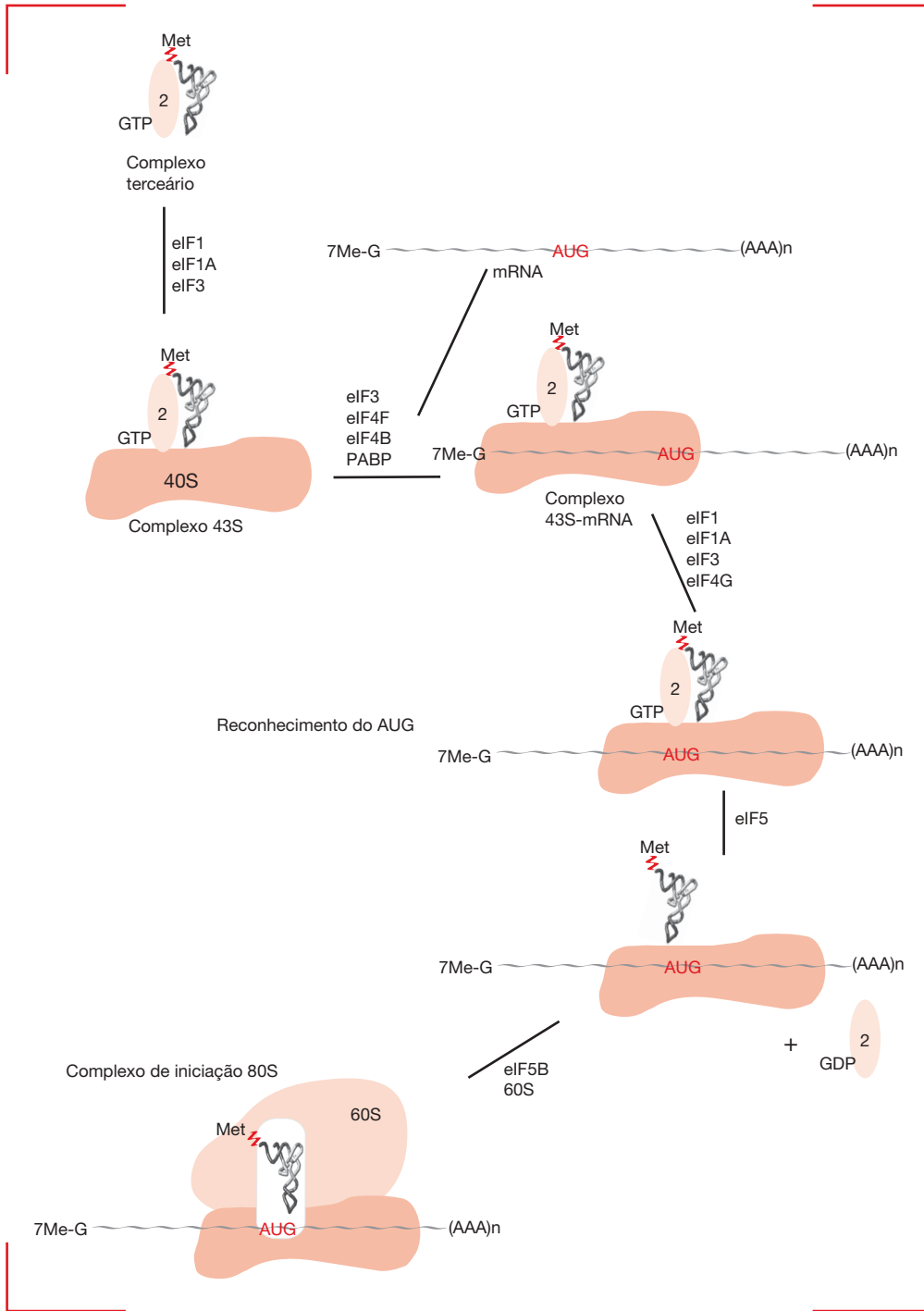


Figura 12.10

Formação do complexo de início da tradução em eucariotos. Representação esquemática da formação do complexo de início da síntese de proteínas. Apenas as etapas principais estão apresentadas, e o fator eIF2 está representado somente pelo seu número (eIF2=2). No mRNA estão marcados CAP (7-MeG) e poli(A) (AAA)n. As setas representam a direção das reações.

liberando o sítio A (aminoacil) para o recebimento dos outros tRNAs, permitindo a etapa de alongamento da cadeia polipeptídica.

O reconhecimento do início da síntese de proteínas é uma etapa decisiva para a célula, pois erros na tradução são muito danosos. Os organismos investiram muitos recursos para assegurar que o códon iniciador correto seja utilizado durante a tradução. Isto pode ser observado no

número de fatores utilizados no início da síntese de proteínas em leveduras, onde aproximadamente metade dos fatores iniciadores está envolvida com o reconhecimento e a seleção do códon iniciador.

4.2 Alongamento da cadeia polipeptídica

Durante a síntese de proteínas, a incorporação de cada aminoácido na cadeia polipeptídica nascente pode ser se-

parada sequencialmente em: decodificação, transferência do peptídeo e translocação.

4.2.1 Decodificação e atividade de peptidil-transferase

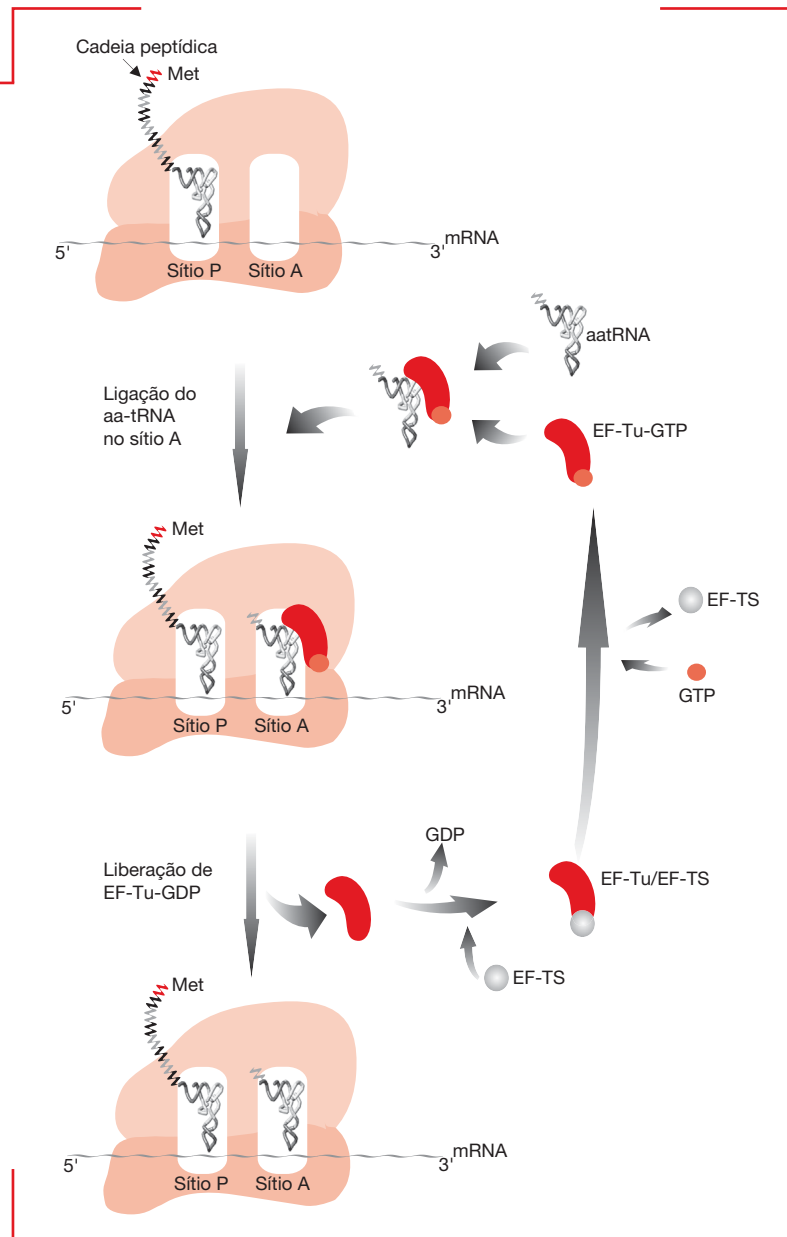
A fidelidade na tradução é assegurada por duas etapas durante o processo de alongamento da cadeia peptídica. A primeira etapa pode ser considerada a seleção do códon correto pela interação códon-anticódon (decodificação). O ribossomo (70S ou 80S) participa desse processo monitorando a seleção do aa-tRNA correto antes dele se posicionar definitivamente no sítio A. Assim como o que ocorre na formação do complexo de iniciação, existe a participação de fatores associados ao alongamento da

cadeia proteica. Em organismos procarióticos, o fator EF-Tu-GTP é responsável pela sua ligação a todos os aa-tRNAs (menos o Met-tRNA iniciador) e a apresentação no sítio A do ribossomo (**Figura 12.11**). O posicionamento do aa-tRNA cognato (correto) pelo reconhecimento do códon resulta na ativação do centro catalítico da subunidade maior contendo atividade GTPásica, que realiza a hidrólise do GTP associado ao fator EF-Tu. Após a hidrólise, o fator EF-TU-GDP é liberado do tRNA e do ribossomo, permitindo a acomodação da região aminoacil do tRNA no sítio correto do ribossomo, o que possibilita a realização da ligação peptídica.

Durante o alongamento da cadeia polipeptídica, o fator EF-Tu-GTP pode apresentar no sítio A qualquer aa-

Figura 12.11

Alongamento da cadeia peptídica. A figura demonstra as etapas da síntese proteica em procariotos e inicialmente a apresentação do aminoacil-tRNA pelo fator EF-Tu-GTP e seu posicionamento no sítio A do ribossomo e, logo após, o pareamento códon-anticódon e a liberação do fator EF-Tu-GDP. A reciclagem do fator EF-Tu pelo fator EF-Ts está demonstrada ao lado direito da figura. As setas representam a direção das reações. Na cadeia peptídica, a metionina (Met) está indicada.



-tRNA. No entanto, a etapa irreversível de hidrólise do GTP ocorrerá com mais rapidez na condição de aa-tRNA cognato posicionado no sítio A. Quando não ocorre o pareamento correto códon-anticódon, mesmo após a liberação do EF-Tu-GDP, o ribossomo ainda pode eliminar o aa-tRNA não cognato. Este seria um mecanismo de correção de erro, permitindo que o crescimento da cadeia peptídica seja reiniciado pelo posicionamento do aa-tRNA cognato. Por fim, o posicionamento do aa-tRNA cognato no sítio A direciona o processo para a realização da ligação peptídica (transferência do peptídeo).

O fator EF-Tu-GDP é reciclado por meio da atividade do fator EF-Ts, gerando novamente EF-Tu-GTP (ver Figura 12.11). O fator EF-Ts desloca o GDP formando um complexo de fatores combinados EF-Tu/EF-Ts e, finalmente, EF-Ts é deslocado pelo GTP regenerando o fator EF-Tu-GTP. Em organismos eucarióticos, os fatores correspondentes são denominados eEF1 α -GTP, responsável pela ligação e apresentação dos aa-tRNAs, e o fator eEF1 $\beta\gamma$ (equivalente ao EF-Ts) que regenera a forma ativa de eEF1 α -GTP.

O posicionamento do aa-tRNA cognato no sítio A resulta na formação do complexo de pré-translocação do ribossomo. Neste complexo, ocorre a formação da ligação peptídica catalisada pela subunidade maior do ribosso-

mo. A atividade responsável pela formação dessa ligação peptídica é denominada peptidil-transferase. O mecanismo de catálise da atividade de peptidil-transferase envolve o ataque nucleofílico do grupamento aínico do aminoácido localizado no sítio A ao grupamento carboxílico do peptídeo posicionado no sítio P. Dessa forma, ocorre a ligação peptídica, sendo o peptídeo transferido para o tRNA posicionado no sítio A do ribossomo (Figura 12.12). Como resultado desta reação, o sítio P fica ocupado pelo tRNA não carregado, e o sítio A com o tRNA carregando a cadeia peptídica que está sendo sintetizada (peptidil-tRNA).

4.2.2 Translocação

A segunda etapa relacionada à precisão na síntese de proteínas durante o alongamento da cadeia peptídica é denominada translocação. Logo após o término da reação de peptidil-transferase ocorre espontaneamente um movimento dos tRNAs. A região da extremidade 3' do tRNA, que está ocupando o sítio P (forma P/P), se movimenta para o sítio E e a região 3', contendo o peptídeo do tRNA no sítio A (forma A/A), passa a ocupar a região do sítio P da subunidade maior (Figura 12.13). Este movimento resulta na formação de uma forma híbrida, em que o tRNA do sítio A gera um híbrido A/P, e o tRNA desacetilado adquire uma forma híbrida P/E. Nessas

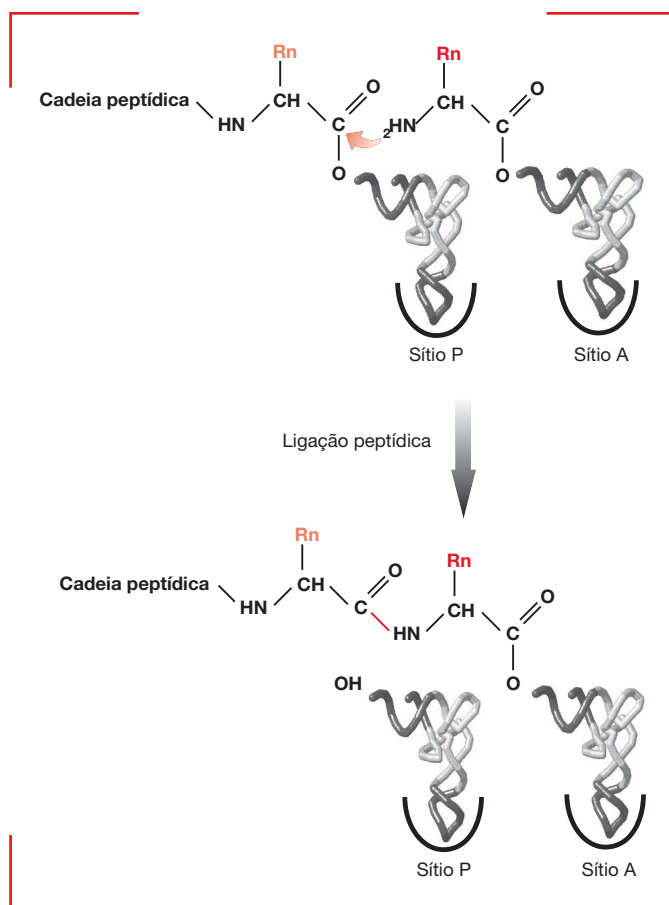


Figura 12.12

Representação esquemática da atividade de peptidil-transferase. Rn representa a cadeia lateral de qualquer aminoácido e a seta indica o ataque nucleofílico originando a ligação peptídica. Os sítios A e P estão representados para demonstrar os sítios do ribossomo.

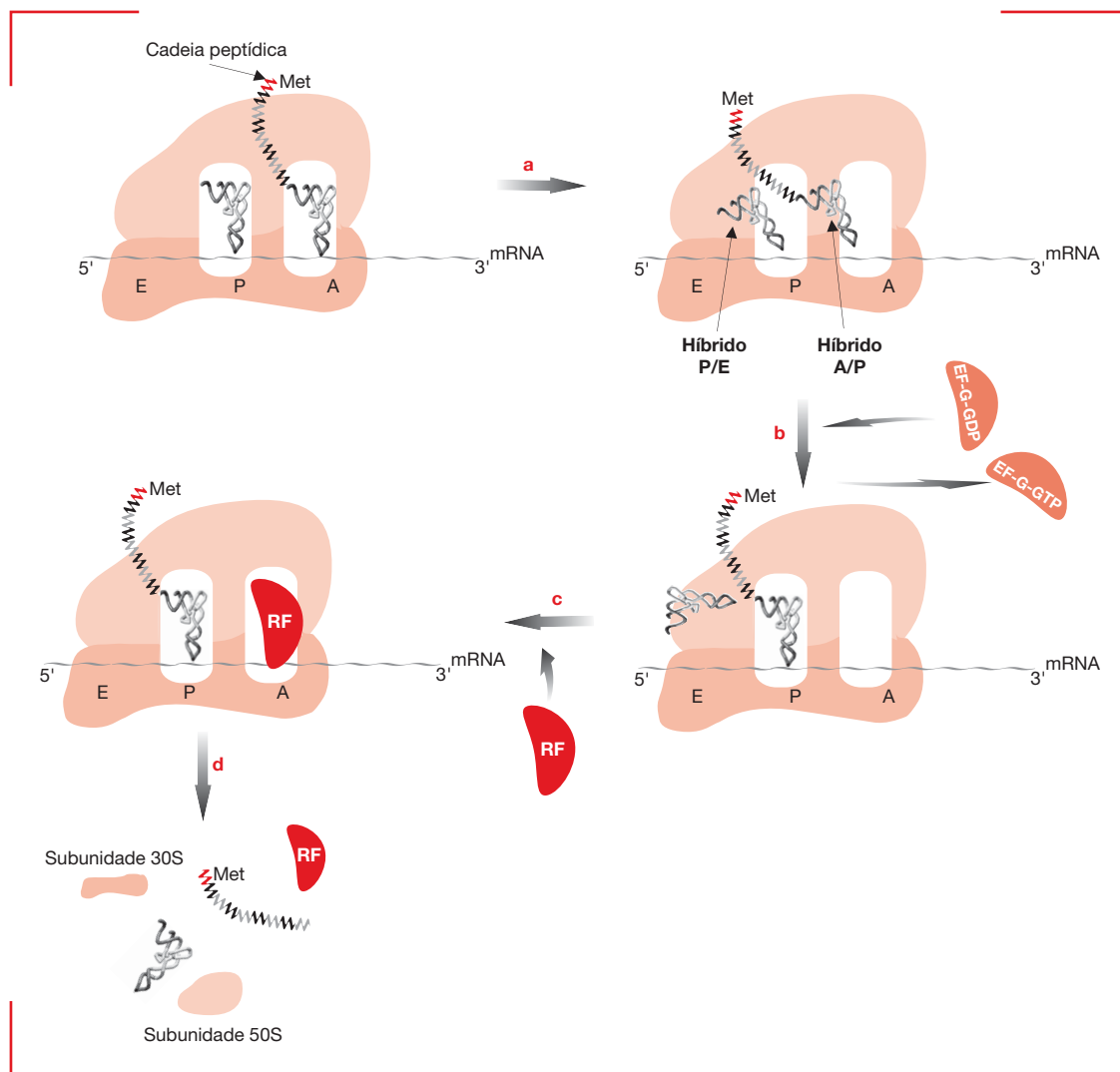


Figura 12.13

Movimento de translocação e término da síntese de proteínas. A figura esquemática demonstra o movimento de translocação durante o alongamento da cadeia peptídica (a e b) e o término da síntese de proteínas (c e d). **(a)** Representa a formação dos híbridos A/P e P/E, gerados durante o movimento espontâneo dos tRNAs em relação à subunidade maior. **(b)** Movimento final de translocação estimulado pelo fator EF-G. O aparecimento de um códon de terminação no sítio A **(c)** propicia a ligação de um fator de terminação (RF), que finaliza a síntese de proteínas liberando o peptídeo, o tRNA, o RF e as duas subunidades do ribossomo. As setas representam a direção das reações. Na cadeia peptídica, a metionina (Met) está indicada.

formas, os tRNAs se movimentam somente em relação à subunidade maior do ribossomo. No entanto, o movimento de translocação total dos tRNAs é estimulado por um fator denominado EF-G em procariotos e EF-2 em eucariotos.

O fator EF-G apresenta algumas características funcionais relevantes para sua ligação ao ribossomo, esti-

mulando o movimento de translocação. 1) o EF-G/GTP apresenta pouca afinidade pelo ribossomo, quando o peptidil-tRNA está localizado no sítio P. A remoção do peptídeo do tRNA aumenta a afinidade de EF-G/GTP pelo ribossomo, estimulando a hidrólise de GTP; 2) o fator EF-G é incapaz de se ligar ao ribossomo na ausência de tRNA no sítio A. As diferenças de afinidade e de seletividade de ligação ao ribossomo asseguram que existirá

uma alternância na ligação dos fatores EF-Tu e EF-G, e não uma competição entre os fatores pelo sítio A do ribossomo. Portanto, para o alongamento da cadeia peptídica, podemos ressaltar as seguintes etapas (ver Figuras 12.11, 12.12 e 12.13):

1. quando o peptidil-tRNA está ocupando o sítio P, ocorre a apresentação do aa-tRNA ligado ao fator EF-Tu/GTP no sítio A do ribossomo;
2. a formação do pareamento códon-anticódon e a hidrólise de GTP com liberação de EF-Tu/GDP;
3. a formação da ligação peptídica pela atividade de peptidil-transferase;
4. o movimento parcial dos tRNAs no ribossomo (somente a região posicionada na subunidade maior), gerando os híbridos A/P e P/E;
5. finalmente ocorre o movimento de translocação total dos tRNAs, estimulado por EF-G/GTP, gerando ribossomos que contêm o peptidil-tRNA no sítio P e o aparecimento no sítio A do próximo códon.

O fator EF-G é uma proteína de aproximadamente 78 kDa contendo cinco domínios funcionais característicos. Os domínios III, IV e V mimetizam a estrutura do complexo terciário composto de EF-Tu/GTP ligado ao aminoácido. Este mimetismo estrutural (EF-G e EF-Tu-aminoácido) está relacionado ao fato de os dois fatores ocuparem o mesmo sítio no ribossomo (sítio A). No entanto, os dois fatores apresentam diferenças em especificidade de ligação ao sítio A do ribossomo. O fator EF-Tu-aminoácido apenas se liga no sítio A quando no sítio P existe um aminoacil ou peptidil-tRNA, ao passo que EF-G necessita da presença do peptidil-tRNA no sítio A (mas na forma híbrida A/P).

A precisão da síntese de proteínas depende de vários fatores, sendo o reconhecimento dos códons pelos diferentes aminoacil-tRNAs e os fatores de término (ver Seção 4.3) fundamentais para a síntese correta da cadeia peptídica. No entanto, outras regiões do ribossomo também participam deste processo. Durante o movimento de translocação, os tRNAs desacetilados estão localizados no sítio E dos ribossomos. Portanto, como a velocidade de dissociação do tRNA do sítio E é equivalente à velocidade de associação do tRNA ao sítio A, a presença de um tRNA no sítio E poderia influenciar a ligação do aminoacil-tRNA no sítio A do ribossomo.

4.3 Término da síntese de proteínas

O término da síntese de proteínas ocorre no momento em que um dos três códons de terminação é exposto no sítio A. O reconhecimento desses códons não é realizado por um tRNA, como nos processo de início e alongamento da cadeia polipeptídica, mas por uma molécula

de proteína (ver Figura 12.13). Em organismos procarióticos, foram identificados três fatores de término principais, sendo denominados RF1, RF2 e RF3. As proteínas RF1 e RF2 são responsáveis pelo reconhecimento dos três códons de terminação, apresentando especificidade sobreposta: RF1 reconhece os códons UAG e UAA, e RF2 reconhece UAA e UGA. Nos organismos eucarióticos, foi identificado somente um fator de terminação (eRF1) capaz de reconhecer os três códons. O fator de término RF3 (procariotos) ou eRF3 (eucariotos) parece apresentar diferentes funções entre esses organismos. Em procariotos, RF3 parece estar envolvido apenas com os eventos finais de tradução, sendo responsável pela liberação dos fatores RF1 ou RF2 do sítio do ribossomo. Em eucariotos, eRF3 forma um heterodímero com eRF1 para realizar a liberação da cadeia peptídica dos três códons de terminação.

Os fatores RF1, RF2 e eRF1 apresentam uma atividade equivalente aos fatores relacionados à etapa de alongamento da cadeia peptídica (EF-Tu ou eEF1- α), contendo um domínio de reconhecimento do códon (localizado na subunidade menor do ribossomo) e outro domínio responsável pela indução da reação de catálise no sítio de peptidil-transferase (localizado na subunidade maior do ribossomo). Análises da estrutura conformacional das proteínas de terminação demonstram que essas (RF) apresentam uma estrutura similar às moléculas de tRNA, indicando um segundo exemplo de mimetismo molecular. No entanto, apesar dos fatores RF ocuparem o mesmo sítio do ribossomo que os aminoacil-tRNAs, as regiões de reconhecimento e interação com o ribossomo são diferentes entre os fatores de término e os tRNAs ligados aos seus aminoácidos cognatos.

4.4 Saída da cadeia peptídica do ribossomo

A etapa final da síntese de proteínas está relacionada com a liberação da cadeia peptídica que poderá se enovelar adquirindo sua estrutura final funcional. A partir de 1980, foi visualizada uma estrutura presente na subunidade maior do ribossomo, denominada “túnel de saída do peptídeo”. Esse túnel apresenta um diâmetro de 12 a 20 angstroms, sendo localizado próximo ao centro da subunidade maior do ribossomo. Logo após a etapa da ligação peptídica, a cadeia polipeptídica nascente se encontra localizada no túnel de saída do ribossomo. Durante a etapa de alongamento do peptídeo, aproximadamente 30 a 50 aminoácidos ficam localizados dentro do túnel, sendo protegidos de degradação. Dentro do túnel, a cadeia peptídica ainda não está na sua estrutura final enovelada, pois devido ao seu diâmetro somente as estruturas em α -hélice das proteínas podem ser acomodadas neste túnel de saída.



Resumo

Neste capítulo, são apresentados os mecanismos gerais utilizados pelas células para sintetizar as suas proteínas. Na tabela do código genético, estão organizadas as trincas de nucleotídeos presentes nas moléculas de RNA mensageiro correspondentes aos diferentes aminoácidos. Entre as 64 possíveis combinações de três nucleotídeos, 61 delas correspondem aos diferentes aminoácidos, sendo as restantes responsáveis pela finalização da síntese da proteína. Os códons que representam o mesmo aminoácido podem variar na terceira posição, sendo possível uma variação total (qualquer das quatro bases corresponde ao mesmo aminoácido) ou parcial, ocorrendo a substituição dentro do grupo das purinas ou pirimidinas.

As moléculas de tRNAs apresentam uma estrutura conservada, sendo, no entanto, reconhecidas por aminoacil-tRNA-sintetases específicas que realizam a ligação ao aminoácido correspondente. O tRNA ligado aos seus respectivos aminoácidos (aa-tRNA) contém a sequência do anticódon que reconhece os códons no mRNA, realizando o pareamento entre as bases (códon-anticódon). Um tRNA iniciador contendo metionina (tRNAⁱ-Met) é utilizado no início da síntese de proteínas, pois apresenta a característica de ligação ao fator iniciador IF2 (procariotos) ou eIF2 (eucariotos) e também de se ligar ao sítio P parcial do ribossomo. Este tRNAⁱ-Met apresenta uma sequência de bases no anticódon complementar ao códon AUG presente no mRNA.

Os ribossomos são formados por moléculas de RNA (rRNA) e diversas proteínas, apresentando uma estrutura conservada entre os diferentes organismos. A estrutura funcional é formada pela associação das duas subunidades (menor e maior). Cada subunidade apresenta diferentes centros ativos, todos relacionados com o processo de síntese de proteínas. Os principais centros ativos são os sítios P (peptidil), A (aminoacil), E (saída do tRNA desacetilado), pepti-

dil-transferase e o sítio de ligação do mRNA. Os três primeiros centros ativos (P, A e E) são formados pela associação das duas subunidades, estando parcialmente representados em cada subunidade. Entretanto, o sítio de atividade de peptidil-transferase está localizado na subunidade maior, e o sítio de ligação do mRNA está presente na subunidade menor do ribossomo.

Durante a síntese de proteínas, os ribossomos carregam dois tRNAs ao mesmo tempo, sendo um posicionado no sítio P (ligado ao peptídeo nascente) e outro no sítio A (aa-tRNA). Portanto, o movimento dos tRNAs durante a síntese do peptídeo ocorre através da entrada no sítio A, onde, após a ligação peptídica, se movimenta para o sítio P e finalmente é eliminado do ribossomo através do sítio E. Além destas moléculas de RNA (mRNA, tRNA e rRNA-ribossomo), fatores proteicos associados são necessários durante todas as etapas da síntese de proteínas.

A síntese de proteínas pode ser dividida em três etapas principais: início, alongamento e término. Para o início da tradução, são necessários a subunidade menor do ribossomo, o tRNA iniciador e os fatores iniciadores IF e eIF. Durante a etapa de alongamento da cadeia peptídica, a subunidade maior deve se associar originando o ribossomo completo. Nesta etapa, ocorre também, a participação de fatores proteicos denominados EF ou eEF. Um dos fatores EF é responsável pela ligação aos aa-tRNAs e sua apresentação ao sítio A do ribossomo. Após a reação de peptidil-transferase, outro fator EF se liga ao sítio A induzindo o movimento total do ribossomo (translocação) e o posicionamento no sítio A do próximo códon do mRNA. O término da síntese proteica ocorre quando no sítio A é exposto a um dos códons de terminação. Esses códons são reconhecidos por fatores de terminação (RF) que são moléculas de proteína, e não de tRNA, indicando o final da síntese do peptídeo.



Leituras recomendadas

Berleant D, White M, Pierce E, Tudoreanu E, Boeszoemenyi A, Shtridelman Y, et al. The genetic code—more than just a table. *Cell Biochem Biophys*. 2009;55(2):107-16.


Kramer G, Boehringer D, Ban N, Bukau B. The ribosome as a platform for cotranslational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol*. 2009;16(6):589-97.

Ling J, Reynolds N, Ibba M. Aminoacyl-tRNA synthesis and translational quality control. *Annu Rev Microbiol*. 2009;63:61-78.

Miluzio A, Beugnet A, Volta V, Biffo S. Eukaryotic initiation factor 6 mediates a continuum between 60S ribosome biogenesis and translation. *EMBO Rep*. 2009;10(5):459-65.

Myasnikov AG, Simonetti A, Marzi S, Klaholz BP. Structure-function insights into prokaryotic and eukaryotic translation initiation. *Curr Opin Struct Biol*. 2009;19(3):300-9.

Nakamoto T. Evolution and the universality of the mechanism of initiation of protein synthesis. *Gene*. 2009;432(1-2):1-6.



Shoji S, Walker SE, Fredrick K. Ribosomal translocation: one step closer to the molecular mechanism. *ACS Chem Biol.* 2009;4(2):93-107.

Simonetti A, Marzi S, Jenner L, Myasnikov A, Romby P, Yusupova G, et al. A structural view of translation initiation in bacteria. *Cell Mol Life Sci.* 2009;66(3):423-36.

Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell.* 2009;136(4):731-45.

Valouev IA, Fominov GV, Sokolova EE, Smirnov VN, Ter-Avanesyan MD. Elongation factor eEF1B modulates functions of the release factors eRF1 and eRF3 and the efficiency of translation termination in yeast. *BMC Mol Biol.* 2009;10:60.

Sérgio Ceroni da Silva | Irene Silveira Schrank | Luiza Amaral de Castro

Controle da Expressão Gênica em Procarionotos

1. Controle no início da transcrição 278

- 1.1 Mecanismos de controle transcricional 278
 - 1.1.1 Reguladores gerais de transcrição 278
 - 1.1.2 Tipos de reguladores de transcrição 279
- 1.2 Controle transcricional exercido por cascata de fatores sigma 280
- 1.3 Óperon como unidade transcricional e de controle da expressão gênica 281
 - 1.3.1 Óperon *lac* de *E. coli* 281
 - 1.3.1.1 Controle negativo da transcrição no óperon *lac* 281
 - 1.3.1.2 Controle positivo da transcrição no óperon *lac* 285
 - 1.3.2 Controle negativo da transcrição do óperon *trp* de *E. coli* 285
 - 1.3.3 Óperon da arabinose de *E. coli* 285

2. Controle no término da transcrição 286

- 2.1 Controle da transcrição por atenuação do óperon *trp* de *E. coli* 287
- 2.2 Controle da transcrição por antiterminador 288

3. Integração de mecanismos regulatórios: controle da expressão gênica no bacteriófago lambda 289

- 3.1 Genoma do bacteriófago lambda 289
- 3.2 Mecanismos de controle da expressão gênica no bacteriófago lambda 290

4. Mecanismos de controle pós-transcricional 290

- 4.1 Proteínas como moléculas reguladoras 292

5. RNAs como moléculas reguladoras 293

- 5.1 Estrutura secundária de mRNA como mecanismo regulatório: ribocomutadores (*riboswitches*) 293
- 5.2 Pequenos RNAs (sRNAs) 294
 - 5.2.1 sRNAs e o sequestro de proteínas reguladoras 295
 - 5.2.2 sRNAs: controles positivo e negativo da tradução 296

6. Outros mecanismos de controle da expressão gênica 297

- 6.1 Variação de fase e controle da expressão gênica 297
- 6.2 Regulação gênica na defesa bacteriana de vírus ou plasmídeos (CRISPR) 298

Desde que o pediatra austro-alemão Theodor Escherich publicou, em 1885, o artigo descrevendo o isolamento e a identificação de uma bactéria do cólon de lactantes, por ele chamada de *bacterium coli commune*, a *Escherichia coli* tornou-se uma das formas de vida celular melhor estudada e conhecida. Além disso, *E. coli* também tornou-se um dos mais importantes organismos-modelo para diversas áreas do conhecimento, principalmente a biologia molecular.

Mais de um século depois, em julho de 1995, o primeiro genoma procariótico, da γ -proteobactéria *Haemophilus influenzae*, era completamente sequenciado. Finalmente, dois anos depois, em setembro de 1997, a linhagem “laboratorial” K-12 (sublinhagem MG1655) de *E. coli* teve seus 4.639.675 pb sequenciados. O genoma de *E. coli* K-12 MG16550 apresenta 4.493 genes, dos quais 4.149 codificam cadeias polipeptídicas, 172 correspondem a RNAs estruturais (incluindo os 87 sRNAs – *small RNAs*) e 174 foram identificados como pseudogenes. A expressão dos genes dependerá de condições como, por exemplo, disponibilidade de nutrientes, temperatura e tempo. Em condições específicas apenas uma fração dos mais de 4000 genes é expressa. Além disso, a célula é capaz de bloquear ou ativar a expressão de diferentes genes, como resultado de resposta aos estímulos internos e externos. Os mecanismos que regulam o processo de transcrição do DNA para sintetizar RNA, e, posteriormente, a sua tradução para gerar proteínas, são conhecidos como regulação da expressão gênica.

Uma visão geral sobre as questões referentes à expressão diferencial dos genes de procariotos pode ser apreciada se observado o comportamento de suas enzimas. Em bactérias, duas classes genéricas de enzimas, diferentes em sua concentração sob diversas condições metabólicas, podem ser categorizadas: as enzimas constitutivas e as induzíveis. As **enzimas constitutivas** são produzidas a velocidades e quantidades constantes, independentemente do estado metabólico do organismo. Os exemplos mais comuns de enzimas constitutivas são algumas daquelas envolvidas com a via glicolítica, rota essencial para o metabolismo celular. Já as **enzimas induzíveis** estão presentes em quantidades diminutas em um determinado momento, podendo aumentar sua concentração em até 1000 vezes quando seu substrato encontra-se no meio de cultivo, principalmente se esse substrato for uma fonte energética para a célula. O exemplo dessas enzimas é a β -galactosidase de *E. coli*, que catalisa a hidrólise da lactose para formar glicose e galactose, sendo sintetizada em resposta à presença de lactose no meio de cultivo.

Em geral, no processo de indução, é estimulada a síntese de várias enzimas relacionadas à mesma rota metabólica, em um processo de indução coordenada. No exemplo em que a lactose é adicionada a uma cultura de *E. coli*, três polipeptídeos são sintetizados em resposta a esse estímulo externo: a β -galactosidase, a permease e a transacetilase. Essas três enzimas estão envolvidas com o transporte e a utilização da lactose. Estas observações

indicam que mecanismos de controle da expressão gênica devem estar em ação no metabolismo da célula.

Além do mecanismo de indução, existe outro, de caráter oposto, chamado de repressão enzimática. Por exemplo, caso uma cultura de *E. coli* tenha somente amônia (NH_4^+) como fonte de nitrogênio, a bactéria é obrigada a fazer com que todas as rotas anabólicas (sintéticas) dos aminoácidos estejam ativas. A adição de um aminoácido específico faz com que as enzimas envolvidas em sua síntese deixem de ser sintetizadas. Esse processo de repressão está condicionado pelo produto final de uma rota metabólica. Existe, ainda, a repressão por **catabólito**, por exemplo, a repressão causada pela glicose sobre a expressão da β -galactosidase, mesmo que na presença de lactose. Neste caso, a *E. coli* segue a rota glicolítica para obtenção de energia a partir da glicose, em detrimento da rota de degradação da lactose.

O controle da expressão gênica pode ser exercido em diversas etapas, incluindo o início e o término da transcrição, o início e o término da tradução e a estabilidade e a acessibilidade do mRNA ou da proteína sintetizada. Os processos melhor estudados e compreendidos há mais tempo incluem aqueles envolvendo o controle do início da transcrição e da tradução dos mRNAs.

Por ser a RNA-polimerase a enzima-chave do processo de transcrição, é sobre ela que pode ser exercida a regulação primária da expressão gênica. Nos genes cuja expressão é constitutiva (os genes constitutivos), a interação promotor/RNA-polimerase é o fator determinante na transcrição. Nos outros genes, aqueles que codificam enzimas induzíveis, existem diferentes proteínas reguladoras que afetam a transcrição pela RNA-polimerase e que podem ter função ativadora, aumentando a quantidade de mRNA transcrito, ou repressora, bloqueando a transcrição.

1. Controle no início da transcrição

O processo de regulação gênica mais utilizado pelas células bacterianas ocorre no início da transcrição. Dessa forma, as proteínas necessárias em cada etapa da multiplicação celular são sintetizadas, evitando gastos energéticos. O controle da expressão gênica no início da transcrição pode ser realizado por diferentes mecanismos, por exemplo, aqueles que utilizam proteínas regulatórias, atuando diretamente na atividade da RNA-polimerase; ou aqueles que utilizam cascatas de fatores sigma alternativas, que irão reconhecer diferentes sequências promotoras.

1.1 Mecanismos de controle transcricional

1.1.1 Reguladores gerais de transcrição

Os mecanismos de regulação gênica, durante a transcrição, em organismos procarióticos, são bastante comple-

xos, apresentando-se em muitos casos interconectados. As principais moléculas responsáveis por esses mecanismos regulatórios são denominadas **fatores de transcrição**. Os fatores de transcrição são proteínas que ativam ou reprimem a expressão de um determinado gene por sua ligação com sequências específicas (regiões regulatórias) localizadas próximas à região promotora (ver Capítulo 4).

Os fatores de transcrição apresentam três principais mecanismos de atuação: (1) reconhecem e ligam-se à sequências específicas do DNA (ação em *cis*); (2) não apresentam capacidade de ligação ao DNA, mas realizam sua ação por meio da interação com outras proteínas; (3) atuam pela combinação dos dois mecanismos já citados, isto é, apresentam a capacidade de se ligar ao DNA e de interagir com outras proteínas. A maioria dos fatores de transcrição apresenta atividade repressora ou ativadora dos genes. No entanto, existem alguns fatores de transcrição que podem atuar tanto ativando como reprimindo a expressão gênica. Exemplos de atividade desses fatores de transcrição serão discutidos nas Seções 1.3.1, 1.3.2 e 1.3.3.

Os fatores de transcrição podem ser divididos em duas grandes categorias: os específicos para um determinado promotor e os reguladores gerais da transcrição. A maioria dos fatores de transcrição presentes em uma bactéria pertencem à primeira categoria, em que um fator de transcrição regula a expressão de um gene (ou óperon). No entanto, alguns fatores de transcrição (reguladores gerais da transcrição) são responsáveis pela regulação de vários genes ou óperons, gerando uma regulação em cascata. Na bactéria *E. coli*, por exemplo, existem em média sete reguladores gerais da transcrição responsáveis pelo controle da expressão de aproximadamente 50% dos seus genes.

Os fatores de transcrição classificados como reguladores gerais da transcrição podem atuar utilizando diferentes processos. Alguns reguladores gerais ajustam vários genes (ou óperons), pertencentes a diferentes classes funcionais, ou relacionados ao controle em cascata de diversos genes pela ação sequencial dos fatores de transcrição, sendo o fator iniciador um dos reguladores gerais da transcrição. Finalmente, existem os reguladores gerais que atuam em promotores regulados por diferentes cascatas de fatores sigma alternativas.

Um exemplo de regulador geral da transcrição, em *E. coli*, é a CAP de *catabolite activator protein*, – proteína ativadora de catabólito, discutida em detalhes na Seção 1.3.1.2), que regula mais de cem óperons diferentes, entre eles os responsáveis pela utilização de fontes alternativas de açúcares como lactose, arabinose, galactose e outros. CAP realiza a ativação destes diferentes óperons, quando a célula bacteriana está em deficiência de glicose, por meio de sua ligação a uma sequência específica localizada na região regulatória dos óperons.

Outro regulador geral, bastante caracterizado, é a proteína FNR (de *fumarate nitrate reduction* = redução

de fumarato e nitrato). FNR é um fator de transcrição que realiza tanto a repressão como a ativação de genes relacionados com a produção de energia na célula. Este regulador geral da transcrição reprime genes envolvidos na respiração aeróbica e ativa genes necessários para a fermentação e respiração anaeróbica.

Atualmente, com a disponibilidade de genomas de várias bactérias, diferentes metodologias foram desenvolvidas para o estudo dos fatores de transcrição. A combinação de análise *in silico* (bioinformática, ver Capítulo 17) e experimental possibilitará o mapeamento dos genes nos organismos, conduzindo para uma visão geral e integrada dos mecanismos regulatórios presentes em organismos procarióticos.

1.1.2 Tipos de reguladores de transcrição

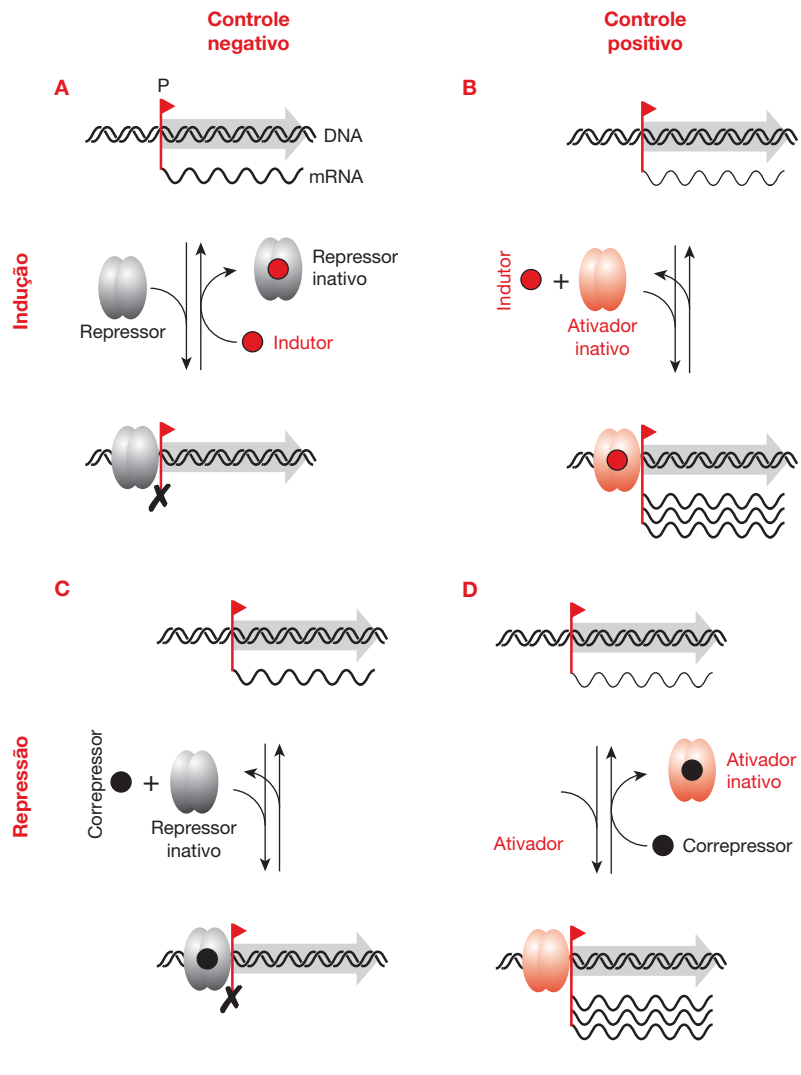
Os dois tipos principais de controle transcricional encontrados em procariotos são: controles negativo e positivo. Ambos os controles podem ainda ser subdivididos, com respeito à presença de um sinal (realizado por uma molécula sinalizadora), realizando indução ou repressão. Na **Figura 13.1** são apresentados esquemas que sumarizam estes quatro tipos de mecanismos.

O processo em que a regulação ocorre por meio de uma proteína repressora que se une ao DNA, é chamado de regulação negativa. Os genes sujeitos ao controle negativo são expressos, a menos que a transcrição seja impedida pela ligação da proteína repressora ao DNA. Os repressores unem-se a sítios específicos no DNA (sítios operadores), que estão localizados muito próximos ou, até mesmo, sobrepondo-se aos promotores. A ligação do repressor ao DNA pode impedir a ligação ou a movimentação da RNA-polimerase ao longo do DNA. A ligação do repressor ao sítio operador é regulada por um sinal (uma molécula pequena que se une ao repressor e causa uma mudança conformacional nele), alterando sua afinidade pelo operador. Essa mudança pode fazer com que um repressor unido ao DNA se desligue dele, permitindo desta forma a ação da RNA-polimerase e a transcrição do DNA (controle negativo com indução, Figura 13.1A). A molécula-sinal pode, também, unir-se a um repressor inativo, e esse complexo, então, irá ligar-se ao DNA e bloquear a ação da RNA-polimerase (controle negativo com repressão, Figura 13.1C). Assim, na regulação negativa, a transcrição pode ser favorecida ou bloqueada, dependendo da ação da molécula-sinal sobre a proteína repressora.

A regulação positiva é mediada por ativadores. No caso dos genes sujeitos ao controle positivo, a expressão depende da presença de proteínas reguladoras ativadoras, que são o contraposto molecular dos repressores. Os ativadores unem-se a sítios específicos que podem estar localizados adjacentes ao promotor, aumentando a afinidade da RNA-polimerase por ele e favorecendo, portanto, a transcrição. Sem a presença do ativador, a ligação da RNA-polimerase ao promotor pode ser quase nula. No controle positivo com indução, o ativador não se une ao DNA enquanto a molécula-sinal não formar um complexo com o ativador (Figura 13.1B). No controle po-

Figura 13.1

Mecanismos de controle transcricional da expressão gênica. Nos esquemas de controle positivo, a linha mais fina usada para representar o mRNA indica que sua transcrição ocorre em um nível basal baixo. O X abaixo do promotor indica bloqueio de transcrição. Níveis altos de transcrição são indicados por várias cópias de mRNA. P = promotor.



sitivo com repressão, a proteína ativadora *per se* liga-se ao DNA e ativa a transcrição, somente dissociando-se do DNA no momento da ligação de uma molécula-sinal (Figura 13.1D).

Portanto, a diferenciação dos sistemas em induzíveis e reprimíveis é feita com base na resposta apresentada a um sinal externo, normalmente uma pequena molécula envolvida na rota metabólica, cujas enzimas/proteínas são codificadas pelos genes que estão sendo regulados. Assim, nos sistemas induzíveis, a expressão dos genes ocorre apenas na presença de um indutor, ao passo que nos sistemas reprimíveis, a expressão ocorre na ausência do correpessor.

1.2 Controle transcricional exercido por cascata de fatores sigma

Conforme discutido no Capítulo 10, a especificidade da ligação da RNA-polimerase nos diferentes promotores é

determinada pela interação entre as várias espécies de fatores sigma (σ) existentes na célula. Portanto, um nível de controle da expressão gênica pode ser exercido pela disponibilidade destes diferentes fatores σ , produzidos em tempos específicos. Assim, um gene cujo promotor dependa de um determinado fator σ para ser reconhecido pela RNA-polimerase, apenas seria transcrito quando esse fator σ estivesse disponível.

O mecanismo de regulação gênica por alternância de cascata de fatores sigmas é utilizado para controlar a expressão de genes em diversas bactérias. Na bactéria *E. coli*, a expressão de diversos genes é regulada pelas cascatas de fatores sigmas alternativas como: genes que codificam proteínas relacionadas à utilização de fontes alternativas de nitrogênio; genes que codificam proteínas responsáveis pela síntese do flagelo. Em *Bacillus subtilis* uma cascata de fatores σ funciona como o mecanismo central de regulação da expressão gênica

no controle da esporulação. Neste processo, os genes são expressos em uma sequência coordenada para formar o esporo, forma resistente da bactéria, em resposta às condições adversas do meio ambiente, tais como pouca disponibilidade de nitrogênio ou fosfato.

Outro exemplo deste tipo de regulação da expressão gênica em procarionotos é o que ocorre no bacteriófago SPO1, que infecta e se multiplica em *B. subtilis*. Os genes desse fago podem ser agrupados em três categorias diferentes com relação a sua expressão temporal: os genes de expressão imediata, os genes de expressão intermediária e os genes de expressão tardia.

Logo após a infecção da bactéria, os genes de expressão imediata são os primeiros genes do fago a serem transcritos, processo que depende do fator σ^A de *B. subtilis* (equivalente ao fator σ^{70} de *E. coli*). É a célula hospedeira, portanto, que fornece o fator σ essencial para iniciar a série de eventos que leva à expressão dos genes do fago (Figura 13.2A). Dentre os genes de expressão imediata está o gene 28 de SPO1, cujo produto corresponde ao fator σ^{sp28} (Figura 13.2B) que, quando em quantidade suficiente, permite a transcrição dos genes 33 e 34 do fago. Os produtos destes genes de expressão intermediária correspondem aos fatores σ^{sp33} e σ^{sp34} , respectivamente (Figura 13.2C). Quando associados, esses fatores formam o fator $\sigma^{sp33/34}$, responsável pela identificação dos promotores dos genes de expressão tardia, cuja transcrição, então, pode ocorrer (Figura 13.2D). Desta maneira, a expressão temporalmente regulada dos diferentes genes do bacteriófago SPO1 depende de uma cascata de fatores σ , também temporalmente regulada.

1.3 Óperon como unidade transcricional e de controle da expressão gênica

Os estudos pioneiros da regulação gênica em *E. coli* foram feitos por François Jacob e Jacques Monod, na década de 1950. Esses pesquisadores e seus colaboradores analisaram a expressão de enzimas envolvidas no metabolismo da lactose, que pode ser usada como fonte de carbono e energia por meio de sua clivagem em glicose e galactose.

A β -galactosidase, enzima que catalisa a clivagem da lactose, assim como outras enzimas envolvidas no metabolismo da lactose, é expressa somente quando esse glicídeo está disponível para a bactéria. Caso contrário, não é gasta energia na síntese de RNAs e proteínas desnecessárias, por isso, a lactose induz a síntese das enzimas envolvidas no seu próprio metabolismo. Além de requerer a β -galactosidase, o metabolismo da lactose envolve os produtos de dois outros genes relacionados: o da lactose-permease, que transporta a lactose para dentro da célula; e o da transacetilase, cuja função no metabolismo da lactose ainda não é completamente conhecida. Com base em experimentos genéticos, Jacob e Monod deduziram o mecanismo pelo qual a expressão destes genes é regulada, formulando o modelo do óperon (ver Capítulo 4), que permanece fundamental para o entendimento da regulação da transcrição.

O óperon é uma unidade de expressão gênica que inclui tanto os genes estruturais (codificando as proteínas/enzimas que atuam no metabolismo celular) como os elementos reguladores, que controlam a sua expressão. Estruturalmente, pelo menos três elementos distintos estão presentes em um óperon: o promotor, sítio no qual ocorre a ligação da RNA-polimerase; as regiões controladoras, nas quais as proteínas reguladoras ligam-se; e os genes estruturais, codificando as diferentes cadeias polipeptídicas.

No momento da transcrição de um óperon, a partir de seu promotor, um único mRNA é sintetizado, contendo as regiões codificadoras (**cístrons**) das diversas cadeias polipeptídicas. Esse mRNA, chamado de **policístrônico**, tem cada um dos cístrons traduzido de forma independente, possuindo seus próprios sítios de ligação ao ribossomo (RBS) e códons de iniciação e de terminação (ver Capítulos 10 e 12).

1.3.1 Óperon *lac* de *E. coli*

1.3.1.1 Controle negativo da transcrição no óperon *lac*

O modelo de regulação gênica que ocorre no óperon *lac*, desenvolvido com base nos experimentos de Jacob e Monod, está ilustrado na Figura 13.3. Os genes codificando a β -galactosidase, a permease e a transacetilase são expressos como uma única unidade. A transcrição do óperon é controlada por *o* (o operador), que está adjacente ao sítio de início da transcrição. O gene *i* codifica uma proteína que regula a transcrição por ligação ao operador.

Uma vez que mutantes *i*⁻ (resultantes em expressão constitutiva) são recessivos, foi concluída sua deficiência na produção de um produto gênico funcional. Estes resultados significam que o produto normal do gene *i* é um repressor, que bloqueia a transcrição, quando ligado no sítio *o* (Figura 13.3A). A adição de lactose leva à indução do óperon pela ligação lactose-repressor, o que impede sua ligação ao sítio operador no DNA (Figura 13.3B). Em mutantes no gene *i*, não sujeitos à indução (que são dominantes sobre *i*⁻), o repressor mutante não se liga à lactose, assim, o óperon não pode ser induzido.

A confirmação deste modelo básico foi realizada por uma variedade de experimentos, incluindo o isolamento do repressor *lac* e a análise de sua ligação no sítio operador do DNA, por Walter Gilbert, na década de 1960. As análises moleculares definiram o operador como sendo correspondente a aproximadamente 30 pb de DNA, iniciando a poucas bases do sítio de iniciação da transcrição. Análises por *footprinting* identificaram essa região como o local de ligação do repressor, com consequente bloqueio da transcrição. Essa sequência, denominada O1, é uma das sequências reconhecida pela proteína repressora, codificada pelo gene *lacI*. Duas outras sequências de ligação ao repressor, chamadas de O2 e O3, estão presentes nas adjacências. Quando o repressor, um tetrâmero, liga-se ao DNA (em O1/O2 ou O1/O3), faz com que o DNA assumira uma estru-

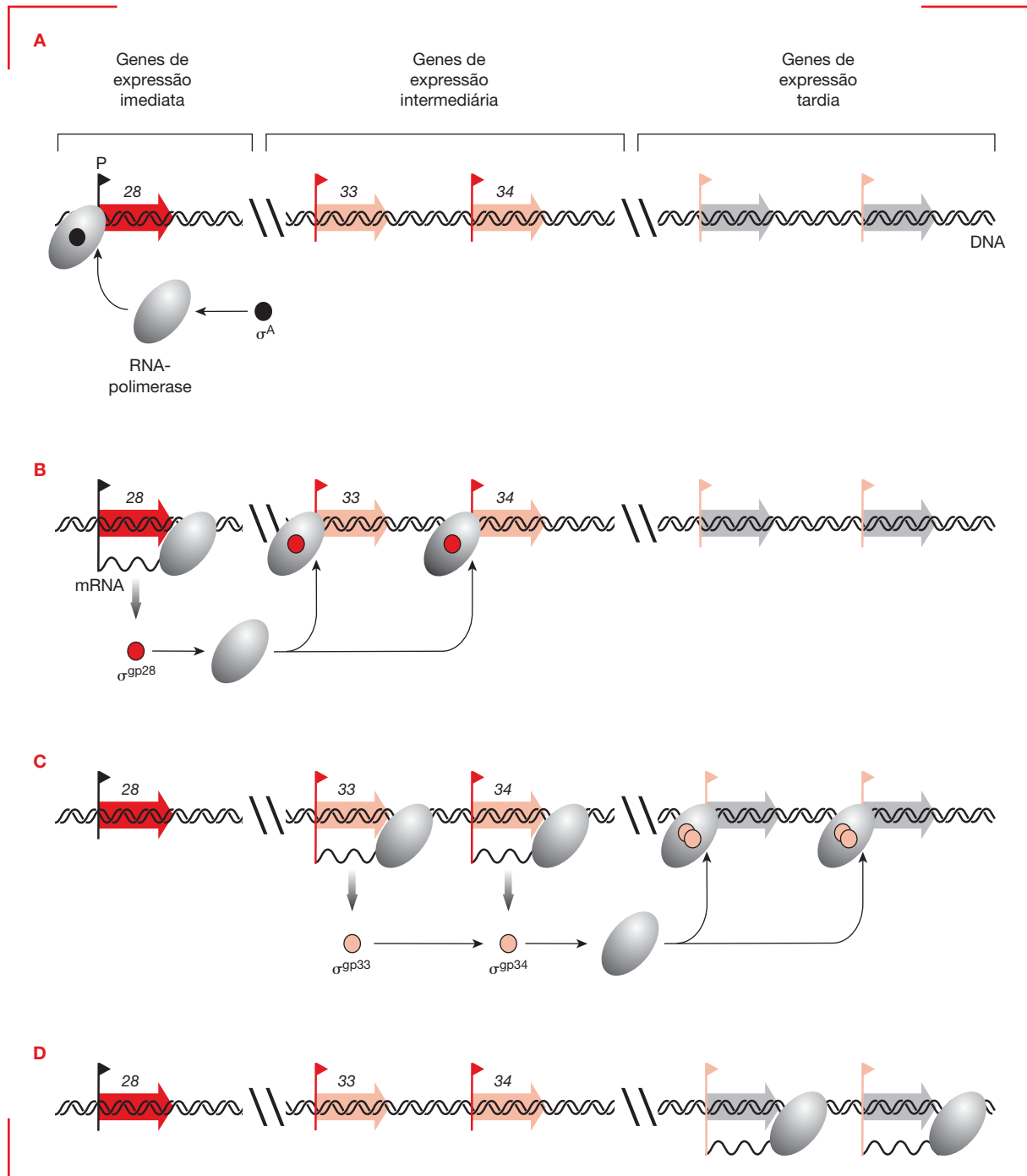


Figura 13.2

Controle da transcrição por fatores σ alternativos no bacteriófago SPO1 de *B. subtilis*. (A) A transcrição do gene 28 do bacteriófago SPO1 é iniciada pela RNA-polimerase, utilizando o fator σ^A de *B. subtilis*. Esse é o primeiro gene do fago a ser expresso. (B) Após a transcrição e a tradução, o produto deste gene (σ^{gp28}) atua como fator σ , permitindo o reconhecimento dos promotores dos genes 33 e 34. (C) As proteínas σ^{gp33} e σ^{gp34} atuam, em conjunto com o fator $\sigma^{gp33/34}$, que permite a identificação dos promotores dos genes de expressão tardia. (D) Transcrição dos genes de expressão tardia de SPO1. Observar que a transcrição do gene 28 (em C e D) e dos genes 33 e 34 (em D) ainda está ocorrendo, mas não é representada na figura. P = promotor.

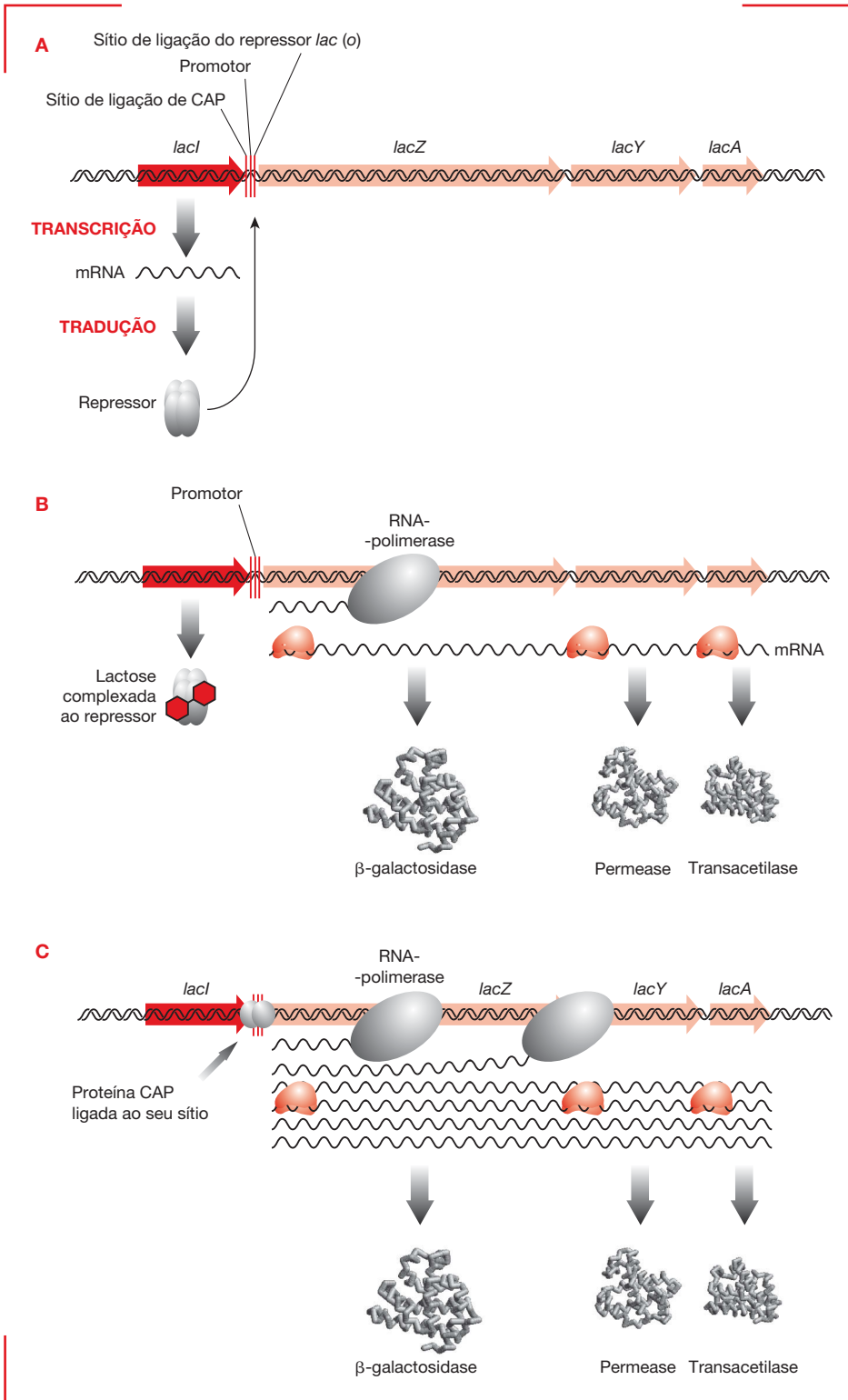


Figura 13.3

Controle do óperon *lac*. (A e B) Controle negativo: O gene *lacI* codifica um repressor que, na ausência de lactose (em cima), liga-se ao operador (*o*) e bloqueia a transcrição dos três genes estruturais (*lacZ*, β-galactosidase; *lacY*, permease; e *lacA*, transacetilase). A lactose induz a expressão do óperon por ligação ao repressor (embaixo), impedindo, assim, que o repressor ligue-se ao operador. (C) Controle positivo: A proteína CAP interage com a subunidade α da RNA-polimerase para ativar a transcrição. Desta maneira, a quantidade de mRNA transcrito do óperon é aumentada na presença de CAP ligada ao seu sítio-específico.

tura em forma de alça (**Figura 13.4A**). No momento em que a lactose encontra-se complexada com o repressor, uma mudança conformacional é sofrida por ele, impedindo-o de se ligar ao DNA no sítio operador. Mutações no operador (σ^c de constitutivo) alteram as suas sequências, impedindo a ligação do repressor e resultando em expressão gênica constitutiva.

O princípio central de regulação gênica, exemplificado pelo óperon da lactose, é de que o controle da trans-

crição é mediado pela interação de proteínas reguladoras com sequências específicas do DNA. Esse modelo geral de regulação é aplicável tanto a células procarióticas como a células eucarióticas. Sequências reguladoras, tais como os operadores, são chamadas de elementos de controle com atuação em *cis*, pois afetam apenas a expressão de genes fisicamente ligados na mesma molécula de DNA. Por outro lado, proteínas como os repressores são chamadas de fatores com atuação em *trans*, pois afetam a

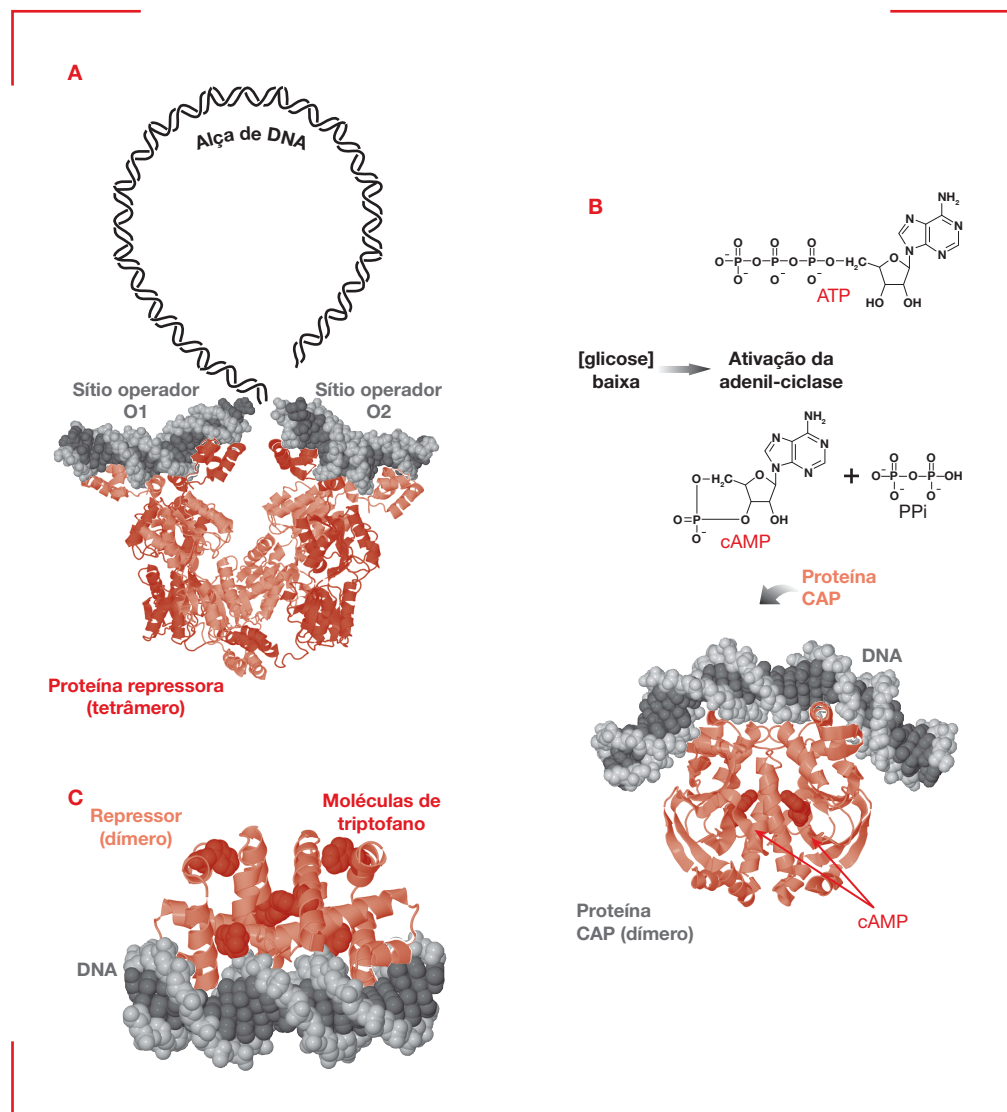


Figura 13.4

Estrutura de proteínas reguladoras e sua interação com o DNA. (A) Repressor LacI: O homotetrâmero formado pelo repressor *lac* está ligado a dois segmentos de DNA, correspondentes aos operadores O1 e O2. (B) Ativador CAP: Níveis baixos de glicose ativam a adenil-ciclase, que converte ATP em AMP cíclico (cAMP). O cAMP liga-se a CAP e estimula sua ligação em sequências reguladoras de vários óperons relacionados com o metabolismo de açúcares de utilização alternativa, como a lactose. (C) Repressor TrpR: A proteína forma um complexo com o triptofano e liga-se ao sítio operador no DNA. As moléculas de triptofano (correpressor) formam um complexo com o triptofano. Esse complexo (dímero) está ligado ao DNA. As proteínas estão representadas em modelo de *cartoon* e o DNA aparece em modelo de preenchimento.

expressão de genes localizados em outras moléculas de DNA da célula, por exemplo, em outro cromossomo ou plasmídeo. O óperon *lac* é um exemplo de regulação negativa, pois a ligação do repressor bloqueia a transcrição. Contudo, essa afirmação não vale para todos os casos, já que vários fatores com atuação em *trans* são ativadores da transcrição.

1.3.1.2 Controle positivo da transcrição no óperon *lac*

Um dos exemplos mais estudados de controle positivo em *E. coli* é o efeito da glicose na expressão dos genes que codificam as enzimas envolvidas na quebra (catabolismo) de outros açúcares (incluindo a lactose), os quais constituem-se em fontes alternativas de carbono e energia. Por exemplo, caso *E. coli* seja multiplicada em um meio contendo tanto lactose como glicose, o óperon *lac* não é induzido, e a glicose é preferencialmente usada pela bactéria. Portanto, a presença da glicose impede que o óperon *lac* seja transcrito em níveis altos, mesmo na presença do indutor normal (lactose).

O efeito da glicose (conhecido como repressão por catabólito) é mediado por um sistema positivo de controle acoplado aos níveis de AMP cíclico (cAMP). Em bactérias, a enzima adenilciclase, que converte o ATP em cAMP, é regulada de forma que os níveis de cAMP aumentam quando os níveis de glicose diminuem. O cAMP liga-se a uma proteína reguladora transcricional chamada de proteína ativadora de catabólito (CAP) (Figura 13.4B) e essa ligação estimula outra, de CAP à sua sequência-alvo no DNA, que no óperon *lac* está situada a aproximadamente 60 pb antes do sítio de início da transcrição. Então, a CAP interage com a subunidade α da RNA-polimerase, facilitando a ligação da enzima ao promotor e ativando a transcrição (Figura 13.3C).

Portanto, a indução do óperon *lac* requer tanto a presença de lactose (o indutor) para inativar o repressor, como a ausência de glicose para aumentar a concentração de cAMP e permitir a ligação da proteína CAP ao DNA. Sendo assim, CAP é um elemento regulador positivo, que responde aos níveis de glicose, ao passo que o repressor *lac* é um elemento regulador negativo, que responde aos níveis de lactose. A ativação por CAP é ineficaz neste sistema, se o repressor *lac* estiver bloqueando a transcrição; contudo, mesmo que o repressor *lac* esteja dissociado do sítio operador, a transcrição não atinge os níveis máximos, a menos que CAP esteja ligada ao DNA.

1.3.2 Controle negativo da transcrição do óperon *trp* de *E. coli*

O óperon *trp* codifica cinco enzimas essenciais para a biossíntese do aminoácido triptofano (Figura 13.5). Em condições normais, ou seja, no trato gastrointestinal das diversas espécies animais, a bactéria *E. coli* não precisa realizar a síntese de seus próprios aminoácidos, obtendo-os diretamente do meio em que se encontra. Assim, os genes estruturais codificados pelo óperon *trp* somente

são expressos quando o triptofano não está disponível à célula em seu meio ambiente, uma vez que a síntese do triptofano adicional seria desnecessária. O triptofano funciona como um regulador negativo de sua própria biossíntese (regulação que pode ser realizada de duas maneiras: por repressão e por atenuação).

A expressão do óperon do triptofano é inibida pela proteína repressora TrpR. O gene que codifica esse repressor (*trpR*) está localizado bastante distante do óperon. O repressor *per se*, em geral, não se liga à região operadora, apenas quando em associação com o correpressor, que é o próprio triptofano (Figura 13.4C). Assim, o triptofano liga-se ao repressor e o complexo triptofano/repressor liga-se à região operadora. Uma vez que essa região se sobrepõe à região promotora, há um bloqueio da transcrição (Figura 13.5B).

O segundo mecanismo de controle da transcrição denominado atenuação será discutido na Seção 2.1.

1.3.3 Óperon da arabinose de *E. coli*

O óperon da arabinose de *E. coli* inclui os genes que codificam as enzimas envolvidas no metabolismo da arabinose: a ribulose-quinase, a arabinose-isomerase e a ribulose-5-fosfato-isomerase, codificadas pelos genes *araB*, *araA* e *araD*, chamados de genes *araBAD* (Figura 13.6).

Além destes genes, o óperon *ara* também inclui dois sítios operadores (*araO₁* e *araO₂*) e o sítio *araI*, que corresponde ao local de ligação para a proteína reguladora AraC. Essa proteína é codificada pelo gene *araC*, localizado a montante dos genes *araBAD*. A transcrição dos genes *araBAD* é feita a partir de um promotor que se sobrepõe ao sítio *araI*, ao passo que o gene *araC* é transcrito a partir de um promotor com sobreposição ao sítio *araO₁*. Entre os sítios *araO₁* e *araI* é encontrado um sítio de ligação para CAP.

A proteína AraC exerce uma complexa função reguladora no sistema, regula sua própria síntese, pela ligação ao operador *araO₁* e pela repressão de sua própria transcrição, quando mais de 40 cópias da proteína acumulam-se na célula. Essa proteína também atua como regulador positivo e negativo dos genes *araBAD*. Se tanto a glicose como a arabinose estão presentes no meio, a proteína AraC liga-se ao operador *araO₂*. Essa ligação faz com que a proteína AraC ligue-se também ao sítio *araI*, formando uma alça de DNA com cerca de 210 pb, configuração que reprime a transcrição dos genes *araBAD*. Quando o meio contém arabinose, mas sem a presença de glicose, aumentam os níveis de cAMP e, portanto, da proteína CAP-ativada (CAP-cAMP), que se liga ao seu local específico, adjacente ao sítio *araI*. A arabinose liga-se à proteína AraC, alterando sua conformação e causando a abertura da alça de DNA. A proteína AraC, com a conformação modificada, liga-se aos sítios *araI* e *araO₁*, atuando como um ativador, juntamente com CAP-cAMP, fazendo com que os genes *araBAD* sejam transcritos.

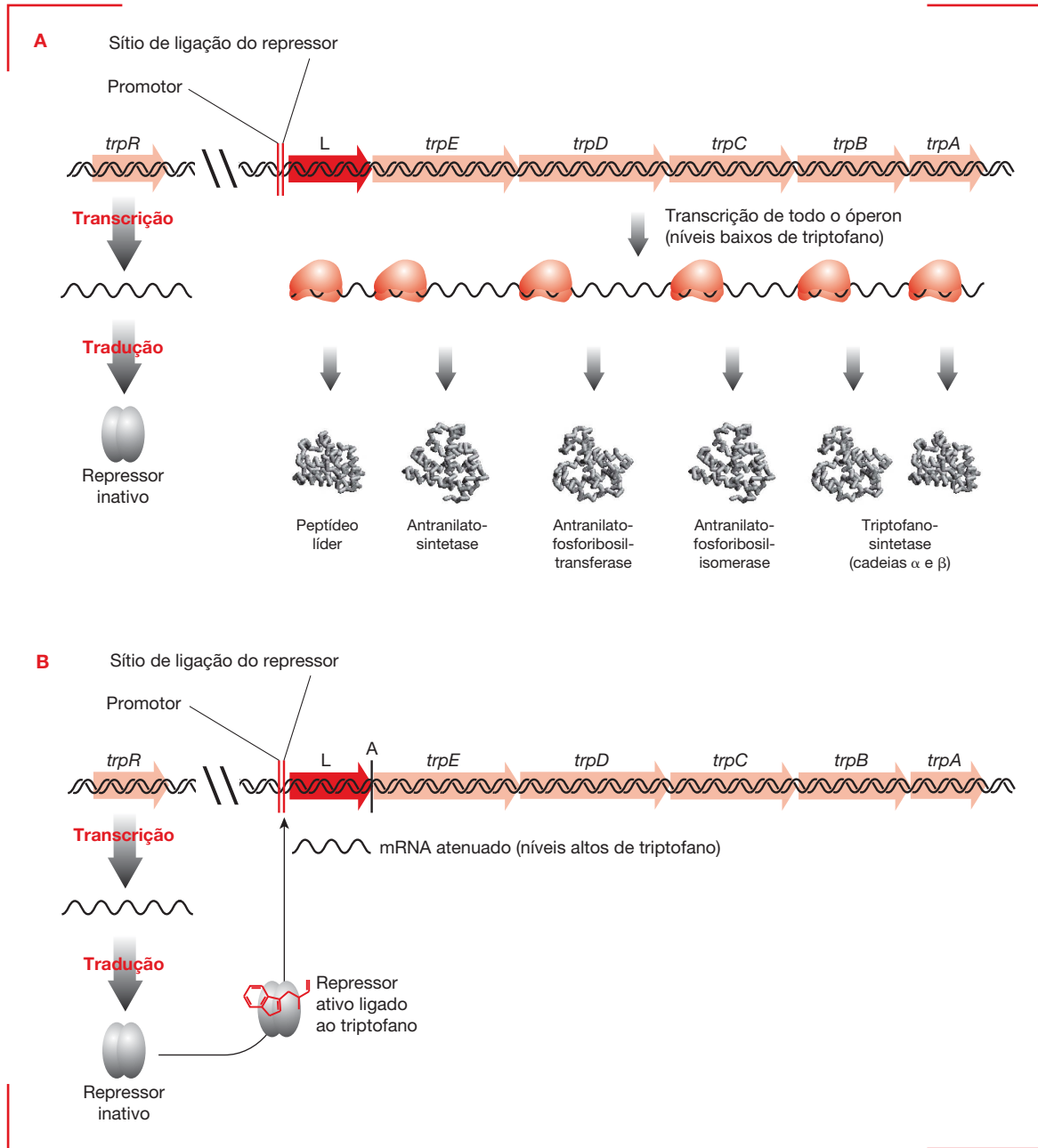


Figura 13.5

O óperon do triptofano. (A) O óperon contém cinco genes estruturais codificando as enzimas envolvidas na biossíntese do triptofano: *trpE*, *trpD*, *trpC*, *trpB* e *trpA*. Além destes genes estruturais, há uma sequência-líder (L), que codifica um pequeno peptídeo que precede os genes estruturais. **(B)** A expressão dos genes é controlada em dois níveis. O gene *trpR* codifica um repressor que, na presença do triptofano, liga-se ao operador (o) e bloqueia a transcrição. Além disso, a expressão é mediada por uma sequência atenuadora (A).

2. Controle no término da transcrição

O controle do término da transcrição pode ser utilizado como um mecanismo para regulação da expressão de

genes em bactérias. Nesta seção serão discutidos dois mecanismos, os controles da transcrição por atenuação e por antiterminador, para exemplificar este processo de regulação gênica.

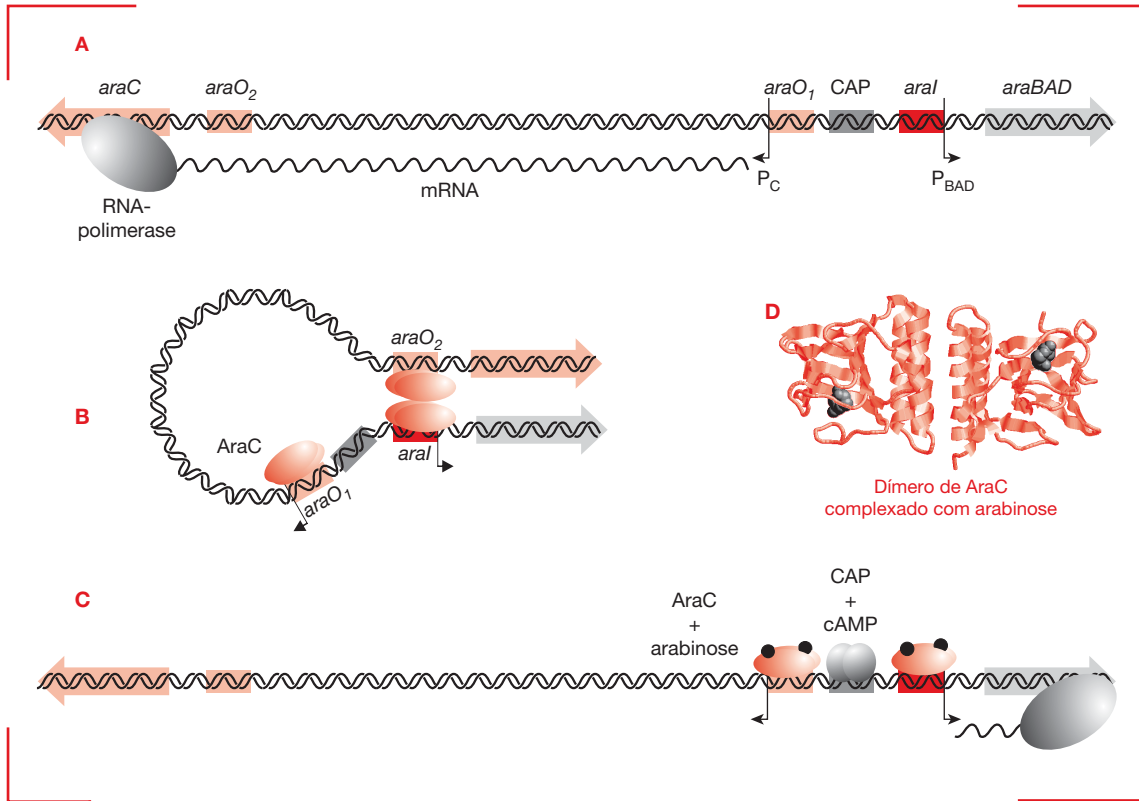


Figura 13.6

Controle do operon da arabinose. (A) Na ausência da proteína AraC, o gene *araC* é transcrito a partir do promotor P_C . (B) Em condições de baixa concentração de arabinose, com níveis altos de glicose, a proteína AraC (homodímero) liga-se a *araI* e *araO₂*, complexo esse que provoca a formação de uma alça na estrutura do DNA, impedindo a transcrição de *araBAD*. AraC também liga-se a *araO₁*, impedindo sua própria transcrição. (C) Quando a arabinose está disponível e a concentração de glicose está baixa, a proteína AraC liga-se a arabinose, passando a ser um ativador da transcrição que atua junto com CAP-cAMP na transcrição dos genes *araBAD*, a partir do promotor P_{BAD} . (D) Dímero da proteína AraC (representada em modelo de cartoon) formando um complexo com a arabinose (representada em modelo de preenchimento).

2.1 Controle da transcrição por atenuação do operon *trp* de *E. coli*

O mecanismo, denominado atenuação da transcrição, regula a expressão de alguns genes pelo controle da capacidade da RNA-polimerase em continuar o alongamento além de determinados sítios. Esse modo de regulação, atuando em nível de término da transcrição, foi primeiramente descrito no operon *trp* de *E. coli*.

A transcrição do operon *trp* pode não prosseguir até o seu sítio normal de término, após o gene *trpA*, caso haja disponibilidade de triptofano, sendo interrompida antes do gene *trpE* (Figura 13.5B). Como discutido anteriormente, o operon *trp* é regulado, em parte, por um repressor que, ligado ao triptofano, bloqueia a transcrição (Figura 13.5B). Contudo, a atenuação transcricional permite um nível adicional de controle, resultando em uma regulação mais estrita do que aquela possível utilizando apenas a repressão da iniciação. O sítio da atenuação está

localizado a 162 nt após o sítio de iniciação da transcrição (Figura 13.7).

A região crítica do mRNA do *trp* contém dois códons adjacentes para triptofano, de maneira que a velocidade de tradução é altamente dependente dos níveis de triptofano. Isto constitui a ligação entre a atenuação transcricional e a disponibilidade de triptofano. Caso os níveis celulares de triptofano sejam baixos, os ribossomos fazem uma pausa devido a pouca disponibilidade dos tRNA^{Trp}, impedindo a formação do grampo de terminação e permitindo que a transcrição do mRNA continue. Caso o triptofano seja abundante, a tradução nesta região segue em velocidade normal, há a formação de um grampo de terminação e a transcrição é terminada.

Inicialmente descoberto na década de 1970, por Charles Yanofsky, o controle por atenuação foi descrito como sendo uma estratégia de regulação nova e atípica, comparada com os tradicionais mecanismos de ativação e repressão da transcrição. Contudo, à medida que diver-

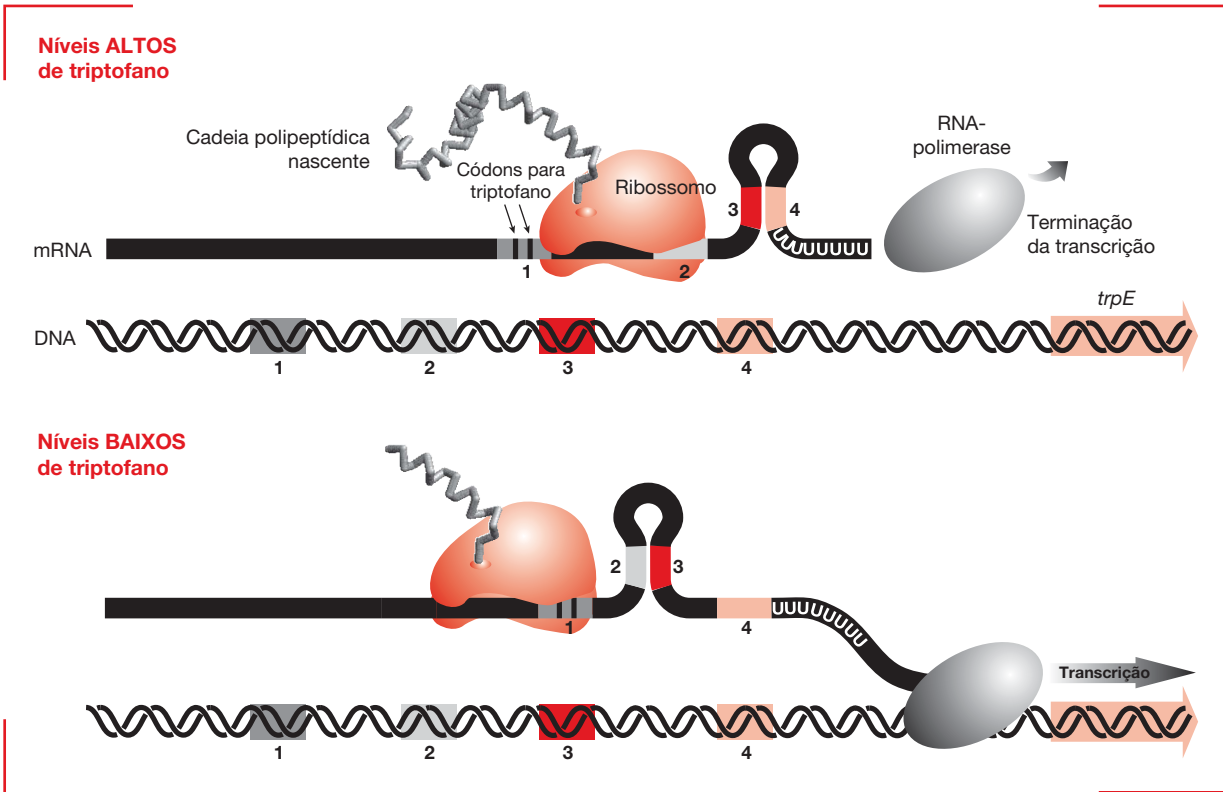


Figura 13.7

Atenuação da transcrição no operon do triptofano. Na região correspondente ao atenuador há quatro sítios que apresentam complementaridade de bases. Assim, os sítios 1, 2, 3 e 4, são capazes de formar estruturas em grampo. Os sítios 2 e 3 também apresentam complementaridade de bases. Como o processo de tradução ocorre simultaneamente com a síntese do mRNA em procariotos, os ribossomos prosseguem ao longo dele, ocupando uma região anterior ao local de transcrição. Em condições com disponibilidade de triptofano, o ribossomo prossegue à velocidade regular, impedindo a formação de grampo entre os sítios 2 e 3. Com isso, é formado um grampo entre os sítios 3 e 4, correspondente a um grampo de terminação. Em condições que a presença de triptofano é limitada, e, portanto, há baixa disponibilidade do tRNA correspondente carregado com esse aminoácido (tRNA^{Trp}), o ribossomo faz uma pausa sobre a região 1. Essa pausa ocorre em decorrência da presença de dois códons adjacentes para triptofano, permitindo o emparelhamento de bases entre as regiões 2 e 3. Assim, quando a região 4 é transcrita, ela não tem a região 3 disponível para hibridizar e formar o grampo de terminação. Com isso, a transcrição pode prosseguir normalmente.

Outros óperons foram sendo estudados, ficou evidente que esse era um mecanismo utilizado amplamente para o controle da expressão gênica. Assim, a expressão gênica em óperons que codificam proteínas envolvidas na biossíntese de outros aminoácidos, é regulada por atenuação.

2.2 Controle da transcrição por antiterminador

O mecanismo de controle da expressão gênica no término da transcrição por um antiterminador foi inicialmente descoberto no bacteriófago lambda, sendo depois identificado como mecanismo regulatório utilizado em outros fagos e óperons bacterianos.

As proteínas antiterminadoras apresentam a característica de interagirem com a RNA-polimerase, impedindo o reconhecimento do término da transcrição (ver Capítulo 10). Este mecanismo é específico para cada gene ou óperon, pois depende da presença, no mRNA, de um sítio de reconhecimento e da ligação do antiterminador. Portanto, como consequência da atividade do antiterminador, ocorre a síntese de um mRNA contendo sua região 3' alterada (mais longa).

Um exemplo da atividade de uma proteína antiterminadora será apresentado na Seção 3.2, demonstrando sua interação com o mecanismo de regulação gênica no início da transcrição.

3. Integração de mecanismos regulatórios: controle da expressão gênica no bacteriófago lambda

O bacteriófago λ multiplica-se apenas no interior de *E. coli*, devendo necessariamente injetar o seu DNA na bactéria hospedeira para a sua multiplicação (ver Capítulo 9). Após a infecção, o genoma do fago pode seguir duas vias: o **ciclo lítico** ou o **ciclo lisogênico** (ver Capítulo 9). Em condições naturais, a opção entre seguir o ciclo lítico ou lisogênico depende das condições do meio. Em um meio rico em nutrientes, como, por exemplo, quando o bacteriófago λ infecta uma *E. coli* intestinal, o estado lítico é preferencial. Por outro lado, em um meio pobre de nutrientes, como é o caso da *E. coli* no solo, o fago prefere o estado lisogênico. Em condições experimentais, o estado a ser seguido depende do estado de equilíbrio entre os fatores do meio intra e extracelular, de fatores genéticos da bactéria hospedeira e do bacteriófago.

3.1 Genoma do bacteriófago lambda

O bacteriófago λ é uma partícula viral constituída por proteínas e DNA em iguais proporções. O DNA de λ , em sua forma isolada da partícula viral, é uma molécula linear, de fita dupla, com 48.502 pb (**Figura 13.8A**). As extremidades da molécula contêm regiões de fita simples, sobressalentes em 5', com 12 nucleotídeos e, que são complementares na sequência de bases (**sítios cos**). Por meio destas regiões complementares o DNA assume a forma circular, depois de injetado na célula hospedeira (**Figura 13.8B**). O genoma do bacteriófago λ codifica aproximadamente 50 proteínas, cujos genes têm um cronograma de expressão bem definido, o que determina a escolha entre o ciclo lítico ou lisogênico. Os genes que codificam as proteínas necessárias para a integração do bacteriófago λ no cromossomo da bactéria estão localizados à esquerda do gene *cIII*, ao passo que os codificadores de proteínas envolvidas no ciclo lítico estão localizados à direita do gene *cII* (**Figura 13.8A**).

Os genes do bacteriófago λ , estão arranjados em três grupos: os iniciais imediatos, os iniciais e os tardios. Os

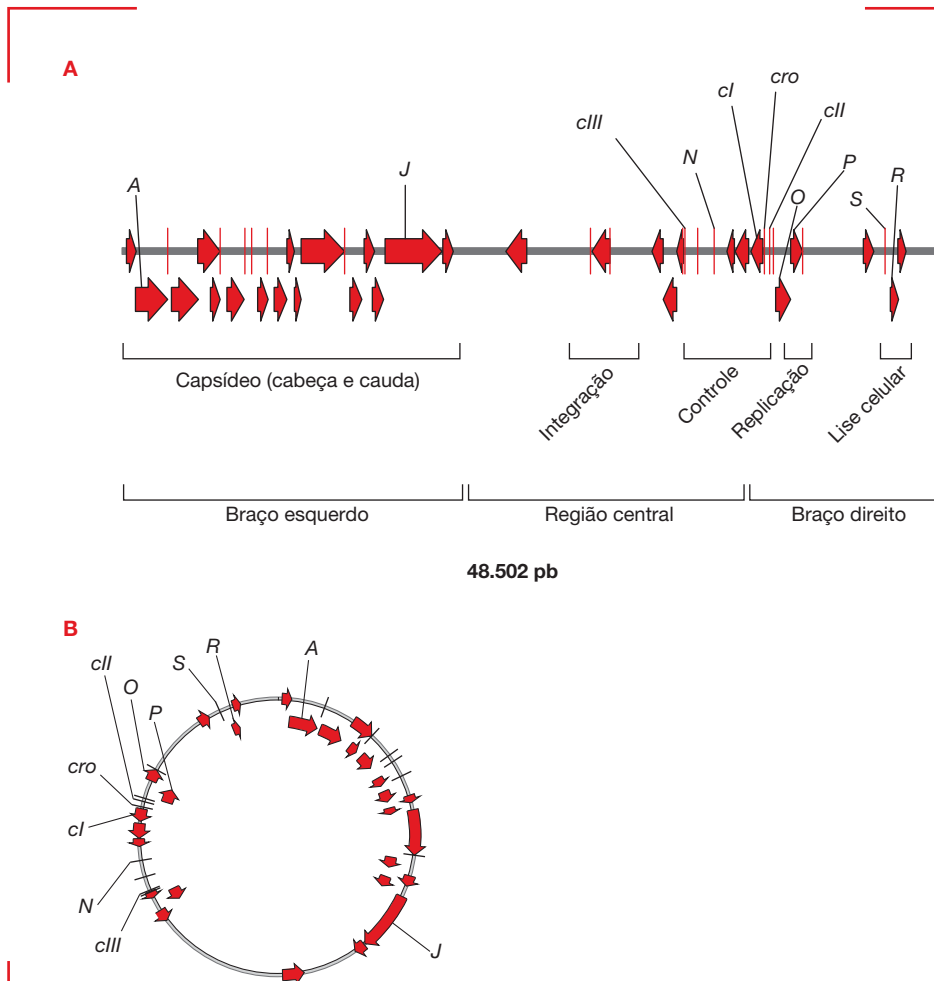


Figura 13.8

Genoma do bacteriófago λ . (A) Mapa linear do genoma, com a localização dos principais genes de λ . Logo abaixo do mapa estão indicadas as funções destes diferentes grupos de genes. (B) Mapa circular do genoma de λ , após ligação dos sítios *cos*. O tamanho das setas é proporcional ao tamanho dos genes.

genes iniciais imediatos codificam proteínas reguladoras, que controlam a transcrição dos genes iniciais. Os produtos dos genes iniciais, por sua vez, regulam a transcrição dos genes tardios, que codificam proteínas necessárias para a montagem do capsídeo viral e para a formação da progênie infectiva de fagos.

3.2 Mecanismos de controle da expressão gênica no bacteriófago lambda

O bacteriófago λ apresenta diversos mecanismos de regulação gênica para realizar seu ciclo de infecção na bactéria *E. coli*. Nesta seção serão discutidos apenas os mecanismos relacionados ao controle do início e do término da transcrição.

Logo após a infecção de *E. coli* por λ , é feita a transcrição dos genes do bacteriófago a partir dos promotores P_L e P_R . A partir da transcrição de P_L é produzida a proteína N, e P_R controla a transcrição do gene *cro* (Figura 13.9A). Esse estágio corresponde ao inicial imediato.

A proteína Cro liga-se ao operador O_{R3} , que se apresenta sobreposto ao promotor P_{RM} , impedindo, desta forma, a transcrição do gene *cI* (responsável pela manutenção da via lisogênica). Em um estágio posterior, chamado de inicial, a proteína N sintetizada, com ação anti-terminadora nos sítios T_L e T_R , permite que a transcrição prossiga para além destes sítios, incluindo nos mRNAs formados os genes *cIII* e *cII* (Figura 13.9B).

Durante o transcurso do ciclo lítico, o produto do gene *cro* está diretamente relacionado com a replicação do genoma do bacteriófago λ , por meio da indução da expressão dos genes *O* e *P*. Já o produto do gene *N* está relacionado com a expressão dos genes da região responsável pelo empacotamento do DNA. Juntos, os genes *A* e *J*, são responsáveis pela síntese das proteínas da cabeça e da cauda do bacteriófago λ . O produto de *N* também controla a expressão dos genes *S* e *R*, envolvidos com a lise da célula hospedeira.

Uma característica importante da proteína Cro é sua capacidade de ligação a O_{R1} , O_{R2} e O_{R3} . Contudo, Cro apresenta uma maior afinidade por O_{R3} , por isso, quando sua concentração é baixa, Cro liga-se apenas a O_{R3} . À medida que a transcrição do gene *cro* prossegue e sua concentração aumenta, a proteína Cro liga-se aos sítios O_{R2} e O_{R1} , que se sobrepõem ao promotor P_R , bloqueando, desta forma, a transcrição do gene *cro* a partir de P_R . Portanto, a proteína Cro autorregula sua síntese.

Durante o ciclo lisogênico, a transcrição também é iniciada nos promotores P_L e P_R . Como no caso do ciclo lítico, a ação antiterminadora de N também é importante para o ciclo lisogênico (Figura 13.9B). O papel do produto do gene *cII* é fundamental neste processo, pois atua como um ativador que estimula a transcrição a partir do promotor P_{RE} (Figura 13.9B). Com isso, o gene *cI* é transcrito, sendo também chamado de repressor de λ , responsável pela manutenção do ciclo lisogênico inibindo a expres-

são dos genes responsáveis pelo empacotamento e pela lise. O produto do gene *cI* reprime a expressão dos genes *N* e *cro* por meio da ligação nas regiões operadoras O_R e O_L (Figura 13.9C e D). O produto do gene *cI* bloqueia a transcrição a partir de quase todos os promotores, com exceção de P_{RM} .

A proteína cI tem uma afinidade maior por O_{R1} , que é adjacente ao gene *cro*. A ligação a O_{R2} dá-se por efeito cooperativo, após ligação a O_{R1} . Uma vez que o sítio O_{R1} apresenta sobreposição parcial com P_R , a síntese do mRNA de *cro* é bloqueada, e essa combinação leva a ativação do promotor P_{RM} . Com isso, há um aumento na transcrição do gene *cI*, que quando tem sua concentração aumentada liga-se a O_{R3} , junto ao promotor, bloqueando sua própria transcrição.

Portanto, o desenvolvimento dos ciclos lítico e lisogênico é controlado pelas quantidades relativas de Cro e cI. Quando Cro é mais abundante que cI (o que ocorre na maioria das vezes) toda a expressão gênica de λ é direcionada para o ciclo lítico. Contrariamente, quando cI é mais abundante (a exceção no ciclo vital de λ), o fago tem seu DNA integrado no genoma bacteriano.

Dentre os fatores do hospedeiro (*E. coli*) que influenciam no direcionamento dos ciclos lítico/lisogênico, podemos citar os genes *hflKC* (previamente chamado de *hflA*) e *ftsH* (chamado de *hflB*). O primeiro gene codifica um complexo proteico de membrana, ao passo que o segundo codifica uma ATPase/protease integral da membrana. Resultados recentes indicam que *HflKC* é um modulador da atividade de *FtsH*, que devido a sua ação proteolítica seria responsável pela degradação de cII, favorecendo, portanto, o desenvolvimento do ciclo lítico. Contudo, a atividade proteolítica poderia ser inibida por cIII, provocando um aumento na concentração de cII e, portanto, um aumento da transcrição dos genes que estão sob o controle de P_{RE} , resultando na transcrição do gene *cI* e o direcionamento para a lisogenia. Esse mecanismo sugere a seguinte explicação para o efeito do *status* nutricional da célula hospedeira sobre o direcionamento dos ciclos lítico/lisogênico: as proteases, em geral, são mais abundantes em uma *E. coli* multiplicando-se exponencialmente e, portanto, a proteína cII (sensível à degradação) tem sua concentração reduzida, conduzindo, via de regra, ao ciclo lítico.

4. Mecanismos de controle pós-transcricional

Muita ênfase sempre foi atribuída (em procariotos) aos mecanismos reguladores de controle em nível de transcrição, envolvendo proteínas ativadoras e/ou repressoras transcricionais. No entanto, vários mecanismos de controle da expressão gênica foram identificados regulando o início da tradução. O controle da tradução está relacionado com o bloqueio do acesso do ribossomo ao mRNA, que pode ocorrer pela presença de estruturas secundárias

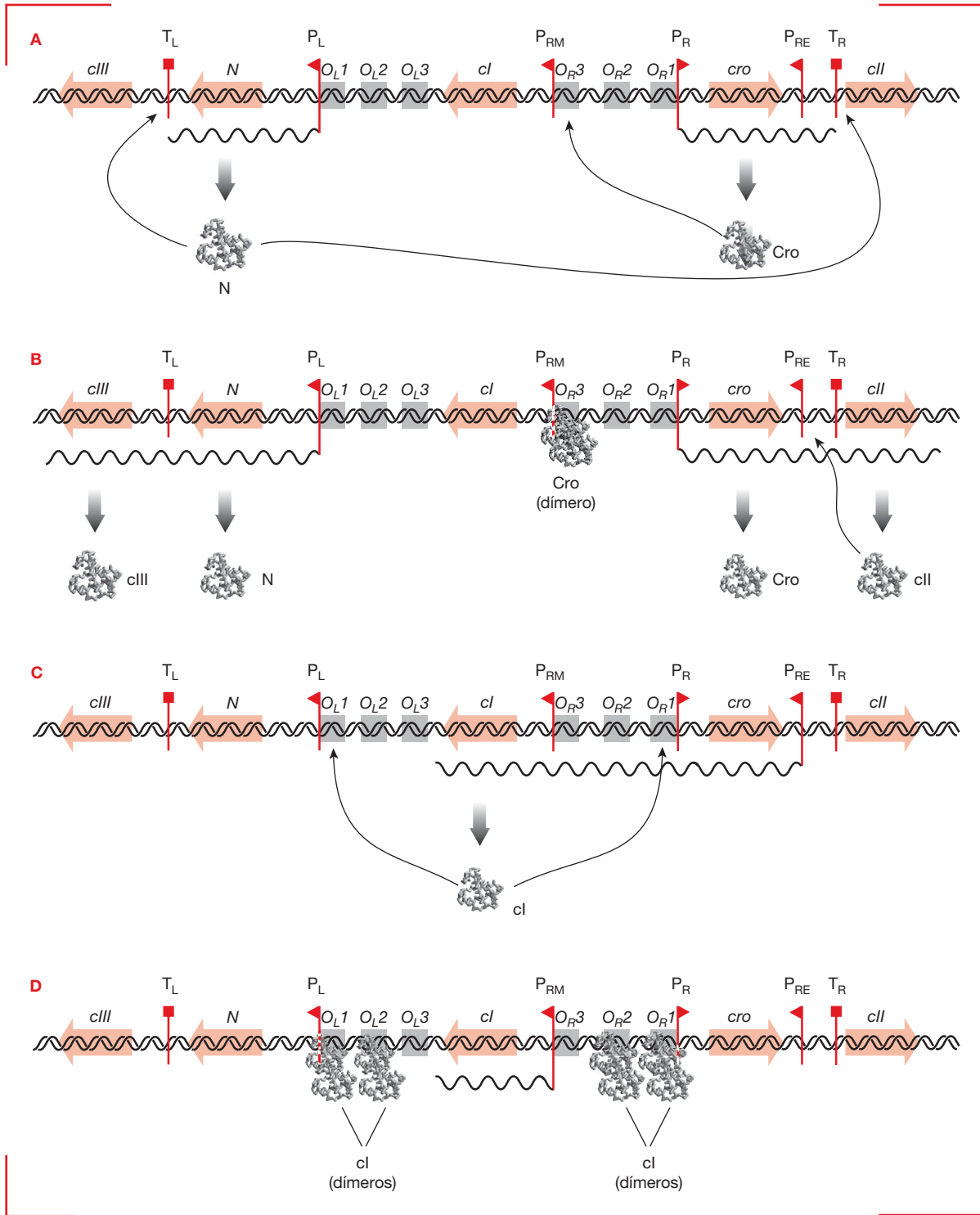


Figura 13.9

Cascata de eventos dos ciclos lítico e lisogênico de λ . (A) A transcrição é iniciada em P_L e P_{RM} , com produção das proteínas N e Cro. A proteína N tem ação antiterminadora em T_L e T_R , ao passo que Cro liga-se a O_{R3} , bloqueando a transcrição a partir de P_{RM} . (B) Em consequência da ação antiterminadora de N, são produzidos as proteínas cIII e cII. Caso cII acumule em quantidade suficiente, a transcrição é ativada a partir do promotor P_{RE} . (C) O produto do gene *cl*, transcrito a partir de P_{RE} , liga-se a O_{L1} e O_{R1} . (D) A transcrição a partir de P_L e P_R é bloqueada, ao passo que cI também ligado a O_{R2} ativa sua própria transcrição a partir de P_{RM} .

na região 5' dos mRNAs ou pela ligação de proteínas ao mRNA impedindo a ligação da subunidade 30S.

Recentemente, diversos mecanismos mediados por RNAs reguladores (também chamados riborreguladores) têm sido descritos e serão discutidos na Seção 5. Estes mecanismos têm-se mostrado fundamentais em processos como a resposta ao estresse e a virulência.

4.1 Proteínas como moléculas reguladoras

A utilização de proteínas como moléculas reguladoras da tradução é um mecanismo empregado por diferentes bactérias no controle da expressão de vários genes. Em geral, a proteína reguladora realiza sua função competindo com a ligação do ribossomo. Dois tipos de ligação da proteína ao seu mRNA podem ser utilizados: um dos mecanismos depende do reconhecimento, pela proteína, de uma estrutura secundária no mRNA; e o outro mecanismo está relacionado à ligação da proteína a uma sequência específica da molécula de RNA.

Na maioria dos mecanismos de controle da tradução por proteínas, o sítio de ligação da proteína compete com a ligação da subunidade menor do ribossomo. Podemos ressaltar alguns exemplos deste tipo de regulação, na qual a proteína controladora compete com a ligação da subunidade ribossômica 30S, reprimindo sua própria síntese: as proteínas treonil-tRNA sintetase, CsrA (de *carbon storage regulator* = regulador do armazenamento de carbono) e ribossômicas de *E. coli*; as proteínas DNA-polimerase e SSB (de *single strand binding* = ligante de

fita simples) do bacteriófago T4; e as proteínas do capsídeo dos bacteriófagos Q β , MS2 e PP7, que reprimem a tradução das suas DNA-replicases, bloqueando a replicação e direcionando o processo para a montagem das partículas virais.

O controle pós-transcricional relacionado à síntese de proteínas ribossômicas é estudado há bastante tempo. Essas proteínas, em um número de aproximadamente 70 por ribossomo, devem ser produzidas em quantidades equivalentes e em velocidades compatíveis com o metabolismo celular. Assim, células em multiplicação rápida apresentam uma taxa de síntese proteica também alta e, portanto, devem ter a síntese de proteínas ribossômicas aumentada. Contrariamente, células com metabolismo inativo devem reduzir a produção de proteínas ribossômicas.

Os genes que codificam as diversas proteínas ribossômicas estão organizados em óperons, que podem conter até 11 genes, sendo o controle da síntese dessas proteínas feito em nível de tradução. Em cada óperon pode ser encontrado, pelo menos, um gene cujo produto tenha atividade de repressão traducional. A proteína ribossômica, com atividade repressora, liga-se ao mRNA, bloqueando, desta forma, a tradução do mRNA (Figura 13.10). Uma vez que, as proteínas ribossômicas possuem uma afinidade muito maior pelo rRNA do que pelo mRNA, a repressão da tradução somente ocorre quando há um excesso de proteínas ribossômicas em relação ao rRNA. Considerando que cada cístron tem seu próprio RBS entenderia-se que, somente aquele onde houve ligação da proteína

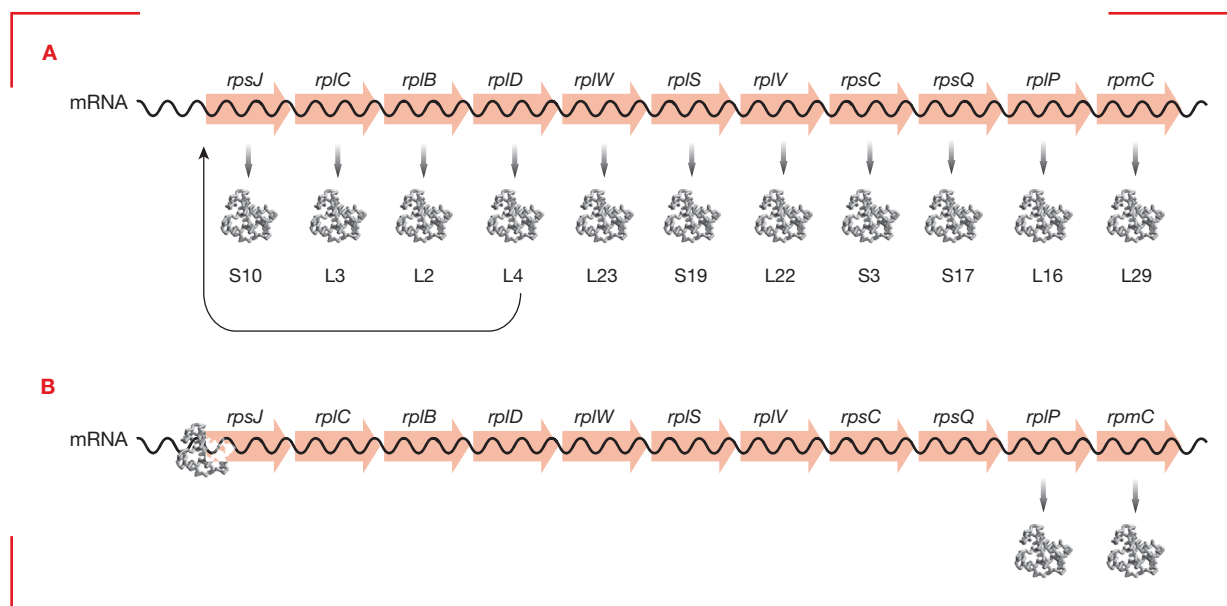


Figura 13.10

Controle da tradução de mRNAs codificando proteínas ribossômicas. (A) O esquema representa o mRNA transcrito a partir do óperon S10, com os diferentes cístrons e suas proteínas indicadas. Após a tradução, a proteína L4, quando em excesso, liga-se a uma região do mRNA anterior a *rpsJ*. **(B)** A ligação de L4 ao mRNA impede a ligação do ribossomo e a tradução da maioria dos genes, com exceção de *rplP* e *rpmC*.

ribossômica tivesse sua tradução bloqueada (por exemplo *rspJ*, na Figura 13.10). No entanto, devido à formação de estruturas secundárias (por complementaridade e emparelhamento de bases) no mRNA, a liberação dos diversos RBS depende da tradução do gene que o antecede. Dessa forma, quando o ribossomo chega ao códon de terminação em um determinado cístron são desfeitas as estruturas secundárias adjacentes do mRNA, expondo a região do RBS à ligação do ribossomo.

A regulação no início da tradução pode ser realizada por proteínas repressoras ou ativadoras do processo. No entanto, a maioria dos mecanismos deste tipo de controle regula negativamente a síntese de proteínas.

5. RNAs como moléculas reguladoras

Em organismos procarióticos, as moléculas de RNA atuam em diversas funções regulatórias, podendo controlar diferentes aspectos do metabolismo das células. A ação regulatória dessas moléculas pode ocorrer tanto na transcrição como na tradução. Os RNAs reguladores pertencem a um grupo heterogêneo de RNAs, podendo ser classificados em estruturais ou sRNAs. A classe de RNA reguladores estruturais é definida pela presença de estruturas secundárias no próprio RNA, que podem agir em *cis* ou *trans*, e a classe de sRNAs é caracterizada por pequenas moléculas de RNA, que apresentam a capacidade de se ligarem a proteínas ou outras moléculas de RNAs.

5.1 Estrutura secundária de mRNA como mecanismo regulatório: ribocomutadores (*riboswitches*)

Os ribocomutadores são uma classe de RNAs reguladores que fazem a transdução de sinal entre metabólitos e a regulação da expressão gênica. Esses ribocomutadores seguem vários dos princípios de regulação já discutidos, ou seja, havendo a necessidade de formação de estruturas secundárias no mRNA, para que haja a regulação da terminação da transcrição (mediada por grampos de terminação), ou a regulação do início da tradução (mediada pela acessibilidade do RBS). A grande diferença dos ribocomutadores é que eles não dependem da mediação por parte de proteínas acessórias intermediárias, interagindo diretamente com os metabólitos a serem regulados. Prescindem, também, da utilização de quaisquer outras moléculas de RNA externo para que a regulação seja exercida; ou seja, todos os fatores necessários para a regulação da expressão gênica estão contidos no próprio mRNA a ser regulado.

Os ribocomutadores funcionam como **aptâmeros** naturais de RNA, sendo parte integrante da porção inicial do mRNA de vários genes e óperons cujos produtos estão envolvidos com o metabolismo da célula, por exemplo, a

regulação da rotas de biossíntese de algumas vitaminas, alguns aminoácidos e algumas purinas. Os ribocomutadores são capazes de reprimir ou ativar seus genes correspondentes, alguns em nível de transcrição e outros em nível de tradução. Os ribocomutadores são, também, domínios estruturados residentes no próprio mRNA, com atuação em *cis*, estando organizados em duas regiões distintas: o domínio sensor, no qual ocorre a ligação do metabólito (a porção que funciona como aptâmero), e a plataforma de expressão, região responsável por alterar a transcrição ou a tradução do gene.

Quando os ribocomutadores atuam sobre o término da transcrição (**Figura 13.11A e C**), a ligação direta de um metabólito com estruturas secundárias presentes na porção inicial do mRNA (no domínio sensor), tanto pode gerar (em A) como pode desfazer (em C) um grampo de terminação, com o respectivo efeito de uma repressão ou de uma ativação gênica.

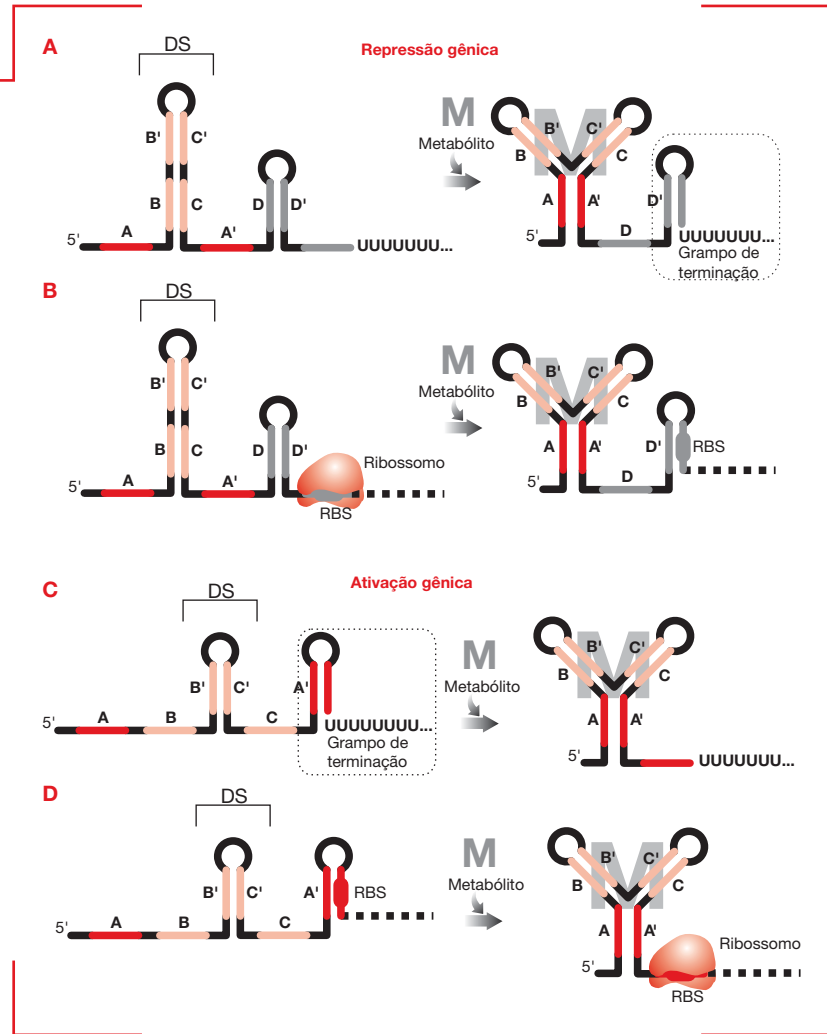
Como exemplo desse modo de controle pode ser citado o ribocomutador da flavina-mononucleotídeo (FMN). A riboflavina (vitamina B2) é um precursor essencial das coenzimas FMN e FAD (flavina-adenina-dinucleotídeo). Em *B. subtilis*, o óperon de biossíntese da riboflavina é composto por cinco genes (*ribGBAHT*), cujos produtos catalisam reações em uma via metabólica em que o substrato inicial é o GTP. A ligação direta de FMN e/ou FAD ao ribocomutador presente na porção inicial do mRNA, sendo transcrito deste óperon, leva à formação de um grampo de terminação que de maneira efetiva bloqueia a expressão do óperon. Mesmo que a transcrição tenha sido iniciada pela RNA-polimerase, a formação do grampo de terminação garantirá a atenuação do mRNA.

No momento em que os ribocomutadores atuam sobre o início da tradução (Figura 13.11B e D), a ligação direta de um metabólito às estruturas secundárias presente na porção inicial do mRNA, tanto pode tornar o RBS inacessível ao ribossomo (em B) como pode desfazer um grampo, que previamente sequestrava o RBS (em D), com o respectivo efeito de uma repressão ou de uma ativação gênica. Como exemplo desse modo de controle pode ser citado o ribocomutador da tiamina-pirofosfato (TPP).

A tiamina (vitamina B1) é a precursora do cofator TPP, essencial para várias enzimas envolvidas no metabolismo de glicídeos. Em *E. coli*, há três óperons relacionados à biossíntese da tiamina, e que estão sujeitos à regulação negativa por ribocomutadores: *thiCEFSGH*, *thiMD* e *sfuABC*. Neste exemplo, à medida que TPP se acumula na célula e sua biossíntese é dispensável, essa mesma TPP liga-se ao mRNA transcrito do óperon *thiMD*, gerando a formação de estruturas secundárias cujo resultado final será o de sequestrar o sítio de ligação ao ribossomo (RBS) do mRNA, bloqueando a expressão gênica por um mecanismo pós-transcricional. No óperon *thiCEFSGH* a plataforma de expressão do ribocomutador presente inclui um grampo de terminação, que é formado apenas com a complexação da TPP (controle transcricional).

Figura 13.11

Controle da expressão gênica por ribocomutadores. (A) A ligação de um metabólito (como um aminoácido, uma vitamina ou uma purina) ao ribocomutador gera a formação de um grampo de terminação, provocando o término da transcrição. (B) A ligação de um metabólito ao ribocomutador gera a formação de um grampo que irá sequestrar o sítio de ligação ao ribossomo (RBS), com consequente repressão gênica. (C) A ligação de um metabólito desfaz um grampo de terminação previamente existente no mRNA e permite a transcrição. (D) A ligação de um metabólito desfaz um grampo previamente existente no mRNA e permite a tradução. (A) e (C) demonstram o controle em nível de transcrição e (B) e (D) em nível de tradução. O esquema representa as sequências de ribocomutadores onde as regiões que apresentam complementaridade de bases estão representadas em cores semelhantes. DS=domínio sensor.



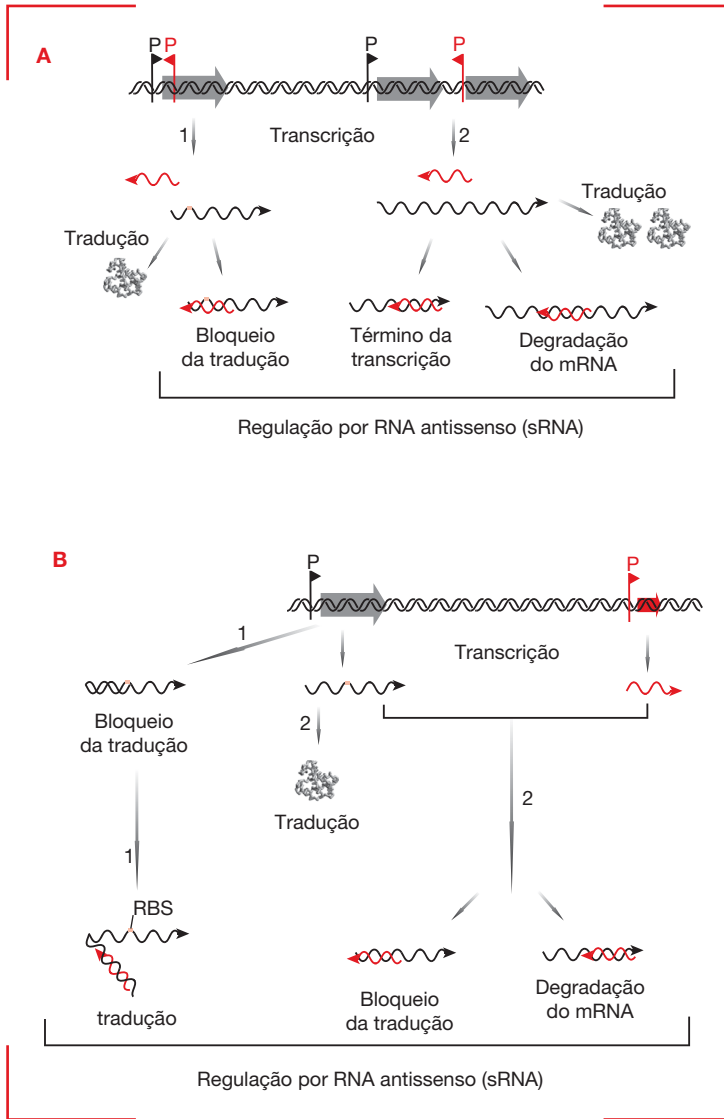
5.2 Pequenos RNAs (sRNAs)

Moléculas de RNA que atuam na regulação gênica de procariotos, foram descobertas na década de 1980, tendo como função típica realizar o pareamento de bases com o mRNA-alvo, inibindo a ligação da subunidade 30S do ribossomo e consequentemente impedindo o início da tradução. A princípio, esses RNAs reguladores foram denominados RNA antissenso, pois apresentavam complementaridade de bases, total ou parcial, com o mRNA a ser regulado. Atualmente, uma nova nomenclatura está sendo utilizada, a de pequenos RNAs (sRNAs), devido à diversidade dos mecanismos de ação destes tipos de RNA. Apesar da maioria dos sRNAs atuarem como repressores da tradução existem alguns que participam nos mecanismos de ativação de genes e também na regulação da replicação de alguns plasmídeos.

Alguns dos mecanismos realizados pelos sRNAs estão representados na **Figura 13.12**. O sRNAs que atuam em *cis* (Figura 13.12A), em geral, realizam a re-

gulação negativa dos genes. O mecanismo clássico de pareamento do sRNA com a região RBS do mRNA, que originou a identificação da regulação por RNA antissenso, depende da localização do sítio promotor do sRNA (Figura 13.12A, 1). Neste caso, o pareamento sRNA-mRNA compete diretamente com a ligação da subunidade menor do ribossomo, impedindo a tradução da proteína. Este tipo de repressão difere do mecanismo de repressão durante a transcrição (Seção 1.3), pois resulta somente em uma diminuição dos níveis da proteína. Quando a organização gênica é diferente, estando o promotor do sRNA localizado entre genes de um óperon (Figura 13.12A, 2), o mecanismo de regulação por pareamento sRNA-mRNA resultará na degradação do mRNA ou na síntese de um transcrito menor que o original.

O mecanismo de ação dos sRNAs, que atuam em *trans*, pode ser tanto por repressão como por ativação. A seleção do mecanismo de regulação dependerá da estrutura gênica e da resposta fisiológica necessária. Neste mecanismo a região do sRNA no DNA está localizada

**Figura 13.12**

Mecanismo de ação dos pequenos RNAs não codificadores (sRNAs). Representação dos mecanismos de regulação realizados pelos sRNAs (ou RNA antissenso) por meio do pareamento com os mRNAs-alvo. **(A)** Representa os mecanismos de repressão por meio da atuação em *cis* de sRNAs. **(B)** Representa os mecanismos de repressão (2) e ativação (1) por meio da atuação em *trans* de sRNAs. As regiões promotoras estão representadas pela letra P (■ –mRNA ou ■ –sRNA) acima de uma seta indicando a orientação da transcrição de cada tipo de RNA. As pontas de setas presentes nos mRNAs (■) e nos sRNAs (■) indicam também a orientação da transcrição.

distante da região codificadora do mRNA a ser regulado (Figura 13.12B). Portanto, em geral, o sRNA apresenta apenas uma complementaridade parcial com o mRNA, mas o mecanismo de ação por repressão é similar ao anteriormente descrito. O pareamento sRNA-mRNA pode ocorrer na região 5' do mRNA, impedindo a ligação do ribossomo, ou na região interna do mRNA, induzindo sua degradação (Figura 13.12B, 2).

Um dos mecanismos de ação por ativação dos sRNAs pode alterar a eficiência da tradução. Na Figura 13.12B, 1, a tradução está reduzida devido à formação de uma estrutura secundária do mRNA que bloqueia o sítio RBS (ver Capítulo 12). O pareamento do sRNA com a região do mRNA compete com esta estrutura secundária, induzindo uma alteração na conformação do mRNA e liberando o sítio RBS para a ligação da subunidade 30S. Como resultado desta alteração estrutural a síntese da proteína ocorrerá com maior eficiência.

5.2.1 sRNAs e o sequestro de proteínas reguladoras

Em *E. coli* e outras espécies bacterianas, a proteína CsrA é componente fundamental de um sistema de regulação geral da célula, envolvido na regulação de fatores de virulência, na motilidade, no sensoriamento populacional (*quorum sensing*), no metabolismo do carbono, na captação de peptídeos e no desenvolvimento de biofilmes. Essa proteína é responsável por reprimir a tradução de genes que são expressos apenas quando as células atingem a fase estacionária na multiplicação. Simultaneamente, CsrA ativa a expressão dos genes expressos na fase de multiplicação exponencial.

Uma vez que os mecanismos de ativação mediados por CsrA ainda não foram elucidados, será discutido apenas o mecanismo de repressão da tradução, que é bastante simples: CsrA liga-se a múltiplos sítios no

mRNA dos genes sendo negativamente regulados, em que pelo menos um desses sítios corresponde ao sítio de ligação ao ribossomo (RBS), impedindo, assim, a tradução do mRNA.

A participação dos RNAs reguladores ocorre no controle da atividade da proteína CsrA. Dois desses RNAs reguladores (CsrB e CsrC), também chamados de sRNAs, apresentam múltiplos sítios de ligação para a proteína CsrA, sequestrando e antagonizando a ação da mesma.

5.2.2 sRNAs: controles positivo e negativo da tradução

Um dos exemplos de RNAs reguladores corresponde ao sRNA DsrA, identificado por sua ação reguladora sobre a síntese do ácido colânico e cápsula em bactérias. DsrA atua em *trans* na ativação e na inibição da síntese de dois reguladores transcricionais (H-NS e RpoS). H-NS corresponde a uma proteína semelhante a histona, sendo responsável por bloquear a transcrição de vários genes bacterianos, ao passo que RpoS corresponde a um fator σ . Na **Figura 13.13** são apresentados os mecanismos de bloqueio e ativação da tradução por DsrA.

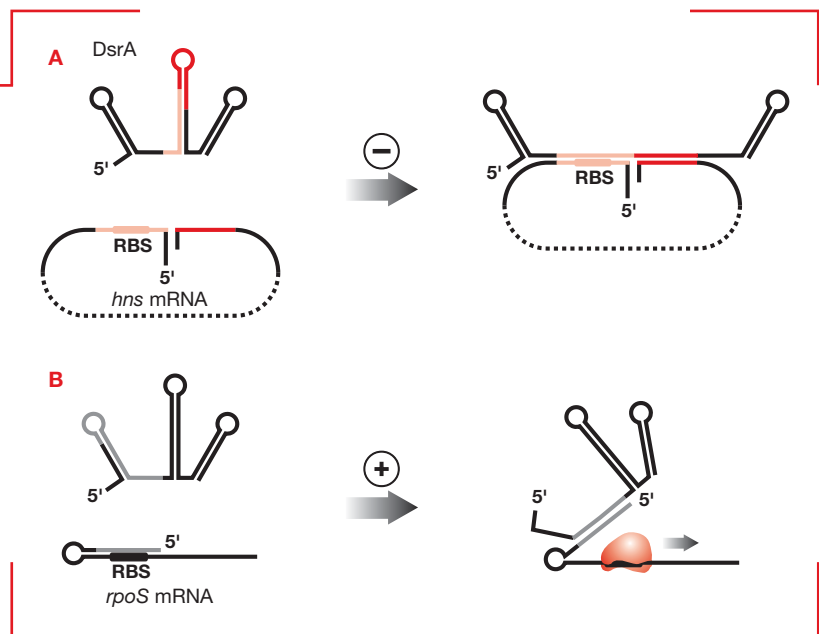
O mecanismo de ação de DsrA depende de sua interação, por complementaridade e emparelhamento de bases, com os mRNAs dos genes sendo controlados. Quando o sRNA DsrA atua bloqueando a tradução, ele interage com a região do mRNA de *hns* onde está localizado o sítio de ligação ao ribossomo (RBS). Ao formar uma estrutura de fita dupla com essa região, o RBS não fica disponível para interagir com a subunidade ribossômica menor

(30S) e iniciar o processo de tradução. Contrariamente, DsrA também pode atuar na ativação da tradução do mRNA de *rpoS*. O mRNA desse gene forma uma estrutura em grampo junto à sua extremidade 5', essa estrutura bloqueia o acesso da subunidade ribossômica 30S ao RBS do mensageiro. Neste caso, DsrA forma uma estrutura secundária estável com a região do mRNA de *rpoS* que estava bloqueando o sítio RBS, liberando-o para o início da tradução. Devido ao mecanismo de ação destes riborreguladores, eles são, também, chamados de RNA antissenso, uma vez que formam estruturas em fita dupla com o mRNA (RNA senso).

Os diferentes níveis de controle da expressão gênica podem interagir em mecanismos com capacidade de atingir um grau elevado de complexidade. Como discutido, o sRNA DsrA é um regulador negativo da tradução do gene *hns* e seu produto, a proteína H-NS, é um regulador transcricional, que reprime a expressão de vários genes de *E. coli* quando essa bactéria encontra-se fora do intestino. Neste papel, H-NS atua como um repressor típico da expressão do gene *micF*. O produto do gene *micF*, o sRNA MicF, atua como um regulador negativo da tradução de *ompF*, agindo como um RNA antissenso que bloqueia a tradução por emparelhamento de bases com a região onde está localizado o RBS. Por fim, a proteína OmpF corresponde a uma **porina de membrana externa** de *E. coli* que, juntamente com a proteína OmpC, controla o transporte passivo de pequenas moléculas hidrofílicas através da membrana. Esse exemplo ilustra como os controles em nível de transcrição e de tradução, envolvendo proteínas e RNAs reguladores, podem estar associados em uma cascata de regulação.

Figura 13.13

Regulação da tradução por DsrA. Mecanismos propostos para a regulação negativa (A) ou positiva (B) dos mRNA de *hns* e *rpoS*. As regiões com complementaridade de bases estão indicadas por cores semelhantes.



6. Outros mecanismos de controle da expressão gênica

Nas seções anteriores foram discutidos os mecanismos de controle na transcrição e no início da tradução, realizados por proteínas ou RNAs, que tradicionalmente são consideradas as principais etapas de regulação gênica. No entanto, cabe ressaltar que pode ocorrer controle da expressão gênica em outras fases destes processos, tanto no tempo de vida (estabilidade) dos mRNAs e das proteínas como no final da tradução. Seria importante salientar também, que novos mecanismos estão sendo descobertos, resultando no aparecimento de diferentes interações entre moléculas ou etapas de regulação gênica (**ver Material Complementar Online 13.1**).

6.1 Variação de fase e controle da expressão gênica

O controle da expressão gênica durante a variação de fase em bactérias pode ocorrer por meio de vários eventos como o de recombinação, combinada com o controle do

início da transcrição, que resulta na alternância da localização do promotor. **Variação de fase** é um processo de adaptação apresentado por muitas bactérias, relacionado com a variação da expressão de proteínas por um mecanismo de repressão e desrepressão de genes específicos.

Um exemplo deste tipo de mecanismo é o processo de variação de fase que ocorre em *Salmonella typhimurium*. O genoma dessa bactéria codifica duas proteínas flagelares (H1 e H2), constituintes do flagelo que é reconhecido pelo sistema imune do hospedeiro. Portanto, a capacidade da *S. typhimurium* em alternar a composição do flagelo, contendo H1 ou H2, permite sua permanência no hospedeiro, pois se evade temporariamente da resposta imune. Neste processo regulatório o gene que codifica o antígeno flagelar H2 forma um óperon com uma proteína repressora que tem seu sítio de ligação na região regulatória do gene codificador de H1. Portanto, quando ocorre a expressão de H2 o gene da proteína H1 está reprimido (**Figura 13.14A1**). No entanto, quando a proteína invertase (Hin) realiza a inversão do fragmento pela ligação aos sítios de recombinação HixL/HixR, o promotor do óperon *h2-rh1* é deslocado e não são sintetizadas as proteínas H2 e a repressora de H1 (Figura 13.14A2). Consequentemente,

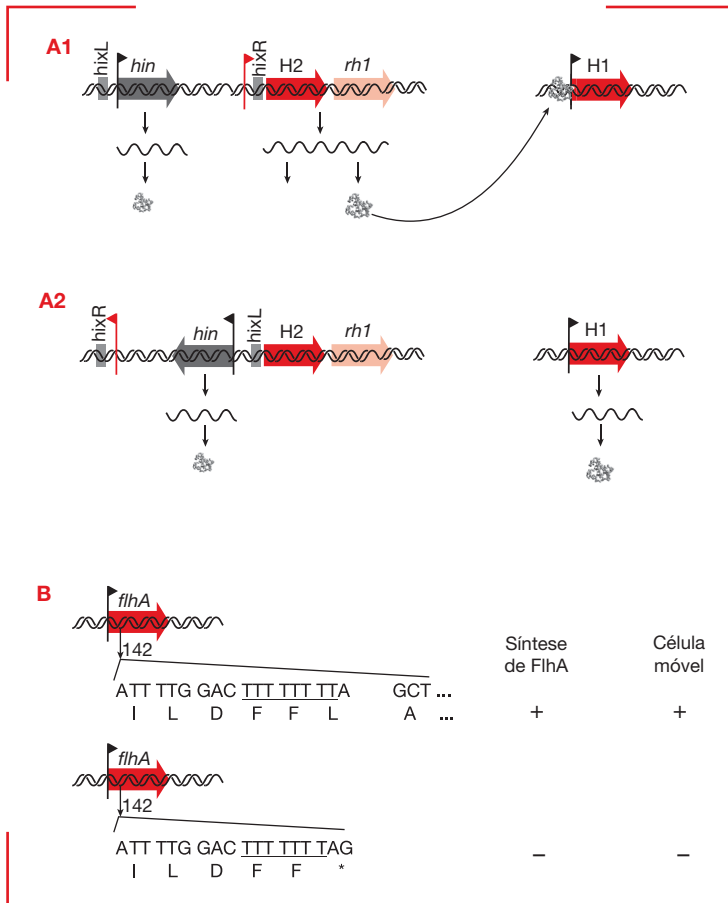


Figura 13.14

Regulação gênica em variação de fase. (A) Representação do processo de recombinação que gera variação de fase em *S. typhimurium*. Os antígenos flagelares estão representado por H1 e H2, o gene da proteína repressora de H1 por *rh1* e o da proteína invertase por *hin*. HixR e HixL são os sítios palindrômicos de ação da invertase, que realiza a inversão do fragmento. O promotor de H2 está marcado pela seta em ■. (B) Representa a mutação no gene *flhA*, resultando em variação de fase em *Campylobacter coli*. A sequência do gene está representada a partir do nucleotídeo 142, considerando o ATG. Abaixo da sequência de nucleotídeos estão os aminoácidos correspondentes. A região homopolimérica está sublinhada. * representa o término da tradução, e ... indicam que a sequência da proteína continua.

ocorre a transcrição e tradução do antígeno H1. Como o gene da proteína invertase é transcrito nas duas posições do fragmento este processo é reversível.

Outro mecanismo, que gera variação de fase, pode ocorrer pela diversidade em regiões contendo **sequências homopoliméricas**. Estas regiões homopoliméricas são sítios-alvo de mutações por inserção ou deleção de bases (ver Capítulo 7), causando mudanças de fase de leitura. As mudanças de fase de leitura (ver Capítulo 12), resultantes das mutações, geram inativações gênicas reversíveis, sendo responsáveis pelos eventos adaptativos relacionados à variação de fase. Em geral, os genomas bacterianos apresentam as sequências homopoliméricas (poli(A) ou poli(T)) localizadas nas regiões 5' dos genes regulados, por serem, provavelmente, mais eficientes para os processos de controle da expressão gênica.

Várias bactérias utilizam este mecanismo de regulação gênica, por exemplo, *B. subtilis*, *C. coli*, *Mycoplasma hominis*, *Listeria monocytogenes*. Existe, também, diversidade em relação ao gene regulado por este mecanismo, podendo estar relacionado à patogenicidade ou condições fisiológicas normais da célula.

O gene *swrA* de *B. subtilis*, relacionado à motilidade em meio sólido, apresenta variação de fase pela presença de uma sequência homopolimérica poli(A) na sua região 5'. O gene *flhA* relacionado à motilidade em *C. coli* também utiliza regulação por variação de fase para alternar entre células bacterianas móveis e imóveis (Figura 13.14B). Nesse gene, ocorre a alternância por mutações do número de repetições na sequência homopolimérica poli(T), resultando em uma proteína FlhA funcional na presença de oito "T" (8poli(T)). As células imóveis contêm, no gene *flhA*, 7poli(T) gerando uma proteína FlhA incompleta pela presença de um códon de terminação precoce (Figura 13.14B). Outro exemplo, é o gene *inlA* que codifica o fator de virulência de *L. monocytogenes*, contendo também na região 5' uma sequência poliA podendo originar, durante a variação de fase, uma proteína InlA inativa. Aparentemente, este mecanismo de variação de fase, dependente de mutações nas sequências homopoliméricas, apresenta-se como um sistema universal para regulação de expressão gênica em procariotos.

6.2 Regulação gênica na defesa bacteriana de vírus ou plasmídeos (CRISPR)

Uma classe específica de RNAs reguladores foi caracterizada e denominada RNA CRISPR (de *clustered regularly interspaced short palindromic repeats* = pequenas repetições palindrômicas agrupadas). Essa classe de RNAs regulatórios permite que a célula procariótica desenvolva

resistência à infecção por bacteriófagos ou evite conjugação plasmidial. CRISPR são sequências conservadas de 24-27 pb que se encontram repetidas no genoma de bactérias e arqueas, sendo separadas por espaçadores contendo sequências de nucleotídeos variáveis. Localizados adjacentes às repetições CRISPR estão os genes codificadores de proteínas CAS (de *Cascade complex of CAS proteins* = proteínas do complexo CAS). Essas proteínas apresentam similaridade estrutural e funcional com endonucleases, exonucleases, helicases e proteínas que se ligam ao DNA ou ao RNA.

Os organismos procarióticos desenvolveram diferentes mecanismos de resistência à infecção viral, como o bloqueio da adsorção ou restrição a injeção do material genético do fago (ver Capítulo 9). O mecanismo de resistência por CRISPR ocorre pela incorporação, no genoma da bactéria ou arquea, do material genético do fago na forma de espaçadores, permitindo ao procarionto desenvolver imunidade ao fago que tiver a mesma sequência.

A região genômica onde está localizada o CRISPR é transcrita como um longo RNA, que será processado pelas proteínas CAS em unidades contendo pequenas sequências repetidas denominadas crRNA. As moléculas de fita simples das crRNAs são, então, direcionadas pela CAS para o pareamento com as moléculas alvo do fago ou do plasmídeo. Finalmente, o material genético do fago ou do plasmídeo será degradado (Figura 13.15A).

Este mecanismo permite que a célula procariótica adquira novas sequências CRISPR a cada infecção viral ou conjugação plasmidial. Nos estágios iniciais de entrada na célula do novo material genético (viral ou plasmidial), as proteínas CAS reconhecem e clivam o material genético na região das sequências CRISPR e as incorporam, no genoma celular, na região 5' do grupo de CRISPRs já existentes. Desta forma, ocorre uma amplificação das sequências CRISPR com variabilidade nos espaçadores, permitindo que a resistência a infecção viral (ou conjugação plasmidial) ocorra para uma diversidade de fagos ou plasmídeos (Figura 13.15B).

O poeta romano Juvenal (*Decimus Iunius Iuvenalis*) escreveu em seu livro VI de Sátiras: *Quis custodiet ipsos custodes? Dezenove séculos depois*, sua pergunta "Quem controla os controladores?" ainda é pertinente, pois apesar de todo o conhecimento que se tem sobre os mecanismos reguladores da expressão gênica, não existe uma resposta definitiva. Resta o consolo de que a visão da intricada teia de mecanismos reguladores é lenta e permanentemente aprofundada, com os dados e as análises acumuladas dos estudos dos genomas, dos transcritomas, dos proteomas e dos metabolomas de *E. coli* e de diversos outros organismos procarióticos.

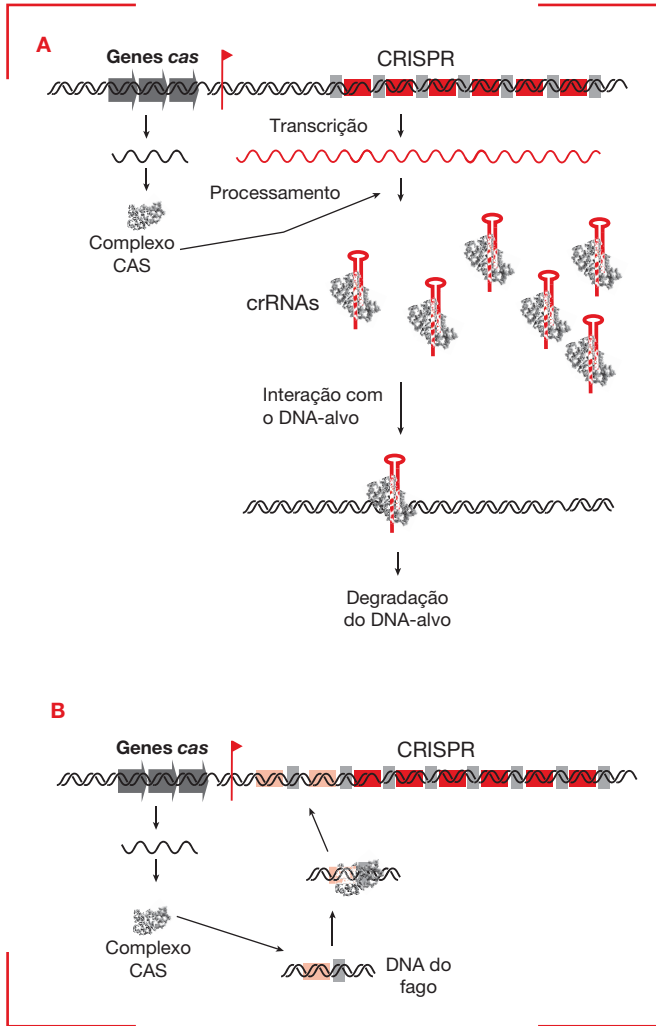


Figura 13.15

Regulação na defesa bacteriana (CRISPR). (A) e (B) representam a região do genoma contendo a região CRISPR com as sequências conservadas (retângulos em ■) separadas pelos espaçadores (retângulos em □) e os genes codificadores de proteínas CAS (setas em ■ demonstrando a orientação da transcrição). A seta vertical em ■ representa o promotor. (A) Demonstra o mecanismo de transcrição da região de CRISPR e das proteínas CAS, a formação do complexo proteico CAS, o processamento do RNA de CRISPR pelo complexo CAS e a interação de CAS com os pequenos RNAs de fita simples, formando os crRNAs (estrutura em grampo em ■ e em □ o complexo CAS). Finalmente, a interação dos crRNAs com o DNA-alvo, gerando sua degradação. (B) Demonstra o mecanismo de incorporação de novas sequências CRISPRs (retângulos em ■) originadas do DNA do fago que foi recentemente incorporado na célula.



Resumo

Dentre o repertório de genes contidos em um genoma procarionótico, aqueles que devem ser expressos pela célula em um determinado momento, ou em resposta a fatores internos ou externos, são regulados por mecanismos bem definidos de controle da expressão gênica. Esses mecanismos podem operar em nível transcricional ou pós-transcricional.

Em nível transcricional, um tipo de controle pode ser exercido pela disponibilidade de diferentes fatores σ , essenciais para o início da transcrição. Ainda em nível transcricional, o controle pode ser exercido por proteínas repressoras (controle negativo) ou ativadoras (controle positivo), que interagem com sítios

específicos no DNA e impedem ou facilitam, respectivamente, a ligação (ou movimentação) da RNA-polimerase. As moléculas efetoras nestes casos são os indutores ou correpressores, moléculas pequenas (metabólitos) cuja disponibilidade no meio modulará a resposta transcricional, pela interação com os repressores ou ativadores. Os genes sujeitos a esse tipo de controle estão muitas vezes organizados sob forma de óperons, conjuntos policistrônicos de controle transcricional. Outro nível de controle transcricional pode ser exercido por meio da formação seletiva de grampos de terminação, acoplados ao processo de tradução, no mecanismo de atenuação da transcrição.

Com relação ao controle pós-transcricional, diversos mecanismos tem sido descritos. Um primeiro envolve a ligação de proteínas à extremidade 5' do mRNA, bloqueando a tradução pelo ribossomo. Uma modulação deste mecanismo poderia ocorrer pela disponibilidade de sRNAs reguladores, com múltiplos sítios de alta afinidade por estas proteínas, e que as removeriam de circulação. Outra forma, ainda, de ação dos RNAs riborreguladores envolve sua ligação, por complementaridade e emparelhamento de bases, à extremidade 5' do mRNA cuja tradução será regulada.

Em um controle positivo da tradução, a ligação do sRNA riborregulador ao seu alvo desfaz uma es-

trutura secundária (grampo) que bloqueava o acesso do ribossomo ao seu sítio de ligação. No controle negativo, a ligação do riborregulador ao alvo bloqueia o sítio de ligação ao ribossomo. Finalmente, um nível diferenciado de controle da expressão gênica é exercido pelos ribocomutadores, capazes de interagir diretamente com metabólitos por meio da formação de estruturas secundárias bem definidas no RNA, e exercer dois níveis de controle: transcricional, mediado pela formação/remoção de grampos de terminação; ou pós-transcricional, regulando o início da tradução pela acessibilidade do sítio de ligação ao ribossomo.



Leituras recomendadas

Babitzke P, Baker CS, Romeo T. Regulation of translation initiation by RNA binding proteins. *Annu Rev Microbiol.* 2009;63:27-44.

Babitzke P, Romeo T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr Opin Microbiol.* 2007;10(2):156-63.

Fröhlich KS, Vogel J. Activation of gene expression by small RNA. *Curr Opin Microbiol.* 2009;12(6):674-82.

Hacker J, Blum-Oehler G. In appreciation of Theodor Escherich. *Nat Rev Microbiol.* 2007;5:902.

Huang HY, Chang HY, Chou CH, Tseng CP, Ho SY, Yang CD, et al. sRNAMap: genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes. *Nucleic Acids Res.* 2009;37:D150-4.

Johansson J, Mandin P, Renzoni A, Chiaruttini C, Springer M, Cossart P. An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*. *Cell* 2002;110(5):551-61.

Loh E, Dussurget O, Gripenland J, Vaitkevicius K, Tiensuu T, Mandin P, et al. A trans-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell.* 2009;139(4):770-9.

Orsi RH, Bowen BM, Wiedmann M. Homopolymeric tracts represent a general regulatory mechanism in prokaryotes. *BMC Genomics.* 2010;11:102.

Serganov A, Polonskaia A, Phan AT, Breaker RR, Patel DJ. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature.* 2006; 441(7097):1167-71.

Waters LS, Storz G. Regulatory RNAs in bacteria. *Cell.* 2009;136(4):615-28.

Zhou D, Yang R. Global analysis of gene transcription regulation in prokaryotes. *Cell. Mol Life Sci.* 2006;63(19-20):2260-90.

Arnaldo Zaha

Controle da Expressão Gênica em Eucariotos

- 1. Os genes podem ser regulados em diferentes etapas da expressão gênica 302**
- 2. Regulação do início da transcrição 302**
- 3. Fatores de transcrição 303**
 - 3.1 Ativadores e repressores da transcrição 303
 - 3.2 Motivos presentes em proteínas que se ligam ao DNA 304
 - 3.2.1 Homeodomínio 304
 - 3.2.2 Dedo de zinco 304
 - 3.2.3 Zíper de leucina 305
 - 3.2.4 Hélice-alça-hélice 305
 - 3.3 Domínios de ativação 305
- 4. Formação de complexos com múltiplos componentes 305**
- 5. Modificações das histonas 306**
 - 5.1 Acetilação e desacetilação 306
 - 5.2 Metilação e desmetilação 307
 - 5.3 Ubiquitinação e desubiquitinação 307
- 6. Remodelamento da cromatina 307**
- 7. Controle da transcrição: o exemplo dos receptores nucleares de hormônios 308**
 - 7.1 Mecanismos de ação dos receptores de hormônios esteroides 309
- 8. Regulação de expressão de locus multigênicos: os genes de globina como modelo 309**
- 9. RNAs não codificadores no controle da transcrição 312**
- 10. Regulação pós-transcricional 312**
 - 10.1 Pequenos RNAs na regulação da expressão gênica 312
 - 10.1.1 miRNAs 313
 - 10.1.2 siRNAs 315
 - 10.1.3 piRNAs 315
 - 10.2 *Turnover* de mRNA 315
- 11. Regulação da expressão no nível da tradução 315**
- 12. Controle de expressão pela localização celular do mRNA 317**

Os organismos eucarióticos, particularmente os multicelulares, apresentam características dependentes de processos controlados, como a transcrição e a síntese de proteínas, que ocorrem nas suas células, possibilitando, por exemplo, que elas se diferenciem para realizar funções específicas.

Nos eucariotos, temos de levar em consideração o fato de apresentarem algumas características importantes em relação à organização do genoma, que afetam a forma como a expressão dos genes é regulada. A partir das informações já estudadas sobre os genes que codificam proteínas em eucariotos, os nucleossomos formados pela associação do DNA às histonas e os processos envolvidos na expressão gênica das células eucarióticas, será analisada a forma como esses aspectos afetam a regulação da expressão gênica.

Com o avanço nos estudos genômicos, a regulação da expressão gênica passou a ser estudada de forma global, em que a transcrição de um grande conjunto de genes pode ser analisada simultaneamente. Além de possibilitar a identificação de genes expressos de forma específica em diferentes órgãos, tecidos e células, esses estudos têm mostrado que muitos RNAs que não codificam proteínas (muitos transcritos a partir dos íntrons), são detectados. Por outro lado, a própria organização espacial dos cromossomos no núcleo parece influenciar a expressão gênica.

Em geral, os processos que ocorrem em uma célula, como replicação do DNA, transcrição, processamento dos RNAs e síntese de proteínas, são apresentados separadamente, de forma didática, para facilitar o entendimento. Entretanto, sabe-se que a maioria desses processos estão integrados.

Neste capítulo, será visto, como a expressão dos genes é regulada em eucariotos, levando-se em consideração as diferentes etapas do processo da expressão gênica, desde a transcrição até a produção de uma proteína funcional. Na etapa de transcrição, as modificações da cromatina são fundamentais na regulação. Também será abordada a forma como determinados RNAs que não codificam proteínas atuam na regulação da expressão gênica.

1. Os genes podem ser regulados em diferentes etapas da expressão gênica

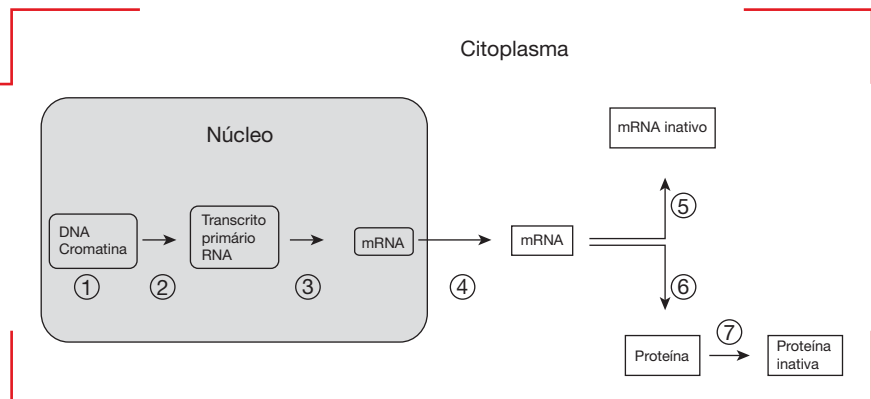
A expressão de um determinado gene em uma célula eucariótica envolve a modificação da cromatina, para possibilitar a sua transcrição e gerar uma molécula precursora de RNA no núcleo da célula. As modificações posteriores desse precursor podem dar origem ao RNA mensageiro maduro, proporcionar o transporte do mRNA maduro para o citoplasma e a tradução dele em uma proteína. A célula apresenta mecanismos de regulação em todas essas etapas da expressão gênica. A **Figura 14.1** mostra, em esquema, as diferentes etapas da expressão gênica. A regulação dessa expressão pode ser realizada em qualquer uma dessas etapas. Um dos níveis importantes da regulação é o início da transcrição, em que sequências reguladoras são reconhecidas pela maquinaria de transcrição, possibilitando a ativação do gene e levando à síntese de RNA. O RNA precursor sintetizado sofre, então, uma série de modificações para gerar o mRNA maduro, que é transportado para o citoplasma da célula, onde será traduzido em uma cadeia polipeptídica.

2. Regulação do início da transcrição

Dois mecanismos, não exclusivos, têm sido propostos para explicar a regulação do início da transcrição dos genes em metazoários. Em um dos modelos propõe-se que, após um determinado sinal, proteínas ativadoras da transcrição ligam-se às sequências reguladoras dos genes para promover o recrutamento da RNA-polimerase II (RNA-pol II). O outro modelo envolveria a regulação da atividade da RNA-pol II após o seu recrutamento, primariamente pela sua fosforilação. Além disso, estudos recentes em escala genômica têm mostrado que a RNA-pol II poderia estar ligada a promotores de muitos genes,

Figura 14.1

Etapas da expressão gênica, com as indicações dos possíveis mecanismos de regulação. ① Alteração da cromatina; ② regulação transcripcional; ③ regulação no processamento; ④ regulação no transporte e localização; ⑤ regulação na estabilidade dos mRNAs; ⑥ regulação na tradução; ⑦ regulação na atividade das proteínas.



que não estariam sendo expressos, antes de algum sinal específico de ativação.

Os genes apresentam regiões importantes que são reconhecidas pela maquinaria transcricional, possibilitando a síntese dos RNAs. As características dessas regiões dos genes, que correspondem aos promotores e a outros elementos reguladores, foram apresentadas no Capítulo 10.

3. Fatores de transcrição

As sequências reguladoras, presentes nos elementos promotores e reforçadores dos genes, são reconhecidas pelos **fatores de transcrição**. Existem, basicamente, duas classes de fatores de transcrição: uma é composta pelos fatores de transcrição basais (FTBs), necessários para a expressão de todos os genes transcritos pelo mesmo tipo de RNA-polimerase, estando, dessa forma, presentes na maioria das células; a outra classe é constituída por fatores de transcrição específicos, responsáveis pela expressão de genes que codificam proteínas restritas para um determinado tipo ou momento celular. Os FTBs já foram descritos no Capítulo 10. Na segunda classe, estão incluídos os ativadores, coativadores, repressores e correpressores da transcrição, discutidos a seguir.

3.1 Ativadores e repressores da transcrição

A transcrição é um processo regulado, que exige a participação de um complexo formado por um grande número de proteínas. Ativadores, coativadores, repressores e correpressores, em geral, são proteínas (fatores de transcrição) que fazem parte desse complexo envolvido na regulação da transcrição.

É importante ressaltar que, em eucariotos, o DNA está complexado com várias proteínas, formando a cromatina (ver Capítulo 3). Muitas das proteínas envolvidas na regulação da transcrição atuam na reorganização da estrutura da cromatina, possibilitando a ativação da transcrição.

Os ativadores, geralmente, são constituídos por domínios funcionais que possibilitam o transporte para o

núcleo, o reconhecimento das sequências reguladoras dos genes, a ativação da transcrição e a interação com outras proteínas. Estudos realizados na década de 1980, por Roger Brent e Mark Ptashne, com GAL4 – uma proteína ativadora de transcrição de levedura –, forneceram as primeiras informações sobre a estrutura dos fatores de transcrição. GAL4 é responsável pela ativação da transcrição de genes que codificam proteínas envolvidas no metabolismo de galactose em *Saccharomyces cerevisiae*. Os pesquisadores demonstraram que era possível construir uma proteína híbrida, contendo a região de ligação ao DNA da proteína LexA (proteína repressora de *Escherichia coli*) e a porção de ativação da transcrição da proteína GAL4. A proteína híbrida LexA-GAL4 foi capaz de reconhecer e se ligar a uma sequência reguladora específica de genes controlados por LexA e ativar a transcrição na levedura. A proteína, porém, só era capaz de fazer isso se a sequência reguladora fosse a do operador de *lexA* e estivesse localizada próxima ao sítio de início da transcrição.

Dessa forma, eles mostraram que a região de GAL4 responsável pela ligação ao DNA poderia ser substituída por uma região equivalente de uma proteína procariótica, sem perda da função de ativação transcricional. Eles propuseram que GAL4 deveria ser constituída por dois domínios, um responsável pela interação com a sequência reguladora (denominada UAS, de *upstream activating sequence* = sequência ativadora a montante) do gene cuja expressão é controlada, e o outro, responsável pela ativação transcricional. A ativação ocorreria pela interação desse último domínio com outras proteínas. A **Figura 14.2** mostra as etapas desses experimentos e a estrutura da proteína GAL4 com os diferentes domínios.

As análises genéticas e bioquímicas dos mecanismos envolvidos na regulação da transcrição mostraram que o controle também pode ser exercido por repressores. Os repressores atuam de maneira contrária aos ativadores, inibindo a transcrição dos genes. A análise da expressão gênica na levedura mostrou, por exemplo, que a expressão constitutiva de determinados genes em linhagens mutantes estava associada à ação de repressores. De forma semelhante aos ativadores, os repressores reconhecem as sequências reguladoras nos genes. No caso da sequência reconhecida por um ativador, uma mutação nessa região do gene, impedindo o reconhecimento pelo ativador, levaria a uma diminuição na sua expressão. Por outro lado,

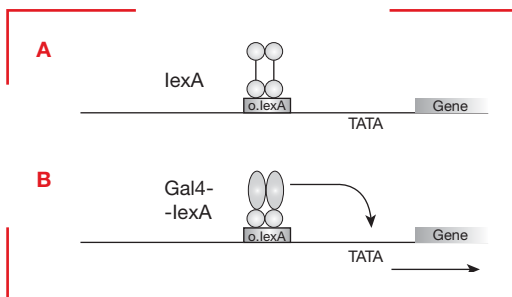


Figura 14.2

Experimento de ativação gênica demonstrando a existência de domínios de ligação ao DNA e de ativação. (A) A proteína LexA completa reconhece o operador de *lexA* (*o.lexA*), mas não é capaz de ativar a transcrição. (B) No caso da proteína híbrida LexA-Gal4, ocorre o reconhecimento do operador *o.lexA*, pela porção de LexA que se liga ao DNA, e a porção da proteína GAL4 promove a ativação da transcrição do gene.

uma mutação na sequência do gene reconhecida pelo repressor levaria a um aumento na transcrição do gene. Assim como os ativadores, os repressores também apresentam domínios bem definidos: um domínio de ligação ao DNA e um domínio de repressão.

Os estudos comparativos das sequências deduzidas de aminoácidos dos fatores de transcrição mostram a existência de classes de motivos estruturais que seriam responsáveis pela ligação ao DNA e motivos que seriam responsáveis pela ativação ou repressão da transcrição.

3.2 Motivos presentes em proteínas que se ligam ao DNA

Grande parte dos fatores de transcrição eucarióticos conhecidos apresenta um dos seguintes motivos estruturais para ligação ao DNA: homeodomínio, dedo de zinco (*zinc finger*), zíper de leucina (*leucine zipper*), hélice-alça-hélice (*helix-loop-helix*) ou hélice “alada” (*winged-helix*). Estudos recentes demonstram que, no genoma humano, devem existir cerca de 1.700 genes que codificam fatores de transcrição que se ligam a sequências específicas de DNA. Três tipos desses fatores representam mais de 80% do repertório de fatores de transcrição, sendo o dedo de zinco tipo C_2H_2 o mais frequente, seguido pelo homeodomínio e por hélice-alça-hélice. A **Figura 14.3** mostra, esquematicamente, as estruturas tridimensionais de alguns desses motivos estruturais.

3.2.1 Homeodomínio

O motivo denominado **homeodomínio** foi descrito inicialmente em proteínas envolvidas no processo de desenvolvimento da *Drosophila melanogaster*. Os genes que

codificam essas proteínas apresentam uma região conhecida como *homeobox*. O *homeobox* é uma sequência de DNA de 180 pb dentro da região codificadora de certos genes, que codifica um homeodomínio de 60 aminoácidos. A estrutura do homeodomínio é constituída por três α -hélices, sendo que a α -hélice 3 realiza o contato com as bases nitrogenadas da molécula de DNA. O nome *homeobox* é atribuído ao fato de que mutações em alguns desses genes resultam em uma transformação homeótica, em que uma estrutura toma o lugar de outra. Por exemplo, a linhagem de *Drosophila* com a mutação *Antennapedia* pode desenvolver patas na cabeça, em vez de antenas. Sequências contendo *homeobox* têm sido descritas em fatores de transcrição nos mais diferentes organismos, sendo que muitos desses fatores estão envolvidos com o desenvolvimento e a diferenciação.

3.2.2 Dedo de zinco

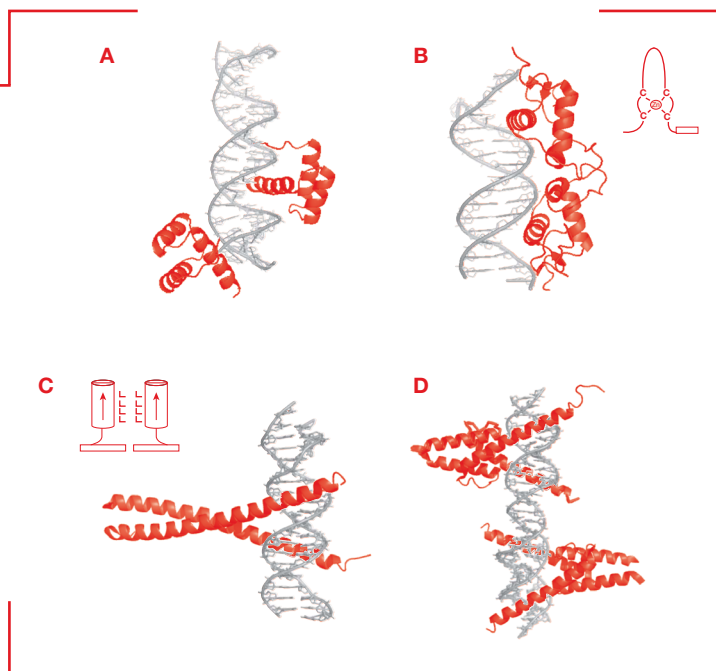
O **dedo de zinco** é um motivo estrutural de ligação ao DNA descoberto no fator TFIIIA, que está envolvido na transcrição de rRNA 5S pela RNA-polimerase III. Esse motivo tem sido encontrado, também, em vários outros fatores de transcrição.

A sequência de consenso de dedo de zinco tipo C_2H_2 é $C-X_{2-4}-C-X_3-F-X_5-L-X_2-H-X_3-H$, em que C é cisteína, X é qualquer aminoácido, F é fenilalanina, L é leucina e H é histidina. As cisteínas e as histidinas constituem o sítio de ligação de zinco (C_2-H_2). Esse é o motivo de dedo de zinco encontrado, com maior frequência, nos fatores de transcrição eucarióticos. O dedo de zinco tipo C_4 apresenta a sequência de consenso $C-X_2-C-X_{13}-C-X_2-C-X_{14-15}-C-X_5-C-X_9-C-X_2-C$, em que cada um dos dois grupos de quatro cisteínas é essencial para a ligação do íon Zn^{2+} . Esses

Figura 14.3

Representação da estrutura tridimensional de alguns motivos estruturais presentes em fatores de transcrição que se ligam ao DNA.

As estruturas representam os complexos dos motivos estruturais com DNA e se baseiam nas estruturas depositadas no *Protein Data Bank* (PDB). As regiões da molécula de DNA (em) e os motivos estruturais das proteínas (em) estão representadas. **(A)** Homeodomínio da proteína codificada pelo gene *Engrailed* de *D. melanogaster*; **(B)** Receptor de glicocorticoide de rato – no detalhe, está representado um dedo de zinco com quatro cisteínas “C” interagindo com o zinco (Zn). **(C)** Fator GCN4 de *S. cerevisiae* – no detalhe, é apresentado o esquema do zíper de leucina (L). **(D)** Fator MyoD de camundongo com o motivo estrutural hélice-alça-hélice.



motivos são encontrados, por exemplo, em receptores de hormônios esteroides.

O número de dedos de zinco, nas diferentes proteínas, pode variar de um a até mais de trinta. Além disso, o sítio de ligação de zinco também pode oscilar, havendo proteínas em que quatro cisteínas se ligam ao zinco, como no caso dos receptores de hormônios esteroides, e outras em que uma histidina e quatro cisteínas formam o sítio de ligação ao zinco, como no caso da proteína codificada pelo gene *gag* de retrovírus.

3.2.3 Zíper de leucina

As sequências que constituem o **zíper de leucina** são caracterizadas por repetições de leucina a cada sete aminoácidos, em uma região de aproximadamente trinta aminoácidos. Esse motivo foi descrito, primeiramente, no fator GCN4 de *S. cerevisiae* e, em seguida, no fator de mamíferos C/EBP e nos produtos dos oncogenes *fos*, *jun* e *myc*. As proteínas apresentam estruturas de α -hélice com 3,6 resíduos por volta, com uma leucina a cada duas voltas localizada em um dos lados da hélice. Em uma proposta inicial, duas proteínas contendo esse motivo seriam capazes de formar dímeros antiparalelos com interdigitação das leucinas, surgindo daí o nome zíper de leucina. No modelo atual, as proteínas formam dímeros com as α -hélices paralelas com interação por meio das leucinas.

No caso do fator GCN4 e dos produtos de *fos* e *jun*, foi demonstrado que o motivo zíper de leucina é responsável pela formação do dímero, e que a região responsável pela ligação ao DNA está situada próxima à porção N-terminal do zíper de leucina, sendo rica em resíduos de lisina.

Algumas proteínas com motivo zíper de leucina podem formar heterodímeros, como, por exemplo, o fator AP-1, constituído de uma proteína *fos* e uma proteína *jun*. Essa possibilidade amplia as chances de combinação de domínios de ativação e repressão, com mudança das propriedades reguladoras de uma molécula ligada a um sítio no DNA.

3.2.4 Hélice-alça-hélice

Alguns fatores de transcrição apresentam um domínio de ligação ao DNA representado por um motivo estrutural constituído de duas α -hélices separadas por uma alça não helicoidal. Uma das α -hélices contém aminoácidos básicos que interagem com o DNA. Esse motivo denominado **bHLH** (de *basic helix-loop-helix*) é encontrado em fatores de transcrição que regulam a especificação do tipo celular, a diferenciação e a morfogênese de uma grande variedade de células. A formação do músculo esquelético, por exemplo, é controlada por uma rede transcricional que utiliza fatores do tipo bHLH.

3.3 Domínios de ativação

Um **domínio de ativação** é a região de uma cadeia polipeptídica capaz de ativar a transcrição. Embora os domínios de ativação apresentem uma considerável diversi-

dade em sua sequência de aminoácidos, muitos mostram uma composição rica em determinados aminoácidos. Por exemplo, os fatores GAL4 e GCN4 de *S. cerevisiae* possuem domínios de ativação ricos em aminoácidos ácidos (ácido aspártico e ácido glutâmico). Outros domínios de ativação são ricos no aminoácido glutamina, como os fatores Oct1, Oct-2, Sp1 e Antennapedia, e outros ainda, em prolina, como os fatores AP-2 e CTF/NF1.

Um dos possíveis mecanismos de ação dos domínios de ativação poderia ser pelo aumento na velocidade da formação do complexo basal no início da transcrição pela RNA-pol II. Sabe-se, por exemplo, que ativadores contendo domínios de aminoácidos ácidos interagem com TFIID, estimulando a sua ligação ao promotor. TFIIB, outro fator envolvido no início da transcrição por RNA-pol II, também interage com domínios de ativação dos fatores de transcrição, estimulando o recrutamento desse fator. A ação dos ativadores sobre os fatores de transcrição basais, para estimular a transcrição, não se restringiria ao aumento do recrutamento deles ao promotor. Os ativadores poderiam atuar alterando a conformação desses fatores e aumentando a sua atividade. Por outro lado, sabe-se que a interação de muitos ativadores com o complexo basal de transcrição ocorre de forma indireta, com a participação de outros fatores. Além disso, muitos genes apresentam outras sequências reguladoras, como os reforçadores, indicando que os componentes envolvidos na transcrição se organizam de forma mais complexa.

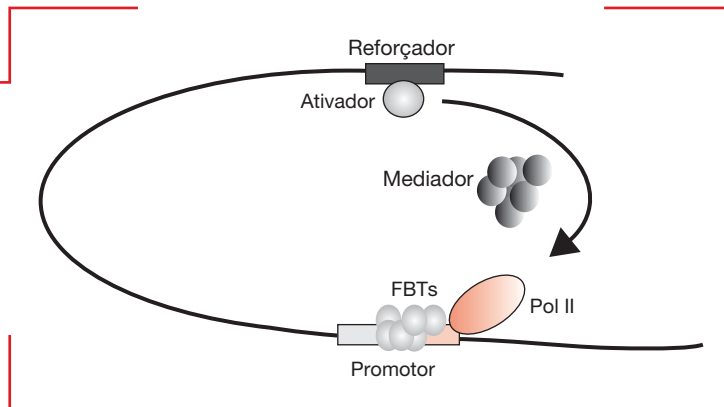
4. Formação de complexos com múltiplos componentes

Junto com os promotores, os elementos reguladores conhecidos como reforçadores exercem uma função muito importante na regulação da expressão gênica. Os reforçadores são reconhecidos por ativadores que, com a participação de outros fatores (coativador, mediador), interagem com o complexo multiproteico basal de transcrição, aumentando o nível da transcrição (**Figura 14.4**).

Os estudos realizados em *S. cerevisiae* pelo grupo de Roger D. Kornberg mostraram a existência de um complexo proteico, que denominaram **mediador**, importante para a regulação da transcrição pela RNA-pol II. O mediador é constituído por mais de 20 subunidades proteicas. Posteriormente, subunidades correspondentes foram descritas em várias espécies de metazoários, inclusive em mamíferos, indicando que o mediador teria surgido muito cedo na evolução dos eucariotos e que teria ocorrência em todas as espécies. Descrito, a princípio, como um componente importante para a ativação da transcrição, o mediador mostrou-se em todas as espécies um componente essencial para a transcrição por RNA-pol II a partir de todos os promotores. Ele interage com o ativador ligado ao reforçador e com a RNA-pol II no complexo de iniciação de transcrição. O mediador é importante tanto para ativação como para repressão da transcrição. A forma como o

Figura 14.4

Modelo de ativação de transcrição mostrando a ação conjunta de ativadores, mediadores e fatores basais de transcrição. FBTs – fatores basais de transcrição; Pol II – Complexo RNA-polimerase II.



mediador atua na transcrição ainda não é bem entendida, mas acredita-se que ele atue principalmente na etapa de formação do complexo de iniciação da transcrição, aumentando a eficiência e a velocidade da formação do complexo, ou afetando o recrutamento da RNA-pol II, TFIID ou outros fatores de transcrição gerais.

Embora muitos dos avanços nos estudos da regulação da expressão gênica em eucariotos tenham surgido de experimentos realizados com o DNA purificado, isto é, sem as proteínas que interagem com o DNA na célula, sabe-se que o estado da cromatina é muito importante para a regulação da expressão gênica.

Como abordado no Capítulo 3, nas células eucarióticas, o DNA é compactado na forma de cromatina. A cromatina é uma estrutura organizada na qual as unidades básicas são os nucleossomos. Geralmente, a estrutura compactada da cromatina não possibilita a interação de proteínas reguladoras com o DNA. Dessa forma, os nucleossomos precisam ser remodelados durante a ativação gênica. O empacotamento do DNA na estrutura da cromatina afeta todas as etapas da transcrição, desde a ligação do ativador e formação do complexo de iniciação até o alongamento e a terminação.

5. Modificações das histonas

Uma das características das histonas é sua estrutura globular, exceto pelas “caudas” aminoterminais que se apresentam desestruturadas. As histonas e, em particular, essas porções aminoterminais caracterizam-se por apresentar um grande número de modificações. Existem diversos tipos de modificações das histonas, que incluem metilação de resíduos de arginina; metilação, acetilação, ubiquitinação, ADP-ribosilação e sumoilação de resíduos de lisina; e, fosforilação de serinas e treoninas. Essas modificações ocorrem tanto nas histonas H2A, H2B, H3 e H4 como em histonas variantes, como H3.1, H3.3 e HTZ.1 (consulte o Capítulo 3 para saber mais sobre as histonas). Algumas das modificações consideradas importantes na regulação da expressão gênica serão discutidas a seguir.

5.1 Acetilação e desacetilação

A acetilação das histonas 3 e 4 (H3 e H4) está associada à ativação da transcrição. A acetilação ocorre em múltiplos resíduos de lisina (K) dessas histonas e, em geral, é realizada pelas histona-acetiltransferases (HATs, de *histone acetyltransferases*), que transferem um grupamento acetílico da acetil-coenzima A para o grupamento α -amínico dos resíduos alvos de lisina. As regiões da cromatina que estão superacetiladas costumam apresentar maior atividade de transcrição.

Alguns modelos foram propostos para explicar o efeito da acetilação das histonas na transcrição. Um desses modelos propõe que a neutralização da carga positiva dos resíduos de lisina reduziria a afinidade dos octâmeros de histona pelo DNA carregado negativamente, possibilitando uma maior exposição do DNA aos fatores de transcrição. Essa hipótese é reforçada pela observação de que as histonas acetiladas são removidas do DNA, com facilidade, tanto *in vivo* como *in vitro*. Além disso, a observação de que alguns coativadores transcricionais apresentam atividade intrínseca de HAT reforçam essa proposta. A acetilação da histona pode ser revertida pela ação da enzima histona desacetilase (HDAC, de *histone deacetylase*). De forma semelhante aos coativadores transcricionais com atividade de HAT, alguns repressores da transcrição apresentam subunidades com atividade HDAC.

A hipótese de que a modificação do DNA aumentaria a sua exposição aos fatores de transcrição poderia se aplicar também aos outros tipos de modificação das histonas, com exceção da metilação. Em um outro modelo, os grupamentos acetílico das histonas modificadas poderiam participar da interação com as proteínas reguladoras. Por outro lado, algumas modificações poderiam alterar a organização da cromatina, como é o caso da acetilação da histona H4 na lisina 16 (H4K16), que inibe a formação das fibras compactas de 30 nm. É importante ressaltar que essas possibilidades não são mutuamente exclusivas, como exemplificado no caso da acetilação de H4K16, que controlaria a estrutura da cromatina e também as interações com proteínas.

5.2 Metilação e desmetilação

As histonas podem ser metiladas tanto nos resíduos de lisina como nos de arginina (R). As cadeias laterais da lisina podem ser mono-, di- ou trimetiladas, ao passo que a cadeia lateral de arginina pode ser apenas mono- ou dimetilada. As histona-metiltransferases catalisam a transferência de grupamentos metílicos de *S*-adenosil-*L*-metionina para o grupamento ω - ou δ -NH₂ dos resíduos de arginina, ou para o grupamento ϵ -NH₂ dos resíduos de lisina presentes nas histonas. A metilação das lisinas nas histonas pode contribuir tanto para as funções de ativação como para repressão da cromatina. Por exemplo, a metilação da histona H3 nos resíduos de lisina nas posições 4, 36 e 79 (H3K4, H3K36, H3K79) está associada com regiões ativas da cromatina, ao passo que a metilação em H3 nos resíduos de lisina nas posições 9 e 27 (H3K9, H3K27) e no resíduo de lisina da histona H4 na posição 20 (H4K20), em geral, é encontrada em regiões silenciosas da cromatina. A metilação dos resíduos de lisina não afeta a carga das histonas, indicando que essa modificação não estaria envolvida na ruptura dos contatos DNA-histonas.

A metilação das lisinas nas diferentes posições das histonas é feita por diferentes histona-metiltransferases. Por exemplo, a metilação em H3K4 é realizada por uma histona-metiltransferase que possui um domínio de 140 aminoácidos, denominado SET (de *Su[var]3-9*, *Enhancer-of-zeste*, *Trithorax*); e, a metilação de H3K79 é catalisada por uma proteína sem SET, a Dot1 (de *Disruptor of telomeric silencing*). A metilação de H3K4 (mono- e dimetilada) pode ser revertida por uma lisina-desmetilase, LSD1 (de *Lysine Specific Demethylase 1*), uma amino-oxidase dependente de FAD, que associa-se a vários complexos repressores, podendo impedir a transcrição dos genes-alvo. No caso de H3K4 trimetilada, a desmetilação tem a participação de uma desmetilase que contém o domínio JmJc (Jumonji-C). Por outro lado, a metilação de H3K79 não pode ser revertida. A metilação de arginina, em mamíferos, ocorre na histona H3 nas posições H3R2, H3R8, H3R17 e H3R26 e na histona H4 na posição H4R3. A metilação é catalisada pela metiltransferase da classe PRMT (de *Protein Arginine (R) Methyltransferase*). Essa modificação pode ter tanto efeito de ativação como de repressão no funcionamento da cromatina.

5.3 Ubiquitinação e desubiquitinação

A **ubiquitina** é uma molécula de 76 aminoácidos, e seu nome deriva do fato de ela ser amplamente distribuída nas células eucarióticas, tanto no citosol como no núcleo. Ubiquitinação é o processo pelo qual a ubiquitina se liga covalentemente a uma proteína-alvo, regulando a estabilidade, a função e/ou a localização da proteína modificada. As consequências da modificação dependem da forma como a ubiquitina está ligada a essa proteína, se na forma monomérica ou como cadeias de poliubiquitina. Em geral, quando a proteína está ligada a uma cadeia de poliubiquitina, ela se liga ao **proteassomo**, um grande complexo de várias subunidades com atividade de protease, que de-

grada o substrato em pequenos peptídeos e recicla a ubiquitina. Nas histonas, a ubiquitinação ocorre em H1, H2A, H2B e H3, sendo que essa modificação está envolvida em diferentes aspectos da estrutura da cromatina, afetando a expressão gênica e a estabilidade do genoma. A histona H2B é monoubiquitinada, sendo, dessa forma, resistente à degradação pelo proteassomo. A H2B modificada tem função importante na manutenção da estrutura da cromatina e na regulação da expressão gênica. Essa modificação é reversível, e a ubiquitina pode ser removida pela ação de uma protease ubiquitina-específica. A monoubiquitinação da histona H2A é uma das modificações mais frequentes nas histonas de mamíferos, estimando-se que cerca de 5 a 15% das histonas H2A estejam na forma monoubiquitinada (uH2A). A forma uH2A tem sido associada, em eucariotos superiores, ao silenciamento de genes envolvidos no desenvolvimento, na inativação do cromossomo X, na progressão do ciclo celular e nas etapas de iniciação, alongamento e transcrição. A ubiquitinação de H2A também é dinâmica, com variação dos níveis de ubiquitinação durante o ciclo celular e com o envolvimento de enzimas responsáveis pela desubiquitinação.

6. Remodelamento da cromatina

A estrutura da cromatina pode ser modificada por complexos proteicos, denominados complexos remodeladores da cromatina, que utilizam a hidrólise do ATP para alterar os contatos histona-DNA. O remodelamento da cromatina pode provocar um desenrolamento transitório do DNA e octâmeros de histonas, com a formação de uma alça de DNA, ou causar o deslocamento dos nucleossomos para posições diferentes (**Figura 14.5**). Essas alterações podem, em princípio, facilitar o acesso dos fatores de transcrição ao DNA.

Os remodeladores atuam em conjunto com outros fatores da cromatina, controlando a compactação e a descompactação de regiões do DNA, como promotores, reforçadores e origens de replicação. Essas regiões precisam ser expostas de forma regulada para possibilitar seus funcionamentos corretos em processos como transcrição, replicação, recombinação e reparação do DNA. Estão descritas pelo menos quatro famílias de complexos remodeladores de cromatina, denominadas SWI/SNF, ISWI, CHD e INO80.

Todos os remodeladores compartilham cinco propriedades básicas: afinidade pelo nucleossomo; domínios que reconhecem modificações covalentes nas histonas; uma subunidade catalítica com domínio de ATPase dependente de DNA, requerida para o remodelamento e que serve como um motor para a translocação do remodelador no DNA, rompendo os contatos histona-DNA; domínios e/ou proteínas que regulam o domínio de ATPase; e, domínios e/ou proteínas para a interação com outros fatores da cromatina ou de transcrição. Embora todos os remodeladores apresentem subunidades catalíticas, com

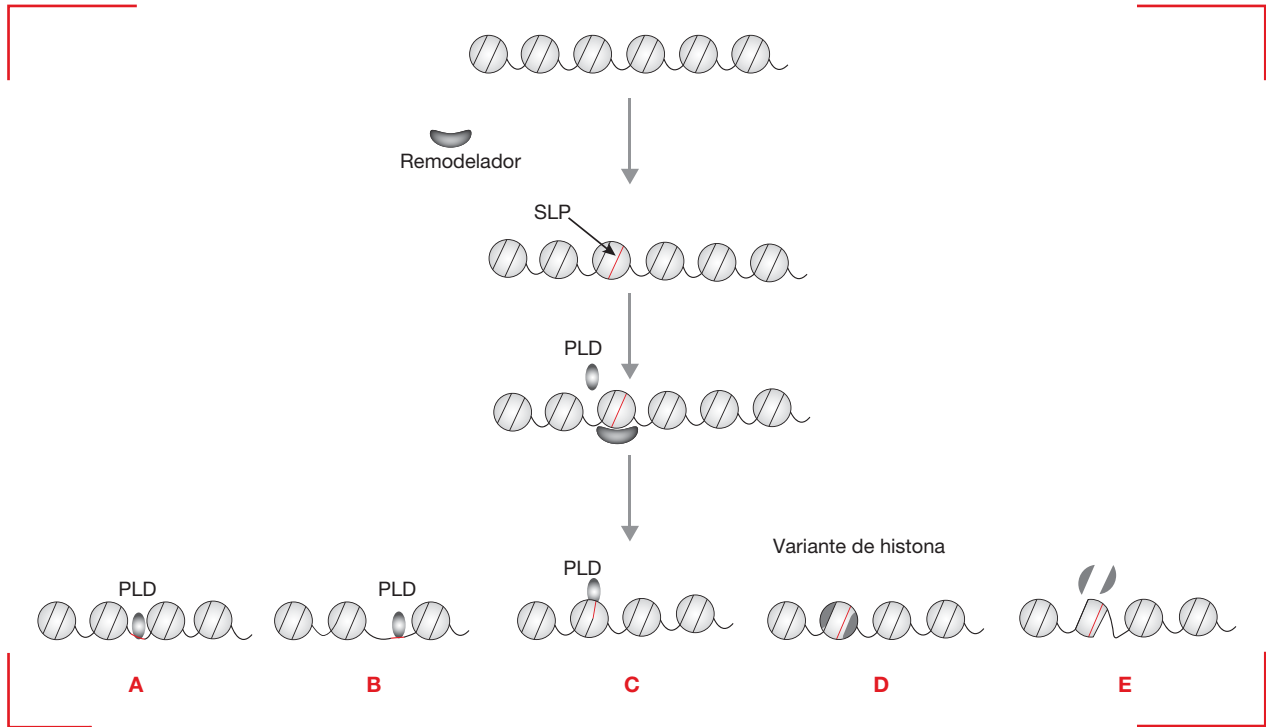


Figura 14.5

Representação esquemática dos possíveis efeitos da ação de remodeladores de cromatina. O remodelador (indicado na figura) pode auxiliar na alteração da estrutura da cromatina de diferentes formas. **(A)** Exposição do sítio de ligação pelo reposicionamento do nucleossomo. **(B)** Remoção do nucleossomo, expondo o sítio de ligação. **(C)** Desenrolamento localizado do DNA no nucleossomo, expondo o sítio de ligação. **(D)** Alteração na composição do nucleossomo, com troca do dímero de H2A-H2B por dímero alternativo contendo uma histona variante. **(E)** Alteração na composição do nucleossomo, com remoção do dímero. SLP – sítio de ligação à proteína; PLD – proteína de ligação ao DNA.

um domínio de ATPase comum, cada membro apresenta domínios flanqueadores diferentes, permitindo a sua classificação nas quatro famílias distintas anteriormente citadas. A **Figura 14.6** mostra as subunidades catalíticas com os domínios de ATPase das diferentes famílias de remodeladores. As diferentes famílias de remodeladores formam complexos que podem conter até mais de 10 subunidades.

7. Controle da transcrição: o exemplo dos receptores nucleares de hormônios

Hormônios lipofílicos, como os hormônios esteroides, o ácido retinoico e os hormônios da tireoide, são capazes de difundir pelas membranas plasmática e nuclear, interagindo com receptores hormonais, denominados receptores nucleares. Esses receptores nucleares são fatores de transcrição que participam da ação de hormônios durante o desenvolvimento, o metabolismo e a homeostasia. O complexo hormônio-receptor nuclear liga-se às sequências reguladoras de determinados genes, controlando a sua expressão.

A estrutura desses receptores apresenta uma região N-terminal variável, contendo um domínio de ativação da transcrição (FA-1); uma região central conservada (DLD), que possibilita a ligação ao DNA; e uma porção C-terminal (DLH), que se liga ao hormônio e contém um segundo domínio de ativação de transcrição (FA-2). FA-2 é importante para recrutar várias proteínas coativadoras, que interagem com proteínas envolvidas no remodelamento da cromatina e na maquinaria geral de ativação transcripcional. DLH também participa nas reações de interação dos receptores (dimerização ou tetramerização) necessárias para a ligação aos elementos de resposta do DNA. O domínio DLD é fundamental para a ligação aos elementos de resposta presentes nos genes regulados pelos receptores nucleares. A estrutura do DLD do receptor de glicocorticoide é uma das melhores estudadas e tem servido de modelo representativo da família dos receptores nucleares. O DLD apresenta dois dedos de zinco, sendo que cada átomo de zinco é coordenado por quatro resíduos de cisteína, além de duas α -hélices, com uma das hélices interagindo com o DNA. A região N-terminal dos receptores nucleares inclui a região com função de ativação FA-1 e é a que apresenta menor conservação na sequência de aminoácidos. A **Figura 14.7A** mostra a estrutura de um receptor nuclear.

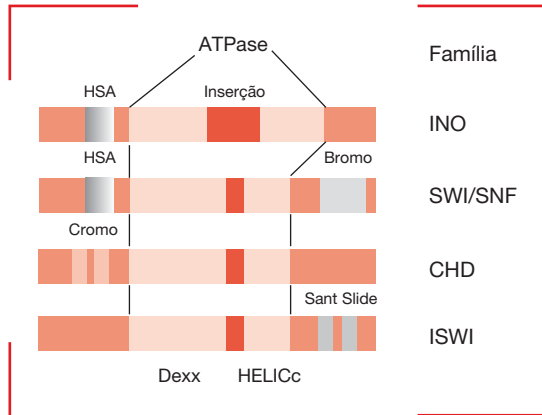


Figura 14.6

Esquema das subunidades catalíticas com os domínios de ATPase de diferentes famílias de remodeladores de cromatina. As diferentes famílias apresentam uma subunidade de ATPase, caracterizada por um domínio de ATPase e dividida nas regiões DExx e HELICc por uma inserção. A família INO apresenta uma inserção longa. As diferenças de cada família são determinadas pelos domínios específicos de cada uma que se apresentam adjacentes ao domínio ATPase. A família INO tem um domínio HSA (Helicase-SANT); a família SWI/SNF apresenta HAS e o bromodomínio (Bromo); a família CHD apresenta os cromodomínios (Cromo) em tandem; e a família ISWI possui o módulo com os domínios SANT-SLIDE. HELICc – domínio C-terminal da superfamília das helicases; DExx – motivo presente em helicase; SANT-SLIDE – domínios encontrados nos remodeladores da família ISWI com capacidade de ligar às histonas. A sequência do domínio localizado na posição mais próxima do C-terminal é bastante diferente da sequência SANT canônica e, por isso, o domínio foi denominado SLIDE; Bromodomínio. Domínio com cerca de 110 aminoácidos capaz de se ligar a resíduos acetilados de lisina. Esse domínio é geralmente encontrado em proteínas que regulam a estrutura da cromatina e a expressão gênica. Cromodomínio – o termo é derivado de “Chromatin Organization Modifier” (Chromo). O domínio é constituído de 3070 aminoácidos e é encontrado em muitas proteínas envolvidas na montagem de complexos proteicos na cromatina.

A purificação de alguns receptores e a clonagem dos genes que codificam esses receptores possibilitaram a análise comparativa das sequências deduzidas de aminoácidos correspondentes. Esse tipo de análise mostrou que essas proteínas apresentam uma conservação bastante grande, tanto na sequência de aminoácidos como na organização dos seus domínios funcionais. A Figura 14.7B mostra um esquema comparativo de diferentes receptores hormonais.

A análise da estrutura dos genes que respondem à ação hormonal mostra a existência de sequências reconhecidas pelo complexo hormônio-receptor. A **Tabela 14.1** mostra as sequências de consenso de alguns elementos que respondem à ação de hormônios.

7.1 Mecanismos de ação dos receptores de hormônios esteroides

No citoplasma, o receptor recém-sintetizado ou sem o hormônio associa-se às **proteínas chaperonas**, sendo a Hsp90 uma das mais importantes do complexo. Estudos *in vitro* mostraram que Hsp40, Hsp70, Hsp90, Hop e p23 constituem-se nos componentes mínimos para a formação de complexos com receptores de glicocorticoides, possuindo capacidade de ligação do hormônio. Na formação do complexo do receptor com as chaperonas, a ligação da Hsp90 só ocorre após outras chaperonas estabelecerem interações com o receptor. A proteína p23, uma cochaperona de Hsp90, atua estabilizando a conformação de Hsp90 e aumentando a capacidade do receptor de se ligar ao hormônio.

Algumas proteínas cochaperonas, como FKBP52, FKBP51, CyP40 e PP5, apresentam um **domínio TPR** (de *tetratricopeptide repeat*), que possibilita a interação dessas proteínas com um sítio de ligação a TPR presente na porção C-terminal de Hsp90. Cada tipo diferente de receptor de esteroide apresenta uma preferência na interação com determinado tipo de cochaperona com domínio TPR. Essas cochaperonas parecem apresentar funções fisiológicas importantes na resposta celular ao hormônio. Por muitos anos acreditava-se que, após a interação com o hormônio, o receptor dissociava-se das outras proteínas, podendo então, atuar de forma independente das chaperonas. Entretanto, estudos mais recentes têm mostrado que a atividade do receptor na ativação gênica depende da atividade da chaperona.

Outros estudos também mostraram que a translocação do receptor do citoplasma para o núcleo envolve a participação de Hsp90 e outras chaperonas ou cochaperonas (**Figura 14.8**).

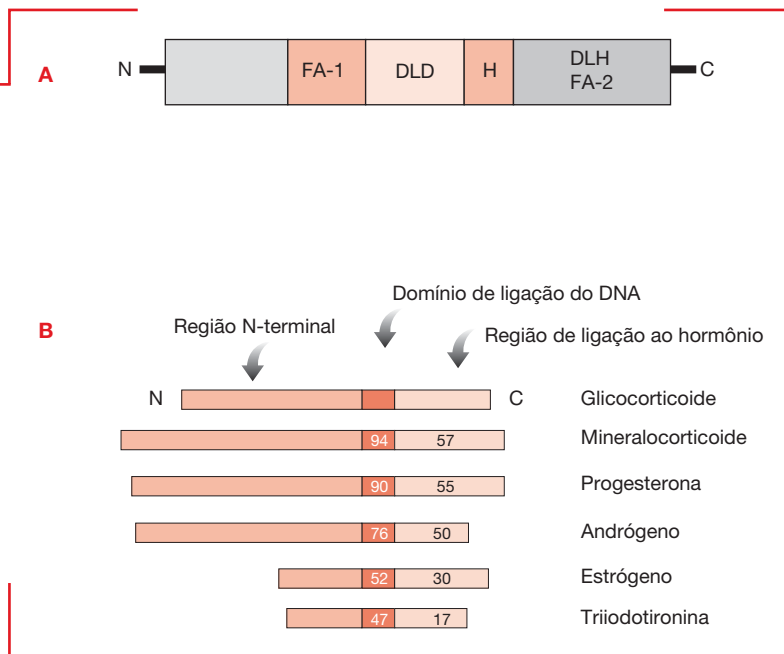
8. Regulação de expressão de locus multigênicos: os genes de globina como modelo

Em eucariotos, muitos genes estão organizados em locus multigênicos, em que são coordenadamente regulados durante a diferenciação e o desenvolvimento. Estudos realizados em diferentes locus gênicos têm mostrado que a regulação da expressão pode envolver a participação de elementos reguladores localizados em regiões bastante distantes dos genes que eles controlam.

A proposição de um novo elemento de controle, diferente daqueles encontrados próximos do sítio de início

Figura 14.7

Estrutura de receptor e similaridade das seqüências dos diferentes domínios dos receptores. (A) Representação esquemática da estrutura de receptor. FA, função de ativação; DLD, domínio de ligação ao DNA; H, região curta que faz a ligação de DLD com DLH; DLH, domínio de ligação do ligante. N, porção N-terminal; C, porção C-terminal. **(B)** Representação esquemática de receptores de hormônios esteroides e de tireoide. As porcentagens de homologia dos domínios de ligação ao DNA e das regiões de ligação ao hormônio dos diferentes receptores estão indicadas. Na região N-terminal, as homologias não são maiores que 15%. N, porção N-terminal; C, porção C-terminal.



de transcrição, a **região de controle de locus (LCR, de locus control region)**, baseou-se principalmente em três observações: (1) as seqüências reguladoras presentes em promotores e reforçadores não eram suficientes para promover a regulação; (2) as descobertas de que deleções de regiões a montante do agrupamento de genes de globina, localizadas a uma distância razoável do gene de β -globina, inativavam a expressão de todo o locus; e (3) a presença de sítios hipersensíveis à DNase em uma região localizada a mais de 20 kb a montante do gene β -globina em eritroblastos. Deleções nessa região resultam em β -talassemias severas, mesmo em pacientes que retêm uma cópia intacta do gene de β -globina.

Existem cinco genes de globina (ϵ , γ^A , γ^G , e β) em seres humanos, sendo que o primeiro gene a ser expresso é o ϵ (embrião), passando para γ^A e γ^G (feto) e, finalmente, para δ e β (adulto). A expressão desses genes é regulada não apenas pelos seus promotores, mas também pelo LCR, contendo sítios hipersensíveis à DNase I. Esses sítios hipersensíveis podem ser identificados *in vitro* pelo tratamento da cromatina com DNase (Capítulo 3). Quando proteínas reguladoras se ligam a uma seqüência reguladora específica, elas deslocam um nucleossomo, expondo o DNA ao ataque da nuclease. Essas proteínas ligam-se ao DNA um pouco antes do gene ser ativado. Nos ensaios

de análise dessas regiões, a nuclease atua nesses sítios, degradando de forma seletiva as seqüências de DNA que estão sendo ativamente transcritas. A **Figura 14.9** mostra um esquema da organização do locus de β -globina e o padrão da expressão dos genes de globina durante o desenvolvimento em seres humanos.

A demonstração da importância da LCR em promover altos níveis de expressão de globina baseou-se nos estudos com camundongos transgênicos para o locus de β -globina. A LCR foi definida como um elemento capaz de controlar a expressão de um transgene, independentemente do seu sítio de integração no genoma. Os estudos realizados em camundongos transgênicos contendo genes humanos com as mais variadas construções, com os diferentes genes do locus de globina e combinações de LCR, promotor e reforçador, forneceram algumas informações sobre como ocorreria a interação entre a LCR e os promotores dos genes individuais, por meio de fatores transcricionais. Os estudos realizados com os genes de globina humanos mostraram que a LCR participa na elevação do nível de transcrição de um gene de cada vez, havendo competição entre os genes para ativação pela LCR. Isso levou à hipótese de que haveria interação física entre a LCR e os genes. Essa hipótese foi comprovada com estudos sobre a organização espacial do locus de β -globina

Tabela 14.1 Sequências de consenso de elementos que respondem à ação de hormônios

Hormônio	Elemento de resposta	Seqüência consenso
Glicocorticoide (andrógeno, mineralocorticoide)	GRE	AGAACANNNTGTTCT
Estrógeno	ERE	AGGTCANNNTGACCT
Hormônio de tireoide, ácido retinoico	TRE	AGGTCA --- TGACCT

Obs.: Os nucleotídeos em negrito representam as seqüências palindrômicas (repetidas e invertidas) de cada elemento. N significa qualquer nucleotídeo. TRE é praticamente idêntica à ERE, com a única diferença de que as repetições invertidas não são separadas por três nucleotídeos.

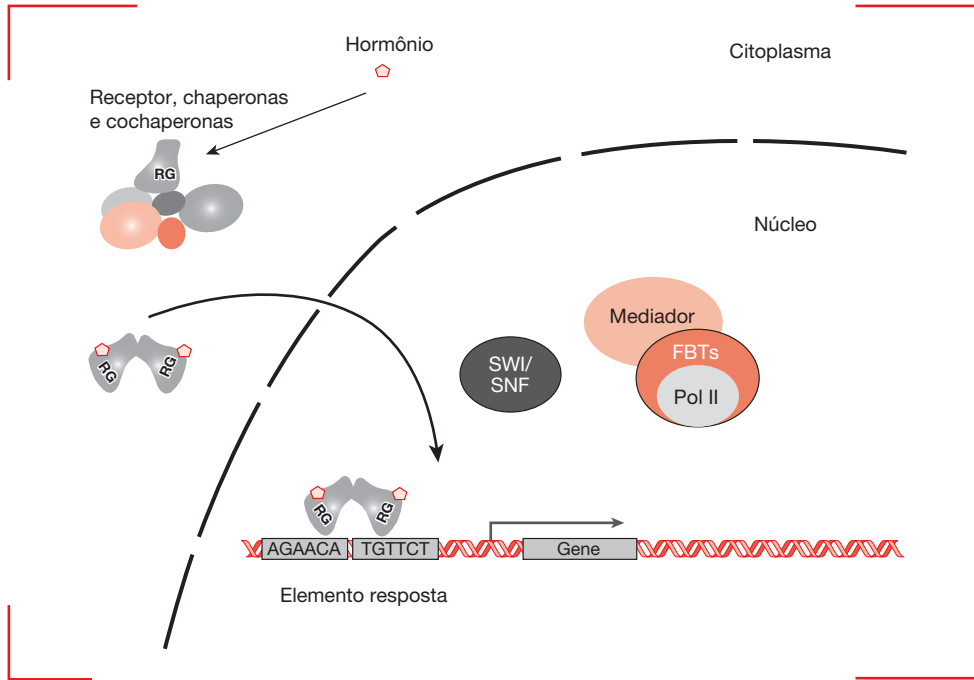


Figura 14.8

Representação esquemática da ativação de um gene pelo complexo receptor-hormônio glicocorticoide. Quando não há hormônio, o receptor está ligado a um complexo proteico que contém Hsp90. Na presença de hormônio, ocorre a sua ligação ao domínio DLH (ver Figura 14.7), causando uma mudança conformacional que libera o receptor do complexo com Hsp90. O complexo RG-hormônio passa para o núcleo, onde o DLD (ver Figura 14.7) do receptor reconhece uma sequência reguladora de um gene, estimulando a transcrição. Em geral, a interação do receptor com o elemento de resposta envolve eventos de remodelamento da cromatina realizado pelo complexo SWI/SNF ou outros sistemas de remodelamento. RG, receptor de glicocorticoide; FBTs, fatores basais de transcrição; Pol II, RNA-polimerase II; SWI/SNF, remodelador de cromatina (ver Figura 14.6).

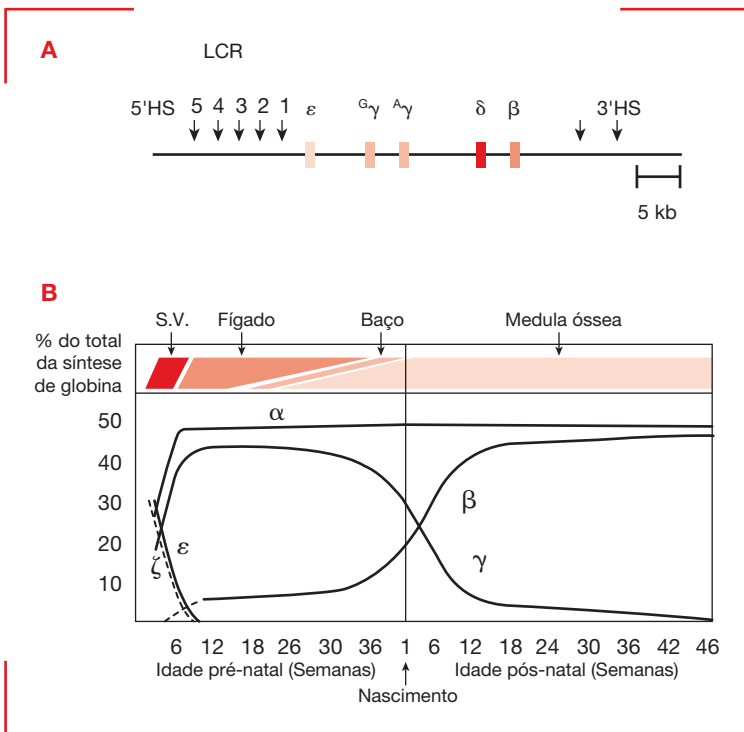


Figura 14.9

Estrutura do locus de β-globina de seres humanos e atividade dos genes durante o desenvolvimento. (A) Organização dos genes de β-globina. O sítios hipersensíveis à DNase (1 a 5) estão indicados. Os genes de globina estão representados por caixas retangulares. (B) Produção das globinas durante o desenvolvimento. S.V. saco vitelino. Os níveis de produção de cada globina durante o desenvolvimento estão indicados. LCR, Locus control region.

e a demonstração de que os genes de globina ativados estavam próximos da LCR. Nas células que expressam globina, há a formação de um *hub* de cromatina ativa (*active chromatin hub*, ACH), (alças formadas na cromatina que possibilitam a interação da LCR com os genes de globina). Muitos fatores, alguns expressos em todas as células (ubíquos) e outros expressos somente em eritrócitos, participam na ativação da expressão dos genes de globina e na formação da estrutura em alça da cromatina no locus.

As LCRs foram posteriormente descritas em vários outros sistemas, como, por exemplo, nos genes que codificam as citocinas IL4 e IL13, CD2 adenosina-desaminase, no agrupamento dos genes de hormônio de crescimento e somatomotropina coriônica em humanos, no gene de lisozima de galinha e no gene da tirosinase de camundongo.

9. RNAs não codificadores no controle da transcrição

Como já analisado, proteínas têm sido identificadas como sendo importantes na regulação da maquinaria de transcrição. Entretanto, essa regulação tem também a participação importante de **RNAs não codificadores** (ncRNA, de *noncoding RNA*). Os ncRNAs são assim chamados porque não apresentam ORFs que codificam proteínas. Recentemente, descobriu-se que mais de 90% do genoma humano pode ser transcrito, gerando RNAs a partir das duas fitas de DNA. Os efeitos reguladores dos ncRNAs podem se manifestar de várias maneiras, uma das mais conhecidas é o efeito que a própria transcrição do ncRNA pode ter sobre os genes vizinhos.

A transcrição de um ncRNA na região promotora de um gene que codifica proteína, localizado mais a jusante, pode impedir que fatores de transcrição se liguem às sequências reguladoras do gene, interferindo, assim, com a transcrição daquele gene. Por outro lado, a transcrição do ncRNA pode atuar na ativação de determinados genes. Esse é o caso, por exemplo, do efeito que a transcrição contínua de ncRNA pode ter, impedindo a ação de proteínas do grupo *Polycomb* (PcG, de *polycomb group*) no silenciamento de determinados genes *Hox*. Os genes do grupo *Polycomb*, descritos inicialmente em *Drosophila*, codificam um conjunto de proteínas que atuam como repressores transcricionais essenciais no desenvolvimento de animais e plantas. Além disso, já foi demonstrado, em *Schizosaccharomyces pombe*, que a transcrição de ncRNAs longos pode induzir alterações na cromatina, importantes para estimular a expressão de genes codificadores de proteínas localizados a jusante. Entretanto, o próprio recrutamento das proteínas PcG para agir no silenciamento dos genes, que pode chegar a mais de mil nos mamíferos, pode ter a participação dos ncRNAs. A ação das proteínas do grupo Trithorax (TrxG, de *Trithorax group*), que atuam na neutralização da atividade da proteínas PcG, mantendo o estado ativo dos genes, parece envolver, também, a participação de ncRNAs longos no recrutamento para os locus-alvo.

10. Regulação pós-transcricional

Uma vez iniciada, a transcrição deve proceder até o sinal de término da transcrição do gene, possibilitando a formação de um produto primário de transcrição. Como estudado no Capítulo 11, esse produto primário deve ser modificado pela adição de CAP na extremidade 5', cauda de poli(A) na extremidade 3' (poliadenilação) e também pela remoção dos íntrons e união dos éxons (*splicing*). Esses processos ocorrem durante a própria transcrição e a terminação está intimamente associada à quebra na extremidade 3' no sítio de poliadenilação. Muitos fatores proteicos associados a esses processos e também fatores que atuam no transporte do RNA têm sido encontrados associados a complexos de RNA-polimerase II, sugerindo que existe uma integração funcional entre as maquinarias envolvidas na transcrição, processamento e transporte, permitindo uma expressão gênica eficiente e regulada. Dessa forma, além da regulação no início da transcrição, a expressão gênica pode ser regulada em nível pós-transcricional. A regulação pós-transcricional pode ocorrer em etapas que envolvem, por exemplo, a adição de 5'-CAP, adição de cauda de poli(A) na extremidade 3', o *splicing* e o transporte do RNA do núcleo para o citoplasma.

Estudos recentes demonstraram que o conjunto de RNAs presente nas células apresenta uma complexidade muito maior do que se imaginava. Além dos RNA ribossômicos (rRNAs), RNAs transportadores (tRNAs), RNAs nucleares de baixo peso molecular (snRNAs) e RNAs mensageiros (mRNAs), as células apresentam os microRNAs (miRNAs) e os RNAs não codificadores (ncRNAs). Como vimos, os ncRNAs podem ter função importante na regulação da transcrição, e os miRNAs têm funções importantes na regulação da expressão gênica em nível pós-transcricional.

Ainda devemos considerar que o processamento alternativo é um mecanismo importante para aumentar a capacidade codificadora das células na produção do seu complemento de proteínas.

Mesmo após terem sido produzidos, os mRNAs estão sujeitos a diferentes mecanismos de regulação que podem envolver a sua localização no citoplasma, estabilidade, *turnover* do mRNA, regulação da tradução por miRNAs, fatores reguladores e a utilização de sítios alternativos de tradução.

10.1 Pequenos RNAs na regulação da expressão gênica

Os RNAs pequenos (20-30 nucleotídeos) podem atuar na regulação da expressão gênica em diferentes níveis, como estrutura da cromatina, transcrição, processamento, estabilidade e tradução do mRNA. Considerando as características relacionadas com a forma como eles são gerados, suas estruturas, as proteínas com as quais se associam e suas funções biológicas, os RNAs pequenos podem ser classificados em três classes principais: RNAs de interferência

curtos (siRNAs, de *short interfering RNAs*), microRNAs (miRNAs) e RNAs que interagem com Piwi (piRNAs).

As formas de ação desses pequenos RNAs levaram ao rápido desenvolvimento de ferramentas para o estudo das funções gênicas, bem como a sua regulação a partir do nocaute gênico pelo silenciamento. Além disso, esses pequenos RNAs estão sendo estudados nos mais diferentes tipos de enfermidades e vêm sendo propostos como uma alternativa para o tratamento de determinadas enfermidades, por exemplo, diferentes tipos de câncer.

10.1.1 miRNAs

Os miRNAs participam da regulação pós-transcricional de uma grande diversidade de mRNAs, atuando de forma importante em muitos processos biológicos, como crescimento celular, divisão e diferenciação celular, metabolismo e desenvolvimento.

Os miRNAs são gerados a partir de estruturas em grampo presentes nos RNAs, pela ação de duas proteínas do tipo RNase III, *Drosha* e *Dicer*. Os miRNAs (cerca de 22 nt) formados associam-se a proteínas Ago da subfamília *Argonauta* e atuam como reguladores pós-transcricionais, tendo os mRNAs como alvos. Os miRNAs são encontrados em praticamente todos os organismos eucarióticos, desde seres humanos até organismos eucarióticos mais simples, como os cnidários. No banco de dados miRBase* podem ser encontradas

* Disponível em: <http://www.mirbase.org>.

sequências anotadas de miRNA dos mais diversos organismos. Existem cerca de 500-1.000 genes de miRNA nos genomas de vertebrados e plantas; e cerca de 100 nos genomas de invertebrados, sendo que cada miRNA pode atuar em sítios-alvo de centenas de mRNAs. Essas características sugerem que os miRNAs possam atuar de forma importante na regulação de uma grande parte do transcriptoma.

Os miRNAs são transcritos pela RNA-polimerase II e os produtos primários, que podem ter vários kb, são processados ainda no núcleo da célula pela proteína *Drosha*. Após o processamento, o precursor do miRNA (pré-miRNA) é transportado para o citoplasma, onde é clivado por *Dicer* para originar o miRNA de fita dupla de 22 nt. Depois disso, esse RNA de fita dupla se associa à proteína Ago, formando o complexo de silenciamento induzido por miRNA (miRISC, de *RNA-induced silencing complex*). Uma das fitas do miRNA permanece associada a Ago, como um miRNA maduro, ao passo que a outra fita é degradada (**Figura 14.10**).

Além de proteínas da subfamília *Argonauta*, outros fatores são necessários para a expressão e função do miRNA, sendo que alguns são específicos da espécie, do mRNA-alvo ou do próprio miRNA. Outros, como a proteína GW182, que contém repetições de resíduos de glicina (G) e de triptofano (W), podem ter uma função mais ampla na ação dos miRNAs e são conservados em diferentes espécies. Em animais, os miRNAs reconhecem os seus alvos por um pareamento imperfeito das bases da

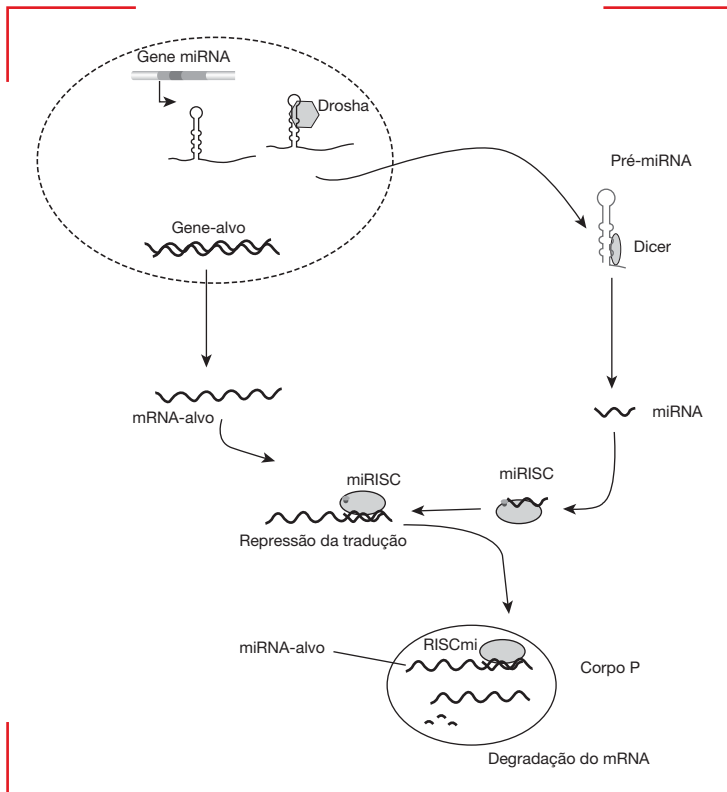


Figura 14.10

Produção de miRNA e potenciais mecanismos de ação sobre os mRNAs-alvo. O precursor do miRNA (pré-miRNA) é transcrito pela RNA-pol II no núcleo, processado pela enzima *Drosha* e exportado para o citoplasma, onde continua a ser processado por *Dicer*. Depois da separação das fitas, o miRNA maduro passa a fazer parte de um complexo (RISC, de *RNA-induced silencing complex*), juntamente com a proteína Argonauta (Ago). O complexo atua na repressão da produção de proteínas bloqueando a tradução ou causando a degradação do mRNA-alvo nos corpos P.

sua extremidade 5' com bases complementares na região 3' não traduzida (3'-UTR) dos mRNAs-alvos. Em geral, os miRNAs inibem a síntese de proteínas, reprimindo a tradução ou promovendo a remoção da cauda de poli(A) (desadenilação) e degradação dos mRNAs-alvo.

Os mecanismos de repressão da tradução pelos miRNAs ainda não estão elucidados em sua totalidade. Existem evidências de que a repressão ocorreria em nível de iniciação da tradução, pela competição entre RISC e eIF4E para interação com CAP na extremidade 5' do mRNA, ou em uma etapa após o reconhecimento de 5'-CAP, pela inibição da associação das subunidades ribossômicas menor e maior. Entretanto, também existem evidências de que a repressão poderia ocorrer após a iniciação da tradução, por exemplo, pelo retardo do alongamento da cadeia polipeptídica pelos ribossomos, acoplado à terminação prematura; ou pela indução de degradação dos polipeptídeos nascentes durante o processo de tradução. Uma possível explicação para as discrepâncias encontradas poderia ser o fato de os miRNAs regularem a tradução por múltiplos mecanismos e a utilização de sistemas biológicos e metodologias distintos poder refletir diferenças reais nos resultados.

Uma outra forma de atuação dos miRNAs sobre os mRNAs-alvo é a indução da desadenilação e aceleração da degradação dos mRNAs. Quando a via do miRNA é inibida ou quando os níveis de miRNA são aumentados em animais ou células em cultivo, há alteração na abundância dos mRNAs-alvo. Os miRNAs aceleram a degradação dos mRNAs-alvo pela remoção da cauda de poli(A)

da extremidade 3'. A desadenilação leva à perda da PABP [*poly(A)-binding protein*, proteína que se liga a poli(A)], levando à remoção de 5'-CAP e expondo a extremidade 5' do mRNA para degradação exonucleolítica.

No citoplasma, existem domínios chamados de **corpos P**, constituídos por várias proteínas que atuam na repressão da tradução, no controle de qualidade, no silenciamento e na degradação dos mRNAs, como as enzimas envolvidas na remoção de 5'-CAP (Dcp1/Dcp2), exonuclease 5'→3' Xrn1 e outras proteínas que participam ou regulam a via de degradação do mRNA. Componentes que atuam na repressão e degradação de mRNAs mediados por miRNA, como proteínas Ago e GW182, miRNAs e os mRNAs-alvo inibidos são encontrados nos corpos P. Junto com a proteína Ago, a proteína GW182 tem uma participação importante no silenciamento mediado por miRNAs. A **Figura 14.11** mostra o esquema de estruturas das proteínas Ago e GW182.

A repetição GW localizada na porção N-terminal da proteína GW182 participa da interação com proteínas Ago. Além disso, junto a uma região rica em glutamina (Q), essa região GW contribui para a localização da proteína GW182 nos corpos P. Entretanto, a localização do mRNA nos corpos P parece não ser essencial para o silenciamento, mas seria uma consequência do processo. Embora no caso da proteína GW182 de *D. melanogaster* existam resultados contraditórios em relação à possível função das porções central e C-terminal no silenciamento, no caso das proteínas GW humanas, a porção C-terminal é suficiente para promover o silenciamento.

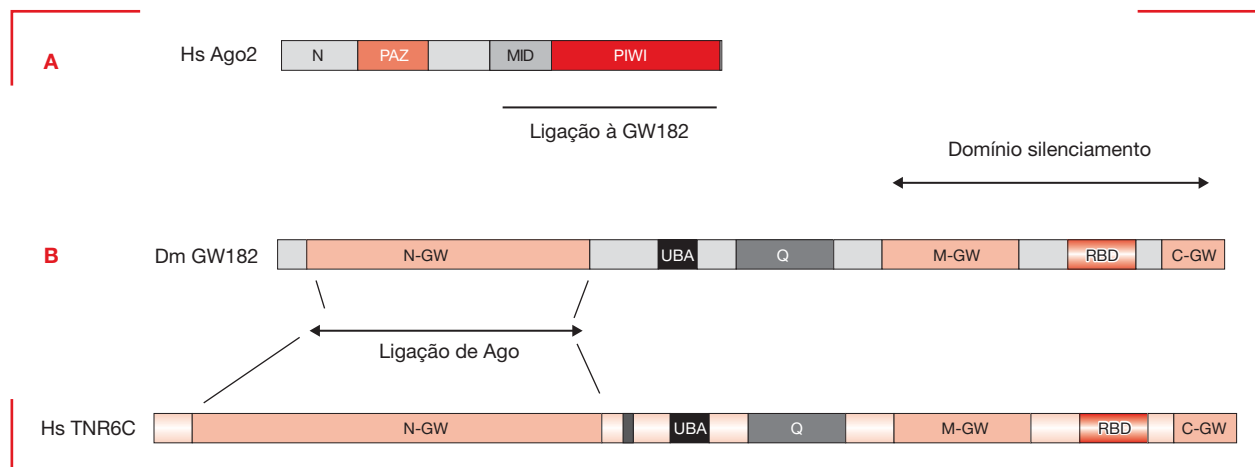


Figura 14.11

Os Representação esquemática da organização dos domínios das proteínas Argonauta (Ago2) e GW182.

(A) Representação esquemática da proteína Argonauta 2 humana. Os quatro domínios característicos de Ago são mostrados: domínio terminal N; PAZ, MID e PIWI. O domínio PAZ serve com sítio de ligação da extremidade 3' dos miRNAs, ao passo que o domínio MID liga-se à extremidade 5' dos miRNAs. O domínio PIWI apresenta um dobramento semelhante ao da RNase H, e, no caso da Ago2, esse domínio está envolvido na clivagem do RNA alvo ligado ao pequeno RNA, uma atividade conhecida como *slicer*. (B) Representação esquemática das proteínas GW182 de *D. melanogaster* (DM GW182) e uma das proteínas de humanos (Hs TNR6C). N-GW, M-GW e C-GW são as regiões contendo dipeptídeos de glicina (G)-triptofano (W). UBA, domínio associado à ubiquitina. Q, região rica em glutamina. RBD, domínio de ligação a RNA. O domínio NGW está envolvido na mediação da interação das proteínas GW182 com o domínio PIWI das proteínas Ago. A região que se estende de M-GW a C-GW está envolvida no silenciamento (indicada na figura).

10.1.2 siRNAs

Os siRNAs foram inicialmente descritos como moléculas de RNA de fita dupla de origem exógena, capazes de silenciar genes por um processo que foi denominado **interferência por RNA** (iRNA, de *interference RNA*). Por algum tempo, considerou-se que os siRNAs seriam originados a partir de vírus, transposons ou transgenes. Posteriormente, mostrou-se, em diversos organismos, que os siRNAs também poderiam ser codificados pelo próprio genoma, ou seja, eram endógenos.

Os siRNAs endógenos (cerca de 21 nt) são semelhantes aos miRNAs, em relação ao tipo de proteína que se associam (Ago) e na sua ação como reguladores pós-transcricionais. Entretanto, eles diferem dos miRNAs pela forma como são produzidos, a partir de RNAs de fita dupla (dsRNA) longos pela ação de *Dicer* e, também, pela sua forma de atuação sobre os elementos transponíveis. Os dsRNAs são convertidos em siRNAs de 21-23 nt pela ação da enzima *Dicer*. A fita antissenso é chamada de **guia**, pois é ela que fará o pareamento com o mRNA-alvo para possibilitar o silenciamento dele; a fita complementar é chamada de **passageira**. Os dsRNAs são incorporados no complexo de silenciamento (RISC) que contém a proteína Ago2, *Dicer* e outras proteínas. O complexo ativado (siRISC) contém a fita guia que possibilitará o pareamento com o mRNA-alvo para ocorrer a clivagem pela proteína Ago. Após a clivagem, o mRNA é degradado, e o complexo RISC é reciclado para a clivagem de novas moléculas do mRNA-alvo.

10.1.3 piRNAs

Os piRNAs (24-30 nt) foram descobertos em estudos dos pequenos RNAs durante o desenvolvimento de *D. melanogaster*. Esses RNAs ligam-se a proteínas Argonauta da subfamília Piwi, sendo abundantes nas células germinativas. Os piRNAs atuam no silenciamento de transposons, protegendo o genoma contra eles. Em *Drosophila*, os piRNAs são produzidos a partir de elementos repetitivos intergênicos dos cromossomos.

10.2 Turnover de mRNA

Como analisado no Capítulo 10, em eucariotos, os mRNAs são produzidos com duas modificações importantes, o CAP na extremidade 5' e a cauda de poli(A) na extremidade 3', que são introduzidas durante a transcrição. Essas estruturas interagem com as proteínas citoplasmáticas eIF4E e PABP, protegendo o RNA contra a ação de exonucleases e promovendo a iniciação da tradução. Para que ocorra a degradação do mRNA, duas possíveis vias são utilizadas: uma que envolve a remoção de 5'-CAP, para que uma exonuclease com ação no sentido 5'→3' degrade o mRNA; e outra que envolve o ataque da ponta 3' desprotegida [sem a cauda de poli (A)] por um complexo de exonucleases, atuando no sentido 3'→5', denominado **exossomo**. A remoção da cauda de poli (A) envolve a participação de desadenilases e, posteriormente, o mRNA pode ser degradado pela via do exossomo

(complexo de 10-12 proteínas, em que pelo menos seis delas apresentam atividade exonucleolítica), ou pela via que envolve a remoção de 5'-CAP, com degradação pela ação de exonuclease, atuando no sentido 5'→3'.

A velocidade de decaimento dos mRNAs é regulada e varia entre os diferentes produtos de transcrição. Em geral, mRNAs que codificam fatores de transcrição, proteínas da maquinaria transcricional e da síntese proteica apresentam meias-vidas curtas, ao passo que mRNAs que codificam proteínas envolvidas no metabolismo central apresentam meias-vidas mais longas. Dependendo das condições em que se encontram as células ou organismos, ocorrem alterações no padrão da expressão dos genes, sendo que, em muitos casos, para cerca de 50% dos genes essa alteração ocorre no decaimento dos seus produtos de transcrição. Esse decaimento é modulado pela variação na composição das proteínas que interagem com os mRNAs (mRNPs), acelerando ou diminuindo a velocidade de decaimento.

Os mRNAs apresentam sequências que estão relacionadas à sua estabilidade, estando localizadas, principalmente, na região 3' não traduzida. Uma das sequências mais estudadas é o elemento rico em AU (ARE, *AU-rich element*), encontrado na região 3' não traduzida de muitos mRNAs que codificam citocinas, proto-oncogenes e fatores de transcrição. Esses elementos ou proteínas associadas a eles modulam a estabilidade dos mRNAs, recrutando proteínas envolvidas em seu processo de decaimento. Não se sabe como determinadas proteínas que estabilizam os mRNAs funcionam, mas acredita-se que elas poderiam atuar removendo os mRNAs dos sítios em que ocorre o decaimento ou competindo com proteínas que provocam a instabilidade dos mRNAs. Uma outra alternativa poderia ser a interação direta das proteínas estabilizadoras com a maquinaria envolvida no decaimento, inibindo a sua ação.

11. Regulação da expressão no nível da tradução

A síntese de proteínas (tradução) representa uma etapa importante na qual a expressão gênica pode ser regulada. Considerando a complexidade na iniciação desse processo, com a participação de inúmeros fatores, essa etapa ofereceria uma maior flexibilidade e diferentes fatores poderiam ser modificados, alterando-se a iniciação do processo. Por outro lado, a síntese de proteínas ocorre de forma localizada em determinadas células, como nos neurônios, ou durante o desenvolvimento. Além disso, a regulação em nível de tradução poderia servir com um mecanismo mais refinado, após a regulação transcricional.

Um dos exemplos mais bem estudados de regulação em nível de síntese de proteínas é aquele envolvendo o controle de tradução dos mRNAs que codificam as subunidades da ferritina e do mRNA do receptor de transferina em vertebrados. Nesses organismos, existe uma rede

de proteínas que promove o transporte, o uso e o armazenamento de ferro. A transferrina (Tf) transporta ferro entre os tecidos e liga-se a receptores de transferrina (TfR) na superfície celular. O complexo transferrina diférrica-TfR é internalizado, resultando na liberação de ferro no citoplasma, na reciclagem da apo-Tf (transferrina sem ferro) e TfR na superfície celular. O ferro liberado no citoplasma é utilizado para a síntese de várias proteínas que contêm ferro, ou armazenado ligado à proteína ferritina.

A modulação coordenada, dependente de ferro, da síntese de TfR e ferritina possui um papel fundamental na homeostasia do ferro. Os mRNAs de ferritina e de TfR apresentam, nas regiões não traduzidas 5' e 3', respectivamente, estruturas do tipo haste-alça, chamadas de **elementos de resposta ao ferro** (IREs, de *iron responsive elements*), reconhecidas por proteínas citosólicas que se ligam ao RNA, denominadas proteínas **reguladoras do ferro** (IRPs, de *iron regulatory proteins*). Quando ligadas ao IRE do mRNA de ferritina, as IRPs bloqueiam a tradução do mRNA, reduzindo a síntese de ferritina. Em contraste, a interação de IRP com os múltiplos IREs no mRNA de TfR estabiliza o mRNA contra a degradação.

Duas IRPs relacionadas foram identificadas (IRP1 e IRP2), essas duas proteínas estão ativas quando há pouco

ferro na célula. Observou-se que IRP1 apresenta grande similaridade com as aconitases mitocondrial e bacteriana, enzimas contendo ferro-enxofre, envolvidas na inter-conversão de citrato e isocitrato. A partir dessa observação teve-se a base para a explicação de como a atividade de IRP1 é regulada. Quando o suprimento de ferro é elevado, IRP1 incorpora um grupamento 4Fe-4S, passando a exibir atividade de aconitase citoplasmática e perdendo a capacidade de ligar-se ao IRE. Na presença de alta concentração de ferro, IRP2 é degradada pela via que envolve ubiquitina-proteassomo. A **Figura 14.12** mostra um esquema das interações entre IRPs e IREs.

O controle da expressão gênica na tradução dos RNAs mensageiros é muito importante, para a oogênese e as fases iniciais da embriogênese. Na maioria das espécies animais, os oócitos contêm mRNAs estáveis, que são armazenados e não traduzidos no citoplasma. A reativação de muitos mRNAs dormentes no oócito parece ser controlada por sequências localizadas na região 3'-UTR. Esse controle envolve tanto a ligação de proteínas reguladoras repressoras como a alteração no tamanho da cauda de poli(A). Os determinantes da região 3', que estão envolvidos tanto no alongamento da cauda de poli(A) como na ativação da tradução, incluem o hexanucleotídeo AAUAAA, também reconhecido pela maquinaria de

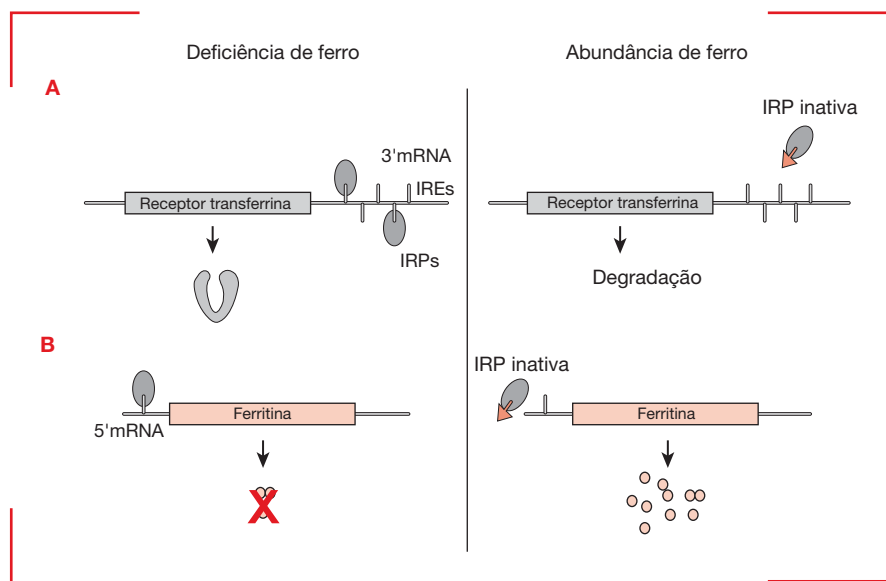


Figura 14.12

Regulação de expressão do receptor de transferrina e de ferritina pelo nível de ferro. (A) Regulação da estabilidade do mRNA do receptor de transferrina. Em condições de baixa concentração de ferro, a IRP é ativada e liga-se aos IREs, estabilizando o mRNA. Em condições de alta concentração de ferro, a IRP está inativa e não se liga aos IREs presentes na região 3' do mRNA de TfR. Existem cinco IREs na região 3' do mRNA de TfR. Sem IRP ligada, o mRNA é degradado por nucleases. (B) Regulação da tradução do mRNA de ferritina. Em condições de baixa concentração de ferro, a IRP é ativada e interage com o IRE, não havendo o início da tradução do mRNA da ferritina. Em condições de alta concentração de ferro, a IRP é inativa e não se liga ao IRE presente na região 5' do mRNA de ferritina, que é traduzido, havendo produção de ferritina.

poliadenilação do transcrito no núcleo e uma sequência rica em U, mais variável, denominada elemento citoplasmático de poliadenilação (CPE, de *cytoplasmic polyadenylation element*).

No caso do mRNA do ativador de plasminogênio do tipo tecidual (tPA, de *tissue-type plasminogen activator*) de camundongo, por exemplo, foi demonstrado que a ligação de uma proteína de aproximadamente 80 kDa a uma região rica em AU na 3'-UTR impede a sua tradução nos oócitos primários. Na fase de maturação meiótica, esse mRNA é traduzido. O modelo proposto para essa regulação sugere que nos oócitos primários em crescimento o RNA poliadenilado (cerca de 250 resíduos de A) recém-sintetizado é exportado do núcleo para o citoplasma. Uma proteína repressora da tradução liga-se a uma sequência na região 3' que determina o silenciamento do mRNA. Essa interação também induz a remoção parcial da cauda de poli(A), deixando uma cauda de aproximadamente 40 As no mRNA dormente. A presença do repressor impediria a ligação de PABP na cauda curta de poli(A), impossibilitando a tradução. Sabe-se que a formação de um complexo de iniciação da síntese proteica envolve a interação de PABP com o fator de iniciação eIF4E, havendo a circularização do mRNA, mostrando a atividade conjunta de 5'-CAP com a cauda de poli(A). Nos oócitos em maturação, o repressor é modificado e liberado do mRNA. PABP pode, então, ligar-se à cauda curta de poli(A), possibilitando a ativação da tradução. Concomitantemente, ocorre a readenilação da cauda de poli(A), promovendo a recuperação do seu tamanho, que é indispensável para a tradução.

12. Controle de expressão pela localização celular do mRNA

Também importante, como evidenciado na embriogênese de *D. melanogaster*, é o acoplamento entre o controle da tradução e a localização do mRNA. A localização do

mRNA é uma forma eficiente de possibilitar uma elevada concentração de proteína em um sítio-específico e, de permitir um controle temporal preciso de síntese local de proteína em resposta a um estímulo externo. Além disso, o acoplamento da tradução com a localização do mRNA impede que a síntese da proteína ocorra em local inapropriado, o que poderia ser prejudicial à célula. Embora tenha sido descoberta inicialmente em ovos de ascídia, a tradução localizada de mRNAs tem sido observada em vários outros tipos de células animais, como neurônios, fibroblastos e em diferentes organismos, como fungos e plantas. Acredita-se que a localização dos mRNAs na célula seja mediada por mecanismos que envolvem o reconhecimento de sequências no mRNA por proteínas, formando-se complexos ribonucleoproteicos. A localização ocorreria pela interação desses complexos com elementos do citoesqueleto.

Os avanços alcançados, sobre a importância da localização celular do mRNA na regulação da expressão gênica, têm sido proporcionados pelos estudos feitos em *D. melanogaster*. Muitos estudos têm sido realizados com os genes envolvidos na definição da polaridade do embrião, como *gurken* (*grk*), *bicoid* (*bcd*), *nanos* (*nos*) e *oskar* (*osk*) (ver Capítulo 15). Embora os estudos com esses genes sejam exemplos específicos, análises recentes mostram que a localização dos mRNAs em sítios subcelulares é muito mais comum do que se acreditava, sendo que no próprio embrião de *Drosophila* foram encontrados milhares de transcritos apresentando padrões distintos de distribuição espacial. Situação semelhante tem sido observada no caso de neurônios e fibroblastos, indicando que esse mecanismo pode ser utilizado nos mais diferentes tipos de células e organismos.

A localização subcelular, em geral, está acoplada à regulação da tradução do mRNA, possibilitando que a proteína codificada por aquele mRNA só seja sintetizada no local apropriado da célula. A regulação da tradução pode envolver a interação com proteínas ligantes de RNA que atuam como repressores da tradução. Essa repressão pode ser revertida pela remoção do repressor ou pelo remodelamento do complexo repressor.



Resumo

A regulação da expressão dos genes em células eucarióticas envolve a modificação da cromatina, possibilitando a ativação da transcrição, as modificações posteriores para dar origem ao RNA mensageiro, o transporte desse RNA para o citoplasma e a sua tradução em uma proteína. Um dos níveis importantes da regulação é o início da transcrição, em que sequências reguladoras são reconhecidas pela maquinaria de transcrição, possibilitando a ativação do gene e levand

do à síntese de RNA. As sequências reguladoras presentes nos promotores e reforçadores dos genes são reconhecidas pelos fatores de transcrição, que podem ser basais, necessários para a expressão de todos os genes transcritos por um determinado tipo de RNA-Pol, ou específicos, que incluem os ativadores, os coativadores, os repressores e os correpressores.

Os ativadores são constituídos de domínios funcionais que possibilitam o transporte para o núcleo, o

reconhecimento das sequências reguladoras dos genes, a ativação da transcrição e a interação com outras proteínas. Grande parte dos fatores de transcrição eucarióticos conhecidos apresenta um dos seguintes motivos estruturais para ligação ao DNA: homeodomínio, dedo de zinco (*zinc finger*), zíper de leucina (*leucine zipper*), hélice-alça-hélice (*helix-loop-helix*) e hélice “alada” (*winged-helix*). Junto com os promotores, os reforçadores exercem uma função muito importante na regulação da expressão gênica, pois eles são reconhecidos por ativadores que, com a participação de outros fatores (coativador, mediador), interagem com o complexo multiproteico basal de transcrição, aumentando o nível da transcrição.

No núcleo da célula eucariótica, a modificação das histonas e o remodelamento da cromatina são fatores importantes no processo de regulação da expressão. As histonas e, em particular, suas porções aminoterminais caracterizam-se por apresentar um grande número de modificações, como metilação de resíduos de arginina; metilação, acetilação, ubiquitinação, ADP-ribosilação e sumoilação de resíduos de lisina; e fosforilação de resíduos de serina e treonina.

A cromatina pode ser modificada por complexos remodeladores, que utilizam a hidrólise do ATP para alterar os contatos histona-DNA. O remodelamento da cromatina pode provocar um desenrolamento transitório entre o DNA e os octâmeros de histonas, facilitando o acesso de fatores de transcrição ao DNA.

A expressão gênica regulada por hormônios lipossolúveis, como hormônios esteroides, ácido retinoico e hormônios da tireoide, é um exemplo de regulação que envolve muitas proteínas. A regulação de lócus multigênico, como o do gene da β -globina, envolve a participação de elementos regulatórios (LCR) localizados em regiões bastante distantes dos genes que eles controlam. A regulação da expressão pode envolver os RNAs não codificadores (ncRNA) e pode ocorrer, também, em etapas posteriores ao início da transcrição, como, por exemplo, nas etapas que envolvem as modificações do RNA precursor. A participação de pequenos RNAs, como os miRNAs e siRNAs, no controle da tradução ou na degradação dos mRNAs contribui de forma importante para a regulação da expressão gênica. Por fim, essa regulação também pode ocorrer na tradução do mRNA ou na sua localização celular.



Leituras recomendadas

Alberts B, Johnson A, Lewis J, Raff R, Roberts K, Walter P. *Biologia molecular da célula*. 5. ed. Porto Alegre: Artmed; 2010.

Bain DL, Heneghan AF, Connaghan-Jones KD, Miura MT. Nuclear receptor structure: implications for function. *Annu Rev Physiol*. 2007;69:201-20.

Bhaumik SR, Smith E, Shilatifard A. Covalent modifications of histones during development and disease pathogenesis. *Nat Struct Mol Biol*. 2007;14(11):1008-16.

Chekulaeva M, Filipowicz W. Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Curr Opin Cell Biol*. 2009;21(3):452-60.

Clapier CR, Cairns BR. The biology of chromatin remodeling complexes. *Annu Rev Biochem*. 2009;78:273-304.

Farnham PJ. Insights from genomic profiling of transcription factors. *Nat Rev Genet*. 2009;10(9):605-16.

Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol*. 2009;10(2):126-39.

Kornberg RD. The molecular basis of eukaryotic transcription. *Proc Natl Acad Sci U S A*. 2007;104(32):12955-61.

Lewin B. *Genes* IX. 9. ed. Porto Alegre: Artmed; 2009.

Lodish H, Berk A, Kaiser CA, Krieger M, Scott MP, Bretscher A, et al. *Molecular cell biology*. 6th ed. New York: W. H. Freeman and Company; 2008.

Martin KC, Ephrussi A. mRNA localization: gene expression in the spatial dimension. *Cell*. 2009;136(4):719-30.

Noordermeer D, de Laat W. Joining the loops: beta-globin gene regulation. *IUBMB Life*. 2008;60(12):824-33.

Rouault TA. The role of iron regulatory proteins in mammalian iron homeostasis and disease. *Nat Chem Biol*. 2006;2(8):406-14.

Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol*. 2009;11(3):228-34.

Wu L, Belasco JG. Let me count the ways: mechanisms of gene regulation by miRNAs and siRNAs. *Mol Cell*. 2008;29(1):1-7.

Vera Lúcia S. Valente Gaiesky

Biologia Molecular do Desenvolvimento

1. Caixa de ferramentas 320

- 1.1 Ovos e diferenças nos padrões de clivagem *versus* conservação dos genes da caixa de ferramentas 323
 - 1.1.1 Genes maternos 324
 - 1.1.2 Genes embrionários (ou zigóticos) 325

2. Genes da família PAX 327

- 3. Genes *otd/Otx* e *ems/Emx* e a formação das partes do cérebro, genes *Dll/Dlx* e a formação dos apêndices e genes *tinman* e *NK2* e a formação do coração 328**

Durante muito tempo a complexidade dos organismos superiores e as diferenças entre seus planos corporais, seus mecanismos de fecundação e o desenvolvimento embrionário dificultaram o entendimento do conjunto de processos atuantes sobre a ontogenia dos eucariotos. É nesse contexto que organismos modelo tão diferentes quanto o verme *Caenorhabditis elegans*, a mosca *Drosophila*, o anfíbio *Xenopus*, o peixe *Danio rerio* e o camundongo *Mus musculus*, entre muitos outros, foram fundamentais para a descoberta dos papéis das famílias de genes codificadores de fatores de transcrição evolutivamente conservadas, que regem o desenvolvimento embrionário dos eucariotos. Essa nova visão da evolução, chamada Evo-Devo, tem revolucionado o pensamento científico, principalmente no que concerne à evolução morfológica.

Alguns dos organismos citados possuíam sua biologia bem conhecida há mais de um século, mas outros foram escolhidos para este fim há apenas poucas décadas. Quando ao sólido conhecimento acumulado sobre a biologia e a genética de organismos como a *Drosophila*, por exemplo, somaram-se as novas ferramentas da biologia molecular, o quebra-cabeça do papel dos genes no desenvolvimento começou a ser decifrado. A revelação da similaridade genética subjacente aos diferentes planos corporais, com base na conservação das famílias de proteínas reguladoras, resultou na quebra de muitos paradigmas, até então vigentes para a biologia. O presente capítulo apresentará um breve resumo dos principais achados sobre os genes de desenvolvimento que passaram a fazer parte da chamada “caixa de ferramentas genéticas” para construir os corpos dos animais.

1. Caixa de ferramentas

Fazem parte da “caixa de ferramentas”, genes e famílias gênicas que codificam os fatores de transcrição responsáveis pelo desenvolvimento embrionário e pela diferenciação. Ao contrário de outros tantos, os genes da caixa de ferramentas são considerados “genes mestres”, já que estão no topo das redes reguladoras e das rotas de sinalização responsáveis pelo comando de programas de desenvolvi-

mento. Esses programas englobam desde o estabelecimento dos planos corporais básicos, como anteroposterior e dorsoventral, até a formação de estruturas, como cérebro, olhos e apêndices. As redes reguladoras coordenadas pelos genes mestres ligam (ou desligam) os genes estruturais requisitados para formar as diferentes estruturas.

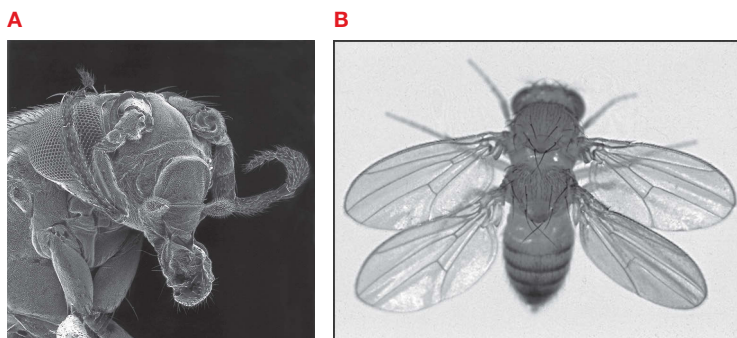
A descoberta da conservação evolutiva das famílias gênicas começou com a utilização da *Drosophila melanogaster* como organismo experimental para o estudo da genética, no início do século XX. Alguns mutantes de *D. melanogaster*, isolados em laboratório, revelaram a existência de programas de desenvolvimento responsáveis pela identidade de certas partes do corpo. Além de serem os responsáveis por estruturas defeituosas, esses mutantes (chamados de mutantes homeóticos) resultavam na produção de estruturas corretas, mas que se desenvolviam em locais errados. As moscas tinham patas no lugar das antenas e dois pares de asas, resultantes da duplicação de parte do seu tórax (**Figura 15.1**). Devido ao aspecto bizarro, esses mutantes foram muito importantes, pois revelaram que, antes dos apêndices se formarem, já havia um “plano de corpo” a ser estabelecido.

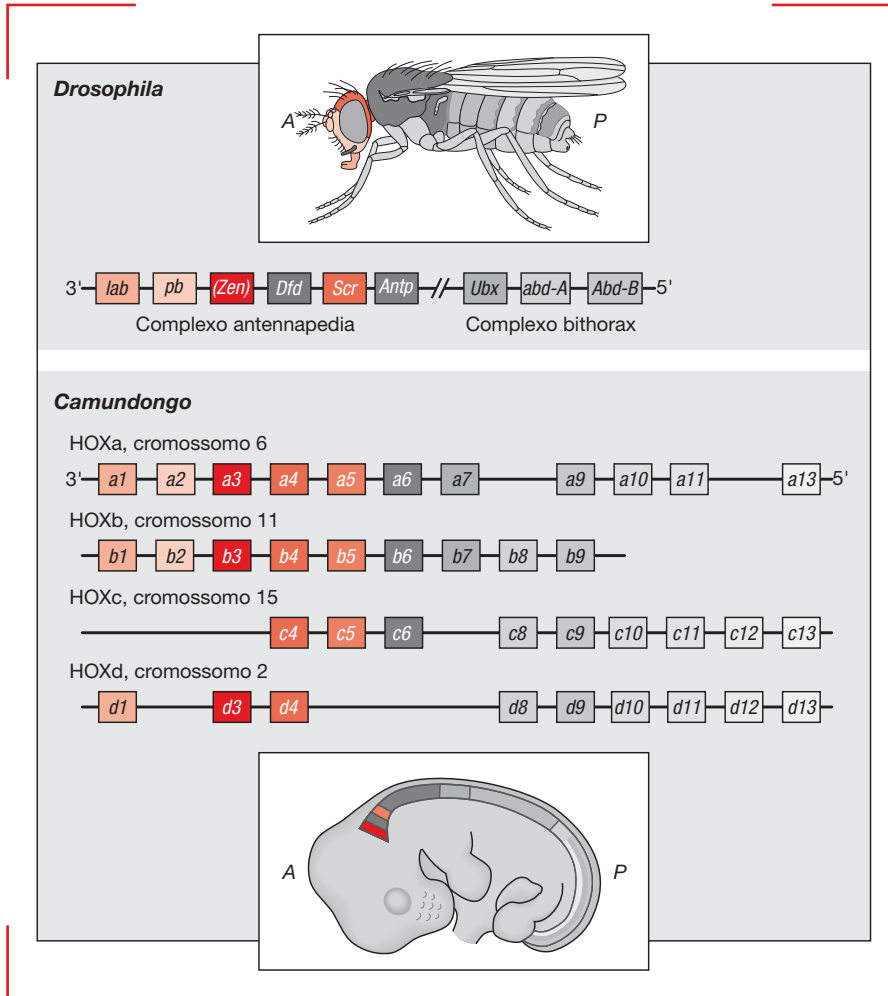
Utilizando os métodos da genética “clássica”, Edward Lewis (Prêmio Nobel de Medicina e Fisiologia em 1995) verificou que os genes responsáveis pelas malformações no plano corporal das moscas localizavam-se no mesmo cromossomo e na mesma ordem física com que deveriam se expressar no corpo do embrião durante o desenvolvimento. Vários outros mutantes, causadores de defeito sobre a organização do corpo das moscas, foram descobertos posteriormente e, mais uma vez, verificou-se que os genes que os codificavam faziam parte do mesmo agrupamento e estavam enfileirados no mesmo cromossomo de *D. melanogaster* (**Figura 15.2**).

Quando foi possível analisar o DNA dos nove genes agrupados no braço direito do terceiro cromossomo (3R) de *D. melanogaster*, foi descoberta a existência de uma sequência de pares de bases (pb) conservada entre todos eles. Essa região, de 180 pb localizada nos terceiros éxons de cada gene, foi denominada Caixa Homeótica (**homeobox**) (**Figura 15.3**). A caixa homeótica codifica cerca de 66 aminoácidos, que correspondem ao cha-

Figura 15.1

Mutantes homeóticos de *D. melanogaster*. (A) *Antennapedia (Antp)*, (B) *Bithorax (Bx)*.



**Figura 15.2**

Os genes *HOM/HOX* e a sua conservação no estabelecimento do plano corporal de *Drosophila* (invertebrado) e camundongo (vertebrado). As mesmas posições dos genes *HOM* (invertebrados)/*HOX* (vertebrados) indicam similaridade estrutural e funcional não só na região da caixa homeótica (*homeobox*), mas também em todo o gene.

mado **homeodomínio** da proteína. O homeodomínio é a parte da proteína que se liga às regiões reguladoras de outros genes. Os genes que possuem o *homeobox* são, portanto, fatores de transcrição.

A partir disso, a descoberta dos *homeoboxes* nos genes de desenvolvimento de *Drosophila* estendeu-se a outros genes, não apenas membros da mesma fileira de genes do cromossomo 3R, mas também espalhados ao longo dos seus quatro pares de cromossomos. Esses genes são importantes para o início e para o progresso do desenvolvimento embrionário. Mais tarde, a caixa homeótica também começou a ser encontrada em genomas de outros eucariotos, tanto nos mais complexos do que as moscas, como cordados (anfioxo) e vertebrados (camundongos, sapos, peixes, aves e seres humanos), quanto nos mais simples (cnidários e vermes). Assim, a próxima etapa foi a verificação de que todos esses organismos de morfologia adulta e complexidade distintas eram “construídos com as mesmas ferramentas”.

A caixa homeótica estava presente nos genes codificadores de proteínas reguladoras do desenvolvimento de todos os animais estudados. Porém, os membros in-

tegrantes da fileira de genes do terceiro cromossomo da mosca ainda estavam enfileirados nos cromossomos dos outros organismos-modelo. Os derivados desses organismos, como muitos dos vertebrados atuais (sapos, camundongos e seres humanos) tem quatro réplicas desta unidade inicial, em quatro cromossomos diferentes (Figura 15.2). Contudo, uma relação estrita entre o aumento da complexidade e o maior número de genes *HOM/HOX* (denominação deste grupo de genes em invertebrados e vertebrados, respectivamente) não é absoluta. Por exemplo, na linhagem que deu origem aos peixes atuais, o conjunto de genes enfileirados desde o ancestral comum sofreu mais um ciclo de duplicação, gerando oito fileiras de genes *HOM/HOX*, cada uma em um cromossomo diferente.

Assim, foi revelada uma regra geral para a construção dos corpos animais, em que os genes homéotícos em sintenia determinam não as estruturas em si, mas a identidade de cada parte do corpo, isto é, dão uma informação posicional às células do embrião. Em sua versão correta, os produtos desses genes “dizem”, a cada parte do corpo, quais as estruturas que devem ou podem se formar naquele local.

Figura 15.3

A Caixa Homeótica (homeobox).

<i>Hox 1.1</i>	Ser	Lys	Arg	Gly	Arg	Thr	Ala	Tyr	Thr	Arg	Pro	Gln	Leu	Val	Glu	Leu	Glu	Lys	Glu	Phe
<i>Xlhbox3</i>	Arg	Lys	Arg	Gly	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln	Thr	Leu	Glu	Leu	Glu	Lys	Glu	Phe
<i>Antennapedia</i>	Arg	Lys	Arg	Gly	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln	Thr	Leu	Glu	Leu	Glu	Lys	Glu	Phe
<i>Fushi tarazu</i>	Ser	Lys	Arg	Thr	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln	Thr	Leu	Glu	Leu	Glu	Lys	Glu	Phe
<i>Ultrabithorax</i>	Arg	Arg	Arg	Gly	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln	Thr	Leu	Glu	Leu	Glu	Lys	Glu	Phe
Hélice 1																				
<i>Xox 1.1</i>	His	Phe	Asn	Arg	Tyr	Leu	Met	Arg	Pro	Arg	Arg	Val	Glu	Met	Ala	Asn	Leu	Leu	Asn	Leu
<i>Xlhbox3</i>	His	Phe	Asn	Arg	Tyr	Leu	Thr	Arg	Arg	Arg	Arg	Ile	Glu	Ile	Ala	His	Val	Leu	Cys	Leu
<i>Antennapedia</i>	His	Phe	Asn	Arg	Tyr	Leu	Thr	Arg	Arg	Arg	Arg	Ile	Glu	Ile	Ala	His	Ala	Leu	Cys	Leu
<i>Fushi tarazu</i>	His	Phe	Asn	Arg	Tyr	Ile	Thr	Arg	Arg	Arg	Arg	Ile	Asp	Ile	Ala	Asn	Ala	Leu	Ser	Leu
<i>Ultrabithorax</i>	His	Thr	Asn	His	Tyr	Leu	Thr	Arg	Arg	Arg	Arg	Ile	Glu	Met	Ala	Tyr	Ala	Leu	Ser	Leu
Hélice 2																				
<i>Xox 1.1</i>	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Met	Lys	Tyr	Lys	Lys	Asp	Gln
<i>Xlhbox3</i>	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Met	Lys	Trp	Lys	Lys	Glu	Asn
<i>Antennapedia</i>	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Met	Lys	Trp	Lys	Lys	Glu	Asn
<i>Fushi tarazu</i>	Ser	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Met	Lys	Ser	Lys	Lys	Asp	Arg
<i>Ultrabithorax</i>	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Met	Lys	Leu	Lys	Lys	Glu	Ile
Hélice 3																				

Em *D. melanogaster*, por exemplo, a maioria das mutações homeóticas pertence a dois complexos gênicos localizados no cromossomo 3R: o complexo **Antennapedia** (*ANT-C*), que afeta as estruturas da cabeça e do tórax, e o complexo **Bithorax** (*Bx-C*). A partir da análise do complexo *Bx-C*, tanto do ponto de vista genético como do molecular, foi estabelecido que a parte posterior de cada segmento do corpo de uma mosca adulta forma-se a partir da ação conjunta de um grupo único de genes homeóticos. Por esse modelo de ação, o segundo segmento torácico, que é o segmento mais anterior a ser controlado pelo complexo *Bx-C*, requer a ativação do menor número de genes homeóticos, mas à medida que os segmentos posteriores vão se diferenciando, cada vez mais genes adjacentes vão sendo requeridos.

O complexo *Bx-C* (ver **Figura 15.4**) é constituído por três genes, cada um deles produz vários produtos proteicos a partir de diferentes padrões de *splicing* alternativos do RNA (ver Capítulo 10). Esses três genes, **Ultrabithorax** (*Ubx*), **abdominal-A** (*abd-A*) e **Abdominal-B** (*Abd-B*), são expressos na epiderme, no Sistema Nervoso Central (SNC) e no mesoderma somático e visceral, mas não no endoderma.

A região central do gene *Ubx* corresponde a uma unidade de transcrição com cerca de 77 kb, que codifica um conjunto de mRNAs de 3,2 e 4,3 kb, diferenciados pela utilização de dois sítios de poliadenilação e pela formação de seis proteínas por *splicing* alternativo. A expressão dos polipeptídeos codificados por esses mRNAs dá-se nas regiões correspondentes aos parassegmentos (regiões de intersecção entre dois segmentos do corpo, envolvendo a parte posterior de um e a parte anterior do outro) 5 e 6, no ectoderma larval e no embrião e, também, nos discos imaginais (conjuntos de células larvais, correspondentes às futuras estruturas adultas, como asas, patas, halteres e antenas). Essa organização estrutural do gene garante a existência de regulação temporal e espacial por *splicing* alternativo, já que os mRNAs de 3,2 kb são expressos mais cedo na embriogênese na epiderme e no mesoderma, e os mRNAs de 4,3 kb expressam-se no SNC durante a fase tardia da embriogênese.

Uma mutação do locus *Ubx* produz a transformação homeótica dos parassegmentos 5 e 6 no parassegmento 4, uma vez que altera a produção de sequências corretas de mRNA do gene. Manifestações parciais do genótipo *Ubx* são obtidas por outros quatro mutantes (***abx***, ***bx***, ***bx^d*** e ***pbx***), que têm posições alteradas de sequências reguladoras específicas. Mutantes *abx* e *bx* diminuem a expres-

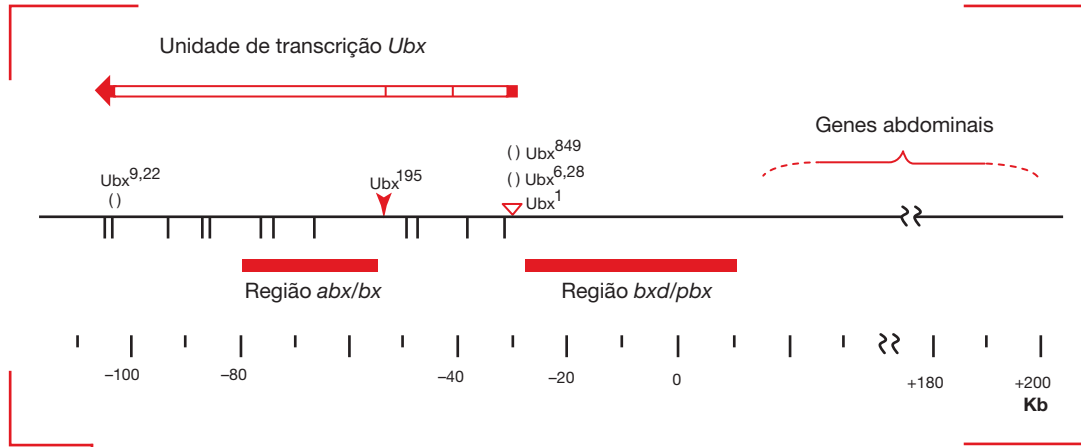


Figura 15.4

Mapa estrutural da região do gene *Ubx* de *D. melanogaster*. Diferentes alelos mutantes são apresentados, indicando a região do gene e qual a alteração encontrada. Inserção (∇) deleção (∇).

são do produto *Ubx* no parassegmento 5, ao passo que os mutantes *bxd* e *pbx* o fazem no parassegmento 6. O estudo desses mutantes ajudou a identificar duas regiões reguladoras que agem em *cis*, uma no parassegmento 5 e outra no parassegmento 6. O grande tamanho dessas regiões (25 e 40 kb, respectivamente) deve-se à presença de muitos elementos reguladores.

Os genes *abd-A* e *Abd-B*, por sua vez, também localizados na região de *Ubx* e responsáveis pela diferenciação dos segmentos abdominais, parecem ter por função reprimir a ação de *Ubx* sobre a região posterior do embrião. A precisão com que os genes homeóticos devem se expressar, garantindo a manutenção da identidade espacial dos segmentos, entretanto, depende da atuação de outros grupos de genes: *polycomb* e *trithorax*. Assim, as proteínas dos genes do grupo *polycomb* mantêm a repressão dos genes homeóticos, ao passo que as do grupo *trithorax* os mantêm ativos nos seus domínios corretos de expressão.

1.1 Ovos e diferenças nos padrões de clivagem *versus* conservação dos genes da caixa de ferramentas

Os oócitos e os embriões dos diversos organismos-modelo são muito diferentes, tanto em tamanho como em quantidade de vitelo e outras características dependentes do tipo de fecundação e desenvolvimento dos embriões. Apesar disso, foi demonstrado que muitos dos genes responsáveis pelo estabelecimento das principais coordenadas e pela polaridade dos ovos são conservados entre animais distintos. As mensagens dos genes responsáveis pela polaridade do ovo e pelo estabelecimento de suas primeiras coordenadas estão depositadas no citoplasma dos oócitos, mesmo antes da sua fecundação, devido à transcrição prévia dos chamados **genes de efeito materno**.

Christiane Nüsslein-Volhard e Erich Wieschaus compartilharam com Edward Lewis o Prêmio Nobel em 1995, no que diz respeito às suas descobertas sobre os genes de efeito materno e os genes que se expressam nos embriões iniciais (**genes zigóticos**), antes mesmo que os genes homeóticos do complexo *HOM/HOX* possam ser expressos. Novamente, foi na *Drosophila* que esses genes foram descobertos.

O oócito de *Drosophila* se desenvolve no interior do ovário, cercado por células que o nutrem (*nurse cells*) com organelas, vitelo e mRNAs de origem materna. São as mensagens codificadas por essas moléculas de RNA que serão processadas quando o desenvolvimento se inicia após a fecundação, e que definirão as primeiras coordenadas do embrião.

Logo depois da fecundação, o zigoto passa por sete ciclos de divisão em um citoplasma comum, formando um sincício. Esses núcleos-filhos vão migrando à medida que se dividem (até cerca de 4.000 núcleos) e alguns se dirigem para o que será a região anterior do embrião. Neste estágio, o embrião corresponde ao blastoderma sincicial. Após mais alguns ciclos de divisão, os demais núcleos do embrião migram para a superfície do ovo e 15 deles irão concentrar-se na sua região terminal (polo posterior), onde, então, se diferenciam em células polares. Algumas dessas células se tornarão células germinais definitivas, ao passo que outras migrarão, retornando à região central do embrião. Quando mais ou menos 6.000 núcleos estiverem na periferia do ovo, as membranas plasmáticas irão separá-los em células individuais, formando uma camada monocelular, chamada de blastoderma celular.

Até esta fase, os diversos núcleos compartilham um citoplasma comum, o que propicia interações dinâmicas entre eles e os mRNAs ou as proteínas diferencialmente posicionadas ao longo do embrião. Essas macromoléculas, dispostas no citoplasma em forma de gradiente, são responsáveis pela chamada informação posicional para os núcleos.

Assim, o citoplasma do ovo é previamente diferenciado antes da fecundação e os núcleos, que vão migrar para cada polo do embrião, passarão a responder às mensagens que haviam sido codificadas pelo genoma da mãe. A descoberta dos genes envolvidos no estabelecimento da polaridade inicial do embrião foi possível pela disponibilidade de mutações, cuja expressão permitiu que se conhecesse o efeito dos genes normais. Muitos desses genes parecem atuar de forma isolada, ou em cooperação com outros genes, em momentos posteriores ao desenvolvimento.

De uma maneira resumida, o esquema da **Figura 15.5** mostra os principais grupos de genes que regem, de forma sincronizada, o início do desenvolvimento na *Drosophila*. Alguns genes desses grupos foram tomados como modelo e, a seguir, serão discutidos em maior detalhe.

1.1.1 Genes maternos

Vários genes de origem materna, atuantes no início do desenvolvimento embrionário de *Drosophila*, são conhecidos. Entre os lócus envolvidos nessa etapa, ***bicoid***, ***torso***, ***nanos*** e ***oskar*** são alguns dos mais estudados.

Os genes expressos pelo genoma da mãe possuem a capacidade de controlar, de forma independente, quatro sistemas atuantes no ovo e no embrião: (1) um sistema determinante do gradiente anteroposterior, mediado pela proteína *bicoid*, traduzida a partir de mRNAs localizados no polo anterior; (2) um sistema atuante na parte posterior do embrião, necessário para a formação do abdome, sob ação do produto do gene *nanos*; (3) um sistema terminal, responsável pela diferenciação da cabeça e da cauda, dependente da ativação da proteína receptora *torso* nas extremidades do embrião; e (4) um sistema de ação dorsoventral, mediado pela proteína receptora *toll*, que atua na linha média do ovo, fazendo com que a proteína materna *dorsal*, distribuída no citoplasma ao longo deste eixo, penetre diferencialmente nos núcleos das regiões dorsal e ventral.

O gene *bicoid*, que é o paradigma dos genes de efeito materno em *Drosophila*, também contém um *home-*

-obox. O mRNA do gene *bicoid* é produzido pelas *nurse cells* do ovário da fêmea e deposita-se no polo anterior do oócito, onde é reconhecido por proteínas localizadas no citoplasma do ovo dessa região. Os produtos de pelo menos três genes são necessários para a localização precisa do mRNA do *bicoid* no embrião: ***staufen***, ***swallow*** e ***exuperantia***.

A proteína *bicoid* foi o primeiro fator de transcrição, que age como morfógeno, a ser bem caracterizado. Essa proteína tem uma meia-vida curta (cerca de 30 min), apesar de desempenhar um papel decisivo para o estabelecimento da polaridade do ovo.

Mutantes para *bicoid* (bem como para outros genes maternos), cujos produtos não são capazes de se distribuir corretamente no citoplasma do ovo, propiciam a reversão da polaridade, podendo formar embriões com duas “cabeças”, que tem efeito letal. Parte da região 3' não traduzida do gene *bicoid* é capaz de formar uma estrutura tridimensional, cuja forma e sequência são reconhecidas pelas proteínas que controlam o seu correto posicionamento.

O padrão da região posterior do embrião, por sua vez, é regulado pela ação de pelo menos nove genes, codificados pelo genoma materno. Mutantes nesses genes, como em *nanos* e *oskar*, ocasionam a perda da formação de segmentos abdominais. O mRNA do gene *nanos* foi identificado, por hibridização *in situ*, no polo posterior do ovo e sua principal função parece ser a de impedir a tradução do mRNA de *bicoid* na parte posterior do embrião e, também, a do mRNA materno de ***hunchback*** (um gene zigótico) na mesma região. Muitos outros genes também são necessários para impedir o mRNA de ***hunchback*** de agir na região posterior do ovo, garantindo, assim, o correto funcionamento de ***nanos***.

Os mutantes maternos que agem de forma dorsoventral não afetam o número de segmentos, mas as camadas germinais que crescem ao longo do eixo anteroposterior, determinando se certos segmentos vão ou não ser formados e afetando, assim, o número total de segmentos. Os

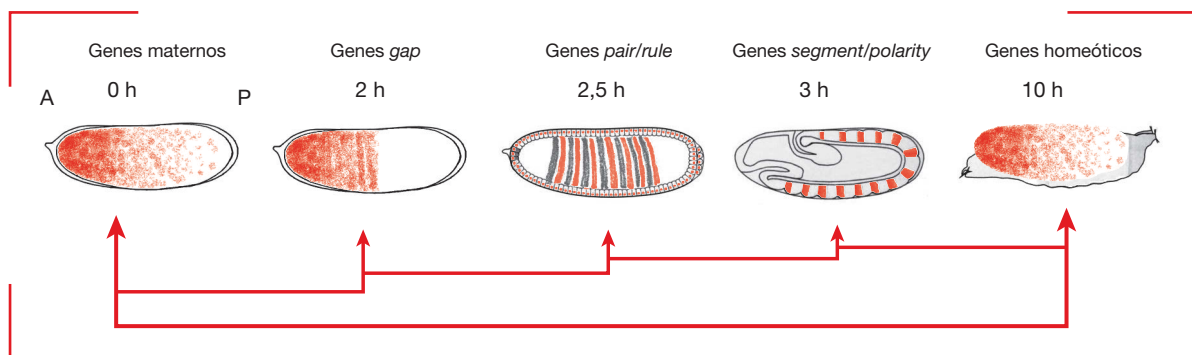


Figura 15.5

Principais grupos de genes e seu tempo aproximado de expressão no embrião de *D. melanogaster*. As regiões em ■ correspondem aos domínios de expressão desses genes. A = polo anterior; P = polo posterior.

mutantes nos genes que agem de maneira dorsoventral no embrião afetam derivados mesodérmicos, transformando em região dorsal aquelas que deveriam ser ventrais. No caso inverso, o embrião tem menos mesoderma e mais ectoderma.

1.1.2 Genes embrionários (ou zigóticos)

Uma vez definido o plano básico de desenvolvimento do embrião pelos genes maternos, começam a ser expressos os genes do próprio embrião (também chamados de zigóticos, por serem expressos após a fecundação). São conhecidos muitos genes que atuam na etapa das duas horas, caracterizada pelo estabelecimento da segmentação do corpo do embrião, que se manterá na larva, e, posteriormente, no indivíduo adulto. Três grupos principais de genes atuam no estabelecimento do padrão de segmentação do corpo da *Drosophila*. São eles:

1. Genes **gap** (lacuna), cuja ativação precede a dos outros e que, por sua vez, são ativados pelos genes maternos. Os genes *gap* (nome originado do mutante cuja manifestação determina falta de segmentos) expressam-se na linha mediana do embrião, dividindo-o em três partes; todos codificam fatores de transcrição que apresentam meia-vida curta (de poucos minutos).

De maneira resumida, pode-se dizer que há no mínimo três funções importantes determinadas pelos genes *gap* e que estão associadas com a formação de padrões. Eles agem sobre a polaridade do corpo, em seguimento aos genes de efeito materno, posicionando a sequência das partes (o que pode ser revertido pelos seus mutantes); agem sobre o espaçamento dos primórdios dos parassegmentos (regiões de interseção entre os 14 segmentos), que nos mutantes são perdidos, fusionados ou erroneamente espaçados; e agem sobre a diferenciação das partes amplas do corpo, como tórax e abdome. Esse grupo de genes, portanto, desempenha um papel intermediário entre os genes maternos e os zigóticos.

O gene *Kruppel* (que tem homologia com o fator de transcrição TFIIIA de *Xenopus*) é um conhecido gene *gap*, ativado pela ligação da proteína *bicoid* em muitos sítios reguladores da sua região promotora. Esse gene é expresso na região central do embrião e atua sobre nove dos 14 parassegmentos do corpo.

É sobre o padrão especificado pelos genes **gap** que os produtos dos genes de outras duas famílias serão posicionados.

2. Os genes **pair rule** (regra dos pares) são ativados pelos genes *gap* e determinam o estabelecimento de um padrão de sete faixas, intercaladas por uma distância correspondente a dois segmentos. Seu tempo de expressão ocorre em torno das três horas, quando o embrião já está na fase de blastoderma celular e seus mutantes induzem a perda repetida de partes específicas ou de segmentos alternados.

Os genes **fushi tarazu** (*ftz*) e **hairy** são paradigmas dessa família de genes e desempenham um papel marcante no estabelecimento da segmentação do embrião de *Drosophila*. Nos embriões mutantes para *fushi tarazu* partes de alguns segmentos são perdidas e as porções incompletas fundem-se com os segmentos adjacentes. Por exemplo, as porções anteriores dos segmentos Mx, T1, T3, A2, A6 e A8 se fundem com a região posterior dos segmentos que os seguem. Isso tem um efeito letal sobre o embrião, que possui só a metade dos 14 segmentos da larva normal.

O gene *fushi tarazu* é regulado de forma ampla pela ação repressiva do produto do gene *hairy*, cuja ação é primariamente localizada nas interfaixas. Na **Figura 15.6** é apresentado um esquema da expressão do gene *hairy* em cada faixa do embrião até então determinada.

Alguns dos elementos reguladores, alvos das proteínas *pair rule* são bastante complexos. Um elemento reforçador (*enhancer*) a montante é o alvo de *ftz*, para o qual contém muitos sítios de ligação. Uma outra região, de 600 pb, inclui o promotor contendo o elemento *ftz*, que é capaz de promover a expressão periódica de *ftz*. Esse elemento, chamado “zebra”, corresponde a um conjunto de sequências-alvo para diversas proteínas reguladoras, responsáveis pela indução do padrão de expressão do gene em forma de faixas.

O gene *ftz* funcional consiste em uma unidade de transcrição de 1,9 kb e uma grande região de controle flanqueadora a 5', de 6,1 kb, necessária para sua expressão correta. Da mesma forma, utilizada para rastrear a expressão de *hairy*, as grandes regiões flanqueadoras 5' de *fushi tarazu* foram fusionadas com as sequências codificadoras do gene da β -galactosidase de *Escherichia coli*. O padrão de expressão da enzima foi comparado com o obtido pela hibridização *in situ*, por coloração com anticorpos fluorescentes e com coloração histoquímica para β -galactosidase. Em todos os casos, o padrão de faixas do tipo “zebra” foi obtido.

Nos últimos estágios embrionários, utilizando as mesmas estratégias metodológicas, verificou-se que a proteína do gene *ftz* também é produzida em precursores neuronais específicos no SNC. O padrão de expressão de *ftz* nesse estágio do desenvolvimento é muito preciso ao ponto de sugerir que a regulação da atividade gênica seja feita em nível de cada célula. No SNC, o padrão de expressão de *ftz* é repetido a cada segmento, ao contrário do padrão “zebra”, que se expressa a cada dois segmentos.

Pelo menos três elementos reguladores da expressão do gene *ftz* são conhecidos na região flanqueadora 5': um elemento a montante, mais distal em relação ao gene; um elemento neurogênico, necessário para a sua expressão no SNC; e o elemento “zebra”, localizado dentro dos 600 pb que antecedem o início da transcrição.

Assim, a aquisição do padrão “zebra” no embrião parece requerer desde a ação de genes de efeito materno (que determinam as coordenadas básicas e estabelecem

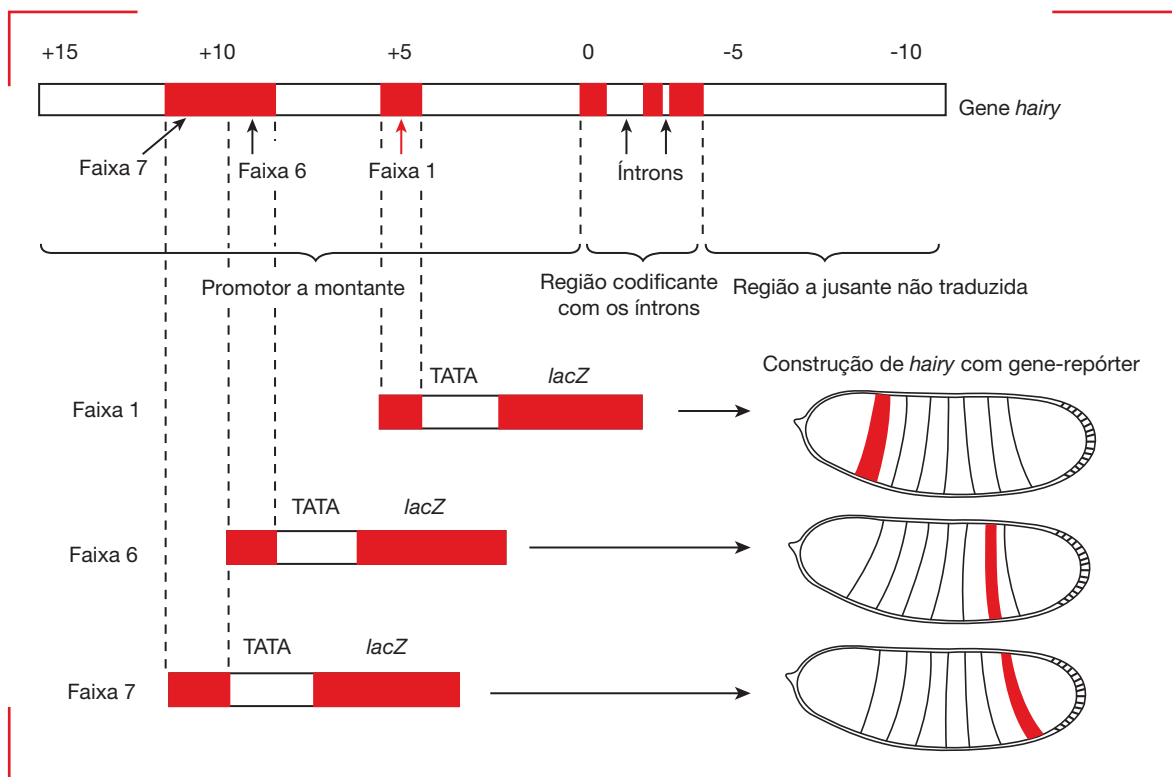


Figura 15.6

Expressão de cada faixa do embrião mediada pelo gene *hairy*. A descoberta dessa especificidade decorreu da utilização de construções obtidas pela fusão de pequenos segmentos do promotor de *hairy* com o gene *lacZ* de *E. coli* responsável pela produção de β-galactosidase. Os locais onde a enzima foi produzida (faixas em ■) correspondem aos da expressão de *hairy*.

um pré-padrão), até a ação coordenada de genes zigóticos de segmentação, por exemplo, *hairy* e *ftz*.

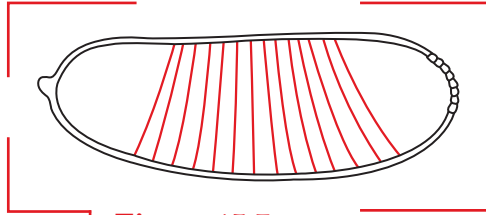
A análise do elemento a montante do gene *ftz* indicou que ele é um *enhancer*, necessário para a expressão em faixas, e que a proteína *ftz* pode interagir com o seu próprio *enhancer*, como ocorre com outros fatores de transcrição, reguladores de sua própria expressão.

- Os genes ***segment-polarity*** (polaridade de segmentos) são expressos em um estágio mais avançado do embrião, ao tempo da gastrulação (três horas e trinta minutos) e determinam a polaridade dos segmentos. Seus alvos são cada um dos segmentos ou uma parte deles. Seus mutantes geralmente são letais, uma vez que mudam o padrão ou a polaridade de cada segmento. Esses genes atuam sobre as bordas dos parassegmentos que haviam sido delineados por *ftz* e *eve*, fixando de forma definitiva o padrão estabelecido pelos genes anteriores.

Entre os genes de polaridade de segmentos, ***engrailed*** é considerado como modelo, já que especifica um claro padrão de compartimentalização. O gene *engrailed* é, a princípio, expresso na região anterior de cada parassegmento em uma única fileira de células, passando depois a ser responsável por maiores domínios de expressão em

linhagens celulares derivadas. Em *D. melanogaster*, o lócus do gene *engrailed* é bastante extenso (cerca de 70 kb) e responsável pela produção de três transcritos, com 3,6, 2,7 e 1,4 kb. A extensão não transcrita desse lócus, portanto, corresponde a regiões de controle.

Embora traços de expressão de *engrailed* sejam detectados muito cedo no embrião de *Drosophila*, eles só se tornam expressivos na gastrulação, quando o padrão característico de 14 faixas correspondentes à região posterior de cada segmento torna-se nítido (ver **Figura 15.7**), além de produzir uma faixa estreita na posição correspondente ao sulco cefálico. Mais tarde (dez horas de desenvolvimento), células de todos os segmentos no SNC começam a expressar *engrailed*. Diferentemente dos genes das classes atuantes nos estágios embrionários anteriores, o gene *engrailed* é expresso de forma contínua através das fases embrionária, larval, pupal e adulta. A expressão de *engrailed* e dos genes homólogos a ele no SNC parece ser comum a vários organismos, desde invertebrados, como vermes, ouriço-do-mar e insetos, até vertebrados, como peixes, aves, camundongos e seres humanos, mas seu papel na segmentação do corpo parece ser exclusivo de artrópodes. A sua persistência em organismos tão diversos, portanto, caracteriza uma ideia de antiguidade evolutiva dessas sequências.

**Figura 15.7**

Padrão final de expressão de engrailed (3 h) no embrião de *D. melanogaster*.

Outros genes integrantes dessa categoria, como **hedgehog** e **wingless**, agem cooperativamente com **engrailed**, mantendo a delimitação dos parassegmentos. Os produtos desses genes são proteínas sinalizadoras, relacionadas com as codificadas por genes que desempenham papéis importantes durante a formação de padrões em vertebrados, como **Sonic hedgehog** e **Wnt**. Sua capacidade de delimitar territórios faz com que esses genes, e outros a eles relacionados, também façam parte da caixa de ferramentas.

Somente a partir dos padrões gerais de polaridade do embrião e segmentação estabelecidos, pelos genes maternos e zigóticos (*gap*, *pair rule* e *segment-polarity*), é que são criadas as condições para a ação de genes com funções mais específicas. Neste momento, os genes seletores dos complexos **HOM/HOX** começam a se expressar, especificando a identidade de cada segmento do corpo adulto. Esses genes, também devem controlar a atividade de muitos outros genes ao longo de todo o desenvolvimento, de forma a manter, em cada segmento, o padrão correto de expressão gênica. Porém, as estruturas vão usar outros componentes da caixa de ferramentas, ou

seja, outras famílias gênicas serão envolvidas nos processos de morfogênese.

A **Tabela 15.1** resume as principais famílias de genes consideradas membros da caixa de ferramentas.

2. Genes da família **PAX**

A família **PAX** codifica um grupo de fatores de transcrição que atua no início do desenvolvimento de eucariotos e que foi conservado ao longo de milhões de anos. Essas proteínas são reguladoras-chave da morfogênese. Como consequência de sua relevância, as mutações nesses genes geram malformações nos seus portadores, sejam eles moscas, sapos, camundongos ou seres humanos. São conhecidos pelo menos nove genes relacionados: **PAX 1, 2, 3, 4, 5, 6, 7, 8 e 9** em vertebrados e oito em invertebrados.

Todas as proteínas **PAX** possuem o chamado domínio *paired* descoberto no gene *paired* de *Drosophila*, membro da classe de genes zigóticos que se liga ao DNA e compartilham um ou mais domínios **HOX** e/ou um octapeptídeo. Dois dos membros dessa família fazem parte da caixa de ferramentas genéticas: **PAX6** e **PAX9**, devido ao seu papel como genes mestres das rotas formadoras de olhos e dentes, além de outras importantes funções.

O papel do **PAX6** como formador de olhos foi estudado pelo grupo de Walter Gehring na Suíça, que construiu moscas transgênicas, utilizando a técnica **GAL4/UAS** para induzir a expressão de um gene **PAX6**. **GAL4** é um gene de levedura, codificador de um fator de transcrição que se liga apenas às sequências reguladoras **UAS**, também de levedura. Duas linhagens de moscas selvagens foram transformadas: uma contendo o elemento **UAS** acoplado com uma cópia da região codificadora do gene

Tabela 15.1 Alguns genes conservados da caixa de ferramentas genéticas para construir os corpos animais

Gene	Função	Ocorrência
Complexo HOM/HOX	Padronização anterior/posterior	Invertebrados/vertebrados
<i>otd/Otx1, Otx2</i>	Cefalização	Invertebrados/vertebrados
<i>ems/Emx1/Emx2</i>	Padronização anterior	
<i>sog/Chordin</i>	Padronização dorsal/ventral	Invertebrados/vertebrados
<i>dpp/BMP4</i>		
<i>Dll/Dlx 1-2</i>	Formação de apêndices	Invertebrados/vertebrados
<i>Dlx 3-7</i>	Eixo proximal/distal	
<i>Dlx 6-5</i>		
PAX6	Formação dos olhos Regulação do SNC Anterior	Invertebrados/vertebrados
PAX9	Regulação do SNC Formação de dentes	Alguns vertebrados
<i>tinman/NK2</i>	Formação de Sistema Vascular Formação do coração	Invertebrados Vertebrados
<i>polycomb</i>	Controle da expressão dos genes HOM/HOX	Invertebrados/vertebrados

PAX 6, e a outra com a sequência codificadora do gene *GAL4* acoplada com o gene da Proteína H, que o controla.

Cruzando as duas linhagens de moscas transformadas, formou-se uma prole que tinha o gene *GAL4* da levedura expresso e a sequência *UAS* ligada ao gene *PAX6*. Como *GAL4* “liga” qualquer gene que tenha *UAS* na região reguladora, as doses extras do gene *PAX6* foram expressas nos tecidos onde a construção se inseriu. Como consequência, foram formadas moscas com até 14 olhos espalhados ao longo do corpo. Construções que utilizaram o cDNA de *PAX6* de vertebrados também produziram moscas com olhos ectópicos, isto é, olhos “extras”, formados em outras partes do corpo. Outros genes atuantes na rota produtora de olhos, também foram sendo descobertos, mas a conservação evolutiva da estrutura, da função e da representatividade *PAX6*, em organismos mais simples (como a planária) e mais complexos (como vertebrados, incluindo os seres humanos) foram demonstradas.

Para fins de comparação, cabe mencionar as diferenças de complexidade entre os genes *PAX6* (seletores, no topo da rede reguladora para a produção de olhos) e o gene estrutural do pigmento **rhodopsin**, a ser expresso nos omatídeos dos olhos compostos da *Drosophila*, na fase final da rede. Essas diferenças são evidentes no tamanho da região codificadora, na existência de sequências reguladoras intercaladas no *PAX6* e nas possibilidades de formação de isoformas, por *splicing* alternativo.

Já o produto do gene *PAX9*, bastante similar ao *PAX1*, parece ter assumido a função de marcador dos locais dos futuros germes dentários, mais tarde na evolução de vertebrados. O produto deste gene, que estabelece complexas associações com produtos de vários outros genes, é ativado no mesênquima da mandíbula de vertebrados pelo *FGF8* (Fator de Crescimento de Fibroblastos 8) e reprimido por *BMP4*, duas conhecidas moléculas reguladoras.

3. Genes *otd/Otx* e *ems/Emx* e a formação das partes do cérebro, genes *Dll/Dlx* e a formação dos apêndices e genes *tinman* e *NK2* e a formação do coração

Apesar das suas diferenças, o desenvolvimento do Sistema Nervoso Central (SNC) da *Drosophila* e do camundongo compartilha mecanismos genéticos similares para a formação do cérebro. Tanto em invertebrados como em vertebrados, a regionalização e a identidade dos neurônios da parte anterior do cérebro são controladas pelos genes *gap otd/Otx* e *ems/Emx*. Já a parte posterior do cérebro envolve os genes *HOX*. Um terceiro domínio intermediário entre ambos, tanto no cérebro

de vertebrados como no de insetos, é caracterizado pela expressão dos genes ortólogos *Pax2/5/8*, sugerindo que os cérebros de protostômios (invertebrados) e deuterostômios (vertebrados) compartilham uma origem evolutiva comum.

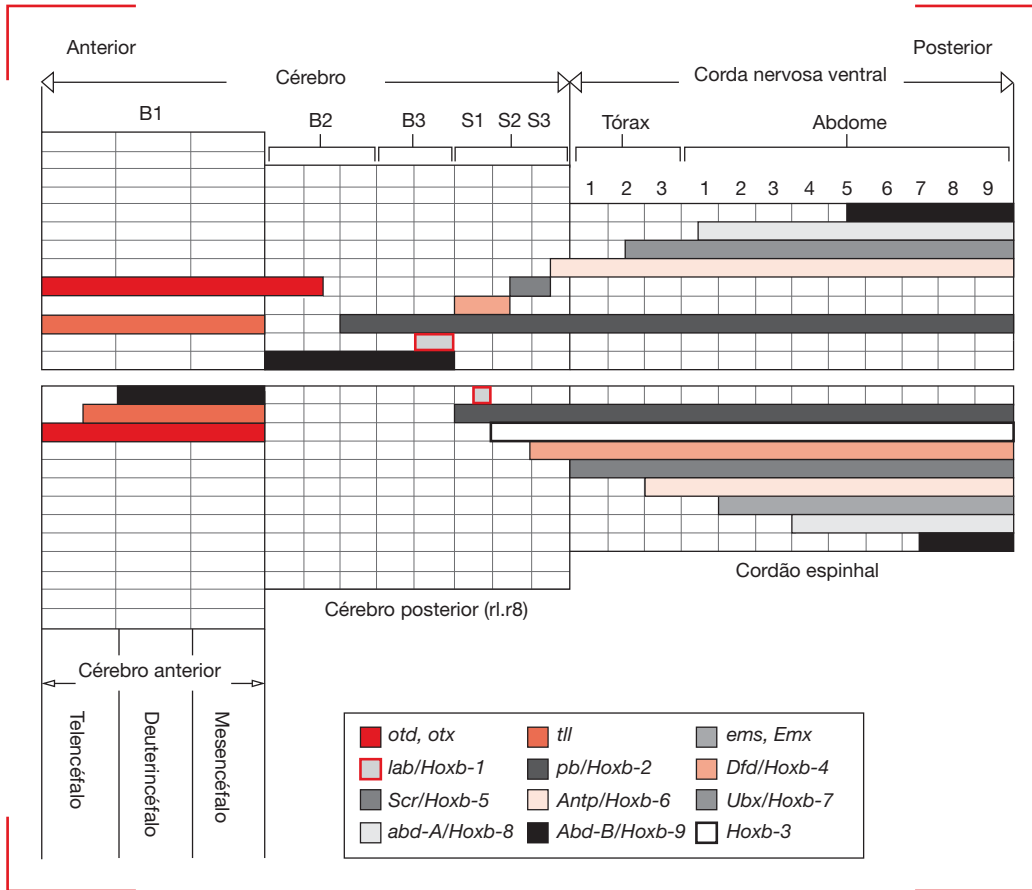
No camundongo e em outros vertebrados, o gene *otd* de *Drosophila* foi duplicado (*Otx1* e *Otx2*), bem como o *ems* (*Emx1* e *Emx2*), e as cópias duplicadas apresentam funções parcialmente sobrepostas.

Experimentos de recuperação “cross-phylum”, isto é, com embriões de *Drosophila* mutantes nulos para *otd* e *ems* transformados com os genes correspondentes de camundongo e de *Xenopus*, tem os defeitos da formação do cérebro da mosca “corrigidos”, mostrando que os membros das famílias *otd/Otx* e *ems/Emx* podem substituir, em parte, o seu gene correspondente no outro filo. A expressão correta de *PAX6* na região neurogênica anterior do embrião, também depende da expressão apropriada de *otd/Otx* e *ems/Emx*, que delimitam o seu campo presuntivo (onde devem se formar olhos e outras estruturas, como partes do focinho de camundongos). Na **Figura 15.8** estão representadas as regiões de expressão de *otd/Otx*, *ems/Emx*, *HOM/HOX* e de outras famílias gênicas, em regiões correspondentes dos embriões de *Drosophila* e camundongos.

São conhecidos, também, genes homólogos envolvidos na dorsoventralização da região neurogênica (ventral em invertebrados e dorsal em vertebrados), como o **sog/Chordin**, mostrando que a inversão do eixo dorsoventral do corpo ocorreu após a separação das linhagens dos protostômios e deuterostômios. Porém, os mecanismos de determinação genética conservados entre invertebrados e vertebrados indicam uma origem monofilética do cérebro. Os genes que atuam a montante e a jusante de *sog/Chordin* também são homólogos, atestando a sua origem comum.

A formação de quaisquer apêndices, tanto em invertebrados como em vertebrados, necessita dos genes *Dll/Dlx*. Em *Drosophila*, uma única cópia de *Dll* é necessária para o crescimento de antenas, patas, asas e outros apêndices, ao passo que, em vertebrados, são conhecidos seis genes derivados de *Dll* por duplicações gênicas. Eles encontram-se agrupados dois a dois em diferentes cromossomos (*Dlx1/2*, *Dlx3/4*, *Dlx5/6*), também possuem um **homeobox** e são mutados em várias síndromes humanas e murinas, cujos fenótipos dos nascidos vivos apresentam malformações dos membros.

Da mesma forma, os diferentes tipos de coração dos vertebrados compartilham com os invertebrados (que não tem coração, mas um vaso ventral pulsante, que bombeia a hemolinfa ao longo do corpo) genes conhecidos como **tinman** em insetos e pelo menos quatro genes *NK2* em vertebrados, surgidos por duplicação gênica. Os genes dessa família também possuem **homeobox**, comprovando a eficiência dos fatores de transcrição que o possuem como membros da caixa de ferramentas envolvidos na morfogênese.

**Figura 15.8**

Principais famílias gênicas envolvidas na especificação do cérebro de *Drosophila* e de camundongo e sua conservação evolutiva.

Com o conhecimento da existência e conservação evolutiva da caixa de ferramentas, o desafio atual da Evo-Devo é identificar os genes que agem a jusante e a montante dos genes mestres seletores e, assim, decifrar cada rede reguladora envolvida na morfogênese. Esforços neste sentido vêm sendo feitos por diversos grupos

de pesquisadores, utilizando várias ferramentas, entre as quais os microarranjos (*microarrays*), que permitem detectar diferenças na expressão de genes em diferentes tecidos e órgãos dos embriões. Outro alvo importante são as sequências reguladoras e como elas surgem e evoluem, gerando novas interações.



Resumo

Genes e famílias gênicas atuantes nos processos de desenvolvimento embrionário, que codificam proteínas reguladoras conservadas do ponto de vista evolutivo, tem sido descobertos nas últimas décadas. A sua conservação, tanto molecular como funcional em organismos

tão diversos como vermes, insetos, roedores e seres humanos, propiciou o surgimento de uma nova abordagem do estudo da evolução morfológica, a chamada Evo-Devo. Esses genes fazem parte da chamada caixa de ferramentas para construção dos corpos dos animais.



Leituras recomendadas

Akam M, Dawson I, Tear G. Homeotic genes and the control of segment diversity. *Development*. 1988;104 Suppl:123-33.

Beachy PA. A molecular view of the Ultrabithorax homeotic gene of *Drosophila*. *Trends Genet*. 1990;6(2):46-51.

Carroll SB. Zebra patterns in fly embryos: activation of stripes or repression of interstripes? *Cell*. 1990;60(1):9-16.

Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*. 2008;134(1):25-36.

Chi N, Epstein JA. Getting your Pax straight: Pax proteins in development and disease. *Trends Genet*. 2002;18(1):41-47.

Gehring WJ. New perspectives on eye development and the evolution of eyes and photoreceptors. *J Hered*. 2005;96(3):171-84.

Gilbert SF. *Developmental biology*. 7th ed. Sunderland: Sinauer; 2003.

Lawrence PA. *The making of a fly: the genetics of animal design*. Oxford: Blackwell; 1992.

Manseau LJ, Schupbach T. The egg came first, of course! Anterior-posterior pattern formation in *Drosophila* embryogenesis and oogenesis. *Trends Genet*. 1989;5(12):400-5.

Nüsslein-Volhard C, Frohnhöfer HG, Lehman R. Determination of anteroposterior polarity in *Drosophila*. *Science*. 1987;238(4834):1675-81.

Panganiban G, Rubenstein JLR. Developmental functions of the Distal-less/Dlx homeobox genes. *Development*. 2002;129(19):4371-86.

Shorrock B. *Drosophila: invertebrate types*. London: Ginn; 1972.

Wolpert L, Jessell T, Lawrence P, Meyerowitz E, Robertson E, Smith J. *Princípios de biologia do desenvolvimento*. 3. ed. Porto Alegre: Artmed; 2008.

Luciane M. P. Passaglia | Arnaldo Zaha

Técnicas de Biologia Molecular

1. Enzimas utilizadas na biologia molecular 333

- 1.1 Enzimas de restrição 333
- 1.2 DNA-ligase 333
- 1.3 Enzimas de modificação do DNA 333
 - 1.3.1 Fosfatase alcalina 334
 - 1.3.2 Polinucleotídeo-quinase 335
 - 1.3.3 Desoxinucleotídeo-terminal-transferase 335
- 1.4 DNA-polimerases 335
 - 1.4.1 DNA-polimerase I 335
 - 1.4.2 Fragmento de Klenow 335
 - 1.4.3 DNA-polimerases termoestáveis 335
 - 1.4.4 Transcriptases reversas 336

2. Princípios básicos de clonagem molecular 336

- 2.1 Construção de moléculas de DNA recombinantes 336
- 2.2 Vetores de clonagem de DNA 336
 - 2.2.1 Plasmídeos 336
 - 2.2.2 O bacteriófago λ como vetor de clonagem 338
 - 2.2.3 Cosmídeos 338
 - 2.2.4 Cromossomos artificiais bacterianos 338
 - 2.2.5 Cromossomos artificiais de levedura 339

- 2.3 Introdução de moléculas de DNA recombinante em células bacterianas 339
 - 2.3.1 Transformação bacteriana 340
 - 2.3.2 Transfecção com DNA de fagos 340
 - 2.3.3 Seleção de transformantes 340
 - 2.3.4 Identificação dos recombinantes 341

3. Bibliotecas de DNA 341

- 3.1 Construção de bibliotecas genômicas 342
- 3.2 Construção de bibliotecas de cDNA 342
- 3.3 Seleção de clones em bibliotecas 344
 - 3.3.1 Hibridização de clones recombinantes com sondas de DNA complementares 344
 - 3.3.2 Seleção imunológica 345

4. Expressão de seqüências clonadas 345

- 4.1 Vetores de expressão procarióticos 345
- 4.2 Vetores de expressão eucarióticos 345
- 4.3 Produção e purificação de proteínas recombinantes 347
 - 4.3.1 Proteínas de fusão e purificação por cromatografia de afinidade 347

5. Técnicas de transformação genética de plantas 348

- 5.1 Transformação de plantas via *Agrobacterium tumefaciens* 348

- 5.2 Transformação de plantas via bombardeamento de DNA 350
- 5.3 Transformação de plantas via eletroporação 350

6. Técnicas de análise de DNA e RNA 350

- 6.1 Eletroforese 350
- 6.2 Hibridização 351
- 6.3 *Southern blot* 351
- 6.4 Sequenciamento de DNA 353
 - 6.4.1 Método de Sanger 353
- 6.5 Sequenciamento automático de DNA 353
- 6.6 Reação em cadeia da polimerase 353
 - 6.6.1 Princípios do método 355
 - 6.6.2 Algumas aplicações da técnica de PCR 355
 - 6.6.2.1 PCR em tempo real quantitativa (qPCR) 355
 - 6.6.2.2 Amplificação de regiões específicas do genoma ou de transcritos 355
 - 6.6.2.3 Análise de polimorfismos de DNA 355
 - 6.6.2.4 Diagnóstico de doenças genéticas 357
 - 6.6.2.5 Medicina forense 357
 - 6.6.2.6 Diagnóstico de doenças infecciosas 357

7.1 Técnicas de análise de genomas e produtos de transcrição em grande escala 357

- 7.1 Sequenciamento de genomas 357
- 7.2 Análise global de transcrição (transcritômica) 357
 - 7.2.1 Etiquetas de sequências expressas 357
 - 7.2.2 Macroarranjos e microarranjos 358
 - 7.2.3 Sequenciamento em grande escala 359

8. Técnicas de análise global de proteínas 359

- 8.1 Técnicas de fracionamento de proteínas associadas à proteômica 359
 - 8.1.1 Eletroforese em gel de poliacrilamida (SDS-PAGE) 359
 - 8.1.2 Eletroforese bidimensional 359
 - 8.1.3 Cromatografia líquida 361
- 8.2 Espectrometria de massas na análise de proteínas 361

No início da década de 1970, uma nova maneira de analisar e explorar as principais moléculas constituintes de uma célula começou a ser praticada. Essas metodologias inovadoras foram chamadas de **tecnologia do DNA recombinante** e, devido a elas, genes específicos puderam ser isolados em quantidade, modificados e reintroduzidos em células e organismos. A manipulação de ácidos nucleicos *in vitro* depende, inicialmente, da disponibilidade de enzimas que possam cortar, ligar e replicar o DNA ou transcrever de forma reversa o RNA. As **enzimas de restrição** cortam moléculas de DNA em fragmentos específicos que podem ser manipulados. As **DNA-ligasas** são enzimas que unem fragmentos de DNA gerados pelas enzimas de restrição. O acesso a vários tipos de enzimas de restrição e de DNA-ligasas possibilitou a transferência sequências específicas de DNA de uma molécula para outra, criando moléculas de DNA não encontradas na natureza, as **moléculas de DNA recombinantes**.

O segundo requerimento para as manipulações genéticas *in vitro* diz respeito ao **pareamento de bases** (ver Capítulo 2), pelo qual ocorre o reconhecimento das moléculas de DNA ou RNA. Técnicas de hibridização, envolvendo sondas de DNA ou RNA complementares, são uma maneira sensível e bastante precisa para se detectar sequências específicas de nucleotídeos. Na tecnologia do DNA recombinante, o pareamento de bases é utilizado na construção de combinações novas de DNA, bem como na detecção de sequências particulares. Essa tecnologia moderna tem fornecido para a medicina e a indústria alternativas eficientes para a produção, em grande escala, de determinadas proteínas, antes disponíveis em quantidades muito reduzidas.

1. Enzimas utilizadas na biologia molecular

1.1 Enzimas de restrição

As enzimas de restrição, também chamadas de **endonucleases de restrição**, reconhecem uma sequência de bases específica na hélice dupla do DNA e cortam ambas as fitas da hélice, em lugares determinados. Elas são indispensáveis na análise da estrutura dos cromossomos, no isolamento de genes e na criação de moléculas novas de DNA que podem ser clonadas.

As enzimas de restrição são encontradas em uma grande variedade de organismos procarióticos e seu papel biológico é o de clivar moléculas de DNA exógenas. O DNA da célula (endógeno) não é degradado, porque os sítios de reconhecimento da sua enzima de restrição são alterados pela adição de um grupamento metílico (CH₃, processo denominado **metilação**) na base nitrogenada de um de seus nucleotídeos em cada uma das fitas. A maior parte dessas enzimas reconhece sequências de quatro a oito pares de bases e hidrolisa uma ligação fosfodiéster em cada fita nesta região. Uma característica marcante de muitos dos **sítios de clivagem** ou **sí-**

tios de restrição é a dupla simetria rotacional, isto é, a sequência de reconhecimento é **palindrômica** e os sítios de clivagem são simetricamente posicionados (ver **Figura 16.1**).

Mais de 3.500 enzimas de restrição, com diferentes especificidades, já foram identificadas e cerca de 150 delas são comumente utilizadas pelos pesquisadores. Sua nomenclatura consiste em uma abreviação de três letras do nome da espécie produtora da enzima (por exemplo: Eco para *Escherichia coli*, Hin para *Haemophilus influenzae*, Hae para *Haemophilus aegyptius*), seguida da designação da linhagem, se necessária, e um número romano (caso mais de uma enzima de restrição seja produzida pela espécie). A Figura 16.1A mostra a especificidade de algumas dessas enzimas.

As enzimas de restrição são utilizadas para cortar moléculas de DNA em fragmentos específicos, facilmente manipulados e que podem conter determinado gene. Um pedaço de DNA, produzido pela ação de uma enzima de restrição, também pode ser clivado, em fragmentos menores, por uma outra enzima.

1.2 DNA-ligase

Para a clonagem de um determinado segmento de DNA, as enzimas de restrição mais utilizadas são aquelas que geram fragmentos com extremidades de fita simples complementares de até quatro nucleotídeos de comprimento, que possuem afinidade uma pela outra e são denominadas **extremidades coesivas**. Fragmentos de DNA contendo extremidades coesivas complementares podem ser unidos pela **DNA-ligase**, que catalisa a formação de uma ligação fosfodiéster entre as duas moléculas (Figura 16.1B). A DNA-ligase requer um grupamento hidroxílico livre na extremidade 3' de uma das cadeias de DNA e um grupamento fosfato na extremidade 5' da outra cadeia. Ela é ativada pela adenilação de um resíduo de lisina no sítio ativo, que, por sua vez, adenila o grupamento fosfato da extremidade 5' do DNA-alvo, ativando-o. Esse grupamento ativado sofre o ataque nucleofílico do grupamento hidroxílico livre da extremidade 3', ocorrendo a formação da ligação fosfodiéster e a liberação de AMP. O bacteriófago T4 e as células eucarióticas utilizam ATP para adenilar a enzima, ao passo que *E. coli* e outras bactérias utilizam NAD⁺ nas suas reações de adenilação (Figura 16.1C).

Existem, também, certas endonucleases de restrição que cortam a molécula de DNA sem criar extremidades coesivas. A DNA-ligase pode unir essas moléculas de DNA com as chamadas **extremidades cegas** (*blunt-ended*) com outras moléculas de DNA que também possuam extremidades desse tipo.

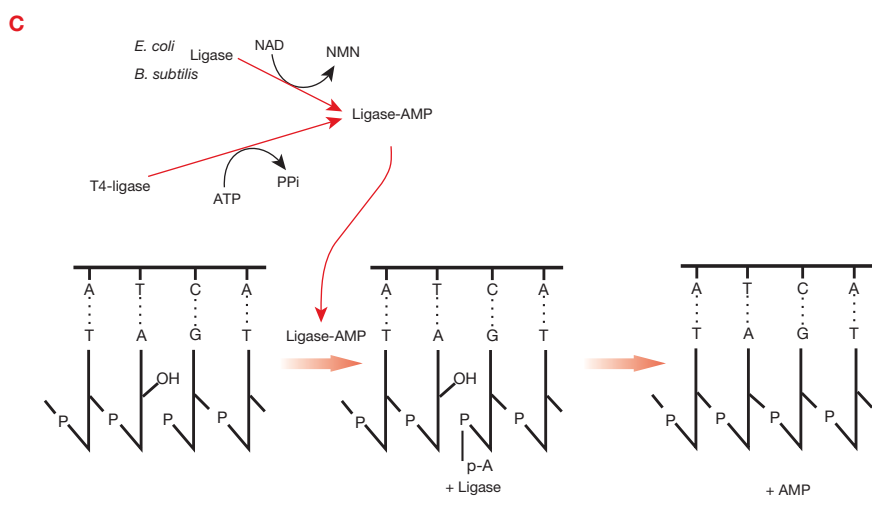
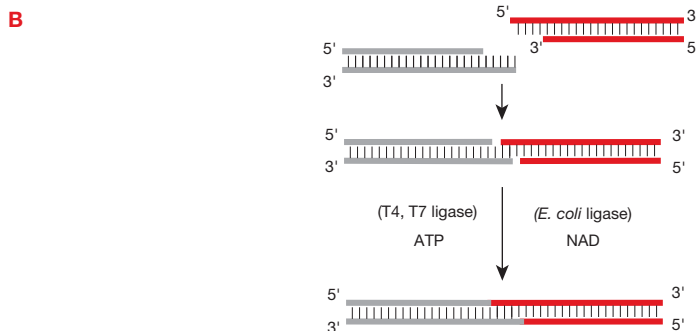
1.3 Enzimas de modificação do DNA

Existem inúmeras enzimas que modificam as moléculas de DNA pela adição ou remoção de grupamentos químicos específicos. As mais importantes são as seguintes:

Figura 16.1

Atividades das enzimas de restrição e da DNA-ligase. (A) Especificidade de algumas enzimas de restrição. As setas indicam o ponto de clivagem na cadeia de DNA. (B) Representação esquemática de uma reação de ligação. (C) Reações que ocorrem durante a ligação de uma cadeia de DNA a outra.

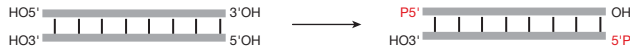
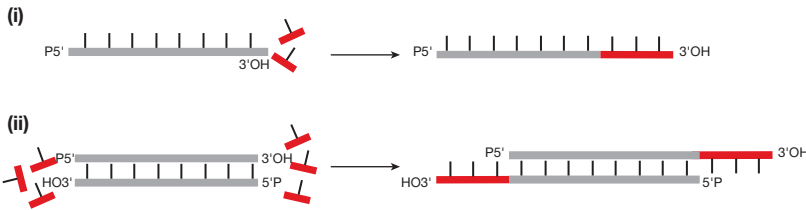
Sequência reconhecida	Nome	Origem
$\begin{array}{c} \downarrow \\ 5' \text{ GGATCC } 3' \\ 3' \text{ CCTAGG } 5' \end{array}$	BamHI	<i>Bacillus amyloliquefaciens</i> H
$\begin{array}{c} \downarrow \\ 5' \text{ GAATTC } 3' \\ 3' \text{ CTTAAG } 5' \end{array}$	EcoRI	<i>E. coli</i> RY13
$\begin{array}{c} \downarrow \\ 5' \text{ GGCC } 3' \\ 3' \text{ CCGG } 5' \end{array}$	HaeIII	<i>Haemophilus aegyptius</i>
$\begin{array}{c} \downarrow \\ 5' \text{ CTGCAG } 3' \\ 3' \text{ GACGTC } 5' \end{array}$	PstI	<i>Providencia stuartii</i>
$\begin{array}{c} \downarrow \\ 5' \text{ CTCGAG } 3' \\ 3' \text{ GAGCTC } 5' \end{array}$	XhoI	<i>Xanthomonas holcicola</i>



1.3.1 Fosfatase alcalina

A fosfatase alcalina é uma enzima produzida em vários órgãos, incluindo ossos, fígado e intestino, sendo encontrada normalmente no sangue de pessoas saudáveis. As

fosfatases alcalinas utilizadas em biologia molecular são produzidas por *E. coli*, tecido intestinal de bovino ou do camarão do ártico. Sua atividade consiste em remover o grupamento fosfato presente em cada extremidade 5' de uma molécula de DNA (Figura 16.2A);

A Fosfatase alcalina**B Polinucleotídeo-quinase****C Desoxinucleotidil-transferase terminal****Figura 16.2**

Reações catalisadas pelas enzimas de modificação do DNA. (A) Fosfatase alcalina remove grupamentos 5'-P. **(B)** Polinucleotídeo-quinase adiciona grupamentos 5'-P. **(C)** Desoxinucleotidil-terminal transferase adiciona desoxirribonucleotídeos às extremidades 3' de cadeias nucleotídicas de (i) fita simples, ou (ii) de fita dupla.

1.3.2 Polinucleotídeo-quinase

A polinucleotídeo-quinase é extraída das células de *E. coli* infectadas com o bacteriófago T4, que possui o gene codificador da enzima em seu genoma. Essa enzima tem o efeito inverso da fosfatase alcalina, adicionando grupamentos fosfatos nas extremidades 5' livres de uma molécula de DNA (Figura 16.2B);

1.3.3 Desoxinucleotidil-terminal-transferase

A desoxinucleotidil-terminal-transferase, extraída do tecido de timo bovino, adiciona um ou mais desoxirribonucleotídeos às extremidades 3'-OH de uma molécula de DNA (Figura 16.2C).

1.4 DNA-polimerases

As DNA-polimerases são enzimas que sintetizam uma nova fita de DNA, complementar a uma fita-molde de DNA ou RNA. A maioria das DNA-polimerases atua no processo de replicação do DNA (ver Capítulo 6) e funciona somente se o molde possuir uma região de fita dupla, que irá atuar como iniciador para a polimerização. Quatro tipos de DNA-polimerases, discutidos nas próximas Seções, são utilizados rotineiramente em biologia molecular.

1.4.1 DNA-polimerase I

A DNA-polimerase I liga-se a uma curta região de fita simples de uma molécula de DNA de fita dupla e, então,

sintetiza uma fita completamente nova, degradando a fita existente à medida que ela prossegue na polimerização. A DNA-polimerase I é uma enzima com dupla atividade, pois polimeriza e degrada o DNA.

1.4.2 Fragmento de Klenow

Essa enzima surge da modificação da DNA-polimerase I, devido à remoção de sua porção com atividade nucleolítica (os 323 primeiros aminoácidos da enzima). A enzima resultante desse processo, denominada fragmento de Klenow, retém a capacidade de polimerização, mas é incapaz de degradar o DNA. Ou seja, ela é capaz de sintetizar uma fita de DNA complementar a partir de um molde de fita simples, mas não consegue continuar a síntese se a lacuna da fita simples estiver preenchida. A principal aplicação do fragmento de Klenow está no sequenciamento de DNA pelo método de Sanger.

1.4.3 DNA-polimerases termoestáveis

A **Taq-DNA-polimerase** é uma enzima do tipo da DNA-polimerase I, extraída da bactéria *Thermus aquaticus*. Esse organismo vive em fontes termais e muitas de suas enzimas, incluindo a Taq-DNA-polimerase, são termoestáveis (resistentes à desnaturação pelo calor). Essa característica especial da Taq-DNA-polimerase a torna adequada para utilização em metodologias como a PCR (Seção 6.6), que envolve etapas de aquecimento a temperaturas de mais de 90°C.

Uma outra versão, mais robusta, de DNA-polimerase termoestável, também utilizada em PCR, é a **Pfu-DNA-polimerase**, uma enzima encontrada no organismo hipertermófilo *Pyrococcus furiosus*. Essa enzima tem a menor taxa de erro conhecida na síntese de DNA, sendo, por isso, usada em aplicações que exigem grande fidelidade. A enzima catalisa a incorporação de nucleotídeos na direção 5'→3' na presença de magnésio a 70-80°C. A *Pfu*-DNA-polimerase tem atividade exonucleásica 3'→5', o que lhe permite corrigir erros introduzidos durante a polimerização. Assim, essa DNA-polimerase tem capacidade para efetuar a revisão da polimerização (*proofreading*).

Portanto, a sua taxa de erro, que é inferior a $2,6 \times 10^{-6}$ por nucleotídeo em cada ciclo, é cerca de 7 a 10 vezes menor do que a da *Taq*-DNA-polimerase (que não é capaz de efetuar a revisão).

1.4.4 Transcriptases reversas

As transcriptases reversas são DNA-polimerases RNA-dependentes envolvidas na replicação de vários tipos de vírus, cujos genomas são constituídos de RNA. Como o próprio nome indica, elas fazem uma transcrição reversa, utilizando RNA como molde para sintetizar uma fita de **DNA complementar (cDNA)**, sendo muito úteis na construção de bibliotecas de cDNA a partir de populações específicas de mRNA (ver Seção 3).

2. Princípios básicos de clonagem molecular

2.1 Construção de moléculas de DNA recombinantes

O principal resultado das aplicações da tecnologia do DNA recombinante é o isolamento e a propagação, em um organismo, de moléculas idênticas de DNA. Essa abordagem, chamada de **clonagem molecular**, envolve: (1) uma **molécula de DNA recombinante** é formada pela ligação de um inserto de DNA, originado da clivagem do DNA de interesse, a uma outra molécula de DNA, chamada de **vetor** ou **veículo de clonagem**. Esse vetor de clonagem, ou simplesmente **vetor**, é assim chamado porque é ele que transporta o **inserto** de DNA para o interior de uma célula hospedeira, onde ele poderá ser replicado; (2) a molécula de DNA recombinante (inserto e vetor unidos) é introduzida dentro de uma célula hospedeira apropriada. O processo de introdução de DNA em células é chamado de **transformação** (ver Seção 2.3.1). A célula hospedeira, contendo uma única molécula de DNA recombinante, divide-se várias vezes, formando uma colônia de células. Cada célula da colônia possui, pelo menos, uma cópia da molécula recombinante. Essas células são chamadas de **transformantes** ou **células transformadas**. Essas transformantes podem ser distinguidas das células que não receberam a molécula de

DNA recombinante pela presença de um gene marcador, presente no DNA do vetor de clonagem.

2.2 Vetores de clonagem de DNA

Uma molécula de DNA necessita apresentar várias características para ser capaz de atuar como um vetor da clonagem gênica. Uma das características principais é a capacidade de replicar dentro da célula hospedeira, de maneira que numerosas cópias da molécula de DNA recombinante possam ser produzidas e transmitidas para as células-filhas. Para isso, os vetores devem conter sequências permitindo a sua **replicação autônoma** dentro da célula hospedeira. Além disso, os vetores mais utilizados possuem um **sítio único de clonagem**, um sítio de restrição que não se repete na molécula, permitindo a inserção sítio-específica de uma outra molécula de DNA (inserto). Alguns vetores contêm sítios únicos para várias endonucleases de restrição posicionados lado a lado, em um segmento denominado **sítio múltiplo de clonagem**.

Um vetor de clonagem precisa ser relativamente pequeno, em forma ideal com menos que 10 kb de tamanho, uma vez que moléculas maiores tendem a se quebrar durante os processos de purificação e também são mais difíceis de serem manipuladas. Dois tipos de moléculas de DNA que satisfazem esses critérios podem ser encontrados em células bacterianas: os plasmídeos e os DNAs de fagos.

2.2.1 Plasmídeos

Plasmídeos geneticamente modificados foram os primeiros vetores desenvolvidos, sendo ainda muito utilizados para a clonagem. Esses vetores plasmídeos são derivados de plasmídeos que ocorrem naturalmente em bactérias e alguns organismos eucarióticos unicelulares, como as leveduras. Plasmídeos são moléculas de DNA circular, de fita dupla, extracromossômicas e que possuem capacidade de replicação autônoma (ver Capítulo 9). Os plasmídeos utilizados como vetores foram bastante modificados pela engenharia genética e muitos estão agora disponíveis com uma ampla variedade de características úteis. Por exemplo, embora apenas uma cópia de um plasmídeo entre em uma célula bacteriana hospedeira, uma vez dentro dela, alguns plasmídeos aumentam o seu número para centenas de cópias. Além disso, com frequência, eles transportam genes que são **marcadores genéticos auxotróficos** ou que conferem resistência aos antibióticos, como a ampicilina e a tetraciclina. Esses genes são utilizados como forma de distinguir as células hospedeiras que receberam o vetor daquelas que não receberam.

Um dos plasmídeos mais utilizados em biologia molecular é o **pUC18 (Figura 16.3A)**, que possui características úteis para um vetor. Os plasmídeos componentes da série pUC de vetores possuem, entre outras sequências, o gene que codifica resistência à ampicilina, a origem de replicação do DNA plasmidial e o gene *lacZ'*,

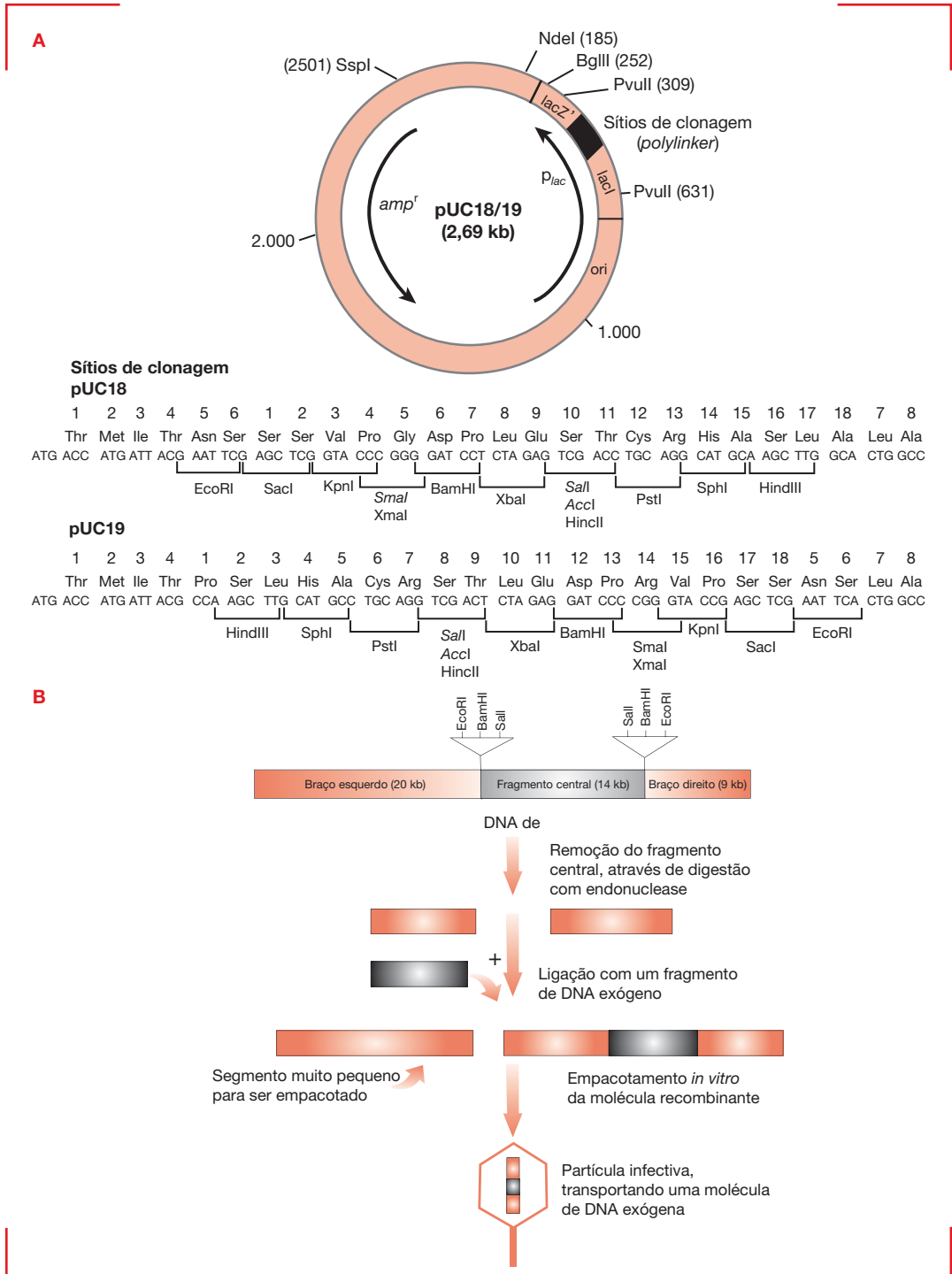


Figura 16.3

Mapas dos plasmídeos pUC18/19 e esquema da utilização de um derivado do bacteriófago λ como vetor de clonagem. (A) Mapas dos plasmídeos pUC18/19 mostrando os sítios de clivagem para diversas enzimas de restrição (os números indicam a posição no mapa) e os sítios únicos para enzimas de restrição localizados na região de clonagem. A seta em ■ indica a posição e orientação do gene que confere resistência ao antibiótico ampicilina (*amp^r*) e a seta em ■ indica a orientação de parte do gene que codifica a β -galactosidase (*lacZ'*). O repressor LacI é codificado pelo gene *lacI* e *ori* indica a origem de replicação do plasmídeo. (B) Esquema da utilização do vetor derivado de bacteriófago λ como vetor de clonagem. O processo de empacotamento seleciona moléculas de DNA contendo insertos de tamanho adequado para serem empacotados. O desenho mostra a representação esquemática do vetor λ EMBL4. O fragmento central (*stuffer*), de 14 kb, pode ser substituído por fragmentos de DNA de 9 a 24 kb.

que codifica a subunidade α da β -galactosidase (β -gal) de *E. coli*. Eles também apresentam vários sítios de clonagem inseridos em uma região apropriada do plasmídeo, que não inativa esse gene. A inserção de um fragmento de DNA em um destes sítios de clonagem causa a interrupção do gene *lacZ*, levando à perda da função da β -gal.

Os plasmídeos pUC18 e pUC19 possuem a mesma região contendo os múltiplos sítios de clonagem, mas em orientação oposta. As bactérias contendo os plasmídeos recombinantes são selecionadas pela coloração branca das colônias, ao passo que aquelas que receberam o **vetor parental ou selvagem** (sem um inserto clonado) apresentam colônia de cor azul, resultado da ação de uma β -gal ativa. Esses plasmídeos possuem um tamanho de 2.643 pb e estão presentes em um alto número de cópias por célula.

2.2.2 O bacteriófago λ como vetor de clonagem

Os fagos pertencem à categoria dos elementos genéticos móveis (ver Capítulo 9) e, por serem capazes de replicação autônoma, podem ser utilizados como vetores de clonagem. Um dos fagos mais usado como vetor em experimentos de clonagem molecular é o **bacteriófago λ** . O DNA do bacteriófago λ foi modificado pela tecnologia do DNA recombinante, originando alguns dos mais versáteis vetores para a clonagem. Segmentos de DNA inseridos em um genoma viral podem ser introduzidos dentro de uma célula hospedeira, como uma partícula viral, por meio de um processo denominado **transfecção** (ver Seção 2.3.2). *In vivo*, um fago infectivo, é composto por um capsídeo e uma cauda (formados por proteínas do fago) e de DNA. As proteínas do fago são produzidas separadamente, a partir do DNA viral que está replicando, e, somente após, é que o capsídeo e o DNA são reunidos para formar a partícula viral.

A reação natural de empacotamento do DNA do fago necessita, como substrato, longas moléculas lineares de DNA, chamadas de concatêmeros, que são formadas por uma série de genomas virais completos unidos (ver Capítulo 6). Esses genomas são separados pela clivagem de sequências específicas, denominadas **sítios ou sequências *cos*** (coesivas). As sequências *cos*, de 12 pb, devem ser repetidas a cada 35 a 50 kb, para que o empacotamento do DNA dentro de um capsídeo de fago ocorra. Se a distância entre dois sítios *cos* não for apropriada, o DNA não será empacotado na forma de uma partícula infectiva.

Existe uma porção do genoma viral que não é necessária para que o fago infecte uma célula e replique-se normalmente, portanto, este segmento, chamado de ***stuffer*** (do inglês *to stuff* = rechear), pode ser substituído por um DNA exógeno. Devido a isso, derivados do bacteriófago λ , destinados à clonagem de insertos de DNA, puderam ser construídos. Um desses derivados, λ EMBL4 (Figura 16.3B), contém dois sítios de clivagem para cada

uma das enzimas de restrição EcoRI, Sall e BamHI, flanqueando um *stuffer* de 14 kb, que, por exemplo, pode ser removido após a clivagem com a enzima BamHI. Os dois segmentos de DNA remanescentes (braços) compreendem apenas 60% do genoma viral completo. Essa quantidade é muito pequena e não pode ser empacotada, a menos que moléculas de DNA adicionais sejam ligadas a esses fragmentos, originando genomas recombinantes. Os tamanhos desses genomas recombinantes podem variar de 75 a 105% do comprimento do genoma viral normal, permitindo que insertos com diversos tamanhos (de 9 até 23 kb, dependendo do vetor utilizado) sejam clonados. As partículas de fagos contendo moléculas recombinantes podem ser introduzidas na célula hospedeira por infecção.

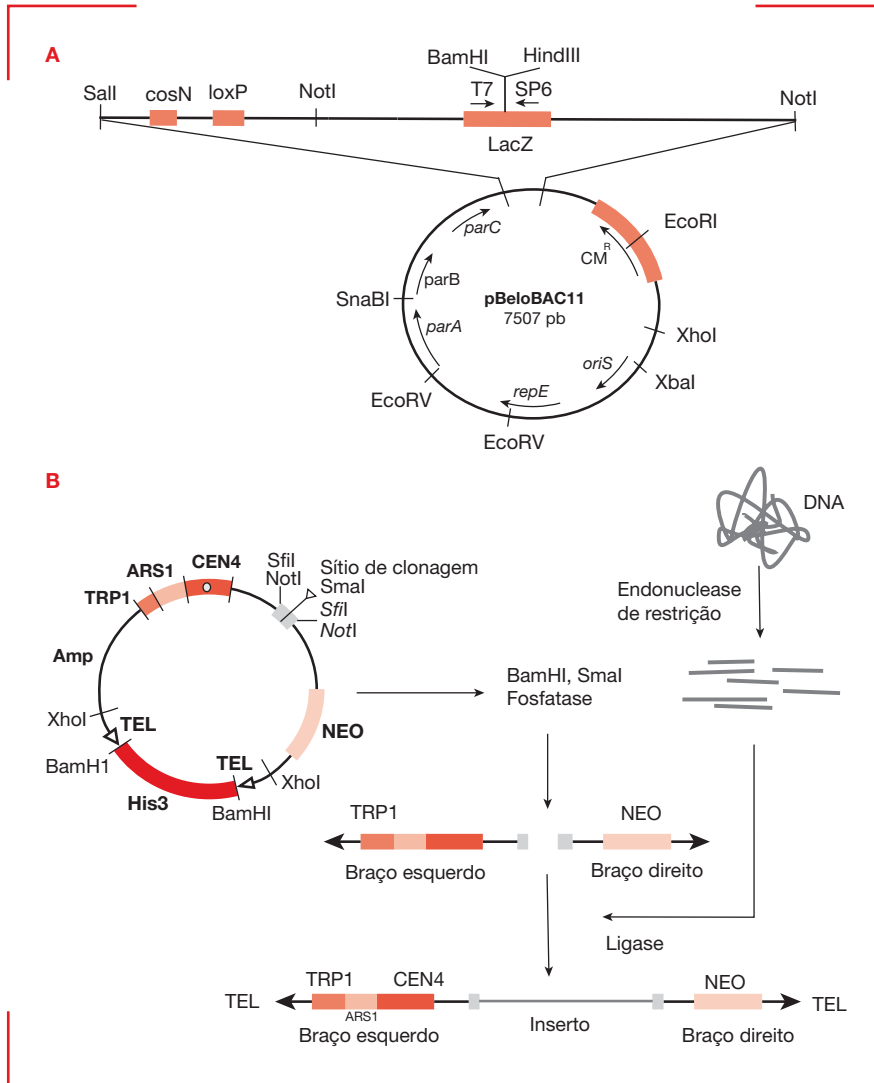
2.2.3 Cosmídeos

Plasmídeos utilizados como vetores são adequados à clonagem de fragmentos de DNA, apresentando uma variação de tamanho desde poucas centenas até 9.000 pb. Para a clonagem de fragmentos maiores devem ser usados preferencialmente os **cosmídeos**, que são plasmídeos bacterianos possuindo uma origem de replicação, uma marca de resistência a antibiótico, um ou mais sítios de clonagem e as sequências *cos* do bacteriófago λ . A clonagem de fragmentos grandes de DNA em um cosmídeo produz recombinantes, que podem ser empacotados nos capsídeos do bacteriófago λ e utilizados para infectar células de *E. coli*. Como o reconhecimento de sítios *cos* requer que segmentos de DNA, de pelo menos 35 a 50 kb, sejam inseridos entre dois sítios *cos* adjacentes, os cosmídeos resultam em excelentes vetores para a clonagem de fragmentos de DNA relativamente grandes.

Uma variação dos cosmídeos são os **fosmídeos**, que contêm a origem de replicação do **plasmídeo F** de *E. coli* (ver Capítulo 9), além das sequências *cos* do bacteriófago λ . A diferença para os cosmídeos é que, devido à origem de replicação de F, o número de cópias de fosmídeos por célula hospedeira é menor, o que é uma vantagem na clonagem de fragmentos maiores, pois diminui a chance de ocorrência de rearranjos por recombinação entre os insertos.

2.2.4 Cromossomos artificiais bacterianos

O sistema de clonagem em **cromossomos artificiais bacterianos (BAC, de *bacterial artificial chromosomes*)** (Figura 16.4A) também se baseia no plasmídeo F de *E. coli*. Fragmentos de DNA maiores que 300 kb podem ser clonados no vetor e mantidos estávelmente na bactéria hospedeira. Pela alta eficiência de clonagem, fácil manipulação e estabilidade do fragmento clonado, o sistema de clonagem em BAC tem sido adotado para construção de bibliotecas genômicas (ver Seção 3.1) de organismos com genomas complexos. O sistema BAC foi utilizado, por exemplo, na construção de bibliotecas usadas no sequenciamento dos genomas humano, da planta *Arabidopsis thaliana*, do díptero *Drosophila melanogaster* e de outros organismos.

**Figura 16.4**

Exemplos da utilização de pBAC e pYAC. (A) Mapa físico de pBeloBAC11. O sítio de clonagem está localizado no fragmento do gene *lacZ*, o que permite a identificação de clones recombinantes pela coloração da colônia branca ou azul. Os genes *oriS* e *repE* estão envolvidos na replicação unidirecional do Fator F, ao passo que os genes *parA* e *parB* mantêm o número de cópias em torno de uma a duas por cromossomo bacteriano. O sítio *cosN* do bacteriófago λ fornece uma posição fixa para clivagem específica com a terminase do bacteriófago λ . O sítio *loxP* do bacteriófago P1 pode ser utilizado de forma semelhante. Os promotores de Sp6 e T7 possibilitam a produção de sondas de RNA. As flechas indicam a direção da transcrição dos genes. Cm^R é o gene de resistência a cloranfenicol. **(B)** Representação esquemática de construção de recombinantes utilizando um vetor pYAC. O vetor é clivado com as enzimas BamHI e SmaI e posteriormente tratado com fosfatase alcalina. Os braços esquerdo e direito do vetor são ligados aos fragmentos de DNA e as moléculas recombinantes são introduzidas na levedura. CEN, centrômero; His3, marcador genético entre as sequências TEL – sua perda indica linearização eficiente; NEO, marcador de seleção; TEL, sequências teloméricas; TRP1, marcador de seleção. Colônias coloridas indicam a presença de β -gal ativa, capaz de degradar o substrato (X-gal). Colônias brancas indicam que a enzima não foi sintetizada (ver Seção 2.3.4).

2.2.5 Cromossomos artificiais de levedura

No final da década de 1980, foi desenvolvida uma metodologia que permitia a clonagem de fragmentos mais longos de DNA (de 100 a 1.000 kb), utilizando como vetores **cromossomos artificiais de levedura (YACs)**, de *yeast artificial chromosome*, Figura 16.4B), desenvolvidos para a levedura *Saccharomyces cerevisiae*. Os YACs são plasmídeos (pYACs) que contêm marcadores genéticos para seleção (TRP1 e NEO) e componentes funcionais de um cromossomo eucariótico, isto é: elementos para a replicação autônoma (ARS), um centrômero (CEN) para segregação correta durante a divisão celular e os telômeros (TEL), necessários para estruturação e estabilização de moléculas lineares. Os pYACs têm sido muito importantes na construção de bibliotecas genômicas de organismos com genomas complexos, por exemplo, os mamíferos. As bibliotecas são construídas e mantidas na levedura *S. cerevisiae*. Entretanto, um dos problemas com as bibliotecas de YACs (ou pYACs) é que muitos dos seus recombinantes são quiméricos, isto é, possuem insertos

compostos por fragmentos de DNA que não são colineares com o DNA genômico. Além disso, os fragmentos de DNA clonados podem sofrer rearranjos por recombinação.

2.3 Introdução de moléculas de DNA recombinante em células bacterianas

Após a clonagem de um gene em qualquer vetor, muitos pesquisadores analisam suas características por meio da reintrodução desse gene em diferentes tipos celulares. Além do propósito primário da transformação, que é a introdução da molécula de DNA recombinante em uma célula em que ela será replicada, produzindo muitas cópias da molécula, existem inúmeras outras razões para a realização do processo de transformação, dentre as quais: (1) os elementos genéticos, responsáveis pela regulação da expressão do gene, podem ser determinados pelas mutações nesse gene e pela análise subsequente de sua atividade sob condições fisiológicas variadas; (2) os efeitos ocasionais de um gene sobre o crescimento celular

podem ser investigados pela produção de linhagens celulares que, ao expressarem o gene, definem o seu fenótipo; (3) podem ser produzidas linhagens celulares para superexpressão do gene, permitindo a purificação de seu produto para caracterização bioquímica ou para produção em grande escala para utilização industrial. Esses e outros objetivos requerem a capacidade de introduzir, de maneira eficiente, o DNA para o interior de uma célula. Inúmeros métodos para a introdução de DNA em células hospedeiras já foram descritos e testados. Nesta seção, serão vistos brevemente os métodos mais conhecidos e eficazes de introdução de DNA em células bacterianas.

2.3.1 Transformação bacteriana

Muitos métodos para transformação bacteriana tiveram base nas observações de Mandel e Higa, em 1970, que demonstraram que bactérias tratadas com soluções geladas de cloreto de cálcio e, logo após, sendo brevemente aquecidas, podiam ser transfectadas com DNA do bacteriófago λ . O mesmo método foi utilizado para a transformação de bactérias com DNA plasmidial e DNA cromossômico de *E. coli*. Em princípio, o tratamento induz a um estado de “**competência**” transitório nas bactérias receptoras, durante o qual elas se tornam permeáveis a moléculas de DNA livres, provenientes de uma grande variedade de fontes. As variações dessa técnica básica, que têm sido descritas, destinam-se a aumentar a eficiência de transformação de diferentes linhagens bacterianas com plasmídeos e outros vetores. Bactérias tratadas conforme o protocolo original de Mandel e Higa rendem de 10^5 - 10^6 colônias transformadas por micrograma de DNA plasmidial. Essa eficiência pode ser aumentada de 100 a 1.000 vezes pela exposição de linhagens de *E. coli*, geneticamente mais suscetíveis à transformação, a combinações de cátions divalentes por períodos de tempos mais prolongados e tratando as células com diversos compostos, como os agentes redutores. As melhorias nas eficiências de transformação têm ocorrido fundamentalmente como consequências de experimentações empíricas.

Células bacterianas podem, de forma alternativa, serem transformadas por **eletroporação**. Esse método, desenvolvido a princípio para introdução de DNA em células eucarióticas, tem sido utilizado em *E. coli* e bactérias que não se prestam ao tratamento convencional com cloreto de cálcio, resultando em um aumento na eficiência de transformação, que pode chegar a 10^9 - 10^{10} transformantes por micrograma de DNA. O método baseia-se na introdução de DNA em células previamente tratadas com uma solução de baixa concentração salina para reduzir a força iônica da suspensão celular. Essas células, suspensas em uma solução de glicerol 10%, são submetidas a um campo elétrico de alta voltagem, que, ao abrir pequenos poros na membrana celular, propicia a entrada do DNA na célula através dessas aberturas. As maiores eficiências de transformação têm sido obtidas com uma combinação de campo de força, duração do pulso elétrico e concentração do DNA, sem que resultem em uma alta frequência de morte celular.

2.3.2 Transfecção com DNA de fagos

A introdução de DNAs de fagos derivados do bacteriófago λ em determinadas linhagens de *E. coli* baseia-se na capacidade que o fago apresenta de ligar-se aos receptores proteicos localizados na membrana externa da bactéria (adsorção). Esses receptores são codificados pelo gene bacteriano *lamB* e, normalmente, utilizados para o transporte de maltose para o interior da célula. Devido ao fato da síntese desses receptores ser induzida por maltose e reprimida por glicose, as culturas de bactérias que irão ser infectadas pelo fago, no processo denominado **transfecção**, devem ser cultivadas em um meio contendo maltose e sem glicose. A adsorção das partículas virais aos receptores ocorre de forma rápida e eficiente (em poucos minutos) tanto em temperatura ambiente como a 37°C . No entanto, a penetração do DNA do fago na célula e os eventos subsequentes, que levarão ao ciclo lítico (ver Capítulo 9), não ocorrem eficientemente em temperatura ambiente, o que impede a formação das **placas de lise** (zonas no meio de cultura desprovidas de células, as quais foram lisadas pela infecção viral) quando as culturas são mantidas nessas condições.

2.3.3 Seleção de transformantes

A transformação de células bacterianas competentes é um processo ineficiente, por mais cuidadosa que tenha sido a preparação das células. Embora a exposição de bactérias competentes a uma nanograma de um vetor plasmidial, do tipo pUC18, por exemplo, possa resultar em 1.000-10.000 transformantes, isso representa a entrada de apenas 0,01% de todas as moléculas disponíveis nas células. Além disso, 10.000 transformantes correspondem a uma proporção muito pequena do número total de células que está presente em uma cultura competente (da ordem de 10^8 a 10^{12} células por mililitro). Essa observação sugere a existência de alguma maneira eficiente para se distinguir uma célula que recebeu um plasmídeo dentre as milhares que não o receberam. A solução é utilizar uma **marca de seleção** contida no plasmídeo.

Uma marca de seleção pode ser um gene que confira à célula transformada uma nova característica, ausente na célula não transformada. Um bom exemplo de marca de seleção é o gene que confere resistência ao antibiótico ampicilina (*amp^R*) do pUC18. Após um experimento de transformação com pUC18, somente as células de *E. coli*, que adquiriram um plasmídeo, serão resistentes ao antibiótico e capazes de formar colônias em uma placa de cultivo com meio sólido, contendo ampicilina. As células não transformadas, que serão sensíveis ao antibiótico (*amp^S*), morrerão e, portanto, não produzirão colônias em tal meio seletivo.

A maioria dos plasmídeos utilizados como vetores de clonagem apresenta pelo menos um gene que confere às células hospedeiras resistência a antibiótico, com a seleção dos transformantes sendo executada por meio do cultivo das bactérias transformadas em um meio sólido contendo o antibiótico relevante. No entanto, as enzimas

que conferem resistência ao antibiótico devem estar presentes nas células transformadas no momento em que elas são expostas ao antibiótico no meio de cultura. Isso pressupõe que os genes que codificam essas enzimas já tenham sido transcritos e seus mRNAs traduzidos. Dessa forma, imediatamente após o choque térmico e antes das células serem submetidas à ação seletiva, elas devem ser colocadas, por um curto período de tempo, em um pequeno volume de meio líquido sem o antibiótico. A replicação do plasmídeo poderá, então, ser iniciada, juntamente com a expressão de seus genes. Assim, quando as células forem expostas à presença do antibiótico elas já sintetizarão, em quantidade suficiente, as enzimas que conferem resistência a ele, sobrevivendo à ação seletiva.

2.3.4 Identificação dos recombinantes

O plaqueamento em meio seletivo, discutido na seção anterior, possibilita que as células transformantes sejam distinguidas das não transformantes. A próxima etapa é determinar quais das colônias transformadas são constituídas por células que receberam moléculas de DNA recombinantes (vetor com inserto) e quais contêm moléculas de vetor que se religaram (vetor vazio), ou nem foram clivadas (vetor selvagem). A maioria dos plasmídeos utilizados como vetores de clonagem apresenta os sítios de clonagem em regiões que, uma vez tendo recebido um inserto de DNA, essa inserção provoca a inativação de um dos genes presentes no vetor. Os recombinantes podem, então, ser identificados, porque a característica codificada pelo gene inativado não estará mais presente nas células hospedeiras. Os princípios gerais da **inativação por inserção** podem ser ilustrados por meio de um experimento de clonagem típico, utilizando o plasmídeo pBR322 como vetor.

O **pBR322** é um plasmídeo com 4.362 pb, portador dos genes que conferem resistência à ampicilina e à tetraciclina. A inativação de um destes genes, pela inserção de um DNA exógeno em um dos vários sítios únicos de clonagem do plasmídeo, é uma excelente forma de selecionar recombinantes pelo método de inativação por inserção. A enzima BamHI, por exemplo, cliva o plasmídeo apenas uma vez, dentro da região que contém o agrupamento de genes, conferindo resistência à tetraciclina. Uma molécula de pBR322 recombinante, contendo um inserto clonado no sítio de BamHI, não irá conferir resistência à tetraciclina a sua célula hospedeira, uma vez que um dos genes necessários para expressar essa resistência está interrompido pelo fragmento inserido. As células contendo o pBR322 recombinante ainda apresentam resistência à ampicilina (*amp^R*), mas serão sensíveis à tetraciclina (*tet^S*). Dessa forma, a identificação das colônias que contêm plasmídeos recombinantes é feita por meio da inoculação simultânea de cada colônia em meios sólidos (réplicas) contendo ampicilina ou tetraciclina. Aquelas colônias que se multiplicarem no meio com ampicilina, mas forem incapazes de se multiplicar no meio com tetraciclina, carregam plasmídeos recombinantes.

Uma outra forma de identificar recombinantes utiliza o sistema de identificação visual, por meio da coloração exibi-

da pela colônia transformante. Esse é o caso das clonagens que usam vetores derivados do pUC18, conforme mencionado no início deste capítulo (Seção 2.2.1 e **Figura 16.5**). Os vetores que utilizam essa estratégia possuem um gene que codifica resistência à ampicilina e um gene, *lacZ'*, que codifica uma parte da enzima β -galactosidase (β -gal). A clonagem em pUC18 envolve a inativação por inserção do gene *lacZ'*, e os recombinantes são identificados pela sua incapacidade de sintetizar uma β -gal ativa.

A β -galactosidase é a enzima responsável pela quebra da lactose em glicose e galactose. Ela é codificada pelo gene *lacZ*, presente no cromossomo de *E. coli*. Algumas linhagens de *E. coli*, entretanto, possuem um gene *lacZ* modificado, que não apresenta um segmento, referido como *lacZ'*, codificando uma parte da β -gal (chamada de subunidade α). Esses mutantes só podem sintetizar uma β -gal funcional quando possuem um plasmídeo, como o pUC18, que tem a sequência codificadora da parte da β -gal faltante (a subunidade α), em um fenômeno denominado **α -complementação**. Dessa forma, a seleção das colônias contendo plasmídeos recombinantes é feita em um meio sólido contendo ampicilina e 5-bromo-4-cloro-3-indolil- β -D-galactopiranosídeo (X-gal), um análogo da lactose que, quando degradado por uma β -gal ativa, produz um composto que confere à colônia uma coloração azul escura.

Os plasmídeos da série pUC, como pUC18 e pUC19, também necessitam da presença de IPTG (isopropil-tiogalactosídeo), um análogo da lactose não suscetível à degradação pela β -gal, que é utilizado como indutor da transcrição do gene *lacZ'*, o qual é adicionado ao meio de cultura. Assim, as colônias transformantes que receberam o vetor selvagem (não clivado) ou religado (vazio) produzirão uma β -gal ativa e terão coloração azul. As bactérias transformantes que receberam plasmídeos recombinantes terão o gene *lacZ'* interrompido pelo inserto, serão incapazes de sintetizar uma β -gal ativa e serão brancas.

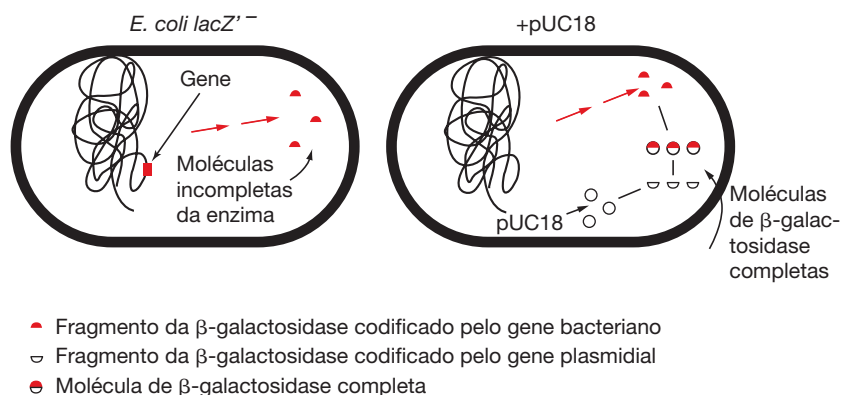
3. Bibliotecas de DNA

A clonagem molecular tem demonstrado ser uma técnica muito útil para o isolamento de fragmentos de DNA de interesse do genoma de um organismo. Os processos de clonagem e isolamento desses fragmentos começam com a construção de uma **biblioteca de DNA**. Uma biblioteca de DNA consiste em todas as moléculas recombinantes, geradas pela ligação dos fragmentos de DNA componentes do genoma de um organismo de interesse, que foi previamente fragmentado, em um vetor apropriado. Entre os vetores utilizados neste processo podemos destacar: bacteriófago λ , plasmídeo, cosmídeo, fosmídeo, BAC e YAC (ver Seção 2.2). O conjunto de clones que compõe uma biblioteca deve ser introduzido em células hospedeiras para que cada recombinante possa replicar *in vivo*. Diferentes tipos de bibliotecas podem ser construídas, dependendo da natureza do inserto, do tipo de vetor utilizado e da finalidade a que se destina cada biblioteca.

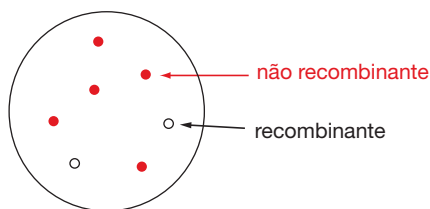
Figura 16.5

A base da inativação por inserção do gene *lacZ'* contido no plasmídeo pUC18. (A) Os genes da bactéria e do plasmídeo se complementam para produzir uma molécula de β -galactosidase funcional. (B) Os plasmídeos recombinantes são selecionados pelo plaqueamento em meio sólido contendo X-gal e IPTG. Colônias coloridas indicam a presença de β -gal ativa, capaz de degradar o substrato (X-gal). Colônias brancas indicam que a enzima não foi sintetizada (ver Seção 2.3.4).

A A função do gene *lacZ'*



B Seleção de recombinantes de pUC8



3.1 Construção de bibliotecas genômicas

As **bibliotecas genômicas** são obtidas pela clonagem de todo o genoma de um organismo. Essas bibliotecas são frequentemente construídas utilizando vetores que suportam a clonagem de grandes insertos, como aqueles citados na Seção 2.2. Em consequência disso, um número menor de clones é requerido para obter uma amostra representativa do genoma. Insertos maiores garantem, também, que qualquer gene completo, incluindo sua sequência codificadora e suas sequências reguladoras (ver Capítulo 4), possa ser obtido de um único clone ou de um pequeno número de clones diferentes.

A estratégia geral para a clonagem de um DNA genômico, em um vetor derivado do bacteriófago λ está representada na **Figura 16.6**. O inserto de DNA é preparado por meio da clivagem parcial do DNA genômico, com uma endonuclease de restrição (Sau3A, por exemplo). Os fragmentos resultantes são separados, de acordo com seus tamanhos, por eletroforese (ver Seção 6.1), e os fragmentos com tamanhos apropriados são purificados do gel para a clonagem. O DNA do vetor de clonagem também é preparado e clivado com uma enzima de restrição compatível.

A seleção dos recombinantes é feita conforme as características do vetor de clonagem utilizado. Devido ao

fato de o DNA genômico ter sido cortado em fragmentos de maneira aleatória, é grande a chance de que os fragmentos dos diferentes clones apresentem regiões de sobreposição, permitindo a ordenação sequencial de clones e possibilitando a análise de extensas regiões do genoma.

3.2 Construção de bibliotecas de cDNA

Uma estratégia importante para a análise de genes e outras sequências de interesse começa com a construção de uma biblioteca de DNA, contendo apenas as sequências transcritas em RNAs que, via de regra, correspondem aos genes funcionais e incluem as sequências codificadoras dos produtos (de RNA ou proteicos) correspondentes. Isto é feito pela extração do RNA total de uma célula e a síntese de uma cópia de **DNA complementar (cDNA)** de cada molécula de RNA presente que tenha uma cauda de poli(A) (**Figura 16.7**). Como visto (Capítulo 11), a cauda de poli(A) na extremidade 3' do RNA, uma característica que se acreditava ser específica dos RNAs mensageiros, está presente também em RNAs nucleares não codificadores de proteínas. A reação de síntese de cDNA é catalisada pela enzima transcriptase reversa de retrovírus, que sintetiza uma cadeia de DNA complementar a partir de uma cadeia-molde de RNA (ver Seção 1.4.4). As moléculas de cDNA de fita simples são convertidas em moléculas de fita

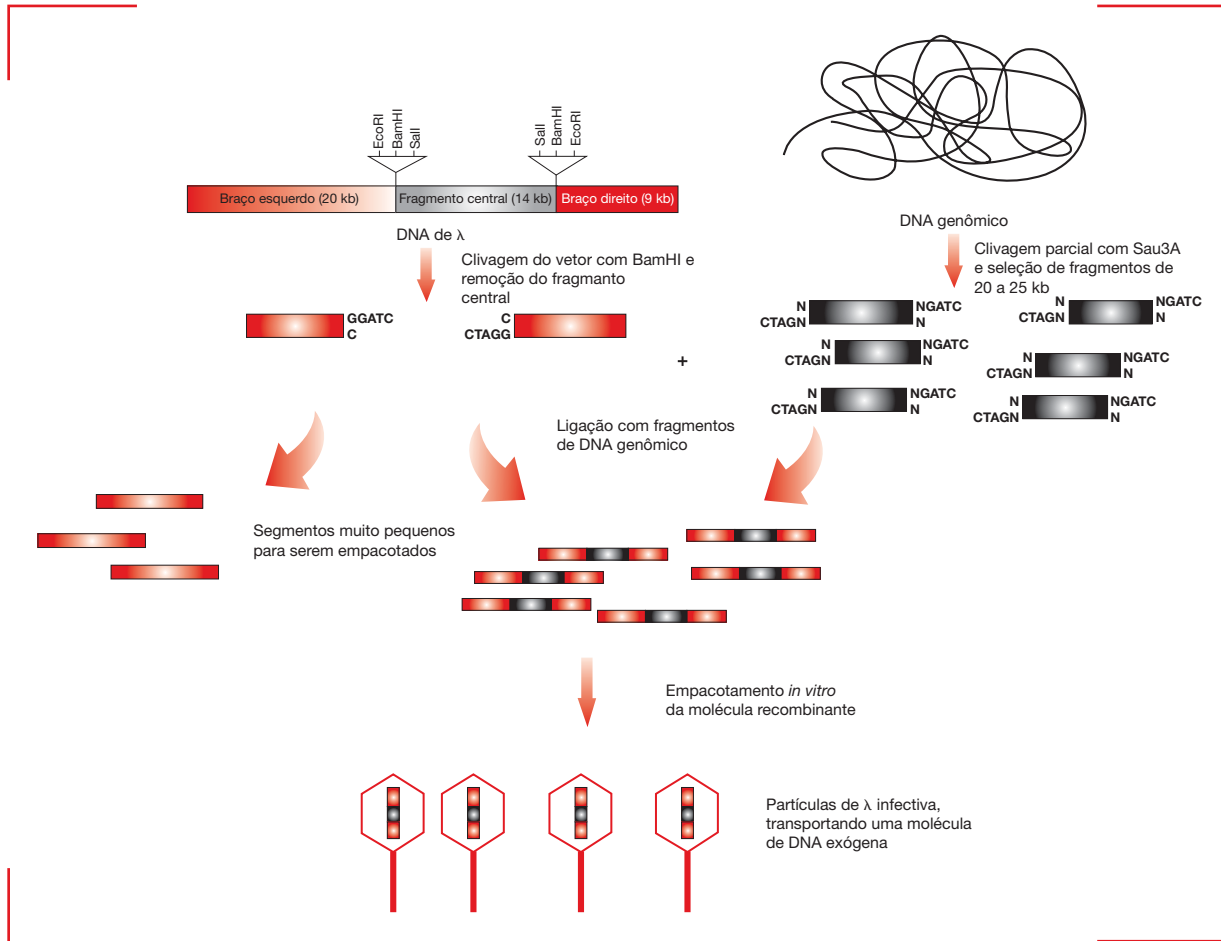


Figura 16.6

Construção de uma biblioteca genômica. O vetor λ EMBL4 pode ser utilizado como vetor de clonagem. O DNA do vetor é clivado com a endonuclease BamHI e o fragmento central de 14 kb é removido. O DNA genômico é clivado parcialmente com a endonuclease Sau3A e os fragmentos de 9 a 23 kb são selecionados. Os DNAs são misturados e ligados. O processo de empacotamento seleciona moléculas de DNA que contêm insertos de tamanho adequado para serem empacotados. Note que as extremidades coesivas dos fragmentos gerados pelas enzimas BamHI e Sau3A são complementares e que dessa forma podem ser ligados.

dupla por uma DNA-polimerase e clonadas em vetores apropriados. Cada clone, obtido dessa maneira, é chamado de clone de cDNA e a coleção de clones derivada de uma preparação de RNA mensageiro constitui uma biblioteca de cDNA. Algumas estratégias possibilitam a construção de bibliotecas enriquecidas com cDNAs completos, **ver Material Complementar Online 16.1.**

Existem diferenças importantes entre clones genômicos e clones de cDNA. Clones genômicos representam uma amostra aleatória de todas as sequências de DNA de um organismo e, com raras exceções, uma biblioteca genômica será sempre a mesma, independentemente do tipo de célula cujo DNA for utilizado para construí-la. Em contrapartida, clones de cDNA contêm somente as regiões do genoma que foram transcritas em uma molécula de RNA e, as sequências presentes numa biblioteca de cDNA dependerão do tipo de célula a partir da qual for extraído o RNA a ser utilizado na sua preparação.

A utilização de uma biblioteca de cDNA tem várias vantagens para a clonagem de genes que codificam proteínas. Uma vez que células especializadas produzem uma grande quantidade de determinada proteína, o mRNA correspondente também é produzido em quantidade equivalente nessas células. Uma biblioteca de cDNA construída a partir de células desse tipo terá, então, grande probabilidade de incluir clones de cDNA que codificam proteínas abundantes naquele tipo celular, facilitando o seu isolamento.

Uma segunda vantagem dos clones de cDNA é a de que eles, em geral, não possuem as sequências intervinientes (íntrons), que fazem parte da maioria dos genes eucarióticos (ver Capítulo 5). Dessa forma, uma sequência isolada de uma biblioteca de cDNA pode ser clonada e expressa em um vetor bacteriano ou de levedura, sem que sejam necessárias as etapas para remoção de sequências intrônicas presentes no DNA genômico.

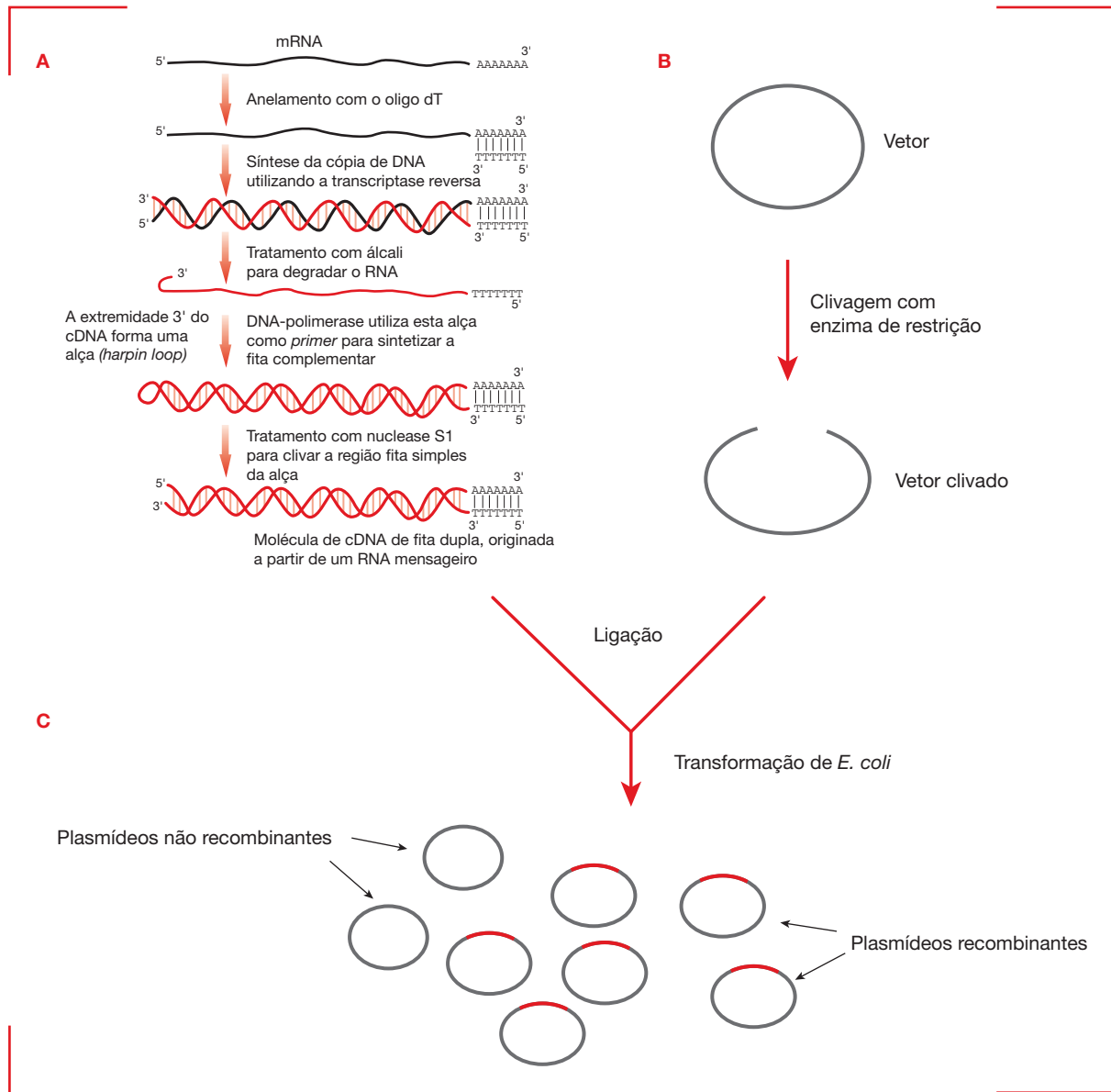


Figura 16.7

Construção de uma biblioteca de cDNA. (A) Os cDNAs complementares são sintetizados a partir dos RNAs que possuem a cauda de poli(A), utilizando o oligo (dT) como iniciador e a transcriptase reversa. Após a degradação do RNA no híbrido cDNA-RNA, a fita complementar do cDNA é produzida, utilizando a DNA-polimerase. Ao DNA de fita dupla sintetizado, adicionam-se caudas de dC. (B) O plasmídeo é linearizado e adicionam-se caudas de dG. (C) Os cDNAs e os plasmídeos são misturados para anelamento das caudas complementares e posteriormente introduzidos em *E. coli* por transformação.

3.3 Seleção de clones em bibliotecas

3.3.1 Hibridização de clones recombinantes com sondas de DNA complementares

Colônias de bactéria, cultivadas em placas com meio de cultura sólido, durante um período de aproximadamente 16 h, a 37°C, são parcialmente transferidas para uma membrana de nitrocelulose. As colônias transferidas, conhecidas como réplicas, são tratadas com álcali e, então,

incubadas com uma sonda de DNA marcada, radioativamente ou com peroxidase, que contém uma parte da sequência do gene de interesse ou outra sequência ortóloga complementar. Se necessário, milhares de clones bacterianos podem ser testados a fim de encontrar um clone que hibridize com a sonda. De forma semelhante, clones recombinantes do bacteriófago λ podem ser analisados após a infecção da linhagem bacteriana hospedeira pela hibridização como uma sonda de DNA.

3.3.2 Seleção imunológica

Uma estratégia alternativa para isolar o clone de interesse se baseia na detecção da proteína codificada pela sequência clonada e expressa na célula hospedeira utilizando técnicas imunológicas. A sequência a ser expressa precisa ser clonada em um vetor especial que possibilite a transcrição e a posterior tradução do mRNA produzido (ver Seção 4). Neste caso, um anticorpo específico é utilizado ao invés de uma sonda de DNA. As proteínas expressas na célula hospedeira podem ser transferidas para uma membrana de nitrocelulose e incubadas com o anticorpo específico. A detecção pode ser feita com a utilização de um antianticorpo conjugado com uma enzima (por exemplo, uma peroxidase), por meio de uma reação que gera um produto colorido.

4. Expressão de sequências clonadas

Muitas vezes é necessário produzir proteínas em grande quantidade, seja para estudos estruturais, funcionais ou para utilização em escala industrial. O desenvolvimento de vetores especiais, os vetores de expressão, tornou isso possível.

4.1 Vetores de expressão procarióticos

Vetores de expressão para hospedeiros procarióticos, como *E. coli* ou *Bacillus subtilis*, são plasmídeos contendo promotores fortes (ver Capítulo 13), sítios de ligação ao ribossomo (ver Capítulo 11) e vários sítios de clonagem localizados imediatamente a jusante desses sinais. Vetores de expressão híbridos têm sido criados a fim de melhorar a eficiência da transcrição e da tradução do gene de interesse. Por exemplo, existem vetores que contêm a sequência promotora -35 do óperon do triptofano, a sequência promotora -10 do promotor do gene *lacUV5* e um forte sítio sintético de ligação ao ribossomo. Como o espaçamento entre o sítio de ligação ao ribossomo e o códon de início de tradução também é decisivo, muitos vetores possuem os sítios de clonagem localizados a jusante do códon de início de tradução. Nestes vetores, a proteína de interesse é gerada em grandes quantidades, mas contendo alguns resíduos extras na extremidade aminoterminal. Esses aminoácidos adicionais geralmente não representam qualquer problema para estudos bioquímicos ou funcionais subsequentes.

Em alguns casos, é necessário regular a produção da proteína pelo vetor de expressão. Vários dos vetores de expressão procarióticos possuem promotores induzíveis. Nestes sistemas, uma sequência operadora é clonada em uma região apropriada do vetor de expressão, ao passo que um gene que codifica um repressor transcricional está presente no próprio vetor ou no genoma do hospedeiro. A inativação do repressor permite a transcrição. Por exemplo, genes sob o controle do promotor *lac* são

induzidos pela adição de IPTG (ver Seção 2.3.4), ao passo que os genes controlados pelo repressor sensível à temperatura do bacteriófago λ são induzidos por uma elevação da temperatura. Promotores induzíveis permitem que a produção de proteínas tóxicas para *E. coli*, seja mantida sob repressão até que as células tenham se multiplicado em grande número. Depois disso, as células podem ser induzidas a expressarem o gene clonado e em seguida serem coletadas para lise e purificação da proteína recombinante. Isso pode evitar que uma proteína recombinante seja degradada pelas proteases da bactéria hospedeira, pois o tempo entre a indução e a coleta das células para processamento é reduzido.

A produção da proteína de interesse pode ser feita fusionando-a (ver Seção 4.3.1) a uma proteína carreadora ou a uma sequência de histidinas. A **glutathione-S-transferase** (GST) e a **poli-histidina** são frequentemente utilizadas como a parte carreadora. Essa fusão permite que a proteína de interesse seja purificada, com base nas propriedades da proteína carreadora.

4.2 Vetores de expressão eucarióticos

Em certos casos, as células procarióticas são incapazes de produzir proteínas funcionais a partir de genes eucarióticos, mesmo que os sinais para a expressão desses genes estejam presentes e todos os esforços a fim de aumentar a eficiência da transcrição e da tradução tenham sido feitos. Isso pode ocorrer, por exemplo, quando o gene clonado inclui códons comuns na espécie de origem, mas pouco utilizados na espécie hospedeira, e que por isso, não possui um repertório adequado de tRNAs para interpretá-los (ver Capítulo 12). Além disso, muitas proteínas eucarióticas devem ser modificadas após a tradução (por exemplo, pela glicosilação) para serem funcionais. Para produção de certas proteínas recombinantes eucarióticas é necessária a clonagem dos genes correspondentes em vetores de expressão que funcionem em eucariotos.

Leveduras são organismos eucarióticos que satisfazem as necessidades básicas de um hospedeiro para a clonagem e a expressão de genes eucarióticos. A levedura *Saccharomyces cerevisiae* constituiu-se em um dos primeiros sistemas de expressão de eucariotos, possibilitando a produção comercial de insulina e vacina contra hepatite B, por exemplo. Em geral, os vetores utilizados são híbridos, contendo tanto sequências derivadas de levedura como de bactéria. As sequências bacterianas incluem a origem de replicação e um gene que confere resistência a um antibiótico, como a ampicilina. Essas sequências permitem a replicação e a manutenção do vetor em *E. coli*. A porção de DNA de levedura inclui também elementos para a seleção das leveduras transformantes (portadoras do vetor), como os genes que codificam a β -isopropilmalato-desidrogenase (LEU2) ou a oritidina-5'-decarboxilase (URA3), que permitem a complementação de mutações auxotróficas na levedura (ver Capítulo 7). Um exemplo de vetor de expressão de *S. cerevisiae* é apresentado na **Figura 16.8A**. Outras espécies de levedura, tais como *Kluyveromyces lactis*, *Pichia*

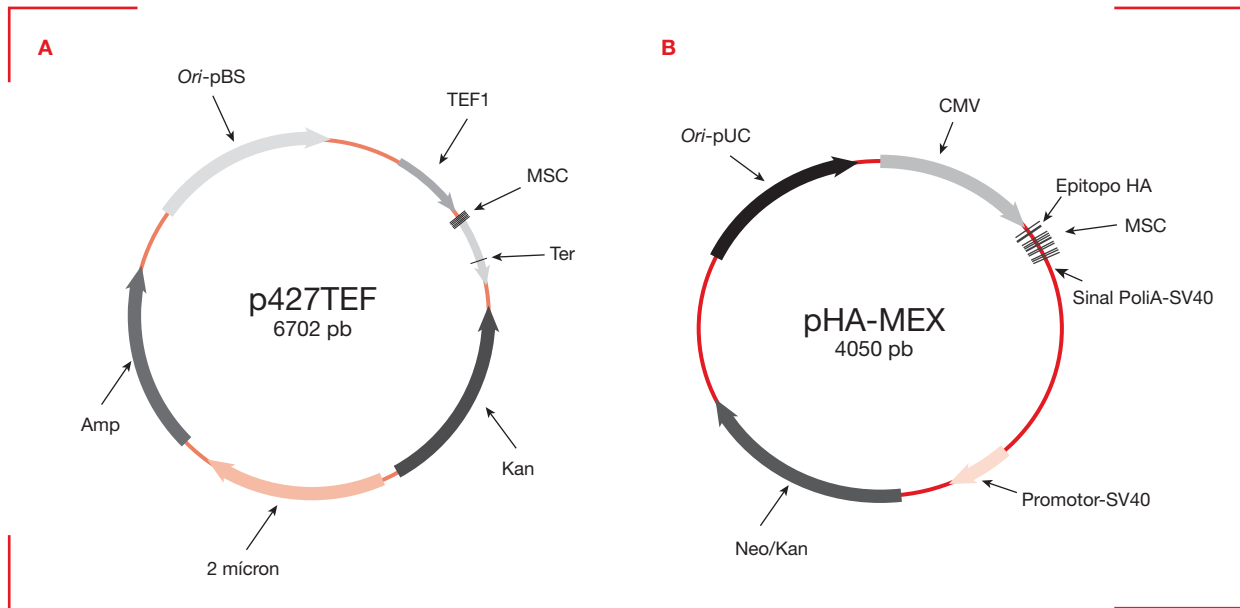


Figura 16.8

Mapas de vetores de expressão *S. cerevisiae*. (A) O mapa do vetor p427TEF é mostrado. TEF1, promotor do gene que codifica o fator de alongamento da tradução; 2 micron, região com origem de replicação que permite a multiplicação do plasmídeo em alto número de cópias (20-100 cópias/célula); MSC, múltiplos sítios de clonagem; AmpR, gene de resistência à ampicilina que permite a seleção em *E. coli*; CYC1, sítio de terminação de transcrição; pBS, origem de replicação do plasmídeo pBluescript que possibilita a propagação do vetor em *E. coli*; Kan, gene de resistência à canamicina que permite seleção em *E. coli* (30 µg/mL de canamicina) e *S. cerevisiae* (200 µg/mL G418). (B) Mapa do vetor pHA-MEX para expressão em células de mamíferos. CMV, promotor do citomegalovírus; NeoR/KanR, gene de resistência a canamicina/neomicina; HA, marca com epitopo de hemaglutinina; pUC, origem de replicação do plasmídeo pUC; SV40, sinal de poliadenilação do poli-SV40; prom-SV40, promotor de SV40. Figura reproduzida com autorização da Dualsystems Biotech AG, Suíça.

pastoris e *Hansenula polymorpha*, também têm sido utilizadas para expressão heteróloga.

Em células de mamíferos, moléculas de DNA recombinante são expressas por vetores de estrutura semelhante. Os elementos essenciais de um vetor de expressão de mamíferos são: (1) um promotor induzível ou constitutivo, capaz de proporcionar alto nível de transcrição; (2) sinais para processamento do mRNA e sua tradução, incluindo a sequência de Kozak, códon de terminação da tradução, sinais de clivagem e poliadenilação do mRNA; (3) terminador de transcrição; (4) marca de seleção para preparação de linhagens estáveis de células; e (5) origem de replicação de procariotos, para possibilitar a propagação do vetor na bactéria *E. coli*. As regiões reguladoras (promotor e reforçador) mais utilizadas para controlar a expressão dos genes são originárias do vírus SV40 e do citomegalovírus humano (CMV). Em alguns vetores, a inclusão da origem de replicação de SV40 possibilita a expressão transitória em determinadas linhagens celulares. Um exemplo de vetor de células de mamíferos é apresentado na Figura 16.8B.

Após a sua construção, o plasmídeo recombinante é introduzido nas células de mamíferos por transfecção mediada por fosfato de cálcio, DEAE-dextran, eletroporação ou lipossomos catiônicos. Os dois primeiros proce-

dimentos produzem condições químicas que propiciam a ligação do DNA na superfície celular. O DNA, então, é endocitado pela célula. A eletroporação utiliza um campo elétrico para abrir poros na célula, e o DNA, presumivelmente, difunde para dentro da célula através desses poros. Os lipossomos catiônicos são feitos de lipídeos catiônicos e fosfolipídeos neutros, que conferem uma carga positiva às partículas, possibilitando a formação lipossomo-DNA.

Independentemente do método utilizado, uma transfecção em sistemas de células de mamíferos pode dar origem a dois tipos de resultados. A transcrição ou a replicação de um gene transfectado pode ser analisada entre um a quatro dias após a introdução do DNA. Isso, em geral, exige a colheita da célula transfectada com a interrupção de sua multiplicação. Esse tipo de transfecção é referido como **transiente ou temporário**. Muitos experimentos, no entanto, necessitam da formação de linhagens celulares possuindo o(s) gene(s) integrado(s) no DNA cromossômico, processo referido como **transfecção estável ou permanente**. Em ambos os processos, muitas vezes os pesquisadores desejam introduzir mais de um gene nas células receptoras, o que pode ser obtido pela transfecção de uma mistura de plasmídeos contendo os genes desejados para as células, processo referido como **cotransfecção**.

4.3 Produção e purificação de proteínas recombinantes

A produção de proteínas recombinantes em *E. coli* tem sido bastante explorada nas últimas décadas. Os sistemas de expressão desenvolvidos para essa bactéria oferecem muitas vantagens em relação àqueles desenvolvidos para outros organismos, principalmente devido ao custo mais baixo dos nutrientes utilizados para sua multiplicação, o tempo curto para acumulação de biomassa, a possibilidade de escalonamento na produção e a facilidade de manipulação do sistema.

4.3.1 Proteínas de fusão e purificação por cromatografia de afinidade

Dentre os diversos sistemas desenvolvidos, alguns destacam-se pela facilidade de purificação, por exemplo,

os vetores que utilizam o gene codificador da GST de *Schistosoma japonicum*, como parte da proteína de fusão. Nesse caso, o gene de interesse é clonado no vetor de expressão, de forma que a proteína de fusão produzida contenha GST na sua porção N-terminal e a proteína de interesse na porção C-terminal (ver Seção 4.1). A proteína pode, então, ser purificada por cromatografia de afinidade utilizando-se, por exemplo, glutationa-sepharose. A proteína de interesse pode ser separada da GST pela clivagem com proteases específicas, uma vez que o vetor é construído de forma a conter sítios para essas enzimas, que se situam entre a porção GST e a proteína de interesse. A **Figura 16.9** mostra um esquema do vetor pGEX4T, as etapas de clonagem, a expressão e a purificação da proteína.

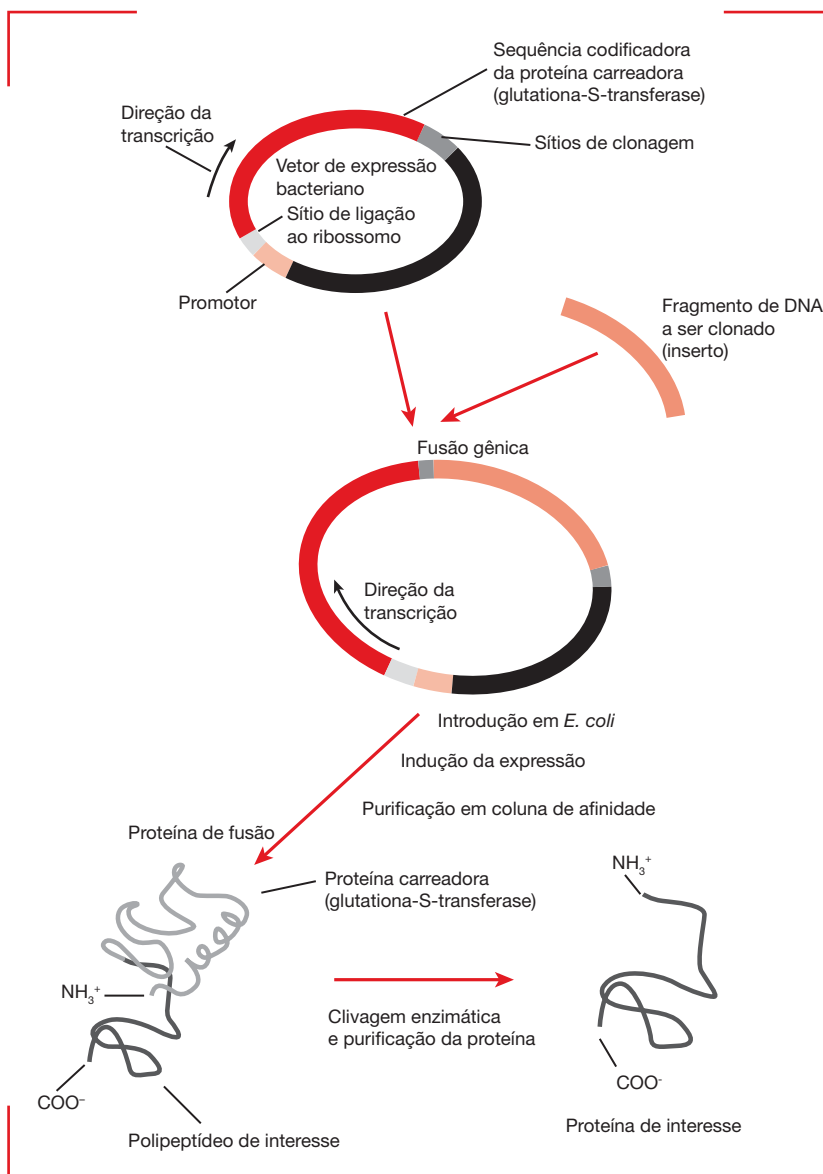


Figura 16.9

Esquema de expressão e purificação de proteína recombinante em *E. coli* com o vetor pGEX4T-1.

A proteína de interesse é produzida como parte de uma proteína de fusão com a glutationa S-transferase (GST). A porção GST possibilita a purificação da proteína recombinante por afinidade, utilizando uma resina contendo glutationa (Sepharose-GSH, por exemplo). A proteína de interesse pode, posteriormente, ser separada da GST pela clivagem enzimática *in vitro*.

5. Técnicas de transformação genética de plantas

Os estudos sobre genética molecular de plantas começaram a ser realizados há pouco tempo, devido às dificuldades em se trabalhar com as plantas de maior interesse econômico. Essas plantas possuem genomas complexos e grandes, em razão da **poliploidia** (algumas com pelo menos quatro cópias de cada cromossomo), o que dificultava muito a identificação de mutações recessivas e a manipulação do DNA. Uma prova disso é o fato do genoma do milho possuir uma quantidade de DNA sete vezes superior à do genoma humano (ver Capítulo 5).

Aos poucos os pesquisadores foram aprendendo a manipular e a selecionar as plantas que mais se prestavam à manipulação genética e, hoje em dia, grandes progressos estão sendo obtidos, principalmente na área de geração de linhagens que apresentam genes exógenos sendo expressos em seus genomas. Esses genes conferem vantagens, seja para a planta, como resistência a determinados insetos ou herbicidas, seja para o homem, como a produção de sementes com maior teor de vitamina A (como é o caso do “arroz amarelo”) ou com maior conteúdo proteico, como é o caso da castanha e da soja. O domínio da tecnologia que permite a modificação genética de plantas, bem como de qualquer organismo de maneira geral, está possibilitando avanços importantes tanto no aumento da produção de alimentos, com a manutenção da mesma área de terra anteriormente ocupada, como na obtenção de produtos medicinais e até mesmo tecidos e órgãos. Aos organismos que sofreram algum tipo de manipulação genética dá-se o nome de **organismos geneticamente modificados (OGMs)** ou **transgênicos**.

Atualmente, os genomas de várias plantas de interesse econômico foram sequenciados ou estão sendo sequenciados, como é o caso do arroz, soja, café, cana-de-açúcar e outras (ver Capítulo 5). Uma vez desvendado o genoma, os pesquisadores podem localizar e modificar, de forma precisa, os genes de interesse e selecionar linhagens que apresentem, por exemplo, características que possibilitem um melhor desenvolvimento e rendimento em determinadas áreas de plantio, ou seja, melhor adaptadas às regiões em que se encontram. Informações atuais e mais abrangentes podem ser encontradas nos endereços citados nas Leituras recomendadas.

A transformação genética de plantas só foi possível depois do domínio de tecnologias que possibilitaram a introdução de DNA nas células vegetais de uma maneira eficiente e de técnicas que permitiram a recuperação e o desenvolvimento das células transformadas para a geração da planta adulta modificada.

5.1 Transformação de plantas via *Agrobacterium tumefaciens*

As plantas são suscetíveis a diferentes tipos de doenças, incluindo alguns tipos de câncer ou tumores. Um tumor

de plantas é chamado de **galha**. Existem muitos tipos de galhas e vários agentes podem induzir a formação de uma galha, incluindo ferimentos, insetos, vírus, uma determinada composição genética e bactérias. Em alguns casos, o tecido contendo a galha pode crescer muito e levar à morte da planta. Dos vários tipos de doenças causadas por galhas, a mais interessante, do ponto de vista biológico, é a chamada **coroa galhada**. O agente causador dessa doença é uma bactéria do solo chamada *Agrobacterium tumefaciens*, um membro da família Rhizobiaceae, da qual também fazem parte as bactérias simbióticas e fixadoras de nitrogênio do gênero *Rhizobium*. *A. tumefaciens* penetra na planta através de aberturas decorrentes de ferimentos e transforma geneticamente suas células. Ela induz a formação de tumores pela transferência de parte de seu material genético para o genoma da planta. A expressão dos genes bacterianos em células transformadas estimula a divisão celular. Como consequência, as células proliferam, formando um tumor, cujo crescimento não pode mais ser regulado pela planta.

Os tecidos do tumor secretam um tipo incomum de aminoácidos, conhecidos como **opinas**. *A. tumefaciens* utiliza essas opinas como fonte de carbono ou nitrogênio para o seu metabolismo. O agente indutor do tumor de *A. tumefaciens* é o **plasmídeo Ti** (de *tumor-inducing* = indutor de tumor). O plasmídeo Ti contém os genes que controlam o catabolismo de um determinado tipo de opina na bactéria e os genes que controlam a síntese dessas opinas nas células transformadas da planta. Existem vários tipos de plasmídeos Ti e cada um deles comanda a síntese e a utilização de um tipo de opina, como a nopalina, a octopina ou a agropina, entre outras.

Somente uma parte do plasmídeo Ti é transferida para a célula vegetal durante a infecção (**Figura 16-10**). Esse fragmento de DNA transferido é chamado de **T-DNA**, que se integra no genoma da planta hospedeira e carrega os genes indutores de tumor (*onc*) e os genes para a síntese de um tipo de opina. Além do T-DNA, duas outras regiões do plasmídeo Ti são importantes para a biologia da coroa galhada: uma região de 30 kb, conhecida como **região *vir***, que contém os genes necessários para a integração do T-DNA no genoma hospedeiro; e uma outra região, onde estão localizados os genes para a utilização da opina produzida pelos genes do T-DNA. Todos os genes do T-DNA são transcritos pela RNA-polimerase II da planta e não são transcritos na bactéria. Os mRNAs transcritos são poliadenilados nas suas extremidades 3' e, portanto, são similares aos demais mRNAs celulares.

O plasmídeo Ti tem sido amplamente utilizado como veículo para a introdução de genes exógenos em plantas. Porém, foi necessário “desarmar” o plasmídeo Ti, para que ele não causasse tumores. Isso foi feito por meio de deleções dos genes do T-DNA que codificavam as enzimas controladoras da síntese dos hormônios auxina e citocininas, os **fito-oncogenes**. Além disso, foi necessário introduzir um gene no T-DNA que permitisse a seleção das células transformadas. Um gene que confere resistência a um determinado antibiótico é frequentemente utilizado

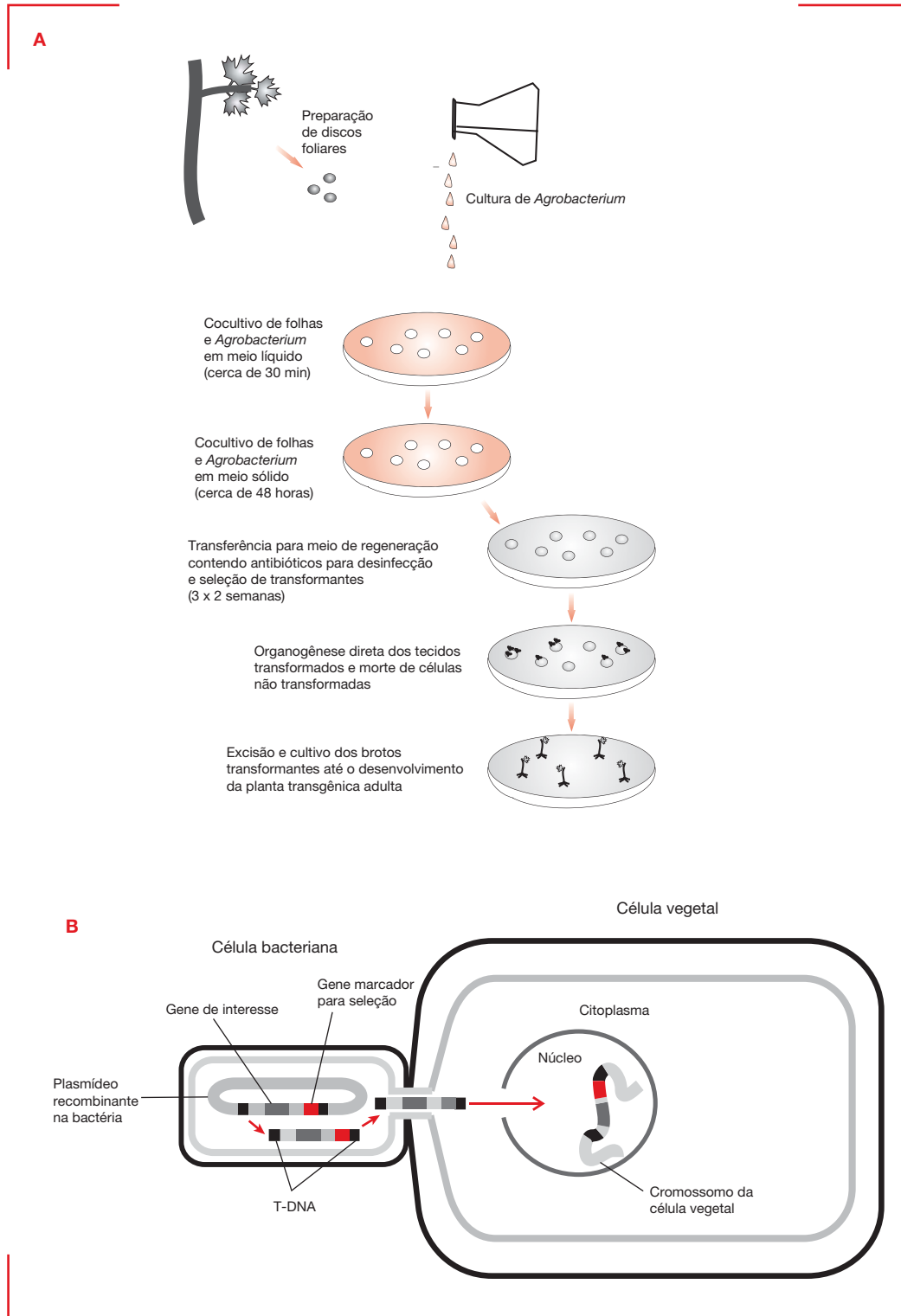


Figura 16.10

Transformação de uma planta por *A. tumefaciens*. (A) Procedimentos básicos para a transformação da planta de interesse; (B) processo de transferência e integração do plasmídeo Ti no genoma da planta hospedeira.

com esse propósito. A Figura 16.10 mostra os procedimentos utilizados para a infecção de células vegetais com *A. tumefaciens* e a posterior introdução de um gene de interesse no genoma da planta.

A metodologia de transformação de células vegetais, envolvendo *A. tumefaciens*, possui várias vantagens em relação às demais técnicas existentes para o mesmo fim, principalmente por utilizar um processo que ocorre naturalmente: (1) a transformação das células é permanente e altamente eficiente, pois apenas uma ou poucas cópias do T-DNA são integradas no genoma da célula vegetal e em um local que não prejudica funções celulares normais; (2) a infecção se dá em células intactas, o que facilita a regeneração da planta; (3) é possível a transferência de grandes moléculas de DNA, com raros eventos de rearranjo; (4) é um processo simples e pouco oneroso. No entanto, nem todas as espécies vegetais são suscetíveis à infecção por *A. tumefaciens* e falsos transformantes aparecem com certa frequência, devido à contaminação bacteriana.

5.2 Transformação de plantas via bombardeamento de DNA

A suscetibilidade limitada das plantas de interesse econômico à infecção por *Agrobacterium* levou os pesquisadores a desenvolver outros métodos para introduzir genes de interesse nessas plantas. Uma destas metodologias consiste na introdução de DNA por um processo denominado **bombardeamento** ou **biobalística**. Esse método utiliza um aparelho chamado de **acelerador de partículas**, no qual o DNA plasmidial, contendo um gene de interesse a ser inserido no genoma da planta e complexado com micropartículas de ouro ou tungstênio, é submetido a altas pressões de gás hélio. Essa pressão funciona como um “tiro”, acelerando o complexo DNA-micropartícula com força e velocidade suficientes para o rompimento simultâneo da parede e da membrana celulares e a penetração na célula vegetal.

O processo de bombardeamento possui a vantagem de servir para qualquer tipo de tecido ou célula e, portanto, ser ilimitado para qualquer espécie vegetal e animal. Como é um método que não envolve qualquer organismo vivo, não há a ocorrência de falsos transformantes. Contudo, a transformação das células tem se mostrado transiente, com uma frequência muito baixa de integrações estáveis no genoma. Esse processo, além de limitar o tamanho do DNA a ser bombardeado, proporciona a ocorrência de rearranjos e múltiplos eventos de integração. Por ser um processo que necessita de equipamento adequado para cada tipo celular, ele se torna oneroso e relativamente complexo.

5.3 Transformação de plantas via eletroporação

A técnica de **eletroporação** (ver Seção 2.3.1) também tem sido utilizada para a introdução de DNA nas células vegetais. Assim como o bombardeamento, essa técnica

também oferece as vantagens de ser aplicada a qualquer espécie vegetal e de não originar falsos transformantes. No entanto, a transformação das células vegetais por eletroporação tem algumas limitações, como o fato de ser transitória, servir apenas para a introdução de pequenas moléculas de DNA e levar à ocorrência de rearranjos no DNA. A eletroporação exige a utilização de **protoplastos** vegetais (células vegetais sem a parede celular), o que dificulta a regeneração das plantas, além de ser um processo complexo e oneroso.

6. Técnicas de análise de DNA e RNA

6.1 Eletroforese

Ácidos nucleicos, bem como proteínas e outras moléculas biológicas, apresentam carga elétrica que, no caso do DNA e do RNA, é negativa, portanto, quando moléculas de DNA ou RNA são colocadas em um campo elétrico, elas migram em direção ao polo positivo. A taxa de migração de uma molécula depende de dois fatores: sua forma e sua razão carga/massa. Quando a eletroforese é realizada em um gel, em geral constituído por **agarose**, **poliacrilamida** ou uma mistura de ambos os polímeros, as moléculas de DNA de tamanhos diferentes migram através da rede de poros do gel com velocidades diferentes, em direção ao eletrodo positivo. Quanto menor a molécula, mais rápida será a sua migração no gel. A **eletroforese em gel**, dessa forma, separa moléculas de DNA de acordo com seus tamanhos.

Em muitos tipos de géis, a mobilidade eletroforética de um fragmento de DNA é inversamente proporcional ao número de pares de bases, até um certo limite. Géis de poliacrilamida são utilizados para separar fragmentos contendo até 1.000 pb, ao passo que géis mais porosos, feitos com agarose, são utilizados para resolver misturas de fragmentos de tamanhos maiores (com até cerca de 20 kb). Uma característica importante dos géis de poliacrilamida é o alto poder de resolução, uma vez que fragmentos de até centenas de nucleotídeos de comprimento diferindo em apenas um nucleotídeo podem ser distinguidos.

A maneira mais fácil de analisar os resultados de um experimento de eletroforese em gel é corar o DNA no próprio gel, com um composto que o torne visível. O **brometo de etídeo** é um desses compostos que, quando uma molécula de DNA está ligada a ele e é submetida à **luz ultravioleta** (UV), ela fluoresce em uma cor avermelhada, desde que uma quantidade suficiente de DNA esteja presente no gel. Contudo, esse procedimento é bastante perigoso porque o brometo de etídeo é um poderoso mutagênico, e a radiação UV pode provocar queimaduras severas. Por essas razões, corantes não mutagênicos, que coram o DNA de verde ou azul (por exemplo, o *sybr green*), algumas vezes não necessitam da luz UV para que os resultados sejam visualizados, estão sendo agora utilizados em muitos laboratórios.

6.2 Hibridização

A parte mais difícil na clonagem de um gene é, muitas vezes, a identificação dos raros clones de uma biblioteca contendo o fragmento de DNA de interesse. Com frequência, esse problema pode ser resolvido devido à complementaridade que existe entre as sequências de DNA ou RNA. Métodos que se baseiam na **hibridização de ácidos nucleicos** levam em consideração a capacidade de cadeias de DNA e RNA para formarem moléculas de fita dupla estáveis, por meio do pareamento de bases complementares. Essa técnica é utilizada para identificar e determinar a localização de sequências específicas de ácidos nucleicos dentro de um universo muito amplo, como um genoma, uma população de moléculas de DNA recombinante ou uma mistura de moléculas de RNA. A estabilidade do dúplex formado está relacionada ao grau de complementaridade existente entre duas fitas de ácido nucleico.

As condições físicas, chamadas de condições de **estringência**, nas quais a hibridização ocorre podem ser alteradas com o objetivo de utilizar uma fita de ácido nucleico como **sonda** para a outra fita total ou parcialmente complementar. Sob condições de estringência severas, que incluem altas temperaturas e baixas concentrações de sais, a sequência utilizada como sonda somente irá hibridizar com uma sequência perfeitamente complementar a ela. Quando a estringência é relaxada, imperfeições no pareamento entre as fitas podem ser toleradas, sem causar uma desestabilização no dúplex formado. Essa estratégia é utilizada para detectar sequências parcialmente similares, do tipo que é esperado que ocorra quando se trabalha com uma sonda de DNA e um DNA-alvo de organismos diferentes (sequências ortólogas, definidas no Capítulo 4). Se a estringência for muito reduzida, a sonda hibridizará em várias regiões parcialmente complementares, e a especificidade será perdida. Um grande número de técnicas que utilizam a capacidade de pareamento entre os ácidos nucleicos tem sido usado para a **triagem** (*screening*) e o isolamento de sequências específicas. A maioria dessas técnicas opera com uma réplica do DNA de interesse imobilizada sobre um suporte sólido, assim como uma membrana de náilon ou de nitrocelulose.

O fragmento que servirá como sonda para a hibridização pode ser marcado por diversas técnicas. Uma dessas técnicas, denominada *nick translation*, consiste em clivar uma das fitas da molécula de DNA de fita dupla que, submetida à atividade exonucleásica da enzima DNA-polimerase (ver Seção 1.4.1), fica com vários segmentos de DNA de fita simples. Essas regiões de fita simples são, então, preenchidas novamente pela ação da própria DNA-polimerase, que vai adicionando os nucleotídeos conforme o molde preservado na outra fita. A substituição dos nucleotídeos preexistentes por nucleotídeos marcados com material radioativo (como ^{32}P ou ^{35}S) torna possível a síntese, *in vitro*, de moléculas de DNA radioativas, que irão servir como sondas nos experimentos de hibridização.

Um método de marcação alternativo, que não envolve a utilização de isótopos radioativos, baseia-se na marcação direta do fragmento de DNA que atuará como sonda com a enzima peroxidase extraída de raiz forte. Isso é obtido por meio da desnaturação completa da sonda, para fornecer fragmentos de fita simples. A peroxidase, que foi complexada com um polímero positivamente carregado, é, então, adicionada e forma uma ligação fraca com a fita de ácido nucleico, devido à atração das cargas positivas e negativas. A adição de uma solução de glutaraldeído promove ligações químicas cruzadas, de forma que a sonda fica covalentemente marcada com a enzima. Uma vez marcada, a sonda pode ser utilizada para a hibridização com um DNA ou RNA-alvo imobilizado sobre uma membrana, sob a temperatura de 42°C, que garante a atividade da enzima.

6.3 Southern blot

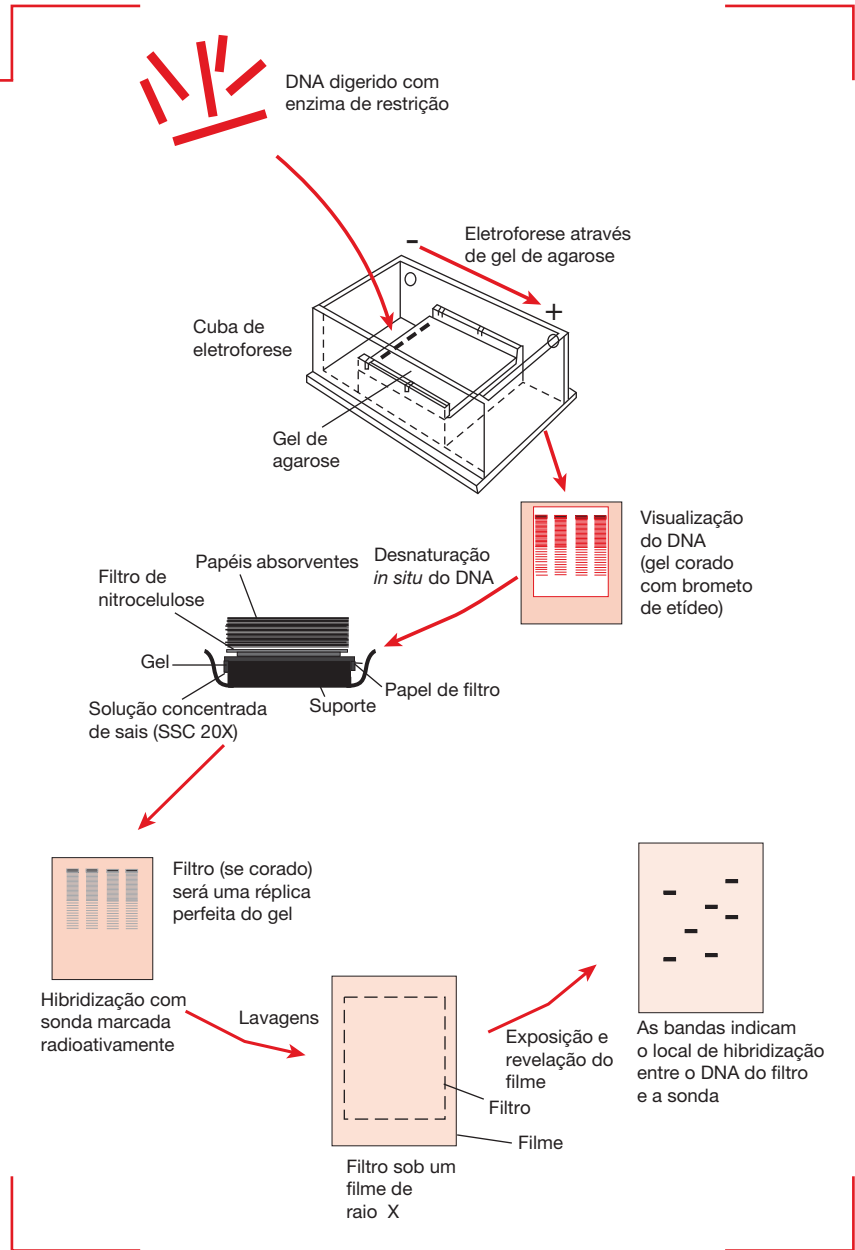
O método de *Southern blot* (**Figura 16.11**) é uma das técnicas de hibridização mais importantes para a detecção de sequências específicas de DNA. O DNA é digerido com uma ou mais enzimas de restrição e os fragmentos resultantes são separados por eletroforese em gel de agarose. Os fragmentos de DNA de fita dupla podem ser visualizados com luz UV, uma vez que o DNA no gel é corado e o gel, fotografado. Os fragmentos de DNA são, então, desnaturados *in situ*, pela imersão do gel em uma solução de hidróxido de sódio e transferidos, por capilaridade, para uma membrana de náilon ou de nitrocelulose, exatamente na mesma posição em que estavam no gel de agarose, criando, na membrana, uma réplica perfeita do gel. O DNA é covalentemente ligado ao filtro pelo calor ou pela luz UV.

A hibridização entre uma sonda, marcada radioativamente ou com peroxidase (pode ser DNA ou RNA), e o DNA desnaturado na membrana é realizada sob as condições de estringência desejadas e, após a lavagem do excesso de sonda, não ligada ao DNA complementar do filtro, a posição na qual a ligação específica ocorreu pode ser detectada por uma autorradiografia do filtro. O padrão de hibridização pode ser, então, comparado com a região do gel original (uma ou poucas bandas), que contém as sequências de DNA de interesse. Quando se utiliza a sonda marcada com peroxidase deve-se, após a hibridização, incubar a membrana com dois reagentes, sendo que um deles fornecerá o substrato para a peroxidase agir (peróxido de hidrogênio). A redução do peróxido de hidrogênio pela enzima é acompanhada de uma reação que produz luz, que, na presença do segundo reagente contendo luminol, é convertida a luz azul. A emissão dessa luz pode ser captada por um filme radiográfico, que mostrará o mesmo padrão de bandas obtido com a marcação radioativa.

A técnica de *Southern blot* permite a identificação de fragmentos de DNA com sequência idêntica ou similar à sonda utilizada. Eventualmente, ela pode também auxiliar no posicionamento relativo de diferentes fragmentos dentro de um segmento maior de DNA (mapeamento).

Figura 16.11

Técnica de Southern blot. Esse método permite a identificação de um segmento de interesse, a partir do genoma total de um organismo ou de uma mistura de fragmentos.



A técnica de *Southern blot* pode, ainda, ser utilizada na identificação de polimorfismos, que determinam a alteração do padrão de clivagem (devido a mutações pontuais em sítios de restrição) obtido a partir de uma determinada região do DNA. Essa metodologia é conhecida como polimorfismo do comprimento do fragmento de restrição, ou **RFLP** (de *restriction fragment length polymorphism*). Estes diferentes padrões são detectados utilizando-se a própria região potencialmente polimórfica como sonda. Padrões de RFLP, obtidos com uma determinada sonda (de uma região de DNA repetitivo), podem ser utilizados no estabelecimento de uma “impressão digital de DNA” (*DNA fingerprint*) que permite a diferenciação entre dois indivíduos quaisquer. Essa técnica é bastante útil em medicina forense,

na identificação de suspeitos ou em testes de paternidade, por exemplo.

Uma técnica semelhante à de *Southern blot* utiliza RNA no lugar do DNA fixado ao filtro, e é conhecida como **Northern blot**. *Northern blot* é uma técnica utilizada na pesquisa em biologia molecular para estudar a expressão gênica, ou seja, verificar se um determinado gene de um genoma é ou não transcrito em RNA e, também, para se quantificar essa transcrição. Essa técnica tem tal nome devido à similaridade de seu procedimento com o *Southern blot* (batizado pelo biólogo britânico Edwin Southern), com a diferença básica de que, em vez de DNA, a substância analisada por eletroforese com uma sonda hibridizadora é RNA. Uma outra diferença no procedimen-

to (quando comparada com o *Southern blot*) é a adição de formaldeído no gel de agarose, que funciona como um desnaturante, evitando que as moléculas de RNA adotem estruturas secundárias por pareamento intracadeia, que alterariam seu padrão de migração eletroforética. Como no *Southern blot*, a sonda pode ser de DNA ou RNA.

6.4 Sequenciamento de DNA

No final da década de 1970 dois métodos foram desenvolvidos para a determinação da sequência de nucleotídeos de moléculas de DNA: (1) o método enzimático de Sanger, ou de terminação de cadeia com didesoxirribonucleosídeo trifosfato; e (2) o método de degradação química de DNA, desenvolvido por Maxam e Gilbert. Embora o segundo método tenha tido uma ampla aplicação inicialmente, o primeiro método tem sido o mais usado, a partir do desenvolvimento de vetores adequados e da utilização de diferentes DNA-polimerases. Portanto, apenas o método de Sanger será discutido neste capítulo.

6.4.1 Método de Sanger

No método desenvolvido por Frederick Sanger, em 1977, o DNA a ser sequenciado, que pode ser de fita simples ou de fita dupla desnaturado, é hibridizado com um oligonucleotídeo (iniciador), com aproximadamente 20 pb. Na presença dos quatro desoxirribonucleosídeos trifosfatados (dATP, dCTP, dGTP e dTTP), um dos didesoxirribonucleosídeos trifosfatados (ddATP, ddCTP, ddGTP ou ddTTP), α - 32 P-dATP ou 35 S-dATP e DNA-polimerase, a fita complementar é sintetizada. Os didesoxirribonucleosídeos trifosfatados não apresentam o grupamento hidroxílico no carbono 3', de forma que, uma vez incorporados, impedem a adição de outro nucleotídeo, terminando, assim, o alongamento da cadeia. Quatro reações separadas são feitas, uma para cada didesoxirribonucleosídeo trifosfato. Após a reação, as amostras são desnaturadas na presença de formamida e analisadas em gel de poliacrilamida contendo ureia. Para leitura de sequências, até uma distância de 50 nucleotídeos do iniciador, utiliza-se um gel com concentração de poliacrilamida de 12 a 20%, ao passo que sequências até 500 pb do iniciador podem ser lidas em géis com concentração de 5 a 6%. A **Figura 16.12A** mostra de forma esquemática as diferentes etapas do sequenciamento pela técnica de Sanger. A diferença entre dATP e ddATP é mostrada na **Figura 16.12B**.

6.5 Sequenciamento automático de DNA

Com o crescimento do interesse em conhecer a sequência completa dos genomas de diferentes organismos, incluindo o do ser humano, técnicas mais sofisticadas para o sequenciamento do DNA tiveram de ser desenvolvidas. Essas novas técnicas visaram à automatização do processo, bem como a redução dos custos destes procedimentos.

Os sequenciadores de DNA analisam automaticamente as fitas de DNA, marcadas com qualquer um dos

quatro didesoxirribonucleotídeos terminadores (ddATP, ddCTP, ddGTP e ddTTP), que são marcados com compostos fluorescentes, utilizados para identificar cada uma das posições dos diferentes nucleotídeos na cadeia (**Figura 16.12.C**). A reação de sequenciamento utiliza o princípio da reação em cadeia da polimerase (ver Seção 6.6). Por meio dos vários ciclos de anelamento dos iniciadores com o molde, extensão da cadeia de DNA (polimerização) e desnaturação, as cadeias truncadas pela incorporação dos didesoxirribonucleotídeos específicos são produzidas.

Utilizando os terminadores marcados, uma única reação é realizada e aplicada em um gel (placa ou capilar). Os fragmentos de DNA marcados são separados no gel de acordo com seu tamanho. Um feixe de *laser* rastreia o gel, excitando os marcadores que emitem luz em um comprimento de onda específico, conforme o fluoróforo utilizado. A emissão de luz é registrada e o padrão do espectro é analisado com ajuda de programas que permitem gerar as sequências de DNA. O uso de terminadores marcados apresenta algumas vantagens em relação ao uso de iniciadores marcados. Como a reação para detecção das quatro bases diferentes (A, C, G e T) é realizada em um único tubo, isso reduz as manipulações.

O desenvolvimento de novas tecnologias de sequenciamento, com base na própria síntese de DNA, no uso de inibidores reversíveis do alongamento da cadeia de DNA e também em procedimentos de ligação de oligonucleotídeos, tem possibilitado uma grande redução no tempo de determinação das sequências e nos custos das reações. Algumas características dessas novas tecnologias são apresentadas no **Material Complementar Online 16.2**.

6.6 Reação em cadeia da polimerase

O surgimento de novas técnicas em biologia molecular tem alterado muito as formas de abordagem dos problemas básicos aplicados nessa área. Dessa forma, o desenvolvimento de técnicas, como as de clonagem molecular, de *Southern blot* e de sequenciamento de DNA, entre outras, permitiu um avanço considerável na área. O desenvolvimento da técnica de amplificação de segmentos de DNA, utilizando a **reação em cadeia da polimerase (PCR, de *polymerase chain reaction*)** ampliou perspectivas em várias áreas do conhecimento.

A PCR é uma técnica considerada simples, pela qual moléculas de DNA ou cDNA são amplificadas milhares ou milhões de vezes de uma forma bastante rápida. Todo o procedimento é realizado *in vitro*, gerando DNA em quantidade suficiente para análises posteriores. A técnica é muito sensível, possibilitando a amplificação de DNA a partir de uma quantidade mínima de amostra. Essa característica tem proporcionado a aplicação da técnica não apenas na pesquisa básica, mas também na pesquisa aplicada, como nos testes de identificação genética, medicina forense, diagnóstico de doenças infecciosas, controle de qualidade industrial, entre outros.

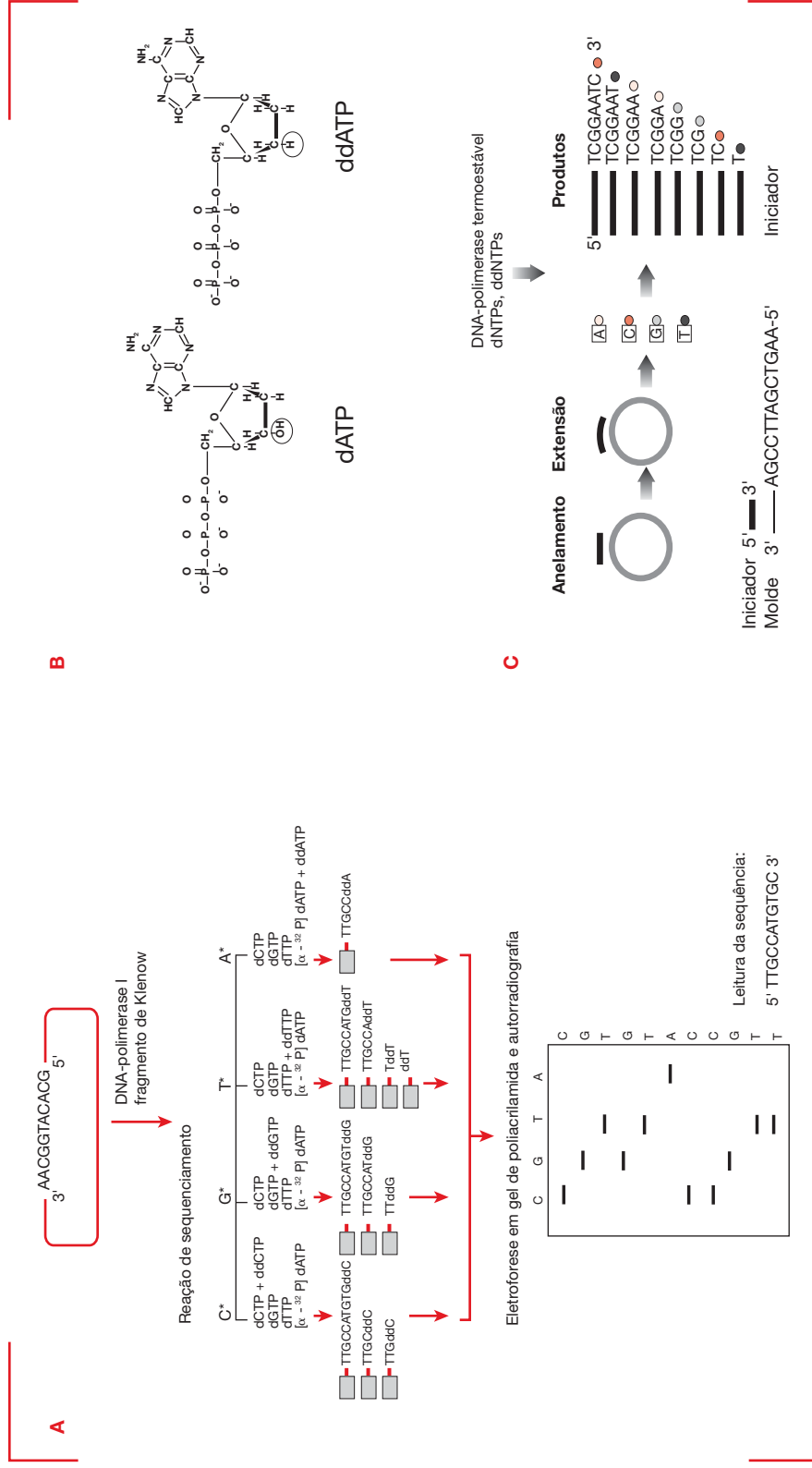


Figura 16.12

Sequenciamento de DNA pelo método de Sanger. (A) Esquema das reações e procedimentos para análise em gel de poliacrilamida; **(B)** representação das estruturas de dATP e ddATP; **(C)** representação esquemática da reação utilizando ddNTPs (terminadores) marcados com fluoróforos.

6.6.1 Princípios do método

A PCR permite a amplificação de segmentos definidos da molécula de DNA (**Figura 16.13**). Em geral, uma reação de amplificação contém o DNA com a sequência-alvo a ser amplificada, uma DNA-polimerase termoestável, dois oligonucleotídeos iniciadores, desoxirribonucleotídeos (dNTPs), tampão de reação e concentração adequada de $MgCl_2$. Os componentes da reação são misturados e a amostra é colocada em um termociclador, um aparelho que possibilita o aquecimento e o resfriamento rápido das amostras. Dessa forma, por exemplo, o aquecimento a $95^\circ C$ possibilita a desnaturação do DNA-alvo, a redução da temperatura para $45-65^\circ C$ (a temperatura a ser escolhida depende da sequência do iniciador, sendo maior para sequências com mais conteúdo de G+C e menor para sequências com menos conteúdo de A+T) proporciona o anelamento dos iniciadores e a temperatura de $68-72^\circ C$ viabiliza a extensão da cadeia de DNA (polimerização). Repetindo-se várias vezes as etapas de desnaturação, anelamento e polimerização (20 a 30 ciclos) possibilitam a amplificação do DNA-molde.

Os primeiros experimentos de PCR foram realizados com o fragmento de Klenow da DNA-polimerase I. Nesse caso, a cada passo de desnaturação a enzima era inativada e mais enzima tinha de ser adicionada. Posteriormente, passou-se a utilizar DNA-polimerases termoestáveis, como a *Taq* e a *Pfu*-DNA-polimerases (Seção 1.4.3), não havendo mais a necessidade de adição de enzima após cada ciclo de desnaturação.

As condições iniciais para a amplificação, tais como concentrações de Mg^{2+} , de enzima e de *primers* e temperatura de hibridização, devem ser definidas para cada par de iniciadores. Cuidados especiais devem ser tomados com possíveis contaminações por DNAs exógenos.

Como a amplificação da região delimitada pelos dois iniciadores ocorre de forma exponencial, após 20 ciclos de amplificação o número de moléculas no final do processo pode chegar a 1 milhão de cópias, considerando-se uma eficiência de 100%. Em geral essa eficiência fica entre 60 e 80%. Por exemplo, quando se usa $1 \mu g$ de DNA genômico total humano para amplificar uma região de 1 kb, essa região corresponde a $0,3 \mu g$ de DNA. Após 30 ciclos de amplificação a quantidade do fragmento de DNA pode chegar a $0,3 \mu g$.

6.6.2 Algumas aplicações da técnica de PCR

6.6.2.1 PCR em tempo real quantitativa (qPCR)

Um avanço importante na técnica de PCR foi o desenvolvimento de metodologias que possibilitaram o acompanhamento da amplificação do DNA em todo o processo e não somente em seu final (PCR convencional). A **PCR em tempo real quantitativa (qPCR)** é um método de detecção e quantificação confiável dos produtos gerados durante cada ciclo de amplificação, os quais são proporcionais à quantidade de molde disponível no início do processo de PCR. É fundamental a presença de um método de detecção da acumulação dos produtos de PCR, e

que o instrumento para realizar a termociclagem esteja capacitado para registrar os resultados de cada ciclo de PCR em tempo real.

Existem diversos trabalhos sobre PCR em tempo real quantitativa e sua variante, a **PCR com transcrição reversa em tempo real quantitativa (qRT-PCR)**.

No caso da qRT-PCR, deve ocorrer, primeiramente, a transcrição reversa (RT, de *reverse transcription*) de moléculas de RNA em cDNA de uma amostra pela enzima transcriptase reversa. A partir da população de moléculas de cDNA é realizada a PCR em tempo real. Muitas das aplicações da qRT-PCR incluem a quantificação dos níveis de expressão de mRNA, a determinação do número de cópias de um fragmento de DNA, a análise do número de cópias e expressão de um transgene, a discriminação alélica, etc.

6.6.2.2 Amplificação de regiões específicas do genoma ou de transcritos

Quando se dispõe de dados sobre a sequência de nucleotídeos de um segmento de DNA a ser analisado é relativamente fácil definir os iniciadores para a amplificação. Quando não se conhece a sequência de nucleotídeos do gene a ser analisado no organismo em estudo, mas existem dados sobre a sequência de nucleotídeos de um gene relacionado de outros organismos, também é possível conseguir a amplificação. Neste caso, a comparação da sequência deduzida de aminoácidos das proteínas codificadas pelos genes dos diferentes organismos permite a seleção de regiões conservadas. Conhecendo os códons mais utilizados para cada aminoácido no organismo em estudo, é possível selecionar um conjunto de iniciadores, em que os oligonucleotídeos apresentem diferenças em uma ou poucas bases, geralmente na terceira base de um códon. Dessa forma, a chance de anelamento correto com o gene em estudo aumenta.

Tanto DNA genômico como cDNA podem ser usados para a amplificação. A amplificação de cDNA pode ser utilizada tanto para conseguir uma sonda específica, como para analisar a expressão do gene em um determinado tipo de célula ou tecido. No caso do cDNA, a amplificação é realizada por RT-PCR (PCR com transcrição reversa), em que o DNA complementar ao RNA é sintetizado, utilizando a transcriptase reversa, e depois amplificado por PCR.

6.6.2.3 Análise de polimorfismos de DNA

A partir da utilização de iniciadores que flanqueiam uma determinada região polimórfica do DNA é possível amplificar um segmento de DNA, que pode ser, então, analisado de diferentes formas, como por sequenciamento direto do DNA ou por clivagem com endonucleases de restrição. A análise dos produtos de PCR por determinação da sequência de nucleotídeos é a mais informativa em relação aos polimorfismos.

Quando se tem informações prévias sobre a presença de polimorfismo em um determinado locus, em nível de sequência de nucleotídeos, métodos rápidos e simples podem

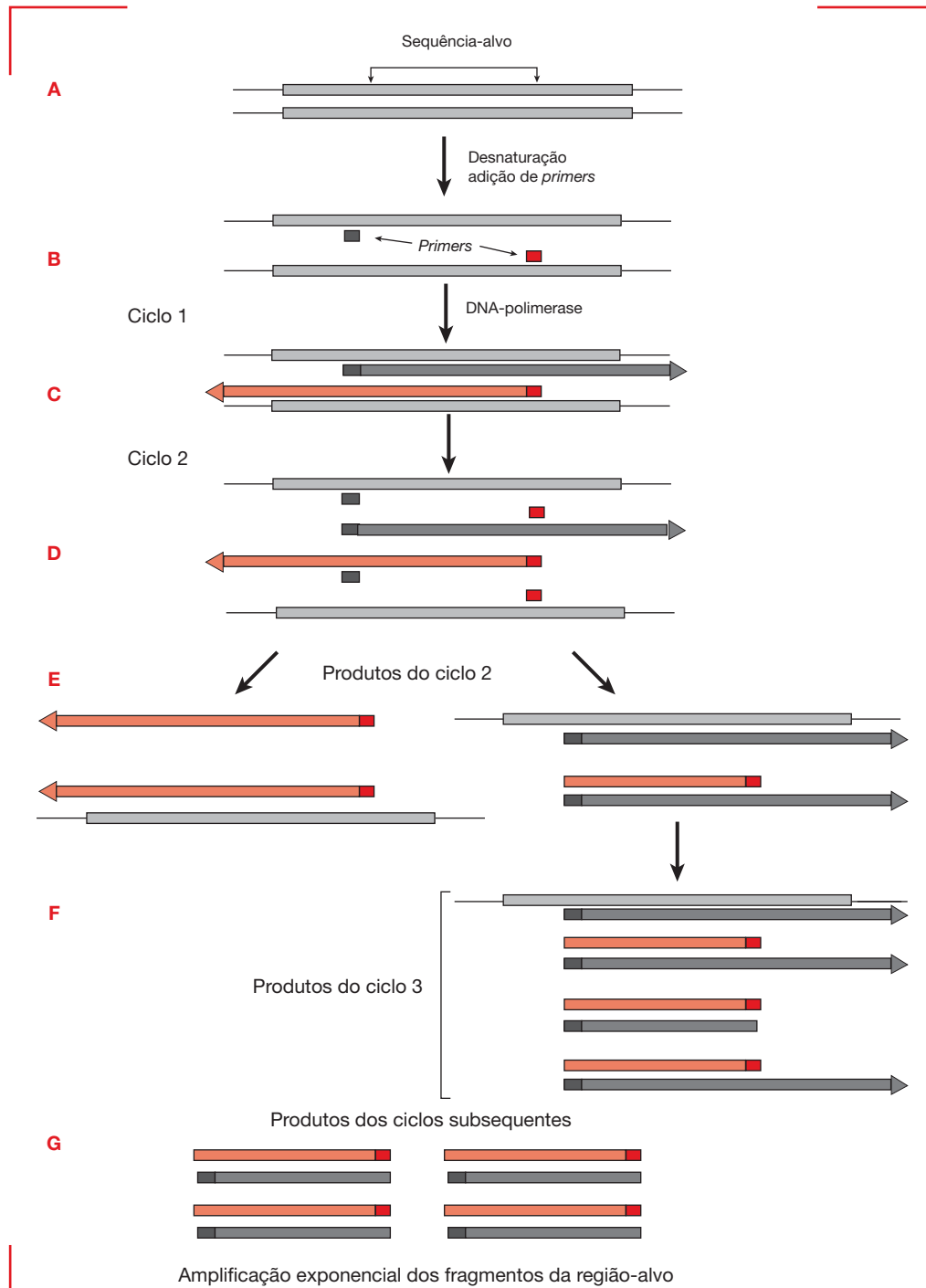


Figura 16.13

Esquema das etapas da reação em cadeia da polimerase. (A) DNA contendo a região-alvo a ser amplificada; **(B)** DNA desnaturado e anelado com os iniciadores (*primers*); **(C)** produtos amplificados após o primeiro ciclo da reação; **(D)** produtos do primeiro ciclo desnaturados e anelados com os iniciadores; **(E)** produtos do segundo ciclo da reação; **(F)** produtos do terceiro ciclo da reação. Somente parte dos fragmentos amplificados é mostrada. **(G)** Produtos dos ciclos subsequentes mostrando a amplificação de fragmentos com tamanhos bem definidos.

ser utilizados. Por exemplo, oligonucleotídeos alelo-específicos podem ser usados para hibridização com os produtos de PCR, aproveitando-se de condições em que a diferença de uma única base pode ser detectada. Esse método tem sido aplicado na detecção de mutações em doenças genéticas, como anemia falciforme, talassemias e fibrose cística, bem como na tipagem de HLA. O polimorfismo pode, também, ser detectado diretamente na PCR. Nesse caso, iniciadores contendo a região 3' correspondente à parte polimórfica são utilizados. Somente ocorre amplificação quando há o pareamento correto das bases na região 3'.

A utilização de um único iniciador com sequência arbitrária pode possibilitar, por PCR, a análise de polimorfismos de DNA. Um número definido de produtos de amplificação é gerado, devido ao fato de as sequências parcialmente complementares à sequência do *primer* estarem em fitas opostas no DNA. Esse método recebeu o nome de **RAPD** (de *random amplified polymorphic DNA* = DNA aleatório polimórfico amplificado).

6.6.2.4 Diagnóstico de doenças genéticas

A técnica de PCR foi utilizada, pela primeira vez, na amplificação de DNA genômico para detecção de mutações associadas à anemia falciforme. Desde então, várias outras doenças, como β -talassemias, distrofia muscular de Duchenne, síndrome de Lesch-Nyham, fenilcetonúria, fibrose cística, doença de Tay-Sachs e doença de Gaucher, têm sido diagnosticadas pela identificação de mutações usando PCR.

Para algumas doenças em que a predisposição genética é bastante alta, como doenças cardiovasculares e doenças autoimunes, o uso da PCR tem sido de grande utilidade. Mutações no gene do receptor de lipoproteína de baixa densidade (LDL) e polimorfismos nos loci da apolipoproteína E (*ApoE*) e apolipoproteína A (*ApoA*) têm sido detectadas por PCR, e relacionadas com o risco de doenças cardiovasculares. Doenças autoimunes, como diabetes melito insulina-dependente, esclerose múltipla e artrite reumatoide, estão associadas a alelos específicos de loci HLA de classe II.

6.6.2.5 Medicina forense

As sequências **VNTR** (de *variable number of tandem repeats* = número variável de repetição em tandem) têm sido analisadas por PCR. Essas sequências nucleotídicas curtas, repetidas de 20 a 100 vezes, apresentam um padrão que varia de pessoa para pessoa, a menos que sejam geneticamente idênticas, como no caso de gêmeos univitelinos. Esse tipo de análise surgiu como uma alternativa mais simples e rápida para utilização em medicina forense, em substituição à técnica de RFLP.

6.6.2.6 Diagnóstico de doenças infecciosas

A técnica de PCR tem tido uma grande influência no desenvolvimento do diagnóstico de doenças provocadas por vírus, bactérias, fungos e protozoários. Tanto a rapidez e sensibilidade como a especificidade, foram aumentadas com a introdução da PCR.

Vírus, como o da imunodeficiência humana tipos 1 e 2 (HIV-1 e HIV-2), do linfotrófico de célula T tipos I e II

(HTLV-I e HTLV-II), que causam leucemia, da hepatite B (HBV), do citomegalovírus (CMV) e do papiloma humano (HPV), bem como protozoários parasitas, tais como *Giardia lamblia*, *Leishmania donovani* e *Trypanosoma*, têm sido detectados por PCR. Infecções bacterianas, como aquelas causadas por *Mycobacterium tuberculosis*, *Neisseria meningitidis* e *Vibrio cholera* também têm sido diagnosticadas por essa técnica.

7.1 Técnicas de análise de genomas e produtos de transcrição em grande escala

Os investimentos iniciados no final da década de 1980 no projeto de sequenciamento do genoma humano estimularam o desenvolvimento de novas tecnologias de sequenciamento e análise de genomas, permitindo uma revolução nos estudos genômicos dos mais diversos organismos. Eles culminaram com a publicação dos dados do sequenciamento completo do genoma humano em 2001, um dos fatos mais marcantes da ciência no início do século.

7.1 Sequenciamento de genomas

Desde o desenvolvimento da técnica de sequenciamento de DNA por Sanger, muitos foram os avanços tecnológicos que permitiram a automação do processo. Como visto anteriormente, novas tecnologias, como a de **pirosequenciamento**, (ver **Material Complementar Online 16.2**), vêm sendo desenvolvidas, possibilitando mais rapidez na decifração das sequências dos genomas e um custo muito menor.

A estratégia mais comum para o sequenciamento de genomas tem sido a de sequenciamento por *shotgun*, utilizando o método de Sanger (Seção 6.4.1). Nessa estratégia, o DNA genômico é quebrado aleatoriamente e os fragmentos, em geral de 2 a 5 kb, são clonados em um plasmídeo. As extremidades dos insertos são sequenciadas e essas sequências são montadas utilizando ferramentas de bioinformática (ver Capítulo 17). Após a **montagem do genoma**, os genes são identificados por um procedimento chamado de **anotação** (ver Capítulo 17). A **Figura 16.14** mostra de forma esquemática os procedimentos utilizados para o sequenciamento de um genoma bacteriano.

7.2 Análise global de transcrição (transcritômica)

7.2.1 Etiquetas de sequências expressas

As **etiquetas de sequências expressas (ESTs, de *expressed sequence tags*)** são sequências curtas de DNA (200-500 pb) geradas a partir do sequenciamento

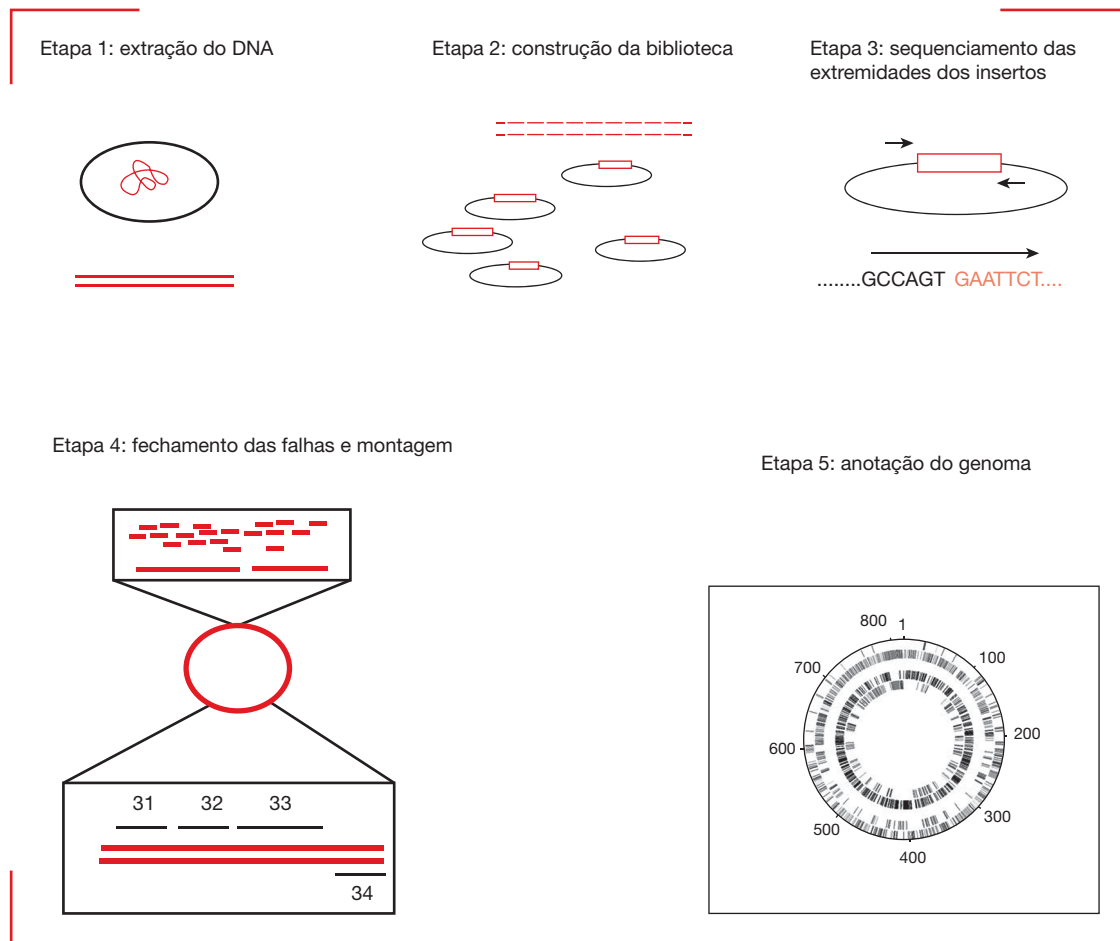


Figura 16.14

Esquema dos procedimentos utilizados para o sequenciamento de um genoma bacteriano. O DNA da bactéria é isolado, fragmentado e clonado em plasmídeo. As extremidades dos fragmentos clonados são sequenciadas, gerando um número suficiente de seqüências que permitem a montagem (em geral 15 mil seqüências a cada 10^6 pb). Após a montagem, o genoma é anotado, isto é, os genes são identificados e suas possíveis funções são anotadas.

das extremidades 3'-OH e 5'-P de clones de cDNA selecionados ao acaso. O método, descrito pelo grupo de Craig Venter, em 1991, vem sendo aplicado para a análise da transcrição de genes nos mais diferentes tipos de organismos. O objetivo do sequenciamento e análise de ESTs é gerar, rapidamente, dados sobre a transcrição dos genes que codificam proteínas de um determinado tipo de célula, tecido, órgão ou organismo. Essas seqüências também ajudam na identificação e posicionamento dos éxons dos genes nos genomas. As ESTs são geradas com rapidez, e bancos de dados públicos contendo ESTs estão disponíveis para livre acesso pela internet.

Uma alternativa às ESTs é a utilização da tecnologia denominada **etiquetas de seqüências expressa de fases abertas de leitura (ORESTES, de open reading frame expressed sequence tags)**. Os detalhes dessa tecnologia são apresentados no **Material Complementar Online 16.3**.

7.2.2 Macroarranjos e microarranjos

As mudanças no padrão de expressão gênica têm uma grande implicação no funcionamento das células, tecidos e organismos. Essas mudanças podem estar relacionadas ao desenvolvimento, à manifestação de doença e à morte celular. A tecnologia de arranjos de DNA permite a análise de variações no padrão de expressão de uma forma rápida e confiável.

Os **macroarranjos de DNA** constituem-se de membranas de náilon (ou outro tipo), em que os segmentos de DNA gerados pela PCR são aplicados ocupando superfícies com diâmetros de 0,5 a 1 mm e com um espaçamento de aproximadamente 2 mm entre eles. Cerca de 25 amostras podem ser aplicadas em uma área de 1 cm^2 . Dessa forma, em uma membrana do tamanho de uma microplaca de titulação (8 x 12 cm), alguns milhares de produtos de amplificação, correspondentes a diferentes genes, podem ser aplicados. Essas membranas são

submetidas a processos de hibridização com sequências de DNA marcadas (por métodos radiativos ou colorimétricos) e os resultados são registrados por equipamentos que possibilitam a detecção da radiação ou da fluorescência emitidas pelas sondas.

Os **microarranjos de DNA** consistem em milhares de sequências de DNA (cDNA ou oligonucleotídeos) imobilizadas em uma superfície de uma lâmina (geralmente de vidro) de aproximadamente 7,5 x 2,5 cm. Milhares de amostras podem ser fixadas por cm² e uma lâmina pode conter, por exemplo, sequências correspondentes a todos os genes que codificam as proteínas de um organismo. A hibridização é feita com alvos (sequências de DNA) marcados com compostos fluorescentes, esses sinais são captados e as imagens processadas para análise, por exemplo, do padrão de expressão gênica.

Os primeiros experimentos utilizando a tecnologia dos **microarranjos de DNA** foram realizados na década de 1990. A tecnologia de microarranjos tem sido bastante utilizada na análise do transcrito de diferentes células, tecidos e organismos, tanto nas análises de questões básicas da expressão gênica em nível transcricional, como na área de pesquisa clínica, visando a análise dos padrões de expressão gênica em diferentes condições patológicas.

7.2.3 Sequenciamento em grande escala

Considerando algumas limitações da tecnologia de microarranjos, como a sensibilidade insuficiente, a hibridização inespecífica e o fato de que essa técnica fornece informações somente sobre os transcritos cujas sequências estejam incluídas no arranjo, o **sequenciamento em grande escala** de cDNAs tem sido considerado como uma alternativa importante para a análise quantitativa da expressão gênica em nível transcricional.

A estratégia de sequenciamento na análise de expressão gênica tem sido associada à tecnologia denominada **SAGE** (de *serial analysis of gene expression* = análise seriada da expressão gênica). Essa tecnologia se baseia na clonagem de sequências curtas (14 a 21 pb) de cDNA interligadas, formando uma série de etiquetas (*tags*), que são posteriormente sequenciadas. Nessa estratégia, assume-se que uma sequência curta de uma *tag* contenha informação suficiente para a identificação do gene correspondente, evidenciando a transcrição. Estima-se que mais de 99,8% das *tags* de 21 pb ocorrem somente uma vez no genoma humano. Uma limitação dessa tecnologia é a necessidade de clonagem e sequenciamento pelo método de Sanger, tornando-a trabalhosa e cara. Utilizando as novas plataformas de sequenciamento de DNA citadas anteriormente neste capítulo, tem sido possível analisar milhões de fragmentos de DNA simultaneamente, gerando gigabases (bilhões de bases) de informações de sequências em uma única “corrida”.

8. Técnicas de análise global de proteínas

As informações geradas pelo sequenciamento dos genomas dos mais diversos organismos formaram a base para a análise e interpretação da função dos genes em grande escala. No caso dos genes que codificam proteínas, uma etapa fundamental é a identificação das proteínas expressas em um determinado momento ou determinada situação, em que se encontram as células, os tecidos, os órgãos ou o organismo. O conjunto dessas **PROTEÍNAS**, expressas por um genOMA, é chamado de **proteoma**.

Nos últimos anos tem existido um grande esforço para identificar o repertório de proteínas expresso por diferentes células, tecidos, órgãos ou organismos e, também, para analisar as interações que ocorrem entre as proteínas e drogas específicas, bem como para descobrir a atividade das proteínas e suas estruturas. Esses esforços fazem parte de uma nova área de pesquisa, denominada de **proteômica**.

As principais técnicas utilizadas na análise proteômica envolvem o fracionamento das proteínas por técnicas eletroforéticas e cromatográficas e sua identificação por espectrometria de massas.

8.1 Técnicas de fracionamento de proteínas associadas à proteômica

A separação de amostras complexas de proteínas ou peptídeos é uma etapa fundamental para a análise proteômica por espectrometria de massas. Os métodos mais utilizados são: (1) eletroforese em gel de poliácridamida em presença do detergente dodecil sulfato de sódio (SDS); (2) eletroforese bidimensional em gel de poliácridamida; e (3) cromatografia líquida, que pode envolver cromatografia de afinidade, troca iônica ou fase reversa.

8.1.1 Eletroforese em gel de poliácridamida (SDS-PAGE)

A **eletroforese em gel de poliácridamida em presença de SDS (SDS-PAGE)**, desenvolvida por Laemmli, em 1970, é uma das técnicas mais utilizadas para separação de proteínas. As proteínas complexadas com SDS são desnaturadas, apresentam carga negativa e a separação ocorre pela diferença das massas moleculares (**Figura 16.15A**). Ela pode ser utilizada para separação de amostras de proteínas com baixa complexidade para posterior análise por espectrometria de massas. Embora seja uma técnica fácil, seu poder de resolução é limitado. Por exemplo, muitas vezes uma banda única de proteína bem definida em um gel SDS-PAGE, proveniente de uma mistura complexa de proteínas, quando analisada por espectrometria de massas mostra-se constituída por várias proteínas, em geral por mais de dez.

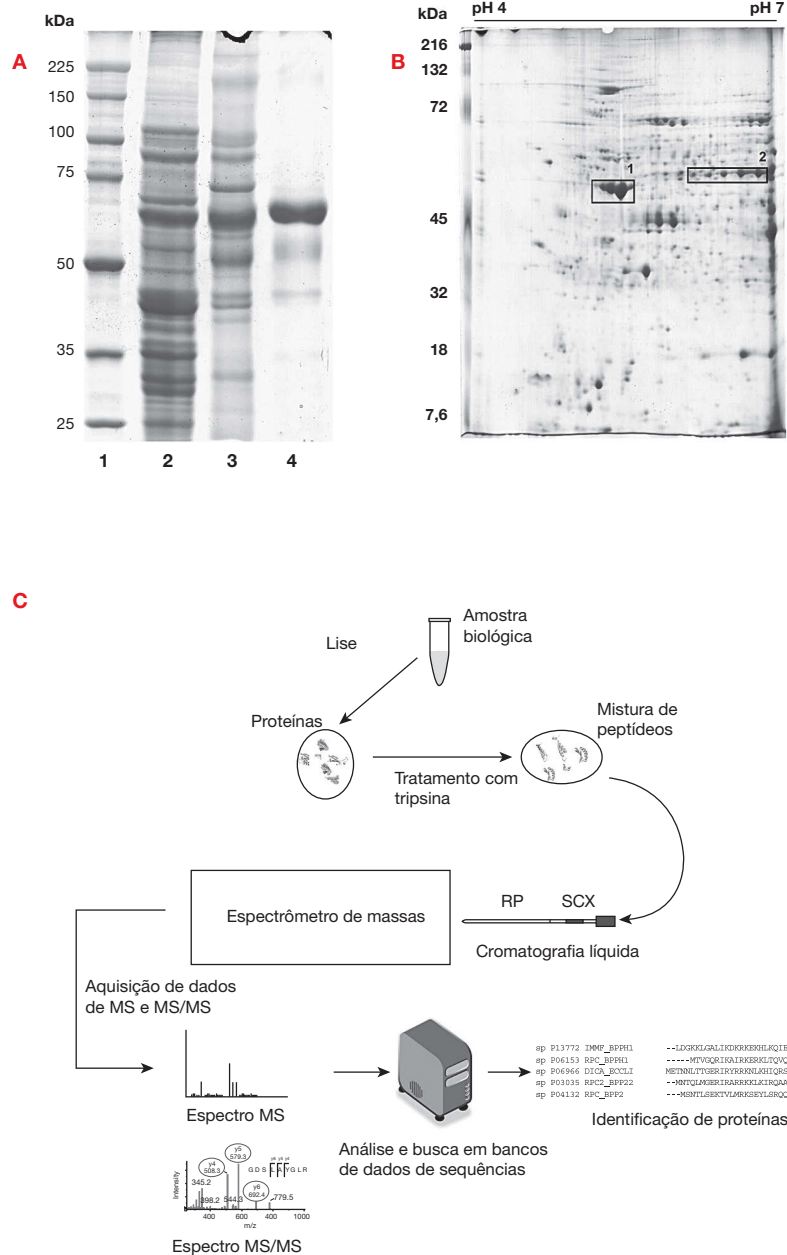
8.1.2 Eletroforese bidimensional

A separação das proteínas por **eletroforese bidimensional (2-DE)** baseia-se na focalização isoeletri-

Figura 16.15

Análise de proteínas por eletroforese em gel de poli-acrilamida e peptídeos por cromatografia líquida.

(A) Análise de proteínas do helminto parasita *Echinococcus granulosus* por SDS-PAGE; (1) Marcadores de massa molecular. (2-4) Extratos de diferentes estágios do parasita. (B) Análise de proteínas do helminto parasita *E. granulosus* em gel 2-DE. Os números mostram grupos de spots identificados como uma única proteína, (1) Actina e (2) Enolase. (C) Representação esquemática das etapas para separação de peptídeos por cromatografia líquida e análise por espectrometria de massas. RP, *reverse phase* (fase reversa), SCX, *strong cation exchange* (troca catiônica forte). As fotografias dos géis em A e B foram cedidas por Karina Lorenzatto (CBiot, UFRGS) e Dra. Karina M. Monteiro (CBiot, UFRGS), respectivamente.



ca, seguida da análise por SDS-PAGE. Na **focalização isoeétrica (IEF)**, de *isoelectric focusing*, a proteína é submetida a um campo elétrico e separada conforme seu ponto isoeletrico (*pI*) que, por sua vez, é o pH no qual a proteína apresenta carga líquida zero. Quando uma proteína é colocada em um meio com um gradiente linear de pH e submetida a um campo elétrico, ela se moverá em direção ao eletrodo com a carga contrária à carga líquida da proteína. Durante essa migração no gradiente de pH, a proteína ganhará ou perderá prótons, alterando sua carga líquida e mudando o padrão de migração. A proteína atingirá um ponto do gradiente

de pH com valor igual ao seu *pI* e, por apresentar carga zero, irá parar de migrar.

Logo depois que a SDS-PAGE começou a ser utilizada, a IEF foi aplicada por O'Farrell, em 1975, na análise de amostras proteicas, como uma etapa anterior à análise por SDS-PAGE, criando a técnica de eletroforese bi-dimensional (2-DE). Um dos grandes avanços na técnica foi a substituição do gel utilizado na IEF pelas tiras com **gradientes de pH imobilizados**, que não se alteram durante o processo eletroforético. Os procedimentos para análise de proteínas por 2-DE são apresentados no **Material Complementar Online, Figura 1**.

A análise do conjunto de proteínas de uma célula, tecido ou organismo por 2-DE, associada à análise por espectrometria de massas (ver Seção 8.2), permite gerar um **mapa proteômico** de referência que pode ser utilizado como ferramenta importante na análise de expressão gênica diferencial, análise de mutantes ou produção de organismo geneticamente modificado. A Figura 16.15B mostra um exemplo de análise de proteínas em géis 2-DE, indicando algumas proteínas que foram identificadas por espectrometria de massas. Nesse tipo de análise, determinadas proteínas foram identificadas em múltiplos *spots*, com diferentes *pI*s, sugerindo possíveis isoformas da proteína ou a presença de modificações pós-traducionais, como a fosforilação.

8.1.3 Cromatografia líquida

Apesar da grande utilidade da 2-DE nas análises proteômicas existem algumas limitações na técnica, como a detecção de proteínas pouco abundantes, proteínas muito grandes ou pequenas, proteínas básicas e de membrana (hidrofóbicas). Além disso, a automação do processo é difícil, impossibilitando a aplicação da técnica em análises de proteínas em grande escala. Como alternativa à 2-DE, vários métodos de **cromatografia líquida (LC, de liquid chromatography)** podem ser utilizados. Na LC, uma solução contendo uma mistura de peptídeos (fase móvel) é passada por uma coluna contendo partículas porosas, com características particulares de ligação (fase estacionária). As interações das proteínas com a fase estacionária possibilitam a sua separação. Na cromatografia de troca iônica, a fase estacionária é carregada e as proteínas são separadas de acordo com a carga. Na cromatografia de fase reversa, a fase estacionária é apolar, ao passo que a fase móvel é polar, possibilitando que as proteínas sejam separadas pelas suas características hidrofóbicas.

A tecnologia multidimensional para identificação de proteínas (MudPIT, de *Multidimensional Protein Identification Technology*) foi desenvolvida como um método para analisar amostras complexas. Essa tecnologia envolve o fracionamento proteico usando cromatografia de troca iônica, seguido por cromatografia de fase reversa (RP, de

reverse phase) e espectrometria de massas em tandem. De forma resumida, o extrato de proteínas de interesse é tratado com tripsina e a mistura de peptídeos é ligada a uma resina de troca catiônica forte (SCX, de *strong cation exchange*). Os peptídeos são eluídos por uma fase móvel de acetato de amônia para uma resina de RP, sendo posteriormente eluídos para o espectrômetro de massas por um gradiente com concentrações crescentes do solvente orgânico acetonitrila em água. A Figura 16.15C mostra de forma esquemática, as etapas da análise de peptídeos separados por LC e a identificação por espectrometria de massas.

8.2 Espectrometria de massas na análise de proteínas

As tecnologias de espectrometria de massas constituem a base para a grande maioria dos estudos envolvendo análises proteômicas. Os avanços nos estudos proteômicos estão relacionados: com o aumento da disponibilidade de dados de sequências de genomas, que surgiu após o sequenciamento dos primeiros genomas, em 1995; os avanços nos métodos de ionização de peptídeos; o desenvolvimento de métodos computacionais para identificação dos peptídeos; e o aumento na resolução dos analisadores dos espectrômetros de massas. Os espectrômetros de massa consistem em uma fonte de íon, que converte as moléculas a serem analisadas em íons em uma fase gasosa; um analisador de massas, que separa as moléculas ionizadas com base na relação massa/carga (m/z); e um detector, que registra o número de íons em cada valor de m/z (**Figura 16.16**).

Os dois tipos principais de ionizadores utilizados nas análises proteômicas são ionização e dessorção a *laser* assistida por matriz (*Matrix-Assisted Laser Desorption Ionization*, MALDI) e a ionização por *electrospray* (*Electrospray Ionization*, ESI). Os tipos de analisadores mais importantes são o tempo de voo (*time-of-flight*, TOF) e o quadrupolo (Q). Os detalhes do funcionamento desses ionizadores e analisadores são apresentados no **Material Complementar Online 16.4**.

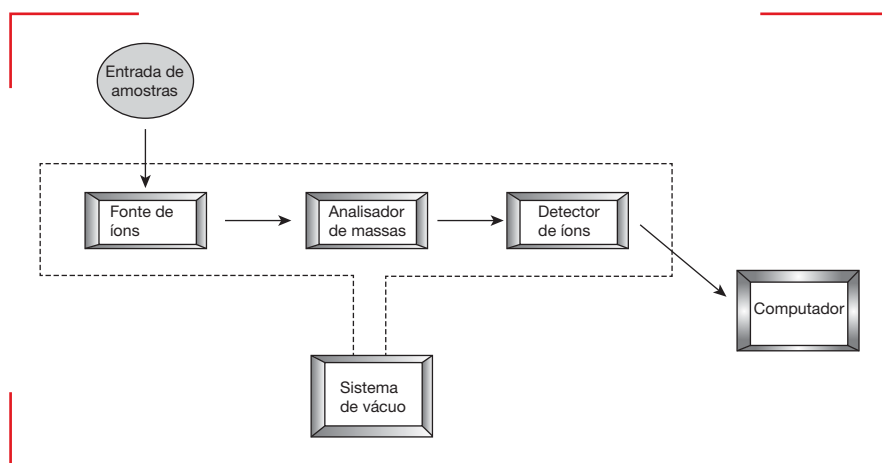


Figura 16.16

Espectrômetro de massas.

Esquema de espectrômetro de massas mostrando a fonte de íon, que converte as moléculas a serem analisadas em íons em uma fase gasosa; o analisador de massas, que separa as moléculas ionizadas com base na relação massa/carga (m/z); e o detector, que registra o número de íons em cada valor de m/z .



Resumo

A tecnologia do DNA recombinante compreende diversas técnicas aplicadas aos estudos de biologia molecular. Enzimas que permitem a manipulação *in vitro* de ácidos nucleicos, como as endonucleases de restrição, as DNA-ligasas, as DNA-polimerases, entre outras, são amplamente utilizadas. Na clonagem molecular, um segmento de DNA de interesse (contendo um gene, por exemplo) pode ser ligado *in vitro* a um vetor e ser introduzido em uma célula, em que um grande número de cópias da sequência de interesse é então produzido *in vivo*. Há vários tipos de vetores, como plasmídeos, fagos, cosmídeos e cromossomos artificiais, adequados à clonagem, à seleção genética de clones e até à expressão de sequências clonadas em sistemas celulares procarióticos ou eucarióticos.

A clonagem pode ser feita em grande escala, de modo a permitir que, em um conjunto de clones (biblioteca), um genoma completo (biblioteca genômica) ou genes que estão sendo expressos (biblioteca de cDNA) estejam representados. Segmentos de DNA também podem ser amplificados *in vitro*, utilizando a reação em cadeia da polimerase (PCR), que tem inúmeras variantes e aplicações. Sequências de DNA clonadas ou amplificadas podem ser fisicamente mapeadas e analisadas por métodos de hibridização ou sequenciamento. Com o aprimoramento das metodologias de sequenciamento de DNA foi viabilizado o sequenciamento completo de genomas e, a partir disso, foram também desenvolvidas técnicas para análise dos produtos de expressão gênica em grande escala, tanto em nível de transcritos (transcritômica) como de proteínas (proteômica).



Leituras recomendadas

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, et al. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*. 1991;252(5013):1651-6.

Brown TA. Clonagem gênica e análise do DNA. 4. ed. Porto Alegre: Artmed; 2002.

Cañas B, Piñeiro C, Calvo E, López-Ferrer D, Gallardo JM. Trends in sample preparation for classical and second generation proteomics. *J Chromatogr A*. 2007;1153(1-2):235-58.

Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods*. 2008;5(7):613-9.

Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science*. 2006;312(5771):212-7.

Finehout EJ, Lee KH. An introduction to mass spectrometry applications in biological research. *Biochem Mol Biol Educ*. 2004;32(2):93-100.

Fraser CM, Eisen JA, Salzberg SL. Microbial genome sequencing. *Nature*. 2000;406(6797):799-803.

Gstaiger M, Aebersold R. Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat Rev Genet*. 2009;10(9):617-27.

Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet*. 2006;7(3):200-10.

J. Craig Venter Institute [Internet]. Rockville: J. Craig Venter Institute; 2011 [capturado 28 ago. 2011]. Disponível em: <http://www.jcvi.org/>.

Macleán D, Jones JD, Studholme DJ. Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol*. 2009;7(4):287-96.

Motoyama A, Yates JR 3rd. Multidimensional LC separations in shotgun proteomics. *Anal Chem*. 2008;80(19):7187-93.

PlantGDB.org [Internet]. [capturado 28 ago. 2011]. Disponível em: <http://www.plantgdb.org/>.

Sambrook J, Russel DW. Molecular cloning: a laboratory manual. 3rd ed. New York: Cold Spring Harbor Laboratory, 2001.

Suzuki Y, Yoshitomo-Nakagawa K, Maruyama K, Suyama A, Sugano S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene*. 1997;200(1-2):149-56.

Ana Tereza Ribeiro de Vasconcelos | Darcy Fontoura de Almeida

Bioinformática na Análise de Genes e Genomas

1. Redes de informações 364

- 1.1 Internet 364
- 1.2 Rede mundial de comunicação 365
- 1.3 HTTP, HTML e URLs 365

2. Alinhamentos de sequências 365

- 2.1 Alinhamentos de pares de sequências 365
 - 2.1.1 Dot Plot 368
 - 2.1.2 FastA 368
 - 2.1.3 BLAST 368
- 2.2 Alinhamentos múltiplos 370

3. Análise de genomas 370

- 3.1 Montagem de genomas 370
- 3.2 Anotação de genomas 371
- 3.3 Análise comparativa de genomas 371
 - 3.3.1 MUMmer 372
 - 3.3.2 ACT 372
 - 3.3.3 PipMaker 372
 - 3.3.4 M-GCAT 372

4. Bancos de dados biológicos 372

- 4.1 Banco de dados primários 373
 - 4.1.1 GenBank 373
 - 4.1.2 EMBL Nucleotide Sequence Database 376
 - 4.1.3 DNA Data Bank of Japan 377
 - 4.1.4 UniProt Knowledgebase 377
- 4.2 Banco de dados secundários 377
 - 4.2.1 PROSITE 377
 - 4.2.2 PRINTS 377
 - 4.2.3 Pfam 378
 - 4.2.4 InterPro 378
- 4.3 Banco de dados de estruturas de proteínas 378
 - 4.3.1 SCOP 378
 - 4.3.2 CATH 378
 - 4.3.3 PDB 379
- 4.4 Bancos de dados de genômica comparativa 379
 - 4.4.1 MBGD 379
 - 4.4.2 xBASE: 379
 - 4.4.3 MicrobesOnline 379
 - 4.4.4 HOGENOM 379
- 4.5 Bancos de dados funcionais e de metabolismo 379
 - 4.5.1 COG 379
 - 4.5.2 KEGG 380

Em meados da década de 1980, o Departamento de Energia (DoE) dos Estados Unidos América, deu início a um projeto cujo objetivo era o de realizar o sequenciamento completo do genoma humano. As sequências produzidas por esse projeto, assim como todas as outras informações pertinentes, foram armazenadas em bancos de dados computacionais que têm sido fonte de variados estudos em biologia e em medicina. O elevado volume de informação nesses bancos de dados exige o uso de métodos computacionais sofisticados e eficientes, que constituem a base para a identificação de novos genes e para a comparação entre novos genomas.

Como resultado daquele projeto foi iniciado o sequenciamento de genomas de organismos-modelo utilizados intensivamente em pesquisas. Entre eles podemos citar os casos da *Escherichia coli*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana*, *Canis familiaris* e *Mus musculus*, todos já concluídos (ver Capítulos 4 e 5).

Surgem desta forma duas áreas de pesquisa que evoluem em conjunto (um sinal da sua transdisciplinaridade): a **bioinformática**, que pode ser descrita como aquisição, análise e armazenamento de informação biológica, principalmente sob a forma de ácidos nucleicos e proteínas; e a **biologia molecular computacional**, que cuida do desenvolvimento de algoritmos e programas computacionais para resolver problemas nessa área. Ambas têm se desenvolvido muito nas últimas três décadas, em razão do número crescente de projetos de sequenciamento de genomas em curso. Um dos objetivos fundamentais dessas áreas é converter a informação contida na sequência de nucleotídeos ou de proteínas dela derivada (linguagem biológica), em conhecimento bioquímico e biofísico (funções estruturais, funcionais e evolutivas). Na realidade, o que se procura é decodificar uma linguagem desconhecida, extraíndo o seu significado biológico. Esta linguagem pode ser decomposta em letras (nucleotídeos e/ou aminoácidos), palavras (motivos) e frases (proteínas). Assim como na estrutura de uma língua, em que a substituição de uma única letra em uma palavra pode alterar o seu significado (p. ex., casa-capa), a substituição de um único aminoácido pode causar uma mudança na função de uma proteína (p. ex., a substituição do ácido glutâmico pela valina na cadeia da hemoglobina A humana causa um tipo particular de anemia).

Pode-se citar, ainda, como um dos objetivos dessas áreas, a localização e a identificação de genes, éxons, íntrons e sequências reguladoras. Para isso, são necessárias comparações de grande abrangência, compreendendo não só sequências de espécies relacionadas como também de outras espécies.

Um outro objetivo importante é o entendimento da estrutura da proteína, pois ela é fundamental para a determinação da função do produto gênico. A função da proteína depende da sua conformação, a estrutura tridimensional (3D) que a proteína assume no meio biológico. Como os padrões ou motivos de estrutura 3D que as proteínas assumem são mais conservados do que os motivos de sequência

primária de aminoácidos, este tipo de homologia é o mais adequado ao seu estudo. Certos motivos podem ter funções similares em diferentes proteínas e, assim, esse tipo de informação pode ser importante para análise de genomas.

Como consequência direta da grande quantidade de informações que são geradas, bancos de dados de sequências de nucleotídeos e de aminoácidos foram criados. Os bancos de dados envolvendo sequências de nucleotídeos, de aminoácidos ou estruturas de proteínas podem ser classificados em bancos de sequências primários, secundários e terciários, e estão disponíveis para serem analisados. Todo ano a revista *Nucleic Acids Research* (NAR) dedica seu primeiro número ao *Database Issue*, em que são publicados atualizações sobre os bancos de dados já existentes e os novos que surgiram durante o ano anterior. Por exemplo, no ano de 2010 surgiram 58 bancos novos e outros 73 foram atualizados. À medida que novas tecnologias de sequenciamento são desenvolvidas, a atualização dos bancos de dados e a quantidade de programas desenvolvidos, de domínio público ou privado, para análises dos resultados, cresce muito e há uma necessidade de atualização constante. Serão descritas neste capítulo somente algumas das técnicas mais tradicionais que são utilizadas para buscas nestes bancos de dados. Com certeza algumas delas já estão em desuso ou poderão ficar desatualizadas em breve.

Os endereços eletrônicos das páginas de instituições, bancos de dados e programas citados no capítulo estão no **Material Complementar Online 17.2**.

1. Redes de informações

1.1 Internet

A internet é uma rede mundial de computadores interligados. Essa rede utiliza um protocolo de comunicações denominado TCP/IP (de *Transmission Control Protocol/Internet Protocol* = protocolo de controle de transmissão/protocolo de interconexão), cujo objetivo é permitir a interação entre vários computadores localizados em diferentes pontos. Nessas redes os computadores são vistos como nós e comunicam-se por meio da transferência de pacotes de dados. Se um dos nós dessa rede se encontra indisponível, o protocolo da rede procura uma via alternativa. Para facilitar esta comunicação entre os nós, cada computador na internet tem um único número de identificação (endereço de IP), descrito no formato ilustrado abaixo:

146.134.8.90

Esse padrão numérico de endereço foi criado para ser inteligível apenas para os computadores. Por isso, foi implementada uma forma simplificada de interação com o usuário, com base em uma hierarquia de domínios em que os números decimais são substituídos por nomes. Assim, o endereço de IP 146.134.8.90 é substituído por banana.lncc.br, em que “banana” identifica uma determinada máquina, “lncc” a localização desta máquina, e

“br” indica o domínio (e possivelmente o subdomínio) ao qual a máquina pertence.

Outras facilidades utilizadas pela computação remota são, por exemplo, o FTP (de *File Transfer Protocol* = protocolo de transferência de arquivos), cujo objetivo é transmitir arquivos entre máquinas, e o *Telnet protocol*, permitindo que usuários situados em localidades diferentes possam se comunicar em tempo real. Pode-se contar, também, com os já bem difundidos correio eletrônico (*e-mail*) e grupos de discussão (*newsgroups*), entre outros, que permitem uma comunicação rápida e eficiente entre os usuários.

1.2 Rede mundial de comunicação

A **Rede Mundial de Comunicação**, também conhecida como **World Wide Web**, **Web**, **WWW** ou **W3**, foi desenvolvida pelo *Conseil Européen pour la Recherche Nucléaire* (CERN), para que informações pudessem ser compartilhadas e difundidas entre grupos da comunidade de física de alta energia dispersos internacionalmente. Porém, o conceito de divisão da informação entre localidades remotas foi adotado por várias áreas e teve um grande impacto na bioinformática. A *Web* é tão popular que, muitas vezes, é entendida como sinônimo de internet. Contudo, o grande desenvolvimento da internet só ocorreu com o surgimento dos **navegadores** (*browsers*) que permitiram o acesso da informação em **sítios** (*sites*) diferentes. Os navegadores são clientes que se comunicam com servidores, utilizando um conjunto padrão de protocolos e convenções. O primeiro ponto de contato entre o cliente e o servidor é a chamada **página inicial** (*home page*). Por meio dessa página, o navegador estabelece uma interface de fácil acesso, permitindo a recuperação de documentos, acessos a arquivos, busca em bancos de dados etc.

Com o passar do tempo um novo desafio surgiu: o conceito de **computação em nuvem** (*cloud computing*), que se refere à utilização da memória, capacidade de armazenamento, cálculo de computadores e servidores compartilhados, interligados pela internet e seguindo o princípio da **computação em grade**. Esse nível de abstração tenta resolver muitas das limitações encontradas quando se lida com um grande conjunto de dados, como é o caso na bioinformática.

1.3 HTTP, HTML e URLs

Os documentos exibidos pelos navegadores exploram técnicas de **hipertexto** e **hipermídia**. Documentos de hipertexto contêm **ligações** (*hyperlinks*) embutidas no texto, com o objetivo de permitir que o usuário seja direcionado a outros documentos. As ligações geralmente são sublinhadas e apresentadas em cor diferente do texto corrente. A comunicação entre as ligações é transparente e o nome dado a este protocolo de comunicação utilizado pelos servidores na *Web* é **HTTP** (de *HyperText Transport Protocol* = protocolo de transporte de hipertexto).

Os documentos em hipertexto são escritos em uma linguagem chamada de **HTML** (de *HyperText Markup Language* = linguagem de marcação de hipertexto), que é uma linguagem de fácil e rápido aprendizado. Estes documentos são acessados por meio de um único endereço ou **URL** (de *Uniform Resource Locators* = localizador padrão de recursos).

2. Alinhamentos de sequências

Quando análises comparativas são feitas, procura-se similaridades e diferenças entre sequências de nucleotídeos ou aminoácidos, com o propósito de inferir analogias estruturais e/ou funcionais e relações evolutivas entre as moléculas envolvidas. A maioria das análises de genes e de genomas se baseia nesse tipo de técnica.

O método comparativo mais utilizado é o **alinhamento de sequências**, que faz uma associação explícita entre resíduos de nucleotídeos ou aminoácidos de duas ou mais sequências. Um dos objetivos do alinhamento é determinar quando duas sequências são suficientemente similares para justificar uma inferência de homologia. Por **similaridade** entende-se uma quantidade observável que expressa esse conceito intuitivo. **Homologia**, por outro lado, refere-se ao fato de dois genes ou seus respectivos produtos partilharem ou não uma história evolutiva comum (ver Capítulo 4).

O objetivo básico do processo de inferência de relações evolutivas (o que a maioria dos métodos de alinhamento tenta) é modelar os mecanismos moleculares pelos quais as sequências evoluíram. O alinhamento de sequências é um dos métodos mais utilizados na área da bioinformática.

2.1 Alinhamentos de pares de sequências

Alinhar duas sequências consiste em estabelecer uma correspondência entre as bases dessas sequências, de modo que a ordem não seja violada. Por ordem entende-se que, se as bases na posição n_1 e n_2 ($n_1 < n_2$) de uma sequência estão respectivamente associadas às bases nas posições m_1 e m_2 da outra sequência, então $m_1 < m_2$. Não é necessário que todas as bases de uma sequência estejam associadas às da outra sequência no alinhamento, mas é desejável que o número de associações seja o maior possível.

Uma maneira muito simples de visualização do alinhamento de sequências consiste na utilização de **lacunas** (*gaps*) entre as bases das sequências, de modo que o alinhamento ocorra verticalmente, como ilustrado abaixo:

SLNSGYHFC	SLNSG---YHFC
SFQETFLSFHFC	SFQETFLSFHFC

Para buscar alinhamentos ótimos é necessário definir um critério, medido na forma de um escore, pelo qual alinhamentos são quantificados. O algoritmo ótimo para uma determinada definição de escore deverá procurar o alinhamento que maximiza o valor desse escore. Defi-

nições de escores diferentes darão origem a algoritmos variados. Por exemplo, um dos critérios mais comuns de definição de escore é aquele que atribui o valor +1 a uma correspondência (*match*) de caracteres, -1 a uma falta de correspondência (*mismatch*) e -2 quando há a ocorrência de uma lacuna. Dessa forma, a sequência do exemplo anterior tem o escore de -7, como ilustrado a seguir:

```

S L N S G - - - Y H F C
|           | | |
S F Q E T F L S F H F C
+1 -1 -1 -1 -1 -2 -2 -2 -1 +1 +1 +1 = -7
    
```

A busca de algoritmos que forneçam o alinhamento ótimo para um determinado critério de escore tem sido um amplo tema de pesquisa. Convém salientar que buscas exaustivas são proibitivas em termos de esforço computacional, mesmo para sequências moderadamente longas.

É necessário compreender que a problemática da medição da similaridade entre sequências não se reduz à medição da correspondência entre os resíduos dessas sequências, porque há necessidade de se considerar o significado biológico. Isto significa que existem casos em que os resíduos devem ser colocados em posições não idênticas em ambas as sequências. Este fato torna o problema do alinhamento muito mais complexo, pois há a necessidade de distinguir alinhamentos com alto escore e de significância puramente matemática daqueles com escore moderado ou baixo, porém de alto significado biológico.

Na tentativa de compatibilizar escore e significado biológico, surgiu a ideia de ponderar a correspondência

entre resíduos diferentes, levando em consideração as taxas de substituição de nucleotídeos ou aminoácidos observadas ao longo de grandes distâncias evolutivas. Esta ponderação pode ser descrita pela matriz de escores (**Figura 17.1**), tecnicamente denominada de matriz de substituição.

A título de ilustração, o mesmo exemplo de alinhamento anteriormente descrito possui um escore de 24.

```

S L N S G - - - Y H F C
|           | | |
S F Q E T F L S F H F C
2 2 1 0 0 -2 -2 -2 -2 6 9 12 = 24
    
```

Dentre as matrizes mais conhecidas, destacam-se a PAM (Point Accepted Mutation) e a BLOSUM (BLOCKS of Amino Acid SUBstitution Matrix). A BLOSUM é uma matriz de substituição utilizada para o alinhamento de sequências de proteínas. Essas matrizes são usadas para alinhamentos entre sequências de proteínas evolutivamente divergentes. Elas se baseiam em alinhamentos locais e levam em conta BLOCOS correspondentes a regiões muito conservadas de famílias de proteínas (que não têm falhas no alinhamento da sequência). Em seguida são contadas a frequência relativa dos aminoácidos e as probabilidades de substituição para cada uma das 210 possíveis substituições dos 20 aminoácidos.

A matriz PAM foi introduzida por Margaret Dayhoff, em 1978, com base em 1.572 mutações observadas em 71 famílias de proteínas relacionadas. Cada matriz tem os vinte aminoácidos em suas vinte linhas e colunas. O

C	12																			
S	0	2																		
T	-2	-1	3																	
P	-3	1	0	6																
A	-2	1	1	1	2															
G	-3	1	0	-1	1	5														
N	-4	1	0	-1	0	0	2													
D	-5	0	0	-1	0	1	2	4												
E	-5	0	0	-1	0	0	1	3	4											
Q	-5	-1	-1	0	0	-1	1	2	2	4										
H	-3	-1	-1	0	-1	-2	2	1	1	3	6									
R	-4	0	-1	0	-2	-3	0	-1	-1	1	2	6								
K	-5	0	0	-1	-1	-2	1	0	0	1	0	3	5							
M	-5	-2	-1	-2	-1	-3	-2	-3	-2	-1	-2	0	0	6						
I	-2	-1	0	-2	-1	-3	-2	-2	-2	-2	-2	-2	-2	2	5					
L	-6	-3	-2	-3	-2	-4	-3	-4	-3	-2	-2	-3	-3	4	2	6				
V	-2	-1	0	-1	0	-1	-2	-2	-2	-2	-2	-2	-2	2	4	2	4			
F	-4	-3	-3	-5	-4	-5	-4	-6	-5	-5	-2	-4	-5	0	1	2	-1	9		
Y	0	-3	-3	-5	-3	-5	-2	-4	-4	-4	0	-4	-4	-2	-1	-1	-2	7	10	
W	-8	-2	-5	-6	-6	-7	-4	-7	-7	-5	-3	2	-3	-4	-5	-2	-6	0	0	17
C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W	

Figura 17.1

Exemplo de matriz de substituição.

valor em uma determinada posição da matriz representa a probabilidade de uma substituição de um aminoácido por outro.

Os métodos de alinhamento de seqüências refletem a visão de dois modelos excludentes. No primeiro, a similaridade é considerada ao longo de toda a extensão das seqüências (**similaridade/alinhamento global**), ao passo que no segundo, as regiões de similaridade constituem uma fração da extensão das seqüências (**similaridade/alinhamento local**). É importante entender essa distinção, porque as seqüências não são uniformemente similares e é inútil realizar um alinhamento global em seqüências que só possuem similaridades locais.

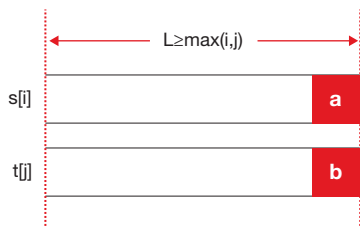
A lógica dos métodos de alinhamento local reside no fato de que sítios funcionais localizam-se em regiões consideradas curtas das seqüências. Assim, a busca por similaridades locais pode produzir resultados mais significativos sob o ponto de vista biológico do que aqueles obtidos por alinhamento ao longo de toda a seqüência.

A maioria dos métodos de alinhamento de seqüências, local ou global, se baseia em técnicas de **programação dinâmica**. Essa área, que pertence a uma mais ampla, denominada **programação matemática**, tem por objetivo construir a solução de problemas de otimização a partir de soluções obtidas para subproblemas semelhantes. O exemplo apresentado a seguir pretende ilustrar esses conceitos de uma forma mais clara.

Suponha que se deseje conhecer o alinhamento global ótimo entre as seqüências $s=AAAC$ e $t=AGC$ de extensões $m=4$ e $n=3$, respectivamente. Por ótimo, entende-se o critério de escore ilustrado no início desta seção, ou seja, o valor +1 a uma correspondência de caracteres, -1 a uma falta de correspondência e -2 quando há a ocorrência de lacuna.

Define-se $s[i]$ e $t[j]$ como sendo os prefixos (subseqüências) de s e t . Para os casos em que i ou j são nulos, tais subseqüências são vazias. A título de ilustração verifique que $s[3]=AAA$ e $t[2]=AG$.

Sendo $A[i,j]$ o alinhamento ótimo entre $s[i]$ e $t[j]$, observe o último caracter de cada uma dessas seqüências:



Existem três possibilidades mutuamente excludentes e exaustivas:

1. Somente a é uma lacuna. Neste caso, o escore de $A[i,j]$ vale o escore de $A[i-1,j]$ menos 2.
2. Somente b é uma lacuna. Neste caso, o escore de $A[i,j]$ vale o escore de $A[i,j-1]$ menos 2.

3. Nem a nem b são lacunas. Neste caso, o escore de $A[i,j]$ vale o escore de $A[i-1,j-1]$ mais 1 (se $a=b$) ou menos 1 (se $a \neq b$).

Conclui-se então que, para a determinação do escore de $A[i,j]$, é necessário o conhecimento apenas dos escores associados a $A[i-1,j]$, $A[i,j-1]$ e $A[i-1,j-1]$. O escore de $A[i,j]$ é determinado pelo maior valor obtido pelos cálculos indicados em (1), (2) e (3). Como os escores associados a $A[0,j]$ ou $A[i,0]$ valem respectivamente $-2 \cdot j$ e $-2 \cdot i$, todos os outros escores podem ser deduzidos destes.

Uma maneira prática de visualizar e resolver este problema de alinhamento consiste em criar uma matriz de $(m+1)$ linhas e $(n+1)$ colunas, cujo elemento (i,j) representa o alinhamento entre as subseqüências $s[i]$ e $t[j]$. Embora o objeto de interesse seja o do escore de $A[m,n]$, este valor só pode ser determinado após a propagação dos valores dos escores inicialmente conhecidos (**Figura 17.2**).

Este exemplo ilustra, para um caso particular, o procedimento básico associado à programação dinâmica, ferramenta utilizada em quase todos os métodos de alinhamento. Convém salientar que, alterando o conceito de escore, o método de alinhamento se altera para ser mais eficaz na determinação do valor máximo deste escore. Embora não discutido aqui, o intuito real não é o valor do escore, mas sim o alinhamento ótimo associado a esse escore, pois assim é permitido extrair das seqüências as possíveis propriedades biológicas de interesse.

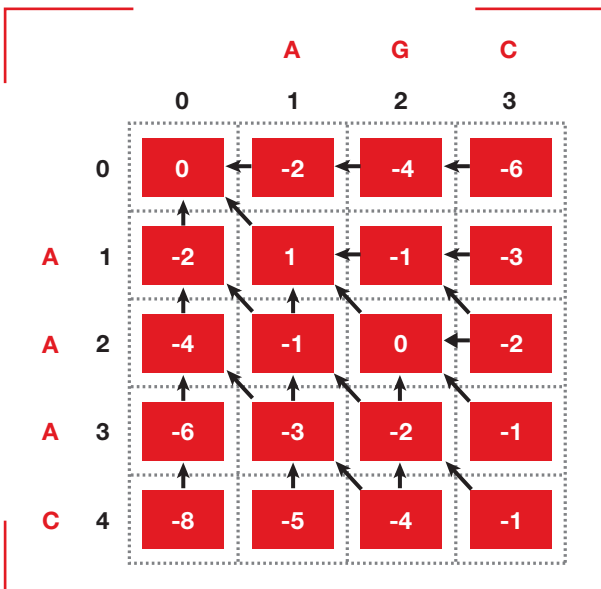


Figura 17.2

Procedimento de cálculo do escore associado ao alinhamento ótimo. O retângulo localizado na linha i e na coluna j possui o escore associado ao alinhamento ótimo $A[i,j]$ entre as subseqüências $s[i]$ e $t[j]$. As setas indicam a origem dos valores ótimos.

2.1.1 Dot Plot

O Dot Plot é um método gráfico que permite a comparação de duas sequências biológicas e a identificação de regiões de semelhança entre elas. Esse método compara duas sequências, organizando uma no eixo-x e outra sobre o eixo-y. Quando os resíduos de ambas as sequências se encontram na mesma posição, um ponto é desenhado na posição correspondente. Este tipo de gráfico considera a direção da sequência e uma vez que os pontos foram plotados, eles se combinam para formar linhas. Mudanças como inserções, deleções e mutações podem ser facilmente visualizadas (ver exemplo no Capítulo 4).

2.1.2 FastA

O algoritmo **FastA**, descrito por Lipman e Pearson, em 1985, é um método de busca de similaridades locais por meio do emprego de matriz de substituição. Concentra-se na busca de correspondências de subsequências curtas de tamanho k (k -uplas), que podem contribuir para a correspondência total usando implementações computacionalmente eficientes.

Visando à eficiência de processamento, este programa utiliza o padrão observado de correspondência de resíduos. Em vez de analisá-los individualmente, o programa procura segmentos com alta incidência destas correspondências. Por meio de um procedimento **heurístico** atribui-se um escore a cada um destes segmentos de alta incidência, que serão compostos ao final da busca para produzir um escore global.

Somente um alinhamento é fornecido para cada uma das sequências do banco de dados utilizado, possuindo um alto escore de alinhamento com a sequência de teste. Dessa forma, como outros alinhamentos importantes podem não ser observados por este enfoque, essas correspondências devem ser reanalisadas por outros programas, tais como o programa LALIGN.

Além dos escores citados, o programa **FastA** apresenta ainda uma estimativa da significância estatística de cada alinhamento encontrado. Essa significância é expressa por:

1. Probabilidade p de que o escore associado ao alinhamento tenha sido obtido por acaso. Quanto mais próximo de 0, maior é a probabilidade de que a correspondência seja real. Quanto mais próximo de 1, maior é a possibilidade de que esta correspondência seja espúria.
2. Valor médio E do número de correspondências de k -uplas que se espera observar por acaso, quando se investiga um banco de dados de um tamanho determinado.

2.1.3 BLAST

A família de programas **BLAST** (de *Basic Local Alignment Search Tool*), descritos por Altschul, em 1990, tornou-se popular devido a sua implementação eficiente em **computação paralela**, que permitiu a busca de sequên-

cias por meio de servidores públicos. O algoritmo BLAST é bastante simples e se baseia no conceito de par de segmentos. Dadas duas sequências, um **par de segmentos** é definido como um par de subsequências de mesmo comprimento, que formam um alinhamento sem lacunas. Este algoritmo determina todos os pares de segmentos, entre as sequências do banco de dados e a sequência em exame, que apresentam escore acima de um valor de limiar pré-fixado pelo usuário. Esses pares de segmentos, denotados de HSP (de *highscoring pairs* = pares com altos escores), são apresentados como saída do programa, assim como seus valores de probabilidade p (Seção 2.1.2).

É importante salientar que, embora tais HSPs possam ser obtidos por meio de técnicas de programação dinâmica, o algoritmo BLAST não as utiliza, optando pela procura de **sementes** que são pares de sequências muito curtas entre as sequências em estudo. Essas sementes são então estendidas para esquerda e direita, sem o uso de lacunas, até que extensões com escores máximos sejam alcançadas. Nem todas as extensões são investigadas, porque o programa compara os seus escores com um limiar cuidadosamente escolhido. Em consequência, existe uma chance pequena de que extensões adequadas não sejam selecionadas, mas, na prática, esta limitação é bastante aceitável.

Similaridades entre sequências são os mais frequentes e fundamentais tipos de análises realizadas no banco de dados GenBank. O programa BLAST apresenta as seguintes opções, chamadas de família BLAST:

BLASTp: Pesquisa de sequências de aminoácidos *versus* banco de dados de proteína.

BLASTn: Pesquisa de sequências de nucleotídeos *versus* banco de dados de nucleotídeo.

BLASTx: Pesquisa de sequências de nucleotídeos em todas as seis fases de leitura *versus* banco de dados de proteína.

tBLASTn: Pesquisa de sequência de aminoácidos *versus* banco de dados de nucleotídeos traduzido dinamicamente em todas as seis fases de leitura.

tBLASTx: Pesquisa de sequências de nucleotídeos traduzidos em todas as seis fases de leitura de proteínas *versus* banco de dados de nucleotídeos.

As consultas utilizando o programa BLAST (**Figura 17.3**) podem ser realizadas na página *Web* do NCBI ou localmente no computador do usuário, pois o programa BLAST é gratuito. Seu acesso está disponível via FTP.*

Algumas melhorias têm sido feitas no pacote de programas da família BLAST, por exemplo, o MEGA-BLAST e o BLAST genomas. Nesses casos o algoritmo para pesquisa de alinhamento de sequências de nucleotídeos é utilizado, concatenando as várias sequências-alvo (*query*). Dessa forma espera-se diminuir a quantidade de tempo gasto para percorrer o banco de dados. O que acontece

* Conteúdo do diretório disponível em: <ftp://ftp.ncbi.nlm.nih.gov/blast/>.

Sequences producing significant alignments:	Score (Bits)	E Value
gb AAM82168.2 VirB6 [Bartonella quintana]	589	2e-166
ref YP_032624.1 virB6 protein [Bartonella quintana str. Toul...	588	3e-166
ref YP_034055.1 virB protein [Bartonella henselae str. Houst...	541	3e-152
ref YP_002972361.1 VirB6 protein [Bartonella grahamii as4aup...	471	4e-131
ref YP_001610002.1 VirB6 protein [Bartonella tribocorum CIP ...]	450	9e-125
ref YP_001610346.1 VblB6 protein [Bartonella tribocorum CIP ...]	165	5e-39
ref YP_002972607.1 VblB6 protein [Bartonella grahamii as4aup...	157	1e-36

```
>|ref|YP_001610346.1| G VblB6 protein [Bartonella tribocorum CIP 105476]
emb|CARO2351.1| G VblB6 protein [Bartonella tribocorum CIP 105476]
Length=311

GENE ID: 5829874 vblB6 | VblB6 protein [Bartonella tribocorum CIP 105476]
(ID or fewer PubMed links)

Score = 165 bits (418), Expect = 5e-39, Method: Compositional matrix adjust.
Identities = 96/249 (38%), Positives = 149/249 (59%), Gaps = 10/249 (4%)

Query 1  MNTVSGLSAISAPLMLASLIIFLYGYNVMTGRVLSMHSLLNNVVKIVVVTAMATMA 60
      M+ T++ LSS++ APL L+ ++I GYN++ GR S+ + + K+ +V ++AT A
Sbjct 24  MDKTINHLSSSLGAPLKLSCITLYIFMGVNIIVGRSSMPLMEFIVITFKLGIIVVSLATKA 83

Query 61  DTFNTYVKDIFFGDLAMAIGMALNSWPASANVFDYLLKTSA-----RYQEVLAAMWF 113
      +N +V IFF DL HAI H          NV+D ++ R A          +Y + F
Sbjct 84  ALYMEWVSGIFPHDLPHALANVTQGAHVDRVVDVDMHIGKAGAVLDAAMKVTGLTQHGHP 143

Query 114  LEKIMVGLLGSMLMIMAVIVFCIGGFIVOMFAOVALVMIIGLGPLFISLYLFNATREKPTDA 173
      + + G + L+I A FC+ GF+V +FA+ L +++ +GPLFISLY+P++TRKPT+A
Sbjct 144  IATWLAGFI-CLVISAF--FCLVGFVVSFLAKLGLFLVLSIGPLFISLYHFSSTRKPTFA 200

Query 174  WITTLIRFTLLQVLVHNLGCTIMCKTIIHVLNGTVDYVFLFPPVWVVISIVGAILFRALPG 233
      W+ + HF IL VLV+LG + KI + V + D+I+          ++ I G LF LP
Sbjct 201  WLGQVABFILLVYLVLLGGLYVKISHTVFEQVDMIFPTLLIQFLVVGIGGIFLFLRLPD 260

Query 234  IASALSSGG 242
      IASAL+SGG
Sbjct 261  IASALASGG 269
```

Figura 17.3

Alinhamento Blast. A sequência de entrada *virB6* é chamada de **objeto (query)** e, comparada com todas as sequências depositadas no banco de dados escolhido, chamadas de **sequências-alvo (subjects)**. Cada alinhamento retornado pelo BLAST vem acompanhado do identificador, da descrição da sequência, do escore, que representa a pontuação do alinhamento, e do valor esperado (**E-value**), que representa a probabilidade desse alinhamento ter ocorrido ao acaso. Quanto menor o valor E, melhor o resultado. Nesta figura, estão destacados os resultados para o gene *VblB6*. Nucleotídeos ou aminoácidos não idênticos, mas com propriedades físico-químicas preservadas, são considerados como positivos. Inserções de espaço para obtenção de melhor alinhamento são denominadas **lacunas (gaps)**.

é uma otimização do programa que alinha as sequências que têm pouca divergência.

O pacote BLAST contém, ainda, módulos que utilizam informações de alinhamentos múltiplos para melhorar a especificidade das pesquisas do banco de dados. O **PSI-BLAST** (de *Position Specific Iterative BLAST* = Blast interativo posição-específica) é um aprimoramento do programa BLAST original que implementa perfis com o objetivo de aumentar a especificidade das pesquisas no banco de dados. O processo tem início com uma única sequência. O PSI-BLAST pesquisa o banco de dados em busca de alinhamentos locais, utilizando o BLAST com intervalos. Em seguida, constrói um alinhamento múltiplo e um perfil com o comprimento da sequência da consulta

original. Esse perfil é usado para pesquisar novamente o banco de dados de proteínas em busca de alinhamentos locais. Esse procedimento pode ser repetido o número de vezes desejado, entretanto, erros no alinhamento podem ser aumentados pela repetição.

PHI-BLAST (de *Pattern Hit Initiated-BLAST* = BLAST iniciado por ocorrência padrão) utiliza uma sequência e um padrão pré-selecionado, encontrado nessa sequência, como entrada para consultar um banco de dados de sequências de proteínas. O padrão deve ser escrito na sintaxe do **PROSITE** (ver Seção 4.2.1), que é descrita em detalhes no sítio do servidor PHI-BLAST.*

* Disponível em: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.

2.2 Alinhamentos múltiplos

Embora a comparação de pares de sequências seja fundamental para o estudo de homologies, a análise de grupo de sequências que formam uma família de genes requer a capacidade de estabelecer conexões entre mais de dois membros de um grupo com o objetivo de avaliar as características deste grupo. O processo de alinhamento múltiplo pode ser entendido como a melhoria da relação sinal-ruído em um conjunto de sequências que, em última instância, facilita a elucidação de temas biologicamente significantes.

Sob o ponto de vista operacional, um alinhamento múltiplo é uma generalização de um caso de alinhamento de um par de sequências (Seção 2.1), e seu objetivo é transformar, por meio de lacunas, todas as sequências em outras de mesma extensão. Essas novas sequências podem ser vistas na forma de uma tabela, em que cada linha está associada a uma sequência, como ilustrado abaixo:

YDGGAVEAL	YDGGAV-EAL
YDGGEAL	YDGG---EAL
FEGGILVAL	FEGGILEVAL
FDGILVQAV	FD-G I LVQAV
YEGGAVVQAL	YEGGAVVQAL

O problema fundamental do alinhamento múltiplo é a definição precisa da qualidade do alinhamento. Existe uma grande quantidade de maneiras de definir estes **escores de qualidade do alinhamento múltiplo**, mas dois deles se destacam:

1. um no qual se define o escore do alinhamento múltiplo como a soma dos escores do alinhamento de todos os pares não ordenados dentre as sequências;
2. e outro no qual se define uma sequência, denominada consenso, e o escore do alinhamento múltiplo é definido como a soma dos escores entre as sequências constituintes e a consenso.

Em qualquer um destes casos, o alinhamento ótimo é obtido pela maximização do escore. É interessante notar que a complexidade computacional do alinhamento múltiplo cresce de forma exponencial com o número de sequências envolvidas, tornando assim imperativo que algoritmos para este tipo de alinhamento tenham de ser bastante eficientes.

O programa **ClustalW** é um dos mais utilizados para o alinhamento múltiplo de sequências e encontra-se disponível gratuitamente para todas as plataformas computacionais mais comuns. Esse programa se baseia no conceito de alinhamento progressivo, que determina os alinhamentos para cada par de sequências e constrói uma matriz de distâncias que reflete esses alinhamentos. Essa matriz, por sua vez, é empregada na determinação de uma árvore filogenética-guia, utilizada como peça chave para a determinação do alinhamento múltiplo.

3. Análise de genomas

Assim como o sequenciamento de DNA pelo método de Sanger introduziu uma mudança de paradigma, que permitiu avanços na biologia molecular, o **sequenciamento de DNA**, utilizando as novas tecnologias, está prestes a mudar radicalmente a natureza da pesquisa biológica. Essas novas tecnologias possibilitaram novas questões e prometem mudanças nas áreas da genômica comparativa, da biologia de sistemas, da metagenômica, entre outras. Por outro lado, algumas áreas da bioinformática como montagem, anotação e comparação de genomas, que pareciam estar resolvidas, do ponto de vista computacional, tornaram-se novamente um problema. Isto deve-se ao fato de que estas novas técnicas de sequenciamento fornecem uma grande quantidade de dados da sequência de DNA, que são coletadas em um ritmo surpreendente, com redução de custos, esforço e tempo, quando comparado ao sequenciamento pelo método de Sanger (ver Capítulo 16).

Um dos maiores desafios da bioinformática moderna é a análise de genomas completos, ou seja, a identificação de genes preditos computacionalmente, a associação de uma função e o fornecimento de subsídios para o delineamento de experimentos que possam testar essas predições e compará-las com outras já existentes.

3.1 Montagem de genomas

O sequenciamento completo de genomas foi possível após avanços tecnológicos que incluem a metodologia conhecida como *shotgun* (Capítulo 16). A técnica de Sanger permitia obter sequências de pequenos segmentos de DNA, que podiam variar de tamanho conforme o vetor no qual eram clonados (ver Capítulo 16). Nos sequenciadores da nova geração a fase da clonagem foi suprimida. Entretanto, nos dois casos, o sequenciamento é feito de maneira aleatória e, em seguida, programas computacionais são utilizados para fazer a sobreposição de sequências (*reads*), em um processo chamado de montagem (*assembly*). O objetivo desse processo é a obtenção de uma sequência consenso de bases contíguas (*contigs*). Os *contigs*, por sua vez, podem ser agrupados formando os *scaffolds* ou *super-contigs*.

Existem diversos programas que realizam o processo de montagem: Phred/Phrap/Consed, CAP/PCAP, Celera Assembler, *Genome Analyzer* (Illumina) e o *GS De Novo Assembler* (Roche). Cada programa utiliza algoritmos diferentes para obter a sequência contígua de DNA. De maneira geral, os programas seguem o modelo que utiliza um algoritmo de programação dinâmica para fazer o alinhamento das sequências comuns, como o algoritmo de Smith-Waterman, que compara os segmentos de todos os tamanhos possíveis e otimiza a medida de similaridade (ver Seção 2.1). Para garantir um resultado de montagem confiável é necessário que cada posição de base do cromossomo seja representada várias vezes, fazendo com que o número de bases sequenciadas atinja entre dez a vinte vezes o tamanho do genoma original. Muitos pro-

blemas podem ocorrer durante a montagem, tais como agrupamento de regiões repetidas, regiões que tiveram baixa qualidade durante o sequenciamento, compressão de base e regiões com baixa cobertura, já que o sequenciamento é aleatório. Cada um desses problemas deve ser tratado de forma individual, por exemplo, com o desenho de iniciadores específicos, ressequenciamento de *reads* de baixa qualidade, entre outras.

3.2 Anotação de genomas

Depois de sequenciar um genoma, o próximo passo é a identificação de todos os genes que codificam proteínas (**proteoma**) e a função das proteínas codificadas, pela busca por similaridades em banco de dados (ver Seção 4). Esta etapa na análise genômica, em que a informação biológica é agregada, chama-se **anotação**. A identificação de genes em procariotos é mais simples, devido à ausência de íntrons (ver Capítulo 4). Programas computacionais identificam códons de iniciação de proteínas (a trinca de nucleotídeos ATG) e códons de terminação na mesma fase de leitura (ver Capítulo 12). Em eucariotos, devido à presença frequente de íntrons (ver Capítulo 5), essa identificação é mais complexa, pois devem ser consideradas as junções íntron-éxon. Nos dois casos, os métodos de predição de genes envolvem o uso de modelos matemáticos, como os modelos de Markov ocultos (HMMs, de Hidden Markov Models), redes neurais, entre outras, e o uso de genes do próprio organismo ou de organismos filogeneticamente próximos, para treiná-los. Como os métodos preditivos são parcialmente precisos, a identificação de genes é facilitada quando existem sequências de ESTs, pois a presença dessas sequências confirma que o gene está sendo transcrito (ver Capítulo 16).

A sequência predita de aminoácidos das proteínas codificadas pelos genes identificados é então usada como sequência-alvo, para buscas de similaridades em **banco de dados secundários** (ver Seção 4.2). Quando existe uma correspondência da sequência de proteína, com uma ou mais sequências em um ou mais desses bancos de dados, essa informação serve para identificar a função do gene e para validar a sua predição. Pseudogenes e genes que perderam a função, também podem ser identificados nessa análise.

Outro ponto importante é a identificação de genes **ortólogos**, ou seja, genes similares em dois ou mais organismos que devem ter a mesma função e história evolutiva (ver Capítulo 4). Nessa etapa, procura-se montar um catálogo dos genes presentes no organismo estudado, atribuindo-lhes nomes e associando-os às prováveis funções. Utilizando esta abordagem, em média, pode-se associar uma função a 73% das ORFs encontradas em genomas bacterianos. Geralmente, para sequências com identidade maior que 50%, esses métodos são eficientes e podem ser usados para predizer relações evolutivas consideradas óbvias entre as sequências. Entretanto, essa técnica apesar de bastante eficaz pode levar a uma sequência de erros, podendo comprometer todo o sistema de anotação automática, pois é possível a propagação

desses erros. Identificados os genes e as suas respectivas funções, deve-se tentar relacioná-los, de modo a obter um mapa funcional do organismo estudado, identificando as rotas bioquímicas que estão completas ou incompletas no organismo e as rotas alternativas. Com isso, abordagens paralelas envolvendo buscas em bancos de dados funcionais e de metabolismo devem ser realizadas, garantindo assim uma melhor anotação funcional de genomas.

Embora a **metabolômica** (ramo da genômica funcional que estuda o metabolismo) seja bastante promissora, ainda existem vários desafios que influenciam a sua implementação, visto que não é possível analisar toda a gama de metabólitos por meio de um único método analítico. O sucesso dessa análise reside na interpretação da importância biológica das medições dos produtos químicos identificados nas amostras. Os dados, geralmente, consistem de medições realizadas em indivíduos sob várias condições, e essas medidas podem ser digitalizadas em espectros ou em uma lista de metabólitos. Na sua forma mais simples, isso gera uma matriz com linhas correspondentes aos indivíduos e colunas correspondentes às concentrações de metabólitos. Os dados são analisados por métodos estatísticos, como a análise de componentes principais (PCA, de *principal component analysis*) e o método dos mínimos quadrados parciais (PLS, de *partial least squares*).

Para proteínas com identidade menor que 50% outras análises necessitam ser realizadas, como por exemplo, buscas em **bancos de dados de estruturas** (ver Seção 4.3). Em projetos de **sequenciamento de transcritomas**, a tarefa consiste em identificar os genes expressos (transcritos) no organismo, em uma determinada condição, além de identificar os genes que apresentam variantes de *splicing* (ver Capítulo 11).

A análise automática de genomas usando a bioinformática é considerada uma ferramenta eficiente, prática e rápida. Assim, o desenvolvimento de métodos confiáveis para anotação funcional de genomas é de grande importância. Vários sistemas foram desenvolvidos e estão disponíveis, por exemplo, nos sites *Web* do *International Nucleotide Sequence Database Collaboration* (INSD) e em outras instituições que desenvolvem sua própria metodologia para atender suas necessidades específicas. No Brasil foi desenvolvido o sistema automático SABIA (**ver Material Complementar Online 17.1**), utilizado para montar e anotar vários genomas de procariotos e de eucariotos, ESTs e metagenomas.

3.3 Análise comparativa de genomas

Quando se dispõe do genoma completo de dois organismos semelhantes, uma pergunta pertinente é: como estes genomas se alinham? A comparação de pares de genomas ou de múltiplos genomas é útil para muitos estudos, desde a elucidação de questões básicas de biologia evolutiva até questões clínicas muito específicas, como a identificação de polimorfismos genéticos. Serão apresentados alguns programas de computador que são utilizados para essa finalidade.

A chamada **genômica comparativa** inclui a comparação de todas as proteínas preditas, ou seja, o proteoma de dois ou mais organismos, a localização dos genes e o número e a localização de sequências repetitivas. Outra área da genômica comparada é o grau de diversidade na população de uma espécie (SNPs, de *single nucleotide polymorphism* = polimorfismo de base única), variações de micro satélites, níveis de expressão gênica e a associação dessas variações nas sequências em resposta ao meio ambiente e a doenças.

Embora exista uma vasta literatura de algoritmos eficientes para o alinhamento de pares de sequências, quando as dimensões das sequências a serem comparadas atingem valores da ordem de milhões de pares de bases, praticamente todos estes métodos falham em atingir um resultado. As razões deste fracasso são de ordem combinatorial: os programas de computador a eles associados esgotam a memória ou apresentam um tempo de execução proibitivo. Talvez, mais importante ainda, seja o fato de que os programas de alinhamento existentes foram projetados para descobrir inserções, deleções e mutações pontuais nas sequências, mas não foram concebidos para procurar mudanças em grande escala, como diferenças em repetições em tandem e inversões em grande escala.

A partir de 1999, soluções para estes problemas começaram a aparecer na literatura, envolvendo técnicas bastante sofisticadas. Uma das técnicas que se mostrou mais eficiente utiliza uma estrutura de dados denominada **árvore de sufixo**. Essa estrutura, aliada ao alinhamento de Smith-Waterman e outros (ver Seção 2.1), permitiu que os alinhamentos de vários milhões de nucleotídeos pudessem ser efetuados.

A seguir, são descritos alguns programas comumente utilizados em genômica comparativa.

3.3.1 MUMmer

O **MUMmer** é um sistema para alinhar genomas completos ou em andamento, gerando visualizações coloridas de regiões similares. Inicialmente desenvolvido por um grupo do *TIGR (The Institute for Genomic Research)*, hoje é um sistema de livre acesso. O MUMmer baseia-se no algoritmo computacional chamado de árvore de sufixo, o que facilita o rápido tratamento pelo sistema de um grande número de comparações de pares de sequências. Esse sistema fornece informações de uma única mudança de nucleotídeo, inversões, translocações e homologies entre os genes. Se as espécies são muito divergentes para um alinhamento de sequência de DNA, o programa PROmer pode gerar alinhamentos com base nas traduções dos seis quadros de leitura de ambas as sequências de entrada.

3.3.2 ACT

A ACT (de *Artemis Comparison Tool* = ferramenta de comparação do Artemis) foi desenvolvida pelo *Sanger Institute* da Inglaterra, é um sistema que permite uma visualização interativa de comparações entre sequências completas de genoma e anotações associadas.

A **WebACT** é uma base de dados de comparações entre todas as sequências de procariotos disponíveis publicamente, permitindo a visualização online de comparações entre até cinco sequências genômicas. A comparação pode ser gerada com diversos programas como BLASTN, TBLASTX (Seção 2.1.3) ou MUMmer (Seção 3.3.1), sendo possível identificar regiões de similaridade, inserções e rearranjos em qualquer nível, desde grandes regiões dos genomas até diferenças de pares de bases.

3.3.3 PipMaker

O **PipMaker**, desenvolvido por um grupo da *Pennsylvania State University*, é um programa que compara duas sequências de DNA e produz um gráfico de identidade percentual como resultado. Esse programa é útil na identificação de padrões de similaridade em larga escala. Usar o PipMaker é relativamente simples, o processo inicia-se com dois arquivos no formato FASTA, gerando um conjunto de instruções para mascarar as sequências repetitivas (repeatMaker). Esse procedimento reduz o número de ocorrências não informativas na comparação de sequências. Após a realização do procedimento, um arquivo com a posição numérica dos genes conhecidos é submetido ao servidor *Web* do PipMaker e os resultados são enviados por e-mail.

3.3.4 M-GCAT

O **M-GCAT**, desenvolvido por um grupo de pesquisadores da *Technical University of Catalonia*, Barcelona, é um programa que permite a rápida visualização e o alinhamento das regiões mais conservadas de vários genomas, (normalmente procariótico). O M-GCAT se baseia em uma abordagem bastante eficaz para comparação múltipla de genomas, pois usa um gráfico de sufixo comprimido, o que permite construir vários quadros de alinhamento de genomas em espécies estreitamente relacionadas. O M-GCAT exhibe sensibilidade e acurácia consistentes com o BLAST e velocidade de processamento dos dados similar ao MUMmer.

4. Bancos de dados biológicos

Com a disponibilidade dos navegadores e a popularidade da internet como força mundial de comunicação, foram criadas redes em biologia molecular nos EUA e na Europa, para atender à demanda em pesquisas nesta área. Essas redes tornaram-se importantes, principalmente a partir da década 1980, quando os bancos de dados começaram a proliferar e tiveram de ser alocados e administrados por instituições de vários países. Dessa forma, surgiram vários tipos de bancos de dados, que foram agrupados de acordo com as informações que armazenam.

Os **bancos de dados secundários** são aqueles que derivam dos primários, ou seja, foram formados usando as informações depositadas nos **bancos primários**, bancos que armazenam sequências de nucleotídeos ou aminoácidos. As sequências homólogas podem

ser agrupadas por meio de um alinhamento múltiplo, e nelas destacadas as regiões conservadas e **motivos**, que geralmente estão envolvidos com a função biológica da proteína. Os motivos têm sido explorados de diferentes maneiras, para construir padrões de caracterização para famílias de proteínas parálogas e ortólogas (ver Capítulos 4 e 5). Assim, como existem vários bancos de dados de sequências primárias e uma grande variedade de formas de analisá-los, a informação armazenada nos **bancos de motivos** também seguiu esta tendência.

Existem, ainda, os **bancos de dados de estruturas** de proteínas, que contêm a predição de estrutura proteica resolvida pela cristalografia e modelagem computacional. Os **bancos de dados de genômica comparativa** e os **bancos de dados funcionais e de metabolismo** procuram estabelecer relações entre os genomas de organismos evolutivamente próximos, visando à identificação de suas características comuns por meio de análise de polimorfismos, filogenia e interações metabólicas. Dessa forma, a quantidade de informação armazenada **em bancos de dados biológicos** é muito grande.

Um dos pontos importantes nesse campo é o desenvolvimento de métodos computacionais, que possam auxiliar o processo de inferência da função biológica; estes métodos são denominados **métodos de análise de sequências**. Quando esses métodos envolvem bancos de dados, o desafio é identificar quais sequências estão inter-relacionadas (verdadeiros positivos) e quais não estão (verdadeiros negativos). Uma vez que, esta inter-relação é julgada com base em um escore e o estabelecimento de um limiar, é possível que sequências não inter-relacionadas sejam consideradas relacionadas (falsos positivos) ou que o oposto ocorra (falsos negativos). A qualidade dos métodos de avaliação de inter-relacionamentos é avaliada pelas probabilidades de ocorrências de falsos positivos e falsos negativos. Assim, os bancos de dados de diagnóstico de assinaturas para famílias de proteínas, domínios e sítios funcionais são ferramentas importantes para classificação computacional da função de sequências ainda sem caracterização bioquímica. A seguir, serão descritos os principais bancos de dados do ponto de vista biológico.

4.1 Banco de dados primários

O INSD, criado em 1988, é composto por três instituições: *National Center for Biotechnology Information* (NCBI), uma divisão da *National Library of Medicine* (NLM) do *National Institutes of Health* (NHI, em Bethesda, Maryland); o *European Molecular Biology Laboratory* (EMBL), que é um banco de dados do *European Bioinformatics Institute* (EBI, Hinxton, Inglaterra); e o *DNA Data Bank of Japan* (DDBJ, Mishima, Japão). As informações inseridas diariamente em cada uma das instituições é repassada para as demais. Assim, os três bancos compartilham todas as informações virtuais a qualquer momento. As três instituições, junto com o *Universal Protein Resource* (UniProt), que é um repositório central de sequências de proteínas curadas, são os maiores repositórios de sequências de nucleotídeos e de proteínas.

4.1.1 GenBank

Desde 1992, o NCBI tem também a função de manter o banco de dados **GenBank** que é um banco de dados de sequências primárias. Periodicamente, o NCBI News publica um gráfico (**Figura 17.4**), no qual é possível observar a evolução do banco e associar o seu crescimento aos eventos relevantes, não só do ponto de vista biológico como do tecnológico.

Em dezembro de 2010, o **GenBank** continha 122 bilhões de nucleotídeos. O crescimento do GenBank é exponencial e aproximadamente a cada 35 meses tem dobrado o seu tamanho. Essas sequências estão separadas em 20 divisões (**Tabela 17.1**).

Desde 2009 o NCBI mantém o *Short Read Archive** como um repositório de dados, no qual são disponibilizadas as sequências geradas pelas novas metodologias de sequenciamento. O SRA aceita dados das novas gerações de sequenciadores, incluindo o Genome Sequencer FLX System -454 (Roche), o Illumina Genome Analyzer (Illumina Inc.), o SOLiD (Applied Biosystems) e o HeliScope (Helicos BioSciences Corporation). Esses métodos podem gerar desde centenas de megabases até gigabases de dados em um único experimento e farão com que o tamanho desse banco aumente ainda mais.

Uma das facilidades produzidas pelo NCBI é denominada **Entrez**, uma ferramenta capaz de recuperar de forma fácil e eficiente a informação que está depositada nos 38 bancos de dados e divisões do **GenBank**, contendo mais de 400 milhões de registros (**Tabela 17.2**). A estrutura do **Entrez** se baseia no conceito de *hard link* e de vizinhança. **Hard links** são conexões entre entidades localizadas em bancos de dados diferentes, ao passo que vizinhança é uma coleção de conexões entre entidades de um mesmo banco de dados que guardam entre si uma relação de proximidade. Assim, buscas entre entidades relacionadas ficam muito eficientes e rápidas.

O GenBank e o DDBJ usam o GBFF (de *GenBank Flatfile*) como formato para a representação de sequências biológicas (o EMBL utiliza um formato parecido, mas com alguns tipos de prefixos como linhas de identificação) (**Figura 17.5**).

O GBFF é dividido em três partes, cada uma contendo informações específicas sobre as sequências. A primeira parte contém um cabeçalho com informações gerais, a segunda contém as descrições das características (*features*) e a terceira contém a sequência de DNA propriamente dita. O cabeçalho possui a seguinte composição:

1. Uma primeira linha contém:

- Nome do locus.

É a primeira identificação do organismo e pode ter até 10 caracteres, todos maiúsculos, sendo que o

* Disponível em: www.ncbi.nlm.nih.gov/Traces/sra/.

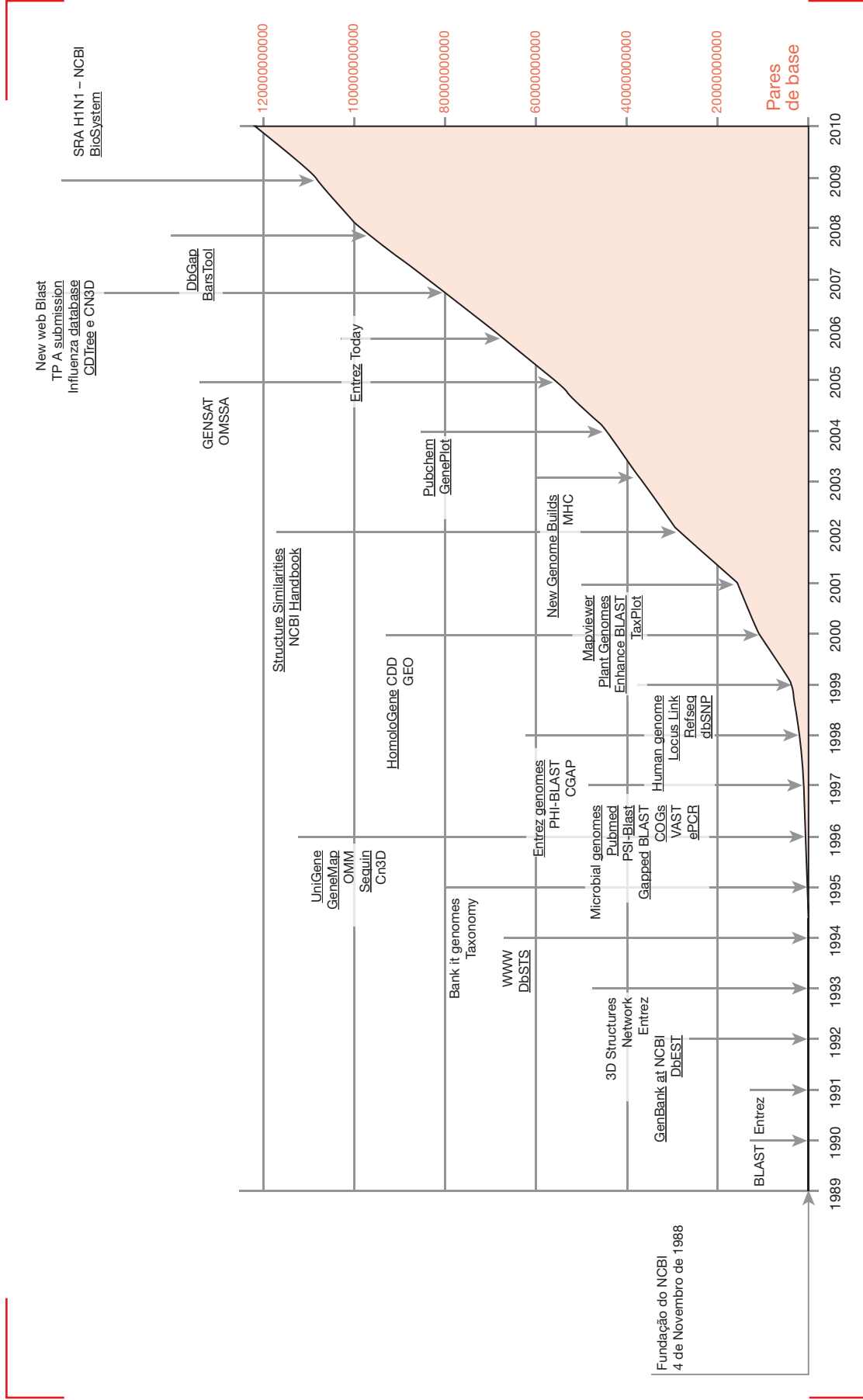


Figura 17.4

Curva de crescimento do Banco de Dados GenBank nos últimos anos.

Tabela 17.1 Divisões/organismos do banco de dados GenBank

Divisões	Organismos ou natureza das sequências
BCT	Bactérias
INV	Invertebrados
MAM	Mamíferos
PHG	Fagos
PLN	Plantas, fungos e algas
PRI	Primatas
ROD	Roedores
VRT	Outros vertebrados
VRL	Outras sequências virais
NEW	Novas sequências
SYN	Sequências sintéticas
UNA	Sequências não anotadas
EST	Etiquetas de sequências expressas
GSS	Sequências geradas de DNA genômico por técnicas variadas
HTG	Sequências de genomas incompletos geradas por sequenciamento de alto desempenho
PAT	Sequências patenteadas
STS	Sítios com etiquetas de sequências expressas e mapeadas
HTC	Sequências de cDNA geradas por sequenciamento de alto desempenho
ENV	Sequências de amostras do meio ambiente
CON	Sequências construídas

primeiro é obrigatoriamente uma letra e os demais podem ser quaisquer caracteres alfanuméricos. Em geral, as três primeiras letras estão relacionadas com o organismo e as outras com o nome do gene.

- Extensão da sequência em pares de bases.
 - Tipo de molécula (geralmente de DNA ou RNA).
 - Divisão no GenBank em 3 letras maiúsculas.
 - Data em que a sequência foi submetida.
2. A segunda linha apresenta uma definição resumida do organismo e do gene que está descrito.
 3. A terceira linha contém o número de acesso (*accession number*), que é o elemento mais importante de identificação da sequência. Esse número permanece invariável, mesmo que a sequência seja atualizada e, por essa razão, é frequentemente utilizado em artigos científicos.
 4. A quarta linha apresenta o NID (*gi – geninfo identifier*), o identificador único da sequência. A letra-prefixo identifica o banco de dados do qual a sequência é originária. Este número pode ser mudado quando a sequência é alterada.
 5. A quinta linha exibe a versão da sequência que indica a existência de mudança e a sua identificação.

Tabela 17.2 Bancos de dados e número de registros do GenBank

Banco de dados	Registros
Nucleotide	78 783 103
EST	62 838 170
PubChem Substance	61 056 228
SNP	59 806 469
GEO Profiles	42 751 725
Protein	28 475 324
GSS	25 787 403
PubChem Compound	25 668 433
PubMed	19 076 621
Probe	10 187 129
Gene	6 261 420
UniGene	3 645 645
PubMed Central	1 834 865
NLM Catalog	1 394 522
Taxonomy	525 252
UniSTS	524 629
Protein Clusters	413 052
3D Domains	280 897
Books	237 535
MeSH	211 794
Cancer Chromosomes	134 570
Homologene	123 767
PopSet	101 569
Biosystems	96 559
GENSAT	91 458
dbGaP	62 335
Structure	59 329
CDD	34 735
Journals	23 939
GEO	Datasets 21 358
OMIM	20 548
Site Search	25 070
Genome	10 777
SRA	6562
Projects ^a	5234
OMIA	2599
PubChem Bioassay	1691
Peptidome	79

^aAnteriormente conhecido como Genome Projects.

6. As sexta e sétima linhas mostram as palavras-chave e os organismos-fonte que descrevem detalhadamente a sequência, usando a taxonomia disponível no NCBI.
7. Um conjunto de linhas, de tamanho variável, apresenta as referências bibliográficas, citando diversas informações, dentre as quais, o local de descrição da sequência e sua identificação nos bancos de informações *MEDLINE* e *PubMed* (este último, somente a partir de 1998).

```

LOCUS      BACLEXA      1242 bp      DNA           BCT           05-MAY-1993
DEFINITION Bacillus subtilis lexA gene, 3' end.
ACCESSION  M64684
NID        g143151
VERSION    M64684.1  GI:143151
KEYWORDS   lexA gene.
SOURCE     Bacillus subtilis DNA.
  ORGANISM Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE  1 (bases 1 to 1242)
AUTHORS    Raymond-Denise,A. and Guillen,N.
TITLE      Identification of dinR, a DNA damage-inducible regulator gene of
            Bacillus subtilis
JOURNAL    J. Bacteriol. 173, 7084-7091 (1991)
MEDLINE    92041604
FEATURES   Location/Qualifiers
  source   1..1242
            /organism="Bacillus subtilis"
            /db_xref="taxon:1423"
  gene     390..1007
            /gene="lexA"
  CDS      390..1007
            /gene="lexA"
            /codon_start=1
            /transl_table=11
            /protein_id="AAA22573.1"
            /db_xref="PID:g289288"
            /db_xref="GI:289288"
            /translation="MTKLSKRQLDILRFIKAEVKSKGYPPSVREIGEAVGLASSSTVH
            GHLARLETKLIRRDPKPRRAIEILDEEVDIPQSQVNVVPVIGKVTAGSPITAVENIE
            EYFPLPDRMVPPDEHVFMLEIMGDSMIDAGILDKDYVIVKQONTANNGEIVVAMTEDD
            EATVKRFYKEDTHIRLQPENPTMEPIILQNVSILGKVIGVFRVH"
BASE COUNT      388 a      241 c      257 g      356 t
ORIGIN
  1 atcagaggtt tgaagtgtat tttatcagc taccattca ataaaatcat tttgtttat
  61 cttttttgta tcggctacct gatcagcaat tgaccagagt gtgtcgcctt gctggacttc
  121 tattttaaca tactgattaa gctcttgccc gctgcttgta tatgacagca taaggataaac
  181 cgcgctcaaa atcactgtaa acagaccgac aaaaataata gattctttac tcatgatcat
  241 aacctccaac aggaatgttt gttcgcattt ttataatttc agtatatacy aacaaacggtt
  301 tctgtcaatt gttttttcga acctatgttt gtactgtaag gaaaatcatg ctataatctt
  361 cttaaaatcg acgttttgag gtgcgaaaaa tgacgaagct atcaaaaagg caacttgata
  421 tcctccgttt tattaagca gaggttaaat caaaaggata tccgccttcc gtgagagaga
  481 tcggagaggc tgtcgggctt gcgtccagtt ctactgtcca cggccatttg gcccgtttg
  541 aaacaaaagg gctgatcaga cgagatccga caaaaccaag agcaatagaa attcttgatg
  601 aagaagtaga cattccgcaa agccaggctg ttaatgtacc ggtcatcggg aaagtacgg
  661 cgggatctcc tattacagca gtgaaaaata ttgaagaata tttccgctc cctgaccgca
  721 tggttcctcc agacgagcat gtatttatgc tggagattat gggagacagt atgatagatg
  781 ctggtattct cgataaggac tatgtcatcg taaacagca aaacacagcg aacaatgggg
  841 aaattgtagt tgcgatgaca gaagatgatg aagcaactgt aaaacgattc tataaagagg
  901 ataccacat tcgcttgacg cctgagaacc caacgatgga acctatcatt ttgcagaatg
  961 tcagcattct agggaaagtg attggcgtgt tcagaaccgt ccattaatct attgatctat
  1021 cagatgcata aaataacccc ctcttcctt acggagaggg ggtttgtgt tcaaaaaaag
  1081 caggccgcc ctgacttttg ggtgcgagcc cggaatagga ggaaaaagaa catcatgttt
  1141 acaccataat tatccaccaa aagattgttg aagtcaagaa attttcttc ctttaggaaa
  1201 cttttatcaa aatatgaaca tttgaatatt agacgcattc tt
  //

```

Figura 17.5

Exemplo da descrição de uma sequência no GenBank.

Na parte reservada às características (*features*), encontram-se as tabelas de descrição que contêm as informações mais importantes sob o ponto de vista biológico. Atualmente existem 13 tipos diferentes de características, seguindo uma nomenclatura determinada.*

* Disponível em: <http://www.ncbi.nlm.nih.gov/collab/FT/index.html>.

4.1.2 EMBL Nucleotide Sequence Database

A EMBL Nucleotide Sequence Database (EMBL Banco de dados de sequências de nucleotídeos) é o membro europeu do INSD. Esse banco permite o acesso direto a centenas de sequências de genomas concluídos, bem como as respectivas sequências deduzidas de aminoáci-

dos das regiões codificadoras de proteínas. Um sistema de anotação automática, a partir do projeto **Ensembl Genome Browser** (EMBL-EBI e *Wellcome Trust Sanger Institute*, 2000), fornece visualizações gráficas e acesso a vários outros bancos de dados. O objetivo desse sistema foi o de ampliar as abordagens utilizadas para o entendimento da biologia e da evolução humana para outros organismos.

O aumento na quantidade de dados depositados nos bancos de nucleotídeos, devido ao progresso das técnicas de sequenciamento, provocou uma mudança radical no **European Nucleotide Archive** (ENA). Ele foi acrescido do **Trace Archive**, para o armazenamento de sequências oriundas de instrumentos com capilares, e do **Sequence Read Archive** (SRA), para as sequências provenientes de tecnologias de última geração. Desse modo, se conseguiu a transformação do ENA em um arquivo completo de sequências de nucleotídeos.

Para o estudo de proteomas, o sistema possui um link direto para o *Universal Protein Resource* (UniProt).

4.1.3 DNA Data Bank of Japan

O **DNA Data Bank of Japan** (DDBJ – Banco de Dados de DNA do Japão), é o terceiro membro da INSD. É característico do DDBJ o fato de que mais de 90% dos dados são originados de pesquisadores sediados no Japão e do *Japan Patent Office* (JPO – Serviço de Patentes do Japão); o restante provém da China, da Coreia e de Taiwan.

Em razão da avalanche de dados causada pelos sequenciadores de última geração, um sistema duplo está sendo usado pelo DDBJ para a divulgação mais rápida dos seus bancos de dados aos interessados. Trata-se, primeiro, da liberação diária de novos dados em formato de *blogs* ou de manchetes à maneira de notícias; e segundo, de um sistema de visualização de palavras contidas em textos de documentos ou em páginas da rede, para ressaltar de forma gráfica (nuvens de palavras) de apreensão imediata, que representa a estatística da importância relativa dos programas em curso na instituição.

Em 2009, o DDBJ igualou-se aos seus dois parceiros (NCBI e EBI), quando passou a operar com dados gerados por tecnologias de sequenciamento de última geração. O formato utilizado, semelhante aos empregados pelo NCBI (SRA, de *Short Read Archive*) e pelo EBI (ERA, de *European Read Archive*), é designado **DRA** (*DDBJ Sequence Read Archive*). Os números de acesso carregam entre parênteses o prefixo DR, característico do banco, seguido por um número de seis dígitos (p. ex., DR000001). São aceitos os dados provenientes das três plataformas mais utilizadas (454, Illumina e SOLiD), em panilhas de arquivo Excel.

Além disso, o DDBJ desenvolveu uma ferramenta automática de anotação desse tipo de material (*DDBJ Read Annotation Pipeline*), compreendendo dois sub-processos: análise básica para mapeamento e montagem *de novo*, e análise de alto nível para combinar anotações

automáticas e manuais (p. ex., detecção de polimorfismos de um único nucleotídeo – SNPs e ESTs).

4.1.4 UniProt Knowledgebase

Os bancos de dados **Swiss-Prot Knowledge Database**, **TrEMBL** e **PIR** se juntaram para formar o **Universal Protein Resource (UniProt)**, que é um repositório central de sequências de proteínas e de anotação funcional. Esse é um banco de dados não redundante, concentrando informações que são incorporadas à sequência de proteínas por meio de programas de anotação automática e de predição de estrutura e domínios. Por fim, é feita uma detalhada anotação manual, curada por especialistas que atuam na anotação de proteínas humanas, bacterianas, vegetais, de fungos, entre outras. O UniProt mantém um alto nível de anotações, como a descrição e a função da proteína, estrutura dos seus domínios, as modificações pós-tradução, além de ter uma estrutura que facilita o acesso computacional a diferentes campos de informações.

4.2 Banco de dados secundários

Em geral, o objetivo dos **bancos de dados secundários** consiste em procurar sequências com função desconhecida em uma biblioteca de padrões para determinar se elas possuem ou não características predefinidas, o que permite dizer se elas pertencem ou não a uma determinada família. Se a função desta família for conhecida, a procura em **banco de motivos** pode levar a inferir sua função biológica.

4.2.1 PROSITE

O PROSITE foi o primeiro banco de estruturas a ser desenvolvido e é mantido atualmente pelo *Swiss Institute of Bioinformatics*. Baseia-se na filosofia de que famílias de proteínas podem ser fácil e simplesmente caracterizadas, por meio de um único motivo, o mais conservado, obtido pelo alinhamento múltiplo de sequências já conhecidas. Esses motivos geralmente estão associados às respectivas funções biológicas. No PROSITE, os motivos são codificados como **expressões regulares** ou **padrões** (*patterns*), o processo usado para obter estes padrões envolve a construção de alinhamentos múltiplos e a inspeção manual para a identificação de regiões conservadas. A informação assim obtida é transformada em expressões de consenso e o resultado é utilizado como a semente de buscas no SWISS-PROT. Esses resultados são manualmente testados para a avaliação do desempenho dos padrões. O ideal é que ocorram apenas os chamados verdadeiros positivos, sem que falsos positivos sejam detectados. Padrões com excesso de ocorrência de falso positivo têm o seu desempenho comprometido e deverão ser objeto de uma análise posterior mais detalhada. Este ciclo é repetido até que um padrão ótimo possa ser definido.

4.2.2 PRINTS

O PRINTS foi desenvolvido e é mantido pelo grupo de biologia computacional da University of Manchester,

UK, e é um compêndio de **impressões digitais (fingerprints)**. Uma impressão digital é um conjunto de motivos conservados usado para caracterizar uma família de proteínas. O PRINTS se baseia no fato de que uma inspeção detalhada dos alinhamentos de sequências revela que a maioria das famílias de proteínas é caracterizada por vários motivos conservados. Por isso, faz-se a utilização desses motivos como assinaturas de determinadas famílias de proteínas. **Impressões digitais** são inerentemente superiores aos métodos com base em um único motivo, pois levam em consideração as suas imbricações.

4.2.3 Pfam

O **Pfam** foi desenvolvido no *Wellcome Trust Sanger Institute* e tem como base um banco de dados de sequência chamado **Pfamseq**, que contém informações oriundas do UniProt. Assim como existem diferentes formas de utilização de motivos para a caracterização de famílias de proteínas, também existem diversos métodos de utilização do alinhamento completo de sequências para construir famílias de discriminadores. Uma alternativa ao uso de perfis é a codificação do alinhamento por meio de HMMs.

HMMs são modelos probabilísticos que consistem em coleções de estados intercomunicantes, nos quais as transições entre esses estados ocorrem conforme um mecanismo probabilístico. Os estados intercomunicantes constituem, em geral, uma cadeia linear que possui os significados correspondência, remoção e inserção. Atribui-se o estado correspondência a cada coluna conservada no alinhamento das sequências. O estado inserção, por sua vez, permite inserções relativas ao estado correspondência, ao passo que o estado remoção permite que posições correspondentes (*matched*) sejam ignoradas. Assim, a construção de um HMM para um alinhamento múltiplo entre sequências requer que cada posição dentro do alinhamento esteja associada a um dos três estados acima definidos.

4.2.4 InterPro

InterPro foi desenvolvido e é mantido pelo EMBL-EBI. É um recurso integrado de documentação para famílias, domínios, regiões e sítios de proteínas. O objetivo do InterPro é a obtenção de assinaturas de proteínas e, para isto, combina uma série de bancos de dados (PROSITE, PRINTS, Pfam, ProDom, SMART, TIGRFAMs, Gene3D, SUPERFAMILY, PIR Superfamily e PANTHER), que utilizam diferentes metodologias e graus de informações biológicas sobre as proteínas bem caracterizadas. Ao unir as bases de dados, capitaliza de cada banco de dados suas características individuais, produzindo um poderoso banco de dados integrado e uma ferramenta de diagnóstico. Cada entrada de **InterPro** inclui descrição funcional, anotação, referências bibliográficas e *links* para o banco de dados relevante.

4.3 Banco de dados de estruturas de proteínas

Muitas proteínas compartilham similaridades estruturais que refletem, em alguns casos, origens evolutivas comuns. O processo evolutivo envolve substituições, inserções e deleções na sequência de aminoácidos. Para proteínas distantemente relacionadas, essas mudanças podem ser extensas, produzindo desdobramentos nos quais os números e as orientações de estruturas secundárias variam de forma considerável.

Em uma tentativa de compreender melhor o relacionamento sequência estrutura e os processos evolutivos subjacentes que dão origem a diferentes famílias, foi criada uma grande variedade de esquemas de classificação de estruturas. Essa informação está armazenada nos **bancos de dados de estruturas** proteicas. A natureza da informação apresentada nesses bancos é dependente da filosofia do enfoque e, conseqüentemente, também é dependente dos métodos usados para identificar e avaliar as similaridades estruturais. Por exemplo, famílias estruturalmente derivadas com base em algoritmos que buscam e conglomeram por associação de motivos comuns serão diferentes daquelas geradas por procedimentos baseados na comparação de estruturas globais. Por sua vez, os resultados destes procedimentos automáticos também diferirão daqueles que se baseiam em inspeção visual, nos quais as ferramentas computacionais utilizadas são aquelas que têm por objetivo tornar viável este tipo de comparação.

4.3.1 SCOP

O **SCOP (Structural Classification of Proteins = Classificação estrutural de proteínas)** é um banco de dados mantido pelo Laboratory of Molecular Biology e Centre for Protein Engineering (MRC, *Medical Research Council*), que tem por objetivo descrever as relações estruturais e evolutivas entre proteínas de estruturas conhecidas. Na medida em que as ferramentas automáticas de comparação de estruturas hoje existentes não podem identificar, de forma confiável, essas relações, o SCOP foi concebido e construído, de modo a utilizar uma combinação de processos de inspeções manuais e métodos automáticos. Existem muitos níveis na hierarquia, mas os principais são família, superfamília e dobra. Como a posição exata das fronteiras entre estes níveis são, em certa medida, subjetivas, o **SCOP** tem uma classificação evolutiva conservadora: se existir qualquer dúvida sobre o parentesco, novas divisões na família e níveis superfamília serão criados.

4.3.2 CATH

O **CATH (de Class, Architecture, Topology, Homology = Classe, Arquitetura, Topologia e Homologia)** é um banco de dados de classificação hierárquica de domínios de estruturas de proteínas, mantido pela *University College London* (UCL). Esse recurso é produzido por métodos automáticos e, depois, passa por uma anotação manual. Diferentes categorias dentro da classificação são

identificadas por meio de números únicos e nomes descritivos, que permitem a manipulação eficiente dos dados.

Cada proteína é classificada em domínios estruturais, divididos em superfamílias homólogas, que são grupos de domínios relacionados evolutivamente. O processo de classificação utiliza uma combinação de técnicas manuais e automatizadas, incluindo algoritmos computacionais, evidências estatísticas e empíricas, revisão de literatura e análise de especialistas.

4.3.3 PDB

O **PDB** (de *Protein Data Bank* = Banco de dados de proteínas), também mantido pelo *EBI*, é um dos principais recursos para obtenção de informações estruturais. Trata-se de um compêndio, contendo estruturas de proteínas que foram resolvidas experimentalmente. Cada proteína possui um resumo, que fornece descrição da resolução, número de cadeias de proteínas, ligantes e íons metálicos, estruturas secundárias, interações ligantes, entre muitos outros. Essas informações são fundamentais, não somente para a visualização das estruturas, como também para o desenho, sendo o único recurso das informações 1D, 2D e 3D. Recursos deste tipo tornam-se cada vez mais importantes na medida em que técnicas de visualização sofisticadas são disponibilizadas.

4.4 Bancos de dados de genômica comparativa

4.4.1 MBGD

O MBGD (Microbial Genome Database for Comparative Analysis = Banco de dados de genomas microbianos para análise comparativa) faz parte de uma plataforma de genômica comparativa, mantida pelo *National Institutes of Natural Sciences* do Japão. Este banco é constantemente atualizado e a identificação dos grupos de genes ortólogos é feita de forma automática usando métodos de clusterização hierárquico. Uma característica proeminente do MBGD é que ele permite aos usuários criarem grupos de ortólogos, utilizando um subgrupo específico de organismos. Algumas funcionalidades importantes do MBGD são: a atribuição de anotação funcional para cada grupo de ortólogos, incluindo *links* externos para outros banco de dados; uma interface para a escolha de um conjunto de genomas, com base em propriedades fenotípicas; e a possibilidade do usuário adicionar seus próprios genomas sem nenhuma anotação (nesse caso, programas de predição de genes são executados antes da identificação dos ortólogos).

4.4.2 xBASE:

O **xBase** é um banco de dados mantido pela *University of Birmingham*, que aprimora as informações experimentais e de bioinformática de vários genomas bacterianos. O xBase provê ao usuário um acesso fácil à anotação e à comparação de genomas por meio de uma interface *Web*, sendo possível verificar o contexto genômico, a utilização de códons, os dados de anotação do GenBank e do Uniprot

e analisar prováveis ortólogos e alinhamentos de sequências com o programas MUMmer (Seção 3.3.1). Além disso, podem ser visualizados genomas completos, identificados motivos proteicos conservados e agrupados genes conforme suas propriedades, como conteúdo GC, entre outras.

4.4.3 MicrobesOnline

O **MicrobesOnline**, mantido pelo *Virtual Institute for Microbial Stress and Survival* (Instituto Virtual para Estresse Microbiano e Sobrevivência), é um banco de dados que inclui vários genomas completos de bactérias, arqueas e fungos, além de dados de expressão de diversos organismos. O MicrobesOnline inclui um navegador de genômica comparativa, com base em árvores filogenéticas para cada família de genes. Além de oferecer ferramentas para identificação de motivos conservados, buscas rápidas de perfil filogenético, comparações de vias metabólicas, predições de óperons, o MicrobesOnline possibilita a identificação de genes corregulados, pela procura de genes baseado no seu perfil de expressão.

4.4.4 HOGENOM

O HOGENOM é um banco de genes homólogos, desenvolvido por pesquisadores da *Université Claude Bernard, Lyon 1*. Este banco contém informações de bactérias, arqueas e eucariotos unicelulares, permitindo a visualização do alinhamento e de árvores filogenéticas de genes homólogos entre as espécies. O principal objetivo do HOGENOM é a identificação de aglomerados de genes ortólogos em uma ampla variedade de táxons. Para construir as famílias gênicas é realizado uma pesquisa de semelhança entre todas as proteínas usando o programa BLAST (Seção 2.1.3).

4.5 Bancos de dados funcionais e de metabolismo

O termo **genômica funcional** é, muitas vezes, utilizado para se referir as várias abordagens que levam à compreensão das propriedades e funções dos genes de um organismo e dos produtos dos seus genes. A comparação do proteoma de genomas completos e a identificação de vias bioquímicas, por exemplo, são instrumentos importantes para descobrir as características metabólicas dos organismos. A seguir, serão discutidos os dois bancos de dados mais utilizados para essas análises.

4.5.1 COG

O banco de dados **COG** (de *Cluster of Orthologous Groups* = Conjunto de grupos ortólogos) do NCBI é um banco de dados com base no conceito de ortologia. O COG é construído por meio da comparação de todas as sequências de proteínas codificadas em vários genomas completos. Cada grupo deve conter sequências de proteínas de, pelo menos, três genomas distintos. A premissa do COG é que as proteínas conservadas nesses genomas, a partir de organismos diversos, representam funções antigas, que foram conservadas ao longo da evolução. A abordagem

utilizada para a construção de COG não substitui uma análise filogenética abrangente. No entanto, ele oferece um atalho rápido e conveniente para delinear um grande número de famílias que, provavelmente, são ortólogas.

4.5.2 KEGG

A **KEGG** (de *Kyoto Encyclopedia of Genes and Genomes* = Enciclopédia de Genes e Genomas), desenvolvida pelo DDBJ, é um dos melhores recursos sobre funções celulares e consiste em 16 bancos de dados principais, classificados em sistemas de informação, informações genômicas e informações químicas. A KEGG é utilizada

principalmente em estudos sobre redes moleculares (mapas e vias metabólicas) e listas hierárquicas (hierarquias funcionais). Este banco de dados funcional fornece uma visão metabólica geral com uma ilustração em forma de um mapa, ao invés de somente texto, o que o torna mais fácil de ser usado. A KEGG fornece listas de números EC e suas enzimas correspondentes separadas por nível, muitos *links* úteis para sítios que descrevem enzimas e a nomenclatura detalhada do ligante. O banco de dados LIGAND, associado à KEGG, é um recurso útil para identificar pequenos ligantes envolvidos em vias bioquímicas. A KEGG pode ser pesquisada por homologia de sequências, palavras-chave e identidade química.



Resumo

A bioinformática é uma área multidisciplinar, envolvendo biologia, matemática, computação e estatística, e que tem, como um dos seus desafios, a análise dos dados gerados pelos diversos projetos de sequenciamento de genomas. Como a **anotação** (predição da função de um gene) é geralmente realizada por meio da sua comparação com sequências depositadas em vários bancos de dados, é importante conhecer a forma de organização dos conteúdos dos bancos de dados, para a realização da anotação correta de genomas. Essas informações estão armazenadas em vários bancos de dados públicos, que contêm sequências de DNA e de proteínas (**bancos primários**), de motivos proteicos conservados (**bancos secundários** e **bancos de dados de estruturas de proteínas**). A busca pode ainda ser realizada em **bancos funcionais e de metabolismo**, e o objetivo é sempre o de adicionar a informação biológica às sequências de DNA.

Se, por um lado, a genética experimental não cessa de fornecer notáveis progressos nas técnicas de sequenciamento, por outro, isso leva ao acúmulo de múltiplos genomas para estudos cada vez mais detalhados. A genômica comparativa tornou-se uma ferramenta poderosa para identificação de novas famílias gênicas e de estudos evolutivos. Algoritmos eficientes para o alinhamento de pares de sequências, do tipo BLAST, encontram-se otimizados e são, rotineiramente, utilizados. Entretanto, o desenvolvimento de métodos rápidos e eficientes para comparar dois ou mais genomas são cada vez mais necessários. Os recursos apresentados neste capítulo, encontram-se (a maioria) disponíveis na internet. Assim, a genética e a genômica modernas enfrentam o desafio de aplicar esses novos recursos ao estudo da dinâmica da atuação dos genes e dos genomas, para alcançar um melhor entendimento de sua biologia.



Leituras recomendadas

Barnes M, editor. Bioinformatics for geneticists: a bioinformatics primer for the analysis of genetic data. Chichester: Wiley; 2007.

Baxevanis AD, Ouellette BFF, editors. Bioinformatics: a practical guide to the analysis of genes and proteins. Hoboken: Wiley; 2005.

Claverie JM, Notredame C. Bioinformatics for dummies. 2nd ed. Hoboken: Wiley; 2007.


Horner DS, Pavesi G, Castrignanò T, De Meo PD, Liuni S, Sammeth M, et al. Bioinformatics approaches for

genomics and post genomics applications of nextgeneration sequencing. Brief Bioinform. 2010;11(2):181-97.

Koboldt DC, Ding L, Mardis ER, Wilson RK. Challenges of sequencing human genomes. Brief Bioinform. 2010;11(5):484-98.

Mount DW. Bioinformatics: sequence and genome analysis. 2nd ed. New York: Cold Spring Harbor Laboratory; 2004.

Nagarajan N, Pop M. Sequencing and genome assembly using next-generation technologies. Methods Mol Biol. 2010;673:1-17.



Ng PC, Kirkness EF. Whole genome sequencing. *Methods Mol Biol.* 2010;628:215-26.

Nucleic Acids Research. 2011 NAR database summary paper category list [Internet]. Oxford: University Press; c2011 [capturado 28 ago. 2011]. Disponível em: <http://www.oxfordjournals.org/nar/database/c/>.

Ramsden JJ. *Bioinformatics: an introduction*. 2nd ed. London: Springer; 2009.

Saccone C, Pesole G. *Handbook of comparative genomics: principles and methodology*. Hoboken: Wiley-Liss; 2003.

Glossário

5'-CAP: resíduo de 7-metil-guanosina presente na extremidade 5' de mRNAs eucarióticos.

Acetilases: enzimas que catalisam a adição de radicais acetílicos a aminoácido(s) de uma proteína.

Agente alquilante: substância que introduz grupamentos metílicos ou etílicos em determinadas bases do DNA.

Agente mutagênico: agente químico ou físico que provoca mutações.

Algoritmo de Smith-Waterman: é um algoritmo que utiliza programação dinâmica para encontrar alinhamentos locais entre as sequências. Sua principal característica é todos os escores negativos, calculados na programação dinâmica, serem alterados para zero, evitando alinhamentos ruins e ajudando na identificação de alinhamentos locais.

Algoritmo: é um procedimento sistemático para resolver um problema em um número finito de etapas, normalmente envolvendo uma repetição de operações. Uma vez especificado, um algoritmo pode ser escrito em uma linguagem de computador e executado como um programa.

Alinhamento: é o procedimento de comparação entre duas ou mais sequências, analisando uma série de caracteres individuais ou padrões de caracteres que estão na mesma ordem nas sequências. Dos dois tipos de alinhamento, local e global, o alinhamento local é o mais utilizado.

Alquilação: adição de um grupamento metílico ou etílico em uma bases do DNA.

Aminoacil-tRNA: um tRNA ligado a seu aminoácido.

Análise de cluster: é um método para agrupar um conjunto de objetos que são mais similares a partir de um grupo maior de objetos relacionados. As relações se baseiam em algum critério de similaridade ou de diferença. Para análise

de sequências é usado uma pontuação de similaridade ou de distância ou, ainda, uma avaliação estatística dos escores encontrados.

Análogo de base: substância química que mimetiza uma base nitrogenada, podendo ser incorporada em seu lugar, em uma cadeia polinucleotídica em crescimento. A frequência com que o análogo de base muda para a sua forma tautomérica pode ser 50 vezes maior do que a da base original, o que aumenta a incidência de mutações.

Anelamento de fitas simples (SSA, *single-strand annealing*): mecanismo de reparação de uma quebra dupla no DNA que ocorre entre sequências repetidas e orientadas na mesma direção, normalmente causando a deleção de mais de 30 nucleotídeos no sítio onde ocorreu a quebra.

Anelamento: ligação de um oligonucleotídeo a uma molécula de DNA de fita simples por pareamento entre bases complementares.

Anfipático: caráter de uma molécula que apresenta uma região hidrofóbica e uma hidrofílica, permitindo, assim, interações não covalentes em meio aquoso com outras moléculas hidrofóbicas e/ou hidrofílicas.

Anticódon: uma sequência de três nucleotídeos presente no tRNA que é complementar e que interage com o códon por pareamento de bases.

Aptâmeros: oligonucleotídeos ou pequenos peptídeos sintéticos que se ligam a uma molécula-alvo com alta afinidade e de maneira específica.

Archaea: é um dos três domínios nos quais são divididos os seres vivos segundo classificação proposta por Carl Woese e Ralph Wolfe, em 1977. Ele inclui todas as espécies de arqueas, que são organismos procarióticos geralmente quimiotróficos, muitos dos quais habitam ambientes extremos (sendo,

por isso, chamados de extremófilos). Dentre os ambientes extremos onde são encontradas arqueas podem ser citados fontes de água quente, lagos ou mares com elevada salinidade, pântanos e ambientes ricos em gás sulfídrico e com altas temperaturas. As arqueas, juntamente com as bactérias, são classificadas como organismos procarióticos.

Auto-splicing: remoção de íntron e união de éxons realizado pela atividade enzimática do próprio RNA que está sendo processado.

Autorradiografia: método para a detecção de moléculas de DNA marcadas, seja radioativamente ou por quimioluminescência, pela exposição de um filme fotográfico sensível a raios X.

Bactéria: um dos três domínios nos quais são divididos os seres vivos segundo classificação proposta por Carl Woese e Ralph Wolfe, em 1977. Ele inclui todas as espécies de bactérias, que, juntamente com as arqueas, são classificadas como organismos procarióticos.

Bacteriófagos: são vírus que infectam procariotos.

Banco de dados: é um conjunto de registros dispostos em estrutura regular e que pode ser pesquisado ou alterado de forma padronizada, por meio de aplicativos especiais, chamados sistemas gerenciadores de bancos de dados (SGBDs).

Biblioteca de cDNA: coleção de cDNAs clonados produzidos a partir de uma população inteira de mRNA (que pode ser de células isoladas, tecidos, órgãos ou organismos).

Biblioteca genômica: coleção de moléculas de DNA genômico clonadas em número suficiente para incluir todas as sequências do genoma de um determinado organismo.

Bivalentes: cromossomos homólogos em sinapse, na primeira etapa da meiose (meiose I).

box SOS: motivo na sequência reguladora dos genes SOS onde o repressor LexA se liga para impedir a transcrição desses genes em nível elevado.

Brometo de etídeo: reagente químico fluorescente que se intercala entre os pares de bases de uma molécula de DNA de fita dupla; utilizado para visualização de DNA.

Capsídeo: revestimento proteico externo de uma partícula viral que envolve o genoma viral.

Catabolismo: parte do metabolismo que se refere à assimilação ou processamento da matéria orgânica adquirida pelos seres vivos para fins de obtenção de energia.

Catabólito: produto resultante do catabolismo.

cauda de poli(A): conjunto de resíduos de adenilato presentes na extremidade 3' da maioria dos mRNAs eucarióticos.

cDNA: ver DNA complementar.

Chaperonas de proteínas: as chaperonas de proteínas formam uma família proteica encontrada em todos os organismos. Elas se ligam a outras proteínas, auxiliando no seu enovelamento para que assumam uma conformação correta. A família das chaperonas inclui chaperonas moleculares e chaperoninas. As chaperonas moleculares se ligam a proteínas total ou parcialmente desenoveladas e as

estabilizam, impedindo que elas sejam degradadas, ao passo que as chaperoninas atuam facilitando o enovelamento das proteínas.

Chaperonas: proteínas que se ligam a outras proteínas que estão incorretamente enoveladas ou organizadas, auxiliando no enovelamento correto dessas proteínas.

Ciclo lisogênico: um dos dois ciclos usados por vírus para sua reprodução, em que o vírus insere seu genoma na célula hospedeira e é incorporado ao genoma da célula infectada.

Ciclo lítico: um dos dois ciclos usados por vírus para sua reprodução, no qual o genoma viral é inserido na célula hospedeira, é replicado de maneira independente e comanda a síntese das proteínas virais. Assim, são produzidas novas partículas virais, ocorre a lise da célula hospedeira e os novos fagos são liberados.

Cístron: segmento de DNA (ou RNA) que codifica uma proteína ou uma cadeia polipeptídica (ver Região codificadora).

Clonagem: a produção de muitas cópias idênticas de uma molécula de DNA por replicação em um vetor apropriado; também denominado de clonagem gênica ou clonagem molecular; a geração de células (ou indivíduos) geneticamente idênticas entre si e aos seus progenitores.

Clone: uma população de células ou organismos idênticos. Na clonagem molecular, um clone contém moléculas de DNA recombinantes idênticas.

Códon: uma sequência de três nucleotídeos das moléculas de RNA mensageiro (mRNA) ou da região codificadora do gene correspondente que codifica um determinado aminoácido.

Cofatores enzimáticos: pequena molécula orgânica ou íon metálico necessário para a atividade catalítica de uma enzima.

Competentes: em procedimentos de transformação bacteriana são chamadas de competentes bactérias que foram tratadas para aumentar a sua capacidade de receber moléculas de DNA exógenas.

Complexo sinaptonêmico: organela que consiste em uma faixa tríplice de nucleoproteína formada entre os cromossomos homólogos pareados na fase de paquíteno da prófase da meiose I.

Computação paralela: é caracterizada pelo uso de várias unidades de processamento para executar uma computação de forma mais rápida. É baseada no fato de que o processo de resolução de um problema pode ser dividido em tarefas menores, que podem ser realizadas simultaneamente – em paralelo – por algum tipo de coordenação.

Concatâmeros: associações lineares de múltiplas cópias de um genoma formando longas cadeias. Concatâmeros são formados, por exemplo, em processos de replicação de certos fagos.

Conformação nativa: conformação tridimensional necessária para que uma proteína seja funcional.

Conjugação: é a transferência de material genético entre células bacterianas pelo contato direto célula-célula ou por apêndices chamados fimbrias ou pilus sexuais.

Conversão gênica: é a alteração de uma fita de um DNA heterodúplex, a fim de torná-lo complementar à outra fita em qualquer posição em que existam bases mal-pareadas. Processo de recombinação não recíproca em que um alelo de um heterozigoto é convertido no seu alelo correspondente.

Cópias gênicas: duas ou mais cópias idênticas (ou praticamente idênticas) de um mesmo gene em um genoma. Em geral, essas cópias são originadas a partir de um ou mais eventos de duplicação de segmentos cromossômicos. As cópias geradas são mantidas idênticas ou muito similares, em razão de alguma pressão seletiva para isso.

Corpo P: é um complexo citoplasmático formado por agregados de proteínas e mRNAs específicos com duas funções: manter determinados mRNAs em uma forma de tradução reprimida reversível, de forma que os mRNAs podem sair do corpo P e serem traduzidos; e recrutar mRNAs que estão direcionados para desadenilação e degradação pela via de remoção de CAP/*Xrn1*.

Cosmídeo: tipo de vetor de clonagem contendo o sítio *cos* do bacteriófago λ inserido em um plasmídeo, utilizado para a clonagem de fragmentos de DNA de até 40 kb de tamanho.

Cromatina: é uma estrutura nucleoproteica organizada, constituída de DNA associado a proteínas específicas, como as histonas (em eucariotos e algumas arqueas).

Cromossomo artificial bacteriano (BAC, *bacterial artificial chromosome*): tipo de vetor para a clonagem de fragmentos de DNA de até cerca de 200 kb de tamanho em *E. coli*. Um BAC contém uma origem de replicação plasmidial, um sítio múltiplo de clonagem e um marcador de seleção.

Cromossomo artificial de levedura (YAC, *yeast artificial chromosome*): tipo de vetor para a clonagem de fragmentos de DNA grandes (normalmente maiores que 300 kb) em levedura. Um YAC é uma molécula linear com um telômero em cada extremidade, um centrômero, uma sequência de replicação autônoma, um marcador de seleção e um sítio múltiplo de clonagem.

Cromossomo: em uma acepção moderna, é cada uma das moléculas de DNA que constitui o genoma de um organismo. Assim, procariotos típicos têm um único cromossomo (em geral, na forma de uma molécula de DNA circular) e células de eucariotos têm vários cromossomos (o número varia de acordo com a espécie), cada um deles formado por uma molécula de DNA linear.

Crossing-over: ou sobre cruzamento, troca de material cromossômico (partes dos braços cromossômicos) entre cromossomos homólogos mediante a quebra e a reunião. A troca de material entre cromátides não irmãs durante a meiose é a base da recombinação genética.

Dedo de zinco: domínio proteico contendo resíduos de cisteínas e histidinas, ou somente cisteínas que coordenam um íon zinco (Zn^{2+}), de modo a formar uma projeção semelhante a um dedo; daí o nome dedo de zinco. Os fatores de transcrição contendo domínio(s) de dedo de zinco são os mais frequentes entre todos os fatores transcripcionais.

Depurinação: resultado da perda de uma base nitrogenada (normalmente uma purina) de um nucleotídeo na cadeia

de DNA, criando um sítio abásico (ou apurínico ou apirimidínico).

Desacetilases: enzimas que catalisam a remoção de radicais acetílicos que foram adicionados a aminoácido(s) de uma proteína por uma acetilase.

Desaminação: no DNA resulta na remoção do grupamento amínico das bases nitrogenadas. Pode ocorrer espontaneamente, por hidrólise, ou pela ação do ácido nitroso.

Desnaturação de moléculas de ácido nucleico: quebra, por meios químicos ou físicos, das pontes de hidrogênio envolvidas no pareamento de bases.

Desoxirribonuclease: uma desoxirribonuclease é uma enzima que degrada DNA, catalisando a quebra de ligações fosfodiéster entre os desoxirribonucleotídeos sem especificidade de sequência.

Didesoxirribonucleotídeo (ddNTP): um nucleotídeo modificado que possui um 3'-H no açúcar desoxirribose, em vez de um 3'-OH. Um ddNTP pode ser incorporado a uma cadeia crescente de DNA, mas não permite que a síntese dessa cadeia prossiga, uma vez que não será possível a formação de uma ligação fosfodiéster com um outro nucleotídeo.

Difração por raios X: é uma metodologia empregada para descrever o arranjo dos átomos em um cristal. Basicamente, um feixe de raios X incide sobre uma amostra cristalina, os raios são absorvidos e difratados e a energia e os ângulos de difração resultantes são detectados em um espectrômetro de raios X. A análise destes valores permite determinar a posição espacial (coordenadas x, y e z) para cada átomo no cristal. Em biologia molecular esta metodologia é utilizada para descrever a estrutura (conformação espacial ou tridimensional) de ácidos nucleicos e proteínas.

Dímero de pirimidinas: produto da ação da luz UV, em que duas bases pirimídicas adjacentes de uma cadeia de DNA ficam ligadas entre si.

DNA complementar (cDNA): cópias de DNA sintetizadas a partir de moldes de RNA em uma reação catalisada pela enzima transcriptase reversa.

DNA-ligase: enzima que catalisa a formação de uma ligação fosfodiéster entre a extremidade 5' de uma cadeia de DNA e a extremidade 3' de uma outra cadeia de DNA, durante a replicação e a reparação do DNA, também utilizada para a construção de moléculas de DNA recombinantes.

DNA-polimerase da família Y: DNA-polimerase translesão que introduz nucleotídeos de forma aleatória em uma cadeia de DNA, quando a cadeia molde contém uma lesão.

DNA-polimerase de desvio (*by pass DNA-polymerase*): DNA-polimerase capaz de sintetizar o DNA, mesmo quando a fita-molde contém uma lesão.

DNA-polimerase: enzima que sintetiza uma nova fita de DNA a partir de uma fita-molde de DNA.

DNA recombinante: qualquer molécula de DNA que tenha sido construída em um tubo de ensaio e contenha sequências provenientes de duas ou mais moléculas de DNA distintas, frequentemente de organismos diferentes.

DNA-replicase: é a DNA-polimerase que realiza a síntese das duas novas fitas de DNA durante a replicação.

DnaA: proteína responsável pelo reconhecimento de origem da replicação em *E. coli*.

Domínio (de proteína): grupo de um ou poucos segmentos polipeptídeos com uma extensão aproximada de 100 a 200 aminoácidos que se enovelam em uma unidade estrutural.

Domínio TPR (de tetratricopeptide repeat): é um motivo estrutural encontrado em múltiplas cópias em diversas proteínas com funções distintas, facilitando interações específicas com outras proteínas parceiras.

Edição de RNA: conjunto de processos pelos quais a sequência nucleotídica de um RNA é alterada em relação à do DNA a partir do qual foi transcrito. Já foi descrita para moléculas de rRNA, tRNA e mRNA em diferentes organismos eucarióticos.

Elementos de transposição (transposons): segmento de DNA que se move de uma posição para outra no genoma por transposição

Elementos genéticos móveis: são sequências de DNA que podem mover-se dentro dos genomas. Os principais elementos genéticos móveis são: plasmídeos, bacteriófagos, elementos transponíveis e íntrons do grupo II.

Eletroforese em gel de agarose: procedimento experimental no qual um campo elétrico é utilizado para movimentar moléculas de DNA ou RNA, negativamente carregadas, por uma matriz de agarose gelificada do polo negativo para o polo positivo.

Eletroporação: método para aumentar a aquisição de DNA por protoplastos ou células competentes pela exposição prévia a alta voltagem elétrica, que resulta na formação temporária de pequenos poros na membrana celular.

Endonucleases de homing: endonucleases de restrição geralmente codificadas por íntrons e envolvidas na mobilidade de íntrons.

Endossimbionte: qualquer organismo que vive no interior do corpo ou das células de outro organismo, em uma relação ecológica designada como endossimbiose. Na endossimbiose, o endossimbionte supre a célula ou o organismo hospedeira(o) com alguma função, enquanto sobrevive às custas dela(e) sem, contudo, prejudicá-la(o) com isso.

Envelope viral: revestimento externo de alguns vírus, que envolve o capsídeo, formado por uma dupla camada de fosfolípidos e proteínas.

Enzima de restrição (endonuclease de restrição): tipo de enzima que cliva moléculas de DNA de fita dupla dentro ou perto de uma sequência nucleotídica específica (sítio de restrição), que, frequentemente, está presente em múltiplas cópias em um mesmo genoma. Essas enzimas são utilizadas para análise do DNA e na construção de moléculas de DNA recombinantes.

Enzima Hin: enzima de recombinação sítio-específica que atua na inversão de um segmento de DNA no genoma da bactéria *Salmonella*.

Enzima: proteína que catalisa (isto é, facilita do ponto de vista termodinâmico) uma reação química em um sistema biológico.

Enzimas constitutivas: enzimas cuja síntese é constante e não dependente da presença de substratos específicos.

Enzimas induzíveis: enzimas cuja síntese é induzida pela presença de um substrato adequado.

Epigenética: é o estudo das mudanças herdáveis no fenótipo por mecanismos que não envolvem mudanças na sequência de DNA.

Epissomos: elementos genéticos extracromossômicos capazes de se integrarem a um cromossomo procariótico. São exemplos de epissomos, os plasmídeos integrativos e os bacteriófagos temperados (capazes de realizar um ciclo lisogênico).

Estrutura de Holliday: na recombinação do DNA, é uma estrutura intermediária vista nas imagens de microscopia eletrônica de transmissão em forma de X, mostrando quatro regiões de DNA de fita simples.

Eucariotos: grupo de organismos que possui a organização celular com núcleo, incluindo o material genético, e outras organelas, delimitados por membranas.

Eucromatina: é a forma menos compactada da cromatina interfásica eucariótica.

Eukarya: um dos três domínios nos quais são divididos os seres vivos segundo classificação proposta por Carl Woese e Ralph Wolfe, em 1977. Os eucariotos são todos os organismos com células compartimentalizadas em organelas, sendo o núcleo a mais típica delas.

Excisionase: enzima de recombinação sítio-específica que atua juntamente a integrase de λ para remover o genoma do fago do cromossomo da bactéria hospedeira e, assim, ingressar no ciclo lítico.

Éxon: sequência de um gene que é transcrita no pré-RNA e mantida, ao contrário dos íntrons, no RNA maduro (processado).

Exonuclease: enzima que remove nucleotídeo a partir de uma das extremidades de uma fita de DNA.

Exossomo: complexo multiproteico dotado de atividade exonucleolítica (3' → 5') envolvido na degradação seletiva de RNAs.

Expressão gênica: inclui todas as etapas envolvidas na geração do produto final de um gene, seja ele nucleotídico (no caso de genes cujo produto final de expressão é um RNA, como genes de rRNAs e tRNAs) ou polipeptídico (no caso de genes que codificam proteínas). As principais etapas da expressão gênica são a transcrição, o processamento de transcritos (principalmente em eucariotos) e a tradução.

Expressões regulares: ferramenta computacional que fornece um método para expressar as variações encontradas em um conjunto de sequências relacionadas, incluindo posições, inserções e repetições, entre outras.

Extremidade cega (blunt-end): extremidade de uma molécula de DNA na qual ambas as fitas terminam na mesma posição nucleotídica, sem qualquer extensão de fita simples.

Extremidade coesiva: extremidade de uma molécula de DNA da fita dupla na qual uma das fitas se projeta além da outra, na forma de DNA de fita simples. Moléculas de DNA

com esse tipo de extremidades, se forem complementares, podem parear entre si.

Famílias de parálogos: famílias de genes ou de outras sequências homólogas (parálogas) geradas por duplicação seguida de divergência.

Fator de transcrição: proteína que, junto à RNA-polimerase, é necessária para iniciar ou regular a transcrição em eucariotos. Um fator geral de transcrição participa da formação do complexo de iniciação da transcrição e é requerido para a transcrição de todos os genes transcritos por um determinado tipo de RNA-polimerase. Os fatores de transcrição específicos estimulam (ativadores) ou inibem (repressores) a transcrição de genes específicos.

Fator liberador: proteína ou proteínas responsáveis por impedir a reativação de origens da replicação em arqueas e eucariotos.

Fita codificadora: no DNA de fita dupla, a fita codificadora de um gene é aquela que possui a mesma orientação (de 5' para 3') da fita de RNA que é transcrita a partir deste gene. Salvo as diferenças inerentes a moléculas de DNA e RNA quanto a suas composições nucleotídicas, as sequências da fita codificadora do gene e de seu transcrito são idênticas.

Fita tardia: a fita tardia na replicação do DNA de fita dupla é aquela que é sintetizada descontinuamente.

Fita-líder: a fita-líder na replicação do DNA de fita dupla é aquela que é sintetizada continuamente.

Fita-molde: em um DNA de fita dupla, é a fita ou porção de uma fita utilizada como molde para síntese de uma cadeia polinucleotídica (de DNA ou RNA) complementar a ela. Na replicação, cada uma das fitas da molécula de DNA parental é utilizada como molde para a síntese de uma nova fita complementar e de orientação oposta. Na transcrição, a fita-molde de um gene é aquela complementar à fita codificadora; ela é interpretada de 3' para 5' para que seja sintetizado um produto de RNA complementar a ela e correspondente, em sequência e orientação, à fita codificadora do gene.

Fosmídeo: vetor que contém a origem de replicação do plasmídeo F de *E. coli* e a sequência *cos* do bacteriófago lambda.

Fotoliase: enzima que executa a reparação por fotorreativação enzimática na presença de luz visível.

Fotorreativação enzimática: mecanismo de reparação de mutações provocadas pela luz UV por absorção de luz visível pela enzima fotoliase.

Fragmento de Klenow: fragmento maior da DNA-polimerase I de *E. coli* produzido a partir de clivagem proteolítica da enzima. O fragmento de Klenow corresponde à porção aminoterminal da enzima, que contém as atividades de polimerase 5'→3' e de exonuclease 3'→5'). O fragmento de Klenow pode ser utilizado *in vitro* para a síntese de novos segmentos de DNA', em substituição aos segmentos preexistentes (que são removidos pela atividade exonucleásica 3'→5').

Fragmentos de Okazaki: segmentos de DNA produzidos durante a replicação da fita descontínua que são, posteriormente, unidos pela enzima DNA-ligase, formando uma cadeia linear única.

Gene interrompido: é um gene que possui íntron(s).

Gene: levando em consideração aspectos moleculares estruturais e funcionais, um gene pode ser simplificada e conceituado como um segmento de uma molécula de DNA que carrega a informação necessária e suficiente para codificar um produto de RNA (p. ex., moléculas de rRNA ou de tRNA) ou proteico (no caso de genes que codificam mRNAs, os quais são traduzidos em cadeias polipeptídicas). O gene é formado não apenas pelas sequências codificadoras de seus produtos de transcrição e/ou tradução, mas também pelas sequências reguladoras diretamente responsáveis pela sua expressão.

Genes marcadores: alelos que produzem diferenças fenotípicas detectáveis, úteis na análise genética de um organismo ou células.

Genes SOS: conjunto de genes, originalmente descrito em *E. coli*, envolvidos em diversos processos de reparação do DNA e que têm a sua expressão induzida quando o número de lesões é elevado.

Genoma mínimo: conceito idealizado correspondente ao menor genoma capaz de abrigar o número mínimo de genes necessário para a sobrevivência de uma forma de vida celular (que seria a procariótica). A definição de um genoma mínimo (ou de um conteúdo próximo ao mínimo possível para um genoma) depende do modo de vida do organismo, pois o conteúdo mínimo necessário de genes seria diferente, por exemplo, para espécies de vida livre ou para espécies parasitas ou simbióticas.

Genoma nuclear: conteúdo de DNA genômico encontrado no núcleo de células eucarióticas.

Genoma: é toda a informação genética de um organismo. No caso de formas de vida celulares (procariotos e eucariotos) essa informação genética está codificada no DNA de cada célula, o qual constitui um ou mais cromossomos, dependendo da espécie.

Genomas extranucleares: são os genomas de organelas (mitocôndrias e plastídeos).

Genômica: estudo sistemático das sequências completas de DNA (genomas) dos organismos.

Helicase (DNA-helicase): enzima que realiza a separação das fitas de DNA pelo rompimento das pontes de hidrogênio entre os pares de nucleotídeos complementares utilizando a energia do ATP.

Hélice-alça-hélice: tipo de motivo proteico presente em fatores de transcrição. Ele é formado por duas hélices anfipáticas separadas por uma alça não helicoidal e suas características básicas facilitam a ligação ao DNA.

Hemizigoto: indivíduo no qual um gene (ou locus) está presente uma única vez, e não aos pares em células diploides. Normalmente este termo é aplicado aos genes presentes no cromossomo X de machos heterogaméticos.

Heterocromatina: é a forma mais compactada da cromatina interfásica eucariótica.

Heterozigoto: indivíduo com alelos diferentes em um ou mais loci.

Heurística: define-se procedimento heurístico como um método de aproximação das soluções ideais dos problemas.

A heurística assume uma solução próxima da ideal, com base na função de avaliação do resultado. A solução encontrada por uma heurística não é necessariamente a melhor possível, e não há garantia da distância entre a solução produzida pelo método heurístico e a solução ideal.

Hibridização: processo que utiliza a capacidade de pareamento de bases entre uma sonda de DNA de fita simples (marcada radioativamente ou com peroxidase) e moléculas de DNA de fita simples ou de RNA imobilizadas em um suporte físico (membrana de náilon ou de nitrocelulose). Visa à identificação dos raros clones de uma biblioteca que contêm o fragmento de DNA de interesse.

Hidrogenossomo: organela celular com capacidade de produzir hidrogênio molecular e compostos como acetato, dióxido de carbono e ATP. São encontrados em alguns eucariotos anaeróbios, como flagelados, ciliados e fungos criticidomicetos. Apesar de, em sua maioria, não possuírem genomas próprios (só há uma exceção conhecida), os hidrogenossomos são estruturais, funcionais e evolutivamente relacionados a mitocôndrias.

Histonas: são proteínas de caráter básico, que fazem parte da cromatina de eucariotos e de arqueas.

Homeodomínio: domínio de ligação ao DNA encontrado em determinados fatores de transcrição. Foi primeiramente identificado em fatores de transcrição codificados pelos chamados genes homeóticos, que são importantes para o desenvolvimento de *Drosophila melanogaster*. O homeodomínio apresenta três α -hélices, sendo que as α -hélices 2 e 3 fazem parte do motivo hélice-volta-hélice envolvido na ligação ao DNA.

Homologia (de sequências polinucleotídicas e polipeptídicas): homologia entre sequências nucleotídicas de DNA ou RNA ou entre sequências de aminoácidos de cadeias polipeptídicas são identidades significativas (em geral, > 30-40%) ao longo de pelo menos 50-60% de suas extensões em decorrência de uma origem evolutiva comum. Ver também Ortólogos, Parálogos e Xenólogos.

Homozigoto: indivíduo com alelos idênticos de um gene ou genes de interesse.

Hormônios lipofílicos: pequenas moléculas hidrofóbicas capazes de atravessar a membrana plasmática e atuar junto aos receptores correspondentes em processos de regulação transcricional de determinados genes.

Inativação por inserção: estratégia de clonagem molecular em que a inserção de um fragmento extra de DNA, em uma outra molécula de DNA, inativa um gene presente nessa molécula.

Iniciadores (primers): na replicação, são oligonucleotídeos de RNA complementares a uma fita-molde de DNA de fita simples, sintetizados por uma primase e utilizados pela DNA-replicase para o alongamento das novas cadeias sintetizadas.

Integrase de λ : enzima de recombinação sítio-específica responsável pela integração do genoma do bacteriófago λ ao cromossomo da bactéria hospedeira.

Integrase: enzima produzida por retrovírus e retrotransposons que permite integrar o material genético desses vírus

e retroelementos transponíveis no DNA das células hospedeiras.

Inteínas: segmentos de proteínas que têm a capacidade de autoexcisão do restante e de propiciar a união das porções adjacentes da proteína. Íntrons de um precursor de mRNA.

Interfase: fase do ciclo celular entre divisões mitóticas ou meióticas.

Íntron: sequência de um gene que é transcrita no pré-RNA, mas removida (por *splicing*), e está ausente do RNA maduro.

Íntrons do grupo I: sequências curtas de RNA, capazes de auto-*splicing*, que interrompem genes de bacteriófagos, genes que codificam proteína, rRNA e tRNA em bactérias e genes de mitocôndrias e cloroplastos de eucariotos inferiores.

Íntrons do grupo II: RNAs capazes de auto-*splicing* que atuam como elementos genéticos móveis, constituídos de um íntron cataliticamente ativo (ribozima) e uma proteína codificada pelo íntron.

Íntrons spliceossômicos: segmentos de DNA não codificadores, presentes em genes nucleares que codificam proteínas, excisados do pré-mRNA pelo spliceossomo.

Invasão de fita: etapa em que uma das fitas de uma molécula de DNA recombinante invade uma outra molécula de DNA homóloga com a qual irá iniciar um processo de recombinação.

Isoaceptores: diferentes tRNAs que se ligam a um mesmo aminoácido.

Junção heterodúplex: é o ponto no qual duas moléculas de DNA de fita dupla recombinantes são conectadas.

Junção recombinante: na recombinação, é o ponto no qual uma fita de DNA cruza de um dúplice para outro (é o limite da junção heterodúplex).

Lacuna: incompatibilidade no alinhamento de duas sequências causada por uma inserção em uma sequência ou uma deleção em outra.

Laranja de acridina: corante catiônico fluorescente seletivo para ácidos nucleicos. É uma substância mutagênica que se intercala nas cadeias de uma molécula de DNA de fita dupla, provocando mutação por modificação da fase de leitura.

Ligação peptídica: ligação amida entre o grupo amínico de um aminoácido com o grupo carboxílico de outro.

Ligações glicosídicas: ligação covalente formada quando um grupamento hidroxílico do carbono anomérico de um carboidrato reage com uma hidroxila de outro carboidrato.

Macroarranjos de DNA: um conjunto limitado de sondas oligonucleotídicas ligadas a um filtro de náilon. Esse filtro é usado para hibridização com DNA genômico ou RNAs marcados.

Mapeamento de restrição: procedimento para o posicionamento relativo dos sítios de clivagem de uma determinada enzima de restrição em um segmento de DNA clonado, gerando um mapa de restrição do fragmento.

Maturação do RNA: *ver* Processamento de RNA.

Mediador: complexo de proteínas que serve como um adaptador entre fatores de transcrição específicos ligados a

elementos reguladores e a RNA-polimerase II, ligada aos fatores de transcrição gerais.

Meiose: processo de divisão celular no qual o número de ploide de cromossomos é reduzido à metade (número haploide de cromossomos).

Microarranjos de DNA: um conjunto de sondas oligonucleotídicas correspondentes a genes específicos ligadas a uma lâmina de vidro em alta densidade. Essa lâmina é utilizada para hibridização com DNA ou RNA marcados com compostos fluorescentes.

Migração de ramificação: na recombinação, é o processo no qual uma fita de DNA, parcialmente pareada com a sua complementar em um dúplex, estende o seu pareamento deslocando a fita residente que é homóloga.

Mitose: processo no qual uma célula eucariótica separa o conjunto diploide de cromossomos do seu núcleo (previamente duplicado na fase S do ciclo celular) em dois conjuntos diploides (separados em dois núcleos) a serem segregados às duas células-filhas geradas na divisão celular mitótica.

Mitossomo: organela celular encontrada em alguns eucariotos anaeróbios, como microsporídeos (Opisthokonta) e espécies de *Entamoeba* (Amoebozoa) e *Giardia* (Excavata). As funções dos mitossomos não foram ainda totalmente elucidadas, mas, pelo menos em microsporídeos, estas organelas estão envolvidas na biossíntese de agrupamentos de ferro e enxofre para incorporação, como grupos prostéticos, em proteínas. Mitossomos, apesar de não possuírem genomas próprios, são estruturais, funcionais e evolutivamente relacionados a mitocôndrias.

Modificação tautomérica: uma isomerização reversível em uma molécula, ocorrida devido à alteração na localização de um átomo de hidrogênio. Nos ácidos nucleicos, as modificações tautoméricas nas bases dos nucleotídeos podem provocar a substituição de uma base por outra no momento da replicação, sendo, portanto, fontes de mutações.

Molécula unida: em um processo de recombinação, é a estrutura formada quando as duas moléculas de DNA de fita dupla envolvidas se unem pelas fitas que estão sendo trocadas.

Mutação com sentido trocado (*missense mutation*): mutação que, ao alterar um nucleotídeo de um códon, resulta na troca do aminoácido no produto proteico correspondente.

Mutação pontual: alteração em uma ou poucas bases em uma cadeia de DNA.

Mutação que modifica a fase de leitura (*frameshift mutations*): adição ou deleção de um ou poucos nucleotídeos, que causa a mudança da fase de leitura do mRNA no momento da tradução.

Mutação sem sentido (*nonsense mutation*): mutação que troca um códon que correspondia a um aminoácido por um códon de parada, resultando na terminação prematura da tradução do mRNA correspondente.

Mutação silenciosa (*silent mutation*): mutação que altera o códon, mas não altera o aminoácido no produto proteico correspondente devido à degeneração do código genético.

Mutação: processo que produz uma alteração no DNA ou na estrutura do cromossomo; nos genes é a fonte de novos alelos.

Mutante: uma célula ou organismo que possui uma ou mais seqüências nucleotídicas alteradas ou mutadas.

Northern blot: técnica de hibridização para detecção de moléculas de RNA específicas, na qual o RNA é separado por eletroforese em gel, transferido para um suporte sólido, por exemplo, uma membrana de náilon, e, então, hibridizado com sondas complementares marcadas.

Núcleo catalítico da DNA-polimerase III: região da holoenzima DNA-polimerase III composta pelas subunidades α , ϵ e τ , responsável pelas atividades de polimerização e correção de erro durante a replicação.

Núcleo: organela de células eucarióticas delimitada por uma membrana (carioteca) que abriga a maior parte do material genético. No núcleo, o material genético está associado a proteínas formando a cromatina, cujo grau de compactação varia conforme a fase do ciclo celular.

Nucleoide: região distinta e de forma irregular do citoplasma de uma célula procariótica ocupada pelo seu material genético (DNA). O nucleoide não está envolto por uma membrana.

Nucleossomo: partícula nucleoproteica da cromatina de eucariotos e de algumas arqueas constituída por oligômeros de histonas em torno dos quais o DNA se associa.

Oligonucleotídeo iniciador (iniciador ou *primer*): uma molécula de DNA de poucos nucleotídeos (cerca de 15-30), que serve como iniciador para a reação de polimerização de uma cadeia de DNA.

Óperon: organização estrutural típica (mas não exclusiva) de genomas procarióticos, na qual duas ou mais seqüências codificadoras de produtos gênicos (RNA ou proteínas) estão sob controle transcricional de um mesmo conjunto de seqüências reguladoras. Em um óperon, as seqüências codificadoras são transcritas em um único RNA, que, por isso, é dito policistrônico.

ORC: complexo de proteínas responsáveis pelo reconhecimento de origens de replicação em eucariotos.

Organelas: estruturas celulares especializadas, delimitadas por membranas e encontradas em células eucarióticas.

Orientação (de cadeias polinucleotídicas, genes e cadeias polipeptídicas): cadeias polinucleotídicas e polipeptídicas são assimétricas e, por isso, têm orientações definidas. Por convenção, as orientações dessas moléculas são descritas com base no sentido dos seus processos de síntese. Assim, uma molécula de RNA tem uma orientação da sua extremidade 5'-P para a sua extremidade 3'-OH, pois esse é o sentido da sua síntese. Da mesma forma, um gene é orientado de 5' para 3' com base na orientação da sua fita codificadora, que é aquela correspondente ao produto de RNA que pode ser transcrito a partir dele. Cadeias polipeptídicas, por sua vez, são orientadas da extremidade aminoterminal para a extremidade carboxi-terminal, porque este é o sentido no qual são sintetizadas.

Origem de replicação: seqüência de DNA reconhecida por proteínas específicas para dar início a um evento de replicação.

Ortólogos: são genes (ou outras sequências) e produtos gênicos (incluindo proteínas) de espécies diferentes que possuem homologia (ou ortologia) identificável com genes e produtos gênicos correspondentes de outras espécies. A ortologia decorre do fato destas espécies compartilharem uma espécie ancestral comum em suas histórias evolutivas.

Pangenoma: é o repertório gênico global de uma espécie procariótica, incluindo todos os genes encontrados em todas as cepas ou variantes reconhecidas daquela espécie. O pangenoma de uma determinada espécie inclui um subconjunto de genes considerados essenciais e presentes em todas as cepas e variantes da espécie (genoma central), e um subconjunto de genes que estão presentes em algumas cepas ou variantes, mas não em outras (genoma dispensável). O genoma dispensável é mais plástico evolutivamente, e a aquisição ou perda de genes deste componente do pangenoma ocorre essencialmente por meio de mecanismos de transferência gênica horizontal.

Parálogos: são diferentes genes (ou outras sequências) e produtos gênicos (incluindo proteínas) de uma mesma espécie que possuem homologia (ou paralogia) identificável, em virtude de terem se originado devido a um ou mais eventos de duplicação gênica (seguida de divergência).

PCR em tempo real (*real-time PCR*): um método de PCR para medir o aumento na quantidade de DNA, à medida que o DNA é amplificado. Também denominada PCR quantitativa em tempo real (qPCR).

Penalidade de lacunas: é um escore numérico usado em programas de alinhamento de sequências para penalizar a presença de lacunas dentro de um alinhamento.

Peptidoglicano: encontrado em paredes celulares de bactérias, é formado por polissacarídeos que possuem ligações cruzadas com peptídeos.

Perfil: representação matricial de uma região conservada em um alinhamento múltiplo de sequências que permite lacunas no alinhamento.

Pirossequenciamento: é uma técnica de sequenciamento de DNA com base na detecção do pirofosfato (PPi) liberado durante a síntese de DNA.

Placa de lise: área clara, circular e em uma camada de bactérias cultivadas em meio sólido, que resulta da lise de algumas células devido a repetidos ciclos de multiplicação lítica de um fago que as infecta.

Plasmídeo: molécula de DNA de fita dupla extracromossômica que replica de forma autônoma em uma célula hospedeira. Plasmídeos ocorrem naturalmente em muitas bactérias e podem ser manipulados para servirem como vetores de clonagem.

Poliadenilação: processo pelo qual uma cauda de poli(A) é adicionada à extremidade 3' de um RNA.

Poli-A polimerase: enzima responsável pela adição de poli(A) a precursores de mRNAs eucarióticos.

Policitrônico: segmento de DNA ou RNA que apresenta mais de um códon.

Polimerização: na síntese de ácidos nucleicos, é a atividade de síntese de uma cadeia polinucleotídica linear no sentido 5' para 3'.

Polimorfismo de comprimento de fragmento de restrição (RFLP, de *restriction fragment length polymorphism*): variação nos comprimentos dos fragmentos gerados pelo tratamento do DNA com uma determinada enzima de restrição em virtude de um polimorfismo genético.

Porina de membrana externa: tipo de proteína que atua como canal para a passagem de pequenas moléculas hidrofílicas. As porinas podem ser específicas, contendo sítios de ligação para um ou mais substratos, ou inespecíficas, compondo canais aquosos.

Pré-mRNA: transcrito primário ou precursor do mRNA transcrito a partir dos genes que codificam proteínas.

Pré-RNA: produto de transcrição primário de um gene interrompido (por íntrons) que ainda não sofreu processamento por *splicing*.

Primase: tipo de RNA-polimerase que sintetiza pequenos fragmentos de RNA complementares a uma fita-molde de DNA de fita simples (iniciadores) no processo de replicação.

Procariotos: grupo de organismos que inclui bactérias e arqueas. Possui seu material genético condensado no nucleóide, sem estar delimitado por uma membrana. Não possui organelas.

Processamento de RNA: mecanismos de geração de RNAs (mRNAs, tRNAs ou rRNAs) maduros a partir de seus precursores.

Processividade: período (tempo) que uma polimerase permanece ligada a uma fita-molde realizando a síntese de uma cadeia polinucleotídica complementar.

Profago: genoma de um fago lisogênico quando integrado no cromossomo da bactéria hospedeira, replicando como parte integrante do cromossomo.

Programação dinâmica: um algoritmo de programação dinâmica resolve um problema combinando soluções para os subproblemas. Essas soluções são computadas uma única vez e armazenadas em uma tabela ou matriz. A programação dinâmica é utilizada quando um problema tem muitas soluções possíveis e uma solução ótima necessita ser encontrada. Essa técnica é utilizada pelos algoritmos de alinhamentos de sequências.

Prokarya: superclado que inclui todos os procariotos (organismos celulares desprovidos de membranas internas e, por conseguinte, de organelas). São organismos procarióticos, as bactérias (domínio Bacteria) e as arqueas (domínio Archaea).

Proteassomo: é um grande complexo multiproteico envolvido na degradação de proteínas, tanto em procariotos como em eucariotos. O proteassomo pode ser dissociado em dois subcomplexos, um núcleo catalítico, que constitui o componente proteolítico e a partícula reguladora, que parece ser responsável pelo reconhecimento, o desenovelamento e a translocação de substratos poliubiquitinados para o componente catalítico, onde eles são degradados.

Proteoma: conjunto de proteínas codificado por um determinado genoma. O estudo em grande escala das proteínas de células, tecidos, órgãos ou organismos é denominado proteômica.

Pseudogenes processados: são sequências de mRNA processadas – com remoção de íntrons e adição de cauda de poli(A) – que foram reintroduzidas no genoma, em forma não funcional, por retrotransposição.

Pseudogenes: são sequências parálogas ou ortólogas a genes funcionais, mas que, ao longo da evolução, divergiram a ponto de perderem a funcionalidade.

Quebra de fita simples (SSB, *single-strand breaks*): quebra em apenas uma das fitas de uma molécula de DNA de fita dupla.

Quebras duplas no DNA (DSB, *DNA double-strand break*): quebras nas duas fitas de uma mesma molécula de DNA de fita dupla.

Quiasmas: são as fitas cruzadas de cromátides não irmãs observadas no diplóteno da primeira divisão meiótica. Considerados como a evidência citológica da troca de material cromossômico, ou *crossing-over*.

Radiação ionizante: radiação com comprimentos de onda de 0,1 – 1 nm, com alto poder de penetração e ionização de moléculas.

Radiação não ionizante: radiação com comprimento de onda de 1 a 380 nm, com baixo poder de penetração e absorvida pelas bases do DNA, principalmente as pirimídicas.

Reação em cadeia da polimerase (PCR): um método para a produção de dezenas de milhares de cópias de uma sequência específica de DNA, a partir de uma mistura de DNA sem a necessidade de clonagem da sequência em um organismo hospedeiro.

Recombinação genética: processo que conduz à formação de novas combinações de genes nos cromossomos.

Recombinação homóloga: processo que envolve uma troca recíproca de sequências de DNA, isto é, entre dois cromossomos que contêm os mesmos loci genéticos.

Recombinação sítio-específica: processo que ocorre entre duas sequências específicas, como a integração/excisão de um fago, ou na resolução de estruturas cointegradas durante a transposição.

Recombinante: um cromossomo, célula ou indivíduo que tenha combinações não parentais de marcadores genéticos, como resultado de uma recombinação genética. Termo também utilizado para vetores ou outras moléculas de DNA manipuladas *in vitro* que contenham fragmentos de DNA extras, resultantes de uma reação de ligação.

Recombinantes “remendados”: no processo de recombinação, resultam da resolução de uma junção de Holliday por clivagem das fitas permutadas. O dúplex é praticamente inalterado, exceto por uma sequência de DNA em uma das fitas, oriunda do cromossomo homólogo.

Recombinantes combinados: no processo de recombinação, provêm da resolução de uma junção de Holliday por meio de clivagem das fitas não permutadas. As duas fitas de DNA situadas antes do ponto de permuta são oriundas de um cromossomo; o DNA localizado após o ponto de permuta é originado do cromossomo homólogo.

Recombinase: nome atribuído ao grupo de enzimas que realizam recombinação sítio-específica.

Reforçador (do inglês, *enhancer*): sequência reguladora da expressão de genes eucarióticos situada distante do(s) gene(s) por ela influenciados. Um reforçador pode atuar independentemente da sua orientação e influenciar a expressão de um ou mais genes.

Reforçadores de *splicing* exônicos: sequências *cis* presentes em éxons associadas à ativação do *splicing*.

Reforçadores de *splicing* intrônicos: sequências *cis* presentes em íntrons associadas à ativação do *splicing*.

Região codificadora: porção de um gene que inclui sequências representadas em seus produtos de expressão (de RNA e/ou polipeptídico). Na maioria das situações, a região reguladora de um gene está flanqueada por regiões reguladoras da sua expressão.

Região de controle de loci (LCR): região reguladora encontrada a montante de genes, como os do grupamento gênico da β -globina, que tem função importante na regulação da expressão desses genes.

Regiões reguladoras: porções de um gene que incluem sequências responsáveis pela regulação da sua expressão. Em situações típicas, elas se encontram a montante e a jusante da região codificadora.

Renaturação proteica: reenovelamento de uma proteína desnaturada de modo a restaurar sua conformação nativa.

Reparação de genes ativos ou fita-seletivo (TC-NER, *transcription-coupled excision repair*): subcaminho da reparação por excisão de nucleotídeos que prefere a reparação das lesões que estão na fita sendo transcrita.

Reparação de mal-pareamentos (*mismatch repair*): uma forma de reparação do DNA, na qual o mecanismo é capaz de distinguir entre a fita contendo a lesão e a fita que está correta.

Reparação de quebras duplas no DNA (DSBR, *DNA double-strand break repair*): mecanismo que reúne extremidades quebradas de uma mesma cadeia de DNA.

Reparação por excisão de base (BER, *base excision repair*): sistema de reparação de bases incomuns no DNA, por meio do reconhecimento dessas bases pelas DNA-glicosilases específicas.

Reparação por excisão de nucleotídeos (NER, *nucleotide excision repair*): sistema de reparação de lesões no DNA, que utiliza a fita complementar não danificada como molde.

Reparação por recombinação: sistema de reparação que utiliza a informação genética contida na cromátide-irmã ou no cromossomo homólogo para reparar a lesão de uma cadeia de DNA, na qual ambas as fitas apresentam problemas.

Reparação sujeita a erro (*error-prone*): sistema de reparação do DNA que, ao reparar uma lesão, introduz mutações na cadeia.

Réplica-*plate*: procedimento para a transferência de colônias bacterianas ordenadas em uma placa-mãe para uma nova placa. Em geral, um cilindro de madeira, coberto com veludo estéril, é levemente pressionado sobre a superfície da placa-mãe, retirando algumas células de cada colônia, que são, então, utilizadas para inocular uma nova placa. As po-

sições relativas das colônias na placa-mãe e nas placas-filhas permanece inalterada.

Replicação semiconservativa: descreve o processo no qual uma molécula de DNA de fita dupla (parental) dá origem a duas outras moléculas de DNA de fita dupla idênticas, sendo uma das fitas parental a outra recém-sintetizada.

Replicação semidescontínua: mecanismo de síntese de novas moléculas de DNA, a partir de uma molécula parental de DNA de fita dupla, em que a síntese de uma das novas cadeias acontece de forma contínua, ao passo que a síntese da outra nova cadeia ocorre de forma descontínua.

Replicon: molécula de DNA ou segmento de uma molécula de DNA que replica a partir de um evento de replicação iniciado em uma origem de replicação.

Replissomo: conjunto de proteínas que se organizam na forquilha de replicação e atuam na síntese do DNA.

Retrotransposição: processo mediado por um retrovírus ou um retrotransposon no qual uma sequência de RNA é transcrita reversamente em uma cópia de DNA e, reintegrada no genoma.

Retrotransposon: elemento genético mobilizável por retrotransposição.

Retrovírus: vírus com genoma de RNA cuja replicação depende de transcrição reversa e integração da cópia de DNA gerada no genoma da célula hospedeira.

Ribonuclease H: é uma enzima endonucleásica não específica que degrada o RNA em moléculas híbridas RNA-DNA, como as formadas durante a transcrição reversa de moldes de RNA.

Ribonuclease: é uma enzima que degrada RNA catalisando a quebra de ligações fosfodiéster entre os ribonucleotídeos sem especificidade de sequência.

Ribozimas: RNAs que detêm atividade catalítica.

RNA não codificador: RNA que não codifica proteína, mas que pode apresentar função na regulação da expressão gênica.

RNAs não codificadores (ncRNAs): RNAs sem função codificadora de proteínas que podem ser transcritos a partir de uma ou de ambas as fitas de regiões intra ou intergênicas. São produtos de transcrição disseminada.

RNase III: ribonuclease III, enzima degradadora de RNA envolvida em uma diversidade de reações de maturação de RNAs.

RNase P: ribozima envolvida em diversos processos da maturação de tRNAs.

RNase Z: enzima envolvida em diversos processos da maturação de tRNAs.

Roentgen (r): unidade de medida de radiação ionizante, em que uma unidade r corresponde à quantidade de radiação ionizante que produz uma unidade eletrostática de carga em um cm³ de volume.

Segregação pós-meiótica: separação das fitas de uma molécula de DNA recombinante na replicação subsequente, em que uma parte da molécula continha mal-pareamentos

que não foram corrigidos. Dessa forma, cada fita dará origem a um dúplex que irá perpetuar a sua informação.

Senescência celular: eventos que conduzem à morte da célula.

SeqA: proteína sequestradora de *oriC*, que impede a remetilização da origem de replicação em *E. coli*.

Sequência consensual: quando é feito um alinhamento para a comparação de múltiplas sequências homólogas de ácidos nucleicos ou proteínas, o resultado pode ser representado na forma de uma sequência consensual. Na sequência consensual, o nucleotídeo ou o aminoácido em cada uma de suas posições corresponde àquele mais frequentemente encontrado em uma determinada posição nas sequências que foram alinhadas.

Sequenciamento com didesoxirribonucleotídeos: método para o sequenciamento de moléculas de DNA utilizando ddNTPs como terminadores de cadeia. O ddNTP terminal corresponde ao nucleotídeo normal, que estaria presente no ponto em que a cadeia foi precocemente terminada.

Sequências GATC: sítios nos quais a adenina é passível de metilação nas duas fitas do DNA. Essas sequências estão presentes no genoma de *E. coli* e espécies relacionadas.

Sequências homopoliméricas: sequências constituídas de um único tipo de nucleotídeo, por exemplo, poli(A) ou poli(T).

Sequências palindrômicas: sequências de nucleotídeos que se encontram duplicadas e invertidas.

Sexo heterogamético: sexo que produz gametas contendo cromossomos sexuais diferentes; nos mamíferos o sexo heterogamético é o masculino.

Sexo homogamético: sexo que produz gametas que não diferem quanto ao cromossomo sexual; nos mamíferos o sexo homogamético é o feminino.

Silenciadores de *splicing* exônicos: sequências *cis* presentes em éxons associadas à inibição do *splicing*.

Silenciadores de *splicing* intrônicos: sequências *cis* presentes em íntrons associadas à inibição do *splicing*.

Sinapse (cromossômica): pareamento dos cromossomos homólogos na meiose.

Sintenia: conservação da ordem dos genes ao longo do(s) cromossomo(s) quando os genomas de duas espécies diferentes são comparadas.

Síntese translesão (ou síntese *by pass*): síntese de DNA por sobre uma lesão na fita-molde.

Sistema de reparação global (GG-NER): subcaminho da reparação por excisão de nucleotídeos que repara lesões em toda a extensão do genoma de uma célula.

Sítio de restrição: sequência no DNA reconhecida é passível de clivagem por uma enzima de restrição. Muitas enzimas de restrição clivam ambas as fitas de um DNA dentro do sítio de restrição, criando extremidades coesivas ou cegas, dependendo da posição das clivagens.

Sítio frio (*cold spot*): local do DNA com baixa probabilidade de ocorrer um evento mutacional.

Sítio múltiplo de clonagem: região de um vetor de clonagem que contém muitos sítios únicos para diversas enzimas de restrição, também denominado *polylinker*.

Sítio quente (*hot spot*): local no DNA altamente sujeito a um evento mutacional.

Sítios *cos*: segmento de 12 nucleotídeos de fita simples presentes em ambas as extremidades do genoma do bacteriófago λ . Estes dois sítios *cos* apresentam complementaridade em suas sequências, permitindo a recircularização do genoma linear do fago.

snoRNA: RNAs nucleolares pequenos, envolvidos na maturação de rRNAs.

snRNPs: complexos ribonucleoproteicos que compõem o spliceossomo.

Sobrecruzamento: Ver *crossing-over*.

Sonda de hibridização: uma molécula de ácido nucleico marcada que pode ser utilizada para identificar moléculas complementares ou homólogas, pela formação de moléculas híbridas por pareamento de bases estável.

Southern blot: técnica para a detecção de fragmentos de DNA específicos, na qual os fragmentos são separados por eletroforese em gel, transferidos para um suporte sólido, por exemplo, uma membrana de náilon e, então, hibridizados com sondas complementares marcadas radioativamente ou não.

Spliceossomo: um grande complexo constituído de proteínas e RNA, responsável pelo *splicing* dos pré-mRNAs.

Splicing: mecanismo de processamento caracterizado pela excisão de íntrons e ligação de éxons.

Splicing alternativo: mecanismo de *splicing* caracterizado pela excisão seletiva de íntrons.

Splicing constitutivo: mecanismo de *splicing* caracterizado pela excisão de todos os íntrons de um precursor de mRNA.

SSB: proteínas que, em processos como o de replicação, se ligam ao DNA de fita simples impedindo o seu pareamento com a fita complementar ou evitando a sua degradação.

Tecnologia do DNA recombinante: coleção de procedimentos experimentais para a inserção de um fragmento de DNA de um organismo em um DNA de outro organismo, e para a clonagem de novas moléculas de DNA recombinantes.

Telomerase: enzima ribonucleoproteica responsável pela síntese de telômeros em cromossomos eucarióticos. A telomerase adiciona segmentos de DNA repetitivo na extremidade de uma das fitas do DNA, utilizando um segmento de RNA (parte da própria enzima) como molde.

Telômeros: extremidades dos cromossomos de eucariotos.

Topologia: é a área da ciência que estuda as propriedades que são preservadas quando se deforma um objeto sem rompê-lo. Ela é, por exemplo, aplicada ao DNA para descrever o superenrolamento.

Trans-splicing: forma de *splicing* na qual éxons provenientes de dois transcritos primários de RNA distintos são unidos.

Transcrição disseminada (do inglês, *pervasive transcription*): fenômeno de transcrição de sequências não codificadoras de genes e de regiões intergênicas que acontece em eucariotos. A transcrição disseminada gera diversas famílias de RNAs não codificadores.

Transcriptase reversa: enzimas com função de DNA-polimerase RNA-dependente, que são codificadas por genes de retrovírus ou de retrotransposons. O processo de síntese mediado por transcriptases reversas é o inverso do processo de transcrição normal, no qual RNA é sintetizado a partir de um molde de DNA.

Transcritoma: o conjunto completo de transcritos de RNA produzido por um genoma em qualquer momento. O transcritoma é dinâmico e muda sob diferentes circunstâncias, devido aos diferentes padrões de expressão gênica. O estudo do transcritoma é denominado transcritômica.

Transdução bacteriana: transferência de pequenos segmentos de DNA do cromossomo bacteriano de uma célula para outra, mediada por bacteriófagos.

Transfecção: introdução de ácidos nucleicos em células eucarióticas utilizando vetor viral ou plasmídeo.

Transferência gênica endossimbiontista: é a transferência horizontal e a fixação de genes de um endossimbionte no genoma nuclear de sua célula hospedeira. É também chamada de transferência gênica intracelular.

Transferência gênica horizontal: é a transferência de genes de um organismo para o genoma de um outro organismo não pertencente a mesma linhagem celular. É também chamada de transferência gênica lateral.

Transformação bacteriana: processo no qual nova informação genética é adquirida por uma bactéria devido à incorporação de fragmentos de DNA extracelulares.

Transição: a substituição de uma purina por outra purina ou de uma pirimidina por outra pirimidina.

Transposição: evento de mobilização de um elemento transponível; as transposições são mediadas por enzimas como a transposase ou por um conjunto de enzimas como uma transcriptase reversa e uma integrase, as quais são codificadas pelo próprio elemento ou por um elemento móvel relacionado.

Transposon: elemento genético mobilizável por transposição.

Transversão: a substituição de uma purina por uma pirimidina ou vice-versa.

União de extremidades mediada por micro-homologias (MMEJ, *microhomology-mediated end joining*): mecanismo de junção de extremidades de uma mesma cadeia de DNA, por remoção ou adição de 5 a 25 nucleotídeos.

União de extremidades não homólogas (NHEJ, *non-homologous end joining*): mecanismo que repara quebras de fita dupla no DNA. NAEJ é referido como “não homólogo” porque as extremidades quebradas são ligadas diretamente, sem a necessidade de uma molécula de DNA homóloga. Pode ocorrer a remoção ou a adição de 1 a 5 nucleotídeos no momento da ligação.

Variação de fase: processo de adaptação apresentado por muitas bactérias que está relacionado com a variação da expressão de proteínas por meio de um mecanismo de repressão e desrepressão de genes específicos.

Velocidade de sedimentação: é uma medida da massa e da forma de moléculas. É determinada, de forma experimental, pela medida do coeficiente de sedimentação em experimentos de centrifugação. Normalmente é expressa em unidades Svedberg (1 U-S equivale a 10^{-5} segundos) e pode ser utilizada na caracterização de macromoléculas. Por exemplo, o rRNA de 18S tem um coeficiente de sedimentação de 18 Svedbergs.

Vetor de clonagem (ou veículo de clonagem): uma molécula de DNA de fita dupla capaz de replicação autônoma em uma célula hospedeira e na qual um ou mais fragmentos de DNA podem ser inseridos para formar uma molécula de DNA recombinante para clonagem molecular.

Vetor de expressão: vetor de clonagem que permite a expressão de uma sequência codificadora de proteína nele

clonada em uma determinada célula ou organismo hospedeiro.

Vírião: é uma partícula viral completa, constituída pelo material genético viral (genoma de DNA ou RNA) e pelo capsídeo proteico que o envolve. É a forma infectiva de um vírus.

Xenobióticos: são compostos químicos estranhos a um organismo ou sistema biológico. Antibióticos e outras drogas são exemplos de xenobióticos.

Xenólogos: são genes ortólogos que foram adquiridos independentemente, por cada espécie por meio de eventos de transferência gênica horizontal.

Zíper de leucina: domínio proteico encontrado em alguns fatores de transcrição. Cada zíper de leucina é formado por duas α -hélices, e atua na dimerização da proteína que o contém. Na dimerização, as hélices dos zíperes de leucina de dois monômeros proteicos associam-se entre si por interações hidrofóbicas entre os resíduos de leucina localizados em cada lado de cada hélice.

Índice

A

Acetilação e dasacetilação, 306
Ácido
 desoxirribonucleio, 14
 graxo, 10-12
 nitroso, 141-142
 nucleico, 13-14
 ribonucleico, 15
Ácidos nucleicos, estrutura do, 17-36
 DNA, 18-31
 composição química, 18-19
 hélice dupla, 19-25
 estrutura, 19-23
 propriedades químicas, 23-25
 outras estruturas, 31
 curvatura, 31
 estruturas cruciformes, 31
 junções de Holliday, 31
 supertorção, 25-27
 tipos, 29-31
 topoisomerases, 27-29
 RNA, 31-35
 classes, 31-35
 composição química, 32
 estrutura secundária, 32
ACT, 372
Agentes alquilantes, 141-142
Agrobacterium tumefaciens, 348-350
Alças de cromatina e fibra de 300 nm, 53-54
Alongamento da cadeia
 de RNA, 222-226
 polipeptídica, 268-273
Aminoácidos e proteínas, 4-8
Análogos de bases, 140-141

Anelamento de fitas simples, 158-161
Apêndices, formação dos, 319-329
Átomos, representação gráfica de, 35

B

Bactérias
 arqueas, 208-209
 mobilidade plasmidial, 188-190
Bacteriófago, 191-193
 λ como vetor de clonagem, 338
Banco de dados
 biológicos, 372-380
 de estruturas de proteínas, 378-379
 de genômica comparativa, 379-380
 funcionais e de metabolismo, 379-380
 primários, 373-377
 secundários, 377-380
Bases alquiladas, reparação enzimática, 148
Bibliotecas de DNA, 341-345
 bibliotecas de cDNA, construção de, 342-344
 bibliotecas genômicas, construção de, 342
 clones em bibliotecas, seleção de, 344-345
Bioinformática na análise de genes e genomas, 363-382
 alinhamentos de sequências, 365-370
 alinhamentos múltiplos, 370
 pares de sequências, 365-369
 BLAST, 368-369
 Dot Plot, 368
 FastA, 368
 bancos de dados biológicos, 372-380
 banco de dados de estruturas de proteínas, 378-379
 CATH, 378-379
 PDB, 379
 SCOP, 378

- banco de dados funcionais e de metabolismo, 379-380
 - COG, 379-380
 - KEGG, 380
 - banco de dados primários, 373-377
 - DNA Data Bank of Japan, 377
 - EMBL Nucleotide Sequence Database, 376-377
 - GenBank, 373-376
 - UniProt Knowledgebase, 377
 - banco de dados secundários, 377-378
 - InterPro, 378
 - Pfam, 378
 - PINTS, 377-378
 - PROSITE, 377
 - dados de genômica comparativa, 379-380
 - HOGENOM, 379
 - MBGD, 379
 - MicrobesOnline, 379
 - xBase, 379
 - genomas, 370-372
 - análise comparativa de genomas, 371-372
 - ACT, 372
 - M-GCAT, 372
 - MUMmer, 372
 - PipMaker, 372
 - anotação de genomas, 371
 - montagem de genomas, 370-371
 - redes de informações, 364-365
 - HTTP, HTML, URLs, 365
 - Internet, 364-365
 - rede mundial de comunicação, 365
 - Biologia molecular, técnicas de
 - análise de DNA e RNA, 350-357
 - eletroforese, 350
 - hibridização, 351
 - reação em cadeia da polimerase, 353-357
 - aplicações da técnica de PCR, 353-357
 - princípios dos métodos, 355
 - sequenciamento de DNA, 353
 - método de Sanger, 353
 - sequenciamento automático de DNA, 353
 - southern blot, 351-353
 - análise de genoma e produtos de transcrição em grande escala, 357-359
 - análise global de transcrição (transcritômica), 357-359
 - etiquetas de seqüências expressas, 357-358
 - macroarranjos e microarranjos, 358-359
 - sequenciamento em grande escala, 359
 - sequenciamento de genomas, 357
 - análise global de proteínas, 359-361
 - espectrometria de massas, 361
 - fracionamento de proteínas associadas à proteômica, 359
 - cromatografia líquida, 361
 - eletrofase bidimensional, 359-361
 - eletrofase em gel de poliacrilamida, 359
 - bibliotecas de DNA, 341-345
 - bibliotecas genômicas, construção de, 342
 - bibliotecas de cDNA, construção de, 342-344
 - clones em bibliotecas, seleção de, 344-345
 - hibridização de clones recombinantes com sondas de DNA complementares, 344
 - seleção imunológica, 345
 - clonagem molecular, 336-341
 - clonagem de DNA, vetores de, 336-339
 - bacteriófago λ como vetor de clonagem, 338
 - cosmídeos, 338
 - cromossomos artificiais bacterianos, 338-339
 - cromossomos artificiais de levedura, 339
 - plasmídeos, 336-338
 - moléculas de DNA recombinantes, 336
 - moléculas de DNA recombinantes em células bacterianas, 339-341
 - recombinantes, identificação dos, 341
 - transfecção com DNA de fagos, 340
 - transformação bacteriana, 340
 - transformantes, seleção de, 340-371
 - enzimas, 333-336
 - DNA-polimerase, 335-336
 - DNA-polimerase I, 335
 - DNA-polimerases termoestáveis, 335-336
 - fragmento de Klenow, 335
 - transcriptases reversas, 336
 - DNA-ligase, 333
 - enzimas de modificação do DNA, 333-335
 - desoxinucleotidil-terminal-transferase, 335
 - polinucleotídeo-quinase, 335
 - enzimas de restrição, 333
 - expressão de seqüências clonadas, 345-347
 - produção e purificação de proteínas recombinadas, 347
 - fusão por cromatografia de afinidade, 347
 - vetores de expressão eucarióticos, 345-346
 - vetores de expressão procarióticos, 345
 - técnicas, 331-362
 - transformação genética de plantas, 348-350
 - Biologia molecular do desenvolvimento, 319-330
 - caixa de ferramentas, 320-327
 - ovos e diferenças nos padrões de clivagem *versus* conservação dos genes da caixa de ferramenta, 323-327
 - genes maternos, 324-325
 - genes embrionários (ou zigotos), 325-327
 - genes da família PAX, 327-328
 - genes *otd/Otx* e *ems/Emx* e a formação das partes do cérebro, 328-329
 - genes *DLL/Dlx* e a formação dos apêndices, 328-329
 - genes *tinman* e *NK2* e a formação do coração, 328-329
 - BLAST, 368-369
 - Bombardeamento de DNA, 350
- ## C
- Cadeia
 - de açúcar-fosfato, 23
 - de RNA, alongamento de, 222-226
 - polipeptídica, alongamento da, 268-273
 - Caixa de ferramentas, 320-327
 - Carboidratos, 9-10
 - CATH, 378-379
 - Célula
 - constituente moleculares, 1-15
 - ácido
 - desoxirribonucleio, 14
 - graxo, 10-12

- nucleico, 13-14
 - ribonucleico, 15
 - aminoácidos e proteínas, 4-8
 - carboidratos, 9-10
 - esteroides, 12-13
 - estrutura tridimensional da proteína, 8-9
 - estruturas celulares, 2-4
 - célula de procariotos e eucariotos, 3-4
 - organelas, 4
 - fosfoacilgliceróis, 12
 - lipídeos, 10
 - procariotos e eucariotos, 3-4
 - bacteriana, introdução de
 - moléculas de DNA recombinantes, 339-341
 - Cérebro, formação das partes do, 328-329
 - Ciclo celular e origem da replicação, 124-125
 - Círculo rolante, sistemas de replicação por, 130-131
 - Clonagem
 - bacteriófago λ como vetor, 338
 - de DNA, vetores de, 336-339
 - molecular, 336-341
 - Clones
 - em bibliotecas, seleção de, 344-345
 - recombinantes com sondas de DNA complementares, hibridização de, 344
 - Código genético e síntese de proteínas, 255-276
 - código genético, 256-258
 - ribossomos, 263-265
 - composição e estrutura, 263-265
 - sítios ativos, estrutura e dinâmica dos, 265
 - RNAs transportadores, 259-262
 - ativação dos tRNAs e as aminoacil-tRNA-sintetases, 261-262
 - estrutura do tRNAs, 259-261
 - tRNAs supressores, 262-263
 - síntese de proteínas, 265-273
 - alongamento da cadeia polipeptídica, 268-273
 - peptidil-transferase, decodificação e atividade de, 269-271
 - translocação, 271-273
 - cadeia peptídica do ribossomo, saída da, 273
 - início, 266-267
 - procariotos, 267
 - eucariotos, 267-268
 - término da síntese de proteínas, 273
 - COG, 379-380
 - Complexo
 - 43S ao mRNA, 267
 - 48S, 267
 - com múltiplos componentes, formação de, 305-306
 - de pré-iniciação 43S, 267
 - iniciador 80S, 268
 - Composição química, 18-19
 - Conteúdo de G+C, 72-74
 - Conversão gênica, 170
 - Coração, formação do, 328-329
 - Corantes de acridina, 141-142
 - Cosmídeos, 338
 - Cromatina
 - de ordem superior e cromossomos metafásicos, 54-55
 - de organelas, 54-56
 - nucleoides de plastídeos, 55-56
 - eucariótica, 46-54
 - composição molecular, 46-48
 - contexto da, 119-120
 - níveis de organização mais complexos, 48-54
 - alças de cromatina e fibra de 300 nm, 53-54
 - estrutura de cromatina de ordem superior e cromossomos metafásicos, 54-55
 - fibra de 10 nm, 50-52
 - fibra de 30 nm, 52-53
 - nucleossomos eucarióticos, 48
 - nucleoides mitocondriais, 55
 - procariótica, 38-46
 - nucleoide e cromatina bacteriana, 38-42
 - nucleoide e cromatina de arqueas, 42-46
 - transcrição, 228-259
 - Cromatina, 37-56
 - procariótica, 38-46
 - Cromatografia
 - de afinidade, fusão por, 347
 - líquida, 361
 - Cromossomos
 - artificiais bacterianos, 338-339
 - artificiais de levedura, 339
 - metafásicos, 54-55
 - Curvatura, 31
- D**
- Dasacetilação, 306
 - Dedo de zinco, 304-305
 - Defesa bacteriana de vírus ou plasmídeos (CRISPR), 298-299
 - Densidade gênica, 64-65
 - Depurinação
 - lesões hidrolíticas, 137
 - Desaminação
 - lesões hidrolíticas, 137
 - Desmetilação, 307
 - Desoxinucleotidil-terminal-transferase, 335
 - Desubiquitinação, 307
 - Desvios de GC, 75-76
 - Dinâmica evolutiva
 - genomas procarióticos, 79-82
 - Distrofias musculares, 145
 - DNA, 18-31
 - análise (biologia molecular), 350-357
 - análise de poliformismos, 355-357
 - bibliotecas, 341-345
 - bombardamento (biologia molecular), 350
 - Data Bank of Japan*, 377
 - de fagos, transfecção com, 340
 - e RNA, análise de, 350-357
 - sequenciamento, 353
 - sequenciamento automático, 353
 - DNA-polimerase, 113-115, 335-336
 - estrutura, 117-119
 - Doença
 - das células falciformes, 144
 - infeciosa, diagnóstico de, 357
 - Domínios
 - de ativação, 305
 - e macrodomínios, 70-71
 - Dot Plot, 368

E

- E. coli*, 281-286
- Efeitos hidrofóbicos, 22
- Elementos
- de transposição, 193-202
 - características gerais e mecanismos de transposição, 193-196
 - elementos de DNA ou transposons, 196-197
 - retrotransposons, 197-201
 - transposição, consequências biológicas da, 201-202
 - genéticos móveis, 185-203
 - bacteriófagos, 191-193
 - estrutura e classificação, 191
 - mobilização de material genético bacteriano mediada por fagos, 191-193
 - plasmídeos, 186-190
 - estrutura e classificação, 186-188
 - mobilidade plasmidial em bactérias, 188-190
 - transponíveis
 - genomas nucleares eucarióticos, 101-103
- Eletofase
- bidimensional, 359-361
 - biologia molecular, 350
 - em gel de poliacrilamida, 359
- Eletroporação, 350
- Empilhamento das bases, 22
- Enzimas, 333-336
- de modificação do DNA, 333-335
 - DNA-ligase, 333
 - enzimas de restrição, 333
 - que atuam na recombinação homóloga em procariotos, 170-175
 - recombinação homóloga, 176-179
- Erros, fidelidade e correção de, 230
- Escherichia coli* como modelo na replicação, 125-126
- Espectometria de massas, 361
- Esteroides, 12-13
- Estruturas
- celulares, 2-4
 - cruciformes, 31
- Etiquetas de seqüências expressas, 357-358
- Eucariotos
- controle da expressão genética, 301-318
 - promotores na transcrição, 213-214
 - recombinação homóloga, 175-179
 - terminação da transcrição, 227-228
 - transcrição, início da, 217-222
- Excisão
- reparação por, 148-153
- Expressão de locus multigênicos, 309-312
- Expressão de seqüências clonadas, 345-347
- Expressão eucarióticos, vetores de, 345-346
- Expressão genética em eucariotos, controle da, 301-318
- complexos com múltiplos componentes, formação de, 305-306
 - cromatina, remodelamento da, 307-308
 - fatores de transcrição, 303-305
 - ativadores e repressores da transcrição, 303-304
 - motivos presentes em proteínas que se ligam ao DNA, 304-305
 - dedo de zinco, 304-305
 - domínio s de ativação, 305
 - hélice-alça-hélice, 305
 - homeodomínio, 304
 - zíper de leucina, 305
- genes de globina como modelo, 309-312
- histonas, modificações das, 306-307
- acetilação e desacetilação, 306
 - metilação e desmetilação, 307
 - ubiquitinação e desubiquitinação, 307
- pela localização celular do mRNA, 317
- receptores nucleares de hormônio, 308-309
- receptores de hormônios esteroides, 309
- regulação da expressão no nível de tradução, 315-317
- regulação do início da transcrição, 302-303
- regulação dos genes, 302
- regulação pós-transcricional, 312-315
- pequenos RNAs na regulação da expressão gênica, 312-315
 - miRNAs, 313-312
 - piRNAs, 315
 - siRNAs, 315
 - turnover de mRNA, 315
- RNAS não codificados no controle da transcrição, 312
- Expressão genética em procariotos, controle da, 277-300
- mecanismos de controle pós-transcricional, 290-293
 - proteínas como moléculas reguladoras, 292-293
 - mecanismos regulatórios, integração de, 289-290
 - expressão gênica no bacteriófago lambda, 290
 - genoma do bacteriófago lambda, 289-290
 - outros mecanismos de controle, 297-299
 - defesa bacteriana de vírus ou plasmídeos (CRISPR), 298-299
 - variação da fase e controle, 297-291
- RNAS como moléculas reguladoras, 293-296
- pequenos RNAs (sRNAs), 294-296
 - sRNAs e o sequestro de proteínas reguladoras, 295-296
 - sRNAs: controles positivo e negativo da tradução, 296
 - ribocomutadores, 293-294
- término da transcrição, controle do, 286-288
- transcrição por antiterminador, controle da, 288
- transcrição, controle do início da, 278-286
- mecanismos de controle transcricional, 278-280
 - reguladores gerais, 278-279
 - tipos, 279-280
 - exercido por cascata de fatores sigma, 280-281
- óperon como unidade transcricional e de controle de expressão gênica, 281-286
- óperon da arabinose de *E. coli*, 285-288
 - óperon *lac* de *E. coli*, 281-285
 - transcrição do óperon *trp* de *E. Coli*, controle negativo da, 285
 - por atenuação do óperon *trp* de *E. coli*, 287-288
- Expressão gênica, regulação da, 182
- Expressão procarióticos, vetores de, 345
- Extremidade 3' e poliadenilação de mRNAs nucleares, 240-241

F

- Fábricas de transcrição, 230
- Fagos, material genético bacteriano mediado por, 191-193

- Família
de parálogos, 95-99
multigênica eucariótica, 89-90
- FastA, 368
- Fibra
de 10 nm, 50-52
de 30 nm, 52-53
de 300 nm, 53-54
- Fidelidade e correção de erros, 230
- Fitas
contínuas e descontínuas, 115-117
simples, 158-161
- Forquilha de replicação colapsadas (recombinação homóloga), 167
- Fosfoacilgliceróis, 12
- Fotorreativação enzimática, 147-148
- Fragmento de Klenow, 335
- Frequência
mutação do DNA, 140
- Fusão por cromatografia de afinidade, 347
- G**
- GenBank, 373-376
- Genes
agrupamentos organizados, 105
coexpressos, agrupamentos de, 105
da família *PAX*, 327-328
de globina como modelo, 309-312
DLL/Dlx e a formação dos apêndices, 328-329
e genomas (bioinformática), 363-382
embrionários (ou zigotos), 325-327
funcionalmente relacionados, agrupamentos de, 104-105
interrompidos, estrutura básica e ocorrências de, 86-88
interrompidos, evolução dos, 106-107
maternos, 324-325
números e densidade gênica, 94-95
otd/Otx e *ems/Emx* e a formação das partes do cérebro, 328-329
tinman e *NK2* e a formação do coração, 328-329
únicos, 95-99
- Genes e genomas eucarióticos, 85-110
genes eucarióticos, 86-90
domínios transcricionais complexos, 88-89
famílias multigênicas eucarióticas, 89-90
genes interrompidos, estrutura básica e ocorrências de, 86-88
tamanho, 88
genomas de organelas, 107-109
genomas de plastídeos, 109
genomas mitocondriais, 107-109
genomas nucleares eucarióticos, 90-105
composição de sequências, 92-103
dinâmica evolutiva de, 105-107
evolução dos genomas eucarióticos, 105-106
genes interrompidos, evolução dos, 106-107
forma, número e organização em replicons, 92
sintenia, 103-104
agrupamentos organizados de genes, 105
óperons eucarióticos, 105
agrupamentos de genes funcionalmente relacionados, 104-105
agrupamentos de genes coexpressos, 105
tamanho, 90
- Genes e genomas procarióticos, 57-84
genes procarióticos, 58-61
estrutura básica, 58-60
homologia entre genes procarióticos, 60-61
tamanho, 60
genomas procarióticos, 62-79
conteúdo de G+C, 72-74
densidade gênica, 64-65
dinâmica evolutiva, 79-82
principais processos evolutivos, 79-82
processos inovadores, 79-81
processos que determinam perda de sequências, 81-82
principais tendências evolutivas, 82
distribuição dos genes e de outras sequências, 74
desvios de GC, 75-76
localização, 74
orientação, 74
replicação, 74
sintenia, 76-77
superóperons e ultraóperons, 77-79
forma, número e organização em replicons, 63-64
tamanho, 62-63
unidades organizacionais, 65-72
domínios e macrodomínios, 70-71
genes, 67-69
ilhas genômicas, 71-72
motivos, 65-66
óperons, 69-70
repetições, 66-67
- Genética de plantas, 348-350
- Genomas
ampliação de regiões específicas, 355
análise, 370-372
bioinformática, 370-372
circulares replicação do DNA, 126-127
de organelas, 107-109
de plastídeos, 109
do bacteriófago lambda, 289-290
e produtos de transcrição em grande escala (técnica de análise), 357-359
eucarióticos
evolução, 105-106
mitocondriais, 107-109
nucleares eucarióticos, 90-105
dinâmica evolutiva, 105-107
procarióticos, 57-84
- Grupos sanguíneos ABO, 145
- H**
- Hélice dupla, 19-25
Hélice-alça-hélice, 305
- Hibridização
biologia molecular, 351
- HOGENOM, 379
- Homeodomínio, 304
- Hormônio
receptores nucleares, 308-309
esteroide, receptores de, 309
- HTML, 365
HTTP, 365

I

Ilhas genômicas, 71-72
 Internet, 364-365
 InterPro, 378
 Íntrons
 do grupo I, 250
 do grupo II, 249-250
 do grupo III, 250
 e *splicing* de pré-mRNA, 241-249
 íntrons spliceossômicos, 241-243
 spliceossomo e *splicing*, 243-245

J

Junções de Holliday, 31

K

KEGG, 380

L

Lacunas por recombinação, reparação para o preenchimento de, 170
 Lesões
 hidrolíticas, 137-140
 oxidativas, 140
 reversão direta
 reparação do DNA, 147-148
 Lipídeos, 10
 Localização celular do mRNA, 317
 Locus multigênicos, expressão de, 309-312

M

Macroarranjos e microarranjos, 358-359
 Macrodônios, 70-71
 Material genético bacteriano mediada por fagos, 191-193
 MGD, 379
 Mecanismos regulatórios, integração de, 289-290
 Medida forense, 357
 Meiose
 segregação correta dos cromossomos, 175-176
 Metilação e desmetilação, 307
 Método de Sanger, 353
 M-GCAT, 372
 Microarranjos, 358-359
 MicrobesOnline, 379
 miRNAs, 313-312
 Mobilidade plasmidial em bactérias, 188-190
 Modelo de Holliday
 recombinação homóloga, 165-166
 Modificações tautoméricas, 136-137
 Moléculas
 de DNA recombinantes, 336
 de DNA recombinantes em células bacterianas, 339-341
 reguladoras, proteínas como, 292-293
 reguladoras, RNAs como, 293-296
 representação gráfica, 35
 Motivos, 65-66
 mRNA
 localização celular do, 317
turnover, 315

mRNAs

nucleares, 239-241
 precursores, 238-249
 MUMmer, 372
 Mutação e reparação do DNA, 133-162
 mutações, 134-146
 aplicações práticas, 145-146
 classificação, 134-136
 por efeitos fenotípicos, 135-136
 por localização, 134-135
 por tipo de alteração molecular, 135
 espontâneas, 136-140
 erros durante a replicação do DNA, 136-137
 lesões hidrolíticas, 137-140
 lesões oxidativas, 140
 frequência, 140
 induzidas, 140-143
 por radiações, 142-143
 quimicamente, 140-142
 natureza, 134
 no homem, 144-145
 distrofias musculares, 145
 doença das células falciformes, 144
 grupos sanguíneos ABO, 145
 reparação, 146-161

N

Nível de tradução, regulação da expressão no, 315-317
 Nucleoides
 e cromatina bacteriana, 38-42
 e cromatina de arqueas, 42-46
 de plastídeos, 55-56
 mitocondriais, 55
 Nucleossomos eucarióticos, 48

O

Óperon, 69-70
 da arabinose de *E. coli*, 285-286
 eucariótico
 genomas nucleares eucarióticos, 104
lac
 controle negativo da transcrição, 281
 controle positivo da transcrição, 285
lac de *E. coli*, 281-285
 Organelas, 4
 Ovos e diferenças nos padrões de clivagem *versus*
 conservação dos genes da caixa de ferramenta, 323-327

P

Pares de sequências, alinhamento de, 365-369
 PCR em tempo real quantitativa (qPCR), 355
 PDB, 379
 Peptidil-transferase, decodificação e atividade de, 269-271
 Pequenos RNAs, 294-296
 na regulação da expressão gênica, 312-315
 Pfam, 378
 PINTS, 377-378
 PipMaker, 372
 piRNAs, 315
 Plantas, genética de, 348-350

- Plasmídeos, 186-190, 336-338
- Poliformismos de DNA, análise de, 355-357
- Polimerase, reação em cadeia da, 353-357
- Polinucleotídeo-quinase, 335
- Precursos de mRNAs, 238-249
- Procariotos
 - controle da expressão genética, 277-300
 - promotores na transcrição, 211-213
 - síntese de proteínas, 267
 - terminação da transcrição, 226-227
 - transcrição, início da, 214-217
- Processamento e maturação de RNA, 233-254
 - degradação, 252-253
 - mecanismos de, 244-245
 - transcritos primários de rRNAs em eucariotos, 245-246
 - transcritos primários de rRNAs em procariotos, 244-245
 - mecanismos de auto-*splicing*, 248-251
 - íntrons do grupo I, 248
 - íntrons do grupo II, 249-250
 - íntrons do grupo III, 250
 - mobilidade dos íntrons, 250
 - mecanismos de edição, 250-253
 - precursores de mRNAs, 234-249
 - adição de 5'-*cap* em hnRNAs, 234-236
 - poliadenilação de hnRNAs, 236-237
 - splicing* em hnRNAs, 237-240
 - splicing* alternativo, 240-243
 - trans-*splicing*, 243-244
 - rRNAs, transcritos primários de, 244
- Processamento de tRNAs, 246-248
 - clivados e adição de 3' CCA, geração de, 246
 - modificações de bases dos tRNAs, 246-247
 - splicing* em tRNAs, 247-248
 - tRNAs, 246-248
- Processos evolutivos
 - genomas procarióticos, 79-82
- Promotor simples de eucariotos
 - modelo esquemático de organização, 215
- PROSITE, 377
- Proteína, 4-8
 - análise global, 359-361
 - como moléculas reguladoras, 292-293
 - estrutura tridimensional, 8-9
 - RecA, 171-172
 - RecBCD, 172-173
 - RecFOR, 173
 - recombinada, produção e purificação de, 347
 - replicação do DNA, 113-119
 - transcrição, 207-211
- Proteômica, 359
- Pseudogenes, 95-99
- Q**
- Quebras duplas
 - no DNA, reparação de, 156-161
 - recombinação homóloga, 167-169
- R**
- Radiação
 - mutação induzida, 142-143
 - radiação ionizante, 142
 - radiação ultravioleta, 142-143
- Reação em cadeia da polimerase, 353-357
- Receptores
 - de hormônios esteroides, 309
 - nucleares de hormônio, 308-309
- Recombinação genética, 163-184
 - recombinação homóloga
 - associações com a reparação do DNA, 166-170
 - forquilhas de replicação colapsadas, 167
 - lacunas por recombinação, 170
 - quebras duplas pela combinação homóloga, 167-169
 - conversão gênica, 170
 - em eucariotos, 175-179
 - enzimas, 176-179
 - segregação correta dos cromossomos durante a meiose, 175-176
 - enzimas que atuam na recombinação homóloga em procariotos, 170-175
 - outras proteínas, 173-175
 - proteína RecA, 171-172
 - proteína RecBCD, 172-173
 - proteína RecFOR, 173
 - modelo de Holliday, 165-166
 - recombinação homóloga, 164-179
 - recombinação sítio-específica, 179-182
 - integração e excisão do bacteriófago λ , 181-182
 - mecanismos, 180-181
 - troca de fitas, 181
 - regulação da expressão gênica, 182
 - Recombinação homóloga, 156-158, 164-179
 - em procariotos (enzimas), 170-175
 - Recombinação sítio-específica, 179-182
- Rede mundial de comunicação, 365
- Redes de informações, 364-365
- Regulação
 - da expressão gênica, 182
 - dos genes, 302
 - pós-transcricional
 - expressão genética em eucariotos, controle da, 312-315
- Reparação do DNA, 146-161
 - reparação
 - de quebras duplas no DNA, 156-161
 - anelamento de fitas simples, 158-161
 - por recombinação homóloga, 156-158
 - união de extremidades não homólogas, 158
 - união de extremidades mediada por micro-homologias, 158
 - reparação por excisão, 148-153
 - excisão de base, 148-149
 - excisão de nucleotídeos, 149-143
 - reversão direta de lesão, 147-148
 - fotorreativação enzimática, 147-148
 - reparação enzimática de bases alquiladas, 148
 - revisão de leitura e reparação de mal pareamentos, 146-147
 - sistema de reparação pós-replicação, 153-156
 - reparação por recombinação, 153-154
 - reparação sujeita a erro, 155-156
 - sistema SOS, 154-155
 - Reparação
 - enzimática de bases alquiladas, 148
 - por excisão
 - reparação do DNA, 148-153
 - pós-replicação, sistema de 153-156

- Repetições, 66-67
- Replicação do DNA, 111-132
- cromatina eucariótica, contexto da, 119-120
 - dinâmica: *Escherichia coli* como modelo, 125-126
 - erros durante a replicação, 136-137
 - origem, 120-125
 - caracterização, 120-124
 - ciclo celular e origem da replicação, 124-125
 - proteínas envolvidas, 113-119
 - DNA-polimerases, 113-115
 - estrutura das DNA-polimerases, 117-119
 - fitas contínuas e descontínuas, síntese das, 115-117
 - outras proteínas presentes no replissomo, 115
 - replicação unidirecional e bidirecional, 113
 - sistemas de replicação por círculo rolante, 130-131
 - término, 126-130
 - genomas circulares, 126-127
 - genomas lineares, 127-128
 - sistemas inibidores alternativos, 129-130
 - telômeros e telomerase, 128-129
- Replicação
- deslizes de replicação, 136
 - orientação e localização de genes em relação à origem e ao sentido da, 74
 - unidirecional e bidirecional, 113
- Replicons, forma, número e organização em
- genomas nucleares eucarióticos, 92
 - genomas procarióticos, 63-64
- Retrotransposons, 197-201
- Reversão direta de lesão, 147-148
- Revisão de leitura e reparação de mal pareamentos, 146-147
- Ribocomutadores, 293-294
- Ribossomos, 263-265
- sítios ativos, estrutura e dinâmica dos, 265
- RNA, 31-35
- alongamento da cadeia de, 222-226
 - e DNA, análise de, 350-357
 - processamento e maturação de, 233-254
 - técnica de análise de (biologia molecular), 350-357
- RNA-polimerases, 207-208
- RNAPs de procariotos, 208-209
- RNAs
- como moléculas reguladoras, 293-296
 - na regulação da expressão gênica, 312-315
 - não codificados no controle da transcrição, 312
 - transportadores, 259-263
- S**
- SCOP, 378
- Segregação correta dos cromossomos durante a meiose, 175-176
- Seleção imunológica, 345
- Sequenciamento em grande escala, 359
- Sequências
- clonadas, 345-347
 - composição, 92-105
 - elementos transponíveis, 101-103
 - gênicas, 92-99
 - genomas nucleares eucarióticos, 92-103
 - repetidas simples, 99-101
 - sequências intergênicas, 99-103
- Sequestro de proteínas reguladoras
- sRNAs, 295-296
- Sintenia
- genomas nucleares eucarióticos, 103-104
 - genomas procarióticos, 76-77
- Síntese
- de proteínas, 265-273
 - de proteínas, término da, 273
 - translesão, 155-156
- siRNAs, 315
- Sistema
- de reparação pós-replicação, 153-156
 - de replicação por círculo rolante, 130-131
 - inibidor alternativo
 - genomas lineares, 129-130
 - SOS, 154-155
- Southern blot* (biologia molecular), 351-353
- Splicing*
- alternativo, 240-243
 - hnRNAs, 237-240
- sRNAs
- controles positivo e negativo da tradução, 296
 - e o sequestro de proteínas reguladoras, 295-296
- Superóperons e ultraóperons, 77-79
- Supertorção, 25-27
- T**
- Técnica de PCR, 355-357
- Telômeros e telomerase, 128-129
- Tendências evolutivas
- genomas procarióticos, 82
- Terminação, 226-228
- Topoisomerases, 27-29
- transcrição, proteínas envolvidas na, 210
- Transcrição, 205-232
- alongamento da cadeia de RNA, 222-226
 - ciclo, 206-207
 - controle do início da transcrição, 278-286
 - controle do término da transcrição, 286-288
 - cromatina, 228-229
 - fábricas de transcrição, 230
 - do óperon *trp* de *E. Coli*, controle negativo da, 285
 - em grande escala, 357-359
 - fatores de transcrição, 303-305
 - fidelidade e correção de erros, 230
 - início, 214-222
 - eucariotos, 217-222
 - transcrição por RNAPI e RNAPIII, 221-222
 - transcrição por RNAPII, 218-221
 - procariotos, 214-217
 - por antiterminador, controle da, 288
 - por RNAPI e RNAPIII, 221-222
 - por RNAPII, 218-221
 - promotores, 211-214
 - eucariotos, 213-214
 - procariotos, 211-213
 - proteínas envolvidas, 207-211
 - outras proteínas, 210-211
 - fatores de transcrição, 210-211
 - topoisomerases, 210
 - RNA-polimerases, 207-210
 - RNAPs de eucariotos, 209-210
 - RNAPs de procariotos, 208-209
 - regulação do início da transcrição, 302

- terminação, 226-228
 - em eucariotos, 227-228
 - em procariotos, 226-227
 - Transcriptases reversas, 336
 - Transcritos
 - ampliação de regiões específicas, 355
 - primários dos RNAs em eucariotos
 - clivagens, 245-246
 - primários dos RNAs em procariotos
 - clivagens, 244-245
 - Transfecção com DNA de fagos, 340
 - Transformação
 - bacteriana, 340
 - genética de plantas, 348-350
 - via *Agrobacterium tumefaciens*, 348-350
 - via bombardeamento de DNA, 350
 - via eletroporação, 350
 - Transformantes, seleção de, 340-371
 - Translocação
 - alongamento da cadeia peptídica, 271-273
 - Transposons, 196-197
 - Trans-*splicing*, 243-244
 - tRNAs, processamento de, 246-248
 - clivados e adição de 3' CCA, geração de, 246
 - e as aminoacil-tRNA-sintetases, 261-262
 - estrutura, 259-261
 - modificações de bases, 246-247
 - splicing* em íntrons, 247-248
 - supressores, 262-263
 - Troca de fitas, 181
 - recombinação sítio-específica, 181
- U**
- Ubiquitinação e desubiquitinação, 307
 - Ultraóperons, 77-79
 - União
 - de extremidades mediada por micro-homologias, 158
 - de extremidades não homólogas, 158
 - UniProt Knowledgebase, 377
 - URLs, 365
- V**
- Vetores
 - de expressão eucarióticos, 345-346
 - de expressão procarióticos, 345
 - Vírus ou plasmídeos, defesa bacteriana de, (CRISPR), 298-299
 - Vírus, 5
- X**
- xBase, 379
- Z**
- Zigotos, 325-327
 - Zíper de leucina, 305