

14<sup>a</sup> EDIÇÃO

MICRO

BIOLO

GIÁ DE BROCK

MADIGAN • MARTINKO • BENDER • BUCKLEY • STAHL



# 4 Microbiologia molecular

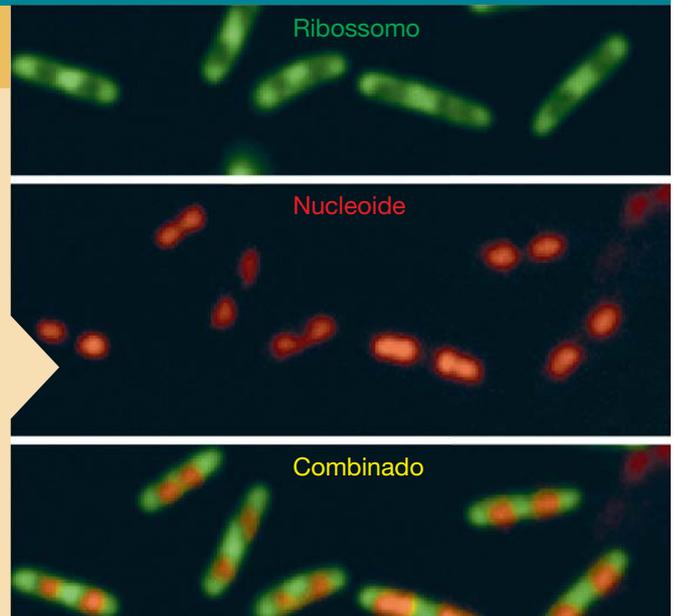
## microbiologiahoje

### A essência da vida: microbiologia molecular

Como você certamente já descobriu, os microrganismos possuem uma variedade surpreendente de capacidades metabólicas. O código genético das células individuais é responsável pelas características distintas observadas em todas as formas de vida. Embora esse depósito de informações necessite ser protegido e passado de geração em geração, a informação também precisa “ganhar vida” para permitir que as células executem esse engenhoso conjunto de atividades fascinantes. Esse fluxo de informação biológica essencial – de DNA relativamente inerte até a síntese de proteínas e enzimas críticas para a sobrevivência celular – é conhecido como o dogma central da vida.

A microbiologia molecular tem sido a base para a compreensão das etapas individuais do dogma central: replicação do DNA, transcrição do DNA em RNA, e tradução do RNA em proteínas. Com o advento das técnicas de ponta, novas descobertas em relação a esses processos biológicos essenciais ainda estão em andamento. Por exemplo, os microbiologistas podem agora identificar a localização de moléculas específicas em células vivas, utilizando marcadores fluorescentes e microscópios de fluorescência de alta resolução. A foto ao lado ilustra o uso do microscópio de fluorescência e a marcação de proteínas em células de *Escherichia coli* em crescimento ativo, para visualização real de RNA-polimerases e ribossomos, duas maquinarias celulares essenciais para o dogma central, em ação. A imagem resultante demonstra que a maioria dos ribossomos, as “fábricas proteicas”, estão localizadas nas porções terminais da célula e em regiões onde ocorre a formação do septo durante a divisão celular (foto superior, em verde), enquanto as RNA-polimerases estão associadas ao DNA cromossômico na região do nucleóide, que está localizado no centro da célula (foto do meio, em vermelho). A sobreposição das duas imagens (foto inferior) nos permite observar que a organização espacial do fluxo de informação biológica, de fato, existe em células bacterianas, apesar da ausência de organelas.

Bakshi, S., A. Siryaporn, M. Goulian, e J. C. Weisshaar. 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Molecular Microbiology* 85: 21–38.



- I O código da vida: estrutura do genoma bacteriano 108
- II Transmissão da informação genética: replicação do DNA 115
- III Síntese de RNA: transcrição 120
- IV Síntese de proteínas 127

Central para a vida é o fluxo de informações. O que instrui a célula a se reproduzir e sobreviver em um determinado ambiente, e quais os processos que prescrevem a produção de células? As células podem ser consideradas máquinas químicas e dispositivos de codificação. Como máquinas químicas, as células são capazes de transformar sua vasta variedade de macromoléculas em novas células. Como dispositivos de codificação, elas armazenam, processam e fazem uso da informação genética. Os genes, os mecanismos pelos quais são transferidos para novas

células, e sua expressão são as bases da biologia molecular e do dogma central da vida. Este capítulo ressalta o código genético celular e as etapas que as células percorrem para transformar essa informação em macromoléculas que desempenham funções celulares. Neste capítulo, será analisado como os processos ocorrem em bactérias, particularmente *Escherichia coli*, o membro do domínio *Bacteria* que é o organismo-modelo para a biologia molecular. Essa bactéria ainda é o organismo mais bem caracterizado entre procarionotos ou eucarionotos.

## I • O código da vida: estrutura do genoma bacteriano

### 4.1 Macromoléculas e genes

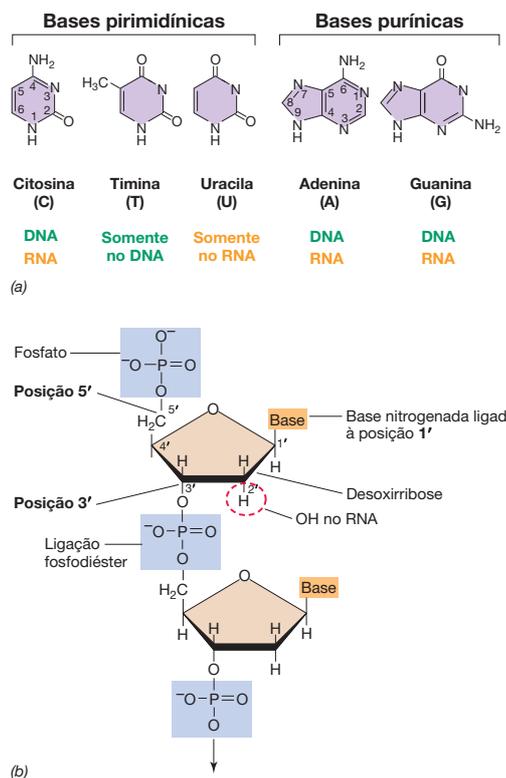
A unidade funcional da informação genética corresponde ao **gene**. Todas as formas de vida, incluindo os microrganismos, contêm genes. Fisicamente, os genes estão localizados nos cromossomos ou em outras moléculas grandes, referidas coletivamente como **elementos genéticos**. Esses elementos compõem o complemento total da informação genética, ou **genoma**, em uma célula ou vírus. Na biologia moderna, as células podem ser caracterizadas de acordo com seu complemento de genes. Assim, se deseja-se entender como os microrganismos funcionam, deve-se compreender como os genes codificam a informação.

A informação genética nas células está contida nos **ácidos nucleicos**, DNA ou RNA. O **ácido desoxirribonucleico**, DNA, carrega o código genético da célula, enquanto o **ácido ribonucleico**, RNA, é uma molécula intermediária que converte esse código em sequências definidas de aminoácidos em proteínas. A informação genética reside na sequência de monômeros nos ácidos nucleicos. Assim, ao contrário dos polissacarídeos e lipídeos que são normalmente compostos por longas unidades repetidas, os ácidos nucleicos são **macromoléculas informacionais**. Uma vez que a sequência de monômeros nas proteínas é determinada pela sequência dos ácidos nucleicos que a codificam, as proteínas também são macromoléculas informacionais.

Os monômeros dos ácidos nucleicos são denominados **nucleotídeos**, conseqüentemente, DNA e RNA são **polinucleotídeos**. Um nucleotídeo é composto por três componentes: um açúcar-pentose (ribose no RNA ou desoxirribose no DNA), uma base nitrogenada e uma molécula de fosfato,  $\text{PO}_4^{3-}$ . As estruturas gerais dos nucleotídeos de DNA e RNA são muito similares (Figura 4.1). As bases nitrogenadas são **purinas** (*adenina* e *guanina*), que contêm dois anéis heterocíclicos fundidos, ou **pirimidinas** (*timina*, *citocina* e *uracila*), que contêm um único anel heterocíclico de seis membros (Figura 4.1a). Guanina, adenina e citosina estão presentes tanto no DNA quanto no RNA. Com poucas exceções, a timina está presente apenas no DNA e a uracila presente apenas no RNA.

As bases nitrogenadas estão ligadas ao açúcar-pentose por meio de uma ligação glicosídica entre o átomo do carbono 1 do açúcar e o átomo de nitrogênio na base, o nitrogênio 1 (nas bases pirimidínicas) ou 9 (nas bases purínicas). Uma base nitrogenada ligada ao seu açúcar, mas sem a presença do fosfato, é denominada **nucleosídeo**. Nucleotídeos são nucleosídeos com a adição de um ou mais fosfatos (Figura 4.1b). Os nucleotídeos

desempenham outras funções, além de seu papel nos ácidos nucleicos. Os nucleotídeos, especialmente o trifosfato de adenosina (ATP) e o trifosfato de guanosina (GTP), são moléculas importantes na conservação de energia (↔ Seção 3.7). Outros nucleotídeos ou derivados atuam em reações redox, como carreadores de açúcares na síntese de polissacarídeos ou como moléculas reguladoras.



**Figura 4.1** Componentes dos ácidos nucleicos. (a) As bases nitrogenadas do DNA e do RNA. Observe o sistema de numeração dos anéis. Ao ligar-se ao carbono 1' do açúcar-fosfato, uma base pirimidínica liga-se por meio de N-1 e uma base purínica liga-se por meio de N-9. (b) Parte de uma cadeia de DNA. Os números no açúcar do nucleotídeo contêm uma apóstrofe (') após os mesmos, uma vez que os anéis das bases nitrogenadas também são numerados. No DNA, um hidrogênio está presente no carbono 2' do açúcar-pentose. No RNA, um grupo OH ocupa essa posição. Os nucleotídeos são unidos por uma ligação fosfodiéster.

## Os ácidos nucleicos: DNA e RNA

O esqueleto dos ácidos nucleicos é um polímero de moléculas de açúcar e fosfato alternadas. Os nucleotídeos são covalentemente ligados por meio do fosfato entre o carbono 3' (3 linha) de um açúcar e o carbono 5' do açúcar seguinte. (Números com marcas de primeira linha referem-se a posições no anel de açúcar; números sem essas marcas referem-se a posições no anel das bases.) A ligação fosfato é denominada **ligação fosfodiéster**, uma vez que o fosfato se conecta a duas moléculas de açúcar por meio de uma ligação éster (Figura 4.1b). A *sequência* de nucleotídeos em uma molécula de DNA ou RNA corresponde a sua **estrutura primária**, e a sequência de bases constitui a informação genética.

No genoma das células, o DNA é *dupla-fita*. Cada cromossomo consiste em duas fitas de DNA, cada fita contendo centenas de milhares a vários milhões de nucleotídeos ligados por ligação fosfodiéster. As fitas são mantidas unidas por meio de ligações de hidrogênio que se formam entre as bases de uma fita e aquelas da outra fita. Quando localizadas adjacentes umas às outras, bases purínicas e pirimidínicas podem formar ligações de hidrogênio (Figura 4.2). As ligações de hidrogênio são mais estáveis termodinamicamente quando a guanina (G) se liga à citosina (C) e a adenina (A) se liga à timina (T). O pareamento de bases específicas, A com T e G com C, garante que as duas fitas de DNA sejam *complementares* em sua sequência de bases; isto é, onde se encontra G em uma fita, encontra-se C na outra fita, e onde se encontra T em uma fita, sua fita complementar apresenta um A.

## Genes e etapas do fluxo informacional

Quando os genes são expressos, a informação genética armazenada no DNA é transferida ao ácido ribonucleico (RNA). Enquanto várias classes de RNA existem nas células, três tipos principais atuam na síntese de proteínas. O **RNA mensageiro (RNAm)** é uma molécula de fita simples que carrega a informação genética do DNA ao ribossomo, a máquina de síntese proteica. Os **RNAs transportadores (RNAts)** convertem

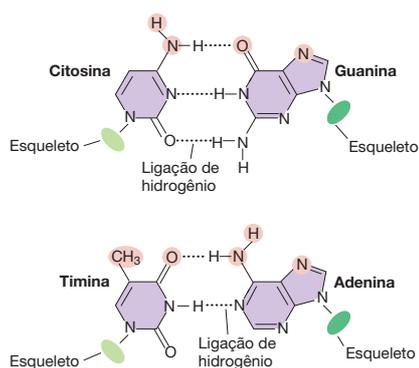
a informação genética nas sequências nucleotídicas do RNA em uma sequência definida de aminoácidos em proteínas. Os **RNAs ribossomais (RNAr)** são importantes componentes catalíticos e estruturais dos ribossomos. Os processos moleculares do fluxo da informação genética podem ser divididos em três estágios (Figura 4.3):

1. **Replicação.** Durante a replicação, a dupla-hélice do DNA é duplicada, produzindo duas cópias. A replicação é realizada por uma enzima denominada *DNA-polimerase*.
2. **Transcrição.** A transferência da informação genética do DNA para o RNA é denominada transcrição. A transcrição é realizada por uma enzima denominada *RNA-polimerase*.
3. **Tradução.** A síntese de uma proteína, utilizando a informação genética contida no RNAm, é denominada tradução.

As três etapas apresentadas na Figura 4.3 são características de todas as células e constituem o dogma central da biologia molecular: DNA → RNA → proteína. Muitas moléculas distintas de RNA são transcritas a partir de uma região relativamente curta da longa molécula de DNA. Em eucariotos, cada gene é transcrito, originando um único RNAm, enquanto em procarionotos, uma única molécula de RNAm pode carrear a informação genética de vários genes; isto é, de mais de uma região codificadora de proteínas. Existe uma correspondência linear entre a sequência de bases de um gene e a sequência de aminoácidos de um polipeptídeo. Cada grupo de três bases em uma molécula de RNAm codifica um único aminoácido, e cada triplete de bases é chamado de **códon**. Os códons são traduzidos em sequências de aminoácidos pelos ribossomos (que consistem em proteínas e RNA) RNAt, e proteínas auxiliares denominadas *fatores de tradução*. Enquanto o dogma central é invariável em células, será visto posteriormente que alguns vírus (que não são células, ⇄ Seção 1.2) violam esse processo de forma interessante (Capítulos 8 e 9).

## MINIQUESTIONÁRIO

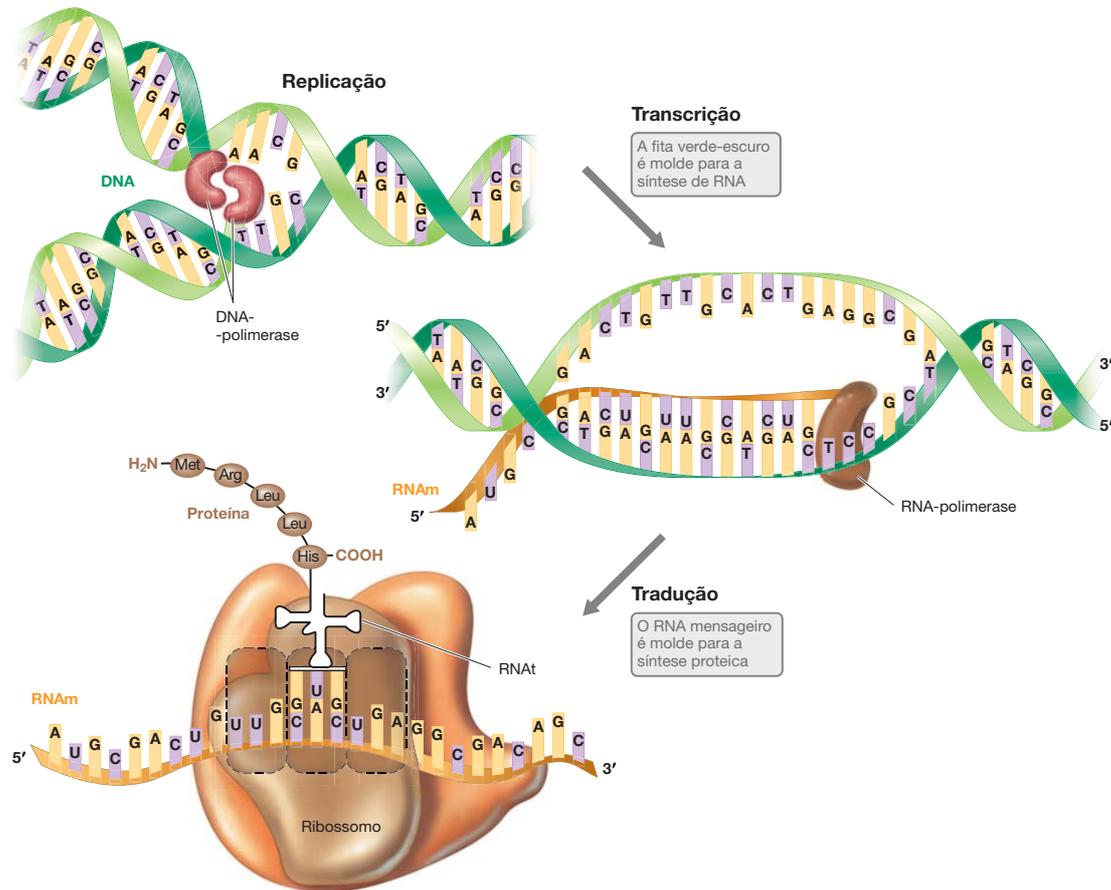
- O que é um genoma?
- Quais componentes são encontrados em um nucleotídeo? De que forma um nucleosídeo se difere de um nucleotídeo?
- Quais as três macromoléculas informacionais envolvidas no fluxo da informação genética?
- Em todas as células, há três processos envolvidos no fluxo da informação genética. Quais são esses processos?



**Figura 4.2** Pareamento específico entre guanina (G) e citosina (C) e entre adenina (A) e timina (T), por meio de ligações de hidrogênio. Esses são os típicos pares de bases encontrados no DNA de dupla-fita. Os átomos encontrados no sulco maior da dupla-hélice, que interagem com as proteínas, encontram-se assinalados em cor-de-rosa. Os esqueletos de desoxirribose-fosfato das duas fitas de DNA são também indicados. Observe as diferentes tonalidades de verde assinalando as duas fitas do DNA, uma convenção utilizada ao longo deste livro.

## 4.2 A dupla-hélice

Em todas as células, o DNA existe como uma molécula de dupla-fita, com duas fitas polinucleotídicas cujas sequências de bases são **complementares**. A complementaridade do DNA ocorre a partir do pareamento específico de bases por meio das ligações de hidrogênio: adenina sempre se liga à timina, e a guanina sempre se liga à citosina. Cada par de bases adenina-timina apresenta *duas* ligações de hidrogênio, enquanto cada par de bases guanina-citosina apresenta *três* ligações. Isso torna os pares GC mais fortes do que os pares AT. As duas fitas da molécula de DNA estão organizadas de forma **antiparalela** (Figura 4.4; as fitas de DNA aparecem ao longo da figura em dois tons de verde). Dessa forma, a fita da esquerda aparece no sentido 5'-3' de cima para baixo, enquanto a fita complementar aparece no sentido 5'-3' de baixo para cima. Embora



**Figura 4.3** Síntese dos três tipos de macromoléculas informacionais. Observe que, para qualquer gene em particular, somente uma das duas fitas da dupla-hélice de DNA é transcrita.

as ligações de hidrogênio individuais sejam bastante fracas, o grande número dessas ligações entre os pares de base de uma longa molécula de DNA confere estabilidade considerável à molécula, o suficiente para manter as duas fitas unidas.

As duas fitas de DNA são torcidas uma sobre a outra, formando uma dupla-hélice (Figura 4.5). Nessa dupla-hélice, o DNA forma dois sulcos distintos, o *sulco maior* e o *sulco menor*. A maioria das proteínas que interage especificamente com o DNA liga-se ao sulco maior, onde o espaço é consideravelmente maior. Devido à regularidade da estrutura da dupla-hélice, alguns átomos de cada base encontram-se sempre expostos no sulco maior (e alguns no sulco menor). As regiões essenciais dos nucleotídeos, importantes nas interações com as proteínas, são apresentadas na Figura 4.2.

### Tamanho e forma das moléculas de DNA

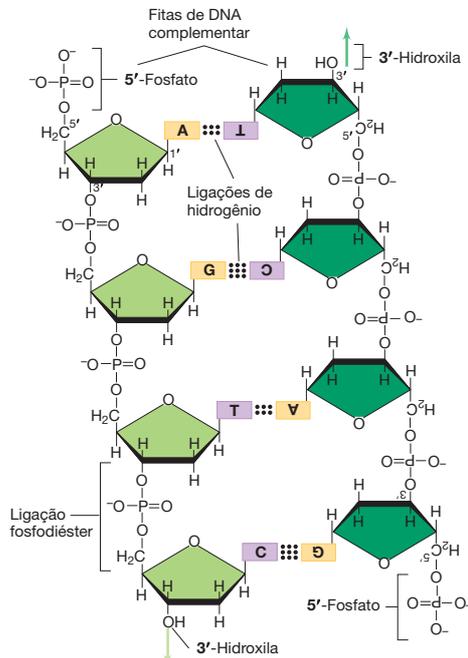
O tamanho de uma molécula de DNA pode ser expresso como o número de bases nucleotídicas ou pares de bases por molécula. Assim, uma molécula de DNA com 1.000 bases contém 1 quilobase (kb) de DNA. Se o DNA estiver na forma de dupla-hélice, utiliza-se a nomenclatura *pares de quilobases* (kpb).

Dessa forma, uma dupla-hélice contendo 5.000 pares de bases teria um tamanho de 5 kpb. A bactéria *Escherichia coli* contém cerca de 4.640 kpb de DNA em seu cromossomo. Uma vez que muitos genomas são bastante grandes, é mais simples falar em termos de milhões de pares de bases, ou *pares de megabases* (Mpb). Assim, o genoma de *E. coli* apresenta 4,64 Mpb.

Cada par de bases apresenta cerca de 0,34 nanômetros (nm) de comprimento ao longo da dupla-hélice, e cada volta da hélice contém aproximadamente 10 pares de bases. Assim, um DNA de 1 kpb apresenta uma hélice com 100 voltas, medindo  $0,34 \mu\text{m}$  de comprimento. O genoma de *E. coli* possui então  $4.640 \times 0,34 \mu\text{m} = 1,58 \text{ mm}$  de comprimento. Uma vez que as células de *E. coli* possuem aproximadamente  $2 \mu\text{m}$  de comprimento, o cromossomo é centenas de vezes maior do que a própria célula!

### Superenovelamento do DNA

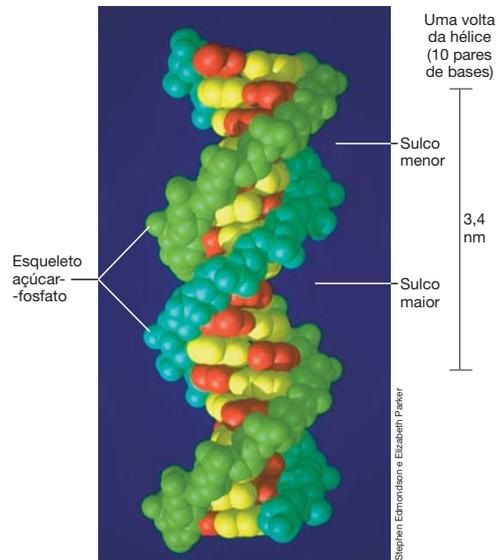
Considerando os cálculos anteriores, como é possível empacotar tanto DNA em um espaço tão diminuto? A solução consiste na imposição de uma estrutura de "ordem superior" para o DNA, na qual o DNA de dupla-fita é adicionalmente



**Figura 4.4** Estrutura do DNA. Natureza complementar e antiparalela do DNA. Observe que uma das cadeias termina em um grupamento 5'-fosfato, enquanto a outra termina em uma 3'-hidroxila. As bases em roxo representam as pirimidinas citosina (C) e timina (T), enquanto as bases em amarelo representam as purinas adenina (A) e guanina (G).

torcido, em um processo denominado *superenovelamento*. A **Figura 4.6** ilustra como ocorre o superenovelamento em uma molécula de DNA circular. Se uma molécula de DNA circular é linearizada, qualquer superenovelamento é perdido e o DNA torna-se “relaxado”. Assim, uma molécula de DNA relaxada apresenta exatamente o número predito de voltas da hélice, a partir do número de pares de bases.

Os superenovelamentos são inseridos ou removidos no DNA por enzimas denominadas *topoisomerases*. A atividade de superenovelamento submete a molécula de DNA a um estado de torção, muito similar à tensão imposta a um elástico quando é torcido. O DNA pode ser superenovelado tanto de forma positiva quanto negativa. No superenovelamento positivo, a dupla-hélice torna-se mais enovelada (apresenta mais do que o número natural de voltas), enquanto no superenovelamento negativo a dupla-hélice torna-se subenovelada (apresenta menos do que o número natural de voltas). O superenovelamento negativo ocorre quando o DNA é torcido ao longo de seu eixo, na direção oposta à da dupla-hélice dextrógira. O DNA superenovelado negativamente corresponde à forma mais predominante encontrada na natureza. No cromossomo de *Escherichia coli*, acredita-se que existam mais de 100 domínios superenovelados, cada qual estabilizado pela ligação de proteínas específicas ao DNA. A inserção de superenovelamentos no DNA requer energia a partir do ATP, enquanto a remoção dos superenovelamentos não re-



**Figura 4.5** Modelo computacional de um pequeno segmento de DNA apresentando o arranjo global da dupla-hélice. Um dos esqueletos de açúcar-fosfato é apresentado em azul e o outro em verde. As bases pirimidínicas são apresentadas em vermelho e as purínicas em amarelo. Observe a localização dos sulcos maior e menor (comparar com a Figura 4.2). Uma volta da hélice contém 10 pares de bases.

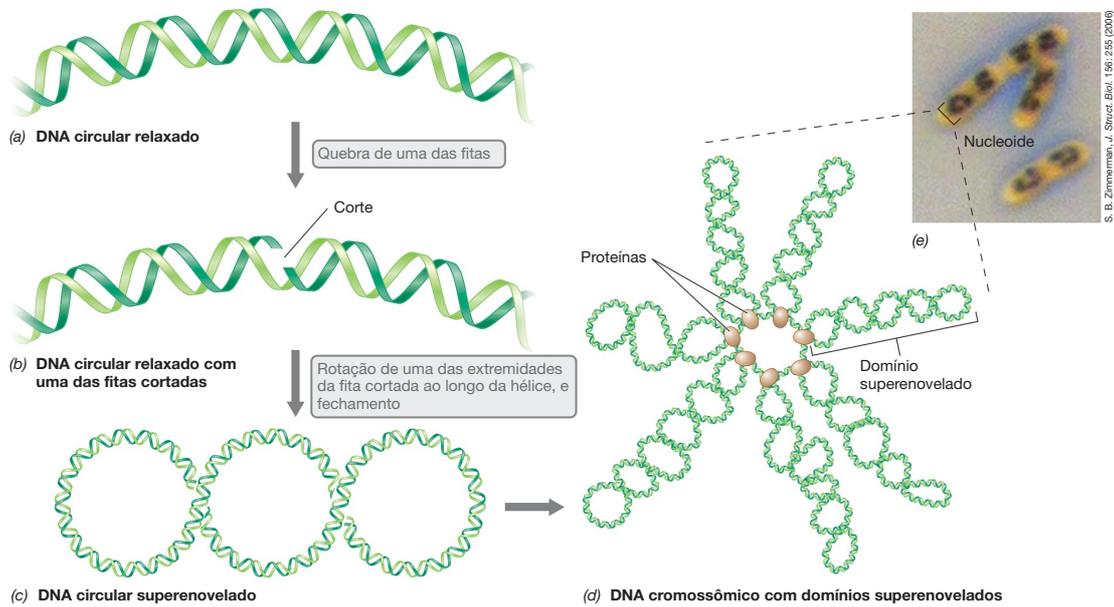
quer. Em bactérias e na maioria de arqueias, há uma enzima denominada **DNA-girase**, que é uma topoisomerase do tipo II, a qual introduz superenovelamentos negativos no DNA fazendo quebras na dupla-fita (**Figura 4.7**). Será visto posteriormente que arqueias, que vivem em temperaturas muito elevadas, possuem cromossomos que são superenovelados positivamente, e essa característica as auxilia na manutenção da estrutura do DNA em tais temperaturas (↔ Seção 16.13). Alguns antibióticos inibem a atividade da DNA-girase. Esses incluem as quinolonas (como o ácido nalidíxico), as fluorquinolonas (como o ciprofloxacino) e a novobiocina.

**MINIQUESTIONÁRIO**

- O que significa antiparalelo em termos de estrutura da dupla-fita de DNA?
- Defina o termo complementar, quando usado para se referir às duas fitas de DNA.
- O que torna os pares GC mais fortes do que os pares AT?
- Por que o superenovelamento é importante? Qual a enzima que facilita esse processo?

**4.3 Elementos genéticos: cromossomos e plasmídeos**

Estruturas contendo material genético (DNA na maioria dos organismos e RNA em alguns vírus) são denominadas *elementos genéticos*. O principal elemento genético de procarionotes é o **cromossomo**. Outros elementos genéticos podem ser encontrados, e desempenham papéis importantes na fun-

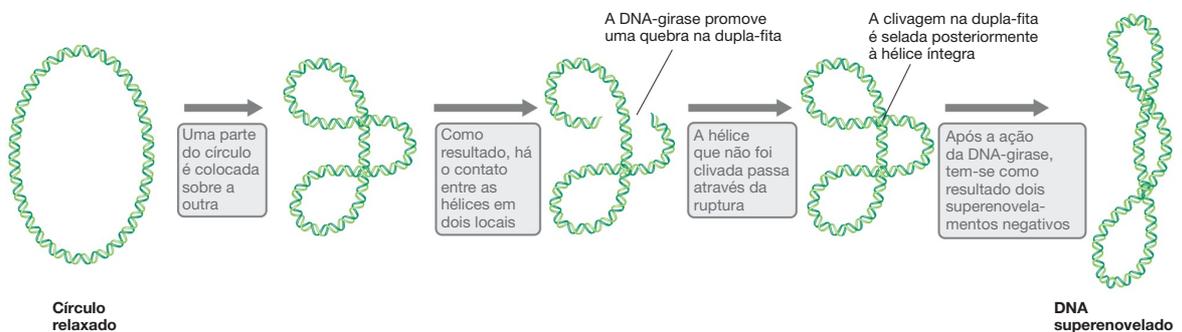


**Figura 4.6 DNA superenovelado.** (a-c) DNA circular clivado, relaxado e superenovelado. Um corte corresponde à quebra de uma ligação fosfodiéster em uma das fitas. (d) Na realidade, o DNA de dupla-fita do cromossomo bacteriano não se organiza como um DNA superenovelado, mas sim como vários domínios superenovelados, conforme apresentado na figura. (e) Imagens

simultâneas de contraste de fase e de fluorescência de *E. coli* ilustrando a localização do nucleóide dentro das células em crescimento. As células foram tratadas com um corante fluorescente específico para DNA e a coloração foi invertida para mostrar os nucleóides em preto.

ção gênica, tanto de procariotos quanto de eucariotos (Tabela 4.1). Esses incluem os *genomais virais*, *plasmídeos*, *genomas de organelas* e *elementos transponíveis*. Um procarioto típico possui um único cromossomo DNA circular, contendo todos (ou a maioria) dos genes encontrados na célula. Embora um único cromossomo corresponda à regra entre os procariotos, há exceções. Alguns poucos procariotos possuem dois ou até três cromossomos. Os genomas eucarióticos apresentam múltiplos cromossomos. Além disso, o DNA de todos os cromossomos eucarióticos conhecidos é linear, ao contrário da maioria dos cromossomos procarióticos, os quais correspondem a moléculas circulares de DNA.

Embora sejam considerados microrganismos, vírus não são células, mas, em vez disso, dependem da célula para a sua replicação. Entretanto, os vírus contêm genomas, de DNA ou RNA, que controlam sua própria replicação e transferência de célula a célula. São conhecidos genomas virais tanto lineares quanto circulares. Além disso, o ácido nucleico dos genomas virais pode ser de fita simples ou dupla. **Plasmídeos** são elementos genéticos que se replicam independentemente do cromossomo. A maioria dos plasmídeos é constituída por DNA de dupla-fita e, embora grande parte deles exiba configuração circular, alguns são lineares. Os plasmídeos são normalmente muito menores que os cromossomos.



**Figura 4.7 DNA-girase.** Introdução de superenovelamento negativo em um DNA circular, pela ação da DNA-girase (topoisomerase II), que realiza quebras na dupla-fita.

**Tabela 4.1** Tipos de elementos genéticos

Organismo	Elemento	Tipo de ácido nucleico	Descrição
Procaríoto	Cromossomo	DNA de dupla-fita	Extremamente longo, geralmente circular
Eucarioto	Cromossomo	DNA de dupla-fita	Extremamente longo, linear
Todos os organismos	Plasmídeo <sup>a</sup>	DNA de dupla-fita	Relativamente curto, circular ou linear, extracromossômico
Todos os organismos	Elemento transponível	DNA de dupla-fita	Sempre encontrado inserido em outra molécula de DNA
Mitocôndria ou cloroplasto	Genomas organelares	DNA de dupla-fita	Comprimento médio, geralmente circular
Vírus	Genoma viral	DNA ou RNA de dupla-fita ou simples	Relativamente curto, circular ou linear

<sup>a</sup>Plasmídeos são incomuns em eucariotos.

**Elementos transponíveis** são segmentos de DNA capazes de moverem-se de um sítio em uma molécula de DNA a outro sítio na mesma molécula, ou em uma molécula distinta de DNA. Os elementos transponíveis não são encontrados como moléculas de DNA independentes, em vez disso, são sempre encontrados inseridos em outras moléculas de DNA. Cromossomos, plasmídeos, genomas virais e qualquer outro tipo de molécula de DNA podem atuar como hospedeiro para um elemento transponível. Elementos transponíveis são encontrados em eucariotos e procariotos, desempenhando importantes papéis nos processos de variação genética (↪ Seção 10.11).

### Disposição dos genes no cromossomo de *Escherichia coli*

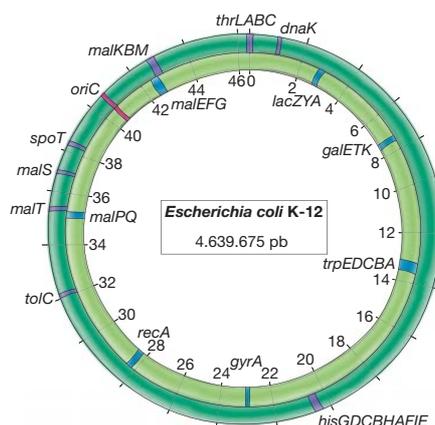
Muitos genomas bacterianos, incluindo aquele de *Escherichia coli*, foram completamente sequenciados, revelando assim o número e a localização dos genes que eles possuem. A linhagem de *E. coli* cujo cromossomo foi originalmente sequenciado, a linhagem MG1655, é um derivado da *E. coli* K-12, a linhagem tradicional utilizada em genética. Um mapa genético correspondente ao cromossomo de 4.639.675 pb é apresentado na **Figura 4.8**, com apenas alguns dos vários milhares de genes no cromossomo de *E. coli* descrito. As distâncias no mapa são apresentadas em 100 pares de quilobases de DNA. Análises genômicas revelaram 4.288 possíveis genes codificadores de proteínas que representam 88% do genoma de *E. coli*. Aproximadamente 1% do genoma é de genes codificadores de RNAt e RNAr. Isso está em contraste com os genomas eucarióticos, que em geral contêm muito mais DNA do que é necessário para codificar todas as proteínas requeridas para a função celular. Por exemplo, no genoma humano, apenas cerca de 3% do DNA total efetivamente codificam proteínas. O DNA eucariótico “extra” está presente entre as sequências codificadoras no DNA (o qual é removido após a transcrição), ou como sequências repetidas, algumas das quais repetidas centenas ou milhares de vezes.

O mapeamento genético dos genes que codificam as enzimas que funcionam na mesma via bioquímica em *E. coli* demonstrou que esses genes estão frequentemente agrupados. No mapa genético, na Figura 4.8, alguns desses grupos são demonstrados. Observe, por exemplo, os grupos gênicos de *gal*, *trp* e *his*. Cada um desses grupos gênicos constitui um

**óperon**, que é transcrito em um único RNA que codifica diversas proteínas individuais. Genes para muitas outras vias bioquímicas em *E. coli* não estão agrupados. Por exemplo, genes para a degradação de maltose (genes *mal*, Figura 4.8) estão espalhados por todo o cromossomo. A análise da sequência do cromossomo de *E. coli* demonstrou que mais de 70% das unidades transcricionais, preditas ou conhecidas, contêm apenas um único gene, e apenas 6% dos óperons possuem quatro ou mais genes. Algumas sequências codificadoras estão em uma fita do cromossomo, enquanto outras estão na fita oposta, e a análise genômica demonstrou que existem números iguais de genes em ambas as fitas. Ao contrário dos procariotos, os eucariotos não possuem óperons.

### Princípios gerais dos plasmídeos

Muitas células procarióticas contêm outros elementos genéticos, em particular, plasmídeos, além do cromossomo. Embora os plasmídeos possuam a sua própria origem de replicação, eles dependem de enzimas cromossomicamente

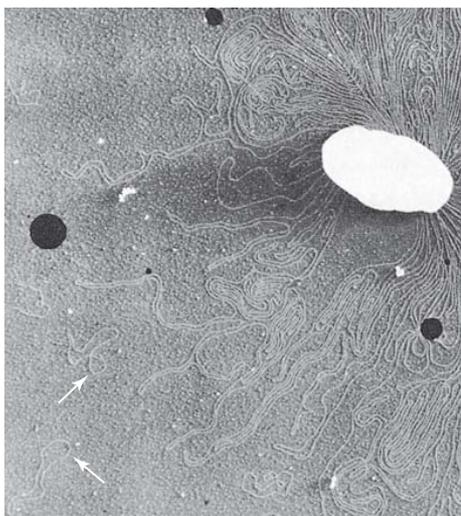


**Figura 4.8** O cromossomo da linhagem K-12 de *Escherichia coli*. As distâncias no mapa são dadas em 100 quilobases de DNA. O cromossomo contém 4.639.675 pares de bases e 4.288 fases de leitura aberta (genes). Dependendo da fita de DNA, a localização de alguns genes e óperons são indicadas. A replicação (Figura 4.3) ocorre em ambas as direções a partir da origem de replicação do DNA, *oriC*, indicada em vermelho.

codificadas para a sua replicação. A maioria dos plasmídeos é geralmente dispensável, uma vez que raramente possuem genes necessários para o crescimento em todas as condições. Em contrapartida, os genes essenciais se encontram nos cromossomos. Ao contrário dos vírus, os plasmídeos não possuem uma forma extracelular e existem dentro das células como DNA livre. Milhares de plasmídeos diferentes são conhecidos. Na verdade, mais de 300 plasmídeos diferentes, de ocorrência natural, foram isolados de linhagens de *Escherichia coli* por si só.

A maioria dos plasmídeos é constituída por DNA de dupla-fita. A maioria exibe configuração circular, porém muitos de configuração linear são conhecidos. Os plasmídeos de ocorrência natural variam em tamanho, de aproximadamente 1 kpb a mais de 1 Mpb. Plasmídeos típicos são moléculas de DNA circular dupla-fita que apresentam menos de 5% do tamanho dos cromossomos (Figura 4.9). A maioria dos plasmídeos de DNA isolados de células é superenovelado, que corresponde à forma mais compacta do DNA quando no interior de uma célula (Figura 4.6). Algumas bactérias podem conter diversos tipos diferentes de plasmídeos. Por exemplo, *Borrelia burgdorferi* (o patógeno da doença de Lyme, ⇨ Seção 30.4) contém 17 plasmídeos circulares e lineares diferentes!

As enzimas de replicação celular também replicam plasmídeos. Assim, os genes carregados pelo plasmídeo estão envolvidos principalmente no controle da iniciação da replicação e na partição dos plasmídeos replicados para as células-filhas. Diferentes plasmídeos também encontram-se presentes em células em números distintos denominados *número de cópias*. Alguns plasmídeos são encontrados em apenas 1 a 3 cópias por célula, enquanto outros podem estar presentes em mais de 100 cópias. O número de cópias é controlado por genes plasmidiais e por interações entre o hospedeiro e o plasmídeo.

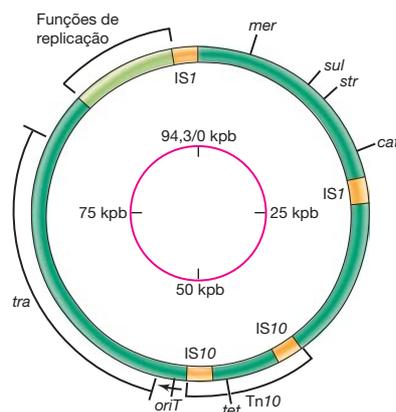


**Figura 4.9** O cromossomo bacteriano e os plasmídeos bacterianos, vistos ao microscópio eletrônico. Os plasmídeos (setas) são as estruturas circulares e são muito menores do que o DNA cromossômico principal. A célula (estrutura grande, branca) foi cuidadosamente degradada para que o DNA permanecesse intacto.

## Tipos de plasmídeos

Embora, por definição, os plasmídeos não codifiquem funções essenciais para o hospedeiro, esses podem carrear genes que influenciam profundamente na fisiologia celular do hospedeiro. Entre os grupos de plasmídeos mais estudados e amplamente distribuídos em células, encontram-se os plasmídeos de resistência, geralmente denominados *plasmídeos R*, que conferem resistência a antibióticos ou vários outros inibidores de crescimento. No geral, os genes de resistência codificam proteínas que inativam os antibióticos ou protegem a célula por algum outro mecanismo. Vários genes de resistência a antibióticos podem ser codificados por um único plasmídeo R; alternativamente, uma célula com múltipla resistência pode conter vários plasmídeos R diferentes. O plasmídeo R100, por exemplo, com 94,3 kpb (Figura 4.10), codifica genes que conferem resistência a sulfonamidas, estreptomicina, espectinomicina, ácido fusídico, cloranfenicol e tetraciclina. O plasmídeo R100 contém ainda vários genes que conferem resistência ao mercúrio. Bactérias patogênicas resistentes a antibióticos são de importância médica considerável, e sua incidência crescente está correlacionada ao uso crescente de antibióticos para o tratamento de doenças infecciosas em seres humanos e animais (⇨ Seção 27.17).

Microrganismos patogênicos possuem uma variedade de características que os permitem colonizarem hospedeiros e estabelecer infecções. As duas principais características envolvidas na virulência (capacidade de provocar doenças) de patógenos são frequentemente codificadas por plasmídeos: (1) a capacidade de o patógeno ligar-se a e colonizar tecidos específicos do hospedeiro, e (2) a produção de toxinas, enzimas e outras moléculas que promovem danos ao hospedeiro. Muitas bactérias produzem ainda proteínas que inibem ou matam espécies estreitamente relacionadas, ou mesmo linhagens diferentes da mesma espécie. Estes agentes, denominados



**Figura 4.10** Mapa genético do plasmídeo de resistência R100. O círculo interno mostra o tamanho em pares de quilobases. O círculo externo mostra a localização dos principais genes de resistência a antibióticos e outras funções-chave: *mer*, resistência ao íon mercúrio; *sul*, resistência a sulfonamida; *str*, resistência à estreptomicina; *cat*, resistência ao cloranfenicol; *tet*, resistência à tetraciclina; *oriT*, origem de transferência conjugativa; *tra*, funções de transferência. As localizações das sequências de inserção (IS) e o transposon Tn10 também são mostrados. Os genes para a replicação plasmidial são encontrados na região de 88 a 92 kpb.

**bacteriocinas**, são análogos dos antibióticos, porém apresentam um espectro de ação mais restrito do que estes. Os genes que codificam as bacteriocinas e as proteínas necessárias ao seu processamento e transporte, e que conferem imunidade ao organismo produtor, são frequentemente carregados por um plasmídeo. Por exemplo, *E.coli* produz bacteriocinas denominadas *colicinas*, que ligam-se a receptores específicos, situados na superfície das células suscetíveis, e promovem a morte celular interrompendo a função da membrana. Outras colicinas exibem atividade de nuclease, degradando DNA ou RNA de linhagens suscetíveis.

Em alguns casos, os plasmídeos codificam propriedades fundamentais para a ecologia da bactéria. Por exemplo, a capacidade de *Rhizobium* em interagir com plantas e formar nódulos radiculares fixadores de nitrogênio necessita de determinadas funções plasmidiais (↔ Seção 22.3). Outros plasmídeos conferem propriedades metabólicas especiais às células bacterianas, como a capacidade de degradar poluentes tóxicos. Algumas características especiais conferidas pelos plasmídeos estão resumidas na **Tabela 4.2**.

#### MINIQUESTIONÁRIO

- O que define um cromossomo em procaríotos?
- O que são vírus e plasmídeos?
- O quão grande, aproximadamente, é o genoma de *Escherichia coli* em pares de bases? Quantos genes ele contém?
- Quais propriedades um plasmídeo R confere à sua célula hospedeira?

**Tabela 4.2** Exemplos de características atribuídas por plasmídeos em procaríotos

Características	Organismos
<b>Produção de antibiótico</b>	<i>Streptomyces</i>
<b>Conjugação</b>	Vasta variedade de bactérias
<b>Funções metabólicas</b>	
Degradação de octano, cânfora, naftaleno	<i>Pseudomonas</i>
Degradação de herbicidas	<i>Alcaligenes</i>
Formação de acetona e butanol	<i>Clostridium</i>
Utilização de lactose, sacarose, citrato ou ureia	Bactérias entéricas
Produção de pigmento	<i>Erwinia</i> , <i>Staphylococcus</i>
Produção de vesícula de gás	<i>Halobacterium</i>
<b>Resistência</b>	
Resistência a antibiótico	Vasta variedade de bactérias
Resistência a metais tóxicos	Vasta variedade de bactérias
<b>Virulência</b>	
Produção de tumor em plantas	<i>Agrobacterium</i>
Nodulação e fixação de nitrogênio simbiótica	<i>Rhizobium</i>
Produção de bacteriocina e resistência	Vasta variedade de bactérias
Invasão de célula animal	<i>Salmonella</i> , <i>Shigella</i> , <i>Yersinia</i>
Coagulase, hemolisina, enterotoxina	<i>Staphylococcus</i>
Toxinas e cápsula	<i>Bacillus anthracis</i>
Enterotoxina, antígeno K	<i>Escherichia coli</i>

## II • Transmissão da informação genética: replicação do DNA

A replicação do DNA é necessária para que as células dividam-se, seja na reprodução, originando novos organismos, como no caso de microrganismos unicelulares, seja na produção de novas células em um organismo multicelular. Para transmitir com sucesso a informação genética de uma célula-mãe para uma célula-filha idêntica, a replicação do DNA precisa ser altamente precisa. Esse processo requer a atividade de uma série de enzimas especiais.

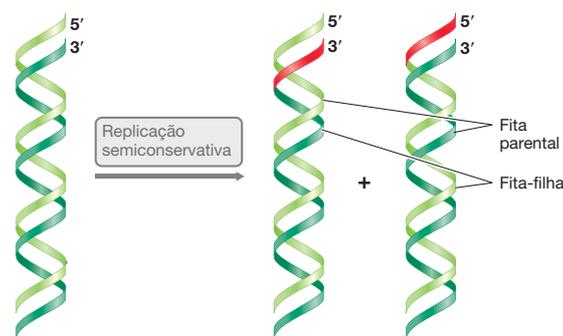
### 4.4 Moldes e enzimas

Como visto anteriormente, o DNA encontra-se nas células como uma dupla-hélice, exibindo pareamento de bases complementares (Figuras 4.3 e 4.4). Quando a dupla-hélice do DNA é aberta, uma nova fita pode ser sintetizada como complemento de cada uma das fitas parentais. Conforme apresentado na **Figura 4.11**, a replicação é um processo **semiconservativo**, significando que as duas duplas hélices resultantes consistem em uma fita recém-sintetizada e uma fita parental. A fita de DNA que é utilizada como molde na produção de uma fita complementar é denominada molde, e na replicação do DNA, cada fita parental corresponde a um molde para uma fita-filha recém-sintetizada (Figura 4.11).

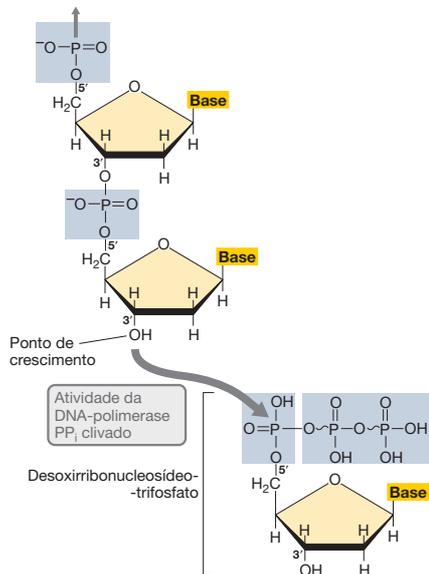
O precursor de cada novo nucleotídeo na fita de DNA corresponde a um desoxinucleosídeo 5'-trifosfato. Durante a inserção, dois fosfatos terminais são removidos, e o fosfato interno é então ligado covalentemente à desoxirribose da cadeia em crescimento (**Figura 4.12**). A adição do nucleotídeo à

fita em crescimento requer a presença de um grupo hidroxil livre, o qual encontra-se disponível somente na extremidade 3' da molécula. Isso leva a um importante princípio, de que a replicação do DNA sempre ocorre *a partir da extremidade 5', em direção à extremidade 3'*, com o grupo 5'-fosfato do nucleotídeo a ser adicionado à cadeia sendo ligado ao grupo 3'-hidroxil do nucleotídeo previamente adicionado.

As enzimas que catalisam a adição dos desoxinucleotídeos são denominadas **DNA-polimerases**. Existem diversos ti-



**Figura 4.11** Visão geral da replicação do DNA. A replicação do DNA é um processo semiconservativo em todas as células. Observe que cada uma das novas duplas-hélices contém uma fita nova (ilustrada em vermelho, no alto) e uma fita parental.

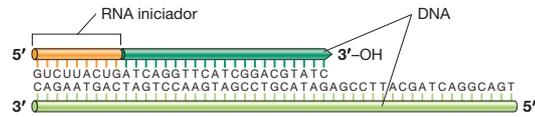


**Figura 4.12** Extensão de uma cadeia de DNA por meio da adição de um desoxirribonucleosídeo-trifosfato à extremidade 3' da cadeia. O crescimento ocorre a partir do 5'-fosfato, em direção à extremidade 3'-hidroxila. A enzima DNA-polimerase catalisa a reação de adição. Os quatro precursores são a desoxitimidina-trifosfato (dTTP), desoxiadenosina-trifosfato (dATP), desoxiguanosina trifosfato (dGTP) e desoxicitidina-trifosfato (dCTP). Na inserção do nucleotídeo, os dois fosfatos terminais do trifosfato são clivados na forma de pirofosfato (PP<sub>i</sub>). Assim, duas ligações-fosfato ricas em energia são consumidas na adição de cada nucleotídeo.

pos de tais enzimas, cada uma apresentando uma função específica. Existem cinco tipos de DNA-polimerases distintas em *Escherichia coli*, denominadas DNA-polimerases I, II, III, IV e V. A DNA-polimerase III (DNA Pol III) é a principal enzima envolvida na replicação do DNA cromossômico. A DNA-polimerase I (DNA Pol I) também está envolvida na replicação cromossômica, embora em menor grau (ver a seguir). As demais DNA-polimerases auxiliam no reparo de DNA danificado (↔ Seção 10.4).

Todas as DNA-polimerases conhecidas sintetizam DNA na direção 5' → 3'. No entanto, nenhuma DNA-polimerase conhecida é capaz de iniciar a síntese de uma nova cadeia; todas essas enzimas somente são capazes de adicionar um nucleotídeo a um grupo 3'-OH preexistente. Dessa forma, para iniciar uma nova cadeia, é necessário um **iniciador**, uma molécula de ácido nucleico à qual a DNA-polimerase pode adicionar o primeiro nucleotídeo. Na maioria dos casos, este iniciador consiste em um pequeno segmento de RNA em vez de DNA (Figura 4.13).

Quando a dupla-hélice é aberta no início da replicação, uma enzima de polimerização de RNA sintetiza o RNA iniciador. Essa enzima, denominada **primase**, sintetiza um pequeno segmento de RNA (cerca de 11-12 nucleotídeos) que é complementar à fita de DNA-molde quanto ao pareamento de bases. Ao final dessa molécula de RNA iniciador, no ponto de crescimento, há um grupo 3'-OH ao qual a DNA-polimerase adiciona o primeiro desoxirribonucleotídeo. Desse modo, a



**Figura 4.13** O RNA iniciador. Estrutura da combinação RNA-DNA, formada na iniciação da síntese de DNA. O RNA iniciador está destacado em cor de laranja.

extensão continuada da molécula ocorre sob a forma de DNA, em vez de RNA. A molécula recém-sintetizada apresenta uma estrutura semelhante àquela ilustrada na Figura 4.13. Eventualmente, o iniciador será removido e substituído por DNA, conforme descrito na próxima seção.

#### MINIQUESTIONÁRIO

- A qual extremidade (5' ou 3') de uma fita de DNA recém-sintetizada a polimerase adiciona uma base?
- Por que há a necessidade de um iniciador na replicação do DNA? De que o iniciador é constituído?

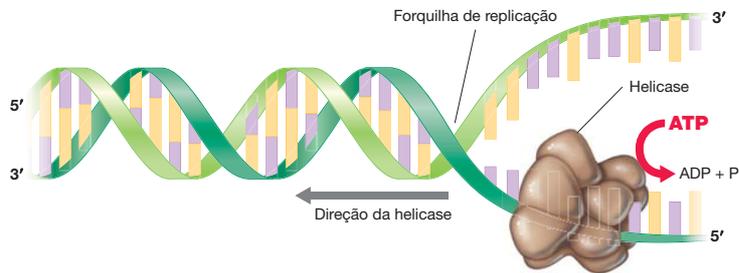
## 4.5 A forquilha de replicação

Grande parte de nosso conhecimento sobre o mecanismo de replicação do DNA foi obtido de estudos realizados com a bactéria *Escherichia coli*; no entanto, é provável que o processo de replicação seja bastante similar em todas as bactérias. Ao contrário, embora a maioria das espécies de arqueias possua cromossomos circulares, muitos dos eventos da replicação do DNA são mais semelhantes àquels observados em células eucarióticas que em *Bacteria*, um reflexo da filiação filogenética entre *Archaea* e *Eukarya* (Figura 1.6b).

### Iniciação da síntese de DNA

Antes da DNA-polimerase ser capaz de sintetizar novo DNA, a dupla-hélice preexistente deve ser desenovelada para expor as fitas-molde. A região de DNA desenovelado, onde ocorre a replicação, é denominada de **forquilha de replicação**. A enzima **DNA-helicase** desenovela a dupla-hélice do DNA, utilizando energia oriunda do ATP, expondo uma pequena região de fita simples (Figura 4.14). As helicases deslocam-se ao longo do DNA e separam as fitas, imediatamente à frente da forquilha de replicação. A região de fita simples é imediatamente complexada a cópias de uma proteína de ligação à fita simples, para estabilizar o DNA na forma de fita simples, e prevenir a reversão à dupla-hélice. O desenovelamento da dupla-hélice pela helicase origina superenovelamentos positivos à frente da forquilha de replicação. Para compensar, a DNA-girase desloca-se ao longo do DNA à frente da forquilha de replicação, e insere superenovelamentos negativos para cancelar o superenovelamento positivo.

Bactérias possuem um único local no cromossomo onde a síntese de DNA é iniciada, a origem de replicação (*oriC*). A *oriC* consiste em uma sequência específica de DNA, com cerca de 250 bases, que é reconhecida por proteínas de iniciação específicas, particularmente uma proteína denominada DnaA (Tabela 4.3), que se liga a essa região e promove a abertura da dupla-hélice. A próxima enzima na cadeia de replicação é a helicase (também conhecida por DnaB), cuja ligação



**Figura 4.14** Uma DNA-helicase desenrolando uma dupla-hélice. Nesta figura, a helicase está desnaturando ou separando as duas fitas antiparalelas de DNA, começando pela extremidade direita e seguindo para a esquerda.

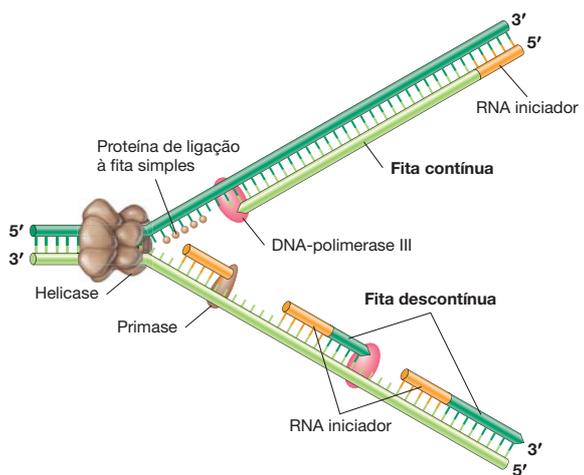
ao DNA é auxiliada pela proteína carregadora de helicase (DnaC). Duas helicases são carregadas, uma em cada fita, voltadas em direções opostas. Em seguida, duas primases, e então duas enzimas DNA-polimerases são carregadas no DNA, atrás das helicases. A iniciação da replicação do DNA é então iniciada nas duas fitas simples. À medida que a replica-

ção prossegue, a forquilha de replicação parece mover-se ao longo do DNA (Figura 4.14).

**Fitas contínua e descontínua**

A Figura 4.15 apresenta detalhes da replicação do DNA na forquilha de replicação. Uma importante distinção na replicação das duas fitas de DNA pode ser feita, devido ao fato de a replicação sempre ocorrer no sentido 5'-3' (5' → 3', havendo sempre a adição de um novo nucleotídeo à extremidade 3'-OH da cadeia em crescimento). Na fita que cresce a partir do 5'-PO<sub>4</sub><sup>2-</sup> em direção ao 3'-OH, denominada **fita contínua**, a síntese de DNA ocorre *ininterruptamente*, uma vez que sempre há uma extremidade 3'-OH livre na forquilha de replicação, à qual um novo nucleotídeo pode ser adicionado. Ao contrário, na fita oposta, denominada **fita descontínua**, a síntese de DNA ocorre de forma *descontínua*, porque não há uma extremidade 3'-OH livre na forquilha de replicação, à qual um novo nucleotídeo possa ser adicionado (Figura 4.15). A extremidade 3'-OH livre dessa fita está localizada na extremidade oposta, distante da forquilha de replicação. Portanto, na fita descontínua, é necessário que a primase sintetize iniciadores de RNA repetidas vezes, para fornecer os grupos 3'-OH livres para a DNA Pol III. Por outro lado, a fita contínua recebe um iniciador somente uma vez, na origem. Como resultado, a fita descontínua é sintetizada como segmentos curtos, denominados *fragmentos*

Tabela 4.3 Principais enzimas envolvidas na replicação do DNA em bactérias		
Enzima	Genes codificadores	Função
DNA-girase	<i>gyrAB</i>	Desenrola superenovelamentos à frente do replissomo
Proteína de ligação à origem	<i>dnaA</i>	Liga-se à origem de replicação para abrir a dupla-hélice
Carregador de helicase	<i>dnaC</i>	Carrega a helicase na origem
Helicase	<i>dnaB</i>	Desenrola a dupla-hélice na forquilha de replicação
Proteína de ligação à fita simples	<i>ssb</i>	Impede o anelamento das fitas simples
Primase	<i>dnaG</i>	Fornece o iniciador para novas fitas de DNA
DNA-polimerase III		Principal enzima de polimerização
Braçadeira deslizante	<i>dnaN</i>	Mantém a Pol III no DNA
Carregador da braçadeira	<i>hoIA-E</i>	Carrega a Pol III na braçadeira deslizante
Subunidade de dimerização (Tau)	<i>dnaX</i>	Mantém as duas enzimas cerne unidas nas fitas contínua e descontínua
Subunidade da polimerase	<i>dnaE</i>	Elongação da fita
Subunidade de revisão	<i>dnaQ</i>	Mecanismo de revisão
DNA-polimerase I	<i>polA</i>	Realiza a excisão do iniciador de RNA e o preenchimento das lacunas
DNA-ligase	<i>ligA, ligB</i>	Sela as quebras no DNA
Proteína Tus	<i>tus</i>	Liga-se à terminação, bloqueando o avanço da forquilha de replicação
Topoisomerase IV	<i>parCE</i>	Separação dos círculos interligados



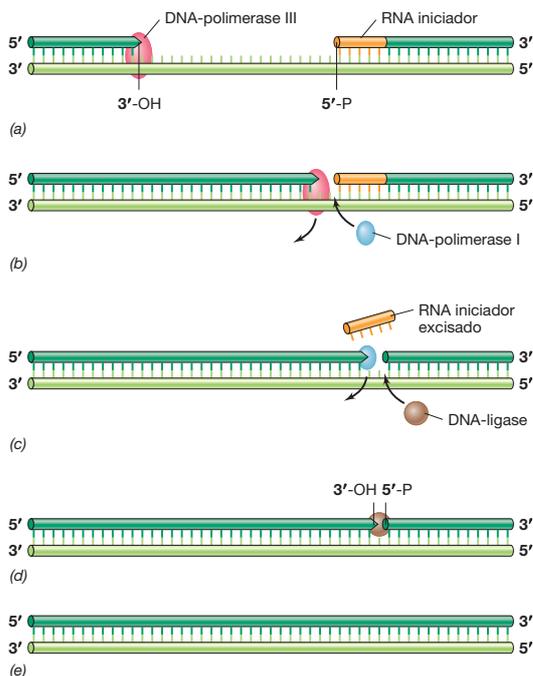
**Figura 4.15** Eventos que ocorrem na forquilha de replicação do DNA. Observe a polaridade e a natureza antiparalela das fitas de DNA.

de Okazaki, em homenagem a seu descobridor, Reiji Okazaki. Esses fragmentos da fita descontínua são posteriormente unidos, originando uma fita contínua de DNA.

### Síntese de novas fitas de DNA

Após sintetizar o RNA iniciador, a primase é substituída pela DNA Pol III. Essa enzima consiste em um complexo de várias proteínas (Tabela 4.3), incluindo a própria enzima cerne da polimerase. Cada molécula de polimerase é mantida no DNA por uma braçadeira deslizante, que circunda e desliza ao longo das fitas simples de DNA que atuam como molde. Consequentemente, a forquilha de replicação contém duas polimerases cerne e duas braçadeiras deslizantes, um conjunto para cada fita. Entretanto, há apenas um único complexo carregador de braçadeira, que atua na montagem das braçadeiras deslizantes no DNA. Após a montagem na fita descontínua, a atividade alongadora da DNA Pol III, catalisada pela DnaE, promove, então, a adição sequencial dos desoxirribonucleotídeos até que atinja o segmento de DNA previamente sintetizado (Figura 4.16). Nesse momento, cessa a atividade da DNA Pol III.

A próxima enzima a participar, a DNA Pol I, possui mais de uma atividade enzimática. Além de promover a síntese de DNA, a Pol I exibe atividade de exonuclease  $5' \rightarrow 3'$ , que remove o iniciador de RNA que o precede (Figura 4.16). Após



**Figura 4.16** União de dois fragmentos na fita descontínua. (a) A DNA-polimerase III promove a síntese de DNA no sentido  $5' \rightarrow 3'$ , em direção ao iniciador de RNA presente em um fragmento previamente sintetizado na fita descontínua. (b) Ao alcançar esse fragmento, a DNA-polimerase III é substituída pela DNA-polimerase I. (c) A DNA-polimerase I continua a síntese de DNA, enquanto retira o iniciador de RNA do fragmento prévio, e a DNA-ligase substitui a DNA-polimerase I após a remoção do iniciador. (d) A DNA-ligase une os dois fragmentos. (e) O produto final, DNA dupla-fita complementar e antiparalelo.

a remoção do iniciador e sua substituição por DNA, a DNA Pol I é liberada. A última ligação fosfodiéster é produzida por uma enzima denominada **DNA-ligase**. Essa enzima sela os cortes feitos no DNA que apresentam um  $5'-PO_4^{2-}$  e um  $3'-OH$  adjacentes (processo que a DNA Pol III não é capaz de realizar) e, juntamente com a DNA Pol I, também participa de processos de reparo de DNA. A DNA-ligase também desempenha importante papel na selagem de DNAs manipulados geneticamente, durante o processo de clonagem molecular (↔ Seção 11.4).

### MINIQUESTIONÁRIO

- Por que existem as fitas contínua e descontínua?
- Como a origem de replicação é reconhecida?
- Que enzimas participam na união dos fragmentos da fita descontínua?

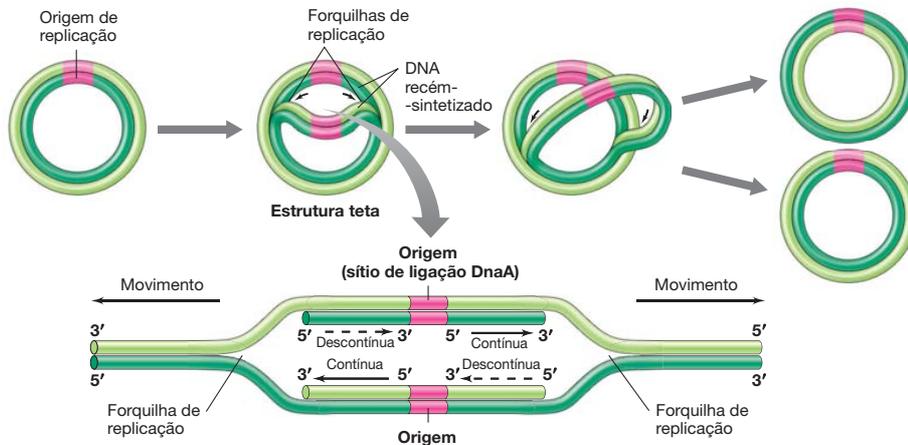
## 4.6 Replicação bidirecional e o replissomo

A natureza circular do cromossomo procariótico cria uma oportunidade de aceleração do processo de replicação. Em *Escherichia coli*, e provavelmente em todos os procariotos contendo cromossomos circulares, a replicação é *bidirecional* a partir da origem de replicação, conforme apresentado na Figura 4.17. Existem, desse modo, *duas* forquilhas de replicação em cada cromossomo, movendo-se em direções opostas. Essas forquilhas são mantidas unidas por duas subunidades da proteína Tau. No DNA circular, a replicação bidirecional leva à formação de estruturas características, denominadas estruturas teta (Figura 4.17).

Durante a replicação bidirecional, a síntese ocorre de modo contínuo e descontínuo em cada fita-molde, permitindo que o DNA se replique o mais rápido possível (Figura 4.17). Embora a DNA Pol III seja capaz de adicionar nucleotídeos a uma fita de DNA em crescimento a uma velocidade de 1.000 por segundo, a replicação do cromossomo de *E. coli* ainda demanda cerca de 40 min. Curiosamente, nas melhores condições de crescimento, *E. coli* é capaz de crescer com um tempo de geração de cerca de 20 min. A solução para esse enigma está no fato de as células de *E. coli*, crescendo com tempos de geração inferiores a 40 min, conterem múltiplas forquilhas de replicação. Isto é, um novo ciclo de replicação do DNA é iniciado antes do ciclo anterior ter sido completado. Esse problema será abordado em mais detalhes no Capítulo 5 (↔ Figura 5.4).

### O replissomo

A Figura 4.15 ilustra as diferenças na replicação das fitas contínua e descontínua e as enzimas que participam do processo. A partir desse diagrama simplificado, pode parecer que cada forquilha de replicação contém diversas proteínas distintas, atuando de forma independente. Na realidade, esse não é o caso. Em vez disso, as proteínas de replicação agregam-se, formando um grande complexo de replicação denominado **replissomo** (Figura 4.18). A fita descontínua do DNA forma uma alça, permitindo que o replissomo mova-se suavemente ao longo de ambas as fitas, e o replissomo literalmente puxa o DNA-molde à medida que a replicação se processa. Assim, é o DNA, e não a DNA-polimerase, que se movimenta durante a replicação. Observe também como a helicase e a primase formam um sub-



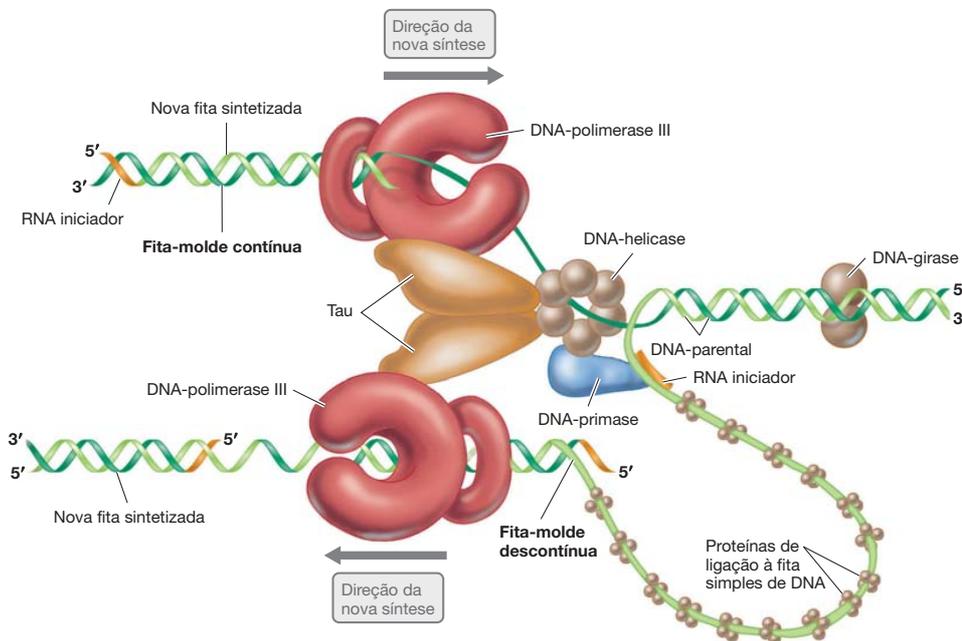
**Figura 4.17** Replicação de DNA circular: a estrutura teta. Em um DNA circular, a replicação bidirecional a partir de uma origem leva à formação de uma estrutura intermediária, semelhante à letra grega teta ( $\theta$ ). A inserção mostra duas forquilhas de replicação no cromossomo circular. Em *Escherichia*

*coli*, a origem de replicação é reconhecida por uma proteína específica, DnaA. Observe que a síntese de DNA está ocorrendo tanto na fita contínua quanto na descontinua em cada uma das novas fitas-filhas. Compare esta figura com a descrição de replissomo mostrada na Figura 4.18.

complexo, denominado *primossomo*, que auxilia sua ação em estreita associação durante o processo de replicação.

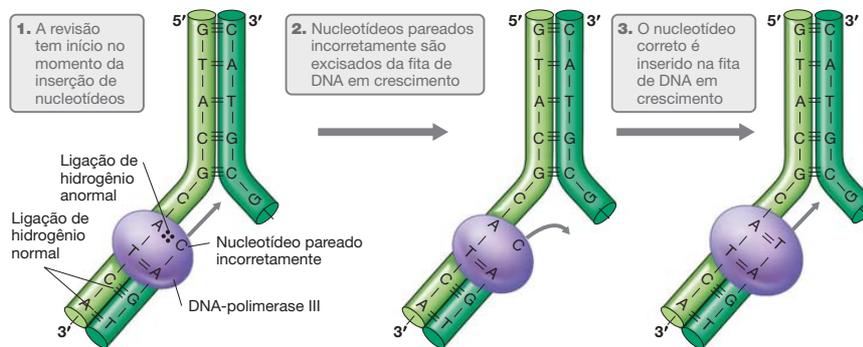
Em resumo, além da DNA Pol III, o replissomo contém várias proteínas essenciais de replicação: (1) DNA-girase, que remove os superenovelamentos; (2) DNA-helicase e primase

(o primossomo), que desenrola e adiciona um iniciador no DNA; e (3) a proteína de ligação à fita simples, que impede as fitas-molde separadas de formarem novamente uma dupla-hélice (Figura 4.18). A Tabela 4.3 resume as propriedades das proteínas essenciais à replicação do DNA em bactérias.



**Figura 4.18** O replissomo. O replissomo consiste em duas cópias de DNA-polimerase III e DNA-girase, além de helicase e primase (que em conjunto formam o primossomo), e muitas cópias da proteína de ligação à fita simples de DNA. As subunidades Tau mantêm os conjuntos de DNA-polimerases e helicase unidos. Imediatamente a montante do restante do replissomo, a DNA-

-girase remove os superenovelamentos no DNA que será replicado. Observe que as duas polimerases estão replicando as duas fitas individuais de DNA em direções opostas. Conseqüentemente, a fita-molde descontinua forma uma alça de modo que todo o replissomo se mova na mesma direção ao longo do cromossomo.



**Figura 4.19** Mecanismo de revisão mediado pela atividade de exonuclease 3' → 5' da DNA-polimerase III. Um pareamento incorreto no par de bases terminal induz a polimerase a realizar uma breve pausa.

Esse é o sinal para que o mecanismo de revisão promova a excisão do nucleotídeo incorreto, após o qual a base correta é inserida pela atividade da polimerase.

### Fidelidade da replicação do DNA: revisão

O DNA replica com uma taxa de erro extremamente baixa. No entanto, quando os erros ocorrem, existe um mecanismo para detectá-los e corrigi-los. Erros durante a replicação de DNA introduzem *mutações*, alterações na sequência do DNA. As taxas de mutação nas células são extremamente baixas, entre  $10^{-8}$  a  $10^{-11}$  erros por par de bases inserido. Essa precisão é possível, em parte, porque as DNA-polimerases têm duas oportunidades para incorporar a base correta em um determinado sítio. A primeira oportunidade ocorre após a inserção de bases complementares às bases opostas da fita-molde pela DNA Pol III, de acordo com as regras de pareamento de bases, A com T e G com C. A segunda oportunidade depende de uma segunda atividade enzimática da DNA Pol I e Pol III, denominada *mecanismo de revisão* (Figura 4.19). No caso da DNA Pol III, uma subunidade proteica distinta, DnaQ, é responsável pela revisão, enquanto na DNA Pol I, uma única proteína realiza a polimerização e a revisão.

A atividade de revisão ocorre quando uma base incorreta foi inserida, uma vez que isso origina um pareamento incorreto de bases. Tanto a DNA Pol I quanto a DNA Pol III possuem atividade exonuclease 3' → 5' que consegue remover os nucleotídeos inseridos de forma incorreta. A polimerase percebe esse fato porque um nucleotídeo inserido incorretamente acarreta uma discreta distorção na dupla-hélice. Após a remoção do nucleotídeo pareado incorretamente, a polimerase tem uma segunda oportunidade de inserir o nucleotídeo correto (Figura 4.19). A atividade exonucleásica de revisão é distinta da atividade de exonuclease 5' → 3' da DNA Pol I, utilizada para remover o iniciador de RNA das fitas contínuas e descontínuas. Apenas a DNA Pol I possui essa última atividade. A revisão por exonuclease ocorre em sistemas de replicação

de DNA em procariotos, eucariotos e vírus. Todavia, muitos organismos apresentam mecanismos adicionais para reduzir os erros realizados durante a replicação de DNA, que operam após a passagem da forquilha de replicação. Serão discutidos alguns desses mecanismos no Capítulo 10.

### Terminação da replicação

Eventualmente, o processo de replicação do DNA é concluído. Como o replissomo reconhece o momento de parar? No lado oposto, em relação à origem, de um cromossomo circular, há um sítio denominado *terminação da replicação*. Nesse ponto, as duas forquilhas de replicação colidem, à medida que os novos círculos de DNA são completados. Na região de terminação encontram-se várias sequências de DNA, denominadas sítios *Ter*, que são reconhecidas por uma proteína denominada Tus, cuja função é bloquear o avanço das forquilhas de replicação. Quando a replicação de um cromossomo circular chega ao fim, as duas moléculas circulares encontram-se ligadas, como elos de uma corrente. Sua separação é catalisada por outra enzima, a topoisomerase IV. Obviamente, é importante que após a replicação do DNA este seja segregado de maneira que cada célula-filha receba uma cópia do cromossomo. Esse processo pode ser auxiliado por FtsZ, uma importante proteína de divisão celular, que auxilia na coordenação de vários eventos essenciais da divisão celular (↔ Seção 5.2).

### MINIQUESTIONÁRIO

- O que é o replissomo e quais são seus componentes?
- Como a atividade de revisão é realizada durante a replicação do DNA?
- Como as atividades do replissomo são interrompidas?

## III • Síntese de RNA: transcrição

**T**ranscrição é a síntese de ácido ribonucleico (RNA) utilizando DNA como molde. Existem três diferenças essenciais na química do RNA e do DNA: (1) o RNA contém o açúcar-ribose em vez de desoxirribose; (2) o RNA contém a base uracila no lugar de timina; e (3) exceto no caso de alguns vírus, o RNA não é encontrado na forma de dupla-fita. A substituição da

desoxirribose pela ribose afeta as propriedades químicas de um ácido nucleico; as enzimas que atuam sobre o DNA geralmente não exibem qualquer efeito no RNA, e vice-versa. Entretanto, a substituição da timina pela uracila não afeta o pareamento das bases, uma vez que essas bases pareiam-se com a adenina com a mesma eficiência.

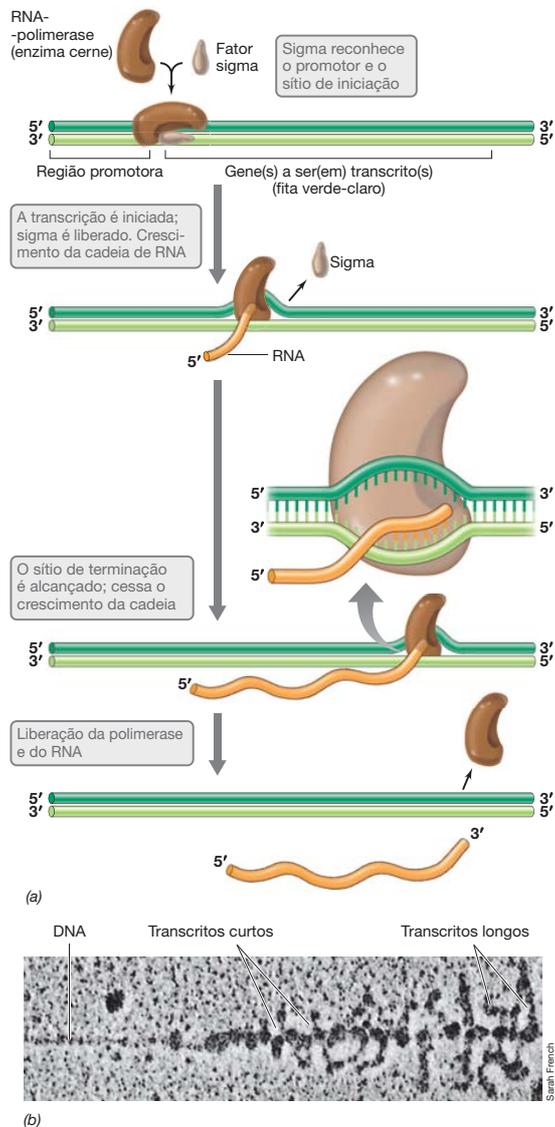
Enquanto o RNA existe predominantemente na forma de fita simples, as moléculas normalmente se dobram sobre si mesmas, em regiões onde o emparelhamento de bases complementares é possível. O termo **estrutura secundária** refere-se a esse processo de dobragem, e o termo **estrutura primária** refere-se à sequência nucleotídica, da mesma forma que no DNA. A estrutura secundária origina moléculas de RNA altamente dobradas e torcidas, cuja função biológica depende criticamente da sua forma tridimensional final.

O RNA desempenha diversos papéis importantes na célula. Como visto anteriormente (Figura 4.3), três tipos principais de RNA estão envolvidos na síntese proteica: RNA mensageiro (RNAm), RNA transportador (RNAt) e RNA ribossomal (RNAr). Vários outros tipos de RNA são também conhecidos, estando a maioria envolvida na regulação (Capítulo 7). Todas as moléculas de RNA são produtos da transcrição do DNA. Deve-se enfatizar que o RNA atua em dois níveis, genético e funcional. No nível genético, o RNAm carrega a informação genética do genoma para o ribossomo. Em contrapartida, o RNAr desempenha papel funcional e estrutural nos ribossomos, e o RNAt desempenha um papel ativo no transporte de aminoácidos para a síntese proteica. Além disso, algumas moléculas de RNA, incluindo RNAr, podem apresentar atividade enzimática. Nesta seção, será analisado como o RNA é sintetizado em espécies de bactérias, utilizando novamente *Escherichia coli* como nosso organismo-modelo.

## 4.7 Transcrição

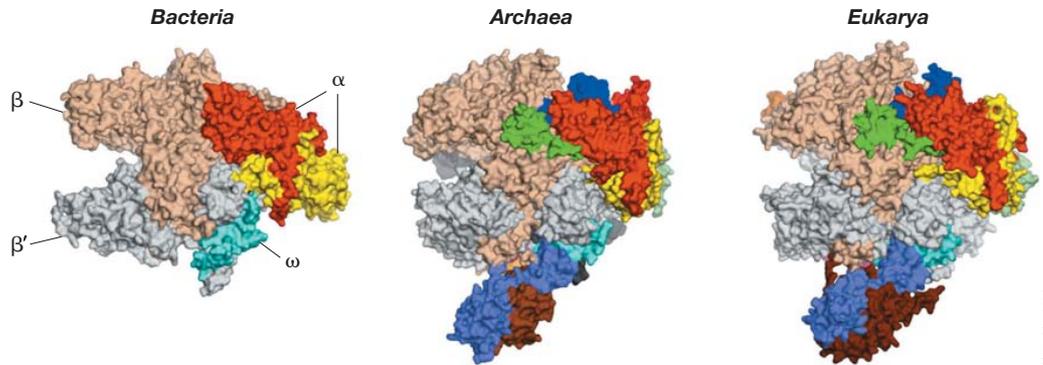
A transcrição da informação genética é realizada pela enzima **RNA-polimerase**. Assim como a DNA-polimerase, a RNA-polimerase catalisa a formação de ligações fosfodiéster, porém, nesse caso, entre ribonucleotídeos rATP, rGTP, rCTP e rUTP, em vez de desoxirribonucleotídeos. A polimerização é conduzida pela energia liberada a partir da hidrólise de duas ligações fosfato ricas em energia, presentes nos ribonucleosídeos trifosfato que estão sendo adicionados. O mecanismo de síntese de RNA é muito semelhante àquele da síntese de DNA (Figura 4.12): durante a elongação de uma cadeia de RNA, os ribonucleosídeos-trifosfato são adicionados ao grupo 3'-OH da ribose do nucleotídeo precedente. Assim, a direção global do crescimento da cadeia ocorre a partir da extremidade 5' em direção à extremidade 3', e a fita recém-sintetizada é antiparalela à fita-molde de DNA, a partir da qual o RNA foi transcrito. O processo global da síntese de RNA é ilustrado na **Figura 4.20**.

A RNA-polimerase utiliza DNA dupla-fita como molde, porém apenas uma das duas fitas é transcrita para cada gene. No entanto, genes estão presentes em ambas as fitas do DNA e, portanto, as sequências de DNA das duas fitas são transcritas, embora em localizações distintas. Contrariamente à DNA-polimerase, a RNA-polimerase é capaz de iniciar novas fitas de nucleotídeos de forma independente; consequentemente, não há a necessidade de um iniciador. À medida que o RNA recém-sintetizado dissocia-se do DNA, o DNA aberto fecha-se de volta para a dupla-hélice original. A transcrição é interrompida em sítios específicos denominados **terminadores de transcrição**. Diferentemente da replicação do DNA, que copia genomas inteiros, a transcrição copia unidades de DNA muito menores, frequentemente correspondendo



**Figura 4.20 Transcrição.** (a) Etapas da síntese de RNA. Os sítios de iniciação (promotor) e de terminação correspondem a sequências nucleotídicas específicas no DNA. A RNA-polimerase move-se ao longo da cadeia de DNA, promovendo a abertura temporária da dupla-hélice, transcrevendo uma das fitas do DNA. (b) Micrografia eletrônica ilustra a transcrição de um gene do cromossomo de *Escherichia coli*. A transcrição está ocorrendo da esquerda para a direita, com os transcritos mais curtos, à esquerda, tornando-se mais longos à medida que a transcrição prossegue.

a um único gene. Esse sistema permite à célula transcrever genes diferentes, em frequências diferentes, dependendo das necessidades da célula para diferentes proteínas. Em outras palavras, a expressão gênica é regulada. Como será visto no Capítulo 7, a regulação da transcrição é um processo importante e elaborado, que utiliza muitos mecanismos diferentes, e é de extrema eficiência no controle da expressão gênica e conservação dos recursos celulares.



**Figura 4.21** RNA-polimerase dos três domínios. Representação da superfície das multissubunidades das estruturas da RNA-polimerase celular de *Bacteria* (esquerda, enzima cerne de *Thermus aquaticus*), *Archaea* (centro, *Sulfolobus solfataricus*) e *Eukarya* (direita, RNA Pol II de *Saccharomyces*

*cerevisiae*). Subunidades ortólogas são destacadas pela mesma coloração. Uma subunidade específica na RNA-polimerase de *S. solfataricus* não é demonstrada nesta figura.

### RNA-polimerase

A RNA-polimerase de *Bacteria*, que possui a estrutura mais simples e sobre a qual muito é conhecido, possui cinco subunidades diferentes, denominadas  $\beta$ ,  $\beta'$ ,  $\alpha$ ,  $\omega$  (ômega) e  $\sigma$  (sigma), estando a subunidade  $\alpha$  presente em duas cópias. As subunidades  $\beta$  e  $\beta'$  (beta linha) são similares, mas não idênticas (Figura 4.21). As subunidades interagem, formando a enzima ativa, denominada holoenzima da RNA-polimerase, porém o fator sigma não se encontra tão fortemente ligado como as demais subunidades, podendo dissociar-se facilmente, levando à formação da enzima cerne da RNA-polimerase,  $\alpha_2\beta\beta'\omega$ . A enzima cerne sozinha sintetiza o RNA, enquanto o fator sigma reconhece o sítio apropriado para a o início da síntese de RNA, no DNA. A subunidade ômega é necessária à montagem da enzima cerne, porém não é requerida para a síntese de RNA. Em *Bacteria*, o fator sigma dissocia-se da holoenzima da RNA-polimerase bacteriana, logo que um segmento pequeno de RNA tenha sido formado (Figura 4.20). A elongação da molécula de RNA é então catalisada pela enzima cerne sozinha. A subunidade sigma somente é necessária para a formação do complexo inicial RNA-polimerase-DNA no promotor.

### Promotores

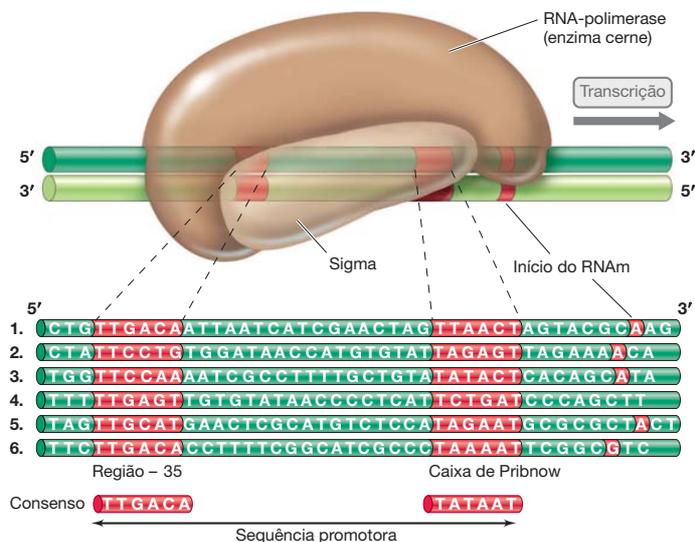
Para iniciar a síntese de RNA corretamente, é necessário que a RNA-polimerase, primeiramente, reconheça os sítios de iniciação no DNA, denominados **promotores** (Figura 4.20). Em bactérias, os promotores são reconhecidos pela subunidade sigma da RNA-polimerase. Uma vez que a RNA-polimerase encontra-se ligada ao promotor, a transcrição pode ocorrer (Figura 4.20). Nesse processo, a dupla-hélice de DNA na região do promotor é aberta pela RNA-polimerase, originando uma bolha de transcrição. À medida que a RNA-polimerase se desloca, desenrola o DNA em curtos segmentos. Esse desenovelamento transitório expõe a fita-molde e permite que ela seja copiada no complemento de RNA. Dessa forma, os promotores podem ser considerados como “direcionadores” da RNA-polimerase para uma direção ou outra ao longo do DNA. Quando uma região do DNA apresenta dois promotores próximos, com direções opostas, a transcrição a partir de

um dos promotores ocorre em uma direção (em uma das fitas de DNA), enquanto a transcrição a partir do outro promotor ocorre na direção oposta (na outra fita).

### Fatores sigma e sequências-consenso

Os promotores são sequências específicas de DNA às quais a RNA-polimerase se liga, e a Figura 4.22 apresenta a sequência de alguns promotores de *Escherichia coli*. Todas essas sequências são reconhecidas pelo mesmo fator sigma em *E. coli*, denominado  $\sigma^{70}$  (o número 70 sobrescrito indica o tamanho dessa proteína, 70 quilodaltos); embora essas sequências não sejam idênticas, a subunidade sigma reconhece duas sequências menores altamente conservadas nos promotores. Essas sequências conservadas estão a montante do sítio onde a transcrição é iniciada. Uma está localizada 10 bases antes do sítio de início da transcrição, a região  $-10$ , ou “caixa de Pribnow”. Embora os promotores sejam ligeiramente diferentes, a comparação de muitas regiões  $-10$  nos fornece a sequência-consenso: TATAAT. A segunda região conservada no promotor encontra-se a aproximadamente 35 bases a montante do sítio de início da transcrição. A sequência-consenso na região  $-35$  corresponde a TTGACA (Figura 4.22). Novamente, a maioria dos promotores é ligeiramente diferente, mas são muito próximos à sequência-consenso.

Em *E. coli*, os promotores que apresentam maior similaridade com a sequência-consenso são geralmente mais eficientes na ligação à RNA-polimerase. Esses promotores mais eficientes são denominados *promotores fortes*, sendo de grande utilidade na engenharia genética, conforme discutido no Capítulo 11. Embora a maioria dos genes de *E. coli* necessitem do fator sigma padrão,  $\sigma^{70}$  (RpoD), para a transcrição, existem diversos fatores sigma alternativos que reconhecem diferentes sequências-consenso (Tabela 4.4). Cada fator sigma alternativo é específico para um grupo de genes necessários em circunstâncias especiais e, portanto, essenciais para a regulação da expressão gênica. Consequentemente, é possível controlar a expressão de diferentes famílias gênicas regulando a presença ou ausência do fator sigma correspondente, e isso ocorre alterando a taxa de síntese ou degradação do fator sigma.



**Figura 4.22** A interação da RNA-polimerase com um promotor bacteriano. Apresentadas abaixo da RNA-polimerase e do DNA encontram-se seis sequências promotoras diferentes identificadas em *Escherichia coli*. São apresentados os contatos da RNA-polimerase com a região - 35 e com a caixa de Pribnow (sequência - 10). A transcrição inicia-se em uma base específica, imediatamente a jusante à caixa de Pribnow. Abaixo das sequências reais das regiões - 35 e da caixa de Pribnow estão as sequências-consenso, derivadas de comparações realizadas entre muitos promotores. Observe que, embora sigma reconheça as sequências do promotor na fita 5' → 3' do DNA (em verde-escuro), a enzima cerne da RNA-polimerase transcreve, de fato, a fita em verde-claro, com sentido 3' → 5', uma vez que a enzima cerne atua apenas na direção 5' → 3'.

### Terminação da transcrição

Em uma célula bacteriana em crescimento, apenas aqueles genes que necessitam ser expressos são normalmente transcritos. Portanto, é importante que a transcrição seja concluída no local correto. A **terminação** da síntese de RNA é governada

por sequências de bases específicas presentes no DNA. Em bactérias, um sinal comum de terminação no DNA corresponde a uma sequência rica em GC, contendo uma repetição invertida com um segmento central não repetido. Quando tal tipo de sequência de DNA é transcrita, o RNA forma uma estrutura em haste-alça, devido a pareamento intrafita de bases. Essas estruturas em haste-alça, seguidas por uma série de adeninas no DNA-molde (e consequentemente, por várias uridinas no RNAm), consistem em eficientes terminadores transcricionais. Esse fato deve-se à formação de um segmento de pares de bases U:A que mantém o RNA e o DNA-molde unidos. Essa estrutura é bastante fraca, pois os pares de bases U:A possuem apenas duas ligações de hidrogênio cada. A RNA-polimerase realiza uma pausa na estrutura em haste-alça, havendo a separação do DNA e do RNA na região contendo a série de uridinas, terminando a transcrição.

O outro mecanismo para a terminação da transcrição em bactérias utiliza um fator proteico específico, conhecido como Rho. O fator Rho não se liga à RNA-polimerase ou ao DNA, porém liga-se fortemente ao RNA e desloca-se para baixo na cadeia em direção ao complexo RNA-polimerase-DNA. Quando a RNA-polimerase realiza uma pausa no sítio de terminação Rho-dependente (uma sequência específica no DNA-molde), esse fator causa a liberação do RNA e da RNA-polimerase do DNA, finalizando a transcrição.

### MINIQUESTIONÁRIO

- Em qual direção ao longo da fita-molde de DNA a transcrição ocorre, e qual enzima catalisa essa reação?
- O que é um promotor? Qual proteína reconhece os promotores em *Escherichia coli*?
- De que forma a expressão de famílias gênicas pode ser controlada grupalmente?
- Quais tipos de estruturas podem levar ao término da transcrição?

**Tabela 4.4** Fatores sigma em *Escherichia coli*

Nome <sup>a</sup>	Sequência de reconhecimento a jusante <sup>b</sup>	Função
σ <sup>70</sup> RpoD	TTGACA	Para a maioria dos genes, o fator sigma principal para o crescimento normal
σ <sup>54</sup> RpoN	TTGGCACA	Assimilação de nitrogênio
σ <sup>38</sup> RpoS	CCGGCG	Fase estacionária, além de estresse oxidativo e osmótico
σ <sup>32</sup> RpoH	TNTCNCCTTGAA	Resposta de choque térmico
σ <sup>28</sup> FliA	TAAA	Para genes envolvidos na síntese flagelar
σ <sup>24</sup> RpoE	GAACCT	Resposta a proteínas dobradas incorretamente no periplasma
σ <sup>19</sup> Fecl	AAGGAAAAT	Para determinados genes no transporte de ferro

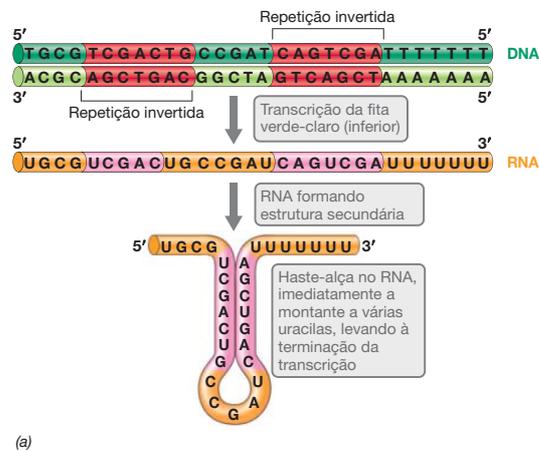
<sup>a</sup>O número sobrescrito indica o tamanho da proteína em quilodáltons. Muitos fatores também possuem outros nomes, por exemplo, σ<sup>70</sup> também é chamado de σ<sup>D</sup>.  
<sup>b</sup>N = qualquer nucleotídeo.

### 4.8 A unidade de transcrição

A informação genética é organizada em unidades de transcrição. Essas unidades são segmentos de DNA que são transcritos em uma única molécula de RNA. Cada unidade de transcrição está conectada por sítios onde a transcrição é iniciada e terminada. Algumas unidades de transcrição contêm apenas um único gene. Outras contêm dois ou mais genes. Esses últimos são referidos como *cotranscritos*, originando uma única molécula de RNA.

### RNA ribossomal e transportador e a longevidade do RNA

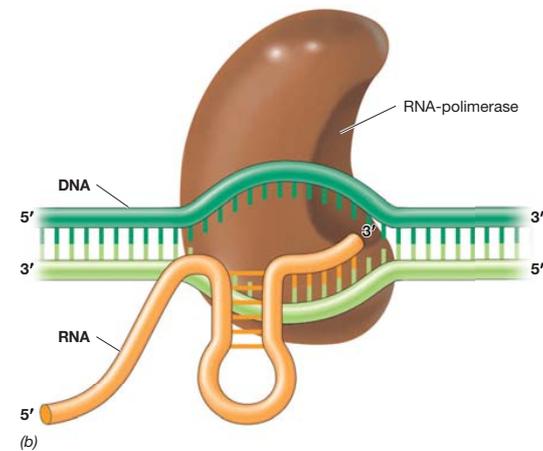
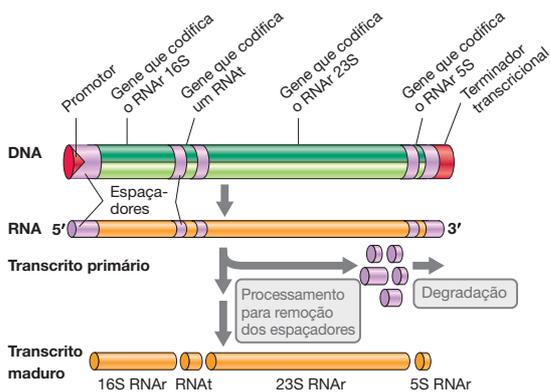
A maioria dos genes codifica proteínas, porém outros codificam RNA que não são traduzidos, como o RNA ribossomal ou o RNA transportador. Existem vários tipos diferentes de RNAr em um organismo. Bactérias e arqueias possuem três tipos: RNAr 16S, RNAr 23S e RNAr 5S (com um ribossomo apre-



**Figura 4.23** Repetições invertidas e a terminação da transcrição. (a) As repetições invertidas no DNA transcrito promovem a formação de uma estrutura em haste-alça no RNA, que promove a terminação da transcrição,

sendando um cópia de cada; Seção 4.14). Conforme ilustrado na **Figura 4.24**, existem unidades de transcrição que contêm um gene de cada um desses RNA, sendo esses genes cotranscritos. Dessa forma, a unidade de transcrição da maioria dos RNAr é mais extensa que um único gene. Em procariotos, os genes de RNAt são frequentemente cotranscritos entre si, ou mesmo com genes de RNAr, conforme apresentado na Figura 4.24. Esses cotranscritos são processados por proteínas celulares específicas, que os clivam em unidades individuais, gerando RNAr ou RNAt maduros (funcionais).

Em procariotos, a maioria dos RNA mensageiros possui meia-vida curta (da ordem de poucos minutos), após a qual são degradados por enzimas denominadas *ribonucleases*. Isso



quando seguida por uma série de uracilas. (b) Diagrama indicando a formação da haste-alça terminadora no RNA dentro da RNA-polimerase.

não é observado nos RNAr e RNAt, que são RNA estáveis. Essa estabilidade pode ser atribuída às estruturas secundárias dos RNAt e RNA que os impedem de serem degradadas por ribonucleases. Por outro lado, o RNAm normal não forma tais estruturas, sendo suscetível a ataque por ribonucleases. A rápida reciclagem dos RNAm procarióticos permite à célula adaptar-se rapidamente a novas condições ambientais, e interrompe a tradução de RNAm cujos produtos não são mais necessários.

### RNAm policistrônico e o óperon

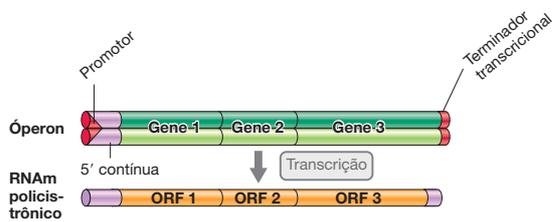
Em procariotos, os genes que codificam diversas enzimas de uma via metabólica em particular, por exemplo, a biossíntese de um determinado aminoácido, são geralmente agrupados. A RNA-polimerase move-se ao longo de tais agrupamentos e transcreve toda a série de genes, originando uma única longa molécula de RNAm. Um RNAm que codifica tal grupo de genes cotranscritos é denominado *RNAm policistrônico* (**Figura 4.25**). RNAm policistrônicos possuem múltiplas *janelas abertas de leitura*, porções do RNAm que de fato codificam aminoácidos (Seção 4.11). Quando esse RNAm é traduzido, vários polipeptídeos são sintetizados, um após o outro, pelo mesmo ribossomo.

Um grupo de genes relacionados, transcritos em conjunto, gerando um único RNAm policistrônico, é referido como um óperon. A organização de genes de uma mesma via bioquímica ou de genes necessários, sob as mesmas condições, em um óperon, permite sua expressão coordenada. Frequentemente, a transcrição de um óperon é controlada por uma região específica do DNA, situada imediatamente a montante à região codificadora de proteínas do óperon. Esse assunto será abordado em mais detalhes no Capítulo 7.

**Figura 4.24** Uma unidade de transcrição do RNAr ribossomal de *Bacteria* e seu processamento subsequente. Em bactérias, todas as unidades de transcrição de RNAr possuem os genes de RNAr na ordem RNAr 16S, RNAr 23S e RNAr 5S (apresentados aproximadamente em escala). Observe que, nessa unidade de transcrição em particular, o “espaçador” entre os genes de RNAr 16S e 23S contém um gene de RNAt. Em outras unidades de transcrição, esta região pode conter mais de um gene de RNAt. Frequentemente, um ou mais genes de RNAt também encontram-se após o gene do RNAr 5S, sendo cotranscritos. *Escherichia coli* possui sete unidades de transcrição de RNAr.

### MINIQUESTIONÁRIO

- Qual o papel do RNA mensageiro (RNAm)?
- O que é uma unidade de transcrição? O que é um RNAm policistrônico?
- O que são óperons e por que eles são úteis aos procariotos?



**Figura 4.25** Óperon e estrutura do RNAm policistrônico em procariontes. Observe que um único promotor controla os três genes dentro do óperon e que a molécula de RNAm policistrônico contém uma fase de leitura aberta (ORF) correspondente para cada gene.

## 4.9 Transcrição em arqueias e eucariotos

Até agora analisou-se a transcrição em bactéria utilizando *Escherichia coli* como sistema-modelo. Embora em arqueias e eucariotos o fluxo geral da informação genética do DNA ao RNA seja o mesmo observado em bactérias, alguns detalhes diferem, e nas células eucarióticas a presença do núcleo complica o encaminhamento da informação genética. Embora arqueias careçam de um núcleo, muitas de suas propriedades moleculares assemelham-se mais a eucariotos do que a bactérias. Essas características do dogma central compartilhadas confirmam que esses dois domínios são mais estreitamente relacionados entre si, do que qualquer um é com bactérias (↪ Seção 1.3). No entanto, arqueias também compartilham similaridades transcricionais com bactérias, como os óperons. Unidades de transcrição em eucariotos incluem apenas um gene. Aqui, são discutidos elementos de transcrição essenciais em arqueias e eucariotos que diferem daqueles de bactérias.

### RNA-polimerase de arqueias e eucariotos

As RNA-polimerases de arqueias e eucariotos são mais similares e estruturalmente mais complexas do que as de bactérias. As arqueias possuem apenas uma única RNA-polimerase, muito semelhante à RNA-polimerase II de eucariotos. A RNA-polimerase de arqueias normalmente é composta por 11 ou 12 subunidades, enquanto a RNA-polimerase II de eucariotos apresenta 12 ou mais subunidades. Isso está em nítido contraste com a RNA-polimerase de bactérias, que possui apenas quatro subunidades diferentes, além da subunidade sigma (de reconhecimento) (Figura 4.21).

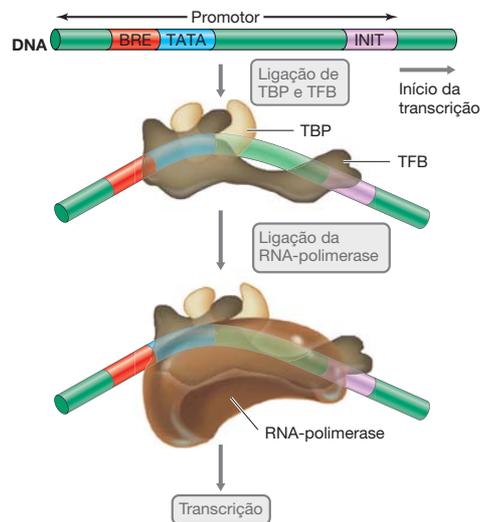
Apreendeu-se na Seção 4.7 a importância do promotor para a transcrição. A estrutura dos promotores de arqueias assemelha-se mais aos promotores eucarióticos, reconhecidos pela RNA-polimerase II eucariótica, do que aos promotores de bactérias. Eucariotos diferem-se de arqueias e bactérias por apresentarem múltiplas RNA-polimerases. Existem no núcleo três RNA-polimerases separadas, que transcrevem diferentes categorias de genes. As mitocôndrias e os cloroplastos também apresentam RNA-polimerases específicas, mas não surpreendentemente, considerando as conexões filogenéticas entre bactérias e organelas celulares eucarióticas (↪ Figura 1.6b), essas estão mais estreitamente relacionadas com a RNA-polimerase de bactérias.

### Promotores e terminadores em arqueias e eucariotos

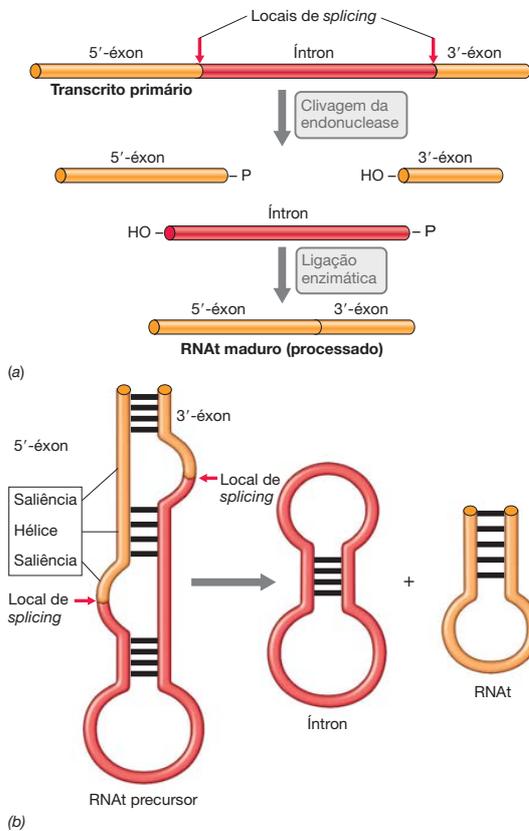
Três sequências de reconhecimento principais são encontradas em ambos os domínios procarióticos, e essas sequências

são reconhecidas por uma série de proteínas denominadas *fatores transcricionais*, similares em eucariotos e arqueias. A sequência de reconhecimento mais importante de promotores de arqueias e eucariotos consiste no “TATA” *box*, de 6 a 8 pares de bases, localizando 18 a 27 nucleotídeos a montante do sítio de início da transcrição (Figura 4.26). Esse é reconhecido pela *proteína de ligação ao TATA* (TBP, *TATA-binding protein*). A montante ao TATA *box* está a sequência do *elemento de reconhecimento B* (BRE, *B recognition element*), que é reconhecida pelo fator transcricional B (TFB, *transcription factor B*). Além disso, uma sequência de elemento iniciador se encontra localizada no ponto de início da transcrição. Uma vez que TBP tenha se ligado ao TATA *box* e TFB tenha se ligado ao BRE, a RNA-polimerase de arqueias pode ligar-se, iniciando a transcrição. Esse processo é similar ao observado em eucariotos, exceto pelo fato desses organismos utilizarem diversos fatores transcricionais adicionais.

Pouco é conhecido sobre a terminação da transcrição em arqueias e eucariotos, com relação ao que se conhece sobre bactérias (Seção 4.7). Alguns genes de arqueias possuem repetições invertidas seguidas por uma sequência rica em AT, similares àquelas observadas em muitos terminadores transcricionais bacterianos. Todavia, essas sequências terminadoras não são encontradas em todos os genes de arqueias. Um tipo possível de terminador transcricional é desprovido das repetições invertidas, porém contém segmentos repetidos de timinas. De alguma forma, essa sequência sinaliza à maquinaria de terminação de arqueias a necessidade de interromper o processo de transcrição. Em eucariotos, a terminação difere dependendo da RNA-polimerase, e frequentemente requer um fator proteico de terminação específico. Proteínas do tipo Rho (Seção 4.7) não foram identificadas em arqueias ou eucariotos.



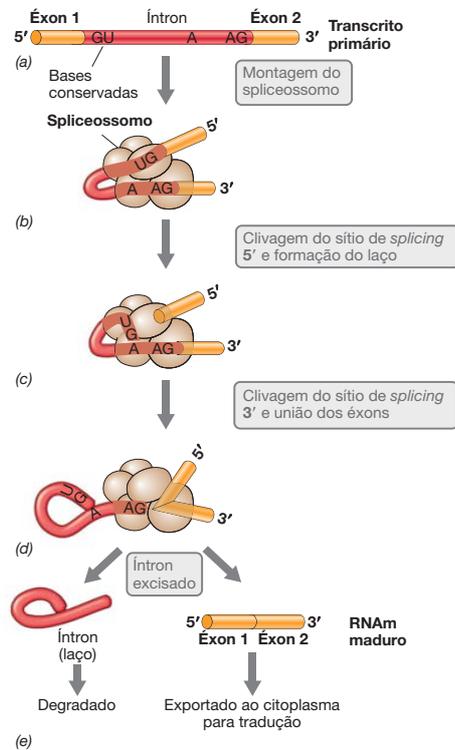
**Figura 4.26** Arquitetura do promotor e transcrição em arqueias. Três elementos promotores são cruciais para o reconhecimento do promotor em arqueias: o elemento iniciador (INIT), o TATA *box*, e o elemento de reconhecimento B (BRE). A proteína de ligação ao TATA (TBP) se liga ao TATA *box*; o fator transcricional B (TFB) se liga a BRE e INIT. Uma vez que TBP e TFB estejam posicionadas, a RNA-polimerase se liga.



**Figura 4.27 Splicing dos introns de arqueias.** (a) Esquema da reação. A remoção dos introns de arqueias é uma reação de duas etapas. Na primeira etapa, uma endonuclease específica excisa o íntron. Na segunda etapa, uma ligase une o 5'-éxon ao 3'-éxon, originando o RNAt maduro e processado. (b) Dobramento do RNAt precursor. Os dois locais de *splicing* (setas vermelhas) são reconhecidos por seus motivos característicos, "saliência-hélice-saliência". Os produtos da reação são o RNAt e um íntron circular.

**Sequências intervenientes em arqueias**

Assim como em bactérias, sequências intervenientes em genes codificadores de proteínas são extremamente raras em arqueias. Esta é uma característica que difere dos eucariotos, em que muitos destes genes são interrompidos em duas ou mais regiões codificadoras, separadas por regiões não codificadoras. Essas moléculas de RNA requerem alterações – conhecidas como **processamento do RNA** – antes de se tornarem *maduras*; ou seja, prontas para desempenharem seu papel na célula. Os segmentos de sequências codificadoras são denominados **éxons**, enquanto as sequências intervenientes, não codificadoras, recebem o nome de **íntrons**. O termo **transcrito primário** refere-se à molécula de RNA originalmente transcrita, antes de os íntrons serem removidos, a fim de gerar o RNAm final, consistindo-se somente em éxons. Vários genes de arqueias que codificam RNAt e RNAr possuem íntrons que devem ser removidos após a transcrição, a fim de gerar os RNAs ou RNAr maduros. Esses íntrons são denominados íntrons de arqueias, uma vez que são processados por um mecanismo diferente daquele que processa os íntrons eucarióticos

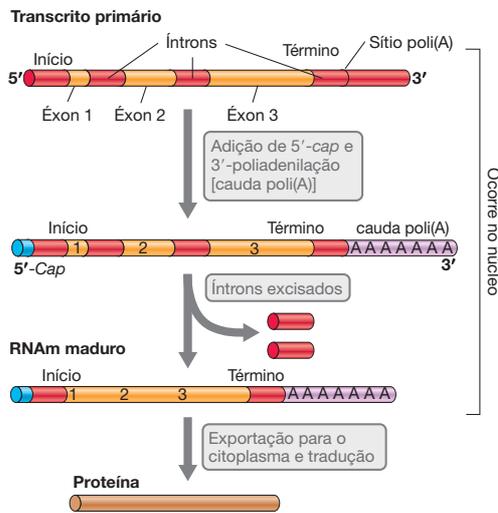


**Figura 4.28 Atividade do spliceossomo.** A remoção de um íntron do transcrito primário de um gene codificador de proteína em um eucarioto. (a) Transcrito primário contendo um único íntron. A sequência GU é conservada no sítio de *splicing* 5', e a sequência AG, no sítio de *splicing* 3'. Existe também um A interno que atua como ponto de ramificação. (b) Várias pequenas partículas de ribonucleoproteínas (ilustradas em marrom) organizam-se no RNA, formando o *spliceossomo*. Cada uma dessas partículas contém pequenos RNA distintos que atuam no mecanismo de *splicing*. (c) O sítio de *splicing* 5' foi clivado, com a formação simultânea de um ponto de ramificação. (d) O sítio de *splicing* 3' foi clivado e os dois éxons unidos. Observe que, globalmente, duas ligações fosfodiéster foram rompidas, porém outras duas foram formadas. (e) Os produtos finais consistem nos éxons unidos (o RNAm) e no íntron liberado.

típicos. Os introns de arqueias são excisados por uma ribonuclease específica que reconhece as junções éxon-íntron (Figura 4.27). Em alguns casos, RNAs em arqueias são montados por meio da união em conjunto de segmentos de dois ou três transcritos primários diferentes.

**Processamento de RNA em eucariotos**

A maioria dos transcritos primários em eucariotos contém íntrons e, dessa forma, requer processamento do RNA antes que esses possam desempenhar seu papel na célula. O processo pelo qual os íntrons são removidos e os éxons unidos em eucariotos é denominado *splicing* (Figura 4.28). O *splicing* do RNA ocorre no núcleo e é realizado por um grande complexo macromolecular, denominado **spliceossomo**. As proteínas do spliceossomo removem os íntrons e unem os éxons adjacentes, originando uma sequência codificadora de proteína contígua no RNAm maduro. Muitos genes, especialmente em animais superiores e plantas, possuem múltiplos íntrons, sendo dessa forma claramente importante que todos eles sejam reconheci-



**Figura 4.29** Processamento do transcrito primário em RNAm maduro em eucariotos. As etapas de processamento incluem a adição do *cap* na extremidade 5', remoção dos introns e encurtamento da extremidade 3' do transcrito, com a concomitante adição da cauda poli(A). Todas essas etapas são realizadas no núcleo. Os locais referentes aos códons de início e término utilizados na tradução estão indicados.

dos e removidos pelos spliceossomos a fim de gerar o RNAm maduro final (Figura 4.29).

Existem duas outras etapas no processamento do RNAm em *Eukarya* que são exclusivas a este domínio. Ambas as etapas ocorrem no núcleo antes do *splicing*. A primeira, denominada adição de *cap*, ocorre antes do término da transcrição. A adição de *cap* consiste na adição de um nucleotídeo de guanina metilado, na extremidade 5'-fosfato do RNAm (Figura 4.29). Esse nucleotídeo *cap* é adicionado na direção oposta em relação ao restante da molécula de RNAm e é necessário para iniciar a tradução. A segunda etapa do processamento consiste na redução da extremidade 3' do transcrito primário, seguida da adição de 100 a 200 resíduos de adenilato, na forma da *cauda poli(A)* (Figura 4.29). A sequência de reconhecimento da cauda poli(A), AAUAAA, está localizada próxima à extremidade 3' do transcrito primário. A cauda poli(A) estabiliza o RNAm e deve ser removida antes que o RNAm possa ser degradado.

#### MINIQUESTIONÁRIO

- Quais são os três componentes principais que formam um promotor de arqueias?
- Qual o efeito que a presença de um núcleo tem no fluxo da informação genética em eucariotos?
- Quais as etapas envolvidas no processamento do RNA eucariótico?

## IV • Síntese de proteínas

Uma vez ocorridas a transcrição e a produção dos RNAs, os transcritos são traduzidos em proteínas. Esse processo requer muitas outras proteínas, outros RNA e uma estrutura celular-chave, os ribossomos. De que forma estes interagem para produzir uma matriz celular proteica é o que será abordado agora.

### 4.10 Polipeptídeos, aminoácidos e a ligação peptídica

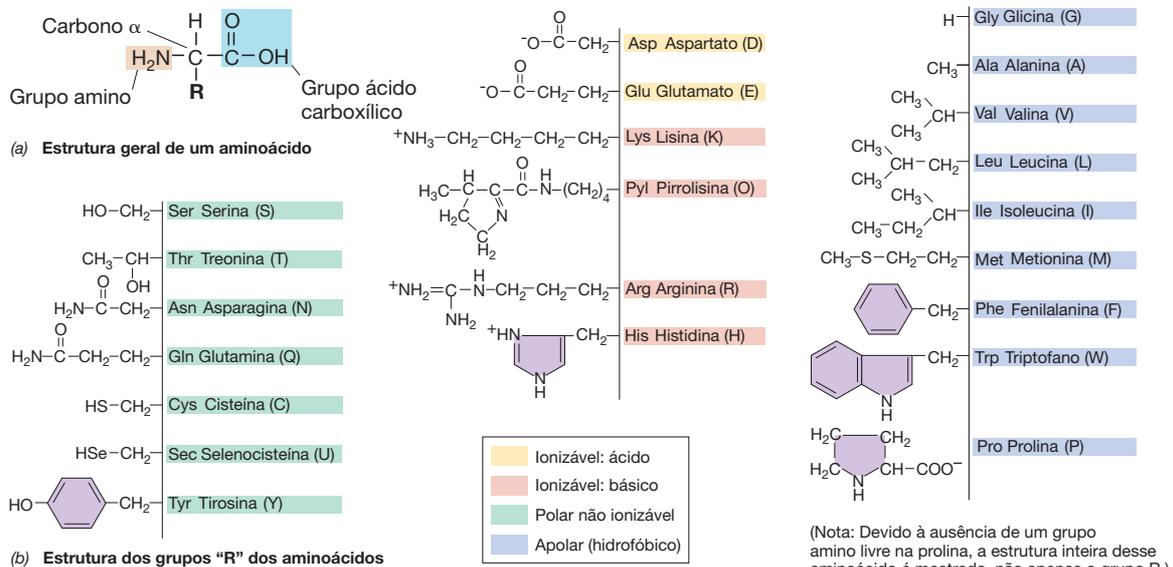
As proteínas desempenham um papel importante na função celular. Duas classes principais de proteínas são as *proteínas catalíticas* (enzimas) e as *proteínas estruturais*. As *enzimas* são catalisadores para as reações químicas que ocorrem na célula. Proteínas estruturais são partes integrantes das principais estruturas da célula: membranas, paredes, ribossomos, e assim por diante. As proteínas reguladoras controlam a maioria dos processos celulares por meio de uma variedade de mecanismos, incluindo a ligação ao DNA, afetando o processo transcricional. No entanto, todas as proteínas apresentam certas características básicas em comum.

As proteínas são polímeros de **aminoácidos**. Todos os aminoácidos contêm um grupo amino ( $-NH_2$ ) e um grupo ácido carboxílico ( $-COOH$ ) que estão ligados ao carbono  $\alpha$  (Figura 4.30a). Ligações entre o carbono do grupo carboxila de um aminoácido e o nitrogênio do grupo amino de outro (com eliminação de água) são conhecidas como **ligações peptídicas** (Figura 4.31). Dois aminoácidos unidos por ligação peptídica constituem um *dipeptídeo*; três aminoácidos, um *tripeptídeo*; e assim por diante. Quando muitos aminoácidos são ligados,

eles formam um **polipeptídeo**. Uma proteína consiste em um ou mais polipeptídeos. O número de aminoácidos difere muito de uma proteína para outra, de 15 até 10.000.

Cada aminoácido possui uma cadeia lateral exclusiva (abreviada como R). Essa cadeia varia consideravelmente, podendo apresentar desde um simples átomo de hidrogênio no aminoácido glicina, até anéis aromáticos na fenilalanina, tirosina e triptofano (Figura 4.30b). Os aminoácidos existem como pares de **enantiômeros**. Esses são isômeros ópticos que apresentam a mesma fórmula molecular e estrutural, exceto pelo fato de um corresponder à “imagem especular” do outro e serem designados como D ou L, dependendo se uma solução pura desvia a luz para a direita ou para a esquerda, respectivamente. Proteínas celulares são compostas apenas L-aminoácidos. Entretanto, D-aminoácidos são ocasionalmente encontrados nas células, mais notavelmente no peptidoglicano, polímero de parede celular (↔ Seção 2.10), e em certos antibióticos peptídicos (↔ Seção 27.14). As células são capazes de realizar a interconversão de enantiômeros por meio da atividade de enzimas denominadas *racemases*.

As propriedades químicas de um aminoácido são dirigidas por sua cadeia lateral. Aminoácidos com propriedades químicas similares são agrupados em “famílias” relacionadas (Figura 4.30b). Por exemplo, a cadeia lateral pode conter um grupo ácido carboxílico, como no ácido aspártico ou ácido glutâmico, o que torna o aminoácido ácido. Outros contêm grupos amino adicionais, tornando-os carregados positivamente e, assim, básicos. Diversos aminoácidos contêm cadeias laterais hidrofóbicas e são denominados aminoácidos apolares. A cisteína contém um grupo sulfidríla ( $-SH$ ). Utili-



**Figura 4.30** Estrutura dos 22 aminoácidos geneticamente codificados. (a) Estrutura geral. (b) Estrutura do grupo R. Os códigos de três letras para os aminoácidos estão dispostos à esquerda dos nomes, e os códigos de uma

letra estão dispostos entre parênteses à direita dos nomes. A pirrolisina, até agora, tem sido encontrada apenas em determinadas arqueias metanogênicas (↔ Seção 16.2).

zando os seus grupos sulfidríla, duas cisteínas podem formar uma ligação dissulfeto (R–S–S–R) que conecta duas cadeias polipeptídicas.

A diversidade de aminoácidos quimicamente distintos torna possível a produção de um número enorme de proteínas exclusivas, com propriedades bioquímicas amplamente diferentes. Se supusermos que um polipeptídeo médio contém 300 aminoácidos, então 22<sup>300</sup> sequências polipeptídicas diferentes são teoricamente possíveis. Nenhuma célula apresenta nem de perto esta enorme quantidade de proteínas distintas. Uma célula de *Escherichia coli* contém cerca de 2.000 tipos diferentes de proteínas, com o número exato sendo altamente dependente dos recursos (nutrientes) e das condições de crescimento empregadas.

A sequência linear de aminoácidos em um polipeptídeo é a sua *estrutura primária*. Esta última determina o dobramento futuro do polipeptídeo, o que, por sua vez, determina a sua atividade biológica (Seção 4.14). As duas extremidades de um

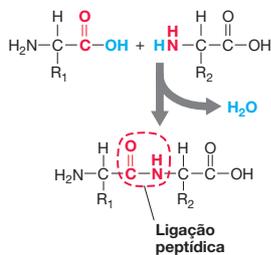
polipeptídeo são denominadas como “C-terminal” e “N-terminal”, dependendo se um grupo ácido carboxílico livre ou grupo amino livre é encontrado (Figura 4.31).

**MINIQUESTIONÁRIO**

- Desenhe a estrutura de um dipeptídeo contendo os aminoácidos alanina e tirosina. Trace a ligação peptídica.
- Qual forma enantiomérica dos aminoácidos é encontrada nas proteínas?
- A glicina não apresenta dois enantiômeros distintos. Por quê?

**4.11 Tradução e o código genético**

Como visto anteriormente, nas duas primeiras etapas de transferência da informação biológica – replicação e transcrição –, ácidos nucleicos são sintetizados em moldes de ácido nucleico. A última etapa, *tradução*, também utiliza um ácido nucleico como molde, porém neste caso o produto é um polipeptídeo em vez de um ácido nucleico. O centro da transferência da informação biológica é a correspondência entre o molde de ácido nucleico e a sequência de aminoácidos do polipeptídeo formado. Essa correspondência está baseada no **código genético**. Um triplete de RNA de três bases, chamado de *códon*, codifica cada aminoácido específico. Os 64 códons possíveis (quatro bases utilizadas em grupos de três = 4<sup>3</sup>) de RNAm são apresentados na **Tabela 4.5**. O código genético é escrito como RNA em vez de DNA, uma vez que é o RNAm que é de fato traduzido. Observe que, além dos códons que especificam os vários aminoácidos, há também códons específicos para o início e o término da tradução. Aqui será abordada a tradução em bactérias, contudo é importante ressaltar que a maquinaria traducional de arqueias e eucariotos é mais estreitamente relacionada entre si do que de bactérias.



**Figura 4.31** Formação da ligação peptídica. R1 e R2 referem-se às cadeias laterais dos aminoácidos. Observe na formação da ligação peptídica que um grupo OH livre está presente na porção C-terminal para formação da próxima ligação peptídica.

**Tabela 4.5** O código genético, expresso por sequências de três bases de RNAm

Códon	Aminoácido	Códon	Aminoácido	Códon	Aminoácido	Códon	Aminoácido
UUU	Fenilalanina	UCU	Serina	UAU	Tirosina	UGU	Cisteína
UUC	Fenilalanina	UCC	Serina	UAC	Tirosina	UGC	Cisteína
UUA	Leucina	UCA	Serina	UAA	Nenhum (sinal de término)	UGA	Nenhum (sinal de término)
UUG	Leucina	UCG	Serina	UAG	Nenhum (sinal de término)	UGG	Triptofano
CUU	Leucina	CCU	Prolina	GAU	Histidina	CGU	Arginina
CUC	Leucina	CCC	Prolina	GAC	Histidina	CGC	Arginina
CUA	Leucina	CCA	Prolina	CAA	Glutamina	CGA	Arginina
CUG	Leucina	CCG	Prolina	CAG	Glutamina	CGG	Arginina
AUU	Isoleucina	ACU	Treonina	AAU	Asparagina	AGU	Serina
AUC	Isoleucina	ACC	Treonina	AAC	Asparagina	AGC	Serina
AUA	Isoleucina	ACA	Treonina	AAA	Lisina	AGA	Arginina
AUG (início) <sup>a</sup>	Metionina	ACG	Treonina	AAG	Lisina	AGG	Arginina
GUU	Valina	GCU	Alanina	GAU	Ácido aspártico	GGU	Glicina
GUC	Valina	GCC	Alanina	GAC	Ácido aspártico	GGC	Glicina
GUA	Valina	GCA	Alanina	GAA	Ácido glutâmico	GGA	Glicina
GUG	Valina	GCG	Alanina	GAG	Ácido glutâmico	GGG	Glicina

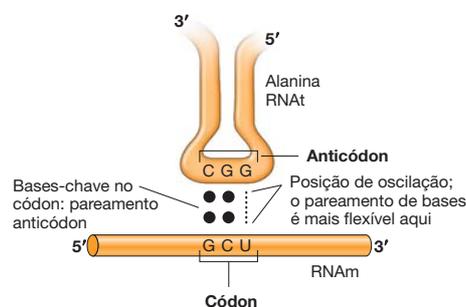
<sup>a</sup>AUG codifica uma N-formilmetionina no início das cadeias polipeptídicas de bactérias.

### Propriedades do código genético

Existem 22 aminoácidos que são codificados pela informação genética carregada no RNAm (alguns outros aminoácidos são formados pela modificação destes após a tradução). Consequentemente, como existem 64 códons, muitos aminoácidos podem ser codificados por mais de 1 códon. Embora conhecer um códon de um determinado local possa identificar inequivocamente o seu aminoácido correspondente, o inverso não é verdadeiro. Tal tipo de código, em que não existe a correlação unívoca entre a “palavra” (ou seja, o aminoácido) e o código (códon), é denominado *código degenerado*. No entanto, conhecendo-se a sequência de DNA e a fase de leitura correta, é possível especificar-se a sequência de aminoácidos presente na proteína. Isso permite a determinação de sequências de aminoácidos a partir de sequências de bases do DNA, o que corresponde ao coração da genômica (Capítulo 6).

Um códon é reconhecido pelo pareamento específico de bases com uma sequência complementar de três bases, denominada **anticódon**, que é parte dos RNAt. Se esse pareamento ocorresse sempre de acordo com o pareamento-padrão de A com U e G com C, pelo menos um RNAt específico seria necessário para reconhecer cada códon. Em alguns casos, esse fato é verdadeiro. Por exemplo, *Escherichia coli* apresenta seis RNAts diferentes que carregam o aminoácido leucina, um para cada códon (Tabela 4.5). Em contrapartida, alguns RNAts podem reconhecer mais de um códon. Assim, embora existam dois códons para a lisina em *E. coli*, há somente um RNAt lisil, cujo anticódon pode parear-se com AAA ou AAG. Nesses casos especiais, as moléculas de RNAt realizam pareamentos-padrão somente nas duas primeiras posições do códon, tolerando um pareamento irregular na terceira posição. Esse fenômeno é denominado **oscilação**, sendo ilustrado na **Figura 4.32**, onde um pareamento entre G e U (em lugar de G com C) é ilustrado na posição oscilante.

Vários aminoácidos são codificados por múltiplos códons, e, na maioria dos casos, os múltiplos códons são estreitamente relacionados em sua sequência de bases (Tabela 4.5). Poderia-se presumir que esses códons múltiplos fossem utilizados com frequência equivalente. No entanto, esse não é o caso, e dados de sequência genômica revelam um importante **uso preferencial de códons** organismo-específico. Em outras palavras, alguns códons são preferencialmente utilizados em relação a outros, embora codifiquem o mesmo aminoácido. O uso preferencial de códons está correlacionado a uma preferência correspondente na concentração de diferentes moléculas de RNAt. Assim, um RNAt correspondente a um códon raramente utilizado estará presente em quantidade relativamente pequena. O uso preferencial de códons deve ser levado em consideração durante a engenharia genética. Por exemplo, um gene de um organismo, cujo uso de códons seja surpreendentemente diferente daquele de outro, pode não ser traduzido.



**Figura 4.32** O conceito de oscilação. O pareamento de bases é mais flexível na terceira base do códon do que em relação às duas primeiras. Somente uma porção do RNAt é ilustrada aqui.

do com eficiência se o gene for clonado nesse último por meio de técnicas de engenharia genética (Capítulo 11).

### Fases abertas de leitura

Um método comum para identificar genes codificadores de proteínas consiste em examinar cada fita da sequência de DNA em busca de **fases abertas de leitura** (ORFs, *open reading frames*). Se um RNA pode ser traduzido, ele contém uma fase de leitura aberta: um códon de início (normalmente AUG) seguido por vários códons, e então um códon de término na mesma fase de leitura que o códon de início. Na prática, apenas ORFs longas o suficiente para codificar um polipeptídeo funcional são aceitas como sequências codificadoras verdadeiras. Embora a maioria das proteínas funcionais exiba extensão de pelo menos 100 aminoácidos, alguns hormônios peptídicos e peptídeos reguladores são muito mais curtos. Consequentemente, nem sempre é possível deduzir somente a partir de dados de sequência se uma ORF relativamente curta é meramente casual ou se codifica um polipeptídeo genuíno, embora curto.

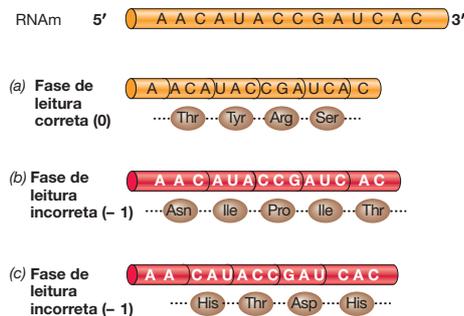
Utilizando-se métodos computacionais, uma longa sequência de bases de DNA pode ser analisada em busca de fases de leitura aberta. Além da busca por códons de início e término, a pesquisa pode incluir também promotores e sequências de ligação ao ribossomo. A pesquisa por ORFs é muito importante na genômica (Capítulo 6). Quando um fragmento de DNA desconhecido é sequenciado, a presença de uma fase de leitura aberta indica que ele pode codificar proteínas.

### Códons de início e término e fase de leitura

O RNA mensageiro é traduzido a partir do **códon de início** (AUG, Tabela 4.5), que codifica uma metionina quimicamente modificada em bactérias, denominada *N-formilmetionina*. Embora um AUG situado no *início* de uma região codificadora codifique *N-formilmetionina*, os AUGs situados no *interior* da região codificadora codificam metionina. Dois RNAs distintos estão envolvidos neste processo (Seção 4.13). Por outro lado, arqueias e eucariotos inserem uma metionina regular como o primeiro aminoácido em um polipeptídeo.

Diante de um código triplo, é fundamental que a tradução seja iniciada no nucleotídeo correto. Caso contrário, toda a fase de leitura no RNAm será alterada, ocorrendo a síntese de uma proteína completamente diferente. Alternativamente, se a alteração introduz um códon de término na fase de leitura, o polipeptídeo será terminado prematuramente. Por convenção, a fase de leitura que é traduzida, originando a proteína codificada pelo gene, é denominada *fase 0* (fase zero). Como pode ser observado na **Figura 4.33**, as outras duas fases de leitura possíveis (-1 e +1) não codificam a mesma sequência de aminoácidos. Portanto, é essencial que o ribossomo encontre o códon de início correto para iniciar a tradução e, uma vez localizado, o RNAm é translocado exatamente três bases a cada vez. Como a correta leitura da fase é garantida?

A fidelidade da fase de leitura é controlada pelas interações entre o RNAm e o RNAr no interior do ribossomo. Em procarionotos, o RNA ribossomal reconhece um AUG específico no RNAm como códon de início, com o auxílio de uma sequência localizada a montante no RNAm, denominada *sítio de ligação ao ribossomo* (RBS, *ribosome-binding site*), ou sequência de



**Figura 4.33** Possíveis fases de leitura em um RNAm. Uma sequência interna de um RNAm é apresentada. (a) Os aminoácidos que seriam codificados se o ribossomo estivesse na fase de leitura correta (denominada fase “0”). (b) Os aminoácidos que seriam codificados por essa região do RNAm se o ribossomo estivesse na fase de leitura -1. (c) Os aminoácidos que seriam codificados se o ribossomo estivesse na fase de leitura +1.

Shine-Dalgarno. Essa exigência de alinhamento explica porque alguns RNAs de bactérias podem utilizar outros códons de início, como GUG. No entanto, mesmo esses códons de início pouco usuais conduzem à incorporação de *N*-formilmetionina como o aminoácido iniciador.

Alguns códons não codificam nenhum aminoácido. Esses códons (UAA, UAG e UGA, Tabela 4.5) são os **códons de término**, uma vez que atuam como sinalizadores do término da tradução de uma sequência codificadora de proteínas no RNAm. Os códons de término são também denominados **códons sem sentido**, pois interrompem o “sentido” do polipeptídeo em crescimento, quando terminam a tradução. Existem algumas exceções a esta regra. Por exemplo, mitocôndrias animais (mas não de plantas) utilizam o códon UGA para codificar triptofano em vez de utilizá-lo como códon de término (Tabela 4.5), enquanto o gênero *Mycoplasma* (*Bacteria*) e o gênero *Paramecium* (*Eukarya*), utilizam determinados códons sem sentido para codificar aminoácidos. Esses organismos simplesmente possuem uma quantidade menor de códons sem sentido, tendo em vista que um ou dois deles são utilizados como códons codificadores (↔ Seção 6.5).

Em alguns poucos casos raros, códons sem sentido codificam aminoácidos incomuns em vez de um dos 20 aminoácidos convencionais. Essas exceções são a selenocisteína e a pirrolisina, o 21<sup>o</sup> e 22<sup>o</sup> aminoácidos codificados geneticamente (Figura 4.30). A selenocisteína e a pirrolisina são codificadas por códons de término (UGA e UAG, respectivamente). Ambas possuem seus próprios RNAs, os quais contêm anticódons capazes de ler esses códons de término. A maioria dos códons de término dos organismos que utilizam a selenocisteína e a pirrolisina indica, de fato, o término. No entanto, códons de término ocasionais são reconhecidos como codificadores de selenocisteína ou pirrolisina. Essa mudança é controlada por uma sequência de reconhecimento situada imediatamente a jusante ao códon de término. A selenocisteína e a pirrolisina são relativamente raras. A maioria dos organismos, incluindo plantas e animais, apresenta um pequeno número de proteínas contendo selenocisteína. A pirrolisina é ainda mais rara. Ela foi encontrada em determinadas arqueias e bactérias, porém foi originalmente descoberta em espécies metanogênicas de arqueias.

## MINIQUESTIONÁRIO

- O que são códons de início e término? Por que eles são importantes para a leitura “em fase” do ribossomo?
- O que é o uso preferencial de códons?
- Se fosse fornecida a você uma determinada sequência nucleotídica, como você encontraria as ORFs?

## 4.12 RNA transportador

Um RNA transportador carrega o anticódon que pareia com o códon no RNAm. Além disso, um RNAt é específico para o aminoácido correspondente ao seu anticódon (ou seja, seu aminoácido *cognato*). O RNAt e seu aminoácido específico são ligados por enzimas específicas denominadas **aminoacil-RNAt sintases**. Para cada aminoácido, existe uma aminoacil-RNAt sintase separada que se liga especificamente ao aminoácido e ao RNAt que possuem os anticódons correspondentes. Essas enzimas asseguram que cada RNAt receba seu aminoácido correto, dessa forma elas precisam reconhecer tanto o RNAt específico quanto seu aminoácido cognato.

## Estrutura geral do RNAt

Existem cerca de 60 RNAt diferentes em células bacterianas, e de 100 a 110 em células de mamíferos. As moléculas de RNA transportador são curtas, de fita simples, apresentando extensa estrutura secundária, e comprimento de 73 a 93 nucleotídeos. Determinadas bases e estruturas secundárias são constantes em todos os RNAt, enquanto outras partes são variáveis. As moléculas de RNA transportador também possuem algumas bases purinas e pirimidinas que exibem ligeiras diferenças, quando comparadas às bases normais encontradas no RNA, uma vez que são quimicamente modificadas. Essas modificações são introduzidas nas bases após a transcrição. Algumas dessas bases pouco usuais são a pseudouridina ( $\psi$ ), inosina, di-hidrouridina (D), ribotimina, metil-guanosina, dimetil-guanosina e metilinosina. O RNAt maduro e ativo também

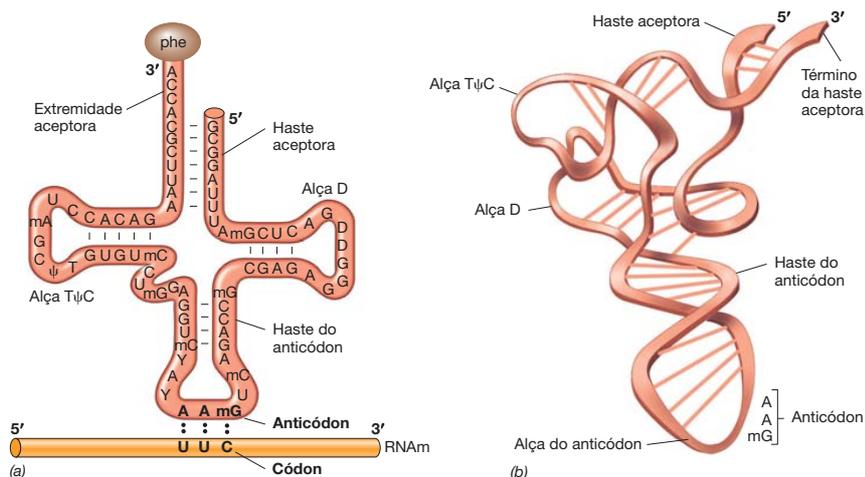
contém extensas regiões de dupla-fita no interior da molécula. Essa estrutura secundária é formada pelo pareamento interno de bases, quando a molécula de fita simples dobra-se sobre si (Figura 4.34).

A estrutura do RNAt pode ser ilustrada na forma de uma folha de trevo, conforme mostrado na Figura 4.34a. Algumas regiões da estrutura secundária do RNAt recebem designações relacionadas com as bases mais frequentemente encontradas naquela região (as alças T $\psi$ C e D), ou a suas funções específicas (alça do anticódon e haste aceptora). A estrutura tridimensional de um RNAt é apresentada na Figura 4.34b. Observe que as bases que aparecem muito distanciadadas no modelo em folha de trevo estão, na realidade, muito próximas, quando observadas no modelo tridimensional. Essa proximidade permite que algumas das bases presentes em uma das alças pareiem-se com bases de outras alças.

## O anticódon e o sítio de ligação do aminoácido

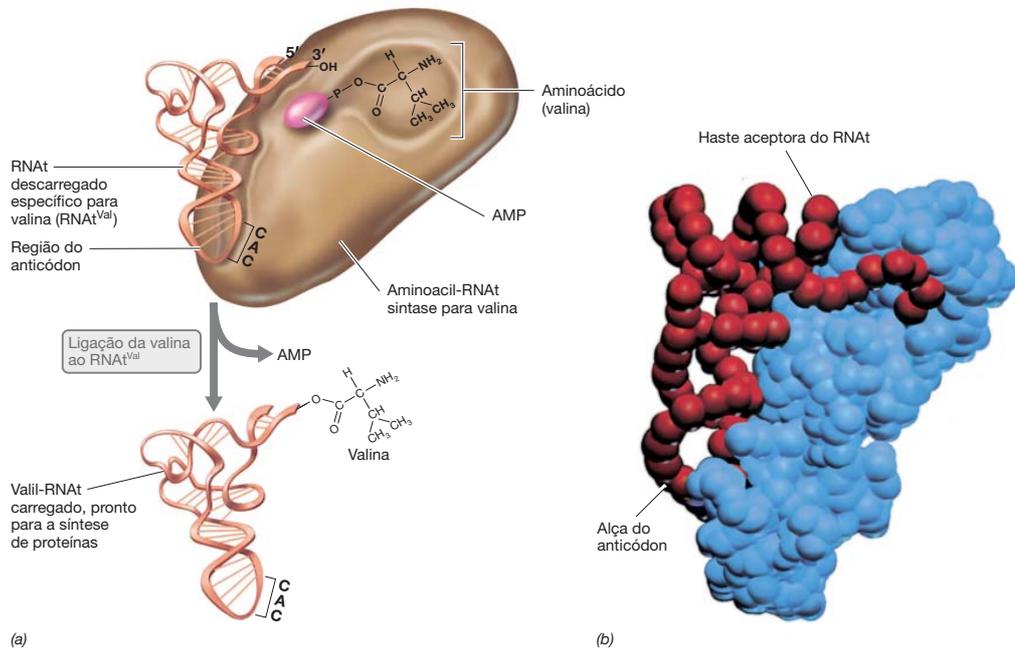
Uma das porções variáveis essenciais da molécula de RNAt corresponde ao *anticódon*, o grupo de três bases que reconhece o códon no RNAm. O anticódon é encontrado na alça do anticódon (Figura 4.34). Os três nucleotídeos do anticódon reconhecem o códon, pareando-se especificamente com suas três bases. Por outro lado, outras porções do RNAt interagem com o RNAr e com os componentes proteicos do ribossomo, as proteínas não ribossomais de tradução, e com a enzima aminoacil-RNAt sintase.

Na extremidade 3' (haste aceptora) de todos os RNAt existem três nucleotídeos não pareados. A sequência desses três nucleotídeos é sempre citosina-citosina-adenina (CCA), sendo eles absolutamente essenciais para a atividade do RNAt. Curiosamente, no entanto, na maioria dos organismos, esses três nucleotídeos (3'CCA) não são codificados pelos genes de RNAt no cromossomo. Em vez disso, cada nucleotídeo é adicionado um por vez, por uma enzima denominada *enzima de adição de CCA*, utilizando CTP e ATP como substratos. O aminoácido cognato é então covalentemente ligado à ade-



**Figura 4.34** Estrutura de um RNA transportador. (a) A estrutura convencional em folha de trevo do RNAt fenilalanina de levedura. O aminoácido liga-se à ribose do A terminal, na extremidade aceptora. A, adenina; C, citosina; U, uracila; G,

guanina; T, timina;  $\psi$ , pseudouracila; D, di-hidrouracila; m, metil; Y, uma purina modificada. (b) Na realidade, a molécula de RNAt dobra-se de tal forma que a alça D e as alças T $\psi$ C aproximem-se, associando-se por meio de interações hidrofóbicas.



**Figura 4.35** Aminoacil-RNAt sintase. (a) Mecanismo de ação de uma aminoacil-RNAt sintase. O reconhecimento do RNAt correto por uma sintase em particular envolve contatos entre sequências específicas de ácido nucleico, presentes na alça D e haste aceptora do RNAt, e aminoácidos específicos da sintase. Neste diagrama, é apresentada a valil-RNAt sintase catalisando a eta-

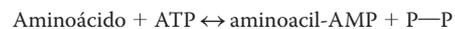
pa final da reação, onde ocorre a transferência da valina, na forma de valil-AMP, para o RNAt. (b) Modelo computacional apresentando a interação da glutamínil-RNAt sintase (azul), com seu respectivo RNAt (vermelho). Reproduzida com a permissão de M. Ruff *et al.* 1991. *Science* 252: 1682-1689. © 1991, AAAS.

nosina terminal da extremidade CCA do seu RNAt correspondente, por meio de uma ligação do tipo éster ao açúcar-ribose. Conforme será visto, a partir deste local no RNAt, o aminoácido é incorporado à cadeia polipeptídica em crescimento presente no ribossomo, em vez de um mecanismo que será descrito na próxima seção.

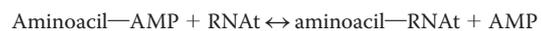
### Reconhecimento, ativação e carregamento dos RNAts

O reconhecimento do RNAt correto por uma aminoacil-RNAt sintase envolve contatos específicos entre regiões-chave do RNAt e da sintase (Figura 4.35). Conforme esperado, devido à sua sequência única, o anticódon do RNAt é importante no reconhecimento pela sintase. Entretanto, outros sítios de contato entre o RNAt e a sintase também são importantes. Estudos da ligação do RNAt à aminoacil-RNAt sintase, onde bases específicas do RNAt foram alteradas por meio de mutações, demonstraram que apenas um pequeno número de nucleotídeos essenciais em um RNAt estão envolvidos no reconhecimento. Esses outros nucleotídeos essenciais ao reconhecimento geralmente encontram-se na haste aceptora ou alça D da molécula de RNAt (Figura 4.34). Deve-se enfatizar que a fidelidade desse processo de reconhecimento é crucial, pois se um aminoácido incorreto for ligado ao RNAt, ele será inserido no polipeptídeo em crescimento, levando à síntese de uma proteína defeituosa.

A reação específica entre o aminoácido e o RNAt, catalisada pela aminoacil-RNAt sintase, é iniciada pela ativação do aminoácido por uma reação envolvendo ATP:



O intermediário aminoacil-AMP formado normalmente permanece ligado à enzima RNAt sintase até colidir com a molécula de RNAt apropriada. Em seguida, conforme apresentado na Figura 4.35a, o aminoácido ativado é ligado ao RNAt, originando um RNAt carregado:



O pirofosfato ( $\text{PP}_i$ ) formado na primeira reação é clivado por uma pirofosfatase, originando duas moléculas de fosfato inorgânico. Como o ATP é utilizado gerando AMP nessas reações, é necessário um total de duas ligações fosfato ricas em energia para carregar um RNAt ao seu aminoácido cognato. Após a ativação e o carregamento, o aminoacil-RNAt deixa a sintase e liga-se ao ribossomo, onde o polipeptídeo é de fato sintetizado.

#### MINIQUESTIONÁRIO

- Qual a função do anticódon de um RNAt?
- Qual a função da haste aceptora de um RNAt?

### 4.13 Síntese de proteínas

É vital para o funcionamento adequado das proteínas que o aminoácido correto seja adicionado nas posições adequadas da cadeia polipeptídica. Essa é a função da maquinaria de síntese proteica, o ribossomo. Embora a síntese proteica seja um

processo contínuo, ela pode ser dividida nos seguintes passos: *início*, *elongação* e *término*. Além de RNAm, RNAt e ribossomos, o processo requer determinadas proteínas denominadas fatores de início, elongação e término. O composto rico em energia trifosfato de guanosina (GTP) fornece a energia necessária para o processo.

### Ribossomos

Os **ribossomos** correspondem aos sítios da síntese proteica. Uma célula pode conter vários milhares de ribossomos, sendo esse número relacionado diretamente à taxa de crescimento. Cada ribossomo é composto por duas subunidades. Os procaríotos apresentam as subunidades ribossomais 30S e 50S, originando ribossomos intactos 70S. Os valores S correspondem a *unidades Svedberg*, que se referem ao coeficiente de sedimentação das subunidades ribossomais (30S e 50S) ou de ribossomos intactos (70S), quando submetidos à força centrífuga em uma ultracentrífuga. (Embora partículas maiores exibam valores S maiores, a relação não é linear e, dessa forma, os valores S não podem ser simplesmente somados.)

Cada subunidade ribossomal contém RNA ribossomais e proteínas ribossomais específicas. A subunidade 30S contém o RNAr 16S e 21 proteínas, enquanto a subunidade 50S contém os RNAr 5S e 23S, além de 31 proteínas. Assim, em *Escherichia coli*, são encontradas 52 proteínas ribossomais diferentes, a maioria presente como cópia única por ribossomo. O ribossomo é uma estrutura dinâmica, cujas subunidades associam-se e dissociam-se alternadamente e também interagem com várias outras proteínas. Diversas proteínas essenciais à atividade ribossomal interagem com o ribossomo em vários estágios da tradução. Essas são consideradas “fatores de tradução”, em vez de “proteínas ribossomais” por si só.

### Início da tradução

Em bactérias, como *E. coli*, o início da síntese proteica começa com uma subunidade ribossomal 30S livre (Figura 4.36). A partir dela, um complexo de início é formado, consistindo na subunidade 30S, no RNAm, no RNAt formil-metionina, e em várias proteínas de início denominadas IF1, IF2 e IF3. GTP é também necessário nesta etapa. Em seguida, uma subunidade ribossomal 50S é adicionada ao complexo de início, formando o ribossomo 70S ativo. Ao final do processo de tradução, o ribossomo novamente se separa nas subunidades 30S e 50S.

Imediatamente precedente ao códon de início no RNAm, há uma sequência de três a nove nucleotídeos denominada sítio de ligação ao ribossomo (RBS na Figura 4.36), que auxilia a ligação do RNAm ao ribossomo. O sítio de ligação ao ribossomo, localizado na extremidade 5' do RNAm, possui bases complementares à sequência presente na extremidade 3' do RNAr 16S. O pareamento de bases entre estas duas moléculas mantém o complexo ribossomo-RNAm unido firmemente, na fase de leitura correta. RNAm policistrônicos possuem múltiplas sequências RBS, uma a montante de cada sequência codificadora. Isso permite que os ribossomos bacterianos traduzam diversos genes em um mesmo RNAm, uma vez que os ribossomos são capazes de encontrar cada sítio de início presente em um mensageiro em vez da ligação ao seu RBS.

O início da tradução sempre começa com a ligação de um aminoacil-RNAt iniciador especial ao códon de início, AUG. Em bactérias, esse corresponde ao RNAt formil-metionina. Após o polipeptídeo ter sido completado, o grupo formil é removido. Consequentemente, o aminoácido N-terminal da proteína completa será a metionina. No entanto, em muitas proteínas essa metionina é removida por uma protease específica.

### Elongação, translocação e término

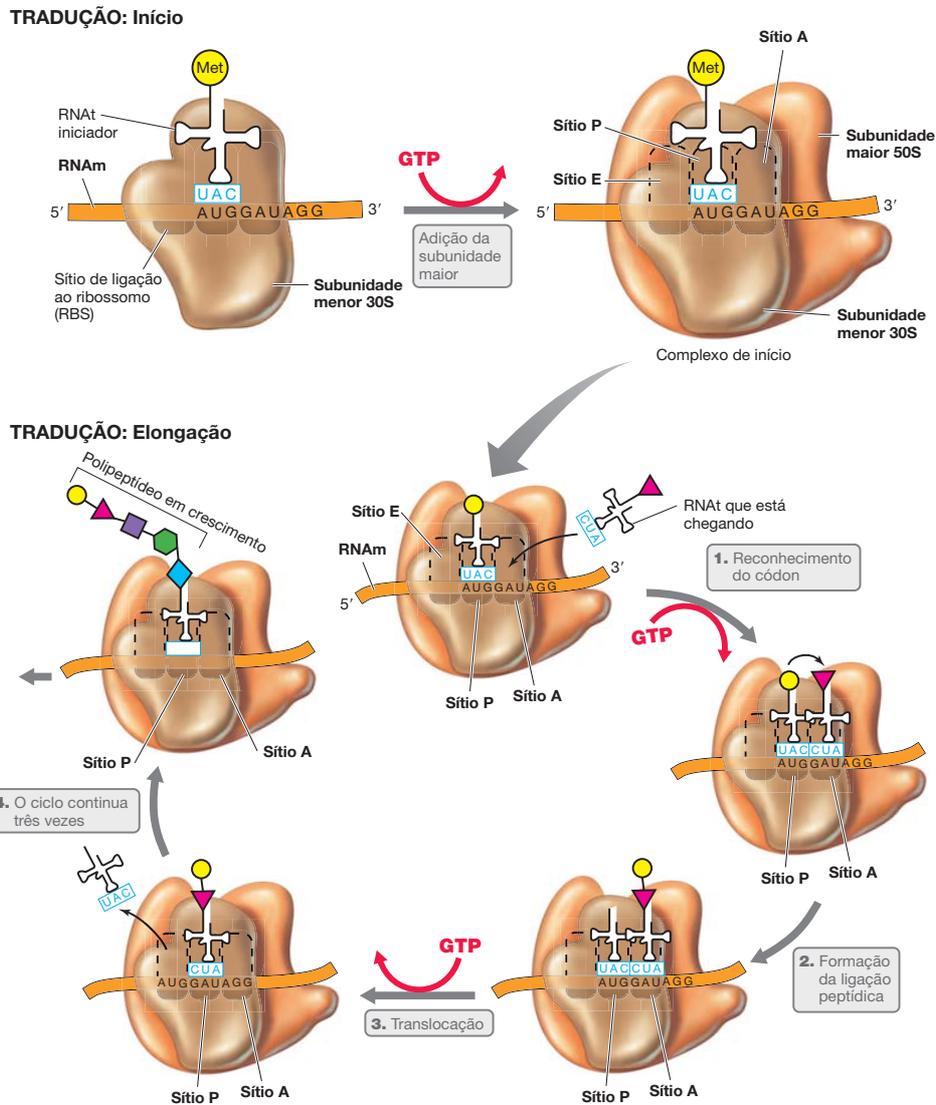
O RNAm passa através do ribossomo primariamente ligado à subunidade 30S. O ribossomo contém outros sítios onde os RNAts interagem. Dois desses sítios localizam-se principalmente na subunidade 50S, sendo denominados sítio A e sítio P (Figura 4.36). O sítio A, sítio aceptor, é o local do ribossomo onde o novo RNAt carregado liga-se inicialmente. O carregamento de um RNAt ao sítio A é auxiliado pelo fator de elongação EF-Tu.

O sítio P, ou sítio peptídico, é o local onde o peptídeo crescente encontra-se ligado pelo RNAt anterior. Durante a formação da ligação peptídica, a cadeia polipeptídica em crescimento move-se em direção ao RNAt localizado no sítio A, enquanto uma nova ligação peptídica é originada. Várias proteínas não ribossomais são necessárias à elongação, especialmente os fatores de elongação EF-Tu e EF-Ts, assim como moléculas adicionais de GTP (visando simplificar a Figura 4.36, os fatores de elongação foram omitidos, sendo apresentada apenas uma porção do ribossomo).

Após a elongação, o RNAt ligado ao polipeptídeo é translocado (movido) do sítio A para o sítio P, deixando o sítio A vazio para receber um outro RNAt carregado (Figura 4.36). A translocação requer o fator de elongação EF-G e uma molécula de GTP para cada evento de translocação. Em cada etapa de translocação, o ribossomo avança por três nucleotídeos, expondo um novo códon no sítio A. A translocação empurra o RNAt, agora livre, para um terceiro sítio, denominado sítio E. A partir desse sítio de saída, o RNAt é liberado do ribossomo. Como era esperada, a precisão da etapa de translocação é crítica à fidelidade da síntese proteica. O ribossomo deve mover-se *exatamente* sobre um códon em cada etapa. Embora, durante esse processo, o RNAm pareça estar deslocando-se por meio do complexo ribossomal, na realidade o ribossomo está movendo-se ao longo do RNAm. Assim, os três sítios ribossomais identificados na Figura 4.36 não são locais estáticos, mas sim partes móveis de uma maquinaria biomolecular.

Vários ribossomos podem traduzir simultaneamente uma única molécula de RNAm, formando um complexo denominado *polissomo* (Figura 4.37). Os polissomos aumentam a velocidade e eficiência da tradução simultaneamente uma vez que cada ribossomo de um complexo polissomal sintetiza um polipeptídeo completo. Observe, na Figura 4.37, como os ribossomos no complexo polissomal que estão mais próximos à extremidade 5' (início) da molécula de RNAm apresentam polipeptídeos menores ligados a eles, uma vez que somente poucos códons foram lidos, enquanto os ribossomos mais próximos à extremidade 3' do RNAm apresentam polipeptídeos praticamente prontos.

O término da síntese proteica ocorre quando o ribossomo alcança um códon de término (códon sem sentido). Nenhum RNAt se liga a um códon de término. Em vez disso, proteínas



**Figura 4.36** O ribossomo e a síntese proteica. *Início* da síntese proteica. O RNAm e o RNAt iniciador, carregando *N*-formilmetionina (“Met”), se liga primeiramente à subunidade menor do ribossomo. Os fatores de início (não apresentados) utilizam energia do GTP para promover a adição da subunidade ribossomal maior. O RNAt iniciador inicia o processo no sítio P. Ciclo de *elongação* da tradução. 1. Os fatores de elongação (não apresentados) utilizam GTP

para instalar o RNAt que está chegando no sítio A. 2. A formação da ligação peptídica é então catalisada pelo RNAr 23S. 3. A translocação do ribossomo ao longo RNAm de um códon para o seguinte requer a hidrólise de outra molécula de GTP. O RNAt que está saindo do complexo é liberado do sítio E. 4. O próximo RNAt a ser carregado se liga ao sítio A e o ciclo se repete. O código genético, expresso na linguagem de RNAm, é mostrado na Tabela 4.5.

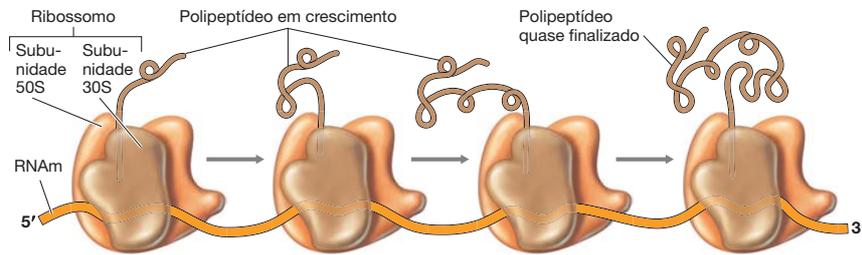
específicas denominadas *fatores de liberação* (RFs, *release factors*) reconhecem o códon de término e clivam o polipeptídeo ligado ao RNAt terminal, liberando o produto completo. Em seguida, as subunidades ribossomais dissociam-se, disponibilizando as subunidades 30S e 50S para formarem novos complexos de início e repetirem o processo.

**Papel do RNA ribossomal na síntese proteica**

O RNA ribossomal desempenha papel vital em todos os estágios da síntese proteica, da iniciação à terminação. Em contra-

partida, o papel das várias proteínas presentes no ribossomo é formar um arcabouço para o posicionamento de sequências essenciais nos RNA ribossomais.

Em bactérias, está claro que o RNAr 16S está envolvido na iniciação por meio de seu pareamento de bases com a sequência RBS no RNAm. Outras interações RNAm–RNAr também ocorrem durante a elongação. Em quaisquer extremidades dos códons dos sítios A e P, o RNAm é mantido em posição por meio da ligação ao RNAr 16S e às proteínas ribossomais. O RNA ribossomal também desempenha papel



**Figura 4.37 Polissomos.** A tradução de um único RNA mensageiro por vários ribossomos forma o polissomo. Observe que os ribossomos mais próximos à extremidade 5' do mensageiro encontram-se em um estágio mais inicial do processo de tradução do que os ribossomos mais próximos à extremidade 3' e, portanto, apenas uma porção relativamente curta do polipeptídeo final foi sintetizada.

na associação das subunidades ribossomais, bem como no posicionamento do RNAt nos sítios A e P do ribossomo (Figura 4.36). Embora os RNAts carregados que chegam ao ribossomo reconheçam o códon correto pelo pareamento códon-anticódon, eles também ligam-se ao ribossomo por meio de interações da haste-alça do anticódon do RNAt, com sequências específicas do RNAr 16S. Além disso, a extremidade acceptora do RNAt (Figura 4.36) pareia-se com sequências do RNAr 23S.

Além de todos esses fatos, a real formação de ligações peptídicas é catalisada pelo RNAr. A reação da peptidil transferase ocorre na subunidade 50S do ribossomo e é catalisada pelo próprio RNAr 23S, em vez de qualquer uma das proteínas ribossomais. O RNAr 23S também desempenha um papel na translocação, e as proteínas EF interagem especificamente com o RNAr 23S. Assim, além de seu papel como esqueleto estrutural do ribossomo, o RNA ribossomal também desempenha um importante papel catalítico no processo de tradução.

### Liberação de ribossomos capturados

Um RNAm defeitivo, desprovido de um códon de término, provoca um problema na tradução. Esse tipo de defeito pode surgir, por exemplo, a partir de uma mutação que removeu o códon de término, síntese defectiva do RNAm, ou quando ocorre degradação parcial do RNAm antes que este seja traduzido. Quando um ribossomo atinge a extremidade de uma molécula de RNAm e não há qualquer códon de término, o fator de liberação não pode se ligar, impedindo a liberação do ribossomo do RNAm. O ribossomo encontra-se efetivamente “capturado”.

As células bacterianas contêm uma pequena molécula de RNA, denominada *RNA<sup>tm</sup>*, que libera ribossomos parados (Figura 4.38). As letras “tm” na denominação referem-se ao fato de o RNA<sup>tm</sup> imitar tanto o RNAt, uma vez que carrega o aminoácido alanina, quanto o RNAm, pelo fato de conter um curto segmento de RNA que pode ser traduzido. Quando o RNA<sup>tm</sup> colide com um ribossomo parado, este liga-se ao lado do RNAm defeitivo. A síntese proteica pode então prosseguir, primeiro pela adição da alanina no RNA<sup>tm</sup> e, em seguida, pela tradução da curta mensagem do RNA<sup>tm</sup>. O RNA<sup>tm</sup> contém um códon de término que permite a ligação do fator de liberação e desmontagem do ribossomo. A proteína sintetizada como resultado dessa operação de resgate é defectiva, sendo subsequentemente degradada. Isso é

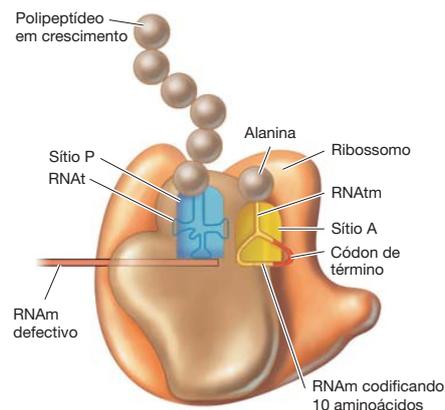
possível por meio da adição da pequena sequência de aminoácidos codificada pelo RNA<sup>tm</sup> à extremidade da proteína defectiva; a sequência é um sinal para que uma protease específica degrade a proteína. Desse modo, pela atividade do RNA<sup>tm</sup>, os ribossomos parados não são inativados, mas liberados para participarem novamente da síntese proteica.

### MINIQUESTIONÁRIO

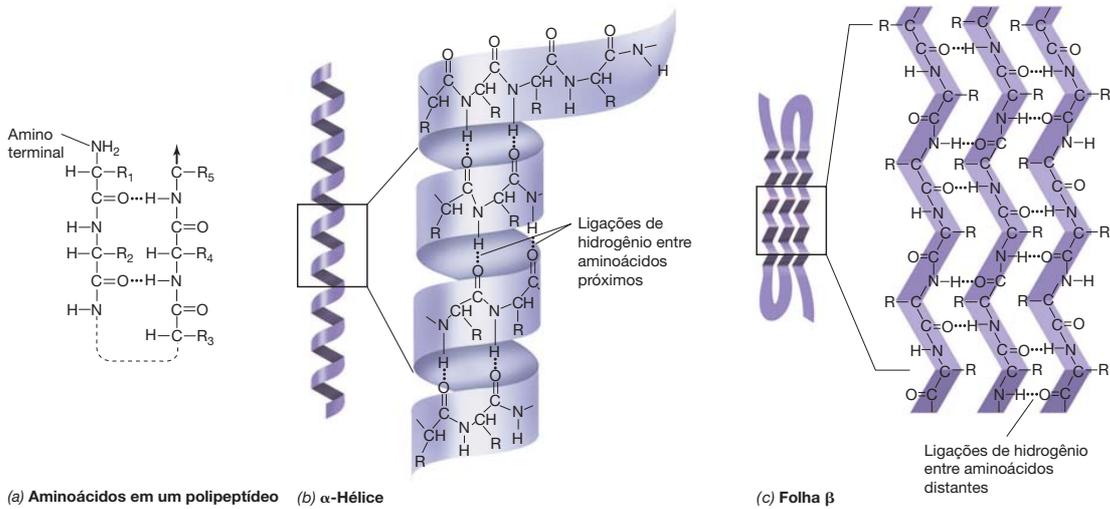
- Quais são os componentes de um ribossomo? Que papéis funcionais o RNAr desempenha na síntese proteica?
- Como uma cadeia polipeptídica completa é liberada do ribossomo?
- Como os RNA<sup>tm</sup> liberam os ribossomos parados?

### 4.14 Dobramento e secreção de proteínas

Para que uma proteína funcione adequadamente, ela deve ser dobrada corretamente após a sua síntese, bem como deve destinar-se ao local correto na célula. Enquanto muitas proteínas se encontram dentro da célula, algumas outras precisam ser transportadas de fora da membrana citoplasmática para o periplasma, ou para dentro das membranas interna e externa para facilitar processos como transporte de íons, açúcares e elétrons. Outras proteínas, como toxinas e enzimas extracelulares (exoenzimas), precisam ser completamente secretadas das células para se tornarem ativas no meio ambiente. Como a célula determina a conformação final e a localização das proteínas? Considera-se estes dois processos relacionados agora.



**Figura 4.38 Liberação de um ribossomo parado pelo RNA<sup>tm</sup>.** Um RNAm defeitivo, desprovido de um códon de término, promove a parada de um ribossomo que contém um polipeptídeo parcialmente sintetizado ligado a um RNAt (azul), no sítio P. A ligação do RNA<sup>tm</sup> (amarelo) ao sítio A libera o polipeptídeo. A tradução então prossegue até o códon de término fornecido pelo RNA<sup>tm</sup>.



**Figura 4.39** Estrutura secundária dos polipeptídeos. (a) Ligação de hidrogênio na estrutura secundária de uma proteína. R representa a cadeia lateral do aminoácido. (b) Estrutura secundária em  $\alpha$ -hélice. (c) Estrutura se-

cundária em folha  $\beta$ . Observe que a ligação de hidrogênio é entre os átomos nas ligações peptídicas e não envolve os grupos R.

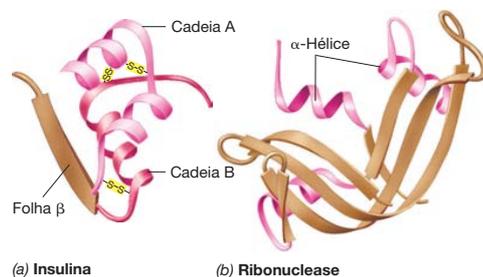
### Níveis de estrutura proteica

Uma vez formado, um polipeptídeo não permanece linear; em vez disso, ele se dobra para formar uma estrutura mais estável. Ligações de hidrogênio, entre os átomos de oxigênio e nitrogênio de duas ligações peptídicas, originam a *estrutura secundária* (Figura 4.39a). Um tipo comum de estrutura secundária corresponde à  *$\alpha$ -hélice*. Para visualizar uma  $\alpha$ -hélice, imagine um polipeptídeo linear enrolado ao redor de um cilindro (Figura 4.39b). Isso posiciona ligações peptídicas próximas o suficiente para permitir a formação de ligações de hidrogênio. O grande número de tais ligações de hidrogênio confere à  $\alpha$ -hélice sua estabilidade inerente. Na estrutura secundária  *$\beta$ -pregueada*, o polipeptídeo dobra-se para trás e para frente, sobre si mesmo, em vez de formar uma hélice. Contudo, como no caso da  $\alpha$ -hélice, o dobramento em uma folha  $\beta$  posiciona as ligações peptídicas de forma que estas possam formar ligações de hidrogênio (Figura 4.39c). Os polipeptídeos podem conter regiões de estrutura secundária  $\alpha$ -hélice e folha  $\beta$ , sendo o tipo de dobramento e sua localização na molécula determinados pela estrutura primária e pelas oportunidades disponíveis de formação de ligações de hidrogênio.

Interações entre os grupos R dos aminoácidos em um polipeptídeo originam mais dois níveis de estrutura. A *estrutura terciária* depende em grande parte das interações hidrofóbicas, com contribuições menores das ligações de hidrogênio, ligações iônicas e ligações dissulfeto. O dobramento terciário origina a forma da estrutura tridimensional global de cada cadeia polipeptídica (Figura 4.40). Muitas proteínas consistem em duas ou mais cadeias polipeptídicas. A *estrutura quaternária* refere-se ao número e ao tipo de polipeptídeos na proteína final. Em proteínas que apresentam estrutura quaternária, cada polipeptídeo é denominado *subunidade* e possui a sua própria estrutura primária, secundária e terciária. Tanto a estrutura terciária quanto a quaternária podem ser estabilizadas por ligações dissulfeto entre os grupos sulfidríla adjacentes de resíduos de cisteína (Figura 4.40). Se os dois resíduos de cisteína

estiverem localizados em diferentes polipeptídeos, a ligação dissulfeto liga covalentemente as duas moléculas. Alternativamente, uma única cadeia polipeptídica pode dobrar-se e ligar-se em si mesma se uma ligação dissulfeto puder ser formada no interior da molécula.

Quando as proteínas são expostas a extremos de calor ou de pH, ou a determinados compostos químicos que afetam seu dobramento, elas podem sofrer **desnaturação**. A desnaturação resulta no desdobramento da cadeia polipeptídica. Quando isso ocorre, as estruturas secundária, terciária e quaternária da proteína são destruídas juntamente com suas propriedades biológicas. No entanto, como as ligações peptídicas não são quebradas, o polipeptídeo desnaturado mantém a sua estrutura primária. Dependendo da intensidade das condições de desnaturação, o polipeptídeo pode dobrar-se novamente após a remoção do agente desnaturante. Contudo, se o redobramento da molécula não for correto, a proteína não será funcional.



**Figura 4.40** Estrutura terciária dos polipeptídeos. (a) Insulina, uma proteína contendo duas cadeias polipeptídicas; observe como a cadeia B contém tanto a estrutura secundária  $\alpha$ -hélice quanto a folha  $\beta$ , e como as ligações dissulfeto (S-S) auxiliam na organização dos padrões de dobramento (estrutura terciária). (b) Ribonuclease, uma proteína grande com diversas regiões de estrutura secundária  $\alpha$ -hélice e folha  $\beta$ .

### Chaperonas auxiliam no dobramento da proteína

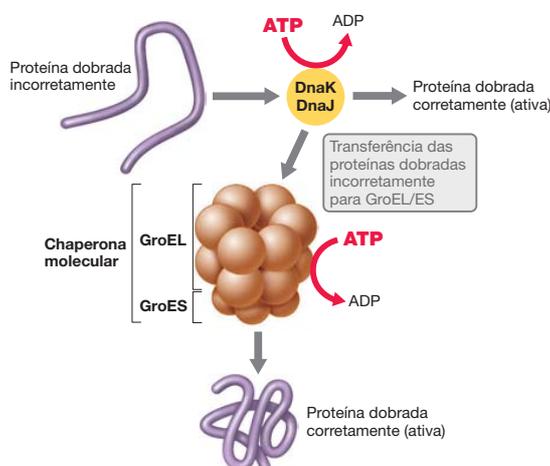
A maioria dos polipeptídeos dobra-se espontaneamente em sua conformação ativa, à medida que são sintetizados. Entretanto, alguns não o fazem e requerem o auxílio de outras proteínas, denominadas **chaperonas** (também conhecidas como **chaperoninas moleculares**), para realizar seu dobramento correto, ou para sua organização em complexos maiores. As próprias chaperonas não se tornam parte das proteínas organizadas, apenas auxiliam no processo de dobramento. De fato, uma importante atividade das chaperonas consiste em impedir a agregação imprópria de proteínas. As chaperonas são amplamente distribuídas em todos os domínios da vida, e suas sequências são altamente conservadas entre todos os organismos.

*Escherichia coli* possui quatro chaperonas essenciais, que são a DnaK, DnaJ, GroEL e GroES. DnaK e DnaJ são enzimas ATP-dependentes que se ligam a polipeptídeos recém-formados, impedindo seu dobramento abrupto, um processo que aumenta o risco de dobramento impróprio (Figura 4.41). Se o complexo DnaKJ for incapaz de dobrar a proteína adequadamente, ele pode transferir a proteína parcialmente dobrada às duas proteínas de multissubunidades, GroEL e GroES. Primeiro a proteína é introduzida em GroEL, uma proteína grande em forma de barril que, utilizando a energia da hidrólise de ATP, dobra a proteína corretamente. GroES auxilia nesse processo (Figura 4.41). Estima-se que cerca de 100 ou mais, das várias milhares de proteínas de *E. coli*, requerem auxílio do complexo GroEL/GroES para seu dobramento e, entre elas, aproximadamente uma dúzia é essencial à sobrevivência da bactéria.

Além do dobramento de proteínas recém-sintetizadas, as chaperonas também promovem o redobramento de proteínas celulares parcialmente desnaturadas. Uma proteína pode ser desnaturada por várias razões, porém, muitas vezes, isso ocorre quando um organismo é submetido temporariamente a altas temperaturas. Desse modo, as chaperonas são um tipo de *proteína de choque térmico*, sendo sua síntese intensamente acelerada quando a célula encontra-se sob estresse devido ao calor excessivo (↔ Seção 7.10). A resposta ao choque térmico consiste em uma tentativa da célula de redobrar suas proteínas parcialmente desnaturadas, a fim de que possam ser reutilizadas antes que as proteases as reconheçam como dobradas incorretamente e as destruam, liberando seus aminoácidos para a síntese de novas proteínas.

### Secreção de proteínas

Muitas proteínas estão localizadas na membrana citoplasmática, no periplasma (de células gram-negativas), ou até mesmo externamente à célula. Essas proteínas devem migrar a partir do sítio de sua síntese nos ribossomos para dentro da membrana citoplasmática, ou mesmo atravessar essa estrutura. Proteínas denominadas *translocases* transportam proteínas específicas por meio e para o interior de membranas procarióticas. Por exemplo, o sistema Sec exporta proteínas desdobradas e insere proteínas de membrana integrais na membrana citoplasmática, enquanto o sistema Tat transporta proteínas dobradas no citoplasma através da membrana. Para secretar completamente proteínas para fora da célula, células gram-negativas precisam empregar translocases adicionais para transportar proteínas através da membrana externa. Pelo menos seis tipos de sistemas de excreção distintos já foram identifica-



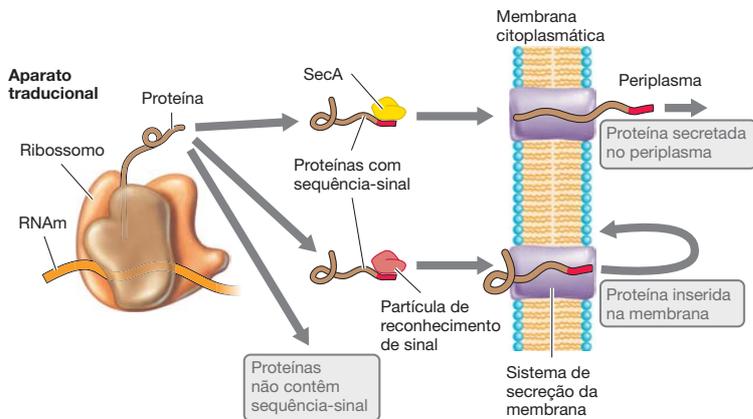
**Figura 4.41** A atividade de chaperonas moleculares. Uma proteína com dobramento incorreto pode ser redobrada pelo complexo DnaKJ ou pelo complexo GroEL-GroES. Em ambos os casos, a energia para o redobramento é proveniente do ATP.

dos, alguns dos quais são empregados por bactérias patogênicas para excretar toxinas ou proteínas nocivas no hospedeiro durante a infecção.

A maioria das proteínas que necessita ser transportada para o interior ou através de membranas é sintetizada com uma sequência de aminoácidos de 15 a 20 resíduos, denominada **sequência-sinal**, no início (N-terminal, Figura 4.31) da molécula proteica. Sequências-sinal são variáveis, mas normalmente apresentam alguns aminoácidos positivamente carregados no começo, uma região central de resíduos hidrofóbicos, e, em seguida, uma região mais polar ao final. A sequência-sinal recebeu esse nome porque “sinaliza” ao sistema secretor da célula que essa proteína em particular necessita ser exportada, e também auxilia a prevenir o dobramento completo da proteína, um processo que poderia interferir na sua secreção. Uma vez que a sequência-sinal corresponde à primeira porção da proteína a ser sintetizada, as etapas iniciais de exportação podem ter início antes que a proteína seja totalmente sintetizada (Figura 4.42).

As proteínas a serem exportadas são identificadas pela *proteína SecA* ou pela *partícula de reconhecimento de sinal* (SRP, *signal recognition particle*) (Figura 4.42). Geralmente, SecA liga-se a proteínas que são totalmente exportadas através da membrana para o periplasma, enquanto SRP liga-se a proteínas que são inseridas na membrana, mas não liberadas na outra face. As SRPs são encontradas em todas as células. Em bactérias, estas contêm uma única proteína e uma pequena molécula de RNA não codificadora (RNA 4.5S). Tanto SecA como SRP encaminham as proteínas que devem ser secretadas para o complexo secretor da membrana, e após essas atravessarem (mediado por SEC) ou serem inseridas na membrana (mediado por SRP), a sequência-sinal é removida por uma protease.

As proteínas transportadas através da membrana citoplasmática na forma desdobrada pelo sistema Sec são dobradas posteriormente (Figura 4.42). Entretanto, há um pequeno número de proteínas, como as proteínas necessárias para o



**Figura 4.42** Exportação de proteínas por meio do sistema secretor principal. A sequência-sinal é reconhecida por SecA ou pela partícula de reconhecimento de sinal, que conduzem a proteína ao sistema de secreção da membrana. A partícula de reconhecimento de sinal liga-se às proteínas que são inseridas na membrana, enquanto SecA liga-se às proteínas que são secretadas através da membrana citoplasmática.

metabolismo energético, que funcionam no periplasma, por exemplo, proteínas ferro-enxofre e várias outras proteínas redox acopladas (↔ Seção 3.10) que devem ser transportadas para fora da célula após terem assumido sua estrutura dobrada final. Em geral, isso se deve ao fato de conterem pequenos cofatores que devem ser inseridos na proteína antes que ela se dobre em sua configuração final. Tais proteínas são dobradas no citoplasma, sendo então exportadas por um sistema de transporte distinto de Sec, denominado *sistema Tat de exportação de proteínas*. O acrônimo Tat é derivado de “*twin arginine translocase*”, uma vez que as proteínas transportadas apresentam uma sequência-sinal curta contendo um par de resíduos de arginina. A presença dessa sequência-sinal em uma proteína dobrada é reconhecida pelas proteínas TatBC, que conduzem a proteína ao TatA, o transportador da membrana. Uma vez que a proteína tenha sido transportada para o periplasma, utilizando a energia fornecida pela força próton-motiva (↔ Seção 3.11), a sequência-sinal é removida por uma protease.

**Sistemas de secreção tipos I a VI**

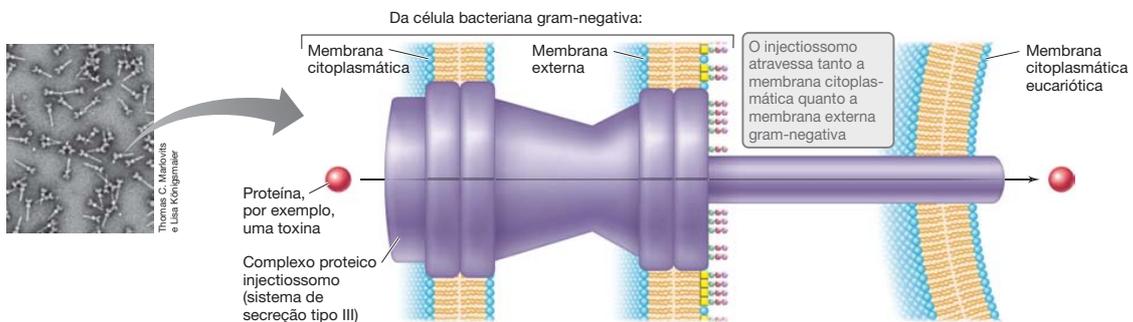
Diversos sistemas adicionais são utilizados por bactérias gram-negativas para transportar proteínas localizadas no interior, ou no meio da membrana externa, para o exterior da célula. Esses mecanismos são os sistemas de secreção tipos I a VI. Cada um desses sistemas é composto por um grande complexo proteico, que forma um canal através de uma ou mais membranas, para a molécula secretada utilizar como passagem (Figura 4.43).

Esses diversos sistemas podem ser agrupados em dois tipos: os de etapa única ou de duas etapas. Os sistemas tipos II e V são considerados mecanismos de duas etapas, uma vez que dependem do sistema Sec ou Tat para transportar as proteínas secretadas, ou uma porção do canal, através da membrana interna. Um segundo grupo de transportadores move proteínas através da membrana externa. Os tipos I, III, IV e VI são sistemas de etapa única, já que formam canais através de ambas as membranas e não necessitam de Sec ou Tat.

Para injetar toxinas proteicas nas células hospedeiras, os sistemas de secreção dos tipos III, IV e VI, também incluem estruturas do exterior da célula que permitem a injeção ou inserção da proteína secretada em outra célula. A estrutura do tipo III completa foi denominada de “*injectiossomo*” devido à sua similaridade a uma seringa em estrutura e função (Figura 4.43).

**MINIQUESTIONÁRIO**

- No que diz respeito às proteínas, defina os termos: estruturas primária, secundária e terciária. De que forma um polipeptídeo se difere de uma proteína?
- Qual a função de uma chaperona molecular?
- Por que algumas proteínas possuem uma sequência-sinal? O que é a partícula de reconhecimento de sinal?
- Por que é importante para bactérias gram-negativas possuírem vias de secreção adicionais?



**Figura 4.43** Secreção de moléculas em bactérias gram-negativas utilizando o sistema tipo III “*injectiossomo*”. O complexo proteico que

constitui o *injectiossomo*. No detalhe: Micrografia eletrônica de *injectiossomos* purificados de *Salmonella enterica* sorovar *typhimurium*.

## CONCEITOS IMPORTANTES

**4.1 •** O conteúdo informativo de um ácido nucleico é determinado pela sequência de bases nitrogenadas ao longo da cadeia polinucleotídica. Tanto o RNA quanto o DNA são macromoléculas informativas, da mesma forma que as proteínas que eles codificam. Os três processos essenciais da síntese macromolecular são: (1) replicação do DNA, (2) transcrição (síntese de RNA) e (3) tradução (síntese de proteínas).

**4.2 •** O DNA é uma molécula dupla-fita que se forma em uma hélice. As duas fitas na dupla-hélice são complementares e antiparalelas. Moléculas muito longas de DNA conseguem ser empacotadas nas células devido ao supernovelamento produzido por enzimas denominadas topoisomerases, como a DNA-girase.

**4.3 •** Além do cromossomo, outros elementos genéticos podem existir nas células. Plasmídeos são moléculas de DNA que existem separadamente do cromossomo e podem conferir uma vantagem de crescimento seletiva em determinadas condições. Os vírus contêm um genoma de RNA ou DNA, e os elementos transponíveis existem como parte de outros elementos genéticos. *Escherichia coli* é o principal organismo-modelo em biologia.

**4.4 •** Ambas as fitas da hélice de DNA são moldes para a síntese de novas fitas (replicação semiconservativa). As novas fitas são alongadas pela adição de desoxirribonucleotídeos à extremidade 3'. DNA-polimerases requerem um iniciador de RNA feito pela enzima primase.

**4.5 •** A síntese de DNA se inicia em sítio denominado origem de replicação. A dupla-hélice é desenovelada pela enzima helicase e estabilizada pela proteína de ligação à fita simples. A extensão do DNA ocorre de forma contínua na fita-líder, porém de forma descontínua na fita tardia, resultando em fragmentos de Okazaki na fita tardia que precisam ser conectados.

**4.6 •** A partir de uma única origem em um cromossomo circular, duas forquilhas de replicação sintetizam DNA simultaneamente, em ambas as direções, até o encontro das forquilhas na região de terminação. As proteínas na forquilha de replicação formam um grande complexo conhecido como replissomo. A maioria dos erros no pareamento de bases que ocorre durante a replicação é corrigida pela função de revisão das DNA-polimerases.

**4.7 •** Em bactérias, promotores são reconhecidos pela subunidade sigma da RNA-polimerase. Fatores sigma alternativos permitem a regulação conjunta de grandes famílias gênicas em resposta às condições de crescimento. A transcrição pela RNA-polimerase continua até que

sejam alcançados sítios específicos denominados terminadores transcricionais. Esses terminadores funcionam ao nível de RNA.

**4.8 •** A unidade de transcrição em procariotos geralmente contém mais de um único gene, que é transcrito em uma única molécula de RNAm, que contém informação para mais de um polipeptídeo. Um grupo de genes que é transcrito em conjunto a partir de um único promotor constitui um óperon.

**4.9 •** O aparato transcricional e a arquitetura do promotor de arqueias e eucariotos possuem muitas características em comum, embora os componentes encontrados em arqueias sejam, frequentemente, relativamente mais simples. Em contrapartida, o processamento do transcrito primário eucariótico é único e possui três etapas distintas: *splicing*, adição de *cap* e adição de cauda poli(A).

**4.10 •** Cadeias polipeptídicas contêm 22 aminoácidos geneticamente distintos codificados, que são conectados por meio de ligações peptídicas. A estrutura primária de uma proteína é a sua sequência de aminoácidos, porém a estrutura de ordem superior (dobramento) de um polipeptídeo determina a sua função celular.

**4.11 •** O código genético é expresso na forma de RNA, e um único aminoácido pode ser codificado por diversos códons distintos, porém relacionados. Além dos códons de término (sem sentido), há também um códon de início específico que sinaliza o início da tradução.

**4.12 •** Enzimas denominadas aminoacil-RNAt sintases ligam aminoácidos ao seu RNAt cognato. Existem um ou mais RNAt para cada aminoácido incorporado em polipeptídeos pelo ribossomo.

**4.13 •** A tradução ocorre no ribossomo e requer RNAm e aminoacil-RNAts. O ribossomo possui três sítios: acceptor, peptídico e de saída. Durante cada etapa da tradução, o ribossomo avança um códon ao longo do RNAm, e o RNAt no sítio acceptor se move para o sítio peptídico. A síntese proteica termina quando um códon de término, que não possui um RNAt correspondente, é alcançado.

**4.14 •** As proteínas necessitam de um dobramento correto a fim de que possam funcionar corretamente, e as chaperonas moleculares auxiliam nesse processo. Muitas proteínas também necessitam ser transportadas para o interior ou através da membrana citoplasmática. Essas proteínas contêm uma sequência-sinal que é reconhecida por translocases celulares. Sistemas de secreção adicionais são empregados por bactérias gram-negativas para secretar proteínas para o interior ou através da membrana externa.

## REVISÃO DOS TERMOS-CHAVE

- Ácido nucleico** DNA ou RNA.
- Aminoácido** um dos 22 monômeros distintos que formam as proteínas; quimicamente, um ácido carboxílico de dois carbonos contendo um grupo amino e um substituinte característico no carbono  $\alpha$ .
- Aminoacil-RNAt sintase** enzima que catalisa a ligação de um aminoácido a seu RNAt cognato.
- Anticódon** sequência de três bases, presente em uma molécula de RNAt, que se pareia com um códon durante a síntese proteica.
- Antiparalelo** em relação ao DNA de dupla-fita, refere-se às duas fitas que se orientam em direções opostas (uma na direção  $5' \rightarrow 3'$  e a fita complementar, na direção  $3' \rightarrow 5'$ ).
- Bacteriocina** proteína tóxica secretada por bactérias, que inibe ou mata outras bactérias relacionadas.
- Chaperona molecular** proteína que auxilia outras proteínas em seu dobramento ou redobramento, a partir de um estado parcialmente desnaturado.
- Código genético** correspondência entre a sequência de ácidos nucleicos e a sequência de aminoácidos de proteínas.
- Códon** sequência de três bases de um RNAm, que codifica um aminoácido.
- Códon de início** códon especial, geralmente AUG, que indica o início de uma proteína.
- Códon de término** códon que indica o final de uma proteína.
- Códon sem sentido** outra denominação para o códon de término.
- Complementar** sequências de ácido nucleico que podem parear-se.
- Cromossomo** elemento genético, normalmente circular em procariontos, que carrega os genes essenciais às funções celulares.
- Desnaturação** perda do dobramento correto de uma proteína, levando (geralmente) à agregação proteica e perda da atividade biológica.
- DNA (ácido desoxirribonucleico)** polímero de desoxirribonucleotídeos unidos por ligações fosfodiéster que carrega a informação genética.
- DNA-helicase** enzima que utiliza ATP para desenovelar a dupla-hélice de DNA.
- DNA-ligase** enzima que sela os cortes no arcabouço do DNA.
- DNA-polimerase** enzima que sintetiza uma nova fita de DNA, na direção  $5' \rightarrow 3'$ , utilizando uma fita antiparalela de DNA como molde.
- Elemento genético** estrutura que carrega a informação genética, como um cromossomo, plasmídeo ou genoma viral.
- Elemento transponível** elemento genético capaz de se mover (transpor) de um local para outro na molécula de DNA do hospedeiro.
- Enantiômero** forma de uma molécula, que é a imagem especular de outra forma da mesma molécula.
- Enzima** uma proteína ou RNA que catalisa uma reação química específica em uma célula.
- Estrutura primária** a sequência precisa de monômeros em uma macromolécula, como um polipeptídeo ou um ácido nucleico.
- Estrutura quaternária** em proteínas, o número e os tipos de polipeptídeos individuais na molécula proteica final.
- Estrutura secundária** o padrão inicial de dobramento de um polipeptídeo ou polinucleotídeo, normalmente ditada por possibilidades de ligações de hidrogênio.
- Estrutura terciária** a estrutura final dobrada de um polipeptídeo que tenha previamente atingido a estrutura secundária.
- Éxons** sequências de DNA codificantes em um gene em divisão (contrário de íntron).
- Fase aberta de leitura (ORF)** sequência de DNA ou RNA que pode ser traduzida, originando um polipeptídeo.
- Fita contínua** a nova fita de DNA, sintetizada ininterruptamente durante a replicação do DNA.
- Fita descontínua** a nova fita de DNA, sintetizada em fragmentos curtos durante a replicação do DNA, os quais são posteriormente unidos.
- Forquilha de replicação** sítio no cromossomo onde ocorre a replicação do DNA e onde as enzimas envolvidas na replicação ligam-se a uma fita simples e desenovelada de DNA.
- Gene** segmento de DNA que especifica uma proteína (via RNAm), um RNAt, um RNAr, ou qualquer outro RNA não codificador.
- Genoma** conjunto total de genes contidos em uma célula ou vírus.
- Iniciador** um oligonucleotídeo ao qual a DNA-polimerase é capaz de adicionar o primeiro desoxirribonucleotídeo durante a replicação do DNA.
- Íntons** sequências de DNA intervenientes não codificantes, em um gene em divisão (contrário de éxons).
- Ligação fosfodiéster** um tipo de ligação covalente que interliga os nucleotídeos em um polinucleotídeo.
- Ligação peptídica** um tipo de ligação covalente que une os aminoácidos em um polipeptídeo.
- Macromolécula informacional** qualquer molécula polimérica grande que carrega a informação genética, incluindo DNA, RNA e proteína.
- Nucleosídeo** uma base nitrogenada (adenina, guanina, citosina, timina ou uracila) mais um açúcar (ribose ou desoxirribose), porém sem fosfato.
- Nucleotídeo** monômero de ácido nucleico contendo uma base nitrogenada (adenina, guanina, citosina, timina ou uracila), uma ou mais moléculas de fosfato, e um açúcar, ribose (no RNA) ou desoxirribose (no DNA).
- Óperon** agrupamento de genes que são cotranscritos, originando um único RNA mensageiro.
- Oscilação** forma menos rígida de pareamento, permitida apenas no pareamento códon-anticódon.
- Pirimidina** uma das bases nitrogenadas dos ácidos nucleicos que contém um anel único; citosina, timina e uracila.
- Plasmídeo** elemento genético extracromossômico que não possui forma extracelular.
- Polinucleotídeo** polímero de nucleotídeos interligados por ligações covalentes denominadas ligações fosfodiéster.
- Polipeptídeo** polímero de aminoácidos interligados por ligações peptídicas
- Primase** enzima que sintetiza o iniciador de RNA utilizado na replicação do DNA.
- Processamento do RNA** conversão do transcrito primário de RNA para a sua forma madura.
- Promotor** sítio no DNA ao qual a RNA-polimerase liga-se e que inicia a transcrição.
- Proteína** um polipeptídeo ou grupo de polipeptídeos que formam uma molécula de função biológica específica.
- Purina** uma das bases nitrogenadas dos ácidos nucleicos que contém dois anéis fusionados; adenina e guanina.
- Replicação** síntese de DNA utilizando uma fita de DNA como molde.
- Replicação semiconservativa** síntese de DNA que gera duas novas duplas-hélices, cada uma consistindo em uma fita parental e uma fita-filha.
- Replissomo** um complexo de replicação de DNA que consiste em duas cópias de DNA-polimerase III, DNA-girase, helicase, primase e cópias da proteína de ligação à fita simples.
- Ribossomo** partícula citoplasmática composta por RNA ribossomal e proteínas, cuja função consiste em sintetizar proteínas.
- RNA (ácido ribonucleico)** polímero de ribonucleotídeos unidos por ligações fosfodiéster que desempenha várias funções nas células, em particular, durante a síntese proteica.

**RNA transportador (RNAt)** pequena molécula de RNA utilizada no processo de tradução que apresenta um anticódon em uma das extremidades e o aminoácido correspondente ligado à outra extremidade.

**RNA mensageiro (RNAm)** molécula de RNA que contém a informação genética necessária à codificação de um ou mais polipeptídeos.

**RNA-polimerase** enzima que sintetiza moléculas de RNA, na direção 5' → 3', utilizando uma fita complementar e antiparalela de DNA como molde.

**RNA ribossomal (RNAr)** tipos de RNA encontrados nos ribossomos; alguns participam ativamente no processo de síntese proteica.

**Sequência-sinal** sequência N-terminal especial, contendo aproximadamente 20 aminoácidos, que sinaliza que uma proteína deve ser exportada através da membrana citoplasmática.

**Spliceossomo** complexo de ribonucleoproteínas que catalisa a remoção dos íntrons dos transcritos primários de RNA.

**Término** término da elongação de uma molécula de RNA em um sítio específico.

**Tradução** processo de síntese de proteínas que utiliza a informação genética contida no RNA como molde.

**Transcrição** síntese de RNA, utilizando um molde de DNA.

**Transcrito primário** molécula de RNA não processada que é o produto direto da transcrição.

**Uso preferencial de códon** uso não aleatório de múltiplos códons que codificam o mesmo aminoácido.

## QUESTÕES PARA REVISÃO

1. Descreva o dogma central da biologia molecular. (Seção 4.1)
2. Os genes foram descobertos antes de sua natureza química ser conhecida. Primeiramente, defina um gene sem mencionar a sua natureza química. Em seguida, nomeie os compostos químicos que constituem um gene. (Seção 4.1)
3. Moléculas de DNA ricas em A-T conseguem separar-se mais facilmente em duas fitas com o aumento da temperatura do que moléculas de DNA ricas em G-C. Explique este fato com base nas propriedades do pareamento de bases AT e GC. (Seção 4.2)
4. Descreva como o DNA, que possui um comprimento muito maior que uma célula quando linearizado, consegue ser acondicionado dentro da célula. (Seção 4.2)
5. Qual o tamanho de um cromossomo de *Escherichia coli* e quantas proteínas aproximadamente este consegue codificar? Quais outros elementos genéticos podem estar presentes em *E.coli*? (Seção 4.3)
6. O que são plasmídeos R e porque eles possuem importância médica? (Seção 4.3)
7. Com relação ao DNA, o que se entende pelos termos semiconservativo, complementar e antiparalelo? (Seção 4.4)
8. Uma estrutura comumente vista no DNA circular durante a replicação é a estrutura teta. Desenhe um diagrama do processo de replicação e demonstre como a estrutura teta pode surgir. (Seções 4.5 e 4.6)
9. Por que erros na replicação do DNA são tão raros? Qual atividade enzimática, além da polimerização, está associada a DNA-polimerase III e como ela minimiza a taxa de erros? (Seção 4.6)
10. Os genes de RNAt apresentam promotores? Eles apresentam códons de início? Explique. (Seções 4.7 e 4.11)
11. Os sítios de início e término da síntese de RNAm (no DNA) são diferentes dos sítios de início e término da síntese proteica (no RNAm). Explique. (Seções 4.7 e 4.11)
12. O que é um óperon e por que ele é benéfico para conectar a expressão de determinados genes? (Seção 4.8)
13. Por que os RNAm eucarióticos precisam ser “processados”, enquanto a maioria dos RNAm procarióticos não necessita? (Seção 4.9)
14. Por que os aminoácidos possuem este nome? Escreva a estrutura geral de um aminoácido. Qual a importância do grupo R para a estrutura final da proteína? Por que o aminoácido cisteína possui uma importância especial para a estrutura proteica? (Seção 4.10)
15. O que se entende por “oscilação” e o que a torna necessária à síntese proteica? (Seções 4.11 e 4.12)
16. O que são aminoacil-RNAt sintases e que tipos de reações elas realizam? Como uma sintase reconhece seus substratos corretos? (Seção 4.12)
17. A atividade enzimática que resulta na formação da ligação peptídica no ribossomo é denominada peptidil transferase. Qual molécula catalisa essa reação? (Seção 4.13)
18. Defina os tipos de estrutura proteica: primária, secundária, terciária e quaternária. Qual destas estruturas é alterada pela desnaturação? (Seção 4.14)
19. Algumas vezes, proteínas apresentando dobramento incorreto podem ser redobradas corretamente, no entanto, algumas vezes isso não ocorre, e estas são destruídas. Quais tipos de proteínas estão envolvidos no redobramento de proteínas dobradas incorretamente? Quais tipos de enzimas estão envolvidos na degradação de proteínas incorretamente dobradas? (Seção 4.14)
20. De que modo a célula identifica quais de suas proteínas devem atuar externamente à célula? (Seção 4.14)

## QUESTÕES APLICADAS

1. O genoma da bactéria *Neisseria gonorrhoeae* é uma molécula de DNA de dupla-fita que contém 2.220 pares de quilobases. Calcule o comprimento dessa molécula de DNA em centímetros. Se 85% dessa molécula de DNA correspondem a fases de leitura aberta de genes que codificam proteínas, e se o tamanho médio das proteínas corresponde a 300 aminoácidos, quantos genes codificadores de proteína são encontrados em *Neisseria*? Que tipo de informação você imagina que estaria presente nos outros 15% de DNA?
2. Compare e diferencie as atividades de DNA e RNA-polimerases. Qual a função de cada uma delas? Quais são os substratos de cada uma? Qual a principal diferença no comportamento das duas polimerases?

3. Qual seria o resultado (em termos de síntese proteica) se a RNA-polimerase iniciasse a transcrição a uma base a montante ao seu sítio normal de início? Por quê? Qual seria o resultado (em termos de síntese proteica) se a tradução fosse iniciada a uma base a jusante ao seu sítio normal de início? Por quê?
4. No Capítulo 10, serão estudadas as mutações, alterações hereditárias na sequência de nucleotídeos em um genoma. Examinando a Tabela 4.5, discuta como o código genético evoluiu de modo a minimizar o impacto das mutações.

# 6 Genômica microbiana

## microbiologiahoje

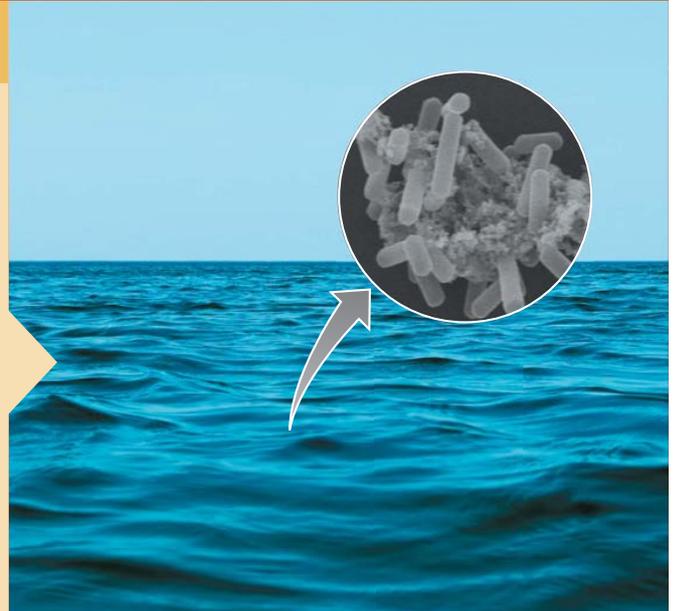
### Genômica e novas arqueias

Até recentemente, três filios de *Archaea* eram conhecidos: Euryarchaeota, Crenarchaeota e Nanoarchaeota. Curiosamente, cada espécie cultivada foi isolada de um ambiente extremo, habitats estritamente anóxicos ou extremamente quentes, salgados ou ácidos. Isso levou muitos microbiologistas a concluir que as arqueias eram principalmente extremófilas e que elas não habitavam oceanos, lagos e solos em número significativo. Mas ecologistas microbianos começaram a questionar essa suposição quando, ao usar microscopia fluorescente, detectaram arqueias somente associadas a Crenarchaeota em amostras de água doce e marinha. Quem eram esses organismos, e como eles ganharam vida?

Um grupo de microbiologistas da Universidade de Washington, em Seattle, tinha um palpite sobre o metabolismo das arqueias e começaram a tentar isolar esses organismos de amostras de água marinha (ver foto). Com persistência, paciência e boa intuição científica, o grupo isolou com sucesso a espécie *Nitrosopumilus*, a primeira arqueia oxidante de amônia (nitrificante) identificada (ver foto em detalhe). Embora muitas espécies de bactérias possam nitrificar, *Nitrosopumilus* consegue oxidar traços de amônia encontrados nas águas oceânicas, o que as bactérias nitrificantes não conseguem. Com as culturas puras destes organismos em mãos, a sua filogenia foi mais profundamente explorada usando as poderosas ferramentas da genômica. Arqueias nitrificantes são realmente apenas “Crenarchaeotas altamente divergentes”.

A genômica é capaz de responder tais questões, e cuidadosas análises dos genomas de duas arqueias nitrificantes<sup>1</sup> claramente mostraram que elas formam seu próprio filo, agora chamado de Thaumarchaeota. Análises genômicas permitiram que todos os genes dessas arqueias pudessem ser comparados aos das outras espécies de arqueias. Além de revelar um quarto filo de *Archaea*, a genômica mostrou as peculiaridades metabólicas das Thaumarchaeota, e isso forneceu uma janela dentro do papel ecológico que pode ser desempenhado em seus habitats deficientes de nutrientes.

<sup>1</sup>Spang, A., et al., 2010. Distinct gene set in two different lineages of ammonia-oxidizing *Archaea* supports the phylum *Thaumarchaeota*. *Trends in Microbiol.* 18: 331–340.



- I Investigando genomas 184
- II Genomas microbianos 191
- III Genômica funcional 198
- IV A evolução dos genomas 206

O **genoma** de um organismo é o conjunto completo de sua informação genética, incluindo seus próprios genes, suas sequências reguladoras e seu DNA não codificador. As análises genômicas resultaram no surgimento da área da genômica, o objeto de estudo deste capítulo. O conhecimento da sequência do genoma de um organismo, além de revelar os seus genes, fornece importantes informações sobre as funções do organismo e a sua história evolutiva. As sequências genômicas também auxiliam no estudo da expressão gênica, a transcrição e tradução da informação genética. A abordagem tradicional do estudo da expressão gênica enfocava um único gene ou um grupo de genes relacionados. Atualmente, a expressão do con-

junto completo de genes de um organismo pode ser examinada em um único experimento.

Avanços na genômica dependem fortemente de melhorias na tecnologia molecular e no poder da computação. Maiores avanços incluem a automatização do sequenciamento de DNA, a miniaturização dos processos de análises e o desenvolvimento de métodos computacionais potentes para análises de DNA e de sequências proteicas. Os novos avanços aparecem a cada ano, reduzindo os custos e aumentando a velocidade na qual os genomas são analisados. A partir de agora, passaremos a abordar os genomas microbianos, algumas técnicas utilizadas na análise desses genomas e o que a genômica microbiana nos revelou até o momento.

## I • Investigando genomas

O termo **genômica** refere-se à área de estudo que envolve o mapeamento, o sequenciamento, a análise e a comparação de genomas. Milhares de genomas de procariotos foram sequenciados, incluindo várias linhagens de espécies importantes de bactérias e arqueias. Devido aos novos avanços no sequenciamento de DNA que aparecem frequentemente, o número de genomas sequenciados continuará crescendo rapidamente. Hoje, o principal entrave na área da genômica são as análises e a visualização de uma grande quantidade de dados de ácidos nucleicos. No entanto, sequências genômicas continuam oferecendo conhecimentos em áreas tão distintas quanto medicina e evolução microbiana.

### 6.1 Introdução à genômica

O primeiro genoma a ser sequenciado foi o genoma de RNA de 3.569 nucleotídeos do vírus MS2 (↔ Seção 9.8), em 1976. O primeiro genoma de DNA sequenciado foi a sequência de 5.386 nucleotídeos do pequeno vírus de DNA de fita simples,  $\phi$ X174 (↔ Seção 9.3), em 1977. Já o primeiro genoma de bactéria a ser sequenciado foi o cromossomo de 1.830.137 pares de bases da *Haemophilus influenzae* publicado em 1995. As sequências de DNA de milhares de genomas procarióticos são atualmente disponibilizadas em banco de dados públicos (para obter uma lista atualizada dos projetos de sequenciamento genômico, procure no endereço <http://www.genomesonline.org/>). A **Tabela 6.1** relaciona alguns exemplos representativos. Eles incluem tanto espécies de bactérias quanto de arqueias, com representantes contendo genomas tanto lineares como circulares. Embora raros, cromossomos lineares são encontrados em várias bactérias, incluindo *Borrelia burgdorferi*, o agente etiológico da doença de Lyme, e *Streptomyces*, um importante gênero produtor de antibióticos. O tamanho dos genomas bacterianos varia de aproximadamente 0,5 a 13 megapares de bases (Mpb) e codificam de 500 a 10.000 genes codificadores de proteínas, respectivamente.

Os genomas de vários organismos superiores, incluindo o genoma humano haploide, que contém aproximadamente 3 bilhões de pares de bases, mas somente cerca de 25.000 genes codificadores de proteínas, estão sendo sequenciados. Os maiores genomas sequenciados até agora, em termos de número total de genes, pertencem à árvore choupou negro (uma espécie de *Populus*), com cerca de 45.000 genes, e ao

protozoário *Trichomonas*, com 60.000 genes estimados codificadores de proteínas, ambos possuem muito mais genes do que os seres humanos.

Os genomas de muitos patógenos foram sequenciados. Em alguns casos, várias linhagens de um patógeno foram comparadas na esperança de revelar quais genes são clinicamente relevantes. Além disso, os hipertermófilos (↔ Seção 5.12) possuem importantes aplicações biotecnológicas, uma vez que as enzimas desses organismos são termoestáveis. De fato, as necessidades das indústrias biomédica e biotecnológica foram determinantes na seleção dos organismos a serem submetidos ao sequenciamento genômico. Atualmente, no entanto, o sequenciamento genômico tornou-se uma atividade tão rotineira e barata, que os projetos deixaram de ser vinculados a razões médicas ou biotecnológicas. Em alguns casos, os genomas de várias linhagens distintas da mesma bactéria foram sequenciados, a fim de revelar o grau de variabilidade genética em uma espécie (pangenoma/genoma nuclear, Seção 6.13). A lista de genomas na Tabela 6.1 inclui organismos modelos também amplamente estudados, como o *Bacillus subtilis* (bactéria esporulante), a *Escherichia coli* (modelo de bactérias gram-negativas) e a *Pseudomonas aeruginosa* (patógeno modelo de bactérias gram-negativas).

#### MINIQUESTIONÁRIO

- Quantos genes tem o genoma humano?
- Cite alguns organismos cujos genomas são maiores do que o genoma humano.

### 6.2 Sequenciamento genômico

Em biologia, o termo **sequenciamento** refere-se à determinação da ordem precisa de subunidades em uma macromolécula. No caso do DNA (ou RNA), a sequência é a *ordem* na qual os nucleotídeos são alinhados. A tecnologia de sequenciamento de DNA está avançando tão rapidamente que dois ou três métodos aparecem todos os anos, embora poucos consigam boa aceitação no mercado ou resistam ao tempo. Isto é bem ilustrado pela redução do custo do sequenciamento de 1 megabase de DNA. Entre os anos de 2001 a 2011 a redução foi de 10.000 vezes! A **Tabela 6.2** resume os métodos de sequenciamento que serão discutidos aqui.

**Tabela 6.1** Genomas procarióticos selecionados<sup>a</sup>

Organismo	Modo de vida <sup>b</sup>	Tamanho (pb)	ORFs <sup>c</sup>	Comentários
<b>Bactéria</b>				
<i>Hodgkinia cicadicola</i>	E	143.795	169	Endossimbionte de cigarra degenerada
<i>Carsonella ruddii</i>	E	159.662	182	Endossimbionte de psilídeo degenerado
<i>Buchnera aphidicola</i> BCC	E	422.434	362	Endossimbionte de afídeo
<i>Mycoplasma genitalium</i>	P	580.070	470	Menor genoma de bactéria não simbiote
<i>Borrelia burgdorferi</i>	P	910.725	853	Espiroqueta, cromossomo linear, causador da doença de Lyme
<i>Rickettsia prowazekii</i>	P	1.111.523	834	Parasita intracelular obrigatório, causador do tifo epidêmico
<i>Treponema pallidum</i>	P	1.138.006	1.041	Espiroqueta, causador da sífilis
<b>Família Methylophilaceae, linhagem HTCC2181</b>				
<i>Aquifex aeolicus</i>	VL	1.551.335	1.544	Hipertermófilo, autotrófico
<i>Prochlorococcus marinus</i>	VL	1.657.990	1.716	Fototrófico oxigênico marinho mais abundante
<i>Streptococcus pyogenes</i>	VL	1.852.442	1.752	Causador da faringite estreptocócica e febre escarlatina
<i>Thermotoga maritima</i>	VL	1.860.725	1.877	Hipertermófilo
<i>Chlorobaculum tepidum</i>	VL	2.154.946	2.288	Modelo de bactéria verde fototrófica
<i>Deinococcus radiodurans</i>	VL	3.284.156	2.185	Resistente à radiação, cromossomos múltiplos
<i>Synechocystis</i> sp.	VL	3.573.470	3.168	Modelo de cianobactéria
<i>Bdellovibrio bacteriovorus</i>	VL	3.782.950	3.584	Predador de outros procariotos
<i>Caulobacter crescentus</i>	VL	4.016.942	3.767	Ciclo de vida complexo
<i>Bacillus subtilis</i>	VL	4.214.810	4.100	Modelo genético de gram-positivos
<i>Mycobacterium tuberculosis</i>	P	4.411.529	3.924	Causador da tuberculose
<i>Escherichia coli</i> K12	VL	4.639.221	4.288	Modelo genético de gram-negativos
<i>Escherichia coli</i> O157:H7	VL	5.594.477	5.361	Linhagem enteropatogênica de <i>E. coli</i>
<i>Bacillus anthracis</i>	VL	5.227.293	5.738	Patógeno, arma biológica
<i>Pseudomonas aeruginosa</i>	VL	6.264.403	5.570	Patógeno oportunista metabolicamente versátil
<i>Streptomyces coelicolor</i>	VL	8.667.507	7.825	Cromossomo linear, produz antibióticos
<i>Bradyrhizobium japonicum</i>	VL	9.105.828	8.317	Fixação de nitrogênio, causa nódulos em soja
<i>Sorangium cellulosum</i>	VL	13.033.799	9.367	Mixobactéria, forma corpos de frutificação multicelulares
<b>Arqueia</b>				
<i>Nanoarchaeum equitans</i>	P	490.885	552	Menor genoma celular não simbiótico
<i>Thermoplasma acidophilum</i>	VL	1.564.905	1.509	Termófilo, acidófilo
<i>Methanocaldococcus jannaschii</i>	VL	1.664.976	1.738	Metanogênico, hipertermófilo
<i>Pyrococcus horikoshii</i>	VL	1.738.505	2.061	Hipertermófilo
<i>Halobacterium salinarum</i>	VL	2.571.010	2.630	Halófilo extremo, bacteriorrodopsina
<i>Sulfolobus solfataricus</i>	VL	2.992.245	2.977	Hipertermófilo, quimiolitotrófico sulfuroso
<i>Haloarcula marismortui</i>	VL	4.274.642	4.242	Halófilo extremo, bacteriorrodopsina
<i>Methanosarcina acetivorans</i>	VL	5.751.000	4.252	Metanogênico que usa o acetato

<sup>a</sup>Informações sobre genomas de procariotos podem ser encontradas no endereço eletrônico <http://cmr.jcvi.org>, um site mantido pelo instituto The J. Craig Venter Institute, em Rockville, MD, e no link <http://www.genomesonline.org>.

<sup>b</sup>E, endossimbionte; P, parasita; VL, de vida livre.

<sup>c</sup>Fases de leitura aberta (ORFs). Os genes que codificam proteínas conhecidas estão incluídos, bem como todas as ORFs que poderiam codificar proteínas com mais de 100 resíduos de aminoácidos. ORFs menores normalmente não são incluídas, a menos que exibam similaridade a um gene de outro organismo, ou se a utilização preferencial de códons for típica do organismo em estudo.

### Sequenciamento de DNA de primeira geração: o método didesoxi de Sanger

O primeiro método amplamente utilizado para sequenciamento de DNA foi o didesoxi inventado pelo cientista britânico Fred Sanger, que foi agraciado com o Prêmio Nobel por essa descoberta. Embora substituído por novas tecnologias de sequenciamento genômico, o método didesoxi é ainda utilizado para algumas aplicações. Sanger introduziu inúmeros conceitos importantes que ainda são utilizados em técnicas mais novas de sequenciamento. Estas incluem sequenciamento através da síntese de DNA em vez da quebra da molécula,

usando didesoxinucleotídeos para bloquear a extensão da cadeia e precursores marcados para detecção.

No sequenciamento através da síntese, pequenos oligonucleotídeos de DNA (geralmente de 10 a 20 nucleotídeos) com sequência definida são usados como iniciadores. Estes são sintetizados artificialmente. **Iniciadores** são pequenos segmentos de DNA ou de RNA que iniciam a síntese de novas fitas de ácido nucleico. Durante a replicação *in vivo*, iniciadores de RNA são utilizados (↻ Seção 4.4), mas na biotecnologia, iniciadores de DNA são utilizados porque eles são mais estáveis do que os de RNA.

**Tabela 6.2** Métodos de sequenciamento de DNA

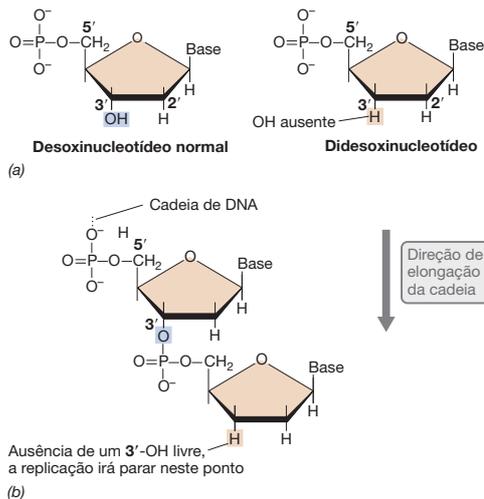
Geração	Método	Características
<b>Primeira geração</b>	Método dideoxi de Sanger (radioatividade ou fluorescência; amplificação de DNA)	Comprimento de leitura: 700-900 bases Usado no projeto do genoma humano
<b>Segunda geração</b>	Pirossequenciamento 454 (fluorescência; amplificação de DNA; massivo em paralelo) Illumina/método Solexa (fluorescência; amplificação de DNA; massivo em paralelo)	Comprimento de leitura: 400-500 bases Usado para sequenciar o genoma de James Watson (finalizado em 2007) Comprimento de leitura: 50-100 bases Genoma do panda gigante (2009; Beijing Genome Institute) Genoma do Denisovan (2010)
	Método SOLiD (fluorescência; amplificação de DNA; massivo em paralelo)	Comprimento de leitura: 50-100 bases
<b>Terceira geração</b>	Sequenciador HelixScope de molécula única (fluorescência; molécula única) Pacific Biosciences SMRT (fluorescência; molécula única, <i>zero-mode waveguide</i> )	Comprimento de leitura: até 55 bases Melhoramento da precisão do DNA de fósseis Comprimento de leitura: 2.500-3.000 bases
<b>Quarta geração</b>	Íon <i>torrent</i> (eletrônico – pH; amplificação de DNA)	Comprimento de leitura: 100-200 bases Sequenciou o genoma do Gordon Moore, cofundador da Intel (autor da lei de Moore), 2011
	Oxford nanoporo (eletrônico – atual; molécula única, em tempo real)	Comprimento de leitura: milhares de bases A unidade portátil MinION é aproximadamente do tamanho de um dispositivo USB

No procedimento de Sanger, a sequência é determinada pela síntese de uma cópia da fita simples do DNA, utilizando a enzima DNA-polimerase. Como vimos anteriormente (☞ Seção 4.4), essa enzima adiciona desoxinucleotídeos trifosfato à cadeia de DNA crescente. Entretanto, no sequenciamento de Sanger, pequenas quantidades de um análogo dideoxinucleotídeo são incluídas em cada uma das quatro reações de incubação, uma para cada uma das quatro bases – adenina, guanina, citosina e timina (Figura 6.1). O análogo dideoxi atua como um reagente específico de *terminação de cadeia*, isso porque o açúcar dideoxi não possui a hidroxila-3', impedindo a elongação da cadeia após a sua inserção. Devido à inserção

aleatória dos dideoxinucleotídeos, fragmentos de DNA de tamanhos variados são obtidos e separados em gel de eletroforese (Figura 6.1).

Inicialmente quatro reações separadas (e quatro canaletas individuais no gel) foram utilizadas para a determinação de cada sequência, uma para cada fragmento terminando com uma das quatro bases. As posições das bandas foram localizadas usando precursores marcados (originalmente radioativos, mas atualmente fluorescentes). Ao alinhar as canaletas dos quatro dideoxinucleotídeos e observar a posição vertical de cada fragmento relativa ao seu vizinho, a sequência de DNA é lida diretamente do gel (Figura 6.2).

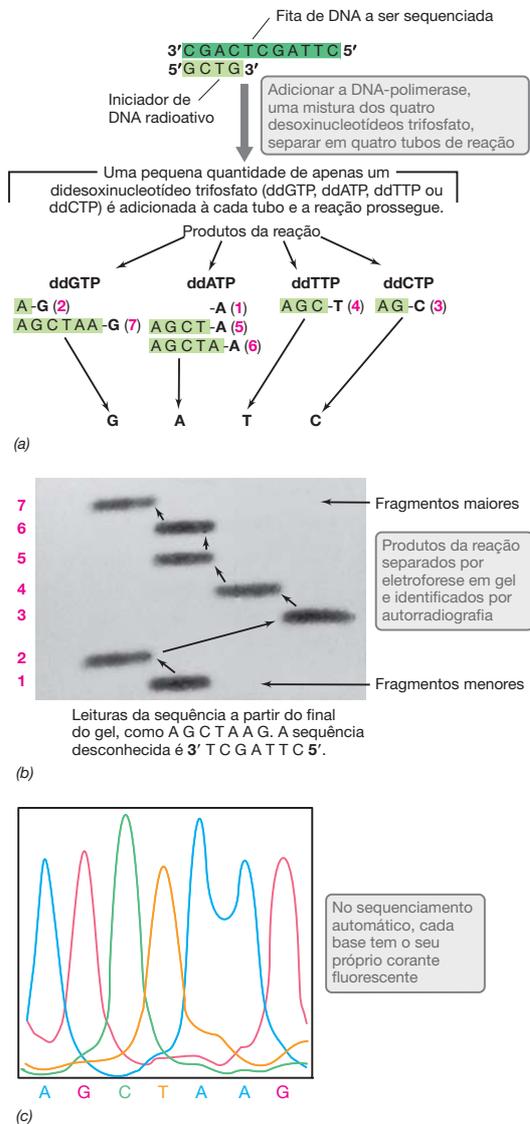
Sistemas de sequenciamento de DNA automáticos utilizam iniciadores (ou nucleotídeos) marcados com corantes fluorescentes em substituição à radioatividade. Os produtos são separados por eletroforese capilar e as bandas são digitalizadas por um *laser* de detecção de fluorescência. Uma vez que cada uma das quatro diferentes reações utiliza um marcador fluorescente de cores distintas, as quatro reações podem ser iniciadas em um único poço. Os resultados são analisados em computador (Figura 6.2).



**Figura 6.1** Dideoxinucleotídeos e o sequenciamento de Sanger. (a) Um desoxinucleotídeo normal possui um grupo hidroxil no carbono 3', ao passo que um dideoxinucleotídeo não o possui. (b) A elongação da cadeia é interrompida onde um dideoxinucleotídeo é incorporado.

### Sequenciamento *shotgun*

O sequenciamento *shotgun* se refere à *preparação* do DNA para o sequenciamento, e não o sequenciamento propriamente dito. A maioria dos projetos de sequenciamento genômico emprega o sequenciamento *shotgun*. A análise de um genoma usualmente começa com a construção de uma **biblioteca genômica** – clonagem molecular de fragmentos de DNA que cobrem todo o genoma (☞ Seção 11.4). Na abordagem *shotgun*, o genoma inteiro, clivado em fragmentos, é clonado. Os fragmentos são então sequenciados. Neste ponto, a ordem e a orientação dos fragmentos de DNA são desconhecidas. As sequências são analisadas por um computador que busca por sequências sobrepostas e monta os fragmentos sequenciados na ordem correta. Por sua própria natureza, grande parte do sequenciamento do método *shotgun* é redundante. Para



**Figura 6.2** Sequenciamento de DNA utilizando o método de Sanger (a) Observe que quatro reações diferentes devem acontecer, uma com cada didesoxinucleotídeo. Devido ao fato de as reações serem feitas *in vitro*, o iniciador para a síntese de DNA pode ser de DNA. (b) Porção do gel contendo os produtos da reação da parte a. (c) Resultados do sequenciamento do mesmo DNA mostrado nas partes a e b, mas dessa vez utilizando um sequenciador automático e marcadores fluorescentes. Os fragmentos de DNA são separados pelo tamanho em uma única coluna capilar, e cada didesoxinucleotídeo marcado com fluorescência é detectado por um *laser* detector.

assegurar que a sequência completa do genoma seja obtida, é necessário sequenciar um número muito grande de clones, muitos dos quais podem ser idênticos, ou praticamente idênticos. Em geral, haverá entre 7 a 10 sequências repetidas para uma região qualquer do genoma. Essa cobertura de 7 a 10 vezes reduz significativamente os possíveis erros na sequência,

uma vez que a redundância no sequenciamento permite que um nucleotídeo consenso seja selecionado em qualquer região da sequência onde haja ambiguidade.

Para o sucesso do sequenciamento *shotgun*, a clonagem deve ser eficiente (há a necessidade de um grande número de clones) e, na medida do possível, os segmentos de DNA clonados devem ser gerados aleatoriamente. Isso pode ser feito através de digestão enzimática do DNA ou por métodos físicos. Os fragmentos de DNA podem ser purificados por tamanho através do gel de eletroforese (↔ Seção 11.1) antes de serem clonados e sequenciados.

### Sequenciamento de segunda geração

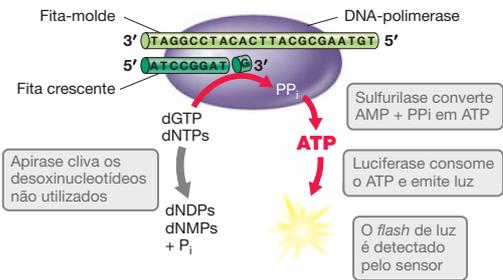
O termo “geração” no sequenciamento de DNA se refere às sucessivas mudanças na tecnologia que conferem um significativo aumento na velocidade combinada com a redução do custo do sequenciamento. A característica que define o sequenciamento de *segunda geração* é o uso de *métodos massivos paralelos*. Em outras palavras, um grande número de amostras é sequenciado lado a lado na mesma máquina. Dois requerimentos principais para que isso ocorra são: a miniaturização e a melhora do poder dos computadores. Métodos do sequenciamento de segunda geração geram dados 100 vezes mais rápido do que os métodos mais antigos. Os três métodos mais utilizados são o pirosequenciamento 454 da Life Sciences, o Illumina/Solexa e o SOLiD/Applied Biosystems.

No sistema 454, a amostra de DNA é fragmentada em segmentos de fita simples de aproximadamente 100 bases cada, sendo cada fragmento imobilizado em uma esfera microscópica. O DNA é amplificado pela reação de polimerização em cadeia (PCR, ↔ Seção 11.3), na qual, ao final do processo, cada esfera conterá uma série de cópias idênticas à fita de DNA. As esferas são então depositadas em uma placa de fibra óptica contendo mais de um milhão de poços, cada um contendo uma esfera.

Assim como no sequenciamento de Sanger (Figura 6.2), o princípio do sequenciamento 454 envolve a síntese de uma fita complementar pela DNA-polimerase (Figura 6.3). Todavia, em vez de ocorrer a terminação de cadeia, no método 454, cada vez que um nucleotídeo é incorporado na fita complementar, uma molécula de pirofosfato é liberada, fornecendo a energia necessária à liberação de luz, pela enzima *luciferase*, também incorporada ao sistema. Os quatro nucleotídeos são aplicados sequencialmente sobre a placa em uma ordem fixa. Assim, cada pulso de luz identifica qual base foi inserida. O método Illumina/Solexa se assemelha ao sequenciamento de Sanger, uma vez que ambos utilizam a síntese de DNA e nucleotídeos de terminação da cadeia. Entretanto, no sistema Illumina, os nucleotídeos que interrompem a cadeia são desoxi (em vez de didesoxi) e podem ser incorporados reversivelmente. Além disso, cada um dos quatro diferentes desoxinucleotídeos carrega a sua própria molécula fluorescente que funciona como um grupo bloqueador para a ligação do 3'-OH, causando, então, a terminação da cadeia.

### Terceira e quarta gerações do sequenciamento de DNA

A característica-chave do sequenciamento de *terceira geração* é o sequenciamento de *moléculas únicas* de DNA. Existem



**Figura 6.3** Mecanismo do pirosequenciamento. Sempre que um novo desoxinucleotídeo é inserido na fita de DNA crescente (setas vermelhas), o pirofosfato (PPi) é liberado e utilizado, pela enzima sulfuriase, para sintetizar ATP de AMP. O ATP é consumido pela enzima luciferase produzindo luz. Desoxinucleotídeos não utilizados são degradados pela enzima apirase (seta cinza).

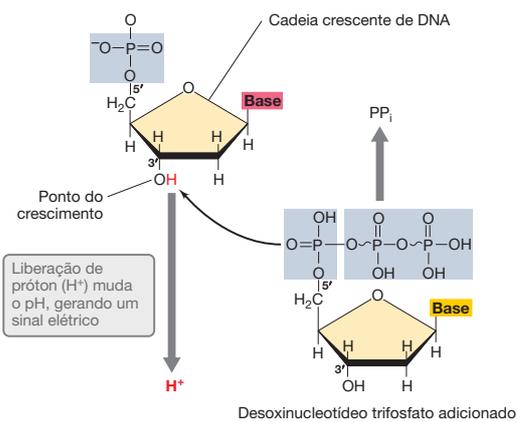
duas abordagens principais: uma baseada em microscopia e a outra em nanotecnologia. No *sequenciador HeliScope de molécula única*, fragmentos de fita simples de DNA com cerca de 32 bases de comprimento são ligados a uma matriz em uma lâmina de vidro. À medida que a fita complementar é sintetizada, os marcadores fluorescentes dos nucleotídeos incorporados são monitorados em um microscópio. O equipamento pode monitorar um bilhão de fragmentos de DNA simultaneamente. Em seguida, um computador monta os fragmentos em uma sequência completa.

O sequenciamento Pacific Biosciences SMRT (molécula única em tempo real) usa uma técnica conhecida como *zero-mode waveguides*. Neste método, a DNA-polimerase estende uma fita crescente pela adição de desoxinucleotídeos marcados com quatro corantes fluorescentes diferentes. Estes desoxinucleotídeos emitem um *flash* de luz quando são incorporados à fita de DNA. Duas características inovadoras são fundamentais para o sequenciamento de moléculas únicas. Em primeiro lugar, as reações são realizadas no interior de nanocontenedores. Estes são minúsculos poços cilíndricos

de metal com 20 nm de largura, que reduzem a interferência da luminosidade o suficiente para permitir a detecção de um *flash* de luz oriundo de um único nucleotídeo. Em segundo lugar, os marcadores fluorescentes são ligados ao grupo pirofosfato que é descartado quando o desoxinucleotídeo é incorporado à cadeia. Assim, etiquetas marcadoras não se acumulam na fita de DNA; em vez disso, cada reação libera uma explosão microscópica de cor.

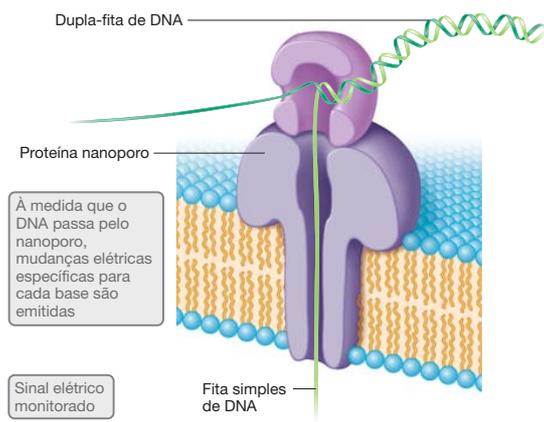
A característica-chave do sequenciamento de *quarta geração*, também chamado de “sequenciamento pós-luz”, é que a detecção óptica não é mais utilizada. O *método ion torrent* não usa o sequenciamento de molécula única. Em vez de usar desoxinucleotídeos marcados, ele mede a liberação de prótons ( $H^+$ ) à medida que o novo desoxinucleotídeo é adicionado à fita crescente de DNA (Figura 6.4a). Um *chip* de silício apelidado de “o menor medidor de pH do mundo” detecta os prótons. O sequenciamento é extremamente rápido por este método e os instrumentos são bem mais baratos do que aqueles utilizados nas metodologias anteriores. Por exemplo, o equipamento *ion torrent* é capaz de sequenciar o genoma humano inteiro – quase 3.000 Mbp – em menos de um dia!

A tecnologia nanoporo (Figura 6.4b) é baseada na maquinaria microscópica que opera na escala de moléculas únicas. Os detectores de DNA do nanoporo são poros extremamente estreitos que permitem somente a passagem de uma única molécula de DNA por vez. O sistema Oxford Nanopore Technologies passa o DNA por um poro biológico em nanoescala feito de uma proteína (Figura 6.4b). À medida que a molécula de DNA transita pelo poro, o detector registra a mudança na corrente elétrica através do nanoporo. Essa mudança é diferente para cada uma das bases ou para combinações destas bases. As principais vantagens da tecnologia nanoporo são a alta velocidade e sua habilidade de sequenciar moléculas longas de DNA (em vez de pequenos fragmentos, como a maioria dos outros métodos). Além disso, muitos nanoporos podem ser montados em uma área muito pequena ou em um *chip*, no



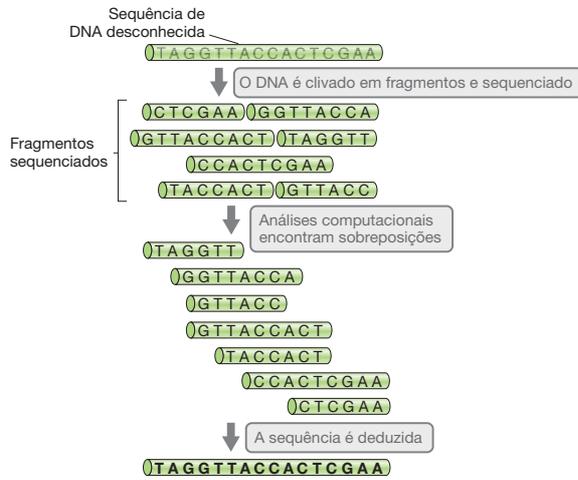
(a) Sequenciamento *ion torrent* semicondutor

**Figura 6.4** Sequenciamento de quarta geração. (a) O sistema de sequenciamento semicondutor *ion torrent* é baseado na liberação de prótons ( $H^+$ ) cada vez que um novo desoxinucleotídeo é inserido na fita crescente de DNA. A mudança de pH resultante é detectada por um eletrodo. (b) No sequenciamento



(b) Sequenciamento nanoporo

nanoport, a dupla-hélice de DNA é convertida em uma fita simples para passar através de um poro. À medida que o DNA atravessa o nanoporo, é causada uma mudança específica na carga elétrica para cada base.



**Figura 6.5** Montagem computacional da sequência de DNA. A maioria dos métodos de sequenciamento de DNA gera uma grande quantidade de pequenas sequências (de 30 a várias centenas de bases) que devem ser montadas. O computador procura por regiões de sobreposição nas pequenas sequências e, então, as organiza para formar uma única sequência global.

qual muitos fragmentos longos de DNA podem ser sequenciados simultaneamente.

**Montagem genômica**

Independentemente de como o DNA é sequenciado, as sequências devem ser montadas para depois serem analisadas. A *montagem* genômica consiste em colocar os fragmentos na ordem correta e eliminar as regiões de sobreposição. Na prática, um computador examina vários fragmentos pequenos de DNA que foram sequenciados e deduz a ordem desses fragmentos pelas sobreposições (Figura 6.5). A montagem gera um genoma apropriado para a *anotação*, processo de identificação de genes e outras regiões funcionais no genoma (discutido na próxima seção).

Algumas vezes, o sequenciamento e a montagem não geram uma sequência genômica completa, havendo lacunas na sequência nucleotídica. Em tais situações, uma variedade de abordagens é utilizada para a obtenção de sequências individuais que cubram as lacunas. Alguns projetos de sequenciamento genômico têm o objetivo de gerar um *genoma fechado*, isto é, a sequência genômica inteira é determinada. Outros projetos são interrompidos no *estágio de versão inicial* (ou rascunho), dispensando o sequenciamento das pequenas lacunas. Uma vez que o sequenciamento e a montagem são procedimentos essencialmente automatizados, enquanto o preenchimento das lacunas não é, a obtenção de um genoma fechado é muito mais dispendiosa e consome mais tempo do que uma versão inicial de uma sequência genômica e normalmente precisa mais da mão de obra humana para completar o serviço.

**MINIQUESTIONÁRIO**

- O que é o sequenciamento *shotgun*?
- Quais são as características que definem os sequenciamentos de terceira e quarta gerações?
- O que é realizado durante a montagem do genoma?

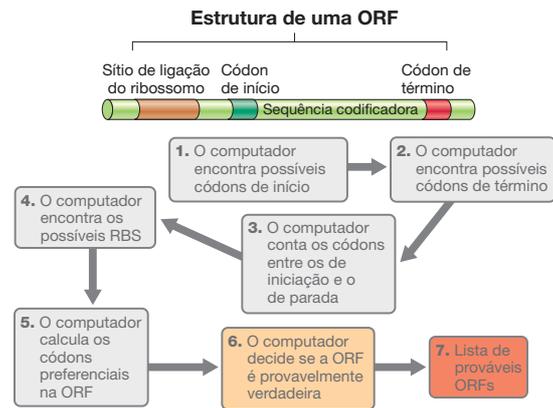
**6.3 Bioinformática e anotação genômica**

Uma vez finalizados o sequenciamento e a montagem, a próxima etapa na análise genômica consiste na *anotação genômica*, a conversão dos dados de sequências crús em uma lista de genes e outras sequências funcionais presentes no genoma. **Bioinformática** se refere ao uso do computador para armazenar e analisar as sequências e as estruturas dos ácidos nucleicos e proteínas. As melhoras nos métodos de sequenciamento (Seção 6.2) estão gerando dados mais rápido do que eles possam ser analisados. Com isso, no momento atual, a anotação é o “gargalo” da genômica.

A maioria de genes de qualquer organismo codifica proteínas, e, na maioria dos genomas microbianos, especialmente os de procariontos, a maior parte do genoma consiste em sequências codificadoras. Devido ao fato de os genomas eucarióticos microbianos possuírem, geralmente, um menor número de íntrons (íntrons, ⇨ Seção 4.9) do que genomas de plantas e animais, e os procariontos serem praticamente desprovidos deles, os genomas microbianos consistem, essencialmente, em centenas a milhares de **fases abertas de leitura** (ORFs, *open reading frames*) separadas por pequenas regiões reguladoras e de terminadores de transcrição. Lembrando que uma fase de leitura aberta é uma sequência de DNA ou RNA que pode ser traduzida produzindo um polipeptídeo (⇨ Seção 4.11).

**Como um computador encontra uma ORF?**

Uma *ORF funcional* é aquela que, de fato, codifica uma proteína na célula. Assim, a forma mais simples de localizar genes que potencialmente codificam proteínas consiste na busca, por meio de um computador, de ORFs na sequência do genoma (Figura 6.6). Embora um dado gene seja sempre transcrito de uma única fita, ambas as fitas são transcritas em algumas partes do genoma (em todos os genomas, com exceção dos menores plasmídeos ou de genomas virais). Com isso, um computador que inspecione ambas as fitas do DNA é requerido.



**Figura 6.6** Identificação computacional de possíveis ORFs. O computador verifica a sequência de DNA procurando primeiro pelos códons de início e de término. Depois ele conta os números de códons em cada janela de leitura interrompida e rejeita aqueles que são muito curtos. A probabilidade de ser uma ORF verdadeira se fortalece quando um provável sítio de ligação ribossomal (RBS, *ribosomal binding site*) é encontrado em uma distância correta na parte anterior à fase de leitura. Os cálculos dos códons preferenciais são utilizados para testar se uma ORF está em conformidade com a utilização dos códons pelo organismo que está sendo examinado.

**Tabela 6.3** Exemplos de códons preferenciais

Códon da arginina <sup>a</sup>	Utilização de cada códon da arginina (%)		
	<i>Escherichia coli</i>	<i>Mosca-da-fruta</i>	<i>Ser humano</i>
AGA	1	10	22
AGG	1	6	23
CGA	4	8	10
CGC	39	49	22
CGG	4	9	14
CGU	49	18	9

<sup>a</sup>A arginina tem seis códons; ver Tabela 4.5.

O primeiro passo para encontrar uma ORF é localizar o códon de início e o de término em uma sequência (↔ Seção 4.11 e Tabela 4.5). No entanto, os códons de início e de término aparecem aleatoriamente dentro da fase de leitura com certa frequência. Com isso, mais pistas são necessárias. A maioria das proteínas celulares contém 100 ou mais aminoácidos, de modo que a maioria das ORFs funcionais possui mais de 100 códons (300 nucleotídeos). No entanto, a simples programação do computador, no sentido de ignorar ORFs com menos de 100 códons, resultará na perda de algumas ORFs genuínas, que correspondem a pequenos genes. Nas espécies do domínio *Bacteria*, a tradução se inicia nos códons de início localizados imediatamente a jusante de uma sequência de ligação do ribossomo (sítio Shine-Dalgarno) no RNAm (↔ Seção 4.13). Portanto, a busca de potenciais sequências de Shine-Dalgarno na sequência de DNA de um genoma procariótico pode auxiliar na definição sobre a funcionalidade de uma ORF e sobre qual códon de início é efetivamente utilizado.

Existe mais de um códon para vários dos 20 aminoácidos comuns (↔ Tabela 4.5), e sendo alguns códons utilizados com maior frequência que outros. Este último é conhecido como **códon preferencial** (utilização de códon) e difere muito entre os organismos. Por exemplo, a Tabela 6.3 mostra o uso diferente dos seis códons existentes para arginina em *Escherichia coli*, seres humanos e moscas-da-fruta. Se a utilização de códons de uma determinada ORF for consideravelmente diferente dos códons consenso de uso preferencial, a ORF pode ser funcional, adquirida por transferência horizontal de genes ou não (Seção 6.12).

Uma ORF é também provavelmente funcional se sua sequência for similar a sequências de ORFs em genomas de outros organismos (independentemente de estas codificarem proteínas conhecidas), ou se alguma parte da ORF apresentar uma sequência que codifica um domínio proteico funcional conhecido. Isso se deve ao fato de proteínas de diferentes células, com funções similares, tenderem a ser homólogas; isto é, elas são relacionadas em um sentido evolutivo, geralmente compartilhando características estruturais e de sequência de aminoácidos (Seção 6.11). Computadores são utilizados na busca de tais similaridades de sequência em banco de dados como o GenBank. Esse banco de dados contém mais de 200 bilhões de pares de bases de sequências e pode ser acessado no endereço eletrônico: <http://www.ncbi.nlm.nih.gov/Genbank/>. A ferramenta de pesquisa mais utilizada é o BLAST (do inglês, *basic local alignment search tool* [ferramenta básica de busca de alinhamento local]), que exibe inúmeras variantes, dependendo da sequência utilizada na busca corresponder a ácido

nucleico ou proteína. Por exemplo, a ferramenta BLASTn pesquisa banco de dados de ácidos nucleicos, utilizando uma sequência de nucleotídeos, enquanto BLASTp pesquisa banco de dados de proteínas, utilizando uma sequência de aminoácidos.

### ORFs não caracterizadas

Embora existam diferenças entre os organismos, na maioria dos genomas o número de genes que possuem o seu papel claramente identificado é de apenas 70% do total das ORFs detectadas. ORFs não caracterizadas (ou desconhecidas) codificam *proteínas hipotéticas*, proteínas que provavelmente existem, embora tenham funções ainda desconhecidas. ORFs não caracterizadas possuem uma fase de leitura ininterrupta de comprimento razoável e os códons de início e de término necessários (Figura 6.6). Entretanto, as proteínas que elas codificam não possuem homologia suficiente com as sequências de aminoácidos de qualquer proteína conhecida para serem identificadas.

Quando a função de um gene é identificada em um organismo, essa função pode ser atribuída às ORFs homólogas em outros organismos. No entanto, a maioria dos genes da síntese de macromoléculas e do metabolismo central essencial para o crescimento do organismo já foram identificados. Portanto, a maior parte das ORFs ainda não identificadas codifica proteínas não essenciais.

Muitos dos genes não identificados em *E. coli* são preditos para codificar proteínas reguladoras ou redundantes. Isso inclui as proteínas necessárias apenas sob condições especiais ou “backups” para enzimas-chave. Entretanto, mesmo em um organismo bem estudado como a *E. coli*, as funções exatas de muitos genes são frequentemente imprevisíveis. Algumas identificações de genes apenas os atribuem a uma família ou função geral (como “transportadores”). Em contrapartida, outros genes são completamente desconhecidos e somente foram preditos pela bioinformática. Além disso, algumas anotações estão realmente incorretas. Na verdade, estima-se que pelo menos 10% dos genes dos bancos de dados foram anotados incorretamente.

### RNA não codificador

Além dos genes que codificam proteínas, existem alguns genes que codificam moléculas que não são traduzidas. Estes genes, portanto, não possuem o códon de início e podem ter múltiplos códons de término dentro da sequência. Além disso, esses genes não possuem os códons preferenciais; consequentemente, eles não serão reconhecidos pelos programas que procuram somente ORFs. Alguns RNAs não codificadores podem ser facilmente localizados porque eles são bem caracterizados e possuem uma sequência altamente conservada. Esses incluem os RNAs e os RNAs. No entanto, muitas moléculas de RNA não codificantes com função reguladora (↔ Seção 7.14) são conservadas somente na estrutura tridimensional, com pouca homologia na sequência. Identificá-los durante a anotação do genoma continua sendo um desafio.

### MINIQUESTIONÁRIO

- O que é uma fase de leitura aberta (ORF)? O que é uma proteína hipotética?
- Como a homologia das proteínas pode ajudar na anotação do genoma?
- Qual é o principal problema na identificação de genes que codificam RNAs que não são traduzidos?

## II • Genomas microbianos

### 6.4 O tamanho do genoma e o seu conteúdo

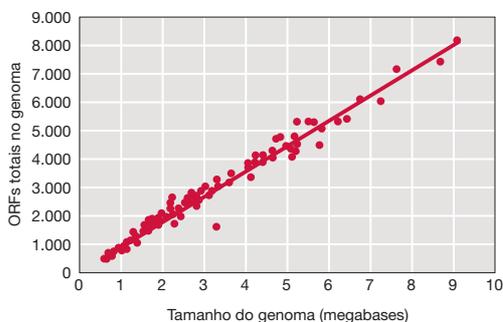
Após o sequenciamento, montagem e anotação, a genômica comparativa pode ser usada para comparar os tamanhos dos genomas, as organizações e os conteúdos dos genes. O *website* The Microbes Online (<http://www.microbesonline.org>) tem aproximadamente 4.000 genomas de microrganismos disponíveis para visualização.

#### Tamanhos dos genomas procarióticos

Os genomas de espécies dos dois domínios procarióticos, *Bacteria* e *Archaea*, exibem uma forte relação entre tamanho de genoma e conteúdo de ORFs (Figura 6.7). Independentemente do organismo, cada megabase de DNA procariótico codifica cerca de 1.000 ORFs. À medida que os genomas procarióticos aumentam de tamanho, há o aumento proporcional no número de genes. Isso difere dos eucariotos, nos quais DNA não codificador (introns, Seção 4.9) pode corresponder a uma grande fração do genoma, especialmente em organismos com grandes genomas.

Análises de sequências genômicas podem fornecer respostas para questões biológicas fundamentais. Por exemplo, quantos genes são necessários para a existência da vida? O recorde do menor genoma para um organismo de vida livre pertence a uma espécie de bactéria, conhecida como HTCC2181, cujo genoma contém 1.304.428 pb e 1.354 genes. A detentora do recorde anterior é a espécie *Pelagibacter ubique*, um heterotrófico marinho, que possui apenas 4.331 pb a mais, sugerindo que esta espécie está perto do limite prático para uma vida independente. HTCC2181 é uma bactéria metilotrófica não cultivável (metilotróficos são organismos que catabolizam compostos de carbono, tal como o metanol) que é comum em ecossistemas marinhos costeiros.

Várias outras espécies de vida livre pertencentes à *Bacteria* e *Archaea* são conhecidas por terem genomas com cerca de 1.400 genes (Tabela 6.1). Esses organismos são extremamente eficientes no uso do seu DNA. Eles possuem poucos ou não possuem introns, inteínas ou transposons, e apresentam os menores espaçamentos intergênicos descritos até o momento. Os maiores genomas de procariotos contêm mais de 10.000 genes, corres-



**Figura 6.7** Correlação entre tamanho de genoma e conteúdo de ORFs em procariotos. Estes dados são oriundos de análises de 115 genomas procarióticos completos e incluem espécies de bactérias e arqueias. Dados de *Proc. Natl. Acad. Sci. (USA)* 101: 3160-3165 (2004).

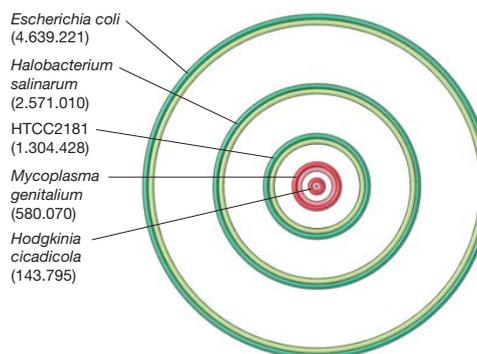
pondendo principalmente a organismos de solo, como as mixobactérias, procariotos que realizam um ciclo de vida complexo (Seção 14.19). A Figura 6.8 mostra, em escala, cinco genomas circulares dos procariotos selecionados para a visualização do qual variável pode ser o tamanho dos genomas procarióticos.

Talvez surpreendentemente, as análises genômicas demonstraram que organismos autotróficos necessitam apenas de alguns poucos genes a mais do que os heterotróficos (Seção 3.3). Por exemplo, *Methanocaldococcus jannaschii* (*Archaea*) é um autotrófico e contém somente 1.738 ORFs. Isso permite que, além de uma vida livre, utilize o dióxido de carbono como fonte única de carbono. *Aquifex aeolicus* (*Bacteria*) é também um autotrófico e contém o menor genoma conhecido de autotróficos, com apenas 1,5 par de megabases (Tabela 6.1). *Methanocaldococcus* e *Aquifex* são também hipertermófilos, com crescimento ótimo em temperaturas acima de 80°C. Assim, um grande genoma não é necessário para sustentar um estilo de vida extremo, incluindo aquele de autotróficos que vivem em água quase fervente.

#### Genomas pequenos

Os menores genomas celulares pertencem a procariotos parasitas ou endossimbiontes (células que vivem dentro de outras células). Os tamanhos de genomas de procariotos parasitas obrigatórios variam de 490 Kpb, para *Nanoarchaeum equitans* (*Archaea*) a 4.400 Kpb, para *Mycobacterium tuberculosis* (*Bacteria*). Os genomas de *N. equitans* e de várias outras bactérias, incluindo *Mycoplasma*, *Chlamydia* e *Rickettsia*, são menores que o maior genoma viral conhecido, o do *Mimivirus* de 1,2\* Mpb (Seção 9.2). A endossimbionte da cigarra degenerada, *Hodgkinia*, tem um pequeno genoma, com menos de 150 Kpb (Figura 6.8, ver também Figura 6.14).

Todos os genomas menores do que 1,2 Mpb são encontrados em bactérias que são dependentes de outras células em algum aspecto para a sua existência. Micoplasmas, com



**Figura 6.8** Comparação do tamanho dos genomas. Os genomas circulares de alguns procariotos são mostrados em escala. O número de nucleotídeos está indicado embaixo de cada espécie. Os círculos verdes indicam organismos de vida livre e os círculos vermelhos indicam parasitas (*Mycoplasma*) e simbiontes de insetos (*Hodgkinia*).

\*N. de R.T. Até o momento da edição deste livro, vírus com genomas ainda maiores que os dos mimivirus já foram isolados e identificados, como é o caso dos *Pandoravirus*, cujo genoma tem aproximadamente 1,9 Mpb.

genomas de pouco mais de 500 Kpb e pouco menos do que 500 genes, possuem os menores genomas entre as bactérias parasitas (Figura 6.8, ver também Figura 6.14). Excluindo-se os endossimbiontes, o menor genoma procariótico é da arqueia *N. equitans*, cerca de 90 Kpb a menos do que o de *Mycoplasma genitalium* (Tabela 6.1). Entretanto, o genoma de *N. equitans* realmente contém mais ORFs que o genoma maior de *M. genitalium*. Isso ocorre porque o genoma de *N. equitans* é extremamente compacto e, praticamente, é desprovido de DNA não codificador. *N. equitans* é um hipertermófilo e parasita de um segundo hipertermófilo, a arqueia *Ignicoccus* (↻ Seção 16.7). Análises do conteúdo gênico de *N. equitans* revelam a ausência de praticamente todos os genes que codificam proteínas envolvidas no anabolismo e catabolismo.

Utilizando *Mycoplasma*, que apresenta cerca de 500 genes, como ponto de partida, vários pesquisadores estimaram que cerca de 250 a 300 genes representam o conteúdo genético mínimo para uma célula viável. Estas estimativas dependem em parte de comparações com outros genomas pequenos. Além disso, foi realizada a mutagenese sistemática para identificar os genes essenciais. Por exemplo, experimentos com *Escherichia coli* e com *Bacillus subtilis*, ambos com cerca de 4.000 genes, indicaram que aproximadamente 300 a 400 genes são essenciais, dependendo das condições de crescimento. Entretanto, nesses experimentos as bactérias receberam muitos nutrientes, permitindo que elas sobrevivessem sem muitos dos genes relacionados com as funções biossintéticas. A maioria dos “genes essenciais” identificados também está presente em outras bactérias e aproximadamente 70% foram encontrados em arqueias e nos eucariotos.

### Genomas grandes

Alguns procariotos apresentam genomas muito grandes, comparáveis àqueles de microrganismos eucarióticos. Como os eucariotos tendem a apresentar quantidades significativas de DNA não codificador, ao contrário dos procariotos, alguns genomas procarióticos apresentam, na realidade, mais genes que microrganismos eucarióticos, apesar de possuírem menos DNA. Por exemplo, *Bradyrhizobium japonicum*, que forma nódulos de raiz fixadores de nitrogênio em soja, apresenta 9,1 Mpb de DNA e 8.300 ORFs, enquanto a levedura *Saccharomyces cerevisiae*, um eucarioto, apresenta 12,1 Mpb de DNA e apenas 5.800 ORFs (ver Tabela 6.5). A bactéria de solo, *Myxococcus xanthus*, também possui 9,1 Mpb de DNA, enquanto seus organismos relacionados apresentam genomas de aproximadamente metade desse tamanho. Acredita-se que eventos de duplicação múltipla de segmentos substanciais de DNA possam ser responsáveis pelos genomas muito grandes, tais como o de *M. xanthus*.

O maior genoma procariótico conhecido até o momento pertence à *Sorangium cellulosum*, uma espécie de mixobactéria (↻ Seção 14.19). Com pouco mais de 13 Mpb em um único cromossomo circular, o seu genoma é cerca de três vezes maior do que da *Escherichia coli*. O genoma da *S. cellulosum* possui uma proporção grande de DNA não codificador para uma bactéria – 14,5% – e, conseqüentemente, possui menos sequências codificadoras do que se era esperado – somente 9.400. No entanto, ela possui mais DNA do que muitos eucariotos, incluindo leveduras e protozoários (*Cryptosporidium* e *Giardia*, respectivamente) (ver Tabela 6.5). A complexa regulação necessária para o estilo de vida de *Sorangium* é vista na grande quantidade de proteínas-quinase (proteínas que fosforilam outras proteínas para controlar a sua atividade). Essa bactéria possui 317 cinases, duas vezes mais do que qualquer outro organismo, incluindo eucariotos.

Contrariamente ao observado em bactérias, os maiores genomas encontrados em arqueias até o momento apresentam cerca de 5 Mpb (Tabela 6.1). De maneira geral, os genomas procarióticos variam de tamanhos correspondentes àqueles dos maiores vírus até os de microrganismos eucarióticos.

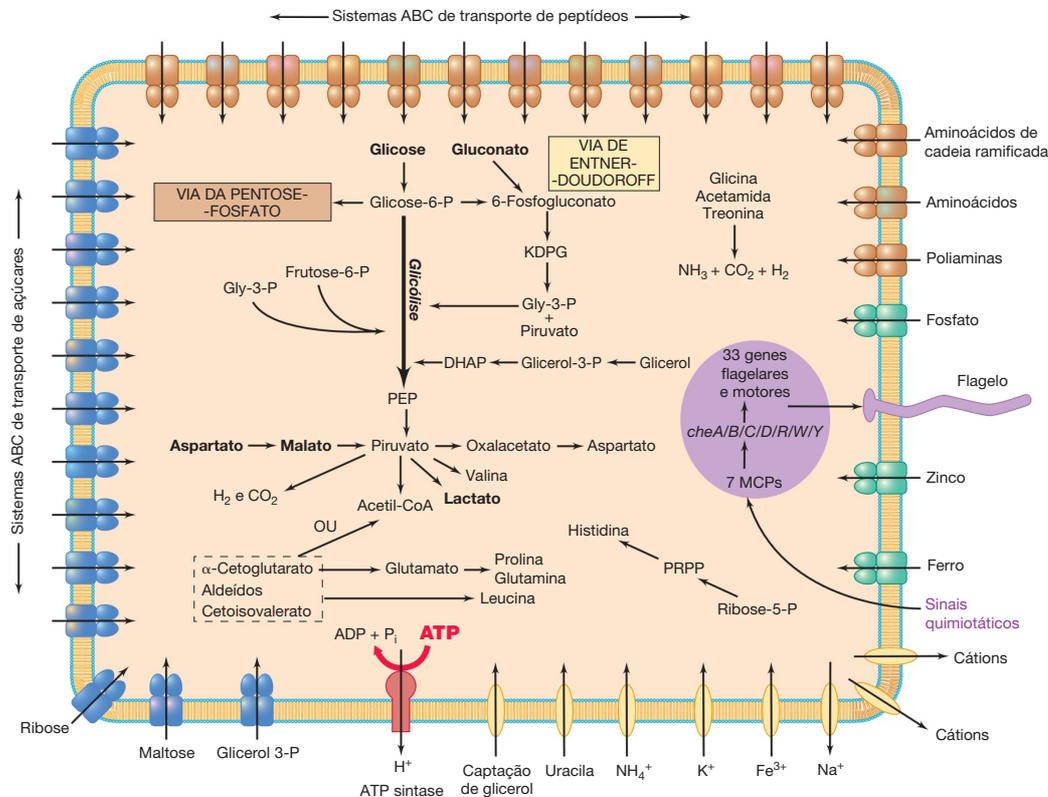
### Conteúdo gênico de genomas bacterianos

O conjunto de genes em um determinado organismo define sua biologia. Inversamente, genomas são moldados pelo estilo de vida de um organismo. Análises comparativas são úteis na busca de genes que codificam enzimas que muito provavelmente existam devido às propriedades conhecidas de um organismo. A *Thermotoga maritima* (Bactéria), por exemplo, é um hipertermófilo encontrado em sedimentos marinhos quentes, e estudos laboratoriais revelaram que esse organismo é capaz de catabolizar um grande número de açúcares. A Figura 6.9 resume algumas das vias metabólicas e dos sistemas de transporte de *T. maritima*, os quais foram deduzidos a partir da análise de seu genoma. Cerca de 7% dos seus genes codificam proteínas envolvidas no metabolismo de açúcares. Como esperado, seu genoma é rico em genes de transporte, particularmente de carboidratos e aminoácidos. Esses dados sugerem que *T. maritima* vive em um ambiente rico em material orgânico.

Poderíamos imaginar, por exemplo, que parasitas obrigatórios, como *Treponema pallidum* (agente causador da sífilis, ↻ Seções 14.20 e 29.12), poderiam necessitar de um número relativamente menor de genes para a biossíntese de aminoácidos, uma vez que os aminoácidos necessários podem ser fornecidos por seus hospedeiros. Esse realmente é o caso, pois o genoma de *T. pallidum* não apresenta genes reconhecíveis para a síntese de aminoácidos, embora sejam encontrados genes que codificam várias proteases, enzimas que convertem os peptídeos adquiridos do hospedeiro em aminoácidos livres. Por outro lado, a bactéria de vida livre *Escherichia coli* apresenta 131 genes para a biossíntese e metabolismo de aminoácidos, enquanto a bactéria de solo *Bacillus subtilis* apresenta mais de 200.

Uma análise funcional dos genes e de suas atividades em vários procariotos é apresentada na Tabela 6.4. Até o momento, um padrão distinto emergiu em relação à distribuição gênica em procariotos. Genes metabólicos são normalmente a classe mais abundante de genes em genomas procarióticos, embora genes envolvidos na síntese de proteínas superem os genes metabólicos em termos de porcentagem, à medida que o tamanho do genoma diminui (Tabela 6.4 e Figura 6.10). Embora muitos genes possam ser dispensados, os genes que codificam proteínas envolvidas nos aparatos de síntese são essenciais. Assim, quanto menor for o genoma maior será a porcentagem de genes que codificam proteínas relacionadas com o processo de tradução. Curiosamente, embora sejam vitais, os genes para a replicação e a transcrição de DNA são a fração menor de um genoma procariótico típico.

A porcentagem de genes de um organismo dedicada a uma ou outra função gênica consiste, em parte, em uma função do tamanho do genoma. Isso está resumido para um grande número de genomas bacterianos na Figura 6.10. Processos celulares centrais, como a síntese de proteínas, replicação de DNA e produção de energia, revelam apenas pequenas variações no número de genes em relação ao tamanho do genoma. Conseqüentemente, a porcentagem relativa de genes envolvidos na síntese de proteínas, por exemplo, aumenta tremendamente em organismos com genoma pequeno. Contrariamente, genes associados à regulação da transcrição aumentam significativamente em organismos com genoma maior. Esse sistema



**Figura 6.9** Visão geral do metabolismo e transporte em *Thermotoga maritima*. A figura apresenta um esquema das capacidades metabólicas desse organismo. Elas incluem algumas das vias envolvidas na produção de energia e no metabolismo de compostos orgânicos, incluindo proteínas de transporte que foram identificadas a partir da análise da sequência genômica. Os nomes dos genes não são mostrados. O genoma contém vários sistemas de transporte do tipo ABC, 12 para carboidratos, 14 para peptídeos e aminoácidos,

e ainda outros para íons. Esses são ilustrados como estruturas de múltiplas subunidades na figura. Outros tipos de proteínas de transporte foram também identificados e são ilustrados como elipses simples. Genes relacionados com o flagelo e com a quimiotaxia estão destacados em lilás, e poucos aspectos do metabolismo de açúcares também estão representados na figura. Esta figura foi adaptada da figura publicada pelo The Institute for Genomic Research (TIGR, Rockville, MD).

regulador extra permite que a célula se adapte com mais flexibilidade às diversas situações ambientais.

Organismos com genomas grandes têm condições de ter muitos genes que codifiquem proteínas relacionadas com os processos metabólicos. Isso provavelmente os torna mais competitivos em seus habitats, os quais, no caso de procariontes com genomas muito grandes, muitas vezes são o solo. O solo é um habitat onde as fontes de carbono e energia frequentemente são escassas, ou disponíveis apenas de forma intermitente e que variam significativamente (↔ Seção 19.1). Assim, um genoma grande que codifica múltiplas opções metabólicas poderia ser fortemente selecionado para tal habitat. Curiosamente, todos os procariontes listados na Tabela 6.1, cujos genomas superam 6 Mpb, habitam o solo.

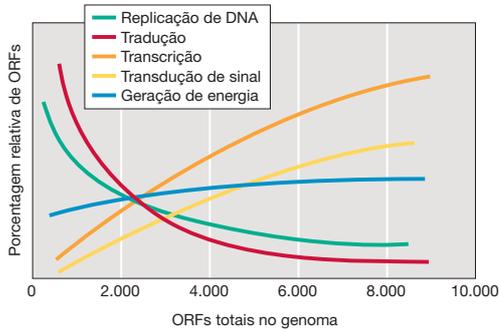
Análises de categorias gênicas foram realizadas em várias arqueias. Em média, espécies de arqueias dedicam uma porcentagem maior de seu genoma à produção de energia e coenzimas do que as de bactérias (esses resultados estão, sem dúvida, ligeiramente distorcidos devido ao grande número de novas coenzimas produzidas por arqueias metanogênicas [↔ Seção 13.20]). Por outro lado, arqueias parecem conter menor número de genes envolvidos no metabolismo de carboidratos ou em funções relacionadas com a membrana citoplasmática, como

transporte e biossíntese de membranas, do que bactérias. Contudo, o resultado dessa descoberta pode estar distorcido pelo fato de as vias correspondentes serem menos estudadas em arqueias do que em bactérias, e muitos dos genes correspondentes de arqueias serem, provavelmente, ainda desconhecidos.

Os dois domínios procarióticos possuem números relativamente grandes de genes cujas funções são desconhecidas, ou codificam somente proteínas hipotéticas, embora em ambas as categorias existam mais incertezas entre as arqueias do que em bactérias. Entretanto, isso pode ser um artefato devido à menor disponibilidade de sequências genômicas de arqueias que de bactérias.

#### MINIQUESTIONÁRIO

- Qual o estilo de vida dos organismos procarióticos que apresentam genomas menores do que aqueles de certos vírus?
- Aproximadamente, quantos genes codificadores de proteínas são encontrados em um genoma bacteriano de 4 Mpb?
- Que organismo provavelmente possui mais genes, um procarionte com 8 Mpb de DNA ou um eucariote com 10 Mpb de DNA? Explique.
- Qual categoria de genes procarióticos existe em maior percentual?



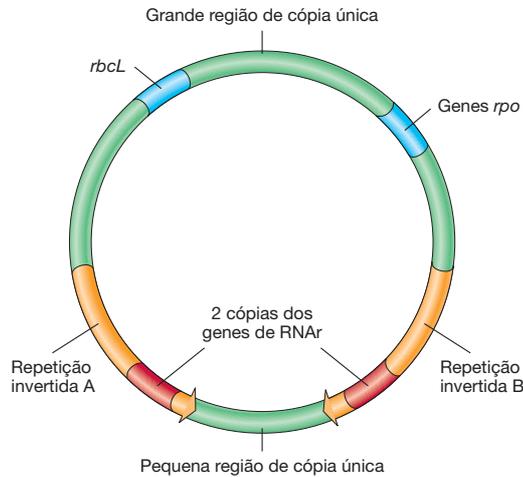
**Figura 6.10** Categoria funcional de genes como porcentagem do genoma. Observe como genes codificadores de produtos para a tradução ou replicação de DNA aumentam em porcentagem nos organismos de genomas pequenos, enquanto genes reguladores transcricionais aumentam em porcentagem nos organismos de genomas grandes. Dados de *Proc. Natl. Acad. Sci. (USA)* 101: 3.160-3.165 (2004).

### 6.5 Genomas de organelas

Mitocôndrias e cloroplastos são organelas derivadas de bactérias endossimbiontes que são encontradas dentro de células eucarióticas (↔ Seções 2.21 e 17.1). Ambos contêm pequenos genomas que possuem propriedades fundamentais semelhantes às dos genomas bacterianos. Além disso, ambos contêm a maquinaria necessária para a síntese proteica, incluindo ribossomos e RNAs de transferência, além de outros componentes necessários para a produção de proteínas funcionais. Mais uma vez, esses componentes são mais estreitamente relacionados com aqueles encontrados em bactérias do que aqueles encontrados no citoplasma eucariótico. Assim, as organelas compartilham muitas características fundamentais em comum com as células procarióticas, às quais são filogeneticamente relacionadas.

#### O genoma do cloroplasto

Células de plantas verdes contêm cloroplastos, organelas responsáveis pela fotossíntese (↔ Seção 13.1). Todos os genomas conhecidos de cloroplastos são moléculas circulares de DNA. Embora existam várias cópias do genoma em cada cloroplasto, elas são idênticas. O genoma típico de cloroplasto corresponde a cerca de 120 a 160 Kpb e contém duas repetições inver-



**Figura 6.11** Mapa de um típico genoma de cloroplasto. Cada repetição invertida contém uma cópia de um dos genes dos três RNAr (5S, 16S e 23S). A subunidade grande de RubisCO é codificada pelo gene *rbcL* e a RNA-polimerase do cloroplasto, pelo gene *rpo*.

tidas de 6 a 76 Kpb que codificam cópias de cada um dos três genes de RNAr (Figura 6.11). Vários genomas de cloroplastos foram completamente sequenciados, e todos são bastante semelhantes. O maior genoma de cloroplasto sequenciado até o momento é o da alga clorofíceia *Floydiella terrestris*. Ele tem um pouco mais de 500 Kpb e contém 97 genes conservados. Cerca de 80% desse genoma consistem nas regiões intergênicas com várias pequenas repetições.

Como já esperado, muitos genes de cloroplasto codificam proteínas para reações de fotossíntese e fixação do CO<sub>2</sub>. A enzima RubisCO catalisa o passo-chave do ciclo de Calvin na fixação do CO<sub>2</sub> (↔ Seção 13.5). O gene *rbcL* que codifica a subunidade grande da RubisCO está sempre presente no genoma do cloroplasto (Figura 6.11), entretanto o gene que codifica a subunidade pequena, *rbcS*, se localiza no núcleo da célula da planta e seu produto proteico deve ser importado do citoplasma para dentro do cloroplasto após sua síntese.

O genoma de cloroplasto também codifica RNAr utilizado nos ribossomos do cloroplasto, RNAr usado na tradução,

**Tabela 6.4** Função gênica em genomas bacterianos

Categorias funcionais	Porcentagem de genes		
	<i>Escherichia coli</i> (4,64 Mpb) <sup>a</sup>	<i>Haemophilus influenzae</i> (1,83 Mpb) <sup>a</sup>	<i>Mycoplasma genitalium</i> (0,58 Mpb) <sup>a</sup>
Metabolismo	21,0	19,0	14,6
Estrutural	5,5	4,7	3,6
Transporte	10,0	7,0	7,3
Regulação	8,5	6,6	6,0
Tradução	4,5	8,0	21,6
Transcrição	1,3	1,5	2,6
Replicação	2,7	4,9	6,8
Outras, conhecidas	8,5	5,2	5,8
Desconhecidas	38,1	43,0	32,0

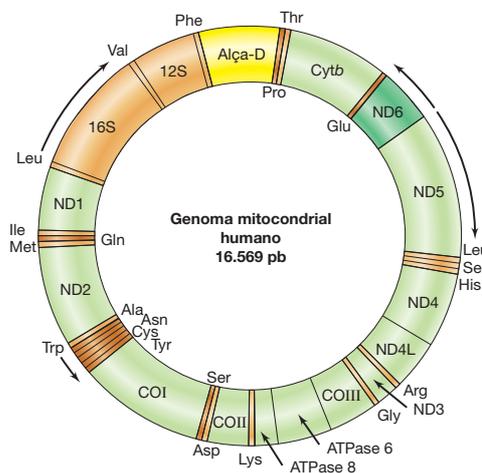
<sup>a</sup>Tamanho de cromossomos, em pares de megabases. Cada organismo listado contém somente um único cromossomo circular.

várias proteínas utilizadas na transcrição e tradução, assim como algumas outras proteínas. Algumas proteínas que atuam no cloroplasto são codificadas por genes nucleares. Acredita-se que eles sejam genes que migraram para o núcleo à medida que o cloroplasto evoluiu de um endossimbionte para uma organela fotossintética. Ao contrário dos procariotos de vida livre, os íntrons (↔ Seção 4.9) são comuns em genes de cloroplastos, sendo principalmente do tipo capaz de sofrer *autosplicing*.

**Genomas mitocondriais e proteomas**

As mitocôndrias são as organelas responsáveis pela produção de energia por meio da respiração e são encontradas na maioria dos organismos eucarióticos (↔ Seções 2.21 e 17.1). Os genomas mitocondriais codificam principalmente proteínas para a fosforilação oxidativa e, assim como os genomas de cloroplasto, também codificam RNAs, RNAs e proteínas envolvidas na síntese proteica. Entretanto, a maioria dos genomas mitocondriais codifica um número menor de proteínas do que os de cloroplastos.

Várias centenas de genomas mitocondriais foram sequenciados. O maior genoma mitocondrial possui 62 genes codificadores de proteínas, enquanto outros codificam apenas três proteínas. As mitocôndrias de quase todos os mamíferos, incluindo o homem, codificam somente 13 proteínas, 22 RNAs e 2 RNAs. A **Figura 6.12** apresenta o mapa do genoma mitocondrial humano, de 16.569 pb. O genoma mitocondrial da levedura *Saccharomyces cerevisiae* é maior (85.779 pb), porém possui apenas oito genes codificadores de proteínas. Além dos genes que codificam RNA e proteínas, o genoma mitocondrial de levedura contém grandes segmentos de DNA extremamente ricos em AT, que não possuem função aparente.

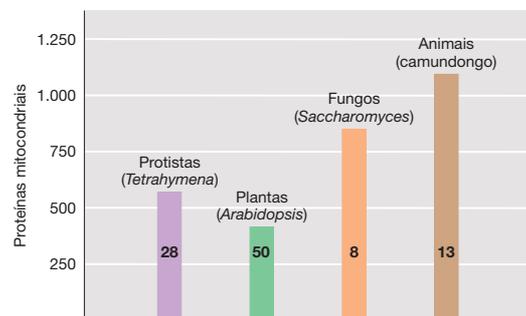


**Figura 6.12** Mapa do genoma mitocondrial humano. O genoma codifica os RNAs, 22 RNAs e várias proteínas. As setas mostram a direção da transcrição para os genes de uma determinada cor, e a designação de três letras para os aminoácidos relativos aos genes de RNA também são mostradas. Os 13 genes codificadores de proteínas estão ilustrados em verde. *Cytb*, citocromo *b*; ND1-6, componentes do complexo NADH desidrogenase; COI-III, subunidades do complexo citocromo oxidase; ATPase 6 e 8, polipeptídeos do complexo ATPase mitocondrial. Os dois promotores estão na região denominada alça-D, uma região também envolvida na replicação do DNA.

Os genomas mitocondriais das plantas são maiores do que os das mitocôndrias das células dos animais, e variam entre 300 Kpb a 2.000 Kpb. Apesar disso, eles somente possuem cerca de 50 genes altamente conservados, sendo que a maioria deles codifica componentes que fazem parte da cadeia respiratória e do aparato de tradução. A variação no tamanho do genoma é decorrente de grande quantidade de DNA não codificador. Os genomas mitocondriais de diferentes espécies de plantas com flores, pertencentes ao gênero *Silene*, variam incrivelmente de tamanho. Os dois maiores possuem aproximadamente 7 e 11 Mpb, tornando-os maiores do que a maioria dos genomas bacterianos.

Contrariamente aos genomas de cloroplastos, que são moléculas simples de DNA circular, os genomas mitocondriais são bastante diversos. Por exemplo, alguns genomas mitocondriais são lineares, incluindo os de algumas espécies de algas, protozoários e fungos. Em outros casos, como na levedura *S. cerevisiae* utilizada em panificação e na produção de cerveja, embora análises genéticas indiquem que o genoma mitocondrial seja circular, parece que a forma predominante *in vivo* é linear, presente em múltiplas cópias. (Lembre-se de que o bacteriófago T4 apresenta um genoma geneticamente circular, porém fisicamente linear, ↔ Seção 8.6.) Por fim, a mitocôndria de diversos fungos e plantas contém pequenos plasmídeos circulares ou lineares além do genoma mitocondrial principal.

As mitocôndrias requerem muito mais proteínas do que elas codificam. Particularmente, mais proteínas são necessárias para a tradução do que aquelas codificadas pelo genoma da organela. As proteínas necessárias para diversas funções da organela são codificadas pelos genes nucleares. A mitocôndria de levedura contém cerca de 800 proteínas diferentes (ver Proteoma, Seção 6.8). Entretanto, somente oito delas são codificadas pelo genoma mitocondrial, o restante das proteínas é codificado pelos genes nucleares (**Figura 6.13**). Os genes que codificam a maioria das proteínas das organelas estão presentes no núcleo, são transcritos lá e traduzidos nos ribossomos 80S no citoplasma das células eucarióticas. As proteínas são, então, transportadas para o interior das organelas. As proteínas codificadas pelo núcleo que são requeridas para a tradução e para a geração de energia na mitocôndria são mais relacionadas com as proteínas homólogas das bactérias do que com as do citoplasma eucariótico, coerente com a história evolutiva da mitocôndria.



**Figura 6.13** Proteomas mitocondriais. Número de proteínas localizadas na mitocôndria de diferentes grupos eucariotos. O número é uma estimativa porque algumas proteínas estão presentes em quantidades muito pequenas. O valor em cada barra colorida representa a quantidade de proteínas codificadas em cada genoma mitocondrial dos organismos.

### A variabilidade no código genético

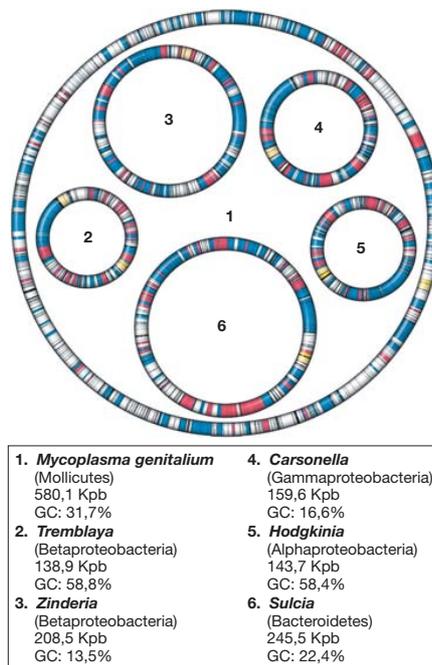
A crença original de que todas as células utilizam o mesmo código genético fez com que esse código fosse considerado universal (↔ Tabela 4.5). Entretanto, descobertas mais recentes mostraram que as mitocôndrias e poucas células utilizam pequenas variações do código genético “universal”. Códigos genéticos alternativos foram primeiramente descobertos nos genomas de mitocôndrias das células animais. Esses códons modificados geralmente utilizam códons de término como códons com sentido. Por exemplo, mitocôndria de animais (mas não de plantas) utiliza o códon UGA para codificar o triptofano em vez de utilizá-lo como códon de término. Mitocôndrias de leveduras também utilizam o UGA para triptofano, além disso, utilizam os quatro códons CUN (sendo N qualquer nucleotídeo) como treonina em vez de leucina. Essas mudanças parecem ter surgido a partir da pressão seletiva para genomas menores; por exemplo, por residirem em ambientes onde muitos nutrientes necessários já estiveram disponíveis. Assim, os 22 RNAs produzidos nas mitocôndrias são insuficientes para a leitura do código genético completo, mesmo considerando-se a oscilação no pareamento-padrão (↔ Figura 4.32). Portanto, o pareamento de bases entre o anticódon e o códon é ainda mais flexível em mitocôndrias do que em células.

Diversos organismos são também conhecidos por utilizar o código genético com ligeiras diferenças. Por exemplo, no gênero *Mycoplasma* (*Bacteria*) e no gênero *Paramecium* (*Eukarya*) certos códons de término codificam aminoácidos. Consequentemente, esses organismos possuem menos códons de término. Alguns fungos utilizam o códon da leucina (CUG) para codificar serina. Entretanto, esses códons se tornaram um pouco ambíguos, como é o caso do CUG que em 97% das vezes é traduzido como serina e apenas 3% das vezes como leucina.

### Simbiontes e organelas

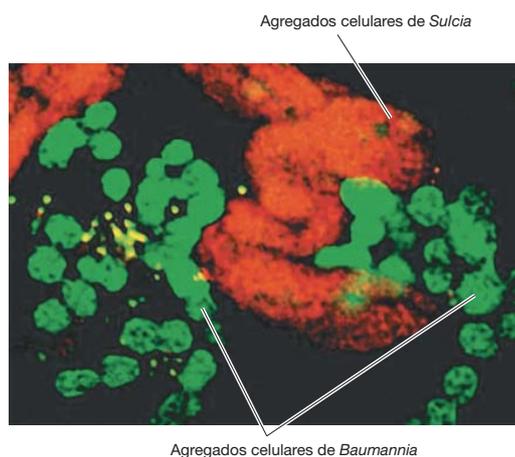
Muitos insetos e alguns vertebrados, incluindo certos nematódeos e moluscos, possuem bactérias simbióticas dentro das suas células. Algumas dessas bactérias simbiontes não são mais capazes de existirem independentemente e por isso apresentam grandes reduções no tamanho dos seus genomas (↔ Seção 22.9). Genomas de simbiontes possuem a mesma variação de tamanho dos genomas das bactérias de vida livre, até cerca de 140 Kpb para *Tremblaya* e *Hodgkinia* (Tabela 6.1 e Figura 6.8), os dois menores exemplos conhecidos (Figura 6.14). Portanto, o genoma de alguns simbiontes contém menos genes do que algumas organelas e alguns vírus. Esses simbiontes dependem totalmente das células dos insetos hospedeiros para a sua sobrevivência e nutrição. Por sua vez, os simbiontes oferecem ao inseto aminoácidos essenciais e outros nutrientes que ele não é capaz de sintetizar.

Alguns insetos possuem duas bactérias simbiontes. Por exemplo, algumas cigarrinhas possuem tanto a *Baumannia cicadellinicola*, que fornece vitaminas e cofatores, além da *Sulcia mulleri*, que fornece muitos dos aminoácidos necessários para o inseto (Figura 6.15). A maioria dos simbiontes são espécies pertencentes a um dos dois maiores grupos de bactérias gram-negativas, os filos Proteobacteria e Bacteroidetes. A maior parte dos genomas de tamanho extremamente reduzido também apresenta um alto conteúdo de AT, por volta de 80%, exceto, paradoxalmente, para os dois organismos com os menores genomas, *Tremblaya* e *Hodgkinia*, que possuem cerca de 40% de AT. Alguns desses genomas altamente reduzidos apa-



**Figura 6.14** Genomas dos organismos simbiontes. Cinco genomas de simbiontes foram desenhados em escala dentro do círculo que representa o genoma do *Mycoplasma*. Azul: genes relacionados com o processamento da informação genética; vermelho: genes relacionados com a biossíntese de vitaminas e de aminoácidos; amarelo: genes que codificam os RNAs; branco: outros genes; lacunas indicam DNA não codificante. Kpb, milhares de pares de bases.

rentemente perderam diversos genes considerados essenciais para a replicação, como o gene que codifica a proteína FtsZ chave para a divisão celular (↔ Seção 5.2). Assim, ainda não se sabe como esses simbiontes conseguem se replicar.



**Figura 6.15** Dois endossimbiontes, *Sulcia* e *Baumannia*, ambos habitam as mesmas células de inseto. As hibridações *in situ* por fluorescência foram feitas utilizando sondas que se ligam seletivamente ao RNAr de *Baumannia* (verde) e de *Sulcia* (vermelho).

Os simbiontes discutidos neste capítulo diferem das mitocôndrias e dos cloroplastos em vários aspectos. Simbiontes são restritos a poucos tecidos, mesmo em um organismo hospedeiro particular. Há pouca evidência para a transferência gênica dos organismos simbiontes para o núcleo da célula hospedeira, assim como proteínas produzidas no citoplasma do hospedeiro não entram nos organismos simbiontes para desenvolverem funções vitais. No entanto, alguns simbiontes são absolutamente necessários para a sobrevivência do hospedeiro e não podem sobreviver fora deles. Isso nos leva a um importante questionamento para o qual ainda não existe resposta: onde está a linha entre um simbionte e uma organela?

### MINIQUESTIONÁRIO

- O que é incomum entre os genes que codificam proteínas mitocondriais?
- O que os genomas dos cloroplastos geralmente codificam?
- O que é incomum entre os genomas de organismos simbiontes de insetos?

## 6.6 Genomas de microrganismos eucarióticos

Um grande número de eucariotos é conhecido e os genomas de vários eucariotos microbianos e superiores já foram sequenciados (Tabela 6.5), e os seus tamanhos variam amplamente. Alguns protozoários unicelulares, incluindo o ciliado de vida livre *Paramecium* (40.000 genes) e o patógeno *Trichomonas*

(60.000), apresentam significativamente mais genes do que os seres humanos (Tabela 6.5). Na verdade, a *Trichomonas* atualmente detém o recorde de maior número de genes entre os organismos. Isso é intrigante uma vez que a *Trichomonas* é um parasita humano, e esses organismos geralmente possuem genomas menores do que os organismos de vida livre porque os parasitas dependem dos seus hospedeiros para algumas ou mesmo várias funções (Seções 6.3 e 6.5).

### Genomas de parasitas microbianos

Apesar do estranho caso de *Trichomonas*, os microrganismos eucarióticos parasitas possuem genomas contendo 10 a 30 Mpb de DNA e entre 4.000 e 11.000 genes. Por exemplo, o genoma do *Trypanosoma brucei*, o agente da doença africana do sono, possui 11 cromossomos, 35 Mpb de DNA e quase 11.000 genes. O mais importante parasita eucariótico é o *Plasmodium*, que causa malária (↔ Seção 17.5). As quatro espécies de *Plasmodium* que infectam os seres humanos possuem genomas de 23 a 27 Mpb contendo 14 cromossomos com cerca de 5.500 genes. Cerca de metade desses genes possui íntrons e um terço codifica proteínas hipotéticas conservadas de função desconhecida. A ameba social de vida livre *Dictyostelium* possui cerca de 12.500 genes (no entanto, observe que *Dictyostelium* possui fases unicelulares e multicelulares em seu ciclo de vida, ↔ Seção 17.8). Como forma de comparação, observe que a ameba patogênica *Entamoeba histolytica*, o agente etiológico da disenteria amebiana, apresenta aproximadamente 10.000 genes.

**Tabela 6.5** Alguns genomas nucleares eucarióticos<sup>a</sup>

Organismo	Comentários	Modo de vida <sup>b</sup>	Tamanho do genoma (Mpb)	Número haploide de cromossomos	Genes codificadores de proteínas
Nucleomorfo de <i>Bigelowiella natans</i>	Núcleo endossimbiótico degenerado	E	0,37	3	331
<i>Encephalitozoon cuniculi</i>	Menor genoma eucariótico conhecido; patógeno humano	P	2,9	11	2.000
<i>Cryptosporidium parvum</i>	Protozoário parasita	P	9,1	8	3.800
<i>Plasmodium falciparum</i>	Malária maligna	P	23	14	5.300
<i>Saccharomyces cerevisiae</i>	Levedura, eucarioto-modelo	VL	12,1	16	5.800
<i>Ostreococcus tauri</i>	Alga verde marinha; menor eucarioto de vida livre	VL	12,6	20	8.200
<i>Aspergillus nidulans</i>	Fungo filamentosos	VL	30	8	9.500
<i>Giardia lamblia</i>	Protozoário flagelado; causa gastroenterite aguda	P	12	5	9.700
<i>Dictyostelium discoideum</i>	Ameba social	VL	34	6	12.500
<i>Drosophila melanogaster</i>	Mosca-da-fruta; organismo-modelo para estudos genéticos	VL	180	4	13.600
<i>Caenorhabditis elegans</i>	Nematoide; organismo-modelo do desenvolvimento animal	VL	97	6	19.100
<i>Mus musculus</i>	Camundongo; mamífero-modelo	VL	2.500	23	25.000
<i>Homo sapiens</i>	Homem	VL	2.850	23	25.000
<i>Arabidopsis thaliana</i>	Planta-modelo para estudos genéticos	VL	125	5	26.000
<i>Oryza sativa</i>	Arroz; principal planta agrícola do mundo	VL	390	12	38.000
<i>Paramecium tetraurelia</i>	Protozoário ciliado	VL	72	>50	40.000
<i>Populus trichocarpa</i>	Choupo negro; uma árvore	VL	500	19	45.000
<i>Trichomonas vaginalis</i>	Protozoário flagelado; patógeno humano	P	160	6	60.000

<sup>a</sup>Todos os dados correspondem a genomas haploides nucleares desses organismos em megapares de bases. Para os genomas maiores, tanto o tamanho quanto o número de genes listados são estimativas devido ao grande número de sequências repetitivas e/ou íntrons nos genomas.

<sup>b</sup>E, endossimbiote; P, parasita; VL, de vida livre.

O menor genoma celular eucariótico conhecido pertence à *Encephalitozoon cuniculi*, um patógeno intracelular de seres humanos e outros animais, causador de infecções pulmonares. O *E. cuniculi* não apresenta mitocôndrias e, embora seu genoma haploide possua 11 cromossomos, o tamanho do genoma é de somente 2,9 Mpb, com aproximadamente 2.000 genes (Tabela 6.5); ele é menor do que muitos genomas procarióticos (Tabela 6.1). Assim como observado em procariotos, o menor genoma eucariótico pertence a um endossimbionte. Conhecido como um *nucleomorfo*, ele corresponde aos restos degenerados de um endossimbionte eucariótico encontrado em certas algas verdes que adquiriram a capacidade fotossintética por endossimbioses secundárias (↔ Seção 17.1). Os genomas de nucleomorfos variam de cerca de 0,45 a 0,85 Mpb.

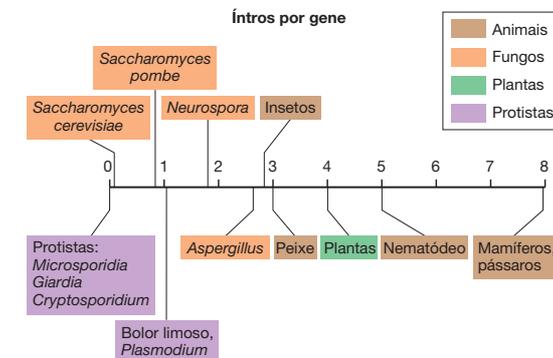
### O genoma da levedura

Entre os eucariotos unicelulares, a levedura *Saccharomyces cerevisiae* é o mais amplamente utilizado como organismo-modelo e também é extensivamente usado tanto na panificação quanto na produção de cerveja. O genoma haploide da levedura contém 16 cromossomos, que variam de 220 Kpb a cerca de 2.352 Kpb em tamanho. O genoma nuclear total (excluindo as mitocôndrias e alguns plasmídeos, além de elementos genéticos semelhantes a vírus) possui aproximadamente 13.400 Kpb. O cromossomo XII da levedura contém um segmento de aproximadamente 1.260 Kpb, contendo 100 a 200 repetições de genes de RNAr de levedura. Além das múltiplas cópias de genes de RNAr, o genoma nuclear da levedura tem aproximadamente 300 genes para RNAs (apenas alguns são idênticos) e cerca de 100 genes para os outros tipos de RNA não codificantes. A levedura tem aproximadamente 600 ORFs, menos do que algumas espécies de bactérias (Tabelas 6.1 e 6.5). Cerca de dois terços das ORFs das leveduras codificam proteínas que não possuem função conhecida.

Quanto dos genes conhecidos da levedura são realmente essenciais? Essa questão pode ser abordada pela inativação sistemática de cada gene por *mutações do tipo nocaute* (mutações que geram um gene não funcional, ↔ Seção 11.5). Mutações do tipo nocaute não podem ser obtidas normalmente em genes essenciais à viabilidade celular em um organismo haploide. Entretanto, leveduras podem ser cultivadas nos estados diploide e haploide (↔ Seção 17.13). A geração de mutações nocaute em células diploides, seguida da investigação de sua ocorrência em células haploides, permite determinar se um gene em particular é essencial à viabilidade celular. Utilizando-se mutações de nocaute, demonstrou-se que pelo menos 900 ORFs (17%) da levedura são essenciais. Observe que esse número de genes essenciais é muito maior do que os aproximadamente 300 genes (Seção 6.4) prognosticados como

## III • Genômica funcional

Apesar do grande esforço necessário à geração de uma sequência genômica anotada, de certa forma o resultado líquido é simplesmente uma “lista de partes”. Para entender o funcionamento celular, devemos conhecer mais do que apenas quais genes estão presentes. Também é necessário investigar a expressão gênica (transcrição) e a função do produto gênico final. Em analogia ao termo “genoma”, todo o conjunto completo de RNAs produzido sob determinadas condições é conheci-



**Figura 6.16** Frequência de íntrons em diferentes eucariotos. O número médio de íntrons por gene é mostrado para uma variedade de organismos eucarióticos.

o número mínimo necessário para procariotos. Entretanto, como os eucariotos são mais complexos que os procariotos, um maior complemento gênico mínimo pode ser esperado.

Por ser um eucarioto, o genoma da levedura contém íntrons (↔ Seção 4.9). Contudo, o número total de íntrons nos genes codificadores de proteínas consiste em apenas 225. A maioria dos genes da levedura com íntrons apresenta um único íntron pequeno próximo à extremidade 5' do gene. Essa situação difere muito daquela observada em eucariotos superiores (Figura 6.16). No verme *Caenorhabditis elegans*, por exemplo, um gene apresenta em média cinco íntrons e, na mosca-da-fruta *Drosophila*, quatro íntrons. Íntrons são também muito comuns nos genes de plantas, apresentando uma média de quatro por gene. Assim, o modelo de plantas superiores, *Arabidopsis*, apresenta em média cinco íntrons por gene, e mais de 75% dos genes de *Arabidopsis* possuem íntrons. Em seres humanos, quase todos os genes codificadores de proteínas apresentam íntrons, não sendo rara a ocorrência de 10 ou mais em um único gene. Além disso, os íntrons humanos são normalmente muito maiores do que os éxons humanos (DNA que realmente codifica proteína). De fato, os éxons constituem somente cerca de 1% do genoma humano, enquanto os íntrons somam 24%.

### MINIQUESTIONÁRIO

- Qual é a variação de tamanho dos genomas eucarióticos?
- Como isso se compara com o dos procariotes?
- Como é possível demonstrar que um gene é essencial?
- O que é incomum em relação ao genoma do eucarioto *Encephalitozoon*?

do como **transcriptoma**. Terminologia semelhante é aplicada aos produtos da tradução, metabolismo e outras áreas afins, adicionando o sufixo “oma”. A Tabela 6.6 resume a terminologia “oma” utilizada neste capítulo.

### 6.7 Microarranjos e o transcriptoma

O transcriptoma refere-se ao estudo global dos transcritos e é feito pelo monitoramento do RNA total gerado sob condições

**Tabela 6.6** Terminologia ooma

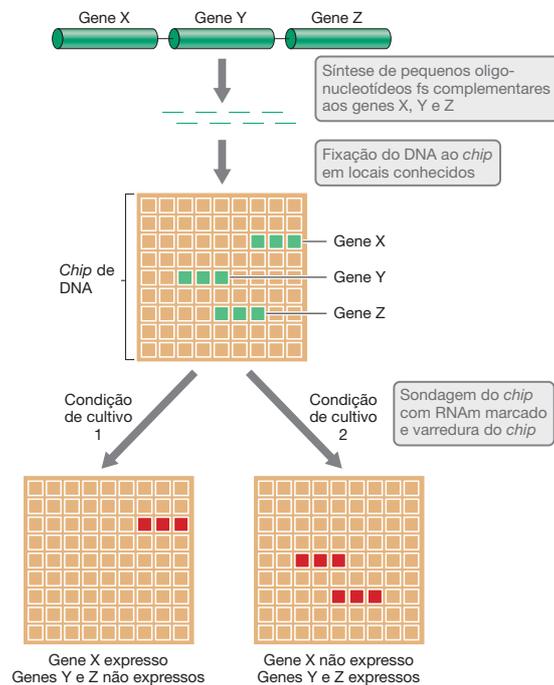
<b>DNA</b>	<p><b>Genoma</b> o conjunto total das informações genéticas de uma célula ou um vírus</p> <p><b>Metagenoma</b> o genoma total de todas as células presentes em um ambiente particular</p> <p><b>Epigenoma</b> número total de possíveis mudanças epigenéticas</p> <p><b>Metiloma</b> número total de sítios metilados no DNA (epigenéticos ou não)</p>
<b>RNA</b>	<p><b>Transcriptoma</b> o total de RNA produzido em um organismo sob determinadas condições específicas</p>
<b>Proteína</b>	<p><b>Proteoma</b> o conjunto de proteínas totais codificadas por um genoma</p> <p><b>Translatoma</b> o conjunto total de proteínas presentes sob condições específicas</p> <p><b>Interatoma</b> o conjunto total de interações entre as proteínas (ou outras macromoléculas)</p>
<b>Metabólitos</b>	<p><b>Metaboloma</b> o conjunto total de pequenas moléculas e metabólitos intermediários</p> <p><b>Glicoma</b> o conjunto total de açúcares e outros carboidratos</p>
<b>Organismos</b>	<p><b>Microbioma</b> o conjunto total de microrganismos em um ambiente (incluindo aqueles associados a organismos superiores)</p> <p><b>Viroma</b> o conjunto total de vírus em um ambiente</p> <p><b>Micobioma</b> o conjunto de fungos em um ambiente natural</p>

de crescimento escolhidas. No caso de genes que ainda não possuem um papel conhecido, a descoberta das condições sob as quais eles são transcritos pode dar pistas sobre as suas funções. Duas abordagens principais são utilizadas: microarranjos, os quais dependem da hibridização do RNA com o DNA, e RNAseq, que depende do sequenciamento de segunda geração (ou gerações mais avançadas).

### Microarranjos e *chip* de DNA

**Microarranjos** são pequenos suportes sólidos aos quais genes, ou, mais frequentemente, partes de genes, são fixados e arranjados espacialmente em um padrão conhecido; eles são frequentemente chamados de *chip* de DNA (Figura 6.17). A tecnologia dos microarranjos requer a hibridização entre RNA-DNA. Quando o DNA sofre desnaturação (i.e., as duas fitas são separadas), a fita simples pode formar moléculas de duas fitas híbridizadas com outra fita de DNA ou RNA pela complementaridade ou quase complementaridade dos pares de bases (↔ Seção 11.2). Esse processo é chamado de *hibridização de ácidos nucleicos*, ou **hibridização**, para abreviar, e é amplamente utilizado na detecção, caracterização e identificação de segmentos de DNA ou RNA. Segmentos de ácidos nucleicos de fita simples cujas identidades são conhecidas e que são utilizados na hibridização são denominados **sondas de ácidos nucleicos** ou, simplesmente, sondas. Para permitir a detecção, as sondas são radioativas ou são marcadas com corantes fluorescentes. Variando as condições, é possível ajustar a “estringência” da hibridização de modo que a complementaridade dos pares de base seja exata, ou próxima disso; ajudando a evitar pareamento inespecífico entre seqüências que são somente parcialmente complementares.

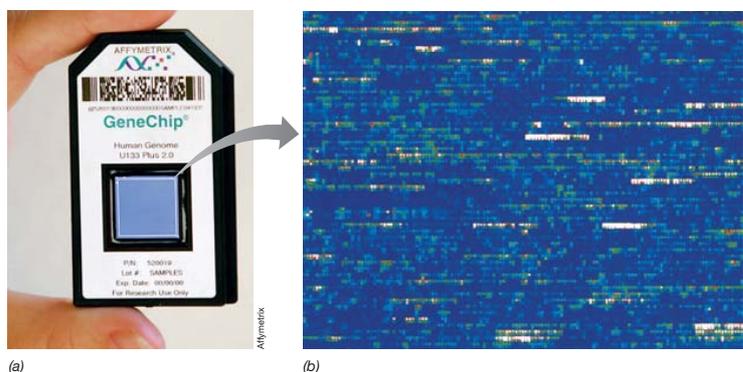
Nos microarranjos, os segmentos gênicos podem ser sintetizados pela reação de polimerização em cadeia (PCR, ↔ Seção 11.3) ou, alternativamente, oligonucleotídeos são conce-



**Figura 6.17** Confeção e utilização de microarranjos. Pequenos oligonucleotídeos de fita simples correspondendo a todos os genes de um organismo são individualmente sintetizados e fixados em locais conhecidos para a produção de um *chip* de DNA (microarranjo). O *chip* de DNA é analisado por hibridização com RNAm marcado por fluorescência e obtido de células cultivadas em uma condição específica, com as sondas de DNA presentes no *chip*, sendo o *chip* varrido por um raio *laser*.

bidos para cada gene, baseados na seqüência genômica. Uma vez ligados ao suporte sólido, esses segmentos de DNA podem ser hibridizados com RNA de células cultivadas em condições específicas, submetidos à varredura e analisados por computador. A hibridização entre um RNA específico e o DNA correspondente, no *chip*, indica que o gene foi transcrito (Figura 6.17; ver também Figura 6.18b). Quando estudamos genes codificadores de proteínas, o RNA mensageiro deve ser mensurado. Na prática, o RNAm está presente em quantidades muito pequenas para ser usado diretamente. Consequentemente, as seqüências de RNAm devem ser amplificadas primeiro. Isto é feito através de uma versão modificada da técnica de PCR capaz de gerar uma fita *complementar de DNA* a partir do RNAm (DNAc, ↔ Seção 11.3).

A fotolitografia, um processo utilizado para produzir *chips* de computador, é também utilizada na produção de *chips* para microarranjos. Os chips possuem normalmente entre 1 a 2 cm e são inseridos em um suporte de plástico que pode ser facilmente manipulado (Figura 6.18a); cada *chip* é capaz de receber milhares de diferentes fragmentos de DNA. Na prática, cada gene é frequentemente representado mais de uma vez no arranjo, a fim de lhe conferir maior confiabilidade. Arranjos genômicos completos contêm segmentos de DNA representando o genoma inteiro de um organismo. Por exemplo, uma empresa comercializa um *chip* de genoma humano que contém o genoma humano completo (Figura 6.18a). Esse *chip* único permite a análise de mais de 47.000 transcritos humanos e apresenta espaço para 6.500 oligonucleotídeos adicionais para uso em diagnósticos clínicos.



**Figura 6.18** Utilização de *chips* de DNA para avaliar a expressão gênica. (a) O *chip* do genoma humano contendo mais de 40.000 fragmentos gênicos. O aumento da parte a para a parte b indica a localização real do microarranjo. (b) *Chip* de levedura hibridizado. A foto revela fragmentos de um quarto do genoma total da levedura *Saccharomyces cerevisiae*, fixado a um *chip* gênico. Cada gene encontra-se em várias cópias e foi hibridizado com DNAC marcado com fluorescência, derivado do RNAm extraído de células de leveduras cultivadas em uma condição específica. O fundo do *chip* é azul. Os locais onde o DNAC foi hibridizado estão indicados por uma graduação de cores até a hibridização máxima, que aparece em branco. Como a localização de cada gene no *chip* é conhecida, quando o *chip* é submetido à varredura, revela quais genes foram expressos.

### Aplicações dos *chips* de DNA: expressão gênica

Os *chips* gênicos podem ser utilizados de diferentes maneiras, dependendo dos genes ligados a eles. A expressão gênica global é monitorada pela fixação, no suporte, de um oligonucleotídeo complementar a cada gene do genoma, utilizando-se, então, a população inteira de RNAs como amostra teste. A Figura 6.18b apresenta uma parte de um *chip* utilizado para avaliar a expressão do genoma de *Saccharomyces cerevisiae*. Esse *chip* facilmente comporta os 6.000 genes codificadores de proteínas de *S. cerevisiae* (Tabela 6.5) de forma que a expressão gênica global nesse organismo pode ser quantificada em um único experimento. Para isso, o *chip* é hibridizado com RNAC ou DNAC derivados de RNAm obtido de células de levedura cultivadas em condições específicas. Para a visualização da ligação, os ácidos nucleicos são marcados com um corante fluorescente e o *chip* é submetido à varredura com um detector a laser de fluorescência. Um padrão distinto de hibridização é observado, dependendo de quais sequências de DNA correspondem a quais RNAs (Figuras 6.18b). A intensidade da fluorescência fornece uma medida quantitativa da expressão gênica. Isso permite que o computador produza uma lista de quais genes são expressos e em que grau. Assim, utilizando *chips* gênicos, o transcriptoma do organismo de interesse, cultivado sob condições específicas, é revelado a partir do padrão e da intensidade dos pontos fluorescentes gerados (Tabela 6.6).

O *chip* gênico de *S. cerevisiae* foi utilizado para o estudo do controle metabólico nesse importante organismo industrial. A levedura pode crescer realizando fermentação ou respiração aeróbia. A análise do transcriptoma pode revelar quais genes são inativados e quais são ativados quando as células de levedura alteram do metabolismo fermentativo (anaeróbio) para o metabolismo respiratório, ou vice-versa. Análises dessa expressão gênica por transcriptomas revelam que a levedura sofre uma importante “reprogramação” metabólica durante a alternância do crescimento anaeróbio para o aeróbio. Alguns genes que controlam a produção de etanol (um produto essencial da fermentação) são fortemente reprimidos, enquanto as funções do ciclo do ácido cítrico (necessário ao crescimento aeróbio) são fortemente ativadas pela alternância. De maneira geral, mais de 700 genes são ativados e mais de 1.000 são inativados durante essa transição metabólica. Além disso, utilizando um microarranjo, o padrão de expressão dos genes com funções desconhecidas é também monitorado durante a alternância da fermentação para a respiração, fornecendo indícios para seu possível papel.

### Aplicações na identificação

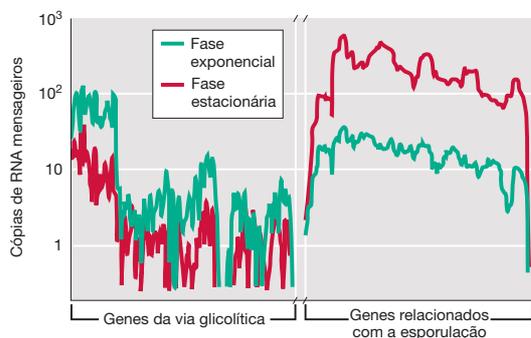
Além de sondar a expressão gênica, os microarranjos podem ser utilizados para identificar microrganismos especificamente. Nesse caso, o arranjo contém um conjunto de sequências de DNA características para cada um dos vários organismos ou vírus. Esse tipo de abordagem pode ser utilizado para diferenciar linhagens intimamente relacionadas a partir das diferenças em seus padrões de hibridização. Isso permite a identificação bastante rápida de vírus ou bactérias patogênicos a partir de amostras clínicas, ou a detecção desses organismos em várias outras substâncias, como alimentos. Por exemplo, *chips* de identificação (ID) foram utilizados na indústria alimentícia para a detecção de patógenos específicos, por exemplo, *E. coli* O157:H7. Outra aplicação importante dos microarranjos de DNA consiste na comparação de genes de organismos intimamente relacionados. Por exemplo, essa abordagem foi utilizada para acompanhar a evolução de bactérias patogênicas, a partir de organismos relacionados, não patogênicos.

Na microbiologia ambiental, os microarranjos foram utilizados para avaliar a diversidade microbiana. Os *filochips*, como são denominados, contêm oligonucleotídeos complementares às sequências do RNAr 16S de diferentes espécies bacterianas, uma molécula amplamente utilizada na sistemática de procariontes (Capítulo 12). Após a extração de DNA ou RNA total a partir de um ambiente, a presença ou ausência de cada espécie pode ser avaliada pela presença ou ausência de hibridização no *chip* (↔ Seção 18.6).

Existem também *chips* de DNA para a identificação de organismos superiores. Um *chip* disponível comercialmente, denominado *FoodExpert-ID*, contém 88.000 fragmentos gênicos de animais vertebrados, sendo utilizado na indústria alimentícia para garantir a pureza do alimento. Por exemplo, o *chip* pode confirmar a presença de carne, conforme citado na embalagem do alimento, e pode também detectar carnes de outros animais que podem ter sido adicionadas como suplementos ou substitutivos dos ingredientes oficiais. O *FoodExpert-ID* pode também ser utilizado para a detecção de subprodutos de vertebrados em rações animais, uma preocupação crescente com o advento das doenças mediadas por príons, como a doença da vaca louca (↔ Seção 9.13).

### Análises de RNA-Seq

Análise de *RNA-Seq* é um método pelo qual todas as moléculas de RNA da célula são sequenciadas. Desde que a sequência do genoma esteja disponível para comparação, esta irá revelar



**Figura 6.19 Análises de RNA-Seq.** Transcriptoma da cultura de *Clostridium* cultivada por 4,5 horas (células na fase exponencial) ou por 14 horas (células na fase estacionária). (1) Segmento de ~5,4-Kb em torno do operon glicolítico *gap-pgk-tpi*, e (2) segmento de ~1,2-Kb em torno do operon de esporulação *cotJC-cotJB*. A produção de endósporos é desencadeada pela escassez de nutrientes (↻ Seção 2.16). Dados de Wang, Y., X. Li, Y. Mao, e H.P. Blaschek. 2011. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. BMC Genomics 12: 479–489.

não somente os genes que foram transcritos, mas quantas cópias de cada RNA foram feitas. RNA-Seq é utilizado tanto para medir a expressão de RNAm quanto para identificar e caracterizar pequenos RNAs não codificadores. RNA-Seq requer alta produtividade do sequenciamento (sequenciamentos de segunda ou terceira geração, Seção 6.2) e é complicado pelo fato de que o RNA mais abundante na célula é o ribossomal (RNAr). Entretanto, métodos estão disponíveis para remover os RNAr ou aumentar os RNAm do total de RNA celular extraído. Além disso, recentes melhorias na tecnologia de sequenciamento permitem sequenciar sem remover o RNAr.

A técnica RNA-Seq está começando a ultrapassar as análises de microarranjos como o método de escolha para estudos globais da expressão gênica. Por exemplo, a **Figura 6.19** mostra uma comparação de RNA-Seq de culturas de uma espécie de *Clostridium* em fases exponencial e estacionária. Clostrídios são bactérias gram-negativas em forma de bastonetes que podem produzir endósporos, o estágio altamente resistente e dormente do ciclo de vida (↻ Seção 2.16). Como se poderia prever, a transcrição de genes da via glicolítica (a principal fonte pela qual o organismo gera ATP) é elevada durante a fase de crescimento exponencial, entretanto a expressão de genes relacionados com a esporulação aumenta na fase estacionária, quando os nutrientes se tornam limitados. O RNA-Seq também está sendo usado para analisar a comunidade microbiana e pode oferecer informações em nível de transcrição quando a sequência do genoma não está disponível para comparação. Neste caso, as sequências detectadas devem ser identificadas pela homologia com as sequências presentes nos bancos de dados.

Como veremos na Seção 6.10, *metagenômica* é a análise genômica de todo o DNA e RNA de um ambiente. A análise de metagenômica utilizando RNA-Seq tem sido explorada pelos laboratórios de cultivo bacteriano em amostras naturais que anteriormente não foram capazes de serem cultivadas em laboratório. Isso foi feito utilizando RNA-Seq para revelar quais genes foram transcritos em altos níveis por uma determinada comunidade microbiana. A análise de sequências, em seguida, identificou as proteínas correspondentes aos RNAs prevalentes. Isso permitiu aos pesquisadores deduzir quais nutrien-

tes as bactérias da amostra estavam utilizando dada a provável atividade enzimática dessas proteínas. Os meios de cultura foram, então, elaborados utilizando essas informações como guia e bactérias que antes eram consideradas não cultiváveis foram cultivadas com sucesso.

#### MINIQUESTIONÁRIO

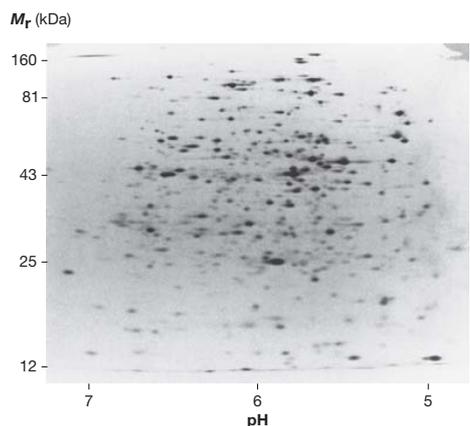
- Por que é útil conhecer como a expressão do genoma inteiro responde a uma determinada condição?
- O que os microarranjos nos revelam sobre o estudo da expressão gênica, que a análise de uma determinada enzima não é capaz?
- De quais avanços tecnológicos o RNA-Seq depende?

## 6.8 Proteômica e o iteratoma

O estudo em escala genômica da estrutura, função e regulação das *proteínas* de um organismo é denominado **proteômica**. O número e os tipos de proteínas presentes em uma célula estão sujeitos a alterações em resposta ao ambiente do organismo ou outros fatores, como os ciclos de desenvolvimento. Como resultado, o termo **proteoma** infelizmente tornou-se ambíguo. Em seu sentido mais amplo, um proteoma refere-se a *todas* as proteínas codificadas pelo genoma de um organismo. Em seu sentido mais restrito, contudo, refere-se às proteínas presentes em uma célula *em um determinado momento*. O termo *translatoma* é utilizado algumas vezes para essa última situação, isto é, para se referir a todas as proteínas expressas sob condições específicas.

### Métodos em proteômica

A primeira abordagem importante da proteômica ocorreu há décadas, com o advento da eletroforese bidimensional (2D) em gel de poliácridamida. Essa técnica permite a separação, identificação e quantificação de todas as proteínas presentes em uma amostra celular. Uma separação em gel 2D de proteínas de *Escherichia coli* é ilustrada na **Figura 6.20**. Na primeira dimensão (a dimensão horizontal na Figura 6.20), as proteínas



**Figura 6.20 Eletroforese das proteínas de *Escherichia coli* em gel bidimensional de poliácridamida.** Autorradiograma das proteínas celulares de *Escherichia coli*. Cada ponto no gel corresponde a uma proteína diferente, radioativamente marcada, permitindo sua visualização e a quantificação. As proteínas desnaturadas foram separadas na direção horizontal pela focalização isoeétrica e na direção vertical pela massa ( $M_r$ ; em quilodáltons), com as proteínas maiores localizando-se na porção superior do gel.

são separadas pelas diferenças em seus pontos isoelétricos, o pH no qual a carga líquida de cada proteína equivale a zero. Na segunda dimensão, as proteínas são desnaturadas de modo a conferir uma carga fixa a cada resíduo de aminoácido. As proteínas são então separadas por tamanho (de forma similar àquela realizada para moléculas de DNA; ↻ Seção 11.1).

Em estudos relacionados à *E. coli* e alguns outros organismos, centenas de proteínas separadas em géis 2D foram identificadas por meio bioquímico ou genético, sendo sua regulação estudada sob diversas condições. Utilizando géis 2D, a presença de uma determinada proteína sob diferentes condições de crescimento pode ser quantificada e relacionada com sinais ambientais. Um método para relacionar uma proteína desconhecida com um determinado gene utilizando o sistema de gel 2D consiste na eluição da proteína do gel e no sequenciamento de uma parte dela, geralmente a partir de sua extremidade aminoterminal. Mais recentemente, proteínas eluídas foram identificadas por uma técnica denominada *espectrometria de massa* (Seção 6.9), habitualmente após a digestão preliminar para gerar um conjunto característico de peptídeos. Essa informação de sequência pode ser suficiente para identificar completamente a proteína. Alternativamente, dados parciais de sequência podem permitir a confecção de sondas ou iniciadores de oligonucleotídeos a fim de localizar o gene que codifica a proteína, a partir do DNA genômico, por hibridização ou reação de polimerização em cadeia. Em seguida, após o sequenciamento do DNA, a identidade do gene pode ser determinada.

Atualmente, a cromatografia líquida vem sendo cada vez mais utilizada na separação de misturas proteicas. Na cromatografia líquida de alta pressão (HPLC, *high-pressure liquid chromatography*), a amostra é dissolvida em um líquido adequado e aplicada sob pressão através de uma coluna empacotada com um material de fase estacionária que separa proteínas de acordo com as diferenças em suas propriedades químicas, como tamanho, carga iônica ou hidrofobicidade. À medida que a mistura atravessa a coluna, ela é separada pela interação das proteínas com a fase estacionária. Frações são coletadas na saída da coluna. As proteínas de cada fração são digeridas por proteases e os peptídeos são identificados por espectrometria de massa.

### Genômica e proteômica comparativas

Embora a proteômica frequentemente requeira experimentação intensa, técnicas *in silico* podem também ser bastante úteis. Uma vez que a sequência de um genoma de um organismo é obtida, ela pode ser comparada com aquela de outros organismos, visando localizar e identificar genes similares àqueles já conhecidos. Nesse procedimento, a sequência mais importante corresponde à *sequência de aminoácidos* das proteínas codificadas. Como o código genético é degenerado (↻ Seção 4.11), diferenças na sequência de DNA não levam necessariamente a diferenças na sequência de aminoácidos.

Proteínas apresentando identidade de sequência superior a 50% frequentemente possuem funções similares. Proteínas com identidades acima de 70% quase certamente desempenham as mesmas funções. Muitas proteínas consistem em módulos estruturais distintos, denominados *domínios proteicos*, cada um com funções características. Tais regiões incluem domínios de ligação a metais, domínios de ligação a nucleotídeos ou domínios para determinadas classes de atividade enzimática, como helicase ou desidrogenase. A identificação de domínios de função conhecida em uma proteína pode revelar muito

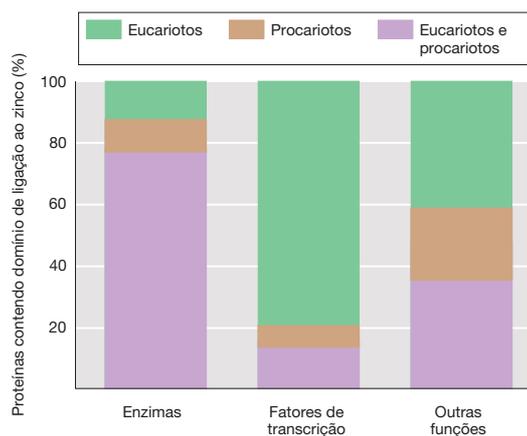
sobre seu papel, mesmo na ausência de homologia completa da sequência. Por exemplo, muitas proteínas possuem o metal zinco como cofator. Eles são encontrados nos sítios ativos das enzimas ou nos domínios de ligação ao DNA. A **Figura 6.21** mostra a distribuição das proteínas que possuem o zinco entre os procariotos e os eucariotos. Mesmo que ambos os grupos sintetizem muitas enzimas que possuem o zinco, a utilização de fatores de transcrição contendo esse metal é predominantemente uma característica dos eucariotos.

A *proteômica estrutural* refere-se à determinação, em escala proteômica, das estruturas tridimensionais (3D) de proteínas. Atualmente, não é possível prever diretamente a estrutura 3D de proteínas a partir de suas seqüências de aminoácidos. Entretanto, estruturas e proteínas desconhecidas podem frequentemente ser modeladas caso a estrutura 3D de uma proteína com identidade de 30% ou maior na seqüência de aminoácidos encontre-se disponível.

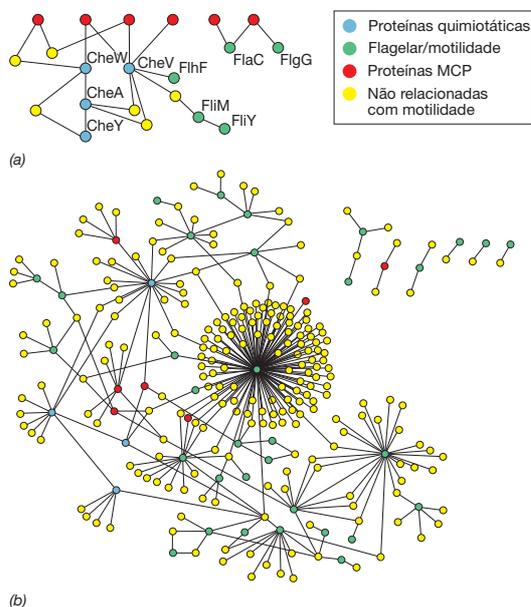
O acoplamento da proteômica e da genômica vem fornecendo importantes indícios de como a expressão gênica em diferentes organismos está relacionada a estímulos ambientais. Além de tais informações trazerem benefícios importantes para a ciência básica, elas também possuem aplicações potenciais na medicina, na análise do meio ambiente, e também na agricultura. Em todas essas áreas, o entendimento da ligação entre o genoma e o proteoma, e como ela é regulada, podem fornecer aos seres humanos o controle sem precedentes na luta contra as doenças e a poluição, assim como benefícios sem precedentes à produtividade agrícola.

### O interatoma

Por analogia aos termos “genoma” e “proteoma”, o **interatoma** é o conjunto total das *interações* entre as macromoléculas que constituem a célula (**Figura 6.22**). Originalmente, o termo interatoma é aplicado para as interações entre proteínas, muitas das quais se organizam em complexos. No entanto, também é possível considerar interações entre diferentes classes de moléculas, como o interatoma entre proteínas e RNA.



**Figura 6.21** **Proteômica comparativa.** As seqüências de 40 bactérias, 12 arqueias e 5 eucariotos contendo domínios de ligação ao zinco foram comparadas pela categoria funcional. As proteínas que contêm o zinco compreendem 5 a 6% do total das proteínas dos procariotos e 8 a 9% dos eucariotos sendo que muitas são enzimas. Os eucariotos também possuem muitos fatores de transcrição contendo o zinco.



**Figura 6.22** Interatoma de proteínas motoras da *Campylobacter jejuni*. Esse diagrama de rede ilustra como os dados do interatoma são representados. (a) Uma subseção da rede destacando as proteínas bem conhecidas na via de transdução de sinal de quimiotaxia (CheW, CheA e CheY) e suas parceiras. MCP, proteínas quimiotáticas aceptoras de metil (↔ Seção 7.8). (b) Interações de alta confiança entre todas as proteínas que possuem funções conhecidas na mobilidade da bactéria. Observe as seis pequenas redes que não fazem parte da rede principal.

Os dados dos interatomos são geralmente expressos em forma de diagramas de rede, em que cada junção representa uma proteína e as linhas de conexão representam as interações. Diagramas de interatomos inteiros podem ser muito complexos e, portanto, diagramas mais específicos, tais como a rede de proteínas motoras da bactéria *Campylobacter jejuni* (Figura 6.22), são mais instrutivos. Esta figura mostra o cerne de interações entre os componentes do sistema de quimiotaxia (↔ Seções 2.19 e 7.8), incluindo todas as outras proteínas que interagem com ele.

#### MINIQUESTIONÁRIO

- Por que o termo “proteoma” é ambíguo, enquanto o termo genoma não o é?
- Quais os métodos experimentais mais comuns utilizados na pesquisa do proteoma?
- O que é um interatoma?

## 6.9 Metabolômica e biologia de sistemas

O **metaboloma** é conjunto completo de *intermediários metabólicos* e outras pequenas moléculas produzidas por um organismo. A metabolômica desenvolveu-se de forma mais lenta que as demais “ômicas”, devido principalmente à imensa diversidade química de pequenos metabólitos. Isso torna a varredura sistemática um desafio em termos técnicos. As primeiras tentativas utilizaram análise de extratos de células marcadas com glicose  $^{13}\text{C}$  por ressonância nuclear magnética (RNM). Todavia, esse método apresenta sensibilidade limitada e o número de compostos que podem ser identificados simultanea-

mente em uma mistura é muito pequeno para a resolução de extratos celulares totais.

### Novas técnicas de espectrometria de massa: MALDI-TOF

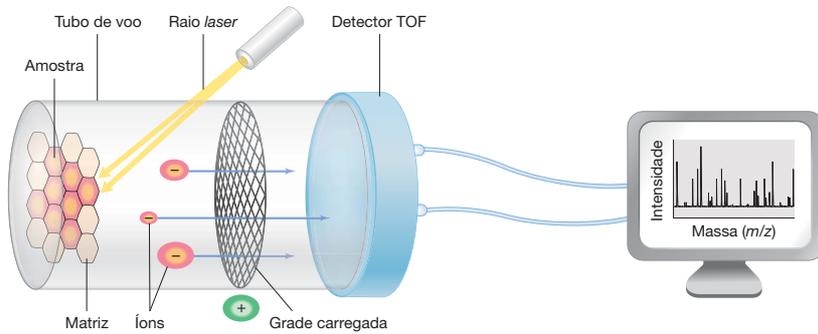
A abordagem mais promissora para a metabolômica é o uso de técnicas recém-desenvolvidas de espectrometria de massa. Essa abordagem não está limitada a classes particulares de moléculas e pode ser extremamente sensível. A massa do carbono 12 é definida como exatamente 12 unidades de massa molecular (daltons). Entretanto, as massas de outros átomos, como o nitrogênio 14 ou oxigênio 16, não apresentam números inteiros exatos. A espectrometria de massa, utilizando resolução de massa extremamente alta, atualmente possível em instrumentos especiais, permite a determinação inequívoca da fórmula molecular de qualquer molécula pequena. Claramente, isômeros apresentarão a mesma fórmula, porém poderão ser distinguidos pelos diferentes padrões de fragmentação durante a execução da espectrometria de massa. A mesma abordagem é utilizada para identificar os fragmentos de peptídeos originados da digestão de proteínas durante as análises de proteomas (Seção 6.8). Neste caso, a identificação de vários oligopeptídeos permite que a identidade da proteína parental seja deduzida desde que sua sequência de aminoácidos seja conhecida.

Na versão da espectrometria de massa, MALDI (do inglês, *matrix-assisted laser desorption ionization*), as amostras são ionizadas e vaporizadas por um *laser* (Figura 6.23). Os íons gerados são acelerados através de um campo elétrico ao longo da coluna até alcançarem o detector. O tempo de voo (TOF, *time of flight*) para cada íon depende da razão entre a massa/carga – quanto menor essa razão, mais rápido o íon se move. O detector mede o TOF de cada íon e o computador calcula a massa e gera a fórmula molecular. A combinação dessas duas técnicas é conhecida como *MALDI-TOF*.

A análise de metaboloma é especialmente útil no estudo de plantas, muitas das quais produzem vários milhares de diferentes metabólitos – mais do que a maioria dos outros tipos de organismos. Isso ocorre devido ao fato de as plantas produzirem muitos *metabólitos secundários*, como aromas, sabores, alcaloides e pigmentos, muitos dos quais são importantes comercialmente. Investigações metabolômicas monitoraram as concentrações de várias centenas de metabólitos na planta modelo, *Arabidopsis*, revelando alterações significativas nas concentrações de muitos desses metabólitos em resposta a alterações térmicas. Caminhos futuros da metabolômica, atualmente em desenvolvimento, incluem a avaliação do efeito de doenças no metaboloma de vários órgãos e tecidos humanos. Tais resultados deverão aumentar significativamente nosso conhecimento sobre como o corpo humano combate as doenças infecciosas e não infecciosas e identificar componentes-chave importantes para nossas defesas. Esses componentes podem possivelmente ser desenvolvidos, tais como fármacos para o tratamento clínico de doenças específicas.

### Biologia de sistemas

O termo **biologia de sistemas** vem sendo amplamente utilizado nos últimos anos para se referir à integração dos diferentes campos de pesquisa para dar uma visão geral de um organismo, ou de uma célula ou até mesmo de uma espécie ou de um ecossistema inteiro. A biologia de sistemas integra todas as “ômicas” que estudamos até agora: genômica, proteômica,



**Figura 6.23** Espectrometria de massa MALDI-TOF. Na espectrometria MALDI, as amostras são ionizadas por um *laser* e os íons atravessam pelo tubo até alcançarem o detector. O tempo de voo (TOF) depende da razão massa/carga ( $m/z$ ) do íon. O computador identifica os íons baseados no tempo de voo; isto é, o tempo que o íon gasta para alcançar o detector. A técnica MALDI-TOF é extremamente sensível e possui alta resolução.

metabolômica e assim por diante (Figura 6.24). A habilidade do computador de armazenar e analisar grandes quantidades de informação biológica é essencial para a biologia de sistemas, e o entendimento de todo o sistema biológico está evoluindo junto com o poder de armazenamento do computador.

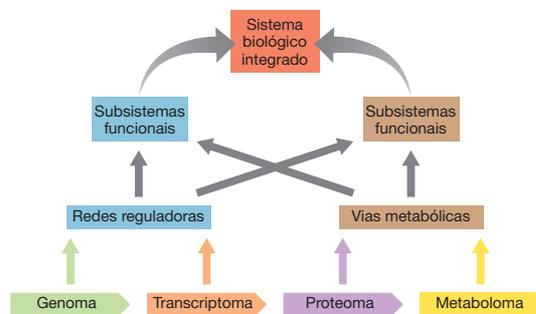
A estratégia básica da biologia de sistemas é a capacidade de compilar uma série de dados das “ômicas” e, em seguida, construir um modelo computacional do sistema em estudo (Figura 6.24). Esses modelos podem permitir a previsão de comportamentos ou propriedades de um organismo em particular que não eram evidentes a partir das observações originais. Essas são chamadas de *propriedades emergentes* de um organismo. Prevê-se que a compreensão das propriedades emergentes de um organismo fornecerá um conhecimento mais profundo sobre a biologia geral desse organismo do que qualquer outro “ômica” estudado sozinho possa oferecer.

#### MINIQUESTIONÁRIO

- Quais são as técnicas utilizadas para monitorar o metaboloma?
- O que é um metabólito secundário?
- Por que a biologia de sistemas depende de um computador potente? O que é uma propriedade emergencial?

## 6.10 Metagenômica

Comunidades microbianas contêm muitas espécies de bactérias e arqueias, sendo que a maioria dessas espécies nunca foi cultivada ou até mesmo identificada. A **metagenômica**, também



**Figura 6.24** Os componentes da biologia de sistemas. Os resultados de várias análises “ômicas” são combinados e, posteriormente, integrados formando uma visão maior da inteira biologia de um organismo.

conhecida como *genômica ambiental*, analisa o conjunto de RNA ou DNA presentes em amostras ambientais contendo organismos que não foram isolados e identificados. Da mesma forma que o total de genes de um organismo constitui o seu genoma, o total de genes dos organismos que habitam um ambiente é conhecido como **metagenoma** (Tabela 6.6). Além da metagenoma baseada no sequenciamento de DNA, análises baseadas em RNA ou proteínas podem ser utilizadas para explorar os padrões de expressão genética em uma comunidade microbiana natural. Com a tecnologia atual esses estudos podem ser feitos até mesmo em uma célula individual (ver Explore o mundo microbiano, “Genômica, uma célula de cada vez”). A genômica de uma única célula será discutida mais adiante no Capítulo 18.

### Exemplos de estudos com metagenômica

Vários ambientes foram estudados pelo projeto de sequenciamento metagenômico em larga escala. Ambientes extremos, tais como água ácida de escoamento de minas, tendem a ter baixa diversidade de espécies. Consequentemente, tem sido possível isolar todo o DNA da comunidade e montá-lo em sequências próximas ao genoma completo de um indivíduo. Por outro lado, ambientes complexos, tais como solos férteis ou ambientes aquáticos, são mais desafiadores, montar os genomas completos é muito mais difícil. Entretanto, uma descoberta surpreendente oriunda dos estudos com metagenômica é que a maioria dos genes destes ambientes não pertence a organismos celulares e sim a vírus. Isso será discutido melhor no Capítulo 9, em que falaremos sobre genômica e filogenia dos vírus.

Mesmo que genomas completos não possam ser montados, informações úteis podem ser obtidas de estudos metagenômicos. Por exemplo, ambientes, podem ser analisados quanto à presença e distribuição de diferentes grupos taxonômicos de bactérias. A abundância relativa das bactérias pode variar bastante de acordo com o ambiente, e a Figura 6.25 ilustra esse fenômeno para os principais subgrupos de Proteobacteria (Capítulo 15), em um local de amostragem perto do Havaí no Oceano Pacífico. Luz, oxigênio, nutrientes e mudanças de temperatura podem ser relacionados com qual subgrupo de protobactérias é mais competitivo em cada profundidade (Figura 6.25). Uma observação curiosa que surgiu com esses estudos de metagenômica é que uma boa parte do DNA em habitats naturais não pertence às células vivas. Por exemplo, cerca de 50 a 60% do DNA dos oceanos são extracelulares, encontrados nos sedimentos do fundo oceânico. Provavelmente esse DNA é depositado quando organismos mortos das camadas superiores do oceano vão para o fundo e desintegram-se. Devido ao fato de que os

# EXPLORE O MUNDO MICROBIANO

Genômica, uma célula de cada vez

UNIDADE 2

Análises modernas do genoma têm ampliado o número de amostras executadas simultaneamente e reduzido o tamanho das amostras. A redução do tamanho necessária para as análises genômicas tornou capaz a análise de até mesmo uma única célula – a técnica é denominada *genômica de célula única* –, e alguns excelentes resultados têm surgido.

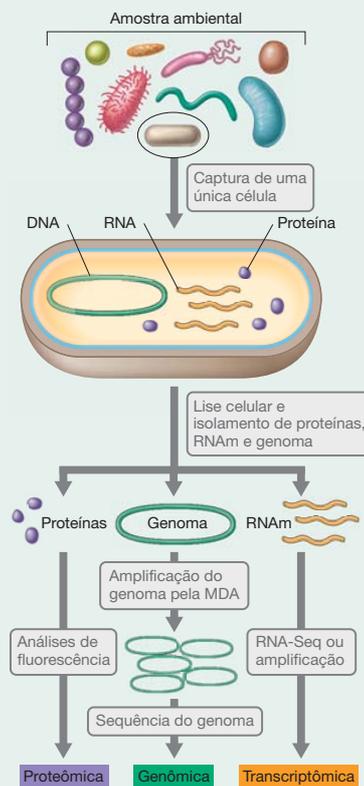
As células individuais podem ser isoladas através de várias técnicas físicas e serem submetidas aos procedimentos genômicos (Figura 1). Sequenciamento do genoma e análises de transcriptoma e proteoma estão sendo feitos com uma única célula. O sequenciamento de DNA de uma única célula se baseia em uma versão altamente modificada da reação de polimerização em cadeia conhecida como *amplificação por deslocamento múltiplo (MDA, multiple displacement amplification)* (Seção 18.11 e Figura 18.32). Essa técnica amplifica fentograma ( $10^{-15}$  g) de DNA presente em uma única célula bacteriana, enquanto para o sequenciamento são necessários microgramas de DNA (um bilhão de vezes mais). Do mesmo modo, o RNA pode ser analisado por RNA-Seq ou por uma amplificação através de uma versão modificada da técnica de PCR. As análises proteômicas de uma única célula são mais complicadas, mas análises que empregam métodos fluorescentes muito sensíveis estão disponíveis para este propósito.

O DNA de células únicas isoladas do solo ou de vários outros habitats tem sido sequenciado. Utilizando genômica de célula única, genes metabólicos presentes em um ambiente podem não só ser apenas identificados, mas atribuídos a uma determinada espécie. Portanto, a genômica de célula única pode revelar qual organismo dentro de uma comunidade microbiana está degradando quais nutrientes. Por exemplo, a genômica de célula única tem

sido utilizada para analisar a degradação de hidrocarbonetos em ambientes poluídos, levando a um melhor entendimento de quais organismos estão fazendo o que durante todo o processo. Da mesma forma, plasmídeos e vírus podem ser atribuídos ao hospedeiro correto quando uma única célula é sequenciada.

Uma descoberta surpreendente em estudos de célula única foi que os níveis de proteína e transcritos variam muito de uma célula para outra em uma cultura pura de bactéria em crescimento, presumivelmente como resultado da transcrição e da tradução que horas ocorrem em explosão e em outras horas devagar. Isso é especialmente verdade para proteínas expressas em baixos níveis. Consequentemente, e contra a intuição, para genes individuais em uma célula única existe pouca relação entre o número de cópias de um RNAm e sua proteína correspondente em qualquer tempo. Isso é em parte devido a diferença entre o tempo de vida média da proteína e da molécula de RNAm. Considerando que a maioria das proteínas sobrevive mais do que a geração celular e o RNAm em bactérias é degradado com 2 a 3 minutos após a sua síntese. Portanto, níveis de RNAm em qualquer tempo são determinados pela taxa de transcrição nos últimos minutos, já os níveis de proteínas refletem a síntese ao longo de uma hora ou mais.

A genômica de célula única tem um futuro brilhante na sondagem de várias facetas da biologia de um organismo em uma célula individual em vez de uma população de células. O método já desafiou suposições anteriores sobre a uniformidade bioquímica das células na fase exponencial das culturas, e é provável que muitas outras questões surjam e a genômica de célula única será ideal para responder a essas perguntas que não seriam possíveis com a cultura de células. A genômica de célula



**Figura 1** Genômica de célula única. Uma única célula isolada de uma amostra ambiental pode ser a fonte de uma diversidade de análises "ômicas".

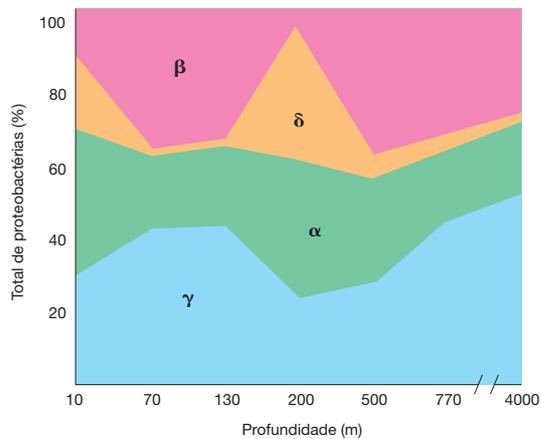
única é também um excelente exemplo de como métodos científicos desenvolvidos com um objetivo em mente (ou seja, a análise genômica de uma população de células) pode ser modificado por cientistas criativos para responder questões científicas que antes não eram possíveis.

ácidos nucleicos são os maiores repositórios de fosfato, o DNA é o principal contribuinte para o ciclo global do fósforo.

## Metagenômica e estudos de "biomas"

É estimado que o corpo humano contenha cerca de 10 trilhões ( $10^{13}$ ) de células, mas cada um de nós também carrega por volta de dez vezes mais células procarióticas do que humanas. Essa coleção de células procarióticas recebe o nome de *microbioma humano*. A maior parte desses procariotos habita o intestino e a maioria pertence a dois grupos bacterianos, os *Bacteroidetes*

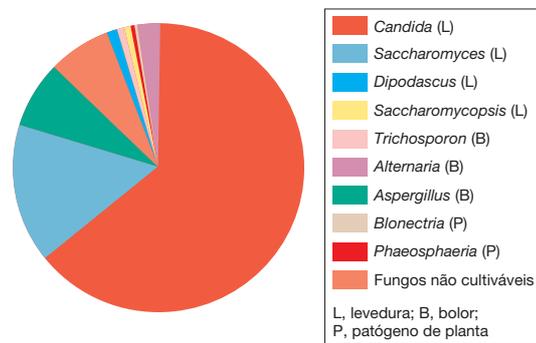
e os *Firmicutes* (Capítulo 15). Um achado fascinante é que a composição do microbioma intestinal relaciona-se com a obesidade tanto em seres humanos quanto em modelos murinos experimentais. Quanto maior a proporção de *Firmicutes* (principalmente *Clostridium* e bactérias relacionadas), mais gordo é o ser humano ou o camundongo. O mecanismo sugerido é que espécies de *Firmicutes* convertem mais fibras em cadeias curtas de ácidos graxos que podem ser absorvidos pelo hospedeiro. Assim, o hospedeiro absorve mais gordura da mesma quantidade de comida. Além disso, embora seja um importante organis-



**Figura 6.25** Metagenômica de Proteobacteria no oceano. A distribuição de acordo com a profundidade dos principais subgrupos (alfa  $\alpha$ , beta  $\beta$ , gama  $\gamma$  e delta  $\delta$ ) de Proteobacteria no Oceano Pacífico está demonstrada na figura. Muitos outros grupos de bactérias estão também presentes (não demonstrado). Dados adaptados de Kembel, S.W., J.A. Eisen, K.S. Pollard, e J.L. Green. 2011. *PLoS One* 6: e23214.

mo-modelo na biologia, a bactéria *Escherichia coli* compreende apenas cerca de 1% da população total de bactérias intestinais.

Estudos recentes sobre o microbioma intestinal humano e murino revelaram várias espécies de fungos (Figura 6.26), anteriormente não detectadas; estes compõem o que chamamos de *micobioma* (o prefixo “mico” significa fungo). Muitos destes são leveduras comuns, como *Saccharomyces* e *Candida*, embora alguns dos fungos intestinais detectados, tais como



**Figura 6.26** O microbioma murino. Os dados mostram a população fúngica do intestino do camundongo. O gráfico mostra que os fungos mais comuns são as leveduras. Dados adaptados de Iliev, I.D., et al. *Science* 336: 1314–1317 (2012).

*Aspergillus* e *Trichosporon*, são graves patógenos potenciais (Figura 6.26). Além disso, embora os fungos intestinais constituam menos do que 1% do microbioma, já se sabe que certas condições como a síndrome do intestino irritável se relaciona fortemente com populações específicas de fungos. Portanto, a metagenômica é uma grande promessa para investigar possíveis conexões entre populações microbianas específicas e certas doenças em seres humanos e outros animais.

#### MINIQUESTIONÁRIO

- O que é um metagenoma?
- Como o metagenoma é analisado?
- Como o microbioma humano e micobioma se diferem?

## IV • A evolução dos genomas

Além do entendimento de como os genes atuam e os organismos interagem com o ambiente, a genômica comparativa pode também revelar as relações evolutivas entre os organismos. A reconstrução das relações evolutivas a partir das sequências genômicas auxilia a distinção entre características primitivas e derivadas, assim como pode resolver ambiguidades nas árvores filogenéticas baseadas em análises de um único gene, como o RNAr da subunidade menor (↔ Seção 12.4). A genômica é também um elo para a compreensão das formas primitivas de vida e, eventualmente, pode responder a questão mais fundamental da biologia: como surgiu a vida?

### 6.11 Famílias gênicas, duplicações e deleções

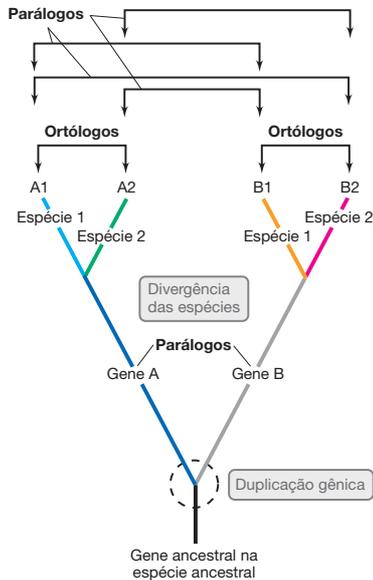
Genomas de origem procariótica e eucariótica frequentemente contêm cópias múltiplas de genes cujas sequências são relacionadas devido ao compartilhamento de um ancestral evolutivo; esses genes são denominados *genes homólogos* ou *homólogos*. Grupos de genes homólogos são chamados de **famílias gênicas**. Assim, não surpreende que genomas maiores tendam a conter um número maior de membros individuais de uma determinada família gênica.

#### Parálogos e ortólogos

A genômica comparativa revelou que muitos genes surgiram pela *duplicação* de outros. Tais homólogos devem ser subdivididos de acordo com suas origens. Genes cuja similaridade é resultante da duplicação gênica em algum momento na evolução de um organismo são denominados **parálogos**. Genes encontrados em um organismo que são similares a genes de outro organismo porque descenderam de um mesmo ancestral comum são denominados **ortólogos** (Figura 6.27). Ortólogos são frequentemente não idênticos porque divergiram devido à especiação ou a eventos evolutivos mais distantes. Como exemplo de genes parálogos temos os codificadores de várias isoenzimas diferentes da lactato desidrogenase (LDH) em seres humanos. Essas enzimas, denominadas *isoenzimas*, são estruturalmente distintas, embora sejam altamente relacionadas e realizem a mesma reação enzimática. Por outro lado, a LDH correspondente da bactéria do ácido láctico, *Lactobacillus*, é ortóloga a todas as isoenzimas LDH humanas. Assim, famílias gênicas contêm tanto parálogos quanto ortólogos.

#### Duplicação gênica

Considera-se que a duplicação gênica é o mecanismo de evolução da maioria dos novos genes. Se um segmento de DNA



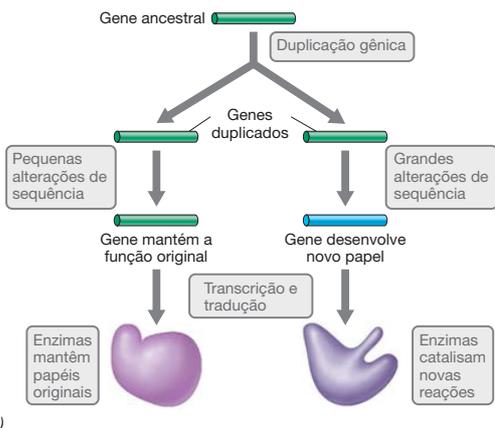
**Figura 6.27 Ortólogos e parálogos.** Esta árvore genealógica ilustra um gene ancestral que foi duplicado e divergiu em dois genes parálogos, A e B. Em seguida, a espécie ancestral divergiu na espécie 1 e espécie 2, ambas apresentando genes A e B (designados A1 e B1, A2 e B2, respectivamente). Cada um desses pares é um parálogo. Entretanto, como as espécies 1 e 2 consistem atualmente em espécies distintas, A1 é ortólogo de A2 e B1 é ortólogo de B2.

duplicado for longo o suficiente para incluir um gene inteiro ou um grupo de genes, o organismo contendo a duplicação possui cópias múltiplas desses genes em particular. Após a duplicação, uma das duplicatas encontra-se livre para evoluir, enquanto a outra cópia continua a suprir a célula com a função original (Figura 6.28a). Desse modo, a evolução pode “testar” uma cópia do gene. Esses eventos de duplicação gênica, segui-

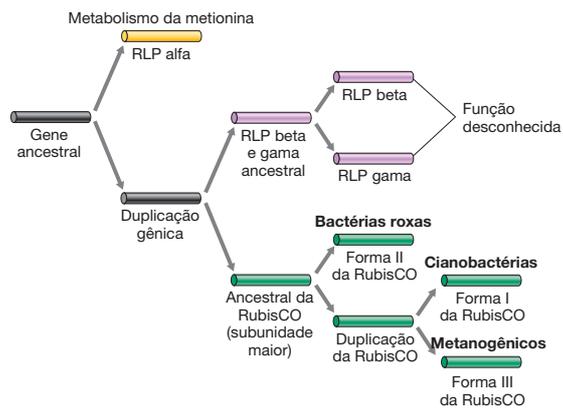
dos pela diversificação de uma das cópias, são considerados os principais eventos que alimentam a evolução. Análises genômicas revelaram inúmeros exemplos de genes codificadores de proteínas que claramente derivaram de duplicação gênica. A Figura 6.28b demonstra isso para a enzima RubisCO, a enzima-chave do metabolismo autotrófico (↔ Seção 13.5). Aqui, um gene ancestral deu origem a enzimas com atividades catalíticas diferentes, porém relacionadas.

As duplicações que ocorrem no material genético podem incluir apenas poucas bases ou até mesmo genomas inteiros. Por exemplo, a comparação de genomas da levedura *Saccharomyces cerevisiae* com outros fungos sugere que o ancestral de *Saccharomyces* duplicou seu genoma inteiro. Isso foi acompanhado por extensas deleções que eliminaram grande parte do material genético duplicado. A análise do genoma da planta-modelo *Arabidopsis* sugere a ocorrência de uma ou mais duplicações do genoma inteiro no ancestral das plantas de floração.

Os genomas bacterianos evoluíram por duplicação do genoma inteiro? A distribuição de genes duplicados e famílias gênicas nos genomas de bactérias sugerem que muitas duplicações frequentes, porém relativamente pequenas, ocorreram. Por exemplo, entre as Deltaproteobacteria, a bactéria de solo *Mycococcus* apresenta um genoma de 9,1 Mpb. Isso corresponde a aproximadamente o dobro dos genomas de outras Deltaproteobacteria típicas, que variam de 4 a 5 Mpb. Entre as Alphaproteobacteria, o tamanho do genoma varia de 1,1 a 1,5 Mpb para membros parasitas, 4 Mpb para *Caulobacter* de vida livre e até 7 a 9 Mpb para bactérias associadas a plantas (Tabela 6.1). Contudo, em todos esses casos, a análise da distribuição gênica aponta para duplicações frequentes em pequena escala, em vez de duplicações do genoma inteiro. Contrariamente, em bactérias parasitas, deleções frequentes sucessivas eliminaram genes não mais necessários a um estilo de vida parasitário, resultando em seus pequenos genomas não usuais (Seção 6.4, Tabela 6.1 e Figuras 6.8 e 6.14).



(a)



(b)

**Figura 6.28 Evolução pela duplicação gênica.** (a) O princípio da duplicação gênica. Após a duplicação, uma cópia do gene é livre para desenvolver uma nova função. (b) A família de genes RubisCO (*rbcL*). A subunidade maior da enzima RubisCO que fixa o CO<sub>2</sub> durante a fotossíntese se dividiu em três formas intimamente relacionadas (I, II e III), sendo que todas mantiveram a

função original (barras verdes). Entretanto, a RubisCO, por sua vez, é derivada de um gene ancestral (barras pretas) de função desconhecida que se dividiu e produziu um gene que codifica uma enzima que faz parte do metabolismo da metionina (barra amarela) e vários outros genes que ainda não possuem funções conhecidas (barras roxas). RPL, proteína do tipo RubisCO.

### Análise gênica em diferentes domínios

A comparação de genes e de famílias gênicas é uma das principais tarefas da genômica comparativa. Uma vez que cromossomos de muitos microrganismos diferentes já foram sequenciados, essas comparações podem ser facilmente realizadas, gerando resultados que, com frequência, são surpreendentes. Por exemplo, genes de arqueias envolvidos na replicação de DNA, transcrição e tradução são mais similares às de eucariotos do que de bactérias. Inesperadamente, entretanto, muitos outros genes de arqueias, por exemplo, aqueles envolvidos em funções metabólicas distintas do processamento de informação, são mais similares aos de bactérias do que de eucariotos. As poderosas ferramentas analíticas da bioinformática permitem que as relações genéticas entre quaisquer organismos sejam rapidamente deduzidas em termos de genes únicos, grupo gênico ou genoma inteiro. Os resultados obtidos até o momento reforçaram o quadro filogenético da vida deduzido originalmente pela análise comparativa de sequências de RNA ribossomal (↔ Seção 12.4) e sugerem que muitos genes de todos os organismos possuem raízes evolutivas comuns. Entretanto, essas análises também revelaram exemplos de fluxo gênico horizontal, uma questão importante a ser discutida em seguida.

#### MINIQUESTIONÁRIO

- O que é um gene homólogo?
- O que é uma família gênica?
- Diferencie genes parálogos de genes ortólogos.

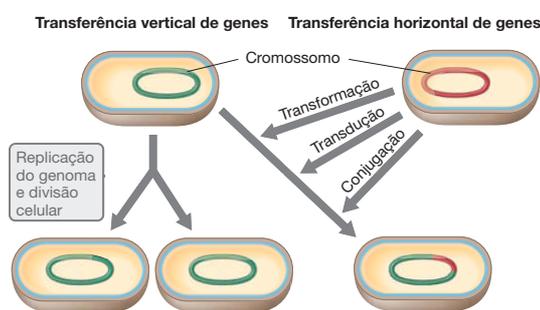
## 6.12 Transferência horizontal de genes e estabilidade do genoma

A evolução baseia-se na transferência de características genéticas de uma geração para a próxima. Entretanto, em procariotos, a **transferência horizontal de genes** (às vezes chamada de *transferência lateral de genes*) também ocorre e pode dificultar as análises dos genomas.

A transferência horizontal de genes ocorre quando eles são transferidos de uma célula para outra, de maneira distinta do processo hereditário usual (vertical), de célula-mãe para célula-filha (Figura 6.29). Em procariotos, pelo menos três mecanismos de transferência horizontal de genes são conhecidos: *transformação*, *transdução* e *conjugação* (Capítulo 10). O fluxo gênico horizontal pode ser amplo na natureza e algumas vezes cruza até mesmo as fronteiras dos domínios filogenéticos. Contudo, para ser detectável pela genômica comparativa, a diferença entre os organismos deve ser bastante grande. Por exemplo, vários genes com origem eucariótica foram encontrados em *Chlamydia* e *Rickettsia*, ambos patógenos humanos. Particularmente, dois genes codificadores de proteínas do tipo histona H1 foram detectados no genoma de *Chlamydia trachomatis*, sugerindo transferência horizontal a partir de uma fonte eucariótica, possivelmente até mesmo seu hospedeiro humano. Observe que essa situação é oposta àquela na qual genes do ancestral da mitocôndria foram transferidos para o núcleo eucariótico (Seção 6.5).

### Deteção do fluxo horizontal de genes

As transferências horizontais de genes podem ser detectadas em genomas uma vez que os genes tenham sido anotados (Seção 6.3). A presença de genes que codificam proteínas nor-



**Figura 6.29** Transferência vertical de genes versus transferência horizontal. Transferência vertical de genes ocorre quando as células se dividem. Já a transferência horizontal de genes ocorre quando a célula doadora transfere seus genes para a célula receptora. Nos procariotos, a transferência horizontal ocorre através de um dos três mecanismos: transformação, transdução e conjugação.

malmente encontradas somente em espécies distantes é um sinal de que os genes foram originados por transferência horizontal. Entretanto, outro indício de genes horizontalmente transferidos é a presença de um segmento de DNA cujo conteúdo de GC, ou utilização preferencial de códons, diferem significativamente do restante do genoma (Figura 6.29). Utilizando essas informações, muitos prováveis exemplos de transferência horizontal foram documentados nos genomas de vários procariotos. Um exemplo clássico consiste no organismo *Thermotoga maritima*, uma espécie de bactéria, que contém mais de 400 genes (mais de 20% do seu genoma) de origem de arqueias. Desses, 81 foram encontrados em agrupamentos distintos. Isso sugere fortemente que esses genes foram obtidos por transferência horizontal de genes, provavelmente de arqueias termofílicas que partilham os ambientes quentes habitados por *Thermotoga*.

Os genes adquiridos por transferência horizontal geralmente codificam funções metabólicas distintas dos processos moleculares centrais de replicação de DNA, transcrição e tradução, podendo ser responsáveis pelas similaridades de genes metabólicos entre arqueias e bactérias previamente mencionados (Seção 6.4). Além disso, existem vários exemplos de genes de virulência de patógenos que foram transferidos horizontalmente. Obviamente, os procariotos estão trocando genes na natureza e esse processo provavelmente atua realizando um “ajuste fino” no genoma do organismo a uma situação ou um habitat particular. É necessária cautela quando se invoca a transferência horizontal de genes para explicar a sua distribuição. Quanto o genoma humano foi inicialmente sequenciado, mais de 200 genes foram identificados como oriundos de transferência horizontal de genes a partir de procariotos. Entretanto, quando um maior número de genomas eucarióticos tornou-se disponível para análise, foram encontrados homólogos para a maioria desses genes em várias linhagens eucarióticas. Por isso, atualmente acredita-se que a maioria desses genes tenha, de fato, origem eucariótica. Somente cerca de doze genes são atualmente aceitos como fortes candidatos a possuírem origem procariótica relativamente recente. A frase “relativamente recente” nesse contexto refere-se a genes transferidos de procariotos após a separação das principais linhagens eucarióticas (↔ Seção 12.4), e não a genes de possível origem procariótica ancestral, que são compartilhados por todos os eucariotos.

### Evolução do genoma e elementos móveis

Como descrito na Seção 10.11, DNA móvel refere-se a segmentos de DNA que se movem de um local para outro no interior de moléculas de DNA do hospedeiro. A maioria dos DNAs móveis consiste em *elementos transponíveis*, porém genomas virais integrados e integrons são também encontrados. Todos esses elementos móveis desempenham importantes papéis na evolução do genoma (Figura 6.30).

*Transposons* são formas comuns de DNA móvel que podem mover-se entre diferentes moléculas de DNA do hospedeiro, incluindo cromossomos, plasmídeos e vírus, pela atividade da enzima *transposase* (Seção 10.11). Ao deslocar-se, eles podem carrear e transferir horizontalmente genes que codificam uma variedade de características, incluindo resistência a antibióticos ou produção de toxinas. Entretanto, transposons podem também mediar uma variedade de alterações cromossômicas em larga escala (Figura 6.30). Bactérias que estão em rápida mudança evolutiva contêm números relativamente elevados de elementos móveis, especialmente sequências de inserção, elementos transponíveis cujos genes codificam exclusivamente para transposição. A recombinação entre elementos idênticos gera rearranjos cromossômicos como deleções, inversões ou translocações. Acredita-se que tal processo represente uma fonte de diversidade genômica sobre qual seleção pode atuar. Assim, rearranjos cromossômicos que se acumulam em bactérias durante o crescimento em condições de estresse são frequentemente flanqueados por repetições ou sequências de inserção.

Contrariamente, uma vez que uma espécie se estabelece em um nicho evolutivo estável, a maioria dos elementos móveis é aparentemente perdida. Por exemplo, genomas de espécies de *Sulfolobus* (*Archaea*) possuem números extremamente elevados de sequências de inserção e exibem uma alta

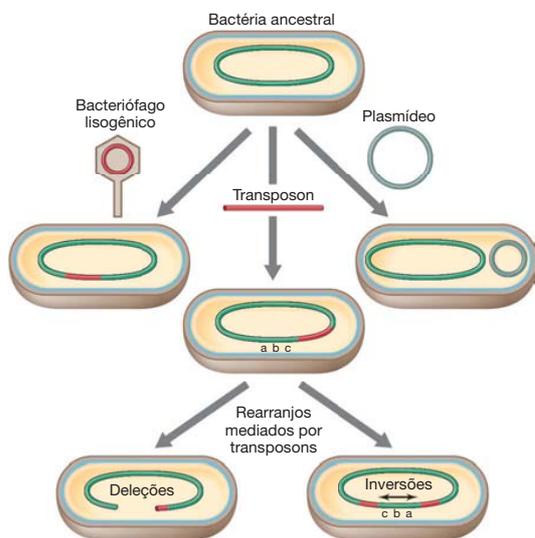
frequência de translocações gênicas. Por outro lado, *Pyrococcus* (*Archaea*) é praticamente desprovido de sequências de inserção, exibindo um número correspondentemente baixo de translocações gênicas. Isso sugere que, por qualquer razão, talvez devido a flutuações nas condições de seus habitats, o genoma de *Sulfolobus* seja mais dinâmico do que o genoma mais estável de *Pyrococcus*.

Rearranjos cromossômicos devido às sequências de inserção aparentemente contribuíram para a evolução de vários patógenos bacterianos. Em *Bordetella*, *Yersinia* e *Shigella*, as espécies mais altamente patogênicas apresentam uma frequência muito maior de sequências de inserção. Por exemplo, *Bordetella bronchiseptica* possui um genoma de 5,3 Mpb, porém não apresenta sequências de inserção conhecidas. Seu relacionado mais patogênico, *Bordetella pertussis*, o agente causador da coqueluche (Seção 29.3), apresenta um genoma menor (4,1 Mpb), mas possui mais de 260 sequências de inserção. A comparação desses genomas sugere que as sequências de inserção são responsáveis por substanciais rearranjos genômicos, incluindo deleções responsáveis pela redução do tamanho genômico em *B. pertussis*.

As sequências de inserção também desempenham um papel na organização de módulos genéticos que geram novos plasmídeos. Assim, 46% do megaplasmídeo de virulência de 220 Kpb de *Shigella flexneri* consistem em DNA de sequências de inserção. Além das sequências de inserção completas, existem muitos fragmentos nesse plasmídeo, sugerindo múltiplos rearranjos ancestrais.

#### MINIQUESTIONÁRIO

- Qual classe de genes é raramente transferida horizontalmente? Por quê?
- Liste os principais mecanismos pelos quais a transferência horizontal de genes ocorre em procariontes.
- Por que os transposons são especialmente importantes na evolução de bactérias patogênicas?

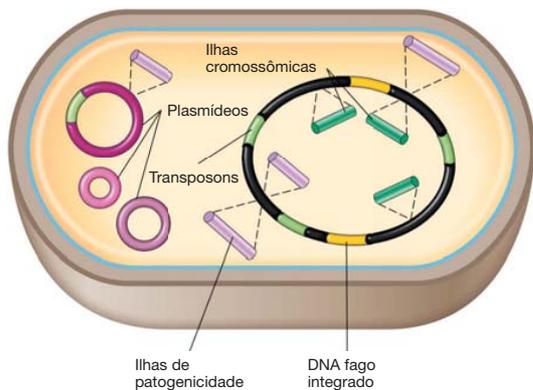


**Figura 6.30** Elementos móveis promovem evolução do genoma. Uma variedade de elementos genéticos móveis pode se mover de um organismo para o outro, adicionando genes ao genoma do organismo receptor. Os mais comuns deles são plasmídeos, bacteriófagos e transposons. Neste último caso, rearranjos cromossômicos, como deleções e inversões do DNA vizinho ao transposon, podem ser mediados pela atividade da enzima *transposase*.

### 6.13 Genoma cerne versus pangenoma

Um dos conceitos mais importantes que surge ao se comparar as sequências genômicas de várias amostras da mesma espécie é a distinção entre o **genoma cerne** e o **pangenoma**. O genoma *cerne* é aquele compartilhado por todas as amostras de uma determinada espécie, entretanto o *pangenoma* inclui o genoma cerne mais todas as sequências extras presentes em uma ou mais amostras, mas não em todas as amostras da mesma espécie (Figura 6.31). Como foi visto, a transferência horizontal de genes de elementos genéticos inteiros, como plasmídeos, vírus ou elementos transponíveis, é possível. Consequentemente, pode haver grandes diferenças na quantidade total de DNA e no conjunto de capacidades acessórias (virulência, simbiose ou biodegradação) entre amostras de uma única espécie bacteriana. Em outras palavras, podemos dizer que geralmente o genoma cerne é o genoma da espécie como um todo, entretanto os outros componentes do pangenoma, frequentemente elementos móveis, estão restritos a amostras particulares dentro de cada espécie.

É difícil definir precisamente o tamanho do pangenoma porque ele aumenta à medida que mais amostras da espécie são sequenciadas. Em alguns casos, tais como os das ente-



**Figura 6.31** Pangenoma versus genoma cerne. O genoma cerne está representado pelas regiões pretas do cromossomo e está presente em todas as amostras da espécie. O pangenoma inclui elementos que estão presentes em uma ou mais amostras, mas não em todas. Cada barra colorida indica uma única inserção. Quando duas barras surgem do mesmo local, elas representam ilhas alternativas que podem ser inseridas em um mesmo sítio. Entretanto, somente uma inserção pode estar presente em uma dada localização. Plasmídeos, assim como cromossomos, podem possuir inserções que não estão presentes em todas as amostras.

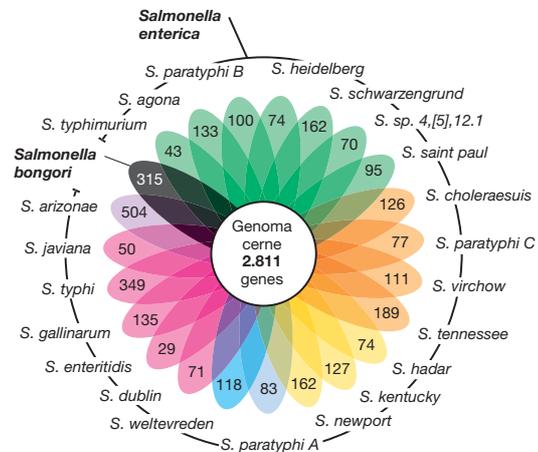
robactérias *Escherichia coli* e *Salmonella enterica*, muitos isolados diferentes têm sido descobertos transportando uma grande variedade de diferentes plasmídeos, transposons e similares. Consequentemente, o pangenoma se torna bem grande. A **Figura 6.32** ilustra o pangenoma dos sorotipos (linhagens) de um importante patógeno humano, *Salmonella enterica*, representado em um gráfico “flowerplot”.

### Ilhas cromossômicas

A comparação do genoma cerne e do pangenoma de determinadas bactérias ou genomas de determinadas espécies com aqueles de organismos relacionados frequentemente revela blocos adicionais de material genético que são partes de cromossomos, em vez de plasmídeos ou vírus integrados. Estes são chamados de **ilhas cromossômicas** que contêm grupos de genes para funções especializadas que não são necessárias para a simples sobrevivência (Figura 6.31). Consequentemente, duas linhagens da mesma espécie bacteriana podem apresentar diferenças significantes no tamanho do genoma.

Não surpreendentemente, as ilhas cromossômicas em bactérias patogênicas têm atraído muitas atenções. Entretanto, são também conhecidas ilhas cromossômicas que carregam genes para a biodegradação de vários substratos derivados da atividade humana, como hidrocarbonetos aromáticos e herbicidas. Além disso, muitos dos genes essenciais à relação simbiótica de rizóbios com plantas na simbiose dos nódulos radiculares (↔ Seção 22.3) são carregados em ilhas cromossômicas. Talvez a ilha cromossômica mais singular seja a ilha de magnetossomo da bactéria *Magnetospirillum*; esse fragmento de DNA carrega os genes necessários à formação de magnetossomos, partículas magnéticas intracelulares utilizadas para orientar o organismo em um campo magnético e influenciar a direção da sua mobilidade (↔ Seção 2.14).

Acredita-se que as ilhas cromossômicas possuam uma origem “exógena”, com base em várias observações. Primeiro,



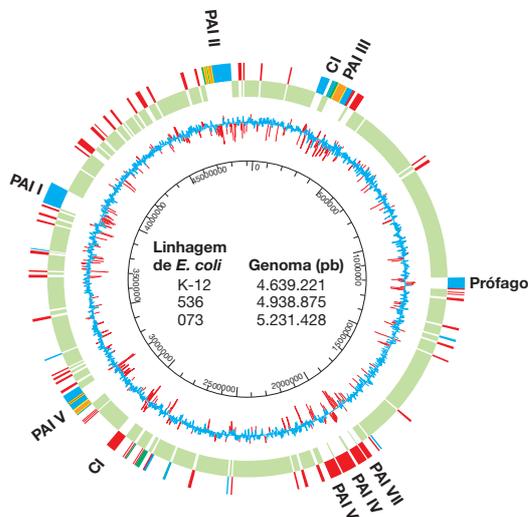
**Figura 6.32** Gráfico “flowerplot” do pangenoma da *Salmonella enterica*. Gráfico de “flowerplot” da família de genes nos sorotipos (linhagens) da bactéria patogênica gram-negativa *Salmonella enterica* (os nomes que circundam o gráfico são imunologicamente sorotipos únicos [S.] de *S. enterica*). A figura apresenta a média de famílias de genes, encontradas em cada genoma, únicas para cada sorotipo. *Salmonella bongori* é uma espécie diferente de *S. enterica*. O sorotipo 4,[5],12,i foi identificado recentemente e ainda não recebeu um nome. Dados de Jacobsen, A., R.S. Hendriksen, F.M. Aarestrup, D.W. Ussery, e C. Friis. 2011. The *Salmonella enterica* pan-genome. *Microb Ecol* 62: 487–504.

essas regiões adicionais são frequentemente flanqueadas por repetições invertidas, implicando que a região inteira foi inserida no cromossomo por transposição (Seção 6.12) em algum momento do passado evolutivo recente. Segundo, em ilhas cromossômicas, a composição de bases e uso preferencial de códons (Tabela 6.3) frequentemente difere de forma significativa do restante do genoma. Terceiro, as ilhas cromossômicas são frequentemente encontradas em algumas linhagens de determinadas espécies, mas não em outras.

Algumas ilhas cromossômicas carregam um gene da enzima integrase e, provavelmente, movem-se de forma análoga aos transposons conjugativos (Seção 6.12). As ilhas cromossômicas estão normalmente inseridas em um gene de RNAt; entretanto, como o sítio-alvo é duplicado durante a inserção, um gene de RNAt intacto é regenerado durante o processo de inserção. Em poucos casos, a transferência de uma ilha cromossômica completa entre bactérias relacionadas foi demonstrada em laboratório; provavelmente a transferência ocorra por qualquer um dos mecanismos de transferência horizontal anteriormente discutidos: transformação, transdução e conjugação (Figura 6.29). Supõe-se que, após a inserção no genoma de uma nova célula hospedeira, as ilhas cromossômicas gradualmente acumulem mutações – tanto mutações pontuais quanto pequenas deleções. Assim, após muitas gerações, as ilhas cromossômicas tendem a perder sua mobilidade.

### Ilhas de patogenicidade e a evolução da virulência

A comparação do genoma de bactérias patogênicas com aqueles de organismos relacionados não patogênicos frequentemente revela ilhas cromossômicas que codificam **fatores de virulência**, proteínas especiais, ou outras moléculas, ou es-



**Figura 6.33** Ilhas de patogenicidade em *Escherichia coli*. Mapa genético da linhagem 536 de *E. coli*, um patógeno do trato urinário, comparado a uma segunda linhagem patogênica (073) e a linhagem selvagem K-12. As linhagens patogênicas contêm ilhas de patogenicidade e, desse modo, seus cromossomos são maiores que os de K-12. Círculo interno, pares de bases nucleotídicas. Círculo irregular, distribuição de GC no DNA; regiões onde o conteúdo de GC difere grandemente da média do genoma estão em vermelho. Círculo externo, comparação dos três genomas: verde, genes comuns a todas as linhagens; vermelho, genes presentes somente nas linhagens patogênicas; azul, genes encontrados somente na linhagem 536; cor de laranja, genes da linhagem 536 presentes em uma localização diferente na linhagem 073. Alguns insertos muito pequenos foram deletados para maior clareza. PAI, ilhas de patogenicidade; CI, ilha cromossômica. Prófago, DNA de um bacteriófago temperado. Observe a relação entre as ilhas genômicas e o conteúdo distorcido de GC. Dados adaptados de *Proc. Natl. Acad. Sci. (USA)*, 103: 12879-12884 (2006).

truturas necessárias para causar doença (Capítulo 23). Alguns genes de virulência são carreados em plasmídeos ou bacteriófagos lisogênicos (↔ Seções 8.8 e 10.7). Entretanto, muitos outros estão agrupados em regiões cromossômicas denominadas **ilhas de patogenicidade** (Figura 6.31 e Figura 6.33).

As ilhas de patogenicidade são as mais conhecidas entre as ilhas cromossômicas. Embora as ilhas de patogenicidade sejam consideradas como uma subclasse das ilhas cromossômicas, ilhas relacionadas geneticamente que compartilham genes homólogos de integração e conjugação podem conter genes de virulência em algumas bactérias e de biodegradação em outras. Por exemplo, a identidade e a localização cromossômica da maioria dos genes de linhagens patogênicas de *Escherichia coli* correspondem às da linhagem laboratorial inócua, *E. coli* K-12, conforme esperado. Contudo, a maioria das linhagens patogênicas contém ilhas de patogenicidade de tamanho considerável, as quais estão ausentes em *E. coli* K-12 (Figura 6.33). Conseqüentemente, duas linhagens da mesma espécie bacteriana podem apresentar diferenças significativas no tamanho do genoma devido à presença ou a ausência de ilhas. Por exemplo, como apresentado na Tabela 6.1, a linhagem entero-hemorrágica de *E. coli* O157:H7 contém 20% mais de DNA e genes do que a linhagem *E. coli* K12.

Ilhas de patogenicidade pequenas que codificam uma série de fatores de virulência estão presentes em certas amostras da bactéria patogênica gram-positiva *Staphylococcus aureus* e podem ser movidas entre as células pelos bacteriófagos (↔ Seção 10.7). As ilhas são menores do que o genoma do fago, e quando são separadas do cromossomo e se replicam, elas induzem a formação de partículas de fagos defectivas que carregam os genes das ilhas, mas que são tão pequenos que não conseguem carregar o genoma do fago. Dessa maneira, as amostras de *S. aureus* que não possuem ilhas podem rapidamente obtê-las e se tornarem mais patogênicas.

#### MINIQUESTIONÁRIO

- Qual é a diferença entre o genoma cerne e o pangenoma?
- O que é uma ilha cromossômica e como identificar se ela é de origem exógena?
- O que é uma ilha de patogenicidade e como ela se move entre espécies bacterianas?

## CONCEITOS IMPORTANTES

**6.1** • Os pequenos vírus foram os primeiros organismos que tiveram os seus genomas sequenciados, mas agora os genomas de muitos procariotos e eucariotos também já foram sequenciados.

**6.2** • A tecnologia de sequenciamento de DNA está avançando rapidamente. O método original de Sanger raramente é utilizado e existem quatro sucessivas gerações de tecnologia de sequenciamento. Avanços nas técnicas têm aumentado a velocidade do sequenciamento de DNA. A técnica “*shotgun*” utiliza a clonagem aleatória e o sequenciamento de pequenos fragmentos do genoma seguido pela montagem do genoma gerada pelo computador.

**6.3** • A análise computacional dos dados de sequenciamento é uma parte vital da genômica. Ferramentas computacionais são utilizadas para armazenar e analisar as sequências e as estruturas das macromoléculas biológicas.

**6.4** • O tamanho dos genomas procarióticos sequenciados varia de 0,15 a 13 Mbp. Os menores genomas de procariotos são menores do que os maiores genomas virais, entretanto os maiores possuem mais genes do que alguns genomas de eucariotos. O conteúdo genético dos procariotos é geralmente proporcional ao tamanho do genoma. Muitos genes podem ser identificados pela sua similaridade genética com genes encontrados em outros organismos. Entretanto, uma porcentagem significativa de genes sequenciados possui funções desconhecidas.

**6.5** • Praticamente todas as células eucarióticas possuem mitocôndrias, e, além disso, as células vegetais ainda possuem cloroplastos. Ambas as organelas contêm genomas de DNA circular que codificam RNAs, RNAs e umas poucas proteínas necessárias para o metabolismo energético. Embora os genomas das organelas sejam independentes do genoma nuclear, as organelas não são. Muitos genes no núcleo codificam proteínas requeridas para as funções das organelas.

**6.6** • A sequência genômica completa de muitos microrganismos eucariotos foram determinadas. O genoma da levedura *Saccharomyces cerevisiae* codifica por volta de 6.000 proteínas, das quais somente cerca de 900 parecem ser essenciais. Relativamente poucos dos genes de levedura que codificam proteínas contêm íntrons. O número total de genes nos microrganismos eucariotos varia de 2.000 (menos do que muitos procariotos) a 60.000 (mais que o dobro dos seres humanos).

**6.7** • Microarranjos são genes ou fragmentos de genes aderidos a um suporte sólido de padrão conhecido; o

RNA<sub>m</sub> é então hibridizado como DNA para determinar os padrões de expressão gênica. Os arranjos são grandes o suficiente para o padrão transcricional de um genoma inteiro (transcriptoma) ser analisado. O RNA-Seq requer o sequenciamento massivo do DNA<sub>c</sub> para analisar o transcriptoma, além de requerer também tecnologias de sequenciamento de terceira e quarta gerações.

**6.8** • A proteômica é a análise de todas as proteínas presentes em um organismo. O objetivo final da proteômica é entender a estrutura, função e regulação dessas proteínas. O interactoma é o conjunto total de interações entre as macromoléculas dentro da célula.

**6.9** • O metaboloma é o conjunto completo de intermediários metabólicos produzidos por um organismo. A biologia de sistemas utiliza dados da genômica, transcriptoma e outras “ômicas” para construir modelos computacionais das atividades moleculares e interações nas células.

**6.10** • A maioria dos microrganismos do meio ambiente nunca foi cultivado. No entanto, análises de amostras de DNA revelaram enormes diversidades de sequências na maioria dos habitats. O conceito de metagenoma envolve o conteúdo genético total de todos os organismos em um ambiente particular.

**6.11** • A genômica pode ser utilizada para estudar a história evolucionária de um organismo. Organismos possuem famílias gênicas, genes com sequências relacionadas. Se os genes surgiram por causa da duplicação gênica, são denominados parálogos, se eles surgiram devido à especiação, eles são denominados ortólogos.

**6.12** • Os organismos podem adquirir genes de outros organismos em seu ambiente pela transferência horizontal de genes, e tais transferências podem até mesmo atravessar os limites dos domínios filogenéticos. Elementos de DNA móveis, incluindo transposons, integrons e vírus, são importantes na evolução genômica e frequentemente carregam genes que conferem resistência a antibióticos e fatores de virulência.

**6.13** • A comparação de genomas de múltiplas amostras da mesma espécie bacteriana mostra um componente conservado (o genoma cerne), além de vários módulos genéticos variáveis somente presentes em certos membros da espécie (o pangenoma). Muitas bactérias possuem relativamente grandes inserções de origem exógena conhecidas como ilhas cromossômicas. Essas contêm grupos de genes que codificam funções metabólicas especializadas ou patogênese e fatores de virulência (ilhas patogênicas).

## REVISÃO DE TERMOS-CHAVE

**Biblioteca genômica** uma coleção de fragmentos de DNA clonados que cobrem todo o genoma.

**Bioinformática** utilização de ferramentas computacionais para adquirir, analisar, armazenar e acessar sequências de DNA e proteínas.

**Biologia de sistemas** a integração de dados da genômica e de outras áreas “ômicas” para construir uma visão global de um sistema biológico.

**Chip genético** pequenos suportes nos quais porções de genes são fixadas e dispostas espacialmente em um padrão conhecido (também chamados de microarranjos).

**Códons preferenciais** proporção relativa dos diferentes códons que codificam o mesmo aminoácido, variam nos diferentes organismos. O mesmo que utilização de códons.

**Família gênica** genes que exibem sequências relacionadas com o resultado de uma origem evolutiva comum.

**Fase de leitura aberta (ORF)** sequência de DNA ou RNA que pode ser traduzida para gerar um peptídeo.

**Genoma** conjunto total da informação genética de uma célula ou um vírus.

**Genoma cerne** parte do genoma compartilhada por todas as amostras da espécie.

**Genômica** disciplina que mapeia, sequencia, analisa e compara genomas.

**Hibridização** junção de duas fitas simples de moléculas de ácido nucleico pela complementaridade dos pares de base para formar uma molécula híbrida de dupla-fita (DNA ou DNA-RNA).

**Homólogos** sequência relacionada em um grau que implica uma ancestralidade genética comum; incluem os ortólogos e parálogos.

**Ilha cromossômica** região do cromossomo bacteriano de origem exógena que contém genes agrupados que conferem alguma propriedade adicional, como virulência ou simbiose.

**Ilha de patogenicidade** região do cromossomo bacteriano de origem exógena que contém genes agrupados que conferem virulência.

**Iniciador** oligonucleotídeo no qual a DNA-polimerase liga o primeiro desoxirribonucleotídeo durante a síntese de DNA.

**Interatoma** o conjunto total de interações entre proteínas (ou outras macromoléculas) em um organismo.

**Metaboloma** conjunto total de pequenas moléculas e intermediários metabólicos de uma célula ou um organismo.

**Metagenoma** conjunto genético total de todas as células presentes em um determinado ambiente.

**Metagenômica** análise genômica de um *pool* de DNA ou RNA obtidos de amostras ambientais contendo organismos que ainda não foram isolados, o mesmo que genômica ambiental.

**Microarranjos** pequenos suportes sólidos aos quais genes, ou partes de genes, são fixados e arranjos espacialmente em um padrão conhecido (também denominados *chips* gênicos).

**Ortólogo** gene encontrado em um organismo, similar àquele de outro

organismo por causa da descendência de um ancestral comum (ver também *Parálogo*).

**Pangenoma** totalidade de genes presentes nas diferentes amostras de uma determinada espécie.

**Parálogo** gene cuja similaridade com um ou mais genes no mesmo organismo é resultado de duplicação gênica (ver também *Ortólogo*).

**Proteoma** conjunto total de proteínas codificadas por um genoma ou conjunto total de proteínas de um organismo.

**Proteômica** estudo em escala genômica da estrutura, função e regulação das proteínas de um organismo.

**Sequenciamento** dedução da sequência de uma molécula de DNA ou RNA, por meio de uma série de reações químicas.

**Sequenciamento *shotgun*** sequenciamento do DNA de pequenos fragmentos do genoma previamente clonados de forma aleatória; o sequenciamento *shotgun* é seguido por métodos computacionais para reconstruir a sequência genômica inteira.

**Sonda de ácido nucleico** fita de ácido nucleico marcada que pode ser utilizada para hibridizar com fitas de ácido nucleico complementares em uma solução.

**Transcriptoma** conjunto de todos os RNAs produzidos em um organismo sob determinadas condições específicas.

**Transferência horizontal de genes** transferência da informação genética entre organismos, diferente da herança vertical a partir de organismo(s) parental(ais).

## QUESTÕES PARA REVISÃO

- Por que os didesoxinucleotídeos atuam como terminadores de cadeia? (Seção 6.1)
- Dê um exemplo do sistema de sequenciamento de primeira, segunda e terceira gerações. (Seção 6.2)
- Quais características são utilizadas na identificação de fases de leitura aberta, utilizando-se dados obtidos do sequenciamento? (Seção 6.3)
- Qual é a relação entre o tamanho do genoma e o conteúdo de fases de leitura aberta no genoma dos procariotos? (Seção 6.4)
- Em relação à proporção do genoma total, qual classe de genes predomina em organismos com genomas pequenos? E naqueles organismos com genomas grandes? (Seção 6.4)
- Quais genomas são maiores: de cloroplastos ou de mitocôndrias? Descreva uma característica incomum de um genoma de cloroplasto e de um mitocondrial. (Seção 6.5)
- Quão maior é o seu genoma em relação ao da levedura? Quantos genes você possui a mais que uma levedura? (Seção 6.6)
- Diferencie os termos genoma, proteoma e transcriptoma. (Seções 6.7 e 6.8)
- O que um gel 2D de proteínas revela? Como os resultados desse tipo de gel podem ser relacionados com a função da proteína? (Seção 6.8)
- Por que a pesquisa do metaboloma está atrasada em relação ao do proteoma? (Seção 6.9)
- Quais são os objetivos da biologia de sistemas? (Seção 6.9)
- Como a expressão gênica pode ser quantificada em bactérias não cultivadas? (Seção 6.10)
- A maior parte da informação genética em nosso planeta não pertence a organismos celulares. Discuta. (Seção 6.10)

- Qual a principal diferença entre a evolução dos genomas procarióticos e eucarióticos em relação à contribuição das duplicações? (Seção 6.11)
- Explique como genes transferidos horizontalmente podem ser detectados em um genoma. (Seção 6.12)
- Explique como os elementos transponíveis promovem a evolução do genoma de bactérias. (Seção 6.12)
- Explique o modo pelo qual se pode esperar que ilhas cromossômicas movam-se entre diferentes hospedeiros bacterianos. (Seção 6.13)
- O que são as ilhas de patogenicidades e por que elas são importantes? (Seção 6.13)

## QUESTÕES APLICADAS

- Além do tamanho do genoma, quais fatores tornam a montagem completa do genoma eucariótico mais difícil do que a montagem do genoma procariótico?
- Descreva como se pode determinar quais proteínas de *Escherichia coli* são reprimidas quando a cultura é transferida de um meio mínimo (que contém somente uma única fonte de carbono) para um meio rico, contendo uma grande quantidade de aminoácidos, bases e vitaminas. Descreva como poderíamos estudar quais genes são expressos durante cada condição de crescimento.
- O gene que codifica a subunidade beta da RNA-polimerase de *Escherichia coli* é considerado ortólogo do gene *rpoB* de *Bacillus subtilis*. O que tal termo significa quanto à relação entre os dois genes? Que proteína deve ser codificada pelo gene *rpoB* de *Bacillus subtilis*? Os genes de diferentes fatores sigma de *Escherichia coli* são parálogos. O que isso indica quanto à relação entre eles?

# 10 Genética de bactérias e arqueias

## microbiologiahoje

### Vírus arcaicos ou agentes secretos de transferência gênica

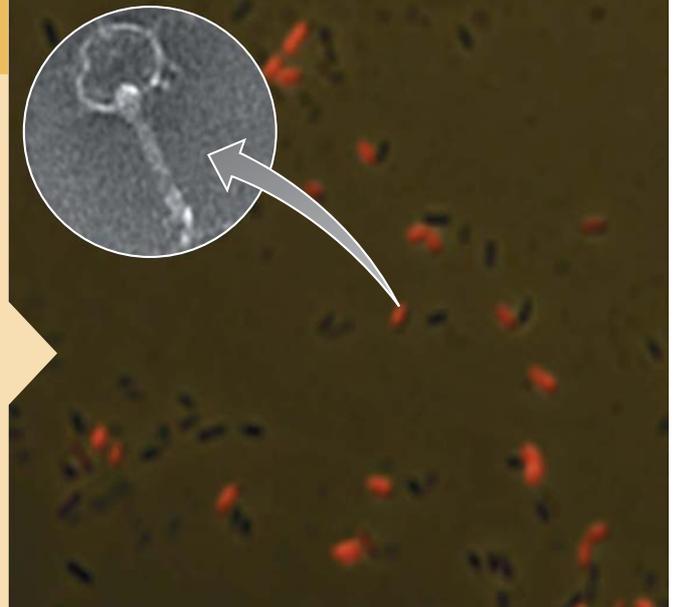
Como arqueias e bactérias adquirem características novas e excitantes que se refletem na grande diversidade do mundo microbiano? Em contraste com os organismos superiores, em procariotos ocorrem vários mecanismos de transferência horizontal de genes. Esta troca de material genético é a base para a adaptação aos diferentes nichos e desempenha um papel fundamental na evolução desses organismos.

Um exemplo da troca desse material genético ocorre por meio de agentes de transferência gênica (GTAs, *gene transfers agents*) – produtos de uma interação incomum entre vírus-hospedeiro. GTAs são o resultado de células microbianas sequestrando vírus defectivos e usando-os especificamente para a troca de DNA. GTAs assemelham-se a minúsculos bacteriófagos com cauda, (em detalhe na fotografia) e contêm pequenos fragmentos aleatórios de DNA do hospedeiro. GTAs não são considerados vírus verdadeiros por não possuírem genes codificando sua própria maquinária de produção e também por não gerarem as características placas virais ou placas de lise.

GTAs têm sido isolados a partir de uma miríade de procariotos incluindo bactérias redutoras de sulfato e arqueias produtoras de metano e são particularmente prevalentes em meio aos procariontes marinhos. Assim, é provável que GTAs sejam amplamente distribuídos na natureza. Geneticistas microbianos conseguiram determinar que um subconjunto de células de bactérias fototróficas *Rhodobacter capsulatus* produzem e liberam GTAs durante a fase estacionária do crescimento ou durante flutuações nos níveis de nutrientes<sup>1</sup>. Isso foi registrado ligando um promotor de um gene essencial para a produção de GTAs a um gene repórter que codifica para uma proteína vermelho-fluorescente; células produzindo GTA se tornaram vermelhas (foto).

Enquanto os bacteriófagos são considerados a entidade mais abundante no planeta Terra, o número destes que realmente podem ser GTAs em vez de vírus é desconhecido. GTAs podem ajudar a explicar a natureza robusta da transferência de DNA entre procariotos, especialmente entre aqueles que habitam os oceanos. Isso também implica outra questão: GTAs desempenham algum papel importante na prevalência de outro fenômeno genético comum, como a resistência bacteriana a antibióticos?

<sup>1</sup>Fogg, P.C., et al. 2012. One for all or all for one: Heterogeneous expression and host cell lysis are key to gene transfer agent activity in *Rhodobacter capsulatus*. *PLOS One* 7: e43772.



- I **Mutação 292**
- II **Transferência gênica em bactérias 299**
- III **Transferência gênica em arqueias e outros eventos genéticos 309**

Inúmeros exemplos de diversidade microbiana são descritos ao longo deste livro. Como essa diversidade surgiu? Apesar de procarionotes se reproduzirem assexuadamente, eles também possuem mecanismos para a troca de informação gênica. Essa troca de genes, juntamente com as inovações genéticas que surgem de alterações aleatórias no código genético, podem conferir uma vantagem que fundamentalmente conduz a diversidade genética.

Neste capítulo iremos discutir os mecanismos pelos quais bactérias e arqueias podem alterar o próprio genoma. Primeiro, descrevemos como surgem alterações no genoma e, então,

consideraremos como os genes podem ser transferidos de um microrganismo para outro por *transfêrencia horizontal de genes*. Enquanto a genética bacteriana é a chave para a diversidade genética e adaptação aos habitats, microrganismos também possuem mecanismos para manter a estabilidade genômica, os quais abordaremos no final desse capítulo. Em conjunto, tanto a alteração quanto a estabilidade na sequência genômica são importantes para a evolução de um organismo (ou de um vírus ou outro elemento genético) e o seu sucesso competitivo na natureza.

## I • Mutação

Todos os organismos contêm uma sequência específica de bases nucleotídicas em seu genoma, que é sua identificação. Uma **mutação** corresponde a uma alteração de natureza *herdável* na sequência de bases daquele genoma, que é passada de uma célula-mãe para sua progênie. As mutações podem promover alterações – algumas vantajosas, algumas prejudiciais, mas a maioria é neutra e sem efeitos – em um organismo. Apesar da taxa de mutação espontânea ser baixa (Seção 10.3), a velocidade em que muitos procarionotes se dividem e seu crescimento exponencial garante com surpreendente rapidez o acúmulo de mutações. Considerando que as mutações geralmente resultam apenas em uma quantidade muito pequena de alteração genética em uma célula, a recombinação genética normalmente gera alterações muito maiores. Em conjunto, mutações e recombinação são o combustível do processo evolutivo.

Começamos considerando os mecanismos moleculares das mutações e as propriedades dos microrganismos mutantes.

### 10.1 Mutações e mutantes

Em todas as células, o genoma é composto por moléculas de DNA de fita dupla. Nos vírus, ao contrário, o genoma pode consistir em DNA ou RNA de fita simples ou fita dupla. Uma linhagem de célula qualquer ou vírus que apresente uma alteração na sequência nucleotídica é denominado **mutante**. Um mutante, por definição, difere de sua linhagem parental quanto ao seu **genótipo**, a sequência nucleotídica do genoma. Além disso, as propriedades observáveis do mutante – seu **fenótipo** – podem também estar alteradas em relação à linhagem parental. Esse fenótipo alterado é denominado *fenótipo mutante*. Frequentemente, uma linhagem isolada de ambientes naturais é denominada **linhagem selvagem**. O termo “linhagem selvagem” pode ser utilizado em referência ao organismo como um todo ou apenas à condição de um gene em particular sendo estudado. Os mutantes podem ser obtidos tanto a partir de linhagens selvagens quanto de linhagens previamente derivadas das selvagens, como, por exemplo, de outro mutante.

#### Genótipo versus fenótipo

Dependendo da mutação, uma linhagem mutante pode ou não apresentar fenótipo distinto de sua linhagem parental. Por convenção, definiu-se em genética bacteriana que o *genótipo* de um organismo é designado por três letras minúsculas, seguidas por uma letra maiúscula (todas grafadas em itálico), para indicar um determinado gene. Por exemplo, o gene *hisC* de *Escherichia coli* codifica uma proteína, denominada HisC, que atua na biossíntese do aminoácido histidina. As mutações no gene *hisC* devem ser designadas como *hisC1*, *hisC2*, e assim por diante, com

os números indicando a ordem de isolamento das linhagens mutantes. Cada mutação em *hisC* deve ser diferente, cada uma podendo afetar a proteína HisC de formas distintas.

O *fenótipo* de um organismo é designado por uma letra maiúscula, seguida por duas letras minúsculas, adicionando-se um sinal de mais (+) ou de menos (–) sobrescrito, para indicar a presença ou ausência daquela propriedade. Por exemplo, uma linhagem His<sup>+</sup> de *E. coli* é capaz de sintetizar sua própria histidina, enquanto uma linhagem His<sup>–</sup> é desprovida dessa capacidade. A linhagem His<sup>–</sup> necessitará da suplementação de histidina para seu crescimento. Retornando ao exemplo anterior, qualquer mutação no gene *hisC* pode levar a um fenótipo His<sup>–</sup>, caso elimine a atividade da proteína HisC.

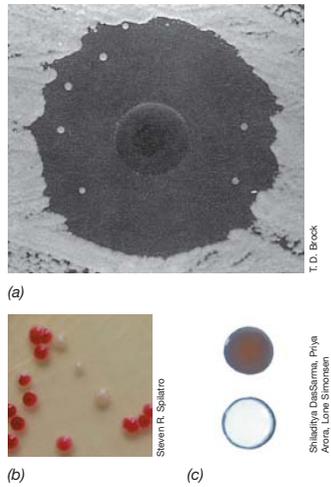
#### Isolamento de mutantes: varredura versus seleção

A princípio, qualquer característica de um organismo pode ser modificada por mutações. No entanto, algumas mutações são selecionáveis, conferindo algum tipo de vantagem ao organismo que as possui, enquanto outras mutações não são selecionáveis, mesmo que promovam claras alterações no fenótipo do organismo. Uma mutação selecionável confere uma clara vantagem à linhagem mutante, sob certas condições ambientais, de forma que a progênie da célula mutante é capaz de crescer de modo a suplantar o parental. Um bom exemplo de mutação selecionável é a resistência a fármacos: um mutante resistente a antibióticos pode crescer na presença de concentrações de antibiótico que inibiriam ou matariam a célula parental (Figura 10.1a), sendo então selecionadas nessas condições. A detecção e o isolamento de mutantes selecionáveis são relativamente simples, a partir da escolha de condições ambientais apropriadas. Portanto, a **seleção** corresponde a uma ferramenta genética extremamente poderosa que possibilita o isolamento de um único mutante a partir de uma população contendo milhões, ou mesmo bilhões, de organismos parentais.

Um exemplo de mutação não selecionável corresponde à perda de coloração em um organismo pigmentado (Figura 10.1b, c). As células despigmentadas geralmente não possuem qualquer vantagem ou desvantagem quando comparadas às células parentais, quando cultivadas em meios sólidos, embora os organismos pigmentados possam exibir alguma vantagem seletiva na natureza. Tais mutações podem ser detectadas pela simples observação visual das várias colônias, em busca daquelas com aspecto “diferente”, processo denominado **varredura**.

#### Isolamento de auxotróficos nutricionais

Embora o processo de varredura seja geralmente mais trabalhoso do que o processo de seleção, existem metodologias



**Figura 10.1** **Mutações selecionáveis e não selecionáveis.** (a) Desenvolvimento de mutantes resistentes a um antibiótico, crescendo no interior da zona de inibição, um tipo de mutação facilmente selecionável. (b) Mutações não selecionáveis. Mutantes de *Serratia marcescens* não pigmentados induzidos por radiação UV. O tipo selvagem apresenta pigmentação vermelho-escuro. Os mutantes brancos ou incolores não sintetizam pigmentos. (c) Colônias de espécies mutantes de *Halobacterium*, um membro de *Archaea*. As colônias selvagens são brancas. As colônias cor de laranja/marrons são mutantes desprovidas de vesículas de gás (↔ Seção 2.15). As vesículas de gás dispersam a luz e mascaram a coloração da colônia.

que permitem a varredura de grandes números de colônias com determinados tipos de mutações. Por exemplo, mutantes nutricionais defectivos podem ser detectados pela técnica de *plaqueamento de réplica* (Figura 10.2). Utilizando-se uma alça ou palito estéreis ou até mesmo braços robóticos, colônias podem ser coletadas a partir de uma placa matriz e inoculadas na superfície de um meio desprovido do nutriente. As colônias da linhagem parental crescerão normalmente, enquanto as

linhagens mutantes não se desenvolverão. Assim, a incapacidade de uma colônia crescer na placa réplica, contendo meio mínimo, indica que ela corresponde a um mutante. A colônia equivalente na placa matriz, a qual corresponde ao espaço vazio encontrado na placa réplica, pode então ser coletada, purificada e caracterizada. Um mutante exibindo uma necessidade nutricional para seu crescimento denomina-se **auxotrófico**, enquanto a linhagem parental da qual ele foi derivado denomina-se **prototrófica**. (Um prototrófico pode ou não corresponder ao tipo selvagem. Um auxotrófico pode ser derivado do tipo selvagem, ou ser originado de um mutante derivado do tipo selvagem.) Por exemplo, mutantes de *Escherichia coli*, com fenótipo His<sup>-</sup>, são denominados auxotróficos para histidina. Exemplos comuns de classes de mutantes e os meios pelos quais elas são detectadas estão listadas na Tabela 10.1.

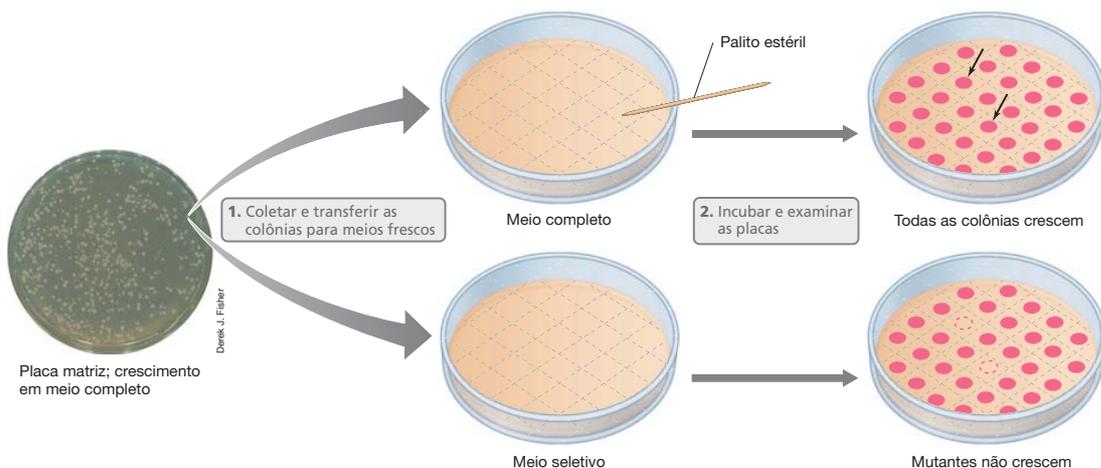
**MINIQUESTIONÁRIO**

- Diferencie os termos “mutação” e “mutante”.
- Diferencie os termos “varredura” e “seleção”.

**10.2 Bases moleculares das mutações**

As mutações podem ser espontâneas ou induzidas. As **mutações induzidas** são aquelas realizadas deliberadamente. Elas ocorrem sem qualquer intervenção humana. Podem decorrer da exposição à radiação natural (raios cósmicos, p. ex.), que promove alterações na estrutura das bases do DNA. Além disso, radicais de oxigênio (↔ Seção 5.16) podem afetar a estrutura do DNA por meio de modificações químicas da molécula. Por exemplo, radicais de oxigênio podem converter a guanina em 8-hidroxiguanina, o que provoca mutações. As mutações espontâneas são aquelas que ocorrem independentemente de uma intervenção externa. A maior parte dessas mutações é decorrente de erros ocasionais no pareamento de bases pela DNA-polimerase durante a replicação do DNA.

Mutações que alteram somente um par de bases são denominadas **mutações pontuais**. As mutações pontuais podem



**Figura 10.2** **Varredura de auxotróficos nutricionais.** O método de plaqueamento pode ser usado para a detecção de mutantes nutricionais. Colônias da placa matriz são transferidas para uma placa réplica contendo um meio diferente para a seleção. As colônias que não se desenvolveram na placa répli-

ca estão indicadas por setas. O meio de seleção é desprovido de um nutriente (leucina), presente na placa matriz. Assim, as colônias assinaladas com uma seta na placa matriz correspondem a auxotróficos para leucina.

**Tabela 10.1** Tipos de mutantes

Fenótipo	Natureza da alteração	Deteção do mutante
Auxotrofia	Perda de uma enzima de uma via biossintética	Incapacidade de crescer no meio desprovido do nutriente
Termossensibilidade	Alteração de uma proteína essencial, tornando-se mais termossensível	Incapacidade de crescer em temperaturas elevadas (p. ex., 40°C), que normalmente permitiam o crescimento
Sensibilidade ao frio	Alteração de uma proteína essencial, que passa a ser inativada em temperaturas baixas	Incapacidade de crescer em temperaturas baixas (p. ex., 20°C), que normalmente permitiam o crescimento
Resistência a um fármaco	Detoxificação do fármaco, modificação do alvo do fármaco, ou permeabilidade ao fármaco	Crescimento em meio contendo uma concentração normalmente inibidora do fármaco
Colônia rugosa	Perda ou modificação da camada lipopolissacarídica	Colônias granulosas, irregulares, em vez de colônias lisas e brilhantes
Ausência de cápsula	Perda ou modificação da cápsula	Colônias pequenas e rugosas, em vez de colônias lisas e grandes
Imobilidade	Perda dos flagelos; flagelos não funcionais	Colônias compactas, em vez de colônias achatadas, com irradiação
Despigmentação	Perda de enzimas da via biossintética, levando à perda de um ou mais pigmentos	Desenvolvimento de novas cores, ou ausência de coloração
Fermentação de açúcar	Perda de enzimas de vias degradativas	Alteração ou perda de cor em meios contendo açúcar e um indicador de pH
Resistência aos vírus	Perda do receptor viral	Crescimento na presença de grandes quantidades de vírus

ser provocadas por substituições de pares de bases no DNA, ou pela perda ou pelo ganho de um único par de bases. Assim como em todas as mutações, a alteração fenotípica decorrente de uma mutação pontual depende do local exato do gene onde ocorreu a mutação, de qual nucleotídeo foi alterado e qual o produto codificado por aquele gene.

**Substituições de pares de bases**

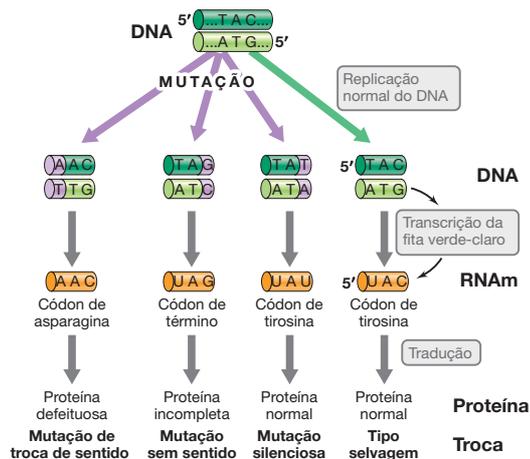
Quando uma mutação pontual ocorre em uma região codificadora de um gene que codifica um polipeptídeo, qualquer alteração fenotípica apresentada pela célula será, quase que certamente, resultante de uma alteração na sequência de aminoácidos do polipeptídeo. O erro introduzido no DNA é transcrito em RNAm e essa molécula defeituosa é, por sua vez, traduzida, originando um polipeptídeo. A Figura 10.3 ilustra as consequências de várias substituições de pares de bases.

Quando os resultados de uma mutação são interpretados, o fato de o código genético ser degenerado deve sempre ser levado em consideração (↔ Seção 4.11 e Tabela 4.5). Devido a essa degeneração, nem todas as mutações na sequência de bases que codifica um polipeptídeo o modificarão. Essa afirmação é ilustrada na Figura 10.3, que apresenta os vários resultados possíveis quando o DNA que codifica um único códon de tirosina em um polipeptídeo sofre uma mutação. Inicialmente, uma alteração no RNA de UAC para UAU não exibe qualquer efeito aparente, uma vez que UAU também corresponde a um códon de tirosina. Embora essas mutações não afetem a sequência do polipeptídeo codificado, essas alterações no DNA correspondem, de fato, a mutações. Essas são um tipo de **mutações silenciosas**, isto é, mutações que não afetam o fenótipo da célula. Observe que mutações silenciosas em regiões codificadoras geralmente estão localizadas na terceira base do códon (os códons de arginina e leucina podem também sofrer mutações silenciosas na primeira posição).

Alterações na primeira ou segunda base de um códon frequentemente promovem alterações significativas no polipeptídeo. Por exemplo, a alteração de uma única base, de UAC para AAC (Figura 10.3), resulta em uma substituição de um aminoácido no polipeptídeo, de tirosina para asparagina, em um sítio

específico. Esse tipo de alteração recebe a denominação **mutação de troca de sentido** (do inglês, *missense mutation*), porque o “sentido” informacional (a precisa sequência de aminoácidos) no polipeptídeo gerado foi alterado. Se a alteração ocorrer em algum ponto crítico da cadeia polipeptídica, a proteína poderá ser sintetizada de forma inativa ou exibir atividade reduzida. Entretanto, nem todas as mutações de sentido trocado necessariamente originam proteínas não funcionais. O resultado final depende da região do polipeptídeo que sofreu a substituição, e como tal alteração está afetando o dobramento e a atividade da proteína. Por exemplo, mutações no sítio ativo de uma enzima têm maior probabilidade de abolir a atividade que mutações que ocorrem em outras regiões da proteína.

Outro resultado possível de uma substituição de um par de bases é a criação de um códon sem sentido (de término). Esse evento resulta em um término prematuro da tradução, originando um polipeptídeo incompleto, provavelmente não funcio-



**Figura 10.3** Possíveis efeitos de substituições de pares de bases em um gene que codifica uma proteína. Três produtos proteicos diferentes podem ser originados a partir de alterações no DNA de um único códon.

nal (Figura 10.3). Mutações desse tipo são denominadas **mutações sem sentido**, devido à troca de códon de um aminoácido (com significado) por um códon sem sentido (↔ Tabela 4.5). A menos que a mutação sem sentido ocorra próxima ao final de um gene, o produto da tradução é considerado truncado ou incompleto. Proteínas truncadas são completamente inativas ou, na melhor das hipóteses, não possuem atividade normal.

Os termos “transição” e “transversão” são utilizados para desenvolver o tipo de substituição de bases em uma mutação pontual. **Transições** são mutações nas quais uma purina (A ou G) é substituída por outra purina, ou uma pirimidina (C ou T) é substituída por outra pirimidina. **Transversões** são mutações nas quais uma purina é substituída por uma pirimidina ou vice-versa.

**Alterações de fase de leitura e outras inserções ou deleções**

Uma vez que o código genético é lido a partir de uma das extremidades da molécula de ácido nucleico, em blocos consecutivos de três bases (i.e., como códons), qualquer deleção ou inserção de um único par de bases resulta em uma alteração da fase de leitura. Essas **mutações de alteração da fase de leitura** podem acarretar sérias consequências. Inserções ou deleções de uma base alteram a sequência primária do polipeptídeo codificado, normalmente de forma significativa (Figura 10.4). Tais microinserções ou microdeleções podem ser resultantes de erros ocorridos durante a replicação. A inserção ou deleção de dois pares de bases também provoca uma alteração de fase; contudo, a inserção ou deleção de três pares de bases adiciona ou remove um códon inteiro. Tal evento resulta na inserção ou deleção de um único aminoácido na sequência polipeptídica. Embora isso possa também ser deletério à atividade da proteína, geralmente não é tão danoso quanto uma alteração de fase, que embaralha toda a sequência polipeptídica após o local da mutação.

Inserções ou deleções podem também resultar na aquisição ou perda de centenas, ou mesmo milhares, de pares de bases. Tais alterações inevitavelmente resultam na perda completa da função gênica. Algumas deleções são tão extensas que podem incluir vários genes. Caso algum dos genes deletados seja essencial, a mutação será letal. Tais deleções não são restauradas por meio de mutações adicionais, mas apenas por eventos de recombinação genética. De fato, uma maneira de se distinguir grandes

deleções de mutações pontuais é o fato de as últimas serem revertidas por mutações adicionais, enquanto as primeiras não o são. Inserções e deleções ainda maiores podem ocorrer devido a erros durante a recombinação genética. Além disso, muitas mutações por inserção são decorrentes da inserção de sequências específicas e identificáveis de DNA denominadas sequências de inserção, um tipo de *elemento transponível* (Seção 10.11). O efeito dos elementos transponíveis na evolução dos genomas bacterianos é discutido em maior detalhe na Seção 6.12.

**MINIQUESTIONÁRIO**

- Por que mutações de alteração da fase de leitura, geralmente, têm consequências mais graves do que as mutações de troca do sentido?
- Mutações de troca de sentido podem ocorrer em genes que codificam RNAt? Por quê?

**10.3 Reversões e taxas de mutação**

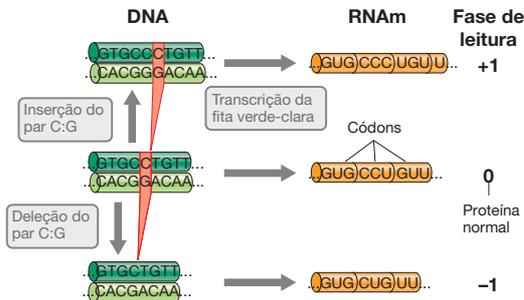
As taxas às quais ocorrem diferentes tipos de mutações variam amplamente. Alguns tipos de mutações ocorrem tão raramente que elas são quase impossíveis de detectar, enquanto outras ocorrem com tanta frequência que apresentam dificuldades para qualquer experimentador tentando manter geneticamente estável uma cultura estoque. Às vezes, uma segunda mutação pode inverter o efeito da mutação inicial. Além disso, todos os organismos possuem uma variedade de mecanismos para o reparo do DNA. Consequentemente, a taxa de mutação observada não depende somente da frequência de alteração na sequência de DNA, mas também sobre a eficiência em que ocorre o reparo do DNA mutado.

**Mutações reversas ou reversões**

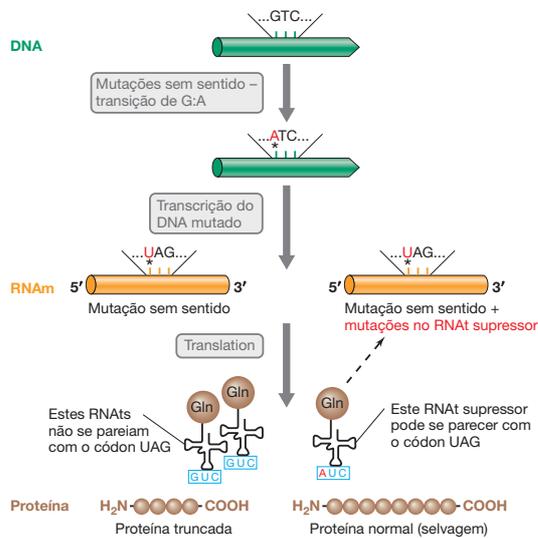
As mutações pontuais são normalmente reversíveis, um processo denominado **reversão**. Uma linhagem revertente é aquela em que o fenótipo original perdido pelo mutante é restaurado. Os revertentes podem ser classificados em dois tipos. Nos **revertentes de mesmo sítio**, a mutação que restaura a atividade ocorre no mesmo sítio onde a mutação original ocorreu. Caso a mutação reversa não ocorra apenas no mesmo sítio, mas também restaure a sequência selvagem, será denominada *revertente verdadeira*.

Nos *revertentes de segundo sítio*, a mutação ocorre em um sítio distinto no DNA. As mutações de segundo sítio podem levar à restauração do fenótipo selvagem se atuarem como *mutações supressoras* – mutações que compensam o efeito da mutação original e restauram o fenótipo original. Várias classes de mutações supressoras são conhecidas. Elas incluem (1) uma mutação que ocorre em algum outro local do gene e restaura a função de uma enzima, como uma segunda mutação de alteração de fase próxima à primeira, restaurando a fase original de leitura, (2) uma mutação em outro gene que restaura a função do gene original mutado, e (3) uma mutação em outro gene, que resulta na produção de uma enzima capaz de substituir a enzima mutada.

Uma interessante subclasse de mutações supressoras compreende as mutações que ocorrem devido a alterações no RNAt. Mutações sem sentido podem ser suprimidas alterando-se a sequência do anticódon na molécula de RNAt, de modo que ela passe a reconhecer um códon de término (Figura 10.5). Tal RNAt modificado é conhecido como *RNAt supressor* e irá inserir seu aminoácido cognato no códon de término, que agora passa a ter significado. Mutações no RNAt



**Figura 10.4** Alterações na fase de leitura de um RNAm, provocadas por inserções ou deleções. A fase de leitura no RNAm é estabelecida pelo ribossomo, que inicia a leitura na extremidade 5' (à esquerda da figura), movendo-se ao longo de unidades compostas por três bases. A fase normal de leitura é referida como fase zero (0), enquanto aquela onde houve a deleção de uma base é denominada -1, e aquela onde uma base foi inserida, de +1.



**Figura 10.5** Supressão de mutações sem sentido. Introdução de uma mutação sem sentido em um gene codificador para uma proteína resulta na incorporação de um códon de término (indicado pelo \*) no RNAm correspondente. Uma única mutação leva à produção de um polipeptídeo truncado. A mutação é suprimida se uma segunda mutação ocorrer no anticódon de um RNAt, um RNAt carregando glutamina neste exemplo, o qual permite o RNAt mutado ou o supressor do RNAt ligar-se ao códon sem sentido.

supressor podem ser letais caso a célula não possua mais de um RNAt para um determinado códon. Um RNAt pode então ser mutado em um supressor, enquanto o outro realiza a função original. A maioria das células possui múltiplos RNAt, de modo que as mutações supressoras são relativamente comuns, pelo menos em microrganismos. Algumas vezes, o aminoácido inserido pelo RNAt supressor é idêntico ao aminoácido original, sendo a proteína completamente restaurada. Em outros casos, um aminoácido diferente é inserido, podendo haver a síntese de uma proteína parcialmente ativa.

### Teste de Ames

O teste de Ames (nomeado em homenagem a Bruce Ames, o bioquímico que desenvolveu o teste) torna prático o uso da detecção de revertentes em grandes populações de bactérias mutantes para testar mutagenicidade de produtos químicos potencialmente perigosos. O procedimento padrão para testar-se a mutagenicidade de compostos químicos é a observação de um possível aumento na taxa de mutações reversas (reversões) em linhagens bacterianas auxotróficas, na presença do agente mutagênico suspeito (Figura 10.6). O teste de Ames avalia a ocorrência de reversões, em vez de mutações primárias (gerando mutantes autotróficos a partir da linhagem selvagem), uma vez que os revertentes podem ser mais facilmente relacionados.

É importante que a linhagem auxotrófica possua uma mutação pontual, pois a taxa de reversão em tal tipo de linhagem é mensurável. Células desse tipo de auxotrófico não crescem em um meio desprovido do nutriente requerido (p. ex., um aminoácido), e mesmo populações muito numerosas podem ser semeadas na placa, sem a formação de colônias visíveis. Entretanto, havendo células com mutações reversas (revertentes), elas serão capazes de originar colônias. Assim, se  $10^8$

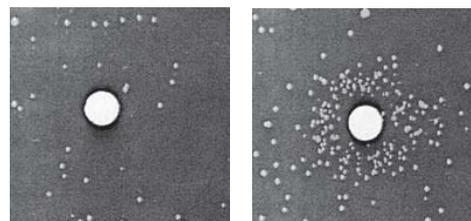
células forem espalhadas na superfície de uma única placa, cerca de 10 a 20 revertentes poderão ser detectados, a partir da visualização das 10 a 20 colônias formadas (Figura 10.6, fotografia à esquerda). No entanto, se a taxa de reversão for *aumentada* pela presença de um mutagênio químico, o número de revertente de colônias é ainda maior. Após incubação durante a noite, a mutagenicidade do composto pode ser detectada por um halo de mutações reversas na área ao redor do disco de papel (Figura 10.6).

Uma ampla variedade de produtos químicos foi submetida ao teste de Ames, e esse tornou-se uma das mais úteis técnicas de varredura para determinar o potencial de mutagenicidade de um composto. Uma vez que alguns mutagênicos podem causar câncer em animais, o teste de Ames também funciona como uma triagem de possíveis carcinógenos.

### Taxas de mutação

Para a maioria dos microrganismos, a frequência de erros durante a replicação do DNA varia de  $10^{-6}$  a  $10^{-7}$  por par de quilibases, durante um único ciclo de replicação. Um gene típico apresenta cerca de 1.000 pares de bases. Assim, a frequência de uma mutação em um determinado gene situa-se na faixa de  $10^{-6}$  a  $10^{-7}$ , por geração. Por exemplo, em uma cultura bacteriana contendo cerca de  $10^8$  células/mL, vários mutantes diferentes para um dado gene poderão, provavelmente, ser encontrados em cada mililitro dessa cultura. Organismos superiores, com genomas muito grandes, tendem a apresentar taxas de erros de replicação cerca de dez vezes inferiores às das bactérias, enquanto vírus de DNA, especialmente aqueles com genomas muito pequenos, podem apresentar taxas de erros 100 a 1.000 vezes superiores às das de organismos celulares. Os vírus de RNA exibem taxas de erros ainda maiores devido a menor taxa de revisão (↔ Seção 4.6) e a falta de mecanismos de reparo de RNA.

Erros em bases únicas durante a replicação do DNA geralmente resultam em mutações de sentido trocado, em vez de mutações sem sentido, uma vez que as substituições de uma única base geram códon que especificam outros aminoácidos (↔ Tabela 4.5). O segundo tipo mais frequente de alteração



**Figura 10.6** Teste de Ames para avaliar a mutagenicidade de um composto químico. As duas placas foram inoculadas com uma cultura de um mutante de *Salmonella enterica* que precisa de histidina para seu crescimento. O meio de cultura não contém histidina, de modo que apenas aquelas células que revertem ao estado selvagem são capazes de crescer. Revertentes espontâneos desenvolvem-se em ambas as placas, porém o agente químico presente no disco de papel-filtro na placa teste (à direita) promoveu um aumento na taxa de mutação, evidenciado pelo grande número de colônias ao redor do disco. Revertentes não são observados muito próximos ao disco de teste, uma vez que a concentração do agente mutagênico é muito elevada, tornando-o letal. A placa à esquerda corresponde ao controle negativo; seu disco de papel-filtro foi adicionado apenas de água.

de códon, provocada por uma alteração em uma base, provoca uma mutação silenciosa. Tal evento ocorre porque a maioria dos códons alternativos para um determinado aminoácido difere apenas em uma modificação de uma base, na terceira posição “silenciosa”. Um determinado códon pode ser modificado, originando qualquer um dos 27 demais códons por uma substituição de uma única base e, em média, cerca de duas delas corresponderão a mutações silenciosas, uma corresponderá a uma mutação sem sentido, e o restante, a mutações de sentido trocado. Há também algumas sequências de DNA, normalmente regiões contendo repetições curtas, que correspondem a *hot spots* (“pontos quentes”) de mutações, pois a frequência de erros da DNA-polimerase é relativamente alta nessas locais. A taxa de erros em *hot spots* é afetada pela sequência de bases das regiões vizinhas.

A menos que um mutante seja selecionável, a detecção experimental desses eventos é difícil, sendo necessária muita habilidade do geneticista microbiano no sentido de aumentar a eficiência de detecção dessas mutações. Como veremos na próxima seção, é possível aumentar significativamente a taxa de mutação pelo emprego de tratamentos mutagênicos. Além disso, a taxa de mutação pode ser alterada em certas situações, como quando sob condições de estresse.

**MINIQUESTIONÁRIO**

- Por que o teste de Ames mede a taxa de reversão em vez da taxa de mutação para a frente?
- Qual a classe de mutação, de troca de sentido ou sem sentido, é mais comum, e por quê?

**10.4 Mutagênese**

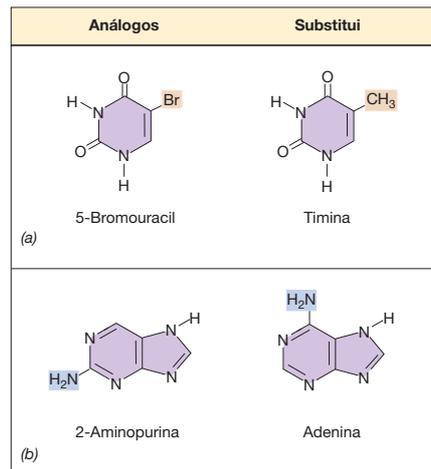
A taxa de mutações espontâneas é muito baixa; no entanto, existem vários agentes químicos, físicos ou biológicos que podem aumentar a frequência de mutações, sendo, portanto, referidos como agentes capazes de induzir mutações. Esses agentes são denominados **mutagênicos**. Passaremos a discutir agora algumas das principais categorias de agentes mutagênicos e suas atividades.

**Agentes químicos mutagênicos**

Uma visão geral dos principais agentes químicos mutagênicos e de seus mecanismos de ação é apresentada na **Tabela 10.2**. Existem várias classes de agentes mutagênicos químicos. Uma das classes corresponde aos *análogos de bases nucleotídicas*, moléculas estruturalmente semelhantes às bases púricas e pirimídicas de DNA, promovendo o pareamento incorreto das bases (**Figura 10.7**). Quando um desses análogos de bases é incorporado ao DNA no lugar de uma base natural, a replicação do DNA pode ocorrer normalmente, na maioria dos casos. Todavia, erros na replicação do DNA ocorrem com maior frequência nesses sítios, devido à incorporação de uma base incorreta na fita de DNA, introduzindo, assim, uma mutação. Durante o processo subsequente de segregação desta fita na divisão celular, a mutação é revelada.

Outros agentes químicos mutagênicos podem induzir *modificações químicas* em uma ou outra base, resultando em um pareamento incorreto ou outras modificações relacionadas (Tabela 10.2). Por exemplo, agentes alquilantes (compostos químicos que reagem com grupos amino, carboxil e hidroxil

<b>Tabela 10.2 Agentes químicos e físicos mutagênicos e seus mecanismos de ação</b>		
<i>Agente</i>	<i>Ação</i>	<i>Resultado</i>
<b>Análogos de bases</b>		
5-Bromouracil	Incorporado como T; ocasional pareamento incorreto com G	AT → GC e ocasionalmente GC → AT
2-Aminopurina	Incorporado como A; ocasional pareamento incorreto com C	AT → GC e ocasionalmente GC → AT
<b>Agentes químicos que reagem com o DNA</b>		
Ácido nitroso (HNO <sub>2</sub> )	Desamina A e C	AT → GC e GC → AT
Hidroxilamina (NH <sub>2</sub> OH)	Reage com C	GC → AT
<b>Agentes alquilantes</b>		
Monofuncionais (p. ex., etil metano sulfonato)	Adiciona um grupo metil em G; pareamento incorreto com T	GC → AT
Bifuncionais (p. ex., mitomicina, gás mostarda, nitrosoguanidina)	Faz ligações cruzadas das fitas de DNA, regiões incorretas excisadas pela DNase	Mutações pontuais e deleções
<b>Corantes intercalantes</b>		
Acridinas, brometo de etídio	Insere-se entre dois pares de bases	Microinserções e microdeleções
<b>Radiação</b>		
Ultravioleta	Forma de dímeros de pirimidina	O reparo pode levar a erros ou deleções
Radiação ionizante (p. ex., raios X)	Radicais livres atacam o DNA, quebrando as cadeias	O reparo pode levar a erros ou deleções



**Figura 10.7 Análogos de bases nucleotídicas.** Estrutura de dois análogos de bases comuns, utilizados na indução de mutações, e as bases normalmente encontradas em ácidos nucleicos, que são substituídas. (a) O 5-bromouracil pode parear-se com a guanina, promovendo substituições de AT por GC. (b) A 2-aminopurina pode parear-se com a citosina, promovendo substituições de AT por GC.

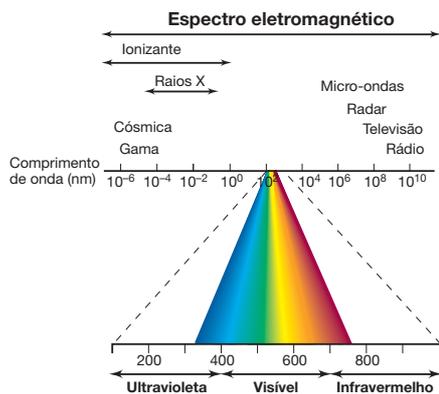
de proteínas e ácidos nucleicos, substituindo-os por grupos alquil), como a nitrosoguanidina, são agentes mutagênicos poderosos que geralmente induzem mutações em frequências superiores às induzidas pelos análogos de bases. Diferentemente dos análogos de bases, que exibem efeito somente quando incorporados durante a replicação do DNA, os agentes alquilantes são capazes de introduzir modificações mesmo em moléculas de DNA que não se encontram em replicação. Tanto os análogos de bases quanto os agentes alquilantes tendem a induzir substituições de pares de bases (Seção 10.2).

Outro grupo de agentes químicos mutagênicos, as acridinas, corresponde a moléculas planares que atuam como *agentes intercalantes*. Esses agentes mutagênicos inserem-se entre dois pares de bases no DNA, afastando-os. Durante a replicação, essa conformação anormal pode levar a inserções ou deleções em moléculas de DNA tratadas com acridina. Dessa forma, as acridinas normalmente induzem mutações de alteração da fase de leitura (Seção 10.2). O brometo de etídio, frequentemente utilizado na detecção de DNA em procedimentos de eletroforese, é também um agente intercalante e, portanto, um agente mutagênico.

### Radiação

Várias formas de radiação são altamente mutagênicas. As radiações eletromagnéticas mutagênicas podem ser divididas em duas categorias principais, *ionizante* e *não ionizante* (Figura 10.8). Embora os dois tipos de radiação sejam empregados em genética microbiana para gerar mutações, as radiações não ionizantes, como a radiação ultravioleta (UV), são mais amplamente utilizadas.

As bases púricas e pirimídicas dos ácidos nucleicos absorvem fortemente a radiação UV, sendo o máximo de absorção para DNA e RNA 260 nm. A morte celular pela radiação UV deve-se, principalmente, à sua ação sobre o DNA. Embora vários efeitos sejam conhecidos, um efeito bem estabelecido corresponde à formação de *dímeros de pirimidina*, em que duas bases pirimídicas adjacentes (citossina ou timina) de uma mesma fita ligam-se covalentemente. Isso resulta no impedimento de a DNA-polimerase ou em uma maior probabilidade de a DNA-polimerase realizar erros na leitura da sequência nesse local.



**Figura 10.8** Comprimentos de onda das radiações. Observe que a radiação ultravioleta consiste em comprimentos de onda imediatamente inferiores aos da luz visível. No caso de qualquer radiação eletromagnética, quanto menor o comprimento de onda, maior a energia. O DNA absorve intensamente em 260 nm.

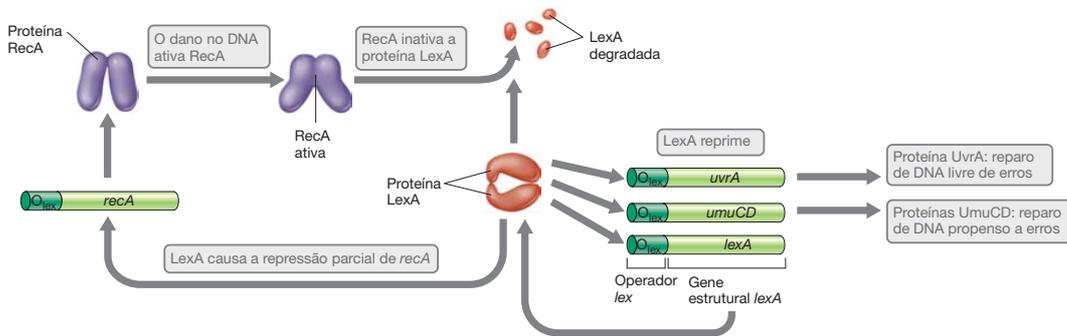
A radiação ionizante é uma forma de radiação mais poderosa do que a radiação UV e inclui os raios de pequeno comprimento de onda, como os raios X, raios cósmicos e raios gama (Figura 10.8). Esses raios provocam a ionização da água e de outras substâncias como o radical hidroxila, OH· (↔ Seção 5.16). Radicais livres reagem com e danificam macromoléculas na célula, incluindo o DNA. Isso causa a quebra nos filamentos de cadeia simples ou dupla que pode levar a grandes deleções ou rearranjos no DNA. Em baixas doses de radiação ionizante, apenas algumas modificações ocorrem no DNA, mas em doses elevadas, múltiplas modificações causam a fragmentação das fitas de DNA que às vezes não podem ser reparadas e, assim, levam à morte da célula.

### Sistemas de reparo de DNA

Lembre-se de que, por definição, uma mutação é uma alteração *hereditária* do material genético. Assim, se um DNA danificado puder ser corrigido antes da divisão celular, não haverá mutação. A maioria das células apresenta uma série de processos distintos para o reparo de DNA, seja para corrigir erros, seja reparar danos. A maioria desses sistemas de reparo de DNA é livre de erros. Entretanto, alguns processos são propensos a erros, introduzindo mutações decorrentes do próprio processo de reparo. Alguns danos no DNA, especialmente danos em larga escala decorrentes de agentes químicos altamente mutagênicos, ou de grandes doses de radiação, podem causar lesões que interferem na replicação. Se tais lesões não forem removidas há tempo, o processo de replicação do DNA irá parar, podendo resultar em quebras letais nos cromossomos. Alguns tipos de dano ao DNA e a interrupção no processo de replicação ativam o **sistema de reparo SOS**. O sistema SOS inicia uma série de processos de reparo, alguns dos quais são livres de erros. Contudo, o sistema SOS também permite o reparo sem a necessidade de um DNA-molde, isto é, sem o pareamento de bases por meio da incorporação aleatória de dNTPs. Assim, conforme o esperado, tal processo resulta na introdução de muitos erros e, com isso, muitas mutações. No entanto, as mutações podem ser menos prejudiciais para a sobrevivência da célula do que a quebra no cromossomo, uma vez que as mutações muitas vezes podem ser corrigidas, enquanto quebras cromossômicas normalmente não podem.

Em *Escherichia coli*, o sistema de reparo SOS regula a transcrição de aproximadamente 40 genes localizados ao longo de todo o cromossomo, e participam da tolerância ao dano no DNA e também no reparo. Na tolerância ao dano no DNA, as lesões permanecem no filamento de DNA, mas são ignoradas por DNA-polimerases especializadas que deslocam-se passando através do dano, um processo denominado síntese translesão. Mesmo na ausência de um molde para permitir a inserção das bases corretas, é menos perigoso preencher as lacunas do que deixar um DNA com quebras ou danificado naquele local. Consequentemente, a síntese translesão gera muitos erros. Em *E. coli*, em que os processos de mutagênese foram detalhadamente estudados, existem duas polimerases de reparo propensas a erro, a DNA-polimerase V, um enzima codificada pelos genes *umuCD* (Figura 10.9), e a DNA-polimerase IV, codificada pelo gene *dinB*. Ambas são induzidas como parte da resposta SOS.

O sistema SOS é regulado por duas proteínas, LexA e RecA. LexA é uma proteína repressora que normalmente impede a expressão do sistema SOS. A proteína RecA, que nor-



**Figura 10.9 Mecanismo da resposta SOS.** Danos no DNA ativam a proteína RecA, que, por sua vez, ativa a atividade de protease de LexA. A proteína LexA normalmente reprime as atividades do gene *recA* e dos genes *umrA* e *umuCD* (as proteínas UmuCD são parte da DNA-polimerase V), envolvidos no

reparo de DNA. No entanto, a repressão não é total. Algumas moléculas da proteína RecA são produzidas, mesmo na presença da proteína LexA. Quando a proteína LexA é inativada, esses genes tornam-se altamente ativos.

malmente participa da recombinação genética (Seção 10.5), é ativada pela ocorrência de danos ao DNA (Figura 10.9). A forma ativada de RecA estimula LexA a se autoinativar, por meio de autoclivagem. Isso leva à desrepressão do sistema SOS, resultando na expressão coordenada de várias proteínas que participam do reparo do DNA. Uma vez que alguns dos mecanismos de reparo do sistema SOS são inerentemente propensos a erros, muitas mutações podem surgir. Assim, uma vez que o dano no DNA foi reparado, o regulon SOS é reprimido, interrompendo a ocorrência de mutações adicionais.

### Alterações na taxa de mutação e suas consequências evolutivas

A alta fidelidade (baixa frequência de erros) na replicação do DNA é essencial para que os organismos mantenham-se estáveis geneticamente. Contudo, a fidelidade perfeita é contraproducente, uma vez que poderia impedir a evolução. Portanto, taxas de mutação existem nas células, porém são muito baixas para serem detectadas. Isso permite ao organismo equilibrar a necessidade de estabilidade genética e as adaptações evolutivas. O fato de organismos filogeneticamente distantes como arqueias e bactérias apresentarem aproximadamente a mesma taxa de mutações poderia ser interpretado como se a pressão evolutiva tivesse selecionado organismos com as menores taxas possíveis de mutação. Entretanto, esse não é o caso. Por exemplo, mutantes de alguns organismos que são hiperprecisos na replicação e reparo do DNA foram selecionados em laboratório. Contudo, os mecanismos otimizados de revisão ou reparo nesses organismos podem ter um custo metabólico significativo; assim, os mutantes hiperprecisos poderiam encontrar-se em desvantagem no ambiente natural.

Em contrapartida, certos organismos podem beneficiar-se desses sistemas otimizados de reparo, permitindo-lhes

ocupar certos nichos na natureza. Por exemplo, a subunidade proteica da DNA-polimerase III envolvida nas atividades de revisão (↔ Seção 4.6) é codificada pelo gene *dnaQ*. Certas mutações nesse gene originam mutantes que ainda são viáveis, mas que exibem taxas de mutação aumentadas. Esses organismos são referidos como linhagens hipermutáveis ou **linhagens mutadoras**. São conhecidas várias mutações que levam a um fenótipo mutador em vários outros sistemas de reparo de DNA. O fenótipo mutador é aparentemente selecionado em ambientes complexos e que sofrem constantes alterações, uma vez que linhagens bacterianas com fenótipos mutadores parecem ser mais abundantes nessas condições. Provavelmente, qualquer desvantagem decorrente do aumento da taxa de mutações seja contrabalançada pela capacidade de gerar um grande número de mutações novas e úteis nesse tipo de ambientes. Em última análise, essas mutações aumentam a adaptabilidade evolutiva da população, tornando o organismo mais apto ao sucesso em um nicho ecológico.

Conforme mencionado anteriormente, um fenótipo mutador pode ser induzido em linhagens selvagens por situações de estresse. Por exemplo, a resposta SOS induz ao reparo propenso a erros. Portanto, quando a resposta SOS é ativada, há um aumento na taxa de mutação. Em alguns casos, tal fenômeno é meramente um subproduto inevitável do reparo de DNA, mas, em outros, a taxa aumentada de mutação pode, por si só, apresentar valor seletivo ao organismo, em relação à sua sobrevivência.

#### MINIQUESTIONÁRIO

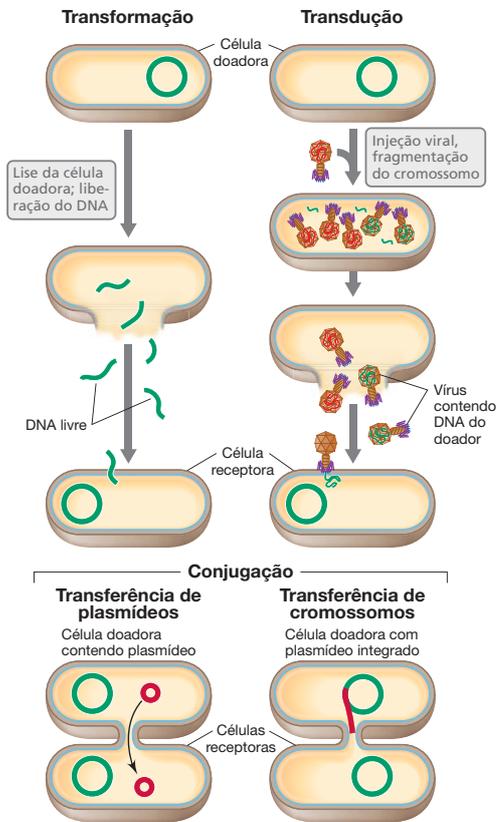
- Como os agentes mutagênicos atuam?
- Por que um fenótipo mutador pode ser útil em um ambiente sujeito a rápidas modificações?
- Qual o significado de reparo do DNA propenso a erros?

## II • Transferência gênica em bactérias

**A**nálises comparativas de genomas de microrganismos estreitamente relacionados, mas que exibem diferentes fenótipos, revelou nítidas diferenças no genoma. Muitas vezes, essas diferenças são resultados da *transferência horizontal de genes*, o movimento de genes entre células que não são des-

cententes diretas (↔ Seção 6.12). A transferência horizontal de genes permite que as células adquiram rapidamente novas características e guia a diversidade metabólica.

Três mecanismos de troca genética são conhecidos em procariotos: (1) *transformação*, em que o DNA livre é liberado



**Figura 10.10** Processos pelos quais o DNA é transferido de uma célula bacteriana doadora para uma outra receptora. Apenas os passos iniciais na transferência são mostrados.

de uma célula, sendo captado por outra (Seção 10.6); (2) *transdução*, no qual a transferência de DNA é mediada por um vírus (Seção 10.7); e (3) *conjugação*, na qual a transferência de DNA envolve o contato célula-célula e um plasmídeo conjugativo na célula doadora (Seções 10.8 e 10.9). Esses processos são comparados na **Figura 10.10**, e deve-se notar que a transferência de DNA normalmente ocorre somente em uma direção, entre o doador e o receptor.

Antes de discutirmos os processos de transferência, devemos considerar o destino do DNA transferido. Qualquer que seja o processo de transferência (transformação, transdução ou conjugação), o DNA captado pela célula pode seguir três caminhos. (1) Ele pode ser degradado por enzimas de restrição, (2) pode ser replicado *per se* (somente se possuir sua própria origem de replicação, como um plasmídeo ou genoma de um fago), (3) ou pode sofrer recombinação com o cromossomo do hospedeiro.

### 10.5 Recombinação genética

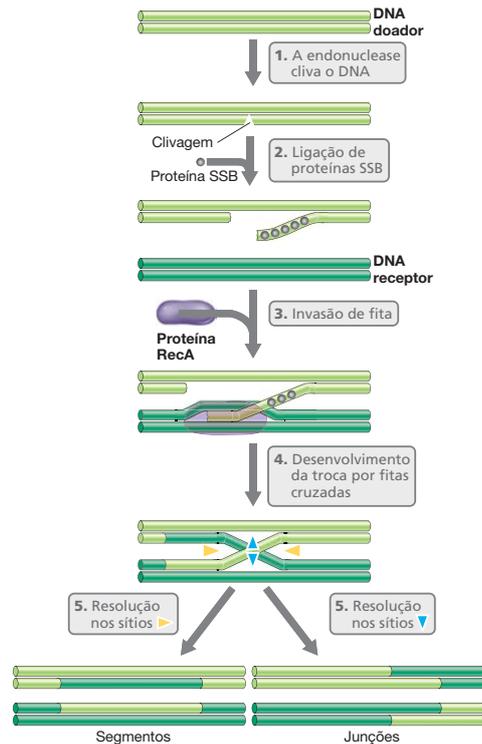
**Recombinação** corresponde à troca física de DNA entre *elementos genéticos* (estruturas que carregam informação genética). Nesta seção, enfocaremos a recombinação *homóloga*, um processo que resulta na troca de sequências homólogas de DNA, a partir de duas origens distintas. Sequências homólogas de DNA são aquelas que apresentam sequências pratica-

mente idênticas, portanto, permitem o pareamento das bases ao longo de uma extensão variável de duas moléculas de DNA. Esse tipo de recombinação está envolvido no processo denominado "*crossing over*", na genética clássica.

#### Eventos moleculares na recombinação homóloga

A proteína RecA, previamente mencionada quando discutimos o sistema SOS (Seção 10.4 e Figura 10.9), é essencial na recombinação homóloga. RecA é essencial a quase todas as vias de recombinação homóloga. Proteínas do tipo RecA foram identificadas em todos os procariontes examinados, assim como em arqueias e na maioria dos eucariontes.

Um mecanismo molecular que proporciona a recombinação homóloga entre duas moléculas de DNA está apresentado na **Figura 10.11**. Uma enzima, denominada *endonuclease*, que cliva o DNA no meio de uma fita, inicia o processo por meio da inserção de um corte em uma das fitas da molécula de DNA doadora. Essa fita cortada é separada da outra fita por proteínas que possuem atividade de helicase (↔ Seção 4.5). O segmento de fita simples resultante se liga a proteínas de ligação a fita simples (Seção 4.5) e, então, à RecA. Esse processo resulta



**Figura 10.11** Versão simplificada da recombinação homóloga. Moléculas de DNA homólogos pareiam-se e permutam porções de DNA. O mecanismo envolve a quebra e reunião das regiões pareadas. Duas das proteínas envolvidas, uma proteína de ligação a fita simples (SSB, *single strand binding*) e a proteína RecA, são apresentadas. As outras proteínas envolvidas não são ilustradas. O diagrama não se encontra em escala: o pareamento pode ocorrer ao longo de centenas ou milhares de bases. A resolução ocorre pela clivagem e reunião de moléculas cruzadas de DNA. Observe que há dois possíveis resultados, segmentos ou junções, dependendo de onde as fitas são clivadas durante o processo de resolução.

na formação de um complexo que promove o pareamento de bases com a sequência complementar da molécula de DNA recipiente. Esse pareamento de bases, por sua vez, causa a remoção da outra fita da molécula de DNA recipiente (Figura 10.11), sendo apropriadamente denominada *invasão de fita*.

A troca das fitas promove a formação de intermediários de recombinação contendo extensas regiões de **heterodúplex**, em que cada fita é oriunda de um cromossomo diferente. Finalmente, as moléculas ligadas são separadas, ou “resolvidas” por resolvasas, que clivam e unem as segundas (previamente não clivadas) fitas. Dependendo da orientação da junção durante a resolução, dois tipos de produtos, segmentos ou junções, são formados, os quais diferem em relação à conformação das regiões de heterodúplex que permanecem após a resolução (Figura 10.11).

### Efeito da recombinação homóloga no genótipo

Para que a recombinação homóloga gere novos fenótipos, é essencial que as duas sequências sejam relacionadas, porém geneticamente distintas. Esse é, obviamente, o caso em uma célula eucariótica diploide, que possui dois conjuntos de cromossomos, um oriundo de cada progenitor. Em procariotos, moléculas de DNA geneticamente distintas, porém homólogas, entram em contato por meio de diferentes mecanismos, sendo a recombinação genética um evento não menos importante. A recombinação genética em procariotos ocorre após a transferência de fragmentos de DNA homólogo de um cromossomo doador para uma célula receptora por transformação, transdução ou conjugação. Somente após a transferência, quando o fragmento de DNA da célula doadora se encontra na célula receptora, pode haver o processo de recombinação homóloga. Uma vez que, em procariotos, apenas um fragmento cromossômico é transferido, se não houver a recombinação homóloga, ele será perdido, pois não possui a capacidade de replicar-se de forma independente. Assim, em procariotos, a transferência é apenas a primeira etapa na geração de organismos recombinantes.

Para que a troca física de segmentos de DNA seja detectada, as células resultantes da recombinação devem ser fenotipicamente diferentes das células parentais (Figura 10.12). Os cruzamentos genéticos geralmente dependem de linhagens receptoras nas quais falta alguma característica selecionável, a qual será adquirida pelas células recombinantes. Por exemplo,

a célula receptora pode ser incapaz de crescer em um determinado meio, enquanto os recombinantes genéticos selecionados possuem tal capacidade. Vários tipos de marcadores selecionáveis, como resistência a fármacos e necessidades nutricionais, foram discutidos na Seção 10.1. A extrema sensibilidade do processo de seleção permite que até mesmo algumas poucas células recombinantes sejam detectadas em uma grande população de células não recombinantes, assim, a seleção é uma importante ferramenta para o geneticista microbiano.

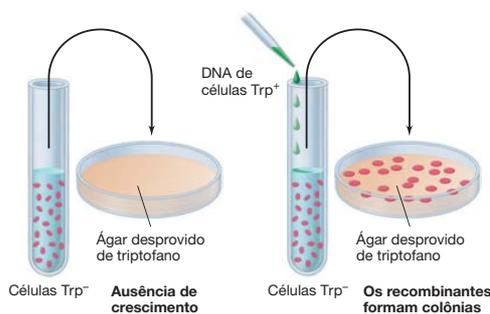
### Complementação

Em todos os três métodos de transferência de genes bacterianos, somente uma porção do cromossomo doador entra na célula receptora. Portanto, a menos que haja recombinação com o cromossomo da receptora, o DNA doador será perdido, uma vez que não é capaz de replicar-se independentemente da receptora. No entanto, é possível manter um estado de diploidia parcial de forma estável, para uso em análises de genética bacteriana. Passaremos a considerar tal tema. Uma linhagem bacteriana que carrega *duas cópias* de qualquer segmento cromossômico é conhecida como diploide parcial ou *merodiploide*. Geralmente, uma cópia encontra-se no cromossomo e a segunda cópia, em outro elemento genético, como um plasmídeo ou em um bacteriófago.

Consequentemente, se a cópia cromossômica de um gene for defectiva devido a uma mutação, é possível introduzir uma cópia funcional do gene em um plasmídeo ou fago. Por exemplo, se um dos genes da biossíntese de triptofano for inativado, isso originará um fenótipo  $Trp^-$ . Isto é, a linhagem mutante será auxotrófica para triptofano e necessitará deste aminoácido para seu crescimento. Todavia, se uma cópia do gene selvagem for inserida na mesma célula, por meio de um plasmídeo ou genoma viral, esse gene codificará a proteína necessária, restaurando, assim, o fenótipo selvagem. Esse processo é denominado *complementação*, pois o gene selvagem *complementa* a mutação, nesse caso, convertendo a célula  $Trp^-$  em  $Trp^+$  (Figura 10.12).

### MINIQUESTIONÁRIO

- Qual proteína, encontrada em todos os procariotos, facilita o pareamento necessário para a recombinação homóloga?
- Por que um fenótipo mutador pode ser útil em um ambiente sujeito a rápidas modificações?
- O que é um merodiploide?



**Figura 10.12** Uso de meio seletivo para detectar recombinantes genéticos raros. No meio seletivo, apenas os recombinantes raros formam colônias, apesar de uma grande população ter sido semeada. Procedimentos desse tipo, que oferecem alta resolução às análises genéticas, são normalmente usados somente com microrganismos. O tipo de permuta genética ilustrada corresponde à transformação.

## 10.6 Transformação

A **transformação** é uma transferência genética a partir da qual o *DNA livre* é incorporado em uma célula receptora, podendo promover alterações genéticas. Vários procariotos são naturalmente transformáveis, incluindo espécies gram-positivas e gram-negativas de bactérias, e também espécies de arqueias (Seção 10.10). Uma vez que o DNA de procariotos é encontrado na célula na forma de uma única grande molécula, quando a célula é submetida à lise branda, o DNA é extravasado. Devido ao seu enorme tamanho (p. ex.,  $1.700 \mu\text{m}$ , em *Bacillus subtilis*), os cromossomos bacterianos são facilmente quebrados. Mesmo após uma extração branda, o cromossomo de *B. subtilis*, contendo 4,2 Mb, é convertido em fragmentos de aproximadamente 10 kb. Uma vez que o DNA correspondente a um gene médio contém cerca de 1.000 nucleotídeos, cada um dos fragmentos do DNA de *B. subtilis* contém, aproxima-

damente, 10 genes. Esse é um tamanho transformável típico. Uma única célula normalmente incorpora apenas um ou poucos fragmentos de DNA, de modo que somente uma pequena proporção dos genes de uma célula é transferida à outra célula, em um único evento de transformação.

### A competência na transformação

Mesmo entre os gêneros transformáveis, apenas algumas linhagens ou espécies são, de fato, transformáveis. Uma célula capaz de captar uma molécula de DNA e ser transformada é referida como *competente*, sendo essa capacidade determinada geneticamente. A competência é regulada na maioria das bactérias naturalmente transformáveis, havendo proteínas especiais que desempenham papéis na captação e no processamento do DNA. Essas proteínas específicas de competência incluem uma proteína de ligação ao DNA, associada à membrana, uma autolisina de parede celular e várias nucleases. Uma via de competência natural em *Bacillus subtilis* – uma espécie facilmente transformável – é regulada por um sistema de *quorum sensing* (um sistema regulador que responde à densidade populacional), (⇌ Seção 7.9). As células produzem e secretam um pequeno peptídeo durante seu crescimento, o qual se acumula, atingindo concentrações elevadas, e induz as células a tornarem-se competentes. Em *Bacillus*, cerca de 20% das células de uma cultura tornam-se competentes, permanecendo nesse estado por várias horas. Entretanto, no gênero *Streptococcus*, 100% das células tornam-se competentes, mas apenas por um breve período, durante o ciclo de crescimento.

A transformação natural com alta eficiência é conhecida apenas em algumas bactérias. Por exemplo, *Acinetobacter*, *Bacillus*, *Streptococcus*, *Haemophilus*, *Neisseria* e *Thermus* são naturalmente competentes e facilmente transformáveis. Ao contrário, a transformação em muitos procariontes, se ocorrer, é muito difícil em condições naturais. *E. coli* e várias outras bactérias Gram-negativas enquadram-se nessa categoria. Contudo, quando células de *E. coli* são tratadas com altas concentrações de íons cálcio e resfriadas por vários minutos, tornam-se adequadamente competentes. Células de *E. coli* tratadas dessa maneira captam DNA de fita dupla, tornando a transformação desse organismo com DNA plasmidial relativamente eficiente. Isso é importante uma vez que a introdução de DNA em *E. coli* – o microrganismo-modelo da engenharia genética – é crucial à biotecnologia, como veremos no Capítulo 11.

A *eletroporação* é uma técnica física utilizada para introduzir DNA em organismos dificilmente transformáveis,

especialmente aqueles que possuem paredes celulares espessas. Na eletroporação, as células são misturadas ao DNA e, em seguida, submetidas a pulsos elétricos curtos de alta voltagem. Esse procedimento torna o envoltório celular permeável, permitindo a entrada do DNA. A eletroporação é um processo rápido, eficiente para a maioria dos tipos celulares, incluindo *E. coli*, muitas outras bactérias, alguns membros de *Archaea* e até mesmo leveduras e determinadas células vegetais.

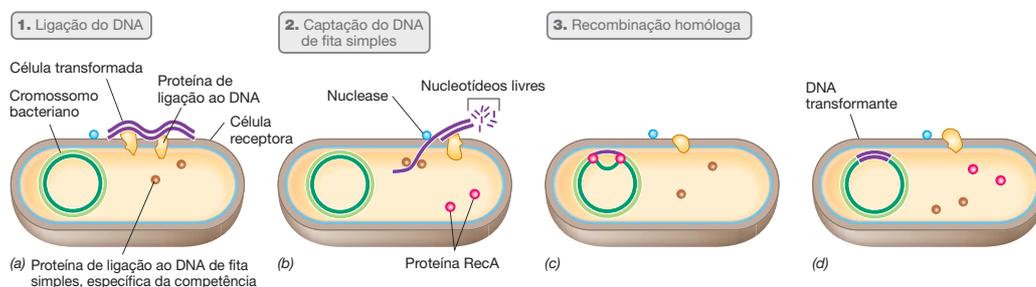
### Captção do DNA na transformação

Durante a transformação, as bactérias competentes ligam o DNA de forma reversível. Todavia, logo em seguida, a ligação torna-se irreversível. Células competentes ligam uma quantidade muito maior de DNA que células não competentes – cerca de 1.000 vezes mais. Como mencionado anteriormente, os tamanhos dos fragmentos transformantes são muito menores do que o genoma completo, sendo esses fragmentos adicionalmente degradados durante o processo de captação. Em *S. pneumoniae*, cada célula é capaz de ligar somente cerca de dez moléculas de DNA de dupla-fita, de 10 a 15 kb cada uma. No entanto, à medida que tais fragmentos são captados, são convertidos em segmentos de fita simples, de aproximadamente 8 kb, sendo a fita complementar degradada. Os fragmentos de DNA presentes na mistura competem entre si no momento da captação e, caso haja a adição de um excesso de DNA desprovido do marcador genético em estudo, há uma redução no número de transformantes.

Curiosamente, a transformação em *Haemophilus influenzae* requer a presença de uma sequência particular, de 11 pb, no fragmento de DNA, para que ocorra a ligação irreversível e posterior captação da molécula. Essa sequência é encontrada no genoma de *Haemophilus* com uma frequência surpreendentemente elevada. Evidências desse tipo, associadas ao fato de que certas bactérias tornam-se competentes em seus ambientes naturais, sugerem que a transformação não corresponde a um artefato laboratorial, desempenhando um importante papel na transferência horizontal de genes na natureza. Ao promover novas combinações de genes, as bactérias naturalmente transformáveis aumentam a diversidade e a adaptabilidade da comunidade microbiana como um todo.

### Integração do DNA transformante

O DNA transformante liga-se à superfície da célula por meio de uma proteína de ligação ao DNA (Figura 10.13). Em seguida, dependendo do organismo, o fragmento inteiro de dupla-fita é captado, ou uma nuclease degrada uma das fitas, sendo a re-



**Figura 10.13** Mecanismo de transformação em uma bactéria gram-positiva. (a) Ligação do DNA de dupla-fita por uma proteína de ligação ao DNA associada à membrana. (b) Passagem de uma das duas fitas para o interior da célula, enquanto a atividade de nucleases degrada a outra fita. (c) Ao pe-

netrar na célula, a fita simples liga-se a outras proteínas específicas, sendo a recombinação com regiões homólogas do cromossomo bacteriano mediada pela proteína RecA. (d) Célula transformada.

manescente captada. Após a captação, o DNA liga-se a uma proteína específica de competência. Essa proteína protege o DNA do ataque de nucleases, até que ele atinja o cromossomo, quando a proteína RecA passa a participar do processo. O DNA é então integrado ao genoma da célula receptora por recombinação (Figuras 10.13 e 10.11). Caso haja a integração de um DNA de fita simples, ocorre a formação de um DNA heterodúplex. Durante o próximo ciclo de replicação cromossômica, são formadas uma molécula de DNA parental e uma molécula de DNA recombinante. Ao ocorrer a segregação, durante a divisão celular, a molécula recombinante encontra-se presente na célula transformada, que está agora geneticamente alterada, em relação à célula parental. Essa discussão é pertinente apenas nos casos envolvendo pequenos fragmentos de DNA linear. Muitas bactérias naturalmente transformáveis são transformadas com baixa eficiência por DNA plasmidiais, uma vez que eles devem permanecer na forma circular e de dupla-fita para que sofram replicação.

#### MINIQUESTIONÁRIO

- Explique por que durante a transformação, apenas um ou pouquíssimos fragmentos de DNA são incorporados pela célula.
- Mesmo em células naturalmente transformáveis, a competência geralmente é induzível. O que isso significa?

### 10.7 Transdução

Na **transdução**, um vírus bacteriano (bacteriófago) transfere o DNA de uma célula para outra. Os vírus podem transferir genes hospedeiros de duas formas. Na primeira, denominada *transdução generalizada*, o DNA derivado de qualquer região do genoma do hospedeiro é empacotado no interior do vírion maduro, substituindo o genoma viral. Na segunda forma, denominada *transdução especializada*, o DNA de uma região específica do cromossomo do hospedeiro encontra-se diretamente integrado no genoma viral – geralmente substituindo alguns dos genes virais. Esse processo ocorre apenas com alguns vírus temperados (↔ Seção 8.8).

Na transdução generalizada, os genes da célula doadora não possuem a capacidade de replicarem-se de maneira independente e também não correspondem a uma porção do genoma viral. Se esses genes doadores não realizarem recombinação com o cromossomo da bactéria receptora, serão perdidos. Na transdução especializada, a recombinação homóloga pode

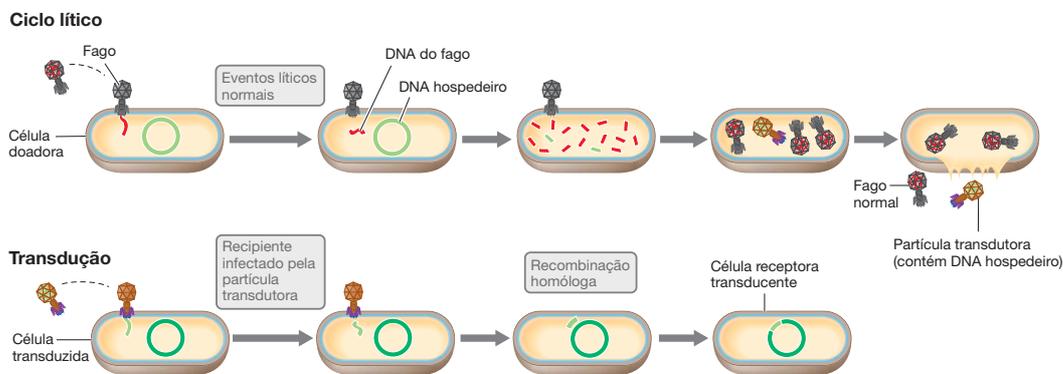
também ocorrer. No entanto, pelo fato de o DNA da bactéria doadora corresponder à parte do genoma de um fago temperado, ele pode integrar-se ao cromossomo da célula hospedeira durante a lisogenia (↔ Seção 8.8).

A transdução ocorre em uma variedade de bactérias, incluindo os gêneros *Desulfovibrio*, *Escherichia*, *Pseudomonas*, *Rhodococcus*, *Rhodobacter*, *Salmonella*, *Staphylococcus* e *Xanthomonas*, bem como em *Methanothermobacter thermoautotrophicus*, uma espécie de arqueias. Nem todos os fagos são capazes de transduzir, assim como nem todas as bactérias são transduzíveis, porém o fenômeno é suficientemente disseminado de modo que, provavelmente, desempenhe um importante papel na transferência gênica na natureza. Exemplos de genes transferidos por bacteriófagos incluem vários genes de resistência a antibióticos entre as linhagens de *Salmonella enterica* sorovar *typhimurium*, genes que codificam para a toxina Shiga em *Escherichia coli*, fatores de virulência em *Vibrio cholerae*, e genes que codificam proteínas que participam da fotossíntese em cianobactérias.

Enquanto a transdução desempenha um papel na transferência horizontal de DNA na natureza, os geneticistas microbianos usam tanto a transdução generalizada quanto a especializada de bacteriófagos para introduzir DNA em células bacterianas-alvo. A transdução pode ser utilizada para introduzir DNA em linhagens em que a transformação e a conjugação não são eficientes. Os bacteriófagos também podem ser usados para produzir grandes fragmentos de DNA para células hospedeiras. Um típico fago com cauda com genoma de DNA de dupla-fita pode empacotar mais de 40 quilobases de pares de DNA. Os bacteriófagos utilizados em laboratório para a transdução geralmente não são líticos porque os genes bacterianos substituíram todos ou alguns genes virais necessários. Para ser selecionado para um evento de transdução, um fago transdutor deve infectar um hospedeiro doador que tem um marcador selecionável.

#### Transdução generalizada

Na transdução generalizada, a princípio qualquer gene do cromossomo doador pode ser transferido para uma célula receptora. A transdução generalizada foi descoberta e intensamente estudada na bactéria *Salmonella enterica* com o fago P22, sendo também estudada em *Escherichia coli* com o fago P1. A **Figura 10.14** ilustra um exemplo de como as partículas transdutoras são formadas. Quando uma célula bacteriana



**Figura 10.14** Transdução generalizada. Observe que os vírions “normais” possuem genes do fago, enquanto a partícula transdutora contém genes do hospedeiro.

é infectada por um fago, um ciclo lítico pode ser iniciado. Todavia, durante a infecção lítica, as enzimas responsáveis pelo empacotamento do DNA viral no bacteriófago podem, acidentalmente, empacotar o DNA da célula hospedeira. O vírion resultante é denominado *partícula transdutora*. Uma vez que as partículas transdutoras são incapazes de promover uma infecção viral (elas não contêm DNA viral), são referidas como *defectivas*. Quando a célula é lisada, essas partículas são liberadas juntamente com os vírions normais (i.e., aqueles contendo o genoma viral). Consequentemente, o lisado contém uma mistura de vírions normais e partículas transdutoras.

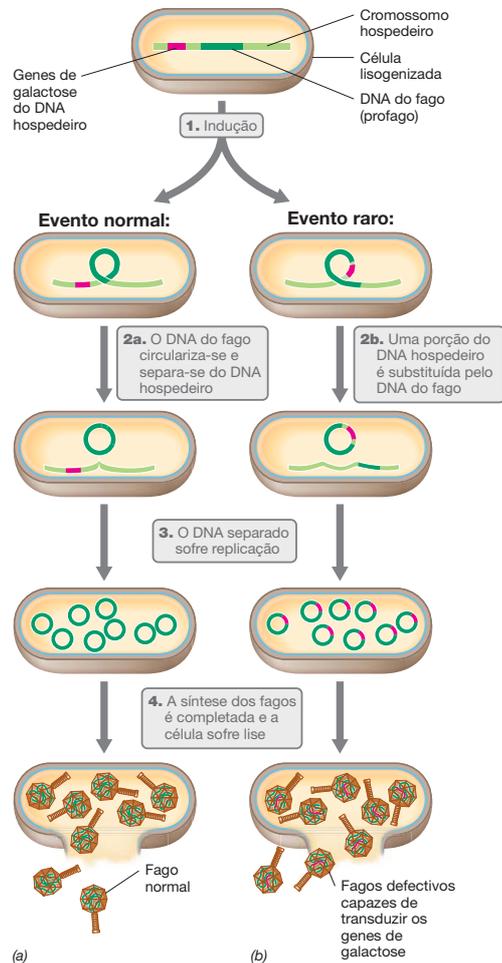
Quando esse lisado é utilizado para infectar uma população de células receptoras, a maioria das células é infectada pelos vírus normais. Contudo, uma pequena parcela da população recebe as partículas transdutoras, que injetam o DNA que receberam da bactéria hospedeira prévia. Embora esse DNA não seja capaz de replicar-se, pode sofrer recombinação genética com o DNA (Seção 10.5) da nova célula hospedeira. Uma vez que apenas uma pequena proporção das partículas do lisado é defectiva, cada uma contendo somente um pequeno fragmento de DNA da célula doadora, a probabilidade de uma determinada partícula transdutora conter um gene em particular é bastante baixa. Normalmente, somente cerca de 1 em  $10^6$  a  $10^8$  células é transduzida com um determinado marcador.

### Lisogenia e transdução especializada

A transdução generalizada permite a transferência de qualquer gene de uma bactéria à outra, porém com baixa frequência. De outra forma, a transdução especializada permite uma transferência extremamente eficiente, porém ela é seletiva, transferindo apenas uma pequena região do cromossomo bacteriano. No primeiro caso descrito de transdução especializada, genes de galactose foram transduzidos pelo fago temperado lambda, de *E. coli*.

Quando o fago lambda lisogeniza uma célula hospedeira, o genoma viral integra-se ao DNA hospedeiro em um sítio específico (↔ Seção 8.8). A região na qual o fago lambda integra-se ao cromossomo de *E. coli* encontra-se adjacente ao agrupamento de genes que codificam as enzimas envolvidas na utilização de galactose. Após a inserção, a replicação do DNA viral passa a ser controlada pela célula bacteriana hospedeira. Quando ocorre a indução, o DNA viral separa-se do DNA hospedeiro por um processo inverso à integração (Figura 10.15). Geralmente, o DNA de lambda é excisado precisamente como uma unidade, porém, ocasionalmente, o genoma do fago é excisado de forma incorreta. Alguns dos genes bacterianos adjacentes a uma das extremidades do prófago (p. ex., o óperon galactose) são excisados juntamente com o DNA viral. Concomitantemente, alguns genes virais são deixados para trás (Figura 10.15b). Esta partícula transdutora pode subsequentemente transferir para uma célula receptora genes para a utilização de galactose. Esta transferência só pode ser detectada se uma cultura bacteriana galactose-negativa ( $\text{Gal}^-$ ) for infectada com tal partícula de transdução e os transdutores  $\text{Gal}^+$  são selecionados.

Para que um vírion de lambda seja viável, há um limite em relação à quantidade de DNA viral que pode ser substituído pelo DNA da célula hospedeira. Uma quantidade suficiente de DNA viral deve ser mantida, a fim de codificar o capsídeo proteico e outras proteínas virais necessárias à lise e lisogenia. Contudo, se um fago auxiliar (em inglês, *helper*) for utilizado juntamente com um fago defectivo em uma infecção mista,



**Figura 10.15** Transdução especializada. (a) Eventos líticos normais e (b) a produção de partículas transdutoras dos genes de galactose, em uma célula de *Escherichia coli*, contendo um prófago lambda.

um número muito menor de genes virais específicos é necessário no fago defectivo. Quando um fago auxiliar é empregado, somente a região *att* (do inglês, *attachment*, que significa ligação), o sítio *cos* (correspondente às extremidades coesivas, importantes no empacotamento) e a origem de replicação do genoma de lambda (↔ Figura 8.17a) são absolutamente necessários à produção de uma partícula transdutora.

### Conversão fágica

A alteração do fenótipo de uma célula hospedeira decorrente da lisogenização é denominada *conversão fágica*. Quando um fago temperado normal (i.e., não defectivo) lisogeniza uma célula e torna-se um prófago, a célula adquire imunidade contra uma nova infecção pelo mesmo tipo de fago. Tal tipo de imunidade pode ser considerada, por si só, uma alteração no fenótipo. No entanto, outras alterações fenotípicas, não relacionadas à imunidade contra fagos, são frequentemente observadas em células lisogenizadas.

Dois casos de conversão fágica foram estudados em detalhe. Um deles envolve uma alteração na estrutura de um polis-



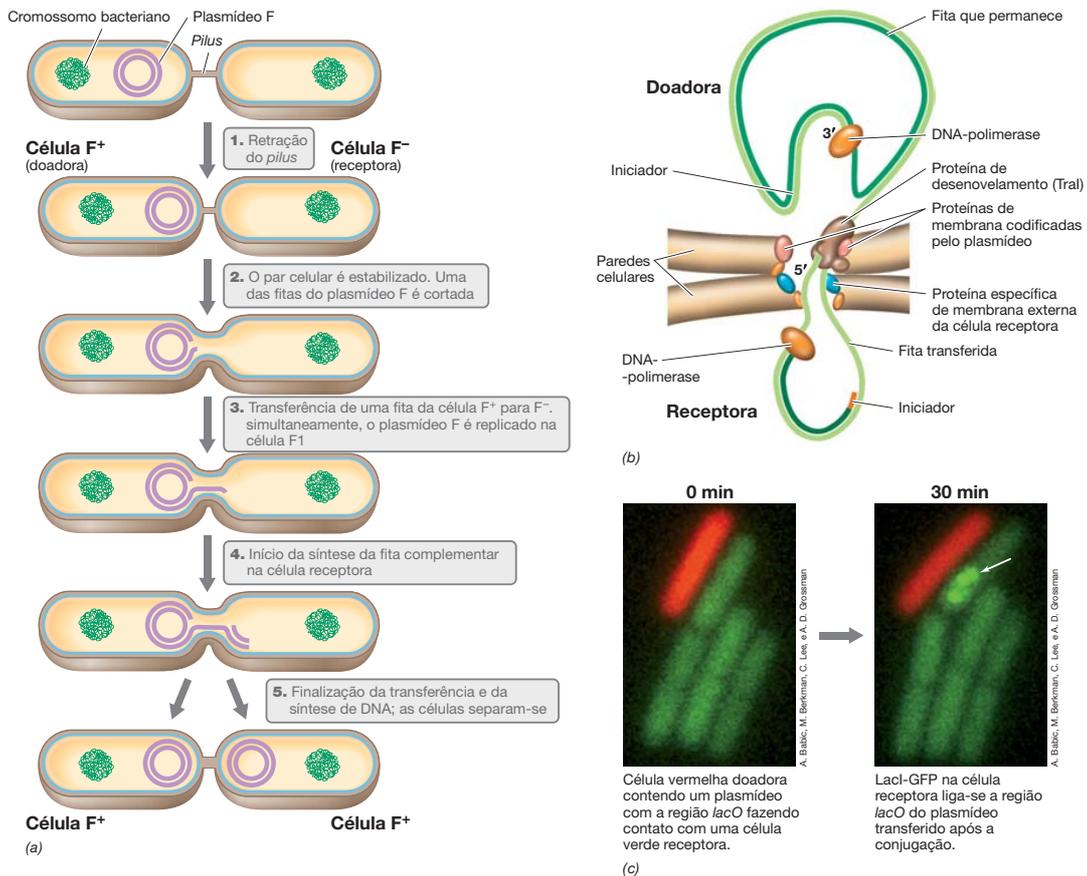
### Mecanismo de transferência de DNA durante a conjugação

A transferência de DNA na conjugação requer a síntese de DNA. Esse DNA não é sintetizado pelo mecanismo convencional de replicação semiconservativa (↔ Seção 4.6), mas sim pela **replicação por círculo rolante**, um mecanismo também empregado por alguns vírus (↔ Seções 8.8 e 9.3), ilustrado na **Figura 10.18**. A transferência do DNA é desencadeada pelo contato entre as células, quando uma fita do DNA plasmidial circular é cortada e então transferida à célula receptora. A enzima que cliva o DNA, Tral, necessária à iniciação do processo, é codificada pelo óperon *tra* do plasmídeo F. Essa proteína também exibe atividade de helicase e, portanto, também participa do desenovelamento da fita a ser transferida. À medida que a transferência ocorre, a síntese de DNA pelo mecanismo de círculo rolante repõe a fita transferida na célula doadora, enquanto uma fita complementar de DNA está sendo sintetizada na célula receptora. Assim, ao final do processo, tanto a célula doadora quanto a receptora terão plasmídeos completos. Na transferên-

cia do plasmídeo F, uma célula doadora contendo F, designada F<sup>+</sup>, acasala-se com uma célula receptora desprovida do plasmídeo, designada F<sup>-</sup>, resultando em duas células F<sup>+</sup> (Figura 10.18).

A transferência do DNA plasmidial é eficiente e rápida; em condições favoráveis, todas as células receptoras que formam pares com as doadoras adquirem um plasmídeo. A transferência do plasmídeo F, que contém aproximadamente 100 kb, leva cerca de cinco minutos. Se os genes plasmidiais puderem ser expressos na célula receptora, ela torna-se uma célula doadora, podendo agora transferir o plasmídeo a outras células receptoras.

Dessa forma, os plasmídeos conjugativos têm a capacidade de disseminar-se em populações bacterianas, comportando-se como agentes infecciosos. Esse fato possui grande importância ecológica, uma vez que esses plasmídeos já foram encontrados em muitas bactérias e em algumas arqueias (Seção 10.10), e por que a introdução de um pequeno número de células portadoras de plasmídeos em uma população de células receptoras poderá converter toda a população em células portadoras de plasmídeo (portanto, doadoras) em um curto período de tempo.



**Figura 10.18** Transferência de DNA plasmidial por conjugação. (a) A transferência do plasmídeo F converte uma célula receptora F<sup>-</sup> em uma célula F<sup>+</sup>. Observe o mecanismo de replicação por círculo rolante. (b) Detalhes do processo de replicação e transferência. (c) Visualização da transferência de DNA por conjugação em *Bacillus subtilis* usando microscopia de fluorescência. A célula doadora expressa constitutivamente uma proteína verde fluorescente,

enquanto a célula receptora fluoresce verde devido à proteína verde fluorescente (GFP) fusionada ao LacI (↔ Figura 7.15). O DNA transferido do doador contém a região do operador *lacO* que liga a LacI-GFP. As setas indicam pontos focais na célula receptora onde LacI-GFP está ligada a uma região *lacO* obtida a partir da conjugação.

**MINIQUESTIONÁRIO**

- Na conjugação, como as células doadoras e receptoras estabelecem contato uma com a outra?
- Explique como a replicação do DNA por círculo rolante permite que tanto o doador quanto o receptor terminem o processo com uma cópia completa do plasmídeo transferido durante a conjugação.
- Por que o plasmídeo F tem duas origens de replicação diferentes?

**10.9 Formação de linhagens Hfr e mobilização do cromossomo**

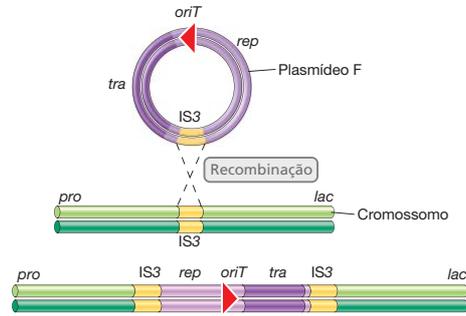
Genes cromossômicos podem ser transferidos por conjugação. Como mencionado anteriormente, o plasmídeo F de *Escherichia coli* pode, sob certas circunstâncias, mobilizar o cromossomo e realizar a sua transferência durante o contato intercelular. O plasmídeo F é, na realidade, um *episomo*, um plasmídeo capaz de integrar-se ao cromossomo da célula hospedeira, e, quando o plasmídeo encontra-se integrado, genes cromossômicos podem ser transferidos. A transferência horizontal de genes por esse mecanismo, seguida da recombinação entre a célula doadora e receptora, pode ser bastante extensa em termos de número e diferentes tipos de genes cromossômicos que podem ser transferidos.

Células que possuem um plasmídeo F não integrado são denominadas  $F^+$ . Aquelas que apresentam o plasmídeo integrado ao cromossomo são denominadas **células Hfr** (do inglês, *high frequency of recombination* [alta frequência de recombinação]). Esse termo refere-se às altas taxas de recombinação genética entre os genes cromossômicos da doadora e os da receptora. Tanto as células  $F^+$  quanto as Hfr são doadoras, mas, contrariamente à conjugação entre uma  $F^+$  e uma  $F^-$ , a conjugação entre uma doadora Hfr e uma  $F^-$  promove a transferência de genes cromossômicos da célula hospedeira. Isso ocorre porque, nessa situação, o cromossomo e o plasmídeo formam uma única molécula de DNA. Conseqüentemente, quando a replicação por círculo rolante é iniciada pelo plasmídeo F, o processo continua e inclui o cromossomo. Assim, o cromossomo é também replicado e transferido. Por essa razão, a integração de um plasmídeo conjugativo fornece um mecanismo de mobilização dos recursos genéticos de uma célula.

Em termos globais, a presença do plasmídeo F resulta em três alterações distintas das propriedades de uma célula: (1) a capacidade de sintetizar o *pilus* F, (2) a mobilização do DNA e sua transferência para outra célula, e (3) a alterações dos receptores de superfície, de modo que a célula deixa de atuar como receptora na conjugação, sendo incapaz de captar uma segunda cópia do plasmídeo F ou de plasmídeos geneticamente relacionados.

**Integração de F e mobilização cromossômicos**

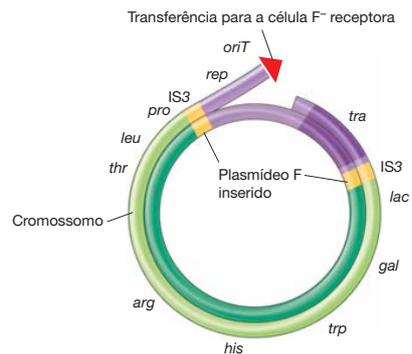
O plasmídeo F e o cromossomo de *E. coli* carregiam várias cópias de elementos móveis denominados *sequências de inserção* (IS, *insertion sequences*; Seção 10.11). Os IS são regiões que exibem homologia entre o DNA do plasmídeo F e o DNA cromossômico. Conseqüentemente, a recombinação homóloga entre um IS do plasmídeo F e um IS cromossômicos correspondente resulta na integração do plasmídeo F ao cro-



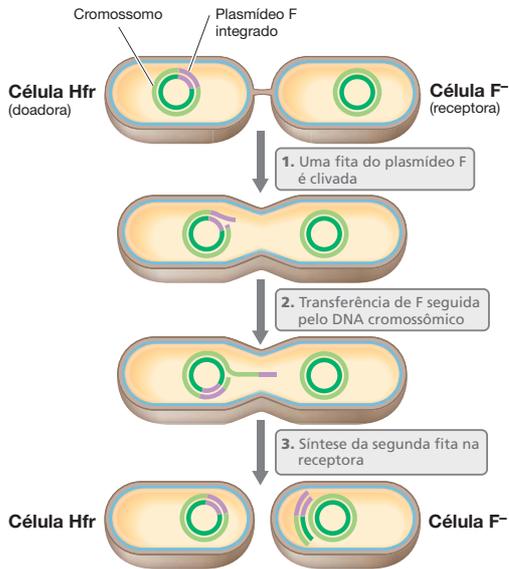
**Figura 10.19** Formação de uma linhagem Hfr. A integração de um plasmídeo F no cromossomo pode ocorrer em vários sítios específicos, onde elementos IS estão localizados. O exemplo apresentado nesta figura corresponde a um IS3, localizado entre os genes cromossômicos *pro* e *lac*. São ilustrados alguns genes do plasmídeo F. A seta indica a origem de transferência, *oriT*, em que a seta corresponde à extremidade-líder. Assim, nesta célula Hfr, *pro* corresponderia ao primeiro gene cromossômico a ser transferido, enquanto *lac* estaria entre os últimos.

mossomo do hospedeiro, como ilustrado na **Figura 10.19**. Uma vez integrado, o plasmídeo deixa de controlar sua replicação, embora o óperon *tra* continue funcionando normalmente, e a linhagem seja capaz de sintetizar *pili*. Quando uma célula receptora é encontrada, a conjugação é desencadeada, conforme descrito para células  $F^+$ , sendo a transferência de DNA iniciada no sítio *oriT* (origem de transferência). Entretanto, como o plasmídeo está agora integrado ao cromossomo, após a transferência de parte do DNA plasmidial, os genes cromossômicos começam a ser transferidos (**Figura 10.20**). Como na conjugação envolvendo apenas o plasmídeo F (**Figura 10.18**), a transferência do DNA também envolve a replicação.

Devido ao fato de que a fita de DNA, em geral, se quebra durante a transferência, apenas parte do cromossomo doador é transferida. Conseqüentemente, a célula receptora não se torna Hfr (ou  $F^+$ ), pois apenas parte do plasmídeo F integrado é transferida (**Figura 10.21**). Não obstante, após a transferência a



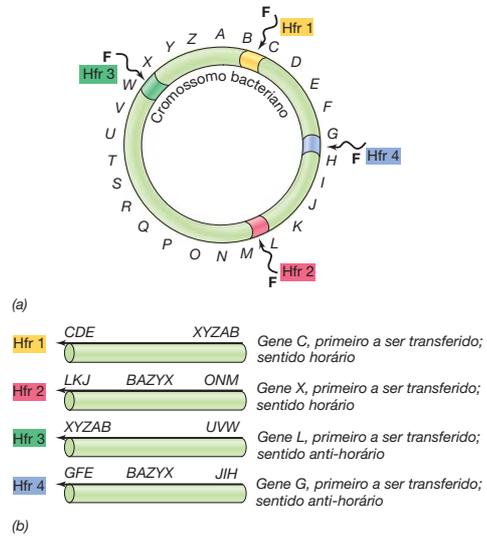
**Figura 10.20** Transferência de genes cromossômicos por uma linhagem Hfr. O cromossomo Hfr é clivado na origem de transferência, situada no interior do plasmídeo F integrado. A transferência do DNA para a célula receptora é iniciada neste ponto. O DNA replica-se durante a transferência da mesma forma que um plasmídeo F livre (**Figura 10.18**). A figura não está em escala; o plasmídeo F integrado corresponde a menos de 3% do cromossomo de *Escherichia coli*.



**Figura 10.21** Transferência de DNA cromossômico por conjugação. A transferência do plasmídeo F integrado em uma linhagem Hfr resulta na co-transferência do DNA cromossômico, uma vez que ele se encontra ligado ao plasmídeo. As etapas de transferência são similares àquelas apresentadas na Figura 10.18a. Todavia, a célula receptora permanece F<sup>-</sup> e recebe um fragmento linear do cromossomo doador ligado à parte do plasmídeo F. Para que o DNA da doadora seja perpetuado, deve ser recombinado no cromossomo da receptora, após a transferência (não ilustrado).

linhagem Hfr permanece Hfr, uma vez que ela retém uma parte do plasmídeo F integrado. Como um cromossomo parcial não pode se replicar, para que o DNA recém-chegado possa “sobreviver”, ele precisa se recombinar com o cromossomo recipiente. Após a recombinação, a célula recipiente pode expressar um novo fenótipo, em razão da incorporação de novos genes. Embora linhagens Hfr transmitam genes cromossômicos frequentemente, elas normalmente não convertem células F<sup>-</sup> em células F<sup>+</sup> ou Hfr, uma vez que o plasmídeo F raramente é transferido em sua totalidade. Em vez disso, um cruzamento entre Hfr e F<sup>-</sup> resulta na Hfr original e em uma célula F<sup>-</sup>, que agora apresenta um novo genótipo. Assim como ocorre durante a transformação e a transdução, a recombinação genética entre genes Hfr e genes F<sup>-</sup> envolve recombinação homóloga na célula recipiente.

Como diversas inserções de sequências distintas estão presentes no cromossomo, muitas linhagens Hfr distintas podem ocorrer. Uma determinada linhagem Hfr sempre doa genes na mesma ordem, iniciando sempre na mesma posição. Entretanto, linhagens Hfr que apresentam diferentes locais de integração do plasmídeo F irão transferir genes em ordens diferentes (Figura 10.22). Em alguns locais de inserção, o plasmídeo F encontra-se integrado com sua origem apontando para uma direção, ao passo que em outros locais de inserção, a origem do plasmídeo aponta na direção oposta. A orientação dos plasmídeo F determina quais genes cromossômicos serão transferidos primeiro, o que ilustra como genes adquiridos através de transferência horizontal e recombinados ao cromossomo recipiente podem ser transferidos para novas células recipientes (Figura 10.22). Utilizando diversas linhagens Hfr



**Figura 10.22** Formação de diferentes linhagens Hfr. Linhagens Hfr distintas transferem os genes em ordens diferentes e a partir de origens distintas. (a) Os plasmídeos F podem ser inseridos nas várias sequências de inserção do cromossomo bacteriano, originando diferentes linhagens Hfr. (b) Ordem de transferência gênica em diferentes linhagens Hfr.

em experimentos de cruzamento, foi possível determinar o arranjo e a orientação da maioria dos genes no cromossomo da *E. coli* (↔ Seção 4.3) muito antes que este fosse sequenciado.

### Transferência de genes cromossômicos para o plasmídeo F

Ocasionalmente, plasmídeos F integrados podem ser excisados do cromossomo. Durante a excisão, genes cromossômicos podem ser algumas vezes incorporados no plasmídeo F liberado. Isso acontece porque ambos, o plasmídeo F e o cromossomo, contêm múltiplas sequências de inserção idênticas, possibilitando a ocorrência de recombinação (Figura 10.20). Plasmídeos F contendo genes cromossômicos são denominados *plasmídeos F'*. Eles diferem dos plasmídeos F normais por conterem genes cromossômicos identificáveis. Quando plasmídeos F' promovem a conjugação, transferem esses genes cromossômicos com alta frequência para a célula hospedeira. A transferência mediada por F' assemelha-se à transdução especializada (Seção 10.7), pelo fato de apenas um grupo restrito de genes cromossômicos ser transferido por um determinado plasmídeo F'. A transferência de um F' conhecido para uma célula receptora permite o estabelecimento de diploides (duas cópias de cada gene) em uma região limitada do cromossomo. Esses diploides parciais (merodiploides) são importantes na realização de testes de complementação (Seção 10.5).

#### MINIQUESTIONÁRIO

- Em uma conjugação envolvendo um plasmídeo F de *Escherichia coli*, como o cromossomo hospedeiro é mobilizado?
- Por que um acasalamento Hfr X F<sup>-</sup> não origina duas células Hfr?
- A quais sítios cromossômicos o plasmídeo F é capaz de integrar-se?

### III • Transferência gênica em arqueias e outros eventos genéticos

Aqui, cobriremos tópicos como a genética de arqueias, que apesar de estar muito atrasada em comparação a bactérias, está começando a emergir juntamente com versões para arqueias das ferramentas genéticas necessárias para análises mais detalhadas. Além disso, discutiremos outros eventos genéticos que ocorrem em bactérias, os quais são importantes conceitos gerais, mas que não participam do fluxo lateral de genes em si.

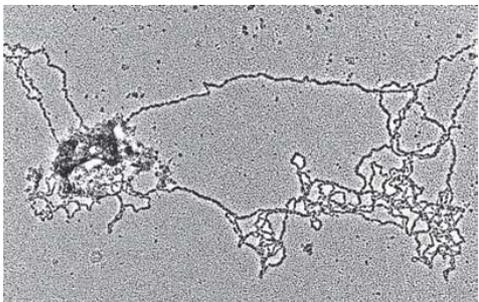
#### 10.10 Transferência horizontal de genes em arqueias

Embora as arqueias contenham um único cromossomo circular, como a maioria dos membros de *Bacteria* (Figura 10.23), e, apesar de os genomas de várias arqueias já estarem totalmente sequenciados e indicarem que ocorre transferência horizontal de genes na natureza, o desenvolvimento de sistemas de transferência genética para esse grupo encontra-se ainda muito atrasado, quando comparado a bactérias. Entre os problemas práticos, podemos citar a necessidade do cultivo de várias arqueias sob condições extremas (Capítulo 16). Assim, as temperaturas requeridas ao cultivo de alguns hipertermófilos promoverão o derretimento do ágar, havendo a necessidade de se utilizar materiais alternativos para confeccionar meios sólidos e obterem-se colônias.

Outro problema está no fato de a maioria dos antibióticos não afetarem as arqueias. Por exemplo, a penicilina não afeta arqueias devido ao fato de suas paredes celulares não possuírem peptidoglicano. A escolha de marcadores selecionáveis para os cruzamentos genéticos é, portanto, muitas vezes limitada. Todavia, a novobiocina (um inibidor da DNA-girase) e a mevinolina (um inibidor da síntese de isoprenoides) são utilizadas na inibição de halófilos, enquanto a puromicina e neomicina (inibidores da síntese proteica) inibem os metanogênicos. Linhagens auxotróficas de algumas arqueias também foram isoladas com o propósito de seleção genética.

#### Sistemas genéticos de arqueias

Nenhuma espécie de *Archaea* tornou-se um organismo-modelo para a genética desse grupo, embora, provavelmente, a maior parte dos trabalhos genéticos tenha sido realizada com determinadas espécies de halófilos extremos (*Halobacterium*,



**Figura 10.23** Um cromossomo de arqueia, observado ao microscópio eletrônico. Cromossomo circular de *Sulfolobus*, um membro hipertermófilo de *Archaea*.

*Haloferax*, ⇨ Seção 16.1), do que com qualquer outra arqueia. Exemplos de transformação, transdução viral e conjugação são conhecidos. Além disso, vários plasmídeos foram isolados de arqueias e alguns foram utilizados na construção de vetores de clonagem, permitindo a análise genética por meio de clonagem e sequenciamento, em vez de cruzamentos genéticos tradicionais. A mutagênese por transposons (Seção 10.11) foi também desenvolvida para determinadas espécies de metanogênicas, incluindo *Methanococcus* e *Methanosarcina*; outras ferramentas, como vetores de clonagem e outros métodos *in vitro* de análise genética foram desenvolvidos visando o estudo da bioquímica pouco usual dos metanogênicos (⇨ Seções 13.20 e 16.2).

A transformação é razoavelmente bem-sucedida em várias arqueias. Os procedimentos de transformação variam para os diferentes organismos. Uma abordagem envolve a remoção de íons metálicos divalentes que, por sua vez, resulta na desorganização da camada glicoproteica da parede celular que envolve muitas células de arqueias, permitindo, assim, acesso ao DNA. No entanto, arqueias com paredes celulares rígidas são de difícil transformação, embora em algumas situações a eletroporação mostre-se eficiente. Uma exceção são as espécies de *Methanosarcina*, organismos com uma parede celular espessa, para os quais sistemas de transformação de alta eficiência foram desenvolvidos, empregando-se preparações lipídicas carregadas com DNA (lipossomos) que liberam o DNA no interior da célula.

Embora os vírus que infectam arqueias sejam abundantes, a transdução viral é extremamente rara. Somente um único vírus de arqueias, que infecta o metanogênico termofílico *Methanobacterium thermoautotrophicum*, transduz os genes de seu hospedeiro. Infelizmente, o pequeno tamanho da população viral liberada (cerca de seis fagos liberados por célula) torna pouco prático o uso desse sistema para a transferência gênica.

#### Conjugação em arqueias

Dois tipos de conjugação foram descritos em arqueias. Algumas linhagens de *Sulfolobus solfataricus* (⇨ Seção 16.10) contêm plasmídeos que promovem a conjugação entre duas células, de modo similar à conjugação de bactérias. Nesse processo, o pareamento das células ocorre antes da transferência do plasmídeo, sendo a transferência do DNA unidirecional. Contudo, a maioria dos genes que codificam essas funções parece ter pequena similaridade com os genes de bactérias gram-negativas. A exceção é um gene similar a *traG*, do plasmídeo F, cujo produto proteico está envolvido na estabilização do par de acasalamento. Dessa forma, parece provável que o real mecanismo de conjugação em arqueias seja bastante diferente daquele de bactérias.

Algumas halobactérias, ao contrário, realizam um novo tipo de conjugação. Nesse processo, não há a participação de plasmídeos de fertilidade, e a transferência do DNA é bidirecional. São formadas pontes citoplasmáticas entre as células conjugantes, aparentemente utilizadas na transferência do DNA. Nenhum sistema de conjugação foi desenvolvido a ponto de poder ser utilizado rotineiramente para a transferência gênica ou análises genéticas. No entanto, tais recursos genéticos provavelmente ainda serão úteis para o desenvolvimento de sistemas genéticos simples nesses organismos.

**MINIQUESTIONÁRIO**

- Por que a seleção de recombinantes em arqueias é normalmente mais difícil do que em bactérias?
- Por que as penicilinas não matam membros de arqueias?

**10.11 DNA móvel: elementos transponíveis**

Como discutimos anteriormente, moléculas de DNA podem mover-se de uma célula para outra; porém, para um geneticista, o termo “DNA móvel” possui um significado especializado. DNA móvel refere-se a segmentos distintos de DNA que se movem como unidades, de um local a outro, no interior de outras moléculas de DNA.

Embora o DNA de certos vírus possa integrar-se e ser excisado do genoma da célula hospedeira, a maioria dos DNA móveis consiste em **elementos transponíveis**. Eles correspondem a segmentos de DNA capazes de deslocarem-se de um sítio para outro. Todavia, os elementos transponíveis são sempre encontrados inseridos em outra molécula de DNA, como um plasmídeo, cromossomo ou genoma viral. Os elementos transponíveis não possuem suas próprias origens de replicação. Eles são replicados quando a molécula de DNA do hospedeiro onde estão inseridos realiza sua replicação.

Os elementos transponíveis movem-se por um processo denominado *transposição*, que é importante para a evolução e análises genéticas. A frequência de transposição é extremamente variável, de 1 em 1.000 a 1 em 10.000.000 por elemento transponível, por geração celular, dependendo do elemento transponível e do organismo. Os elementos transponíveis são amplamente distribuídos na natureza, podendo ser encontrados nos genomas dos três domínios da vida, bem como em muitos vírus e plasmídeos, sugerindo que esses elementos fornecem uma vantagem seletiva por acelerar o rearranjo genômico.

Os dois principais tipos de elementos transponíveis em bactérias são as *seqüências de inserção* (IS) e os *transposons*. Esses dois elementos têm duas importantes características em comum: eles carregam genes que codificam para a transposase, enzima necessária para a transposição, e eles pos-

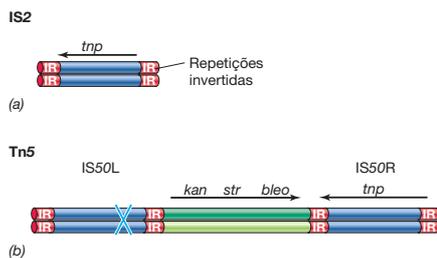
suem nas suas extremidades pequenas repetições terminais invertidas que também são necessárias à transposição. Observe que as extremidades dos elementos transponíveis não se encontram livres, uma vez que são contínuas à molécula de DNA hospedeiro onde foram inseridos. A **Figura 10.24** mostra genético mapas da IS2 elemento de inserção e do transposon Tn5.

**Transposons e seqüências de inserção**

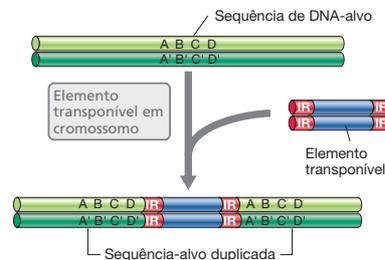
**Seqüências de inserção (ISs)** são o tipo mais simples de elemento transponível. Elas são segmentos curtos de DNA, com cerca de 1.000 nucleotídeos de extensão, geralmente apresentando repetições invertidas de 10 a 50 pb. Cada IS apresenta um número específico de pares de bases em suas repetições terminais. O único gene que as ISs possuem corresponde ao gene de transposase. Várias seqüências de IS distintas foram caracterizadas. Seqüências de inserção são encontradas em cromossomos e plasmídeos de bactérias e arqueias, bem como em certos bacteriófagos. Linhagens individuais da mesma espécie bacteriana variam em relação ao número e à localização dos elementos IS que possuem. Por exemplo, uma linhagem de *E. coli* possui cinco cópias de IS2 e cinco cópias de IS3. Muitos plasmídeos, como o plasmídeo F, também possuem seqüências de inserção. De fato, a integração do plasmídeo F no cromossomo de *E. coli* (Figura 10.19) ocorre devido à recombinação homóloga entre as seqüências de inserção idênticas do plasmídeo F e do cromossomo (Seção 10.9).

Os **transposons** são maiores do que as seqüências de inserção, mas possuem os dois mesmos componentes essenciais: as repetições invertidas em ambas as extremidades e o gene que codifica a transposase (Figura 10.24b). A transposase reconhece as repetições invertidas e desloca o segmento de DNA flanqueado por elas, de um sítio para outro. Consequentemente, qualquer DNA situado entre as duas repetições invertidas é deslocado, sendo ele, de fato, parte do transposon. Os genes incluídos no interior dos transposons variam amplamente. Alguns desses genes, como os genes de resistência a fármacos, conferem propriedades novas e importantes ao organismo que alberga o transposon. Uma vez que a resistência aos antibióticos é importante e de fácil detecção, a maioria dos transposons mais estudados possui genes de resistência a antibióticos como marcadores selecionáveis. Como exemplos temos o Tn5, que carrega a resistência à canamicina (Figura 10.24b) e Tn10, com resistência à tetraciclina.

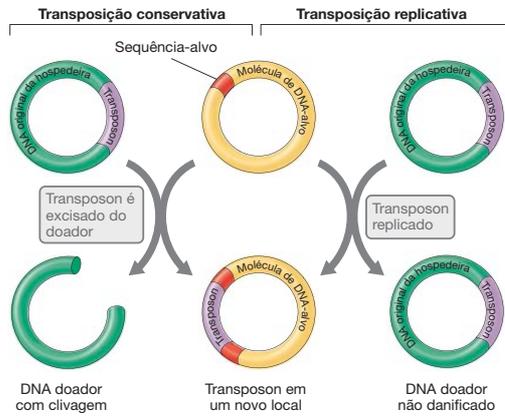
Uma vez que quaisquer genes inseridos entre as repetições invertidas tornam-se parte de um transposon, é possível



**Figura 10.24** Mapas dos elementos de transposição IS2 e Tn5. As setas acima dos mapas indicam a direção da transcrição dos genes presentes nos elementos. O gene *tnp* codifica a transposase. (a) IS2 é uma seqüência de inserção de 1.327 pares de bases, com repetições invertidas de 41 pares de bases em suas extremidades. (b) Tn5 é um transposon composto, com 5,7 kpb, contendo as seqüências de inserção IS50L e IS50R em suas extremidades esquerda e direita, respectivamente. IS50L não é capaz de realizar uma transposição independente, uma vez que possui uma mutação sem sentido, assinalada por uma cruz azul, em seu gene de transposase. Os genes *kan*, *str* e *bleo* conferem resistência aos antibióticos canamicina (e neomicina), estreptomicina e bleomicina.



**Figura 10.25** Transposição. A inserção de um elemento transponível gera uma duplicação da seqüência-alvo. Observe a presença de repetições invertidas (IRs) nas extremidades do elemento transponível.



**Figura 10.26** Dois mecanismos de transposição. O DNA doador (carreando o transposon) é apresentado em verde-claro, e o DNA receptor, carreando a sequência-alvo, é apresentado em amarelo. Tanto na transposição conservativa quanto na replicativa a transposase insere o transposon (em roxo) no sítio-alvo (em vermelho), no DNA receptor. Durante esse processo, o sítio-alvo é duplicado. Na transposição conservativa, o DNA doador permanece com uma clivagem na dupla-fita, no local onde o transposon estava previamente localizado. Ao contrário, na transposição replicativa, os DNA doador e receptor possuem uma cópia do transposon.

construir transposons híbridos que apresentam comportamento complexo. Por exemplo, transposon conjugativos contêm genes *tra* e podem deslocar-se por conjugação entre espécies bacterianas, bem como serem transpostos de um local para outro, no interior de um único genoma bacteriano. Ainda mais complexo é o bacteriófago Mu, que corresponde tanto a um vírus quanto a um transposon (↔ Seção 9.4). Nesse caso, um genoma viral completo está contido no interior de um transposon. Outros elementos genéticos compostos consistem em um segmento de DNA situado entre duas sequências de inserção idênticas. Essa estrutura inteira pode mover-se como uma unidade, sendo denominada *transposon composto*. O comportamento de transposons compostos indica que novos transposons provavelmente surjam periodicamente em células contendo sequências de inserção situadas próximas.

**Mecanismos de transposição**

Ambas as repetições invertidas encontradas nas extremidades dos elementos transponíveis e a transposase são essenciais à transposição. A transposase reconhece, cliva e liga o DNA durante a transposição. Quando um elemento transponível é inserido no DNA-alvo, há a duplicação de uma pequena sequência do DNA-alvo, localizada no sítio de integração (Figura 10.25). A duplicação ocorre devido à clivagem de fitas simples que são realizadas pela transposase. O elemento transponível então se liga às extremidades de fita simples geradas.

Finalmente, enzimas da célula hospedeira reparam as porções de fita simples, resultando na duplicação.

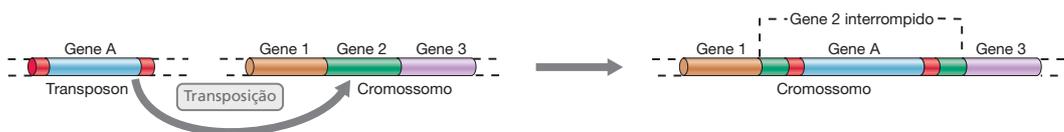
Dois mecanismos de transposição são conhecidos: *conservativa* e *replicativa* (Figura 10.26). Na transposição conservativa, como ocorre com o transposon Tn5, o transposon é excisado de um local, sendo inserido em um segundo local. Com isso, o número de cópias de um transposon conservativo permanece sendo um. De outra forma, durante a transposição replicativa, uma nova cópia é produzida e inserida em um segundo local. Assim, após uma transposição replicativa, uma cópia do transposon permanece no sítio original, havendo uma segunda cópia no sítio novo.

**Mutagênese por transposons**

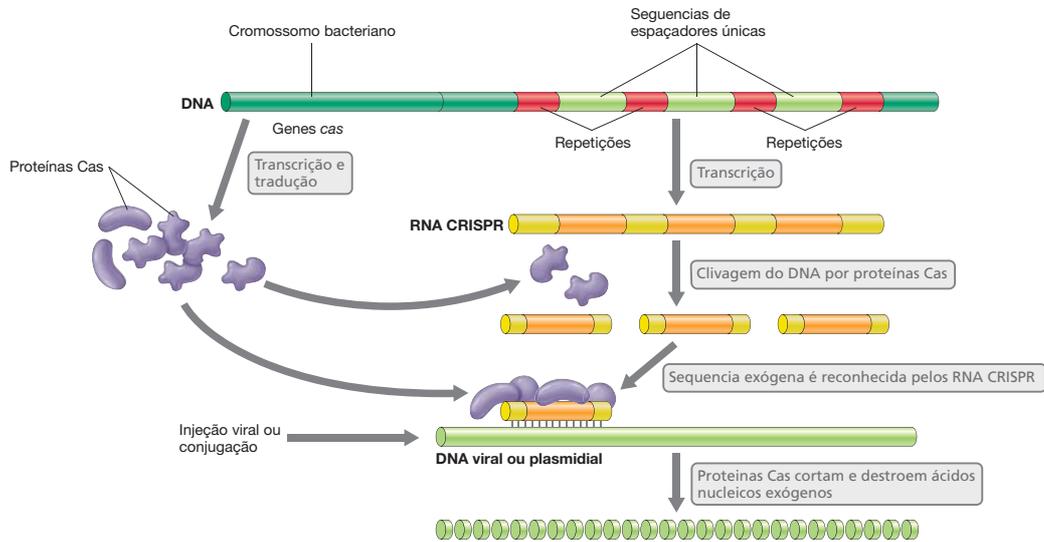
Quando um transposon insere-se no interior de um gene, uma mutação ocorre nesse determinado gene (Figura 10.27). Mutações decorrentes da inserção de um transposon ocorrem naturalmente. No entanto, o uso de transposons para gerar mutações corresponde a uma forma conveniente de criar mutantes bacterianos em laboratório. Geralmente, transposons carreando genes de resistência a antibióticos são utilizados. O transposon é introduzido na célula-alvo por intermédio de um fago ou plasmídeo incapaz de replicar-se naquele hospedeiro em particular. Conseqüentemente, as colônias resistentes ao antibiótico serão oriundas principalmente devido à inserção do transposon no genoma bacteriano.

Como os genomas bacterianos contêm uma pequena quantidade de DNA não codificador, a maioria das inserções de transposons ocorrerá em genes que codificam proteínas. Quando inserido em um gene que codifica uma proteína essencial, a mutação poderá ser letal sob certas condições de cultivo, tornando-se adequada à seleção genética. Por exemplo, se inserções de transposons forem selecionadas em um meio rico, onde os auxotróficos são capazes de crescer, elas podem ser subsequentemente rastreadas em meio mínimo suplementado com vários nutrientes, a fim de se determinar se um nutriente é requerido. Análises adicionais podem ser realizadas com a finalidade de revelar qual gene foi rompido pelo transposon. Mutações auxotróficas devido à inserção de transposons são muito úteis na genética bacteriana. Normalmente, os recombinantes auxotróficos não podem ser isolados por seleção positiva. Contudo, a presença de um transposon com um marcador de resistência a antibióticos permite a seleção positiva.

Dois transposons amplamente utilizados para a mutagênese de *E. coli* e bactérias relacionadas são o Tn5 (Figura 10.24b), que confere resistência à neomicina e canamicina, e o Tn10, que confere resistência à tetraciclina. Muitas bactérias, algumas arqueias e a levedura *Saccharomyces cerevisiae* foram mutagenizadas por meio de mutagênese por transposon. Mais recentemente, os transposons foram empregados para isolar mutações em animais, incluindo camundongos.



**Figura 10.27** Mutagênese por transposon. O transposon move-se para o meio do gene 2. O gene 2 é rompido pelo transposon, sendo inativado. O gene A do transposon será expresso a partir do cromossomo.



**Figura 10.28** A operação do sistema CRISPR. A região CRISPR no cromossomo bacteriano é transcrita em longas moléculas de RNA que depois são processadas em segmentos por algumas das proteínas Cas. Cada segmento espaçador corresponde a encontros anteriores com ácidos nucleicos

exógenos que entraram na célula. Se uma dessas pequenas moléculas de RNA CRISPR (correspondendo a um espaçador) reconhece e para suas bases com o ácido nucleico da transdução ou conjugação, outras proteínas Cas destroem o ácido nucleico exógeno.

#### MINIQUESTIONÁRIO

- Que características as sequências de inserção e os transposons têm em comum?
- Qual a importância das repetições terminais invertidas dos transposons?
- Como os transposons podem ser úteis na genética bacteriana?

### 10.12 Preservação da integridade do genoma: sistema CRISPR de interferência

Bactérias e arqueias produzem endonucleases de restrição (☞ Seções 8.6 e 11.1) que destroem DNA exógeno que entra na célula, mas, em adição, também possuem outro programa de defesa para destruir DNA invasor proveniente de infecções virais e, às vezes, da conjugação. Esse tipo de “sistema imune” em procariotos ajuda a preservar a estabilidade do genoma e é denominado sistema CRISPR (do inglês, *clustered regularly interspaced short palindromic repeats*).

A região CRISPR no cromossomo bacteriano é, por essência, um banco de memória de sequências de ácidos nucleicos que entram na célula, usado para vigilância contra DNA exógeno. Consistindo em grande variedade de segmentos de DNA exógeno que recebe o nome de *espaçadores* em alternância com sequências repetitivas idênticas (Figura 10.28). A sequência espaçadora corresponde a pequenos pedaços de DNA exógeno que tenham previamente invadido a célula. Uma vez que esses espaçadores são recombinados na região CRISPR, o sistema confere resistência para qualquer DNA exógeno (e em algumas vezes RNA) que contenham sequências idênticas ou muito relacionadas com regiões individuais do espaçador. Proteínas do sistema CRISPR executam funções essenciais dessa “imunidade” baseada em RNA.

As proteínas do sistema CRISPR (proteínas associadas a CRISPR ou proteínas Cas) efetuam dois papéis. Algumas

participam na obtenção e armazenamento de segmentos de DNA exógeno como espaçadores, reconhecendo sequências nucleotídicas específicas associadas aos espaçadores. Outras usam a informação da sequência armazenada para reconhecer DNA intruso e destruí-lo. As proteínas Cas são codificadas por genes que estão à montante da sequência de DNA CRISPR (Figura 10.28).

Toda a região CRISPR é transcrita como uma longa molécula de RNA que é clivada no meio de cada sequência repetitiva, pela atividade nucleásica das proteínas Cas. Isso converte a longa molécula de RNA em pequenos RNA de segmentos espaçadores chamados *RNA CRISPR (RNACr)*. Se um desses RNACr parrear com ácidos nucleicos invasores, então o DNA ou RNA exógeno é destruído pela atividade nucleásica de outras proteínas Cas. O sistema CRISPR é amplamente distribuído em ambas arqueias e bactérias. Aproximadamente 90% do genoma sequenciado de arqueias e 70% em bactérias possuem o sistema CRISPR. A utilidade desse sistema foi inicialmente demonstrada na indústria de laticínios em que culturas iniciadoras de fermentação do leite são suscetíveis a infecções explosivas por bacteriófagos. Contudo, uma linhagem de *Streptococcus thermophilus* resistente a bacteriófagos virulentos foi encontrada. A diferença entre essa linhagem de *S. thermophilus* e aquelas suscetíveis a infecção viral está nos espaçadores dentro da região CRISPR. Enquanto não se sabe por que alguns vírus não são alvos do sistema CRISPR, experimentos em laboratório têm demonstrado bacteriófagos que podem superar o reconhecimento por RNACr e proteínas Cas, modificando o próprio genoma por mutação.

#### MINIQUESTIONÁRIO

- Por que o sistema CRISPR é considerado um “sistema imune” em procariotos?
- A que correspondem os espaçadores dentro da região CRISPR?

## CONCEITOS IMPORTANTES

**10.1** • Mutações são alterações herdáveis na sequência de DNA que podem levar a alterações no fenótipo. Mutações selecionáveis são aquelas que conferem ao mutante uma vantagem de crescimento, sob certas condições ambientais, sendo especialmente úteis em pesquisas genéticas. Quando a seleção não é possível, os mutantes podem ser identificados pela varredura.

**10.2** • As mutações, espontâneas ou induzidas, surgem devido às alterações na sequência de bases do ácido nucleico no genoma de um organismo. Uma mutação pontual, que ocorre devido à alteração de um único par de bases, pode levar à substituição de um único aminoácido em um polipeptídeo, ou não promover qualquer alteração, dependendo do códon em particular. Em uma mutação sem sentido, o códon torna-se um códon de término, levando à síntese de um polipeptídeo incompleto. Deleções e inserções provocam mudanças maiores no DNA, incluindo as mutações de alteração de fase de leitura, que frequentemente resultam na completa perda da função gênica.

**10.3** • Diferentes tipos de mutação ocorrem com diferentes frequências. No caso de uma bactéria típica, taxas de mutação de  $10^{-6}$  a  $10^{-7}$  por pares de quilobases, são frequentemente observadas. Embora as RNA e DNA-polimerases introduzam erros com uma frequência similar, os genomas de RNA geralmente acumulam mutações com frequências muito mais elevadas que os genomas de DNA.

**10.4** • Agentes mutagênicos são agentes químicos, físicos ou biológicos que aumentam a taxa de mutações. Esses agentes podem alterar o DNA de diferentes maneiras. Entretanto, alterações no DNA não são consideradas mutações se não forem herdadas. Alguns tipos de dano ao DNA podem levar a célula à morte se não forem reparados; no entanto, existem sistemas de reparo do DNA propensos a erros e outros de alta fidelidade.

**10.5** • A recombinação homóloga ocorre quando sequências de DNA estreitamente relacionadas de dois elementos genéticos distintos são combinadas em um mesmo elemento. A recombinação é um processo evolutivo importante e as células apresentam mecanismos específicos para garantir sua ocorrência.

**10.6** • Determinados procariotos exibem competência, um estado no qual as células são capazes de captar DNA livre, liberado por outras bactérias. A incorporação do DNA doador em uma célula receptora requer a atividade de uma proteína de ligação ao DNA de fita simples, da proteína RecA e de várias outras enzimas. Somente células competentes são transformáveis.

**10.7** • Transdução é a transferência gênica de uma bactéria hospedeira a outra, mediada por um vírus bacteriano. Na transdução generalizada, partículas virais defectivas incorporam, de maneira aleatória, fragmentos de DNA cromossômico da célula hospedeira, embora com baixa eficiência. Na transdução especializada, o DNA de um vírus temperado sofre uma excisão incorreta, carregando genes adjacentes da célula hospedeira; a eficiência de transdução, nesse caso, pode ser bastante alta.

**10.8** • A conjugação é um mecanismo de transferência de DNA em procariotos, a qual requer o contato célula-célula. A conjugação é controlada por genes carregados por determinados plasmídeos (como o plasmídeo F) e envolve a transferência do plasmídeo de uma célula doadora para uma célula receptora. A transferência do DNA plasmidial envolve a replicação pelo mecanismo de círculo rolante.

**10.9** • O cromossomo da célula doadora pode ser mobilizado e transferido a uma célula receptora. Isso requer um plasmídeo F integrado ao cromossomo, originando um fenotipo Hfr. A transferência do cromossomo hospedeiro é raramente completa, mas pode ser utilizada para mapear a ordem dos genes no cromossomo. Plasmídeos F' são plasmídeos F que se encontravam previamente integrados e que sofreram excisão, capturando alguns genes cromossômicos.

**10.10** • O desenvolvimento de sistemas de transferência gênica em arqueias está muito atrasado, em comparação a bactérias. Muitos antibióticos são ineficientes contra arqueias, dificultando a seleção efetiva de recombinantes. Além disso, as condições de cultivo pouco usuais requeridas por várias arqueias também dificultam a experimentação genética. No entanto, todos os sistemas de transferência genética em bactérias – transformação, transdução e conjugação – são conhecidos em arqueias.

**10.11** • Transposons e sequências de inserção são elementos genéticos que podem mover-se de uma região para outra em uma molécula de DNA hospedeira, por um processo denominado transposição, um tipo de recombinação sítio-específica. A transposição pode ser tanto replicativa quanto conservativa. Frequentemente, os transposons carregam genes que codificam resistência a antibióticos e podem ser utilizados como agentes mutagênicos biológicos.

**10.12** • O sistema CRISPR, é um mecanismo em procariotos para a proteção do genoma contra DNA invasor resultante de infecções ou conjugação. Se pequenas moléculas de RNA provenientes de regiões espaçadoras da região CRISPR ligarem-se por complementariedade ao DNA que entra na célula, proteínas Cas destroem o duplex de ácidos nucleicos.

## REVISÃO DOS TERMOS-CHAVE

**Agente mutagênico** agente que provoca mutação.

**Auxotrófico** organismo que desenvolveu uma necessidade nutricional,

frequentemente em decorrência de uma mutação.

**Conjugação** processo de transferência gênica de uma célula procariótica para

outra por um mecanismo que envolve o contato intercelular.

**Célula Hfr** célula cujo plasmídeo F está integrado ao cromossomo.

**Cístron** um gene, de acordo com a definição do teste *cis-trans*; um segmento de DNA (ou RNA) que codifica uma única cadeia polipeptídica.

**Elemento transponível** elemento genético capaz de mover-se (transpor-se) de um sítio para outro em moléculas de DNA hospedeiro.

**Fenótipo** conjunto de características observáveis de um organismo.

**Genótipo** composição genética completa de um organismo; a descrição completa da informação genética de uma célula.

**Heterodúplex** dupla-hélice de DNA composta por fitas simples de duas moléculas de DNA distintas.

**Linhagem mutadora** linhagem mutante na qual a taxa de mutação é aumentada.

**Linhagem selvagem** linhagem bacteriana isolada da natureza.

**Mutante** organismo cujo genoma carrega uma mutação.

**Mutação** alteração hereditária na sequência de bases do genoma de um organismo.

**Mutação de sentido trocado** mutação na qual um único códon é alterado, de modo que o aminoácido na proteína é substituído por um aminoácido diferente.

**Mutação espontânea** mutação que ocorre “naturalmente”, sem qualquer auxílio

de agentes químicos mutagênicos ou radiação.

**Mutação induzida** mutação provocada por agentes externos, como agentes químicos mutagênicos ou radiação.

**Mutação pontual** mutação que envolve um único par de bases.

**Mutação sem sentido** mutação na qual o códon de um aminoácido é trocado por um códon de término.

**Mutação silenciosa** alteração na sequência do DNA que não tem efeito no fenótipo.

**Plasmídeo** elemento genético extracromossômico, que não apresenta uma forma extracelular.

**Recombinação** processo pelo qual moléculas de DNA, de origens distintas, são trocadas, ou ligadas, em uma única molécula de DNA.

**Regulon** conjunto de genes ou óperons que são transcritos separadamente, mas que são controlados de forma coordenada pela mesma proteína reguladora.

**Replicação por círculo rolante** mecanismo de replicação de um DNA circular de dupla-fita que é iniciado pela clivagem e desenovelamento de uma fita e que utiliza a outra fita (ainda circular) como molde na síntese de DNA.

**Reversão** alteração no DNA que reverte os efeitos de uma mutação prévia.

**Seleção** incubação de organismos sob condições em que o crescimento daqueles contendo um determinado fenótipo ou genótipo será favorecido ou inibido.

**Sequência de inserção (IS)** o tipo mais simples de elemento transponível, que carrega somente os genes envolvidos na transposição.

**Transdução** processo de transferência gênica de uma célula hospedeira para outra, mediado por vírus.

**Transformação** transferência gênica bacteriana, envolvendo DNA livre.

**Transição** mutação em que uma base pirimídica é substituída por outra pirimidina, ou uma purina é substituída por outra purina.

**Transposon** tipo de elemento transponível que carrega outros genes, além daqueles envolvidos na transposição.

**Transversão** mutação em que uma base pirimídica é substituída por uma purina, ou vive-versa.

**Varredura** processo que permite a identificação de organismos pelo fenótipo ou genótipo, mas que não inibe ou aumenta o crescimento de determinados fenótipos ou genótipos.

## QUESTÕES PARA REVISÃO

1. Escreva em uma sentença a definição do termo “genótipo”. Faça o mesmo para “fenótipo”. O fenótipo de um organismo muda automaticamente quando ocorre mudança no genótipo? Em ambos os casos, dê um exemplo para corroborar sua resposta. (Seção 10.1)
2. Explique por que uma linhagem de *Escherichia coli* His<sup>-</sup> é auxotrófica, enquanto uma Lac<sup>-</sup> não é. (Dica: Pense em como *E. coli* metaboliza a histidina e a lactose e para o que é usado cada um desses copostos.) (Seção 10.1)
3. O que são mutações silenciosas? A partir de seus conhecimentos sobre o código genético, por que você acredita que a maioria das mutações silenciosas afeta a terceira posição de um códon (Seção 10.2)?
4. Qual a taxa média de mutação em uma célula? Essa taxa pode ser alterada (Seção 10.3)?
5. Dê um exemplo de um agente mutagênico biológico, um agente químico e um agente físico, descrevendo os mecanismos pelos quais cada um provoca uma mutação (Seção 10.4).
6. O que são regiões heteroduplex no DNA e qual processo leva à sua formação? (Seção 10.5)
7. Explique por que células receptoras não captam plasmídeo com sucesso durante a transformação natural. (Seção 10.6)
8. Explique como as partículas de transdução generalizada diferem das partículas de transdução especializada. (Seção 10.7)
9. O que é um *pilus* sexual e que tipo de célula, F<sup>-</sup> ou F<sup>+</sup>, produz tal estrutura (Seção 10.8)?
10. O que uma célula F<sup>+</sup> precisa fazer antes de transferir seus genes cromossômicos (Seção 10.9)?
11. Explique por que a realização de seleção genética é difícil no estudo de arqueias. Dê exemplos de alguns agentes seletivos que atuam de maneira eficiente com arqueias (Seção 10.10).
12. Quais são as principais diferenças entre sequências de inserção e transposons (Seção 10.11)?
13. Explique por que o DNA invasor reconhecido por pequenas moléculas de RNAs expressas da região CRISPR não podem ser completamente exógeno a célula. (Seção 10.12)

## QUESTÕES APLICADAS

1. Um mutante constitutivo corresponde a uma linhagem que produz continuamente uma proteína que, na linhagem selvagem, é induzível. Descreva duas maneiras pelas quais uma alteração em uma molécula de DNA poderia levar ao desenvolvimento de um mutante constitutivo. Como esses dois tipos de mutantes constitutivos são diferenciados geneticamente?
2. Embora um grande número de agentes químicos mutagênicos seja conhecido, não se conhece qualquer agente químico capaz de induzir mutações em um único gene (mutagênese gene-específica). A partir de seus conhecimentos sobre agentes mutagênicos, explique por que a descoberta de um agente químico mutagênico gene-específico é um evento improvável. Nesse sentido, explique como a mutagênese sítio-dirigida é realizada.
3. Por que é difícil transferir grandes quantidades de genes para uma célula receptora em um único experimento de transformação ou transdução?
4. Elementos transponíveis provocam mutações quando inseridos no interior de um gene. Esses elementos interrompem a continuidade de um gene. Poder-se-ia dizer que os íntrons também interrompem a continuidade de um gene, embora ele permaneça funcional. Explique por que a presença de um íntron em um gene não o inativa, enquanto a inserção de um elemento de transposição o faz.

microbiologia**hoje****De patógeno a tumor assassino**

Os avanços na biotecnologia não são apenas de ponta como eles também fornecem importantes informações sobre a biologia básica da vida e também são os pilares para o aperfeiçoamento de produtos naturais. As técnicas moleculares têm sido utilizadas para manipular a produção de biocombustíveis, culturas resistentes à seca e hormônios, tais como a insulina. Mas e quanto às doenças humanas, como os cânceres; também há esperança?

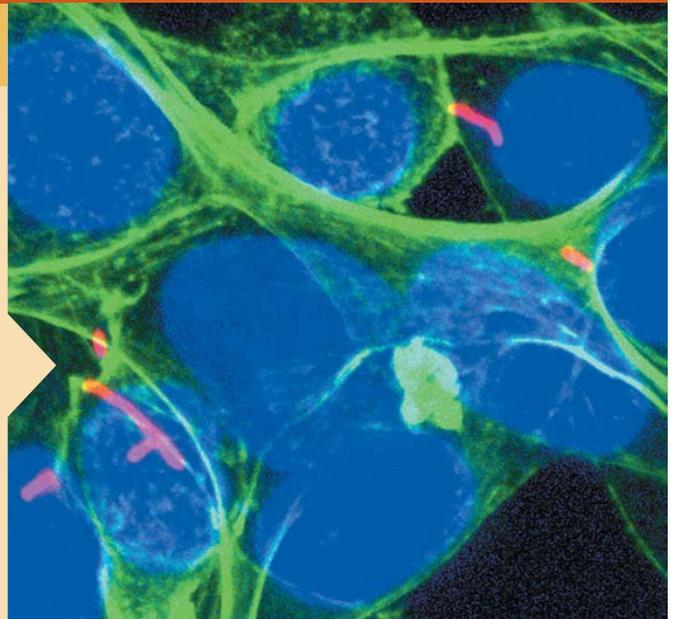
Câncer de pâncreas é uma das principais causas de morte por cânceres. Os atuais tratamentos por radiações e quimioterapias são ineficazes, com taxas desanimadoras de sobrevivência os pacientes. Apesar de fármacos anticâncer estarem disponíveis, eles somente aumentam o tempo de sobrevivência dos pacientes em estágios avançados da doença. Há uma necessidade desesperada de novos tratamentos alternativos e os biotecnologistas estão vindo para socorrer.

*Listeria monocytogenes* é o patógeno agente etiológico da listeriose, uma doença grave veiculada por alimentos. *L. monocytogenes* possui um estilo de vida intracelular que o permite evadir do sistema imune. Entretanto, cientistas descobriram uma linhagem recombinante pouco virulenta que pode ser eliminada pelo sistema imune de células saudáveis, porém não em células tumorais. Isso trouxe a excitante ideia: essa linhagem de *L. monocytogenes* poderia ser usada para carrear agentes anticâncer como os radioisótopos terapêuticos apenas para as células tumorais?

Radioisótopos podem destruir fisicamente as células cancerosas, porém carrear essas moléculas especificamente para as células tumorais pode ser problemático. Usando um engenhoso esquema, cientistas acoplaram radioisótopos de rênio-188 a linhagem de *L. monocytogenes* pouco virulenta.<sup>1</sup> Essa linhagem de *Listeria* (em cor-de-rosa na fotografia) mortal para tumores não só infecta e se multiplica nas células cancerosas de camundongos (em azul), como também reduz a incidência de metástases sem prejudicar as células pancreáticas normais.

Essa pesquisa ilustra como a microbiologia e a biotecnologia podem trabalhar juntas para domar uma bactéria virulenta e convertê-la em um super-herói terapêutico!

<sup>1</sup>Quispe-Tintaya, W., et al. 2013. Nontoxic radioactive *Listeria*<sub>at</sub> is a highly effective therapy against metastatic pancreatic cancer. *Proc. Natl. Acad. Sci.* 110: 8668–8673.



- I Ferramentas e técnicas de engenharia genética 316
- II Clonagem gênica 326
- III Produtos da engenharia genética de microrganismos 333

Neste capítulo, discutiremos as técnicas utilizadas na tecnologia do DNA recombinante, especialmente as usadas para clonar, alterar e expressar genes de maneira eficiente em organismos hospedeiros. As muitas limitações presentes na condução de experimentos genéticos somente *in vivo* (em organismos

vivos) podem ser superadas pela manipulação do DNA *in vitro* (em um tubo de ensaio). Essas técnicas moleculares são as bases da biotecnologia, e até o final desse capítulo, daremos alguns exemplos de como organismos geneticamente modificados podem ter aplicações médicas, agrícolas e industriais.

## I • Ferramentas e técnicas de engenharia genética

A expressão **engenharia genética** refere-se à utilização de técnicas *in vitro* para alterar o material genético em laboratório. Esses materiais alterados podem ser reinseridos no organismo original ou em algum outro organismo. As técnicas básicas de engenharia genética incluem a capacidade de cortar o DNA de interesse em fragmentos específicos e purificar tais fragmentos para manipulações adicionais. Iniciaremos este capítulo abordando algumas das ferramentas básicas de engenharia genética, incluindo enzimas de restrição, separação de ácidos nucleicos por eletroforese, hibridização de ácidos nucleicos, amplificação de DNA e clonagem molecular.

### 11.1 Enzimas de restrição e modificação

Todas as células contêm enzimas que podem modificar quimicamente o DNA de diferentes maneiras. Uma das principais classes dessas enzimas são as *endonucleases de restrição*, ou **enzimas de restrição**. Enzimas de restrição reconhecem sequências de bases específicas no DNA, clivando-o. Embora elas sejam amplamente distribuídas em procariotos (tanto bactérias quanto arqueias), são muito raras em eucariotos. *In vivo*, as enzimas de restrição protegem os procariotos de DNA exógeno hostis, como os genomas virais. Entretanto, enzimas de restrição são também essenciais à manipulação de DNA *in vitro* e sua descoberta deu início à engenharia genética.

#### Mecanismo de ação das enzimas de restrição

Endonucleases de restrição são divididas em três classes principais. As enzimas de restrição do tipos I e III ligam-se ao DNA em seus sítios de reconhecimento, porém clivam o DNA em um local relativamente distante. Por outro lado, as enzimas de restrição do tipo II cortam o DNA no interior da sequência de reconhecimento e, por essa razão, esse grupo de enzimas é muito mais útil à manipulação específica de DNA.

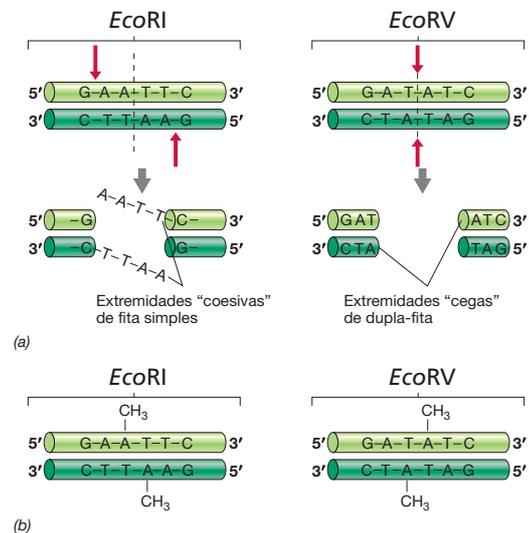
A maioria das sequências de DNA reconhecidas pelas enzimas de restrição do tipo II correspondem a pequenas repetições invertidas de 4 a 8 pares de bases. A **Figura 11.1** apresenta a sequência de 6 pb que é reconhecida e clivada pela enzima de restrição de *Escherichia coli*, denominada *EcoRI* (esse acrônimo refere-se a *Escherichia coli*, linhagem RY13, enzima de restrição I). Os sítios de clivagem estão indicados por setas e os eixos de simetria, por uma linha tracejada. Observe que as duas fitas das sequências de reconhecimento possuem a mesma sequência, se uma for lida a partir da esquerda e a outra, da direita (ou se ambas forem lidas de 5' → 3'). Tais sequências repetidas invertidas são denominadas *palíndromos*. A atividade endonucleásica da *EcoRI* (cujo acrônimo vem de *Escherichia coli* linhagem RY13, enzima de restrição I) faz cortes desiguais nas duas fitas, deixando pequenas porções de

fitas simples, denominadas extremidades "coesivas", nas porções finais dos dois fragmentos. Outras enzimas de restrição, como a *EcoRV*, cortam as duas fitas de DNA de maneira reta, resultando em terminações cegas (**Figura 11.1a**). Como explicado a seguir, esses fragmentos com terminações definidas possuem muitas aplicações, especialmente na clonagem molecular (ver **Figura 11.7**).

Considere as enzimas *EcoRI* e *EcoRV*, que reconhecem sequências específicas de 6 pb (**Figura 11.1**). Qualquer sequência específica de 6 bases pode ocorrer em uma fita de DNA, uma vez a cada 4.096 nucleotídeos (4 bases combinadas em 6, 4<sup>6</sup>), assumindo que o DNA possui uma sequência "randômica". Assim, vários sítios de clivagem de *EcoRI* e *EcoRV* podem estar presentes em uma longa molécula de DNA, como um cromossomo de *E. coli* (ver adiante). As sequências de reconhecimento e os sítios de clivagem de algumas enzimas de restrição estão ilustradas na **Tabela 11.1**.

#### Modificação: proteção contra a restrição

A principal função das enzimas de restrição em procariotos é, provavelmente, proteger a célula da invasão de DNA exó-



**Figura 11.1** Restrição e modificação do DNA. (a) A sequência de DNA reconhecida pela endonuclease de restrição *EcoRI* (painel superior). As setas em vermelho indicam as ligações clivadas pela enzima. A linha tracejada indica o eixo de simetria da sequência. Aparência do DNA após a clivagem da enzima *EcoRI* (Painel inferior). Observe as "extremidades coesivas" de fita simples. (b) A mesma sequência após a modificação pela *EcoRI* metilase. Os grupos metil adicionados pela enzima estão indicados. Eles protegem o sítio de restrição da clivagem por *EcoRI*.

**Tabela 11.1** Sequências de reconhecimento de algumas endonucleases de restrição

Organismo	Nome da enzima <sup>a</sup>	Sequência de reconhecimento <sup>b</sup>
<i>Bacillus globigii</i>	<i>Bgl</i> II	A↓GATCT
<i>Brevibacterium albidum</i>	<i>Bal</i> I	TGG↓C*CA
<i>Escherichia coli</i>	<i>Eco</i> RI	G↓AA*TC <sup>c</sup>
<i>Escherichia coli</i>	<i>Eco</i> RV	GAT↓A*TC <sup>c</sup>
<i>Haemophilus haemolyticus</i>	<i>Hha</i> I	GC*G↓C
<i>Haemophilus influenzae</i>	<i>Hin</i> III	A↓AGCTT
<i>Klebsiella pneumoniae</i>	<i>Kpn</i> I	GGTAC↓C
<i>Nocardia otitidiscaviarum</i>	<i>Not</i> I	GC↓GGC*CGC
<i>Proteus vulgaris</i>	<i>Pvu</i> I	CGAT↓CG
<i>Serratia marcescens</i>	<i>Sma</i> I	CCC↓GGG
<i>Thermus aquaticus</i>	<i>Taq</i> I	T↓CGA*

<sup>a</sup>Nomenclatura: a primeira letra da abreviação de três letras das endonucleases de restrição designa o gênero de onde a enzima se origina, e as outras duas letras, a espécie. O número romano designa a ordem de descoberta das enzimas neste determinado organismo, e outras letras adicionais são designações de linhagens.

<sup>b</sup>As setas indicam os sítios de clivagem enzimática. Os asteriscos indicam o sítio de metilação (modificação). G, guanina; C, citosina; A, adenina; T, timina; Pu, qualquer purina; Py, qualquer pirimidina. Apenas a sequência 5' → 3' é apresentada.

<sup>c</sup>Ver Figura 11.1a.

geno, por exemplo, DNA viral. Se esse DNA estranho penetrar na célula, as enzimas de restrição da célula o destruirão (↔ Seção 8.6). Entretanto, a célula precisa proteger seu próprio DNA da destruição inadvertida por suas próprias enzimas de restrição. Essa proteção é conferida pelas **enzimas de modificação**, que modificam quimicamente nucleotídeos específicos das sequências de reconhecimento do próprio DNA da célula. As sequências modificadas não podem mais ser clivadas pelas enzimas de restrição da célula. Cada enzima de restrição age conjuntamente com uma enzima de modificação correspondente, que compartilha a mesma sequência de reconhecimento. Normalmente, a modificação consiste na metilação de bases específicas no interior da sequência de reconhecimento, de modo que a endonuclease de restrição não pode mais ligar-se. Por exemplo, a sequência reconhecida pelas enzimas de restrição *Eco*RI e *Eco*RV (Figura 11.1a) pode ser modificada pela metilação das duas adeninas mais internas (Figura 11.1b). As enzimas que realizam essa modificação são denominadas *metilases*. Se apenas uma das fitas for modificada, a sequência deixa de ser substrato para as enzimas de restrição *Eco*RI e *Eco*RV.

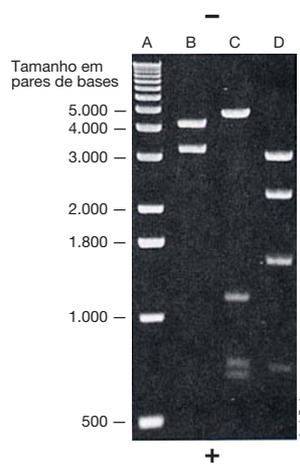
### Eletroforese em gel: separação de moléculas de DNA

A manipulação *in vitro* de ácidos nucleicos frequentemente requer a separação de moléculas de acordo com seu tamanho. Por exemplo, muitas enzimas de restrição cortam as moléculas de DNA em fragmentos que variam de centenas a milhares de pares de base em tamanho. Após a clivagem do DNA, os fragmentos gerados podem ser separados por **eletroforese em gel** e analisados. A eletroforese em gel também é utilizada para verificar o sucesso de uma amplificação de ácidos nucleicos (Seção 11.3).

A eletroforese é um procedimento que separa moléculas carregadas que estão migrando em um campo elétrico. A taxa de migração é determinada pela carga da molécula e por seu tamanho e conformação. Na eletroforese em gel (Figura 11.2a), as moléculas são separadas em um gel poroso, em vez de em uma solução livre. Géis preparados com *agarose*, um polisacarídeo, são utilizados para a separação de fragmentos de DNA. Quando uma corrente elétrica é aplicada, os ácidos nucleicos movem-se por meio do gel, em direção ao eletrodo positivo, devido a seus grupos fosfato carregados negativamente. A presença da malha do gel dificulta o progresso do DNA e as moléculas pequenas ou compactas migram mais rapidamente do que moléculas grandes. Quanto maior a concentração de agarose no gel, maior será a resistência ao movimento de moléculas maiores. Consequentemente, géis de diferentes concentrações são utilizados para separar diferentes intervalos de tamanhos.



(a)



(b)

**Figura 11.2** Eletroforese de DNA em gel de agarose. (a) Amostras de DNA são aplicadas em poços de um gel de agarose submerso. (b) Fotografia de um gel de agarose corado. O DNA foi aplicado nos poços situados no topo do gel (polo negativo), conforme ilustrado, enquanto o polo positivo do campo elétrico localiza-se na parte inferior. A amostra-padrão no poço A possui fragmentos de tamanho conhecido. Usando esses padrões, pode-se determinar os tamanhos dos fragmentos das outras amostras. Bandas coradas menos intensamente na parte inferior do gel são decorrentes dos tamanhos menores desses fragmentos, havendo menos DNA para ser corado.

Após um determinado tempo, o gel pode ser corado com um composto que se liga ao DNA, como *brometo de etídio*, que promove a fluorescência laranja do DNA sob luz ultravioleta (Figura 11.2b). Para determinar o tamanho do DNA de interesse, a migração pode ser comparada a um padrão de tamanho molecular que consiste em fragmentos de DNA com tamanhos conhecidos. Os fragmentos de DNA podem ser purificados a partir do gel e utilizados com diferentes propósitos.

#### MINIQUESTIONÁRIO

- Por que as enzimas de restrição são úteis para os biólogos moleculares?
- Qual é a base para a separação de moléculas por eletroforese?

## 11.2 Hibridização de ácidos nucleicos

Quando o DNA é desnaturado (ou seja, as duas fitas são separadas), as fitas simples podem formar moléculas de fita dupla com outras moléculas de DNA (ou RNA) de fita simples, originando moléculas híbridas que têm sequências de bases complementares (ou quase complementares) (↔ Seção 4.2). Esse processo é denominado *hibridização de ácido nucleico*, ou simplesmente **hibridização**, sendo amplamente utilizado na detecção, caracterização e identificação de segmentos de DNA ou RNA.

Segmentos previamente identificados de ácido nucleico de fita simples, e que são usados em hibridização, são denominados **sondas de ácido nucleico**, ou simplesmente *sondas*. Para permitir a detecção, as sondas podem ser radioativas ou marcadas com compostos químicos coloridos ou que geram produtos fluorescentes (↔ Seção 18.4). Ao variar as condições de hibridização, é possível ajustar o grau de estricção da hibridização de tal forma que o pareamento complementar de bases devem ser quase exato; isso ajuda a evitar o pareamento entre sequências inespecíficas de ácidos nucleicos que sejam apenas parcialmente complementares.

### Southern e Northern blot

A hibridização pode ser bastante útil na detecção de sequências relacionadas em diferentes genomas ou outros elementos genéticos, ou se um gene específico está sendo expresso e o RNA transcrito. No *Southern blotting*, denominado assim em homenagem ao seu inventor, E.M. Southern, sondas de sequência conhecida são hibridizadas com fragmentos de DNA que foram separados por eletroforese em gel (Seção 11.1). Quando o DNA encontra-se no gel, e RNA ou DNA são utilizados como sonda, o procedimento de hibridização é denominado **Southern blot**. Por outro lado, quando o RNA encontra-se no gel, e DNA ou RNA são utilizados como sonda, o procedimento é denominado **Northern blot**.

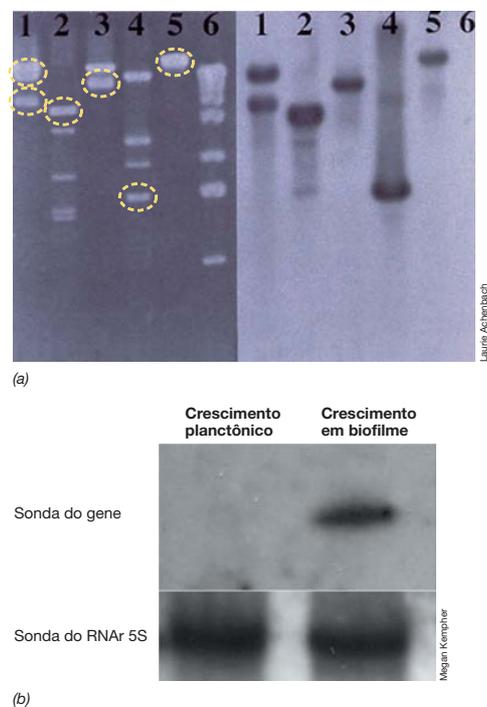
Em um *Southern blot*, os fragmentos de DNA presentes no gel são desnaturados para gerar fitas simples, sendo, então, transferidos para uma membrana sintética. Apesar do RNA ser de fita simples, agentes desnaturantes são adicionados ao gel para evitar a formação de estruturas secundárias (↔ Seção 4.7). A membrana é em seguida exposta a uma sonda marcada. Se a sonda for complementar a algum dos fragmentos, formam-se híbridos e a sonda liga-se à membrana nos locais onde os fragmentos complementares estão situados. A hibridização pode ser detectada pela análise da marcação ligada à membrana. A Figura 11.3a ilustra como um *Southern*

*blot* pode ser utilizado para identificar fragmentos de DNA contendo sequências que hibridizam com a sonda.

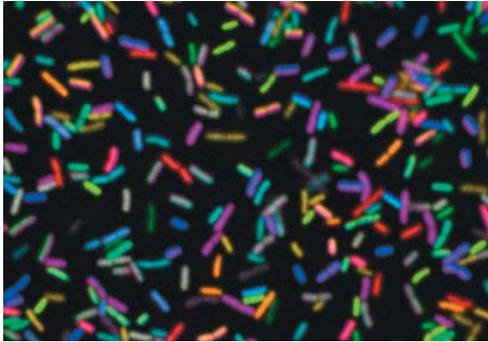
O procedimento de *Northern blot* é semelhante, exceto que moléculas de RNA em vez de DNA são separadas em um gel e transferidas para uma membrana sintética onde são marcadas com sondas. *Northern blot* é também utilizado para identificar RNA mensageiros (RNAm) derivados de genes específicos. A intensidade de um alvo no *Northern Blot* dá uma estimativa aproximada da abundância de RNAm do gene-alvo e pode ser usado para monitorar a transcrição (Figura 11.3b).

### Outros métodos de hibridização

A hibridização é frequentemente utilizada para detectar a presença de genes específicos no genoma que ainda não tenham sido sequenciados, bem como o movimento de elementos genéticos, tais como os transposons (↔ Seção 10.11). Para localizar o local específico de um gene de interesse no genoma, o



**Figura 11.3** Hibridização de ácidos nucleicos. (a) *Southern blotting*. (Painel esquerdo) Moléculas de DNA purificado de vários diferentes plasmídios tratados com enzimas de restrição e depois submetidos a eletroforese em gel de agarose. (Painel direito) Transferência do gel de DNA mostrado à esquerda. Após a transferência, o DNA no gel é hibridizado com uma sonda radioativa. A posição das bandas é visualizada por autorradiografia de raios X. Observe que apenas alguns dos fragmentos de DNA (circulados de amarelo) possuem sequências complementares a sonda marcada. A canaleta 6 contém DNA utilizado como padrão de tamanho molecular e nenhuma das bandas hibridiza com as sondas. (b) *Northern blotting*. (Painel superior) Hibridização e detecção de uma sonda específica de um gene em um ensaio de transferência com o RNA total. A sonda liga-se apenas ao RNA das células crescendo em biofilme, indicando que o gene-alvo não é expresso durante o crescimento planctônico (suspensão). (Painel inferior) Hibridização e detecção de uma sonda radioativa correspondente ao RNAr 5S do mesmo ensaio. A intensidade do sinal indica que foram aplicados no gel quantidades iguais do RNA de cada amostra.



**Figura 11.4** Imagem espectral de fluorescência de 28 diferentes linhagens de *Escherichia coli* marcadas. As células foram marcadas com uma combinação de oligonucleotídeos complementares ao RNAr 16S de *E. coli*, conjugados com fluoróforos.

DNA genômico total pode ser clonado (Seção 11.4). A hibridização das colônias resultantes usando uma sonda de ácido nucleico pode detectar DNA-recombinante nas colônias, como mostrado na Figura 11.8a. Esse procedimento emprega também o *plaqueamento de réplicas*, originando uma duplicata da placa matriz em uma membrana filtrante.

As células presentes no filtro são lisadas em seus locais originais a fim de liberarem seu DNA, sendo o filtro, então, tratado de modo a converter o DNA na forma de fita simples, o qual é fixado ao filtro. Esse filtro é então exposto a uma sonda de ácido nucleico marcada para permitir a hibridização, sendo a sonda não ligada removida por lavagem. O filtro é exposto a um filme de raios X, caso uma sonda radioativa tenha sido utilizada. Após a revelação, o filme de raios X é examinado quanto à presença de sinais. Eles correspondem aos locais, na membrana, onde a sonda radioativa se hibridizou ao DNA de uma determinada colônia. As colônias correspondentes a esses sinais serão, em seguida, selecionadas e estudadas mais detalhadamente.

A hibridização também é a base do método FISH (sigla que em inglês significa hibridização fluorescente *in situ*) (Seção 18.4; Figura 11.4). Usando essa técnica, uma variedade de sinais fluorescentes pode ser covalentemente ligada a sondas de *oligonucleotídeos* (que são pequenas moléculas de fita simples de DNA ou RNA) para uma sequência de DNA alvo específica. Essas sondas podem ser usadas para identificar espécies de bactérias em particular ou linhagens por hibridização com sequências características presentes nos seus genes para o RNA ribossomal 16S ou diretamente com o RNA ribossomal. Esta abordagem permite a identificação de patógenos em amostras clínicas ou bactérias de interesse em amostras ambientais. A Figura 11.4 demonstra o uso simultâneo de oito combinações diferentes de sondas de oligonucleotídeos para distinguir entre 28 diferentes linhagens de *Escherichia coli* cujas sequências de RNAr 16S variam entre as linhagens. As variações na cor dão uma indicação visual da especificidade e potência de sondas de ácidos nucleicos.

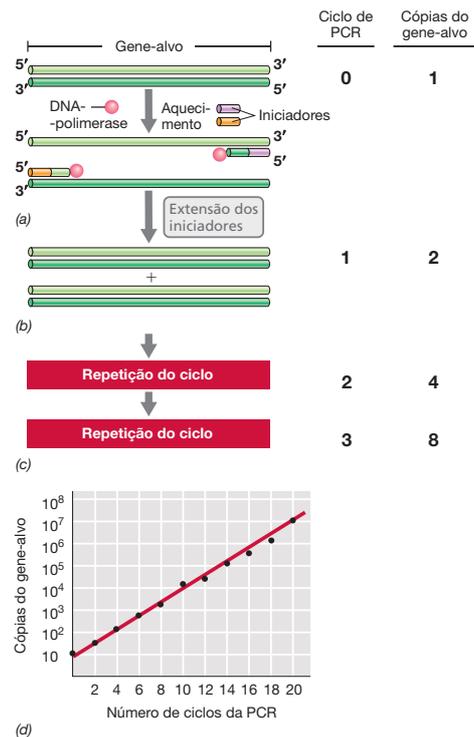
**MINIQUESTIONÁRIO**

- O que é um *Southern blot* e o que ele revela?
- Qual a diferença entre um *Northern blot* e um *Southern blot*?

**11.3 Amplificação de DNA: a reação em cadeia de polimerase**

Cópias de uma sequência específica de DNA são necessárias em diferentes procedimentos moleculares, e a **reação em cadeia de polimerase (PCR)** é o método pelo qual cópias são sintetizadas *in vitro*. A reação em cadeia de polimerase pode copiar segmentos de DNA bilhões de vezes em um tubo de ensaio, um processo denominado *amplificação*, que gera grandes quantidades de genes específicos ou outros segmentos de DNA, os quais podem ser utilizados em um grande número de aplicações em biologia molecular. A PCR utiliza a enzima DNA-polimerase, que naturalmente produz cópias de moléculas de DNA (Seção 4.4). Para iniciar a síntese são utilizados iniciadores oligonucleotídicos artificialmente sintetizados (Seção 11.5), feitos de DNA (em vez de RNA como os iniciadores usados pelas células). Na realidade, a PCR não produz cópias de moléculas inteiras de DNA, promovendo a amplificação de segmentos de até alguns milhares de pares de bases (*o alvo*) a partir de moléculas maiores de DNA (*o molde*).

As etapas em uma amplificação de DNA por PCR podem ser resumidas da forma a seguir (Figura 11.5):



**Figura 11.5** A reação em cadeia de polimerase (PCR). A PCR amplifica sequências específicas de DNA. (a) O DNA-alvo é aquecido para separar as fitas; um grande excesso dos dois oligonucleotídeos iniciadores, um complementar à fita-alvo e outro à fita complementar, são adicionados juntamente com a DNA-polimerase. (b) Após o anelamento dos iniciadores, a extensão deles gera uma cópia do DNA original de PCR que gera 4 e 8 cópias da sequência de DNA original de dupla-fita. (c) Dois ciclos adicionais respectivamente. (d) Efeito da realização de 20 ciclos de PCR, em uma preparação de DNA contendo originalmente 10 cópias do gene-alvo. Observe que o gráfico é semilogarítmico.

1. Dois iniciadores oligonucleotídicos, que flanqueiam o DNA-alvo são adicionados em grande excesso ao DNA-molde que é desnaturado termicamente.
2. À medida que a mistura é resfriada, o excesso de iniciadores complementares ao DNA-molde garante que a maioria das fitas-molde irá parear-se com um iniciador, em lugar de parearem-se umas com as outras (Figura 11.5a).
3. A DNA-polimerase então promove a extensão dos iniciadores, utilizando o DNA original como molde (Figura 11.5b).
4. Após um período de incubação adequado, a mistura é novamente aquecida para separar as fitas. A mistura é então resfriada, para permitir a hibridização dos iniciadores nas regiões complementares dos DNA recém-sintetizados, sendo todo o processo repetido (Figura 11.5c).

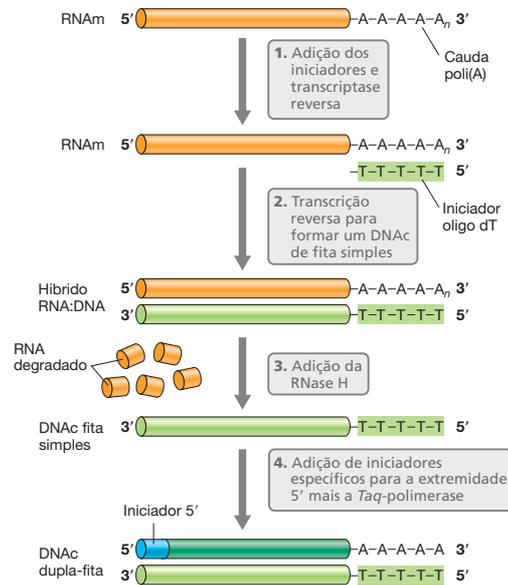
O poder da PCR baseia-se no fato de os produtos de uma extensão a partir de um iniciador atuarem como moldes no próximo ciclo, de modo que cada ciclo duplica a quantidade do DNA-alvo original. Na prática, são realizados geralmente 20 a 30 ciclos, promovendo um aumento de  $10^6$  a  $10^9$  vezes na quantidade da sequência-alvo (Figura 11.5d). Como a técnica consiste em várias etapas altamente repetitivas, as máquinas de PCR, denominadas *termocicladores*, realizam ciclos automáticos de aquecimento e resfriamento. Usando iniciadores específicos de 15 nucleotídeos ou mais e temperaturas de anelamento elevadas, a PCR é tão específica que quase não há pareamento inespecífico, sendo o DNA amplificado praticamente homogêneo.

### Polimerases e PCR em altas temperaturas

Devido às altas temperaturas necessárias para a desnaturação das cópias de fitas duplas de DNA *in vitro*, uma enzima DNA-polimerase termoestável isolada da bactéria termófila de fontes termais *Thermus aquaticus* (↔ Seção 15.20) é utilizada. A DNA-polimerase de *T. aquaticus*, denominado *Taq-polimerase*, é estável a 95°C, não sendo afetado durante a etapa de desnaturação empregada na reação de PCR. A DNA-polimerase de *Pyrococcus furiosus*, um hipertermófilo, com crescimento ótimo a 100°C (↔ Seção 16.4), é denominada *Pfu-polimerase*, sendo ainda mais estável do que a *Taq-polimerase*. Além disso, contrariamente à *Taq-polimerase*, a *Pfu-polimerase* apresenta atividade de revisão (↔ Seção 4.6), o que a torna uma enzima particularmente útil quando a precisão é essencial. Desse modo, a taxa de erro da *Taq-polimerase* sob condições normais é de  $8,0 \times 10^{-6}$  (por base duplicada), enquanto a da *Pfu-polimerase* é de somente  $1,3 \times 10^{-6}$ . Para suprir a demanda de DNA-polimerases termoestáveis para os mercados de PCR e de sequenciamento de DNA, os genes dessas enzimas foram clonados em *E. coli* e produzidos comercialmente em larga escala.

### Aplicações da PCR

A PCR é extremamente útil na obtenção de DNA para a clonagem e o sequenciamento, uma vez que o gene ou os genes de interesse podem ser facilmente amplificados se suas sequências flanqueadoras forem conhecidas. A PCR é também utilizada rotineiramente em estudos comparativos ou filogenéticos envolvendo a amplificação de genes de diferentes origens. Nesses casos, são sintetizados iniciadores dirigidos para regiões do gene que exibem sequências conservadas em uma grande variedade de organismos. Por exemplo, como o RNAr 16S, uma molécula utilizada em análises filogenéticas (↔ Seção 12.5), possui regiões altamente conservadas e regiões altamente variáveis, ini-



**Figura 11.6** PCR transcrição reversa. Etapas na síntese do DNAc a partir de um RNAm eucarioto. A transcriptase reversa sintetiza uma molécula híbrida contendo tanto RNA quanto DNA utilizando o RNAm como molde e iniciadores oligo-T como substrato. Depois, a enzima RNase H hidroliza a porção de RNA da molécula híbrida gerando uma molécula de fita simples de DNAc complementar (DNAc). Após a adição de iniciadores complementares à extremidade de 5' do DNAc, a *Taq*-polimerase produz um DNAc de fita dupla.

ciadores específicos para o gene de RNAr 16S de bactérias e arqueias podem ser sintetizados e utilizados para a detecção desses organismos em um habitat específico. Além disso, se iniciadores com uma maior especificidade forem usados, somente alguns subgrupos pertencentes a cada domínio serão identificados. Essa técnica é intensamente utilizada em ecologia microbiana e revelou a enorme diversidade do mundo microbiano, sendo muitos de seus representantes ainda não cultivados (↔ Seção 18.5).

Devido a sua extrema sensibilidade, a PCR pode ser utilizada na amplificação de quantidades muito pequenas de DNA. Por exemplo, a PCR tem sido utilizada para amplificar e clonar DNA de diferentes origens, como restos humanos mumificados e plantas e animais fossilizados. Pela sua capacidade de amplificar e analisar DNA de misturas celulares, a PCR também tornou-se uma ferramenta comum na microbiologia diagnóstica (↔ Seção 27.10). Por exemplo, a PCR é utilizada como ferramenta forense que permite a identificação de seres humanos a partir de amostras muito pequenas de seu DNA.

Dependendo do objetivo molecular, variações no procedimento de PCR foram desenvolvidas. A **transcrição reversa-PCR (RT-PCR)** pode ser usada para gerar DNA a partir de um molde de RNAm (Figura 11.6). Esse procedimento pode ser usado para monitorar se um gene é expresso ou para produzir um gene eucarioto sem íntrons que pode ser expresso em bactérias como descrito (Seção 11.11) para os hormônios insulina e somatotrofina. A RT-PCR consiste na utilização da enzima dos retrovírus, *transcriptase reversa*, para gerar uma cópia de *DNA complementar* (DNAc) (↔ Seção 9.11). Para quantificar os níveis iniciais de DNA-alvo ou RNA na amostra, o procedimento de *PCR quantitativa (PCRq)* pode ser usado. Essa

técnica utiliza sondas fluorescentes para monitorar durante o processo de amplificação (↔ Figuras 27.18 e 27.19).

A Figura 11.6 ilustra como a transcriptase reversa sintetiza uma fita de DNAc usando uma molécula de RNA como molde. Nesses casos, um iniciador complementar à extremidade 3' do transcrito-alvo é usado pela enzima transcriptase reversa para iniciar a síntese de RNA. Se o molde for um RNAm eucarioto, um iniciador complementar a cauda poli(A) (↔ Seção 4.9) do RNAm pode ser utilizado. A atividade da transcriptase reversa resulta em um ácido nucleico híbrido contendo tanto DNA quanto RNA. RNaseH, uma ribonuclease específica para essas moléculas híbridas, hidrolisa o RNA, restando somente o DNAc como molde para uma PCR convencional usando iniciadores adicionais complementares a extremidade 5'. Modificações desse procedimento podem ser feitas caso a extremidade 5' do RNAm não seja conhecida.

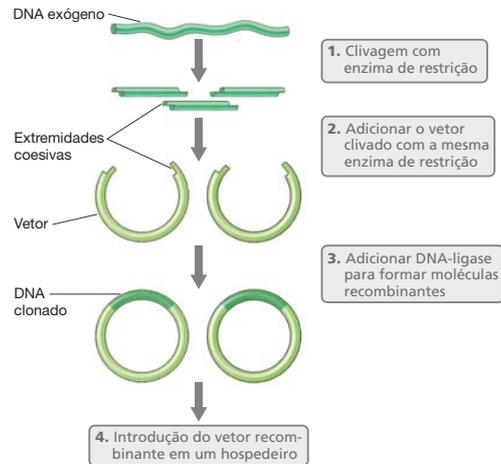
#### MINIQUESTIONÁRIO

- Por que um iniciador deve estar presente em cada extremidade do segmento de DNA a ser amplificado?
- A partir de que organismos as DNA-polimerases termoestáveis são obtidas?
- Em que a RT-PCR difere da PCR-padrão?

### 11.4 Fundamentos da clonagem molecular

Em uma **clonagem molecular**, um fragmento de DNA é isolado e replicado. A estratégia básica da clonagem molecular envolve o isolamento do gene de interesse (ou outro segmento de DNA) a partir de seu local de origem e sua transferência para um elemento genético pequeno e simples, como um plasmídeo ou um vírus, que é denominado **vetor** (Figura 11.7). Clonagem molecular resulta no **DNA recombinante**, uma molécula de DNA que contém duas ou mais origens. Quando o vetor se replica, o DNA clonado carregado por ele é também replicado. Assim, a clonagem molecular inclui a localização do gene de interesse, a obtenção e a purificação de uma cópia do gene e sua inserção em um vetor conveniente. Uma vez clonado com sucesso, o gene de interesse pode ser manipulado de diversas maneiras, podendo eventualmente ser reintroduzido em uma célula viva. Essa abordagem fornece a base de grande parte da tecnologia do DNA recombinante, tendo facilitado sobremaneira a análise detalhada de genomas.

A clonagem molecular tem como objetivo o isolamento de cópias de genes específicos, em sua forma pura. Considere a natureza do problema. Para um organismo geneticamente “simples” como *Escherichia coli*, um gene é codificado, em média, por um DNA de 1 a 2 kb, presente em um genoma de mais de 4.600 kpb (↔ Seção 4.3). Um gene de *E. coli* corresponde, em média, a menos de 0,05% do DNA total da célula. No DNA humano, o problema é muito mais complicado porque, embora as regiões codificadoras de genes não sejam, em média, maiores do que em *E. coli*, os genes são normalmente separados em fragmentos, e o genoma é cerca de 1.000 vezes maior! Como, então, pode-se obter um gene específico? Felizmente, nosso conhecimento sobre a química e enzimologia do DNA nos permitem clivar e ligar, além de replicar moléculas de DNA *in vitro*. Dessa forma, enzimas de restrição, DNA-ligase, a reação de polimerização em cadeia e DNA sintético são importantes ferramentas utilizadas na clonagem molecular.



**Figura 11.7** Etapas principais da clonagem gênica. O vetor pode ser um plasmídeo ou um genoma viral. Clivando-se o DNA exógeno e o vetor com a mesma enzima de restrição, são geradas extremidades coesivas complementares, que permitem a inserção do DNA no vetor.

#### Etapas na clonagem gênica: um resumo

A lista a seguir enumera a sequência de eventos em uma clonagem gênica:

1. **Isolamento e fragmentação do DNA de origem.** O DNA original pode ser DNA genômico de um organismo de interesse. DNA sintetizado de um molde de RNA pela transcriptase reversa (Seção 11.3), um gene ou genes amplificados pela reação de polimerização em cadeia (Seção 11.3) ou mesmo DNA totalmente sintético produzido *in vitro* (Seção 11.5). Se o DNA for genômico, tem que ser inicialmente clivado com enzimas de restrição (Seção 11.1) para gerar uma mistura de fragmentos com tamanho de fácil manuseio (Seção 11.7).
2. **Inserção do fragmento de DNA em um vetor de clonagem.** Vetores de clonagem são pequenos elementos genéticos de replicação independente, utilizados para replicar os genes. A maioria dos vetores é derivada de plasmídeos ou vírus. Os vetores de clonagem são normalmente projetados para permitir a inserção *in vitro* de um DNA exógeno em um sítio de restrição que cliva o vetor, sem afetar a sua replicação (Figura 11.7). Quando o DNA original e o vetor são clivados com uma enzima de restrição que gera extremidades coesivas, a ligação das duas moléculas é extremamente facilitada pelo pareamento das extremidades coesivas. As extremidades abruptas geradas por algumas enzimas de restrição podem ser ligadas diretamente, ou pela utilização de conectores ou adaptadores de DNA sintético. Em ambos os casos, as fitas são ligadas pela *DNA-ligase*, uma enzima que liga covalentemente as fitas do vetor e do DNA inserido. Se a origem do DNA for produto de PCR, a DNA-ligase é usada para ligar o DNA amplificado ao vetor especializado (ver Figura 11.15).
3. **Introdução do DNA clonado em um organismo hospedeiro.** As moléculas de DNA recombinante produzidas em um tubo de ensaio são introduzidas em organismos hospedeiros adequados, onde podem ser replicadas. A transformação (↔ Seção 10.6) é frequentemente uti-

lizada para inserir o DNA recombinante em células. Na prática, esse procedimento gera uma mistura de construções recombinantes. Algumas células contêm o gene clonado de interesse, ao passo que outras podem conter outros genes clonados a partir do mesmo DNA original. Esse tipo de mistura é conhecido como **biblioteca (genômica) de DNA**, porque muitos clones diferentes podem ser purificados a partir dessa mistura, cada um contendo diferentes segmentos de DNA clonados a partir do organismo original. Quando uma biblioteca genômica é construída pela clonagem de fragmentos genômicos aleatórios, o processo é denominado **clonagem shotgun**, que é amplamente utilizado em análises genômicas, como descrito na Seção 11.15 para *mineração genômica*.

### Encontrando o clone correto

Os trabalhos de engenharia genética iniciam-se pela clonagem de um gene de interesse. Mas primeiro é necessário identificar qual colônia do hospedeiro contém o clone correto. Pode-se isolar células hospedeiras contendo um vetor plasmidial pela seleção de um marcador presente no vetor, como a resistência a antibióticos, de modo que somente tais células originarão colônias. No caso de células hospedeiras contendo vetores virais, a simples detecção de placas de lise é suficiente (↔ Seção 8.4). Essas colônias ou placas de lise podem também ser analisadas quanto à presença de vetores contendo DNA exógeno inserido, verificando-se a ocorrência de inativação de um gene do vetor (Seção 11.7). Para a clonagem de um único fragmento de DNA gerado por PCR ou purificado por outros métodos, essas seleções ou varreduras simples geralmente são suficientes.

Uma biblioteca genômica é composta por milhares ou dezenas de milhares de clones e, normalmente, somente um ou poucos deles corresponderão aos genes de interesse. Assim, identificar células contendo o DNA exógeno clonado é apenas a primeira etapa. O maior desafio continua sendo encontrar o clone que possui o gene de interesse. É necessário examinar colônias bacterianas ou placas de lise de células infectadas por vírus crescendo em meio sólido e detectar aquelas poucas que apresentam o gene de interesse. Isso pode ser feito por digestão do DNA com enzimas de restrição ou sequenciamento de plasmídeos extraídos de um grande número de colônias. Outra abordagem é usar a hibridização descrita na Seção 11.2 e como representado na **Figura 11.8a**.

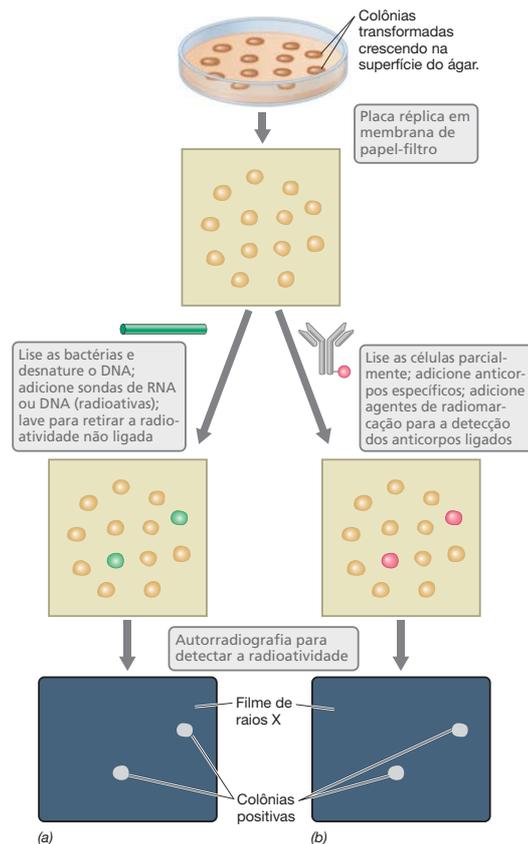
### Genes exógenos expressos no hospedeiro de clonagem

Quando um gene exógeno é expresso no hospedeiro de clonagem, a proteína codificada pode ser procurada em colônias da célula hospedeira. Obviamente, para isso, o hospedeiro em si não deve produzir a proteína em estudo. Assim, quando o gene exógeno é incorporado, sua expressão pode ser detectada. Isso torna a seleção de células contendo os genes clonados relativamente simples, especialmente se a expressão dos genes clonados for de fácil observação.

Anticorpos podem ser utilizados como reagentes para a detecção de uma proteína de interesse. Anticorpos são proteínas do sistema imune que se ligam de maneira altamente específica a uma molécula-alvo, o antígeno (↔ Seção 24.4). Quando anticorpos são utilizados para a detecção, a proteína codificada pelo gene clonado corresponde ao antígeno e é uti-

lizada para a produção de um anticorpo em um animal experimental. Como os anticorpos ligam-se especificamente ao antígeno, quando o antígeno se encontra presente em uma ou mais colônias em uma placa, a localização dessas colônias pode ser determinada observando-se a ligação do anticorpo. Visto que somente uma pequena quantidade da proteína (antígeno) é encontrada nas colônias, apenas uma pequena quantidade de anticorpos liga-se e, desse modo, um procedimento altamente sensível para a detecção de anticorpos ligados deve ser empregado. Na prática, tal procedimento baseia-se no uso de radioisótopos, produtos químicos fluorescentes ou enzimas. Essas e outras técnicas extremamente sensíveis para a detecção de antígenos serão discutidas no Capítulo 27.

Esse método de detecção usando anticorpos é ilustrado na Figura 11.8b. O método de *plaqueamento de réplica* é utilizado para duplicar a placa matriz em um pedaço de filtro de membrana sintética, onde todas as manipulações subsequentes serão feitas nesse filtro. As colônias duplicadas são lisadas para liberar o antígeno de interesse. Em seguida, o anticorpo é adicionado e reage com o antígeno. Os anticorpos não ligados são removidos por lavagem e um reagente radioativo específico para o anticorpo



**Figura 11.8 Encontrando o clone correto.** (a) Métodos para detecção de clones recombinantes por hibridização de colônias com sondas radioativas de ácidos nucleicos. Formação de uma dupla-hélice de DNA liga a sonda de DNA a um ponto específico na membrana. (b) Método para detecção da produção de proteínas utilizando anticorpos específicos contendo um marcador fluorescente ou radioativo.

é adicionado. Uma folha de filme de raios X é depositada sobre o filtro e exposta. Caso uma colônia radioativa esteja presente, será observado um sinal no filme de raios X, após a sua revelação (Figura 11.8a). A localização desse sinal no filme corresponde à localização da colônia que produz a proteína, na placa matriz. Essa colônia pode ser coletada da placa matriz e subcultivada.

A principal limitação desse procedimento é o fato de ser necessária a disponibilidade do anticorpo específico para a proteína em questão. Anticorpos podem ser produzidos pela inoculação do antígeno proteico específico em um animal. Todavia, para que o procedimento seja bem-sucedido, a proteína injetada deve estar pura, caso contrário, anticorpos com várias especificidades serão produzidos. Dessa maneira, é necessária a purificação prévia da proteína, ou reações falso-positivas poderão dificultar a seleção dos clones.

#### MINIQUESTIONÁRIO

- Qual o propósito da clonagem molecular?
- Qual o papel dos vetores de clonagem, enzimas de restrição e DNA-ligase na clonagem molecular?
- Como os genes clonados podem ser identificados?

### 11.5 Métodos moleculares de mutagênese

Como já discutimos anteriormente, os agentes mutagênicos convencionais introduzem mutações *aleatórias* em um organismo intacto (⇔ Seção 10.4). Por outro lado, a mutagênese *in vitro*, mais conhecida como **mutagênese sítio-dirigida**, utiliza técnicas de DNA sintético e de clonagem de DNA para introduzir mutações em genes *em sítios precisamente determinados*. Além de modificar apenas algumas poucas bases, as mutações podem também ser manipuladas por engenharia genética, visando a inserção de grandes segmentos de DNA em locais precisamente determinados.

#### Síntese de DNA

Segmentos de DNA podem ser sintetizados artificialmente e utilizados como iniciadores ou sondas para reações em cadeia de polimerase, hibridização, ou podem fornecer versões alteradas de partes de genes ou regiões reguladoras. Oligonucleotídeos de 12 a 40 bases estão disponíveis comercialmente e oligonucleotídeos com mais de 100 bases de comprimento podem ser gerados, quando necessário. Também é possível sintetizar genes inteiros, se esses codificarem para proteínas pequenas (menos de 600 pb) como é o caso das subunidades da insulina (Seção 11.11).

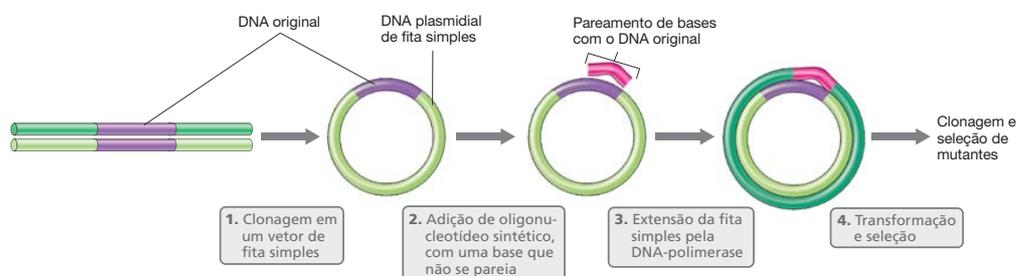
O DNA é sintetizado *in vitro* em um procedimento de fase sólida, no qual o primeiro nucleotídeo da cadeia é ligado a um suporte insolúvel (como as esferas de vidro porosas). Várias etapas são necessárias para a adição de cada nucleotídeo, sendo a química do processo complexa. Após o término de cada etapa, a mistura de reação é eluída do suporte sólido, sendo a série de reações repetida para a adição do próximo nucleotídeo. Uma vez que o oligonucleotídeo apresenta o tamanho desejado, é clivado do suporte de fase sólida pela ação de um reagente específico, e purificado para eliminar subprodutos e contaminantes.

#### Mutagênese sítio-dirigida

A mutagênese sítio-dirigida é uma ferramenta poderosa, pois permite a alteração de qualquer base em um gene específico e apresenta, portanto, várias aplicações na genética. Por alterar sequências gênicas promovendo mudança na sequência de aminoácidos, a mutagênese sítio-dirigida pode ser usada para manipular características de proteínas como atividade enzimática ou afinidade de ligação a proteínas (Seção 11.12). O procedimento básico consiste na síntese de um pequeno iniciador oligonucleotídico de DNA contendo a alteração de base desejada (mutação), o qual é então pareado com uma fita simples de DNA contendo o gene-alvo. O pareamento é completo, exceto na pequena região de pareamento incorreto. Em seguida, o oligonucleotídeo sintético é estendido pela DNA-polimerase, copiando, assim, o restante do gene. A molécula de dupla-fita resultante é inserida em uma célula hospedeira por transformação. Os mutantes são frequentemente selecionados por algum tipo de seleção positiva, como a resistência a antibiótico; nesse caso, o DNA modificado também deve possuir, nas adjacências, um marcador de resistência a antibiótico.

A Figura 11.9 ilustra um procedimento de mutagênese sítio-dirigida. O processo inicia com a clonagem do gene-alvo em um vetor plasmidial. O vetor dupla-fita é desnaturado produzindo fita simples de DNA que permite a ligação do oligonucleotídeo por pareamento de bases com o gene-alvo. Após a extensão pela DNA-polimerase, a molécula de DNA contém uma fita com bases malpareadas. Após a transformação em uma célula hospedeira o vetor de DNA replica, a célula divide-se, e as duas moléculas-filhas possuirão pareamento completo, porém uma vai carrear a mutação e a outra será igual ao selvagem. As progênes bacterianas são rastreadas para as que têm a mutação.

Mutagênese sítio-dirigida pode também ser conduzida por meio de PCR. Neste caso, um pequeno oligonucleotídeo de DNA com a mutação requerida é utilizado como iniciador



**Figura 11.9** Mutagênese sítio-dirigida utilizando DNA sintético. Pequenos oligodesoxirribonucleotídeos sintéticos podem ser utilizados para gerar

mutações. A clonagem em um plasmídeo fornece o DNA fita simples necessário à realização da mutagênese sítio-dirigida.

na PCR. O iniciador contendo a mutação é projetado para parear com o alvo com a base malpareada flanqueada por um número suficiente de nucleotídeos correspondentes para que seja estável durante a reação de PCR. O iniciador mutante é pareado com um iniciador normal, e quando a reação de PCR amplifica o DNA-alvo, a(s) mutação(ões) é(ão) incorporada(s) no produto final amplificado.

### Aplicações da mutagênese sítio-dirigida

A mutagênese sítio-dirigida pode ser utilizada para avaliar a atividade de proteínas contendo substituições conhecidas de aminoácidos. Suponha que a importância de determinados aminoácidos no sítio ativo de uma enzima esteja sendo estudada. A mutagênese sítio-dirigida poderia ser empregada para modificar um determinado aminoácido na enzima, sendo a enzima agora modificada, analisada e comparada à enzima selvagem. Nesse tipo de experimento, o vetor que codifica a enzima mutante poderia ser inserido em uma linhagem hospedeira mutante, incapaz de produzir a enzima original. Consequentemente, a atividade avaliada será decorrente da versão mutante da enzima.

Utilizando a mutagênese sítio-dirigida, os enzimologistas podem associar praticamente qualquer aspecto da atividade enzimática, como a catálise, resistência ou suscetibilidade a agentes químicos ou físicos, ou interações com outras proteínas, a aminoácidos específicos na proteína. Particularmente, a mutagênese sítio-dirigida permitiu que cientistas alterassem a afinidade de ligação ao receptor do hormônio de crescimento bovino, somatotrofina, em seres humanos, para que ele apenas estimulasse o crescimento e não a produção de leite (Seção 11.12).

### Mutagênese por inserção de cassete e ruptura gênica

Para mudar mais que alguns pares de base ou substituir seções de um gene de interesse, fragmentos sintéticos chamados de **cassetes de DNA** (ou cartuchos) podem ser utilizados para mutar o DNA em um processo conhecido como **mutagênese de cassete**. Esses cassetes podem ser sintetizados usando PCR ou por síntese direta de DNA. Também podem substituir seções do DNA de interesse usando sítios de restrição. Contudo, quando sítios de restrição apropriados não estão presentes no local desejado, eles podem ser inseridos por mutagênese sítio-dirigida (Figura 11.9). Os cassetes usados para substituição de genes são normalmente do mesmo tamanho que o DNA selvagem que são substituídos.

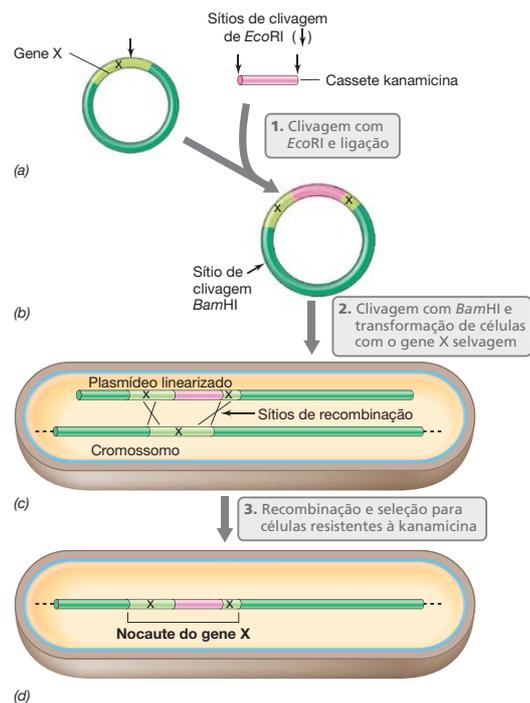
Outro tipo de mutagênese por cassete é denominado **ruptura gênica**. Nessa técnica, os cassetes são inseridos no meio de um gene, interrompendo, assim, a sequência codificadora. Os cassetes utilizados na introdução de mutações por inserção podem ser de qualquer tamanho, muitas vezes correspondendo a um gene inteiro. Para facilitar o processo de seleção, cassetes que codificam resistência a antibióticos são frequentemente utilizados. O processo de ruptura gênica está ilustrado na **Figura 11.10**. Nesse caso, um cassete de DNA contendo o gene de resistência à kanamicina, o cassete Kan, é inserido em um sítio de restrição de um gene clonado. O vetor contendo esse gene mutante é então linearizado pela clivagem com uma enzima de restrição diferente. Finalmente, o DNA linear é transformado na célula hospedeira, e a resistência à kanamicina é selecionada. O plasmídeo linearizado é incapaz de replicar-se, assim, as células resistentes surgem principal-

mente em decorrência de eventos de recombinação homóloga (⇨ Seção 10.5) entre o gene mutado presente no plasmídeo e o gene selvagem presente no cromossomo.

Observe que, quando um cassete é inserido, além de adquirirem resistência à kanamicina, as células *perdem a função do gene* no qual o cassete é inserido. Essas mutações são chamadas de **mutações de nocaute**. Essas mutações assemelham-se às mutações de inserção realizadas por transposons (⇨ Seção 10.11), embora, nesse caso, seja possível escolher exatamente qual gene será mutado. Mutações de nocaute em organismos haploides (como os procaríotos) originam células viáveis somente se o gene interrompido não for essencial. De fato, os nocautes gênicos são uma forma conveniente de determinar se um dado gene é essencial.

### MINIQUESTIONÁRIO

- Por que um suporte de fase sólida é utilizado durante a síntese química de DNA?
- Como uma mutagênese sítio-dirigida pode ser útil para os enzimologistas?
- O que são mutações de nocaute?



**Figura 11.10** Ruptura gênica utilizando mutagênese por inserção de cassete. (a) Uma cópia do gene X selvagem clonado, carreada por um plasmídeo, é clivada com a enzima *EcoRI* e misturada ao cassete de kanamicina. (b) O plasmídeo clivado e o cassete são ligados, gerando um plasmídeo com o cassete de kanamicina como uma mutação de inserção no interior do gene X. Esse novo plasmídeo é clivado com *BamHI* e transformado em uma célula. (c) A célula transformada possui o plasmídeo linearizado contendo o gene X interrompido, além do gene X selvagem em seu próprio cromossomo. (d) Em algumas células, ocorrerá a recombinação homóloga entre as formas selvagem e mutante do gene X. Células capazes de crescer na presença de kanamicina possuem uma única cópia do gene X interrompido.

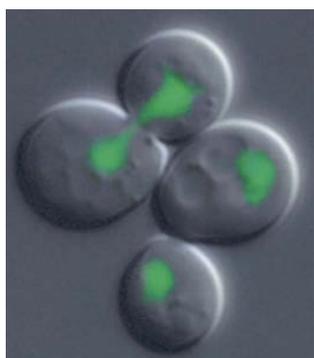
## 11.6 Fusões gênicas e genes repórteres

A manipulação de DNA *in vitro*, além de permitir novas abordagens para a geração de mutações, revolucionou o estudo da regulação gênica. As construções podem ser manipuladas de modo a permitir que uma sequência codificadora de uma fonte (o *repórter*) seja posicionada adjacente a uma região reguladora de outra origem para formar um gene híbrido. Isso pode ser utilizado para estudar a regulação gênica (↔ Seção 7.1), aumentar a expressão de um produto gênico de interesse, ou para analisar a resposta de uma região reguladora a diversas condições.

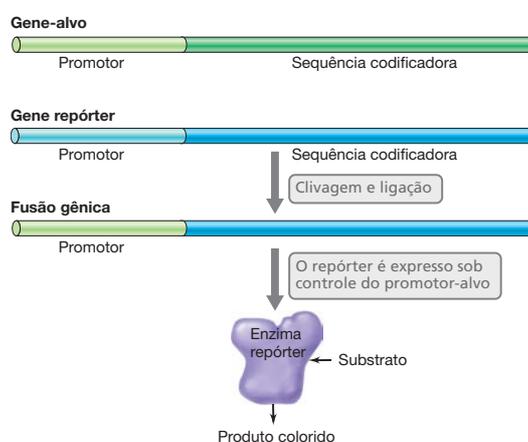
### Genes repórteres

A propriedade essencial de um **gene repórter** consiste na codificação de uma proteína de fácil análise e detecção. Genes repórteres são utilizados com uma grande variedade de propósitos. Eles podem ser usados para sinalizar a presença ou a ausência de um elemento genético em particular (como um plasmídeo) ou a inserção de DNA em um vetor. Eles podem também ser fusionados a outros genes ou ao promotor de outros genes, permitindo o estudo da expressão.

O primeiro gene a ser amplamente utilizado como repórter foi o gene *lacZ* de *Escherichia coli*, o qual codifica a enzima  $\beta$ -galactosidase, necessária para o catabolismo de lactose (↔ Seção 7.3). Células expressando  $\beta$ -galactosidase podem ser detectadas facilmente pela cor de suas colônias em placas indicadoras que contêm o substrato artificial Xgal (5-bromo-4-cloro-3-indolil- $\beta$ -D-galactopiranosídeo); Xgal é clivado pela  $\beta$ -galactosidase, produzindo uma coloração azul (ver Figura 11.14). A **proteína fluorescente verde (GFP, green fluorescent protein)** é amplamente utilizada como um repórter (Figura 11.11). Embora o gene de GFP tenha sido originalmente clonado da água-viva *Aequorea victoria*, a proteína GFP pode ser expressa na maioria das células. Ela é estável e provoca pouca ou nenhuma alteração no metabolismo da célula hospedeira. Quando a expressão de um gene clonado é associada àquela da GFP, a expressão de GFP sinaliza que o gene clonado foi também expresso (Figura 11.11).



**Figura 11.11** Proteína fluorescente verde (GFP). A GFP pode ser utilizada como uma marca para a localização de proteínas *in vivo*. Neste exemplo, o gene que codifica Pho2, uma proteína de ligação ao DNA da levedura *Saccharomyces cerevisiae*, foi fusionado ao gene que codifica GFP, e fotografado por microscopia de fluorescência. O gene recombinante foi transformado em células de leveduras em brotamento, que expressaram a proteína de fusão fluorescente localizada no núcleo.



**Figura 11.12** Construção e uso de fusões gênicas. O promotor do gene-alvo é fusionado à sequência codificadora repórter e o gene repórter é expresso sob condições nas quais o gene-alvo seria normalmente expresso. O repórter ilustrado aqui é uma enzima (como a  $\beta$ -galactosidase) que converte um substrato a um produto colorido, de fácil detecção. Essa estratégia facilita sobremaneira a investigação de mecanismos reguladores.

### Fusões gênicas

É possível manipular construções de modo a consistirem em segmentos de dois genes diferentes. Essas construções são conhecidas como **fusões gênicas**. Se o promotor que controla uma sequência codificadora for removido, a sequência codificadora pode ser fusionada a uma região reguladora diferente, a fim de submeter o gene ao controle de um promotor diferente. Alternativamente, o promotor pode ser fusionado a um gene cujo produto é de fácil detecção. Existem dois tipos de fusões gênicas. Nas **fusões de óperons**, uma sequência codificadora que retém seu próprio sítio e sinais de início da tradução é fusionada aos sinais de transcrição de outro gene. Na **fusão proteica**, genes que codificam duas diferentes proteínas são fusionados de forma a compartilhar os mesmos sinais de início e término da transcrição e da tradução. Após a tradução, a proteína fusionada produz um único polipeptídeo híbrido.

Essa estratégia é muitas vezes utilizada em estudos de regulação gênica, especialmente quando a quantificação do produto gênico natural é difícil, ou demanda muito tempo. A região reguladora do gene de interesse é fusionada à sequência codificadora de um gene repórter, como a  $\beta$ -galactosidase ou a GFP. O repórter é então submetido a condições que possam desencadear a expressão do gene-alvo (Figura 11.12). A expressão do repórter pode, então, ser analisada sob diferentes condições, a fim de determinar-se como o gene de interesse é regulado (↔ Seção 7.1). Ensaios de *controle da transcrição* são feitos fundindo os sinais de início da transcrição em um gene repórter, enquanto o *controle da tradução* é ensaiado pela fusão dos sinais de início da tradução de um gene de interesse a um gene repórter sob o controle de um promotor conhecido.

Fusões gênicas também podem ser utilizadas para avaliar os efeitos de genes reguladores. Mutações que afetam os genes reguladores são introduzidas em células contendo fusões gênicas, e a expressão é quantificada e comparada às células sem as mutações reguladoras. Isso permite a pesquisa rápida de múltiplos genes reguladores que supostamente controlam o gene-

-alvo. Além de utilizar as fusões para monitorar a presença ou a expressão de um determinado gene, proteínas que são facilmente purificadas também podem ser fundidas a proteínas de interesse para auxiliar na purificação (Seção 11.11).

#### MINIQUESTIONÁRIO

- O que é um gene repórter?
- Por que fusões gênicas são úteis em estudos de regulação gênica?

## II • Clonagem gênica

A principal etapa na engenharia genética é a manipulação do DNA com o propósito de clonagem. A clonagem permite aos cientistas isolarem genes de interesse de seus genomas e inserirem esses genes em moléculas carreadoras facilmente manipuladas, ou, de outra forma, estudadas.

### 11.7 Plasmídeos como vetores de clonagem

Os plasmídeos replicam-se independentemente do cromossomo hospedeiro. Além de possuir os genes necessários à sua própria replicação, a maioria dos plasmídeos são vetores naturais, uma vez que frequentemente carregam outros genes que conferem propriedades importantes a seus hospedeiros (↔ Seção 4.3). Como discutido a seguir, os plasmídeos possuem muitas propriedades úteis como *vetores de clonagem*.

Embora, na natureza, os plasmídeos conjugativos sejam transferidos por meio do contato entre duas células (↔ Seção 10.8), a maioria dos vetores de clonagem plasmidiais foi modificada geneticamente a fim de abolir sua transferência conjugativa. Isso impede o deslocamento indesejável do vetor para outros organismos. Entretanto, a transferência do vetor em laboratório pode ser facilitada por transformação quimicamente mediada, ou por eletroporação (↔ Seção 10.6). Dependendo do sistema plasmídeo-hospedeiro, a replicação plasmidial pode estar sujeita a um rígido controle celular, e, nesse caso, somente algumas poucas cópias do plasmídeo são produzidas, ou ela pode estar sujeita a um controle celular relaxado, e nesse caso, um grande número de cópias é produzida. A obtenção de um grande número de cópias é muitas vezes importante na clonagem molecular, e, por meio de uma seleção apropriada do sistema plasmídeo-hospedeiro e da manipulação da síntese de macromoléculas celulares, pode-se obter até centenas de milhares de cópias do plasmídeo em cada célula.

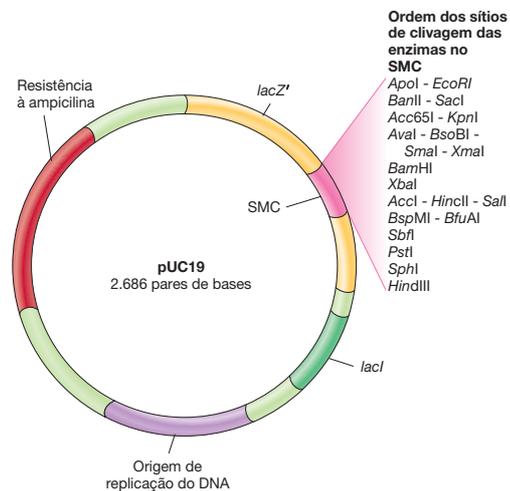
#### Um exemplo de um vetor de clonagem: o plasmídeo pUC19

Os primeiros vetores de clonagem plasmidiais utilizados eram isolados naturais. Em particular, os plasmídeos ColE de *Escherichia coli*, que codificam a colicina E, foram utilizados por serem relativamente pequenos e estarem naturalmente presentes em múltiplas cópias, facilitando o isolamento do DNA. Entretanto, eles foram rapidamente substituídos por plasmídeos resultantes de manipulações *in vitro*. Um vetor de clonagem plasmidial largamente utilizado é o plasmídeo pUC19 (Figura 11.13). Ele foi derivado, após várias etapas de modificação, do plasmídeo ColE1 (↔ Seção 4.3), as quais consistiram na remoção dos genes de colicina e inserção de genes de resistência à ampicilina, bem como em um sistema

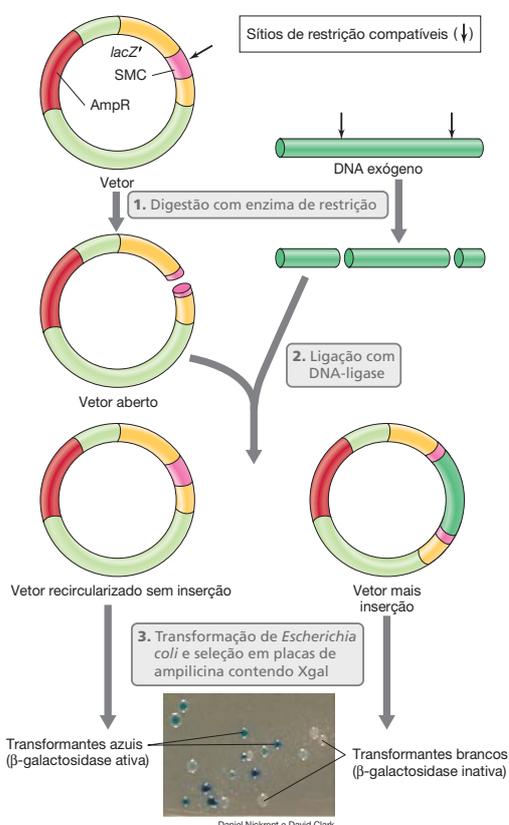
de seleção de coloração azul-branco (ver a seguir). Um segmento de DNA artificial contendo sítios de clivagem para muitas enzimas de restrição, denominado *SMC* ou *sítio múltiplo de clonagem*, encontra-se inserido no interior do gene *lacZ*, um gene que codifica a enzima que degrada lactose, β-galactosidase (↔ Seção 7.3). A presença desse SMC curto não inativa o gene *lacZ*. Sítios de clivagem de enzimas de restrição presentes no SMC não são encontrados no restante do vetor. Consequentemente, o tratamento com cada uma dessas enzimas abre o vetor em um único local, sem clivá-lo em vários fragmentos.

O plasmídeo pUC19 possui uma série de características que o tornam adequado como um veículo de clonagem:

1. Ele é relativamente pequeno, contendo somente 2.686 pb.
2. Ele é estavelmente mantido no hospedeiro (*E. coli*), com um número relativamente elevado de cópias, cerca de 50 cópias por célula.
3. Ele pode ser amplificado a um número muito elevado (1.000 a 3.000 cópias por célula, cerca de 40% do DNA celular) pela inibição da síntese de proteínas com o antibiótico cloranfenicol.

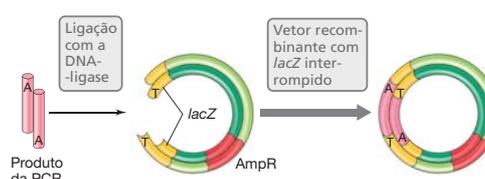


**Figura 11.13** Plasmídeo pUC19, um vetor de clonagem. As características essenciais incluem um marcador de resistência à ampicilina e um SMC, um segmento de DNA contendo vários sítios de clivagem para enzimas de restrição. A inserção do DNA clonado no interior do SMC inativa o gene *lacZ*, que codifica a β-galactosidase e permite a identificação fácil de transformantes pelo sistema de seleção azul-branco.



**Figura 11.14** Clonagem no vetor plasmidial pUC19. Uma enzima de restrição adequada cliva o vetor de clonagem no seu SMC, linearizando-o. A inserção do DNA exógeno inativa a β-galactosidase, permitindo a seleção azul-branco quanto à presença do inserto. A foto na parte inferior apresenta colônias de *Escherichia coli* em uma placa de Xgal. A enzima β-galactosidase pode clivar o Xgal, que normalmente é incolor, originando um produto azul.

4. Ele é facilmente isolado na forma superenovelada, usando técnicas de rotina.
5. Quantidades moderadas de DNA exógeno podem ser inseridas, embora insertos superiores a 10 kb promovam uma instabilidade plasmidial.
6. A sequência completa de bases do plasmídeo é conhecida, permitindo a identificação de todos os sítios de clivagem de enzimas de restrição.
7. O SMC contém sítios únicos de clivagem para uma dúzia de enzimas de restrição, as quais aumentam a versatilidade do vetor.
8. Ele possui um gene que confere ao seu hospedeiro resistência à ampicilina. Isso permite a pronta seleção de hospedeiros contendo o plasmídeo, uma vez que tais hospedeiros ganham resistência ao antibiótico.
9. Ele pode ser facilmente introduzido em células por transformação.
10. Inserção de DNA exógeno no SMC pode ser detectada pela seleção azul-branco devido à presença do gene *lacZ*.



**Figura 11.15** PCR para vetores. O vetor de clonagem linearizado contém resíduos de timinas não pareadas em suas terminações (em inglês, *overhanging*) que pareiam com os resíduos de adenina presentes na extremidade 3' do produto de PCR gerado pela *Taq*-polimerase. A ligação dos dois pedaços do DNA gera um plasmídeo circular contendo o gene *lacZ* interrompido. *AmpR*, gene que codifica para resistência à ampicilina.

### Clonagem de genes em vetores plasmidiais

A utilização de vetores plasmidiais na clonagem de genes é ilustrada na Figura 11.14. Uma enzima de restrição adequada, com um sítio de clivagem no interior do SMC, é escolhida. Tanto o vetor quanto o DNA exógeno a ser clonado são clivados com essa enzima. O vetor é linearizado. Segmentos do DNA exógeno são inseridos no sítio de clivagem aberto, sendo ligados nesta posição pela DNA-ligase. Isso interrompe o gene *lacZ*, um fenômeno denominado *inativação insercional*. Esse processo pode ser utilizado para detectar a presença do DNA exógeno no vetor. Quando o reagente incolor Xgal é adicionado ao meio de cultura, é clivado pela β-galactosidase, gerando um produto azul. Assim, células contendo o vetor *sem* o DNA clonado formam colônias azuis, enquanto as células contendo o vetor *com* um inserto de DNA clonado não produzem a β-galactosidase, sendo, portanto, brancas.

Desse modo, após a ligação do DNA, células de *E. coli* são transformadas com os plasmídeos resultantes. As colônias são selecionadas em meios contendo ampicilina, para a seleção da presença do plasmídeo, e Xgal, para avaliar a atividade da β-galactosidase. As colônias que são *azuis* contêm o plasmídeo sem qualquer DNA exógeno inserido (i.e., o plasmídeo simplesmente sofreu circularização, sem captar o DNA exógeno), enquanto aquelas que são *brancas* contêm o plasmídeo com DNA exógeno inserido, sendo, portanto, selecionadas para análises adicionais (ver Figura 11.20b para um exemplo relacionado com o sistema de seleção azul-branco).

### Outros vetores plasmidiais

Muitos vetores subsequentes têm características similares às daquelas do pUC19, listadas anteriormente, mas também apresentam outras características desejáveis. Por exemplo, alguns vetores foram desenvolvidos especificamente para clonagem de produtos de DNA sintetizados pela enzima *Taq*-polimerase na reação em cadeia de polimerase (PCR) (Seção 11.3). A atividade enzimática da *Taq*-polimerase adiciona um resíduo de adenina na extremidade 3' do seu produto independente de um molde. Vetores linearizados estão disponíveis comercialmente e contêm um resíduo de timina não pareado em suas extremidades (em inglês, *overhanging*) que permite o pareamento de bases com o produto da PCR e sua subsequente ligação usando uma DNA-ligase (Figura 11.15).

Outros vetores foram desenhados para selecionar diretamente os vetores recombinantes por meio da viabilidade

celular, e não pela varredura. Por exemplo, em alguns vetores, o gene carreando o SMC normalmente produz uma proteína que é letal à célula hospedeira. Portanto, apenas as células que possuem um plasmídeo no qual esse gene foi inativado são capazes de crescer.

A clonagem empregando vetores plasmidiais é versátil e amplamente utilizada em engenharia genética, particularmente quando o fragmento a ser clonado é relativamente pequeno. Plasmídeos também são frequentemente utilizados como vetores de clonagem quando se deseja expressar o gene clonado, uma vez que genes reguladores podem ser manipulados por engenharia genética no plasmídeo, a fim de obter-se a expressão dos genes clonados sob condições específicas (Seção 11.9).

**MINIQUESTIONÁRIO**

- Explique por que, em uma clonagem, é necessário utilizar uma enzima de restrição que clive o vetor em somente um local.
- O que é inativação insercional?
- O que é um SMC?

**11.8 Hospedeiros para vetores de clonagem**

Para a obtenção de grandes quantidades de DNA clonado, um hospedeiro ideal deve crescer rapidamente em um meio de cultura de baixo custo. Além disso, o hospedeiro não deve ser patogênico, deve ser capaz de incorporar o DNA, ser geneticamente estável em cultura e possuir as enzimas necessárias à replicação do vetor. Também é útil quando podemos dispor de informações adicionais sobre o hospedeiro, e de uma abundância de ferramentas para sua manipulação genética.

Os microrganismos estão entre os hospedeiros mais úteis à clonagem, pois crescem facilmente e dispomos de muitas informações sobre eles. Esses incluem as bactérias *Escherichia coli* e *Bacillus subtilis* e a levedura *Saccharomyces cerevisiae* (Figura 11.16). Sequências genômicas completas de todos esses organismos encontram-se disponíveis; tais organismos são amplamente utilizados como hospedeiros de clonagem. Entretanto, em alguns casos, outros hospedeiros e vetores especializados podem ser necessários a fim de obtermos o DNA clonado e expresso de forma adequada.

**Hospedeiros procarióticos**

Embora a maioria das clonagens moleculares tenha sido realizada em *E. coli* (Figura 11.16), este hospedeiro apresenta algumas desvantagens. *E. coli* é uma excelente escolha para o trabalho de clonagem inicial, porém é problemática como um vetor de expressão, uma vez que essa bactéria é encontrada no trato intestinal humano, e linhagens selvagens são potencialmente patogênicas (↔ Seção 31.12). Entretanto, várias linhagens modificadas de *E. coli* foram desenvolvidas com finalidades de clonagem e, dessa forma, *E. coli* continua sendo o organismo de escolha para a maioria das clonagens moleculares. Um grande problema com a utilização de qualquer hospedeiro bacteriano, incluindo *E. coli*, é a falta de um sistema para modificar corretamente as proteínas eucarióticas; esse problema pode ser resolvido usando células hospedeiras eucariotas, como discutido a seguir.

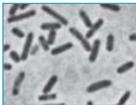
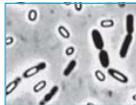
Outro problema com a utilização de *E. coli*, assim como com todas as bactérias gram-negativas, é que elas possuem uma membrana exterior que dificulta a secreção de proteínas. Problema que pode ser resolvido usando a bactéria gram-positiva *B. subtilis* como hospedeiro de clonagem (Figura 11.16). Embora a tecnologia para clonagem em *B. subtilis* seja menos avançada do que em *E. coli*, foram desenvolvidos inúmeros plasmídeos e fagos adequados à clonagem, sendo a transformação um procedimento bem estabelecido em *B. subtilis*. A principal desvantagem na utilização de *B. subtilis* como hospedeiro de clonagem é a instabilidade plasmidial. Frequentemente, a manutenção da replicação plasmidial é dificultada após vários subcultivos do organismo. Além disso, DNA exógenos não são mantidos de maneira eficiente em *B. subtilis*, quando comparados a *E. coli*; assim, o DNA clonado é frequentemente eliminado de maneira inesperada.

Em geral, os organismos utilizados para a clonagem devem apresentar genótipos específicos para serem eficientes. Por exemplo, se o vetor de clonagem utiliza o gene *lacZ* para a varredura dos clones, o hospedeiro deve apresentar uma mutação que inativa seu gene. Essas considerações, e outras, como a capacidade de selecionar os transformantes, devem ser levadas em consideração na escolha de um hospedeiro de clonagem.

**Hospedeiros eucarióticos**

A clonagem em microrganismos eucarióticos concentrou-se na levedura *S. cerevisiae* (Figura 11.16). Vetores plasmidiais, assim como cromossomos artificiais (Seção 11.10), foram desenvolvidos para leveduras. Uma importante vantagem das células eucarióticas como hospedeiras de vetores de clonagem refere-se ao fato de possuírem os sistemas complexos de processamento de RNA e a modificação pós-traducional, necessários à produção de proteínas eucarióticas. Portanto, a presença de tais sistemas torna dispensável a introdução de modificações no vetor ou nas células hospedeiras, as quais seriam necessárias caso o DNA eucariótico fosse clonado e expresso em um hospedeiro de clonagem procariótico.

Em muitas aplicações, a clonagem gênica em células de mamíferos foi realizada. Os sistemas de cultura de células de mamíferos podem ser manipulados de maneira semelhante às culturas microbianas, sendo amplamente utilizados em pesquisas envolvendo genética humana, câncer, doenças infecciosas e fisiologia. Uma desvantagem das células de mamíferos como hospedeiros de clonagem refere-se ao alto custo e a

Bactéria		Eucarioto
<i>Escherichia coli</i>	<i>Bacillus subtilis</i>	<i>Saccharomyces cerevisiae</i>
		
<ul style="list-style-type: none"> <li>Genética bem desenvolvida</li> <li>Muitas linhagens disponíveis</li> <li>Procarioto mais bem conhecido</li> </ul>	<ul style="list-style-type: none"> <li>Facilmente transformável</li> <li>Não patogênico</li> <li>Proteínas secretadas naturalmente</li> <li>A formação de endósporo simplifica a cultura</li> </ul>	<ul style="list-style-type: none"> <li>Genética bem desenvolvida</li> <li>Não patogênico</li> <li>Pode processar RNAm e proteínas de fácil crescimento</li> </ul>
<ul style="list-style-type: none"> <li>Potencialmente patogênica</li> <li>Proteínas retidas no periplasma</li> </ul>	<ul style="list-style-type: none"> <li>Geneticamente instável</li> <li>Genética menos desenvolvida que em <i>E. coli</i></li> </ul>	<ul style="list-style-type: none"> <li>Plasmídeos instáveis</li> <li>Não replica a maioria dos plasmídeos procarióticos</li> </ul>
<span style="background-color: #e0f2f1; padding: 2px;">Vantagens</span>		<span style="background-color: #ffe0b2; padding: 2px;">Desvantagens</span>

**Figura 11.16 Hospedeiros para clonagem molecular.** Um resumo das vantagens e desvantagens de alguns hospedeiros comuns para clonagem.

difícil produção em larga escala. Linhagens celulares de insetos são de cultivo mais simples, havendo o desenvolvimento de vetores a partir de vírus de DNA de insetos, o baculovírus. Em algumas situações, especialmente na agricultura, o hospedeiro de clonagem pode ser uma linhagem de cultura de tecido vegetal, ou até mesmo uma planta inteira. De fato, a engenharia genética apresenta inúmeras aplicações na agricultura (Seção 11.13). Entretanto, independentemente do tipo de hospedeiro eucariótico, é necessário inserir o DNA do vetor em células do hospedeiro. As técnicas para transferir DNA para células eucarióticas não serão discutidas agora, mas incluem transfecção (ver Figura 11.28), microinjeção e eletroporação.

#### MINIQUESTIONÁRIO

- Por que a clonagem molecular requer um hospedeiro?
- Em que situações a utilização de hospedeiros eucariotos para a clonagem molecular é benéfica?

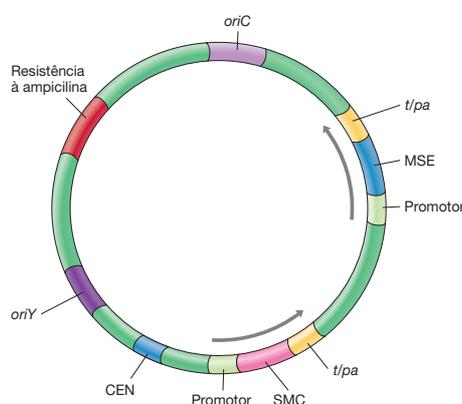
### 11.9 Vetores bifuncionais e vetores de expressão

Uma vez que um gene foi clonado, ele pode ser utilizado para uma variedade de objetivos. Vetores especializados foram desenvolvidos para o uso em diferentes situações. Aqui vamos discutir como transferir um gene clonado entre organismos de diferentes espécies e como otimizar a expressão do gene clonado. Duas classes de vetores são adequadas, vetores *bifuncionais* e vetores de *expressão*.

#### Vetores bifuncionais

Vetores capazes de replicarem-se e serem mantidos de modo estável em dois (ou mais) organismos hospedeiros não relacionados são denominados **vetores bifuncionais**. Assim, genes carregados por um vetor bifuncional podem ser transferidos entre organismos não relacionados. Foram desenvolvidos vetores bifuncionais capazes de replicarem-se tanto em *Escherichia coli* quanto em *Bacillus subtilis*, em *E. coli* e leveduras, em *E. coli* e células de mamíferos, assim como em muitos outros pares de organismos. A importância de um vetor bifuncional está no fato de o DNA clonado em um organismo poder ser replicado em um segundo hospedeiro, sem a necessidade de modificação do vetor para isso.

Muitos vetores bifuncionais foram projetados para permitir a transferência gênica entre *E. coli* e leveduras. Vetores plasmidiais bacterianos foram os primeiros a serem utilizados, sendo modificados para atuarem também em leveduras. Uma vez que as origens de replicação bacteriana não funcionam em eucariotos, é necessário fornecer uma origem de replicação de levedura. Uma vantagem é o fato de as sequências de DNA de origens de replicação serem similares em diferentes eucariotos, de modo que a origem de leveduras é funcional em outros organismos superiores. Quando células eucarióticas dividem-se, os cromossomos duplicados são separados pelos microtúbulos (“fibras do fuso”) ligados aos centrômeros (↔ Seção 2.20). Consequentemente, vetores bifuncionais para eucariotos devem conter um segmento de DNA do centrômero, a fim de serem distribuídos adequadamente na divisão celular (Figura 11.17). Felizmente, a sequência de reconhecimento centromérica de levedura, a sequência CEN, é relativamente pequena e de fácil inserção nos vetores bifuncionais.



**Figura 11.17** Mapa genético de um vetor bifuncional utilizado em levedura. Este vetor contém componentes que permitem sua utilização em *Escherichia coli* e leveduras, e pode ser selecionado em cada organismo: *oriC*, origem de replicação em *E. coli*; *oriY*, origem de replicação de levedura; SMC, sítio múltiplo de clonagem; MSE, marcador de seleção eucariótico; CEN, sequência centromérica de levedura; *t/pa*, sinais de término de transcrição/poliadenilação. As setas indicam a direção de transcrição.

Há também a necessidade de um marcador conveniente para selecionar o plasmídeo na levedura. Infelizmente, as leveduras não são suscetíveis à maioria dos antibióticos eficazes contra bactérias. Na prática, são utilizadas linhagens hospedeiras de leveduras deficientes na produção de algum aminoácido específico, ou base púrica ou pirimídica.

Uma cópia funcional do gene biossintético que é deficiente no hospedeiro é inserida no vetor bifuncional. Por exemplo, se o gene *URA3*, necessário à síntese de uracila, for utilizado, a levedura não crescerá na ausência de uracila, mas o fará apenas se receber a cópia do vetor bifuncional.

#### Vetores de expressão

Organismos apresentam sistemas reguladores complexos e genes clonados são muitas vezes pouco expressos ou não são expressos em uma célula hospedeira exógena. Esse obstáculo pode ser superado pelo uso de **vetores de expressão**, vetores desenvolvidos para permitir ao pesquisador controlar a expressão dos genes clonados. Em geral, o objetivo consiste na obtenção de altos níveis de expressão, especialmente em aplicações biotecnológicas. Entretanto, quando se trata de produtos gênicos potencialmente tóxicos, um nível baixo, porém estritamente controlado, pode ser apropriado.

O controle da expressão é realizado porque vetores de expressão contêm sequências reguladoras que permitem a manipulação da expressão gênica. Geralmente, o controle é transcripcional, uma vez que, para obter-se altos níveis de expressão, é essencial a produção de altos níveis de RNAm. Na prática, altos níveis de transcrição requerem promotores fortes, que permitam a ligação eficiente da RNA-polimerase (↔ Seção 4.7). Entretanto, o promotor nativo de um gene clonado pode não atuar de maneira eficiente no hospedeiro novo. Por exemplo, promotores de eucariotos, ou até mesmo de outras bactérias, não funcionam ou funcionam pouco em *E. coli*. De fato, até mesmo alguns promotores de *E. coli* atuam em baixos níveis em *E. coli*, por apresentarem sequências diferentes da sequência promotora consenso, promovendo uma ligação fraca da RNA-polimerase (↔ Seção 4.7).

Por essa razão, vetores de expressão devem conter um promotor que funcione de maneira eficiente no hospedeiro e que esteja posicionado corretamente, permitindo a transcrição do gene clonado. Entre os promotores de *E. coli* utilizados na construção de vetores de expressão podemos mencionar *lac* (o promotor do óperon *lac*), *trp* (o promotor do óperon *trp*), *tac* e *trc* (híbridos sintéticos dos promotores *trp* e *lac*). Esses são promotores “fortes” em *E. coli* e podem ser regulados especificamente.

### Regulação da transcrição em vetores de expressão

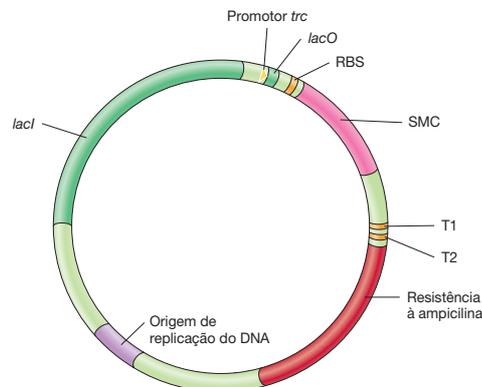
A regulação da expressão de genes clonados é importante. Embora em geral seja desejável a produção de altas concentrações de RNAm (sendo ele traduzido), geralmente não é conveniente que os genes clonados sejam transcritos em elevados níveis em todos os momentos. Em condições ideais, a cultura contendo o vetor de expressão deve crescer até atingir um grande número de células, cada uma contendo inúmeras cópias do vetor. A expressão do(s) gene(s) desejado(s) pode ser ativada por um sinal genético.

A regulação da transcrição por uma proteína repressora (↔ Seção 7.3) é uma maneira útil de controlar um gene clonado. Um repressor forte pode bloquear completamente a síntese de proteínas sob seu controle ligando-se à região do operador. Além disso, quando a expressão do gene é requerida, o repressor pode ser removido pela adição do indutor, permitindo a transcrição dos genes regulados. Para que um sistema repressor-operador possa controlar a produção de proteínas exógenas, o vetor de expressão é desenvolvido de maneira que o gene clonado seja inserido imediatamente a jusante ao promotor e da região operadora escolhidos. Um sítio forte de ligação ao ribossomo é frequentemente incluído entre o promotor e o gene clonado. Isso resulta em um controle do gene clonado pelo promotor e em transcrição e tradução eficientes. O operador e o promotor geralmente são correspondentes (p. ex., o operador *lac* é empregado com o promotor *lac*), embora esse nem sempre seja o caso. Por exemplo, regiões reguladoras híbridas são ocasionalmente utilizadas (p. ex., fusionando o promotor *trp* com o operador *lac* para formar o elemento regulador *trc*).

A Figura 11.18 mostra um vetor de expressão controlado por *trc*. Esse plasmídeo também contém uma cópia do gene *lacI*, que codifica o repressor *lac*. O nível do repressor em uma célula contendo esse plasmídeo é suficiente para impedir a transcrição a partir do promotor *trc*, até que o indutor seja adicionado. A adição de lactose ou indutores *lac* relacionados desencadeia a transcrição do DNA clonado (Figura 12.20). Além de um promotor forte e facilmente regulado, a maioria dos vetores de expressão contém um eficiente terminador de transcrição (↔ Seção 4.7). Isso impede que a transcrição a partir do promotor forte prossiga em outros genes do vetor, o que poderia interferir com a estabilidade do vetor. O vetor de expressão apresentado na Figura 11.18 possui terminadores de transcrição eficientes para interromper a transcrição imediatamente a jusante ao gene clonado.

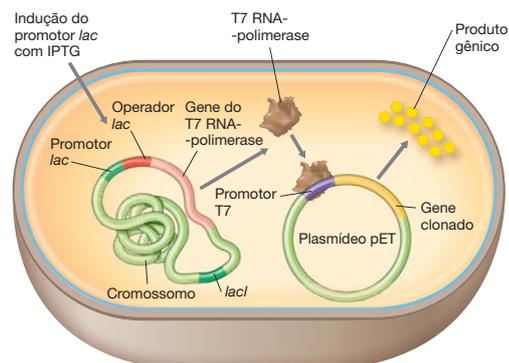
### Regulação da expressão por elementos de controle do bacteriófago T7

Em alguns casos, o sistema de controle transcricional pode não ser um componente normal do hospedeiro. Um exemplo de tal situação é a utilização do promotor e RNA-polimerase do



**Figura 11.18** Mapa genético do vetor de expressão pSE420. Este vetor foi desenvolvido pela Invitrogen Corp., uma empresa de biotecnologia. O SMC contém muitas sequências reconhecidas por enzimas de restrição diferentes, facilitando a clonagem. Esta região e o gene clonado são transcritos pelo promotor *trc*, localizado imediatamente a montante ao operador *lac* (*lacO*). Imediatamente a montante ao SMC há uma sequência que codifica o sítio Shine-Dalgarno (S/D), de ligação ao ribossomo, no RNAm resultante. A jusante ao SMC são encontrados dois terminadores transcricionais (T1 e T2). O plasmídeo também contém o gene *lacI*, o qual codifica o repressor *lac*, e um gene conferindo resistência ao antibiótico ampicilina. Estes dois genes são controlados por seus próprios promotores, os quais não estão apresentados.

bacteriófago T7 para a regulação da expressão. Quando o fago T7 infecta *E. coli*, este codifica sua própria RNA-polimerase, a qual reconhece somente os promotores T7, bloqueando efetivamente a transcrição da célula hospedeira (↔ Seção 9.4). Em vetores de expressão T7, genes clonados são submetidos ao controle do promotor T7, limitando, assim, sua transcrição. Para que a transcrição ocorra, o gene do T7 RNA-polimerase deve também estar presente na célula, sob o controle de um sistema facilmente regulado, como o *lac* (Figura 11.19). Isso é geralmente realizado pela integração do gene do T7 RNA-polimerase com o promotor *lac* no cromossomo de uma linhagem hospedeira especializada.



**Figura 11.19** O sistema de expressão T7. O gene do T7 RNA-polimerase encontra-se em uma fusão gênica controlada pelo promotor *lac* e está inserido no cromossomo de uma linhagem hospedeira especial de *Escherichia coli*. A adição de IPTG induz o promotor *lac*, causando a expressão do T7 RNA-polimerase. Isso promove a transcrição do gene clonado sob o controle do promotor T7, que é carregado pelo plasmídeo pET.

A série BL21 das linhagens hospedeiras é especialmente desenvolvida para ser empregada com a série pET de vetores de expressão T7. Os genes clonados são expressos logo após a ativação da transcrição do gene T7 RNA-polimerase por um indutor *lac*, como o IPTG. Visto que essa enzima somente reconhece promotores T7, apenas os genes clonados serão transcritos. O T7 RNA-polimerase é tão ativo que utiliza a maioria dos precursores de RNA; consequentemente, os genes do hospedeiro, que requerem a RNA-polimerase hospedeira, não são na sua maioria transcritos e, dessa forma, haverá a interrupção do crescimento celular. A síntese proteica nessas células é então dominada pela proteína de interesse. Desse modo, o sistema de controle T7 é extremamente eficiente na produção de grandes quantidades de uma proteína específica de interesse.

### Tradução do gene clonado

Os vetores de expressão devem ser construídos de tal modo que garantam que o RNAm produzido seja de maneira eficiente traduzido. Para que uma proteína seja sintetizada a partir de um RNAm, é essencial que os ribossomos liguem-se ao sítio correto e iniciem a leitura na fase correta. Em bactérias, esse processo é mediado pelo sítio de ligação do ribossomo (sequência Shine-Dalgarno, ↪ Seção 4.11) e por um códon de início próximo, no RNAm. Os sítios bacterianos de ligação do ribossomo não são encontrados em genes eucarióticos, devendo ser introduzidos no vetor, quando altos níveis de expressão de um gene eucariótico são desejados. O vetor apresentado na Figura 11.18 também apresenta tal sítio.

Com frequência, são necessários ajustes para garantir a alta eficiência da tradução após o gene ter sido clonado. Por exemplo, o uso preferencial de códons pode ser um obstáculo (↪ Seção 6.3 e Tabela 6.3). O uso preferencial de códons está relacionado com a concentração celular do RNAt apropriado. Por causa da redundância do código genético, existe mais de um RNAt para a maioria dos aminoácidos (↪ Seção 4.11). Assim, se um gene clonado apresentar um uso preferencial de códons consideravelmente diferente daquele de seu hospedeiro de expressão, ele poderá ser traduzido de modo ineficiente. Mutagêneses sítio-dirigidas (Seções 11.5) podem ser usadas para modificar códons selecionados do gene, tornando-o mais adequado ao padrão de códons preferenciais do hospedeiro.

Finalmente, se o gene clonado possuir íntrons, como genes eucarióticos geralmente apresentam (↪ Seção 4.9), o produto proteico correto não será sintetizado se o hospedeiro for um procarionte. Esse problema também pode ser contornado pelo uso de DNA sintético. Entretanto, o método habitual de gerar um gene sem íntrons consiste em sua obtenção a partir do RNAm (do qual os íntrons foram removidos), utilizando-se a transcriptase reversa para gerar uma cópia de DNAC (ver Figuras 11.6 e 11.23).

### MINIQUESTIONÁRIO

- Descreva alguns dos elementos de um vetor de expressão que otimizam a expressão do gene clonado.
- Descreva os elementos necessários para um vetor bifuncional eficiente.

## 11.10 Outros vetores de clonagem

Os típicos vetores plasmidiais geralmente usados para clonagem molecular são limitados em sua capacidade de DNA que pode ser inserido, com o máximo sendo 10 pares de quilobases (Kpb). Para regiões genômicas grandes, como os óperons e genes eucarióticos, vetores baseados em bacteriófagos, cosmídeos e cromossomos artificiais foram desenvolvidos. Apesar de não serem discutidos em detalhes aqui. Vetores virais são normalmente utilizados em eucariotos multicelulares. Em particular, os retrovírus podem ser utilizados para introduzir genes em células de mamíferos, pois esses vírus replicam-se por intermédio de uma forma de DNA que se integra ao cromossomo do hospedeiro (↪ Seção 9.11).

### Bacteriófago lambda como um vetor de clonagem

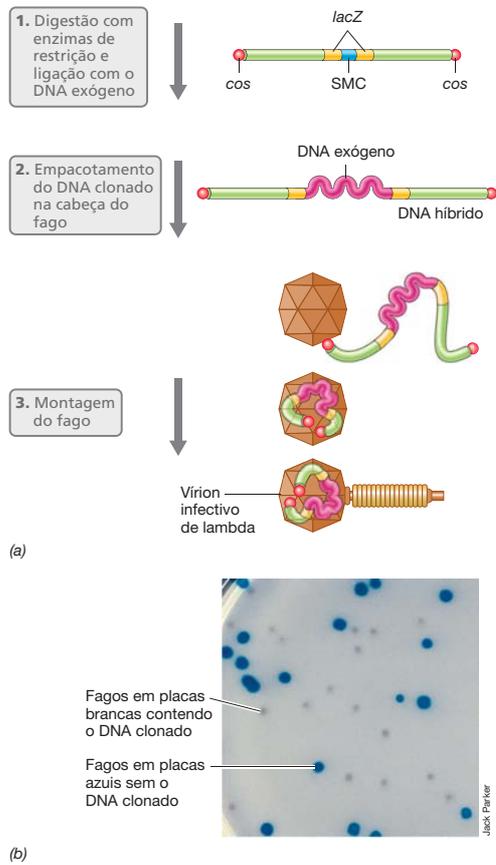
O fago lambda (↪ Seção 8.8) é um vetor de clonagem particularmente útil, pois sua genética molecular é bastante conhecida e é capaz de receber maiores quantidades de DNA do que a maioria dos plasmídeos, sendo esse DNA empacotado de maneira eficiente nas partículas fágicas, *in vitro*. Lembre-se que durante a fase lítica do ciclo do fago, a célula hospedeira *Escherichia coli* é reprogramada para replicar grandes quantidades do DNA lambda (↪ Seção 10.7)

O fago lambda apresenta um grande número de genes, porém um terço do genoma de lambda não é essencial à infectividade e pode ser substituído por um DNA exógeno. Isso permite que fragmentos relativamente grandes de DNA, de até 20 kb, sejam clonados em lambda. Isso é o dobro da capacidade de clonagem de pequenos vetores plasmidiais típicos. Para facilitar o uso do fago lambda como um vetor de clonagem molecular, muitos de seus sítios para enzimas de restrição foram alterados, e um SMC contendo o gene para a  $\beta$ -galactosidase foi adicionado para selecionar os vetores recombinantes.

A Figura 11.20a ilustra a clonagem em vetores lambda. O processo inicial é semelhante à clonagem de DNA em um vetor plasmidial no qual enzimas de restrição e DNA-ligasas são utilizadas. Uma vez que o DNA de interesse foi inserido no DNA do fago lambda, o vetor é empacotado pela adição de extratos celulares contendo proteínas da cabeça e da cauda, permitindo a formação espontânea de partículas fágicas viáveis. Essas partículas podem ser utilizadas na infecção de células hospedeiras adequadas, sendo a infecção muito mais eficiente do que a transformação. Clones individuais podem ser isolados por plaqueamento em uma linhagem de *E. coli* hospedeira e coletadas as placas de lise. Fagos recombinantes podem ser selecionados por varredura para interrupção do gene  $\beta$ -galactosidase usando ágar contendo indicador colorido (Seção 11.7). Placas de lise de fagos que não produzem  $\beta$ -galactosidase podem ser detectadas com facilidade como placas incolores, em um fundo de placas coloridas (ver Figura 11.20b). Procedimentos de hibridização de ácidos nucleicos (Seção 11.2) e sequenciamento do DNA podem ser usados para determinar se o DNA do fago lambda recombinante contém a sequência de DNA exógeno desejada.

### Vetores cosmídeos

Assim como os vetores de substituição, vetores do tipo cosmídeos utilizam genes específicos de lambda e são empacotados em partículas de lambda. *Cosmídeos* são vetores plasmidiais que contêm o sítio *cos* do genoma de lambda, o qual gera extremidades coesivas quando clivado (↪ Seção 8.8). O sítio *cos*



**Figura 11.20** Vetores de clonagem do fago lambda. (a) Inserção de DNA exógeno no DNA lambda modificado para conter um SMC dentro do gene *lacZ* e subsequente empacotamento de um vírion infectivo de lambda recombinante. O tamanho máximo do inserto é aproximadamente 20 kb. (b) Porção de uma placa de ágar contendo Xgal apresentando placas brancas formadas pelo fago lambda contendo o DNA clonado e placas azuis formadas pelo fago sem o DNA clonado.

é necessário ao empacotamento do DNA nos vírions lambda. Os cosmídeos são construídos a partir de plasmídeos, por meio da ligação da região *cos* de lambda ao DNA plasmidial. O DNA exógeno é, então, ligado ao vetor. O plasmídeo modificado, juntamente com o DNA exógeno, podem então ser empacotados em vírions lambda *in vitro*, conforme descrito anteriormente, sendo as partículas fágicas utilizadas na transdução de *E. coli*.

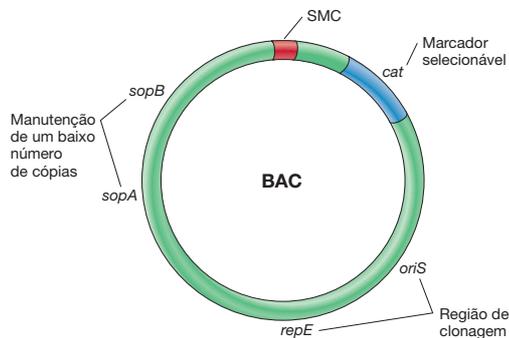
Um das principais vantagens dos cosmídeos está no fato de eles poderem ser utilizados na clonagem de grandes fragmentos de DNA, permitindo insertos de até 50 pares de quilobases. Assim, com insertos maiores, um menor número de clones é necessário à obtenção do genoma completo. A utilização de cosmídeos evita a necessidade de transformar-se *E. coli*, que é especialmente ineficiente com grandes plasmídeos. Os cosmídeos também permitem o armazenamento do DNA em partículas fágicas, em vez de na forma plasmidial. As partículas fágicas são mais estáveis que os plasmídeos, possibilitando, assim, a preservação do DNA recombinante por longos períodos de tempo.

### Cromossomos artificiais

Vetores que permitem a inserção de fragmentos de DNA clonado de aproximadamente 2 a 10 kb são adequados à construção de bibliotecas genômicas, no sequenciamento de genomas procarióticos. Vetores do bacteriófago lambda, capazes de receber insertos de 20 kb ou mais, são também largamente utilizados em projetos genômicos. Entretanto, à medida que o tamanho do genoma a ser sequenciado aumenta, maior é o número de clones necessários à obtenção da sequência completa. Portanto, para a construção de bibliotecas de DNA de microrganismos eucarióticos ou de eucariotos superiores, como o homem, é útil dispormos de vetores capazes de comportar segmentos muitos extensos de DNA. Esses vetores permitem que o tamanho inicial da biblioteca seja manejável. Tais vetores foram desenvolvidos, sendo denominados **cromossomos artificiais**.

O plasmídeo F de *E. coli*, de ocorrência natural, que contém 99,2 kb de DNA (↔ Seção 10.8) e plasmídeos derivados, denominados plasmídeos F', são capazes de carrear grandes quantidades de DNA cromossômico (↔ Seção 10.9). Devido a essas propriedades desejáveis, o plasmídeo F foi utilizado para a construção de vetores de clonagem, denominados **cromossomos artificiais bacterianos (BACs, bacterial artificial chromosomes)**. A Figura 11.21 revela a estrutura de um BAC baseado no plasmídeo F. O vetor apresenta somente 6,7 kb e contém apenas alguns genes do plasmídeo F necessários à replicação e à manutenção de um número baixo de cópias. O plasmídeo também contém o gene *cat*, que confere resistência ao cloranfenicol, e um SMC contendo vários sítios de restrição, para a clonagem de DNA. DNAs exógenos maiores que 300 kb podem ser inseridos e mantidos estavelmente em um vetor BAC desse tipo.

Historicamente, os primeiros cromossomos artificiais correspondiam a **cromossomos artificiais de leveduras (YACs, yeast artificial chromosomes)** (Figura 11.22). Esses vetores replicam-se em leveduras como cromossomos normais, porém possuem sítios em que fragmentos muito grandes de DNA podem ser inseridos. Para atuarem como cromossomos eucarióticos normais, os YACs devem conter (1) uma origem de replicação de DNA, (2) telômeros, para a replicação do DNA nas regiões terminais do cromossomo, e (3) um centrômero, para a segregação durante a mitose. Eles também devem



**Figura 11.21** Mapa genético de um cromossomo artificial bacteriano. O BAC esquematizado na figura contém 6,7 kb. A região de clonagem contém vários sítios únicos de enzimas de restrição. Esse BAC contém apenas uma pequena porção do plasmídeo F, de 99,2 kb.



**Figura 11.22** Um cromossomo artificial de levedura contendo DNA exógeno. O DNA exógeno foi clonado no vetor, em um sítio de restrição *NotI*. Os telômeros são indicados por TEL e o centrômero, por CEN. A origem de replicação é assinalada por ARS (sequência de replicação autônoma). O gene utilizado na seleção é denominado *URA3*. O hospedeiro, no qual o clone é trans-

formado, possui uma mutação em *URA3*, requerendo uracila para seu crescimento (Ura<sup>-</sup>). Células hospedeiras contendo esse CAL (cromossomo artificial de levedura) tornam-se Ura<sup>+</sup>. O diagrama não se encontra em escala; o DNA do vetor contém somente 10 kb, ao passo que o DNA clonado pode apresentar até 800 kb.

conter um sítio de clonagem e um gene que possa ser utilizado na seleção, após a transformação da célula hospedeira, normalmente a levedura *Saccharomyces cerevisiae*. A Figura 11.22 apresenta um diagrama de um vetor YAC, no qual um DNA exógeno foi clonado. Os vetores YACs possuem somente cerca de 10 kb e podem comportar insertos de DNA variando de 200 a 800 kb.

#### MINIQUESTIONÁRIO

- Por que a habilidade de empacotar DNA recombinante em partículas compostas por fagos, *in vitro*, é útil?
- O que os acrônimos BAC e YAC significam?
- O cromossomo artificial de levedura comporta-se como um cromossomo em uma célula de levedura. O que torna isso possível?

## III • Produtos da engenharia genética de microrganismos

A engenharia genética tem sido utilizada para transformar microrganismos em pequenas fábricas para a produção de produtos valiosos como combustíveis, substâncias químicas, fármacos e hormônios humanos como a insulina. Até este ponto, discutimos as técnicas utilizadas para manipular, clonar e expressar o DNA. Agora discutiremos como estas técnicas podem ser diretamente aplicadas para a **biotecnologia**, incluindo alguns dos principais desafios que existem com a expressão de genes eucariotos em bactérias e a subsequente purificação dos produtos proteicos. Também abordaremos a alteração genética de organismos superiores e suas aplicações na agricultura e medicina.

### 11.11 Expressão de genes de mamíferos em bactérias

Alguns problemas encontrados nos vetores de expressão foram mencionados na Seção 11.9, e existem vários outros obstáculos a serem vencidos para clonar um gene de mamífero em bactérias. Esses problemas incluem (1) os genes eucariotos precisam ser colocados sob o controle de um promotor bacteriano (Seção 11.9); (2) todos os íntrons (↔ Seção 4.9) devem ser removidos; (3) utilização preferencial de códons (↔ Seção 4.11) pode exigir edições para as sequências dos genes; (4) muitas proteínas de mamíferos exigem modificações após sua tradução para tornarem-se funcionais e a maioria dessas não pode ser realizada por bactérias. A seguir, abordaremos algumas soluções para esses desafios.

#### Clonagem do gene via RNAm

O procedimento-padrão para a obtenção de um gene eucariótico desprovido de íntrons consiste na sua clonagem via o seu RNAm. Uma vez que os íntrons são removidos durante o processamento do RNAm, o RNAm maduro corresponde a uma sequência codificadora ininterrupta (↔ Seção 4.9 e Figura 4.29). Os tecidos que expressam o gene de interesse frequentemente contêm grandes quantidades de RNAm de interesse, embora outros RNAm também estejam presentes. Contudo, em certas

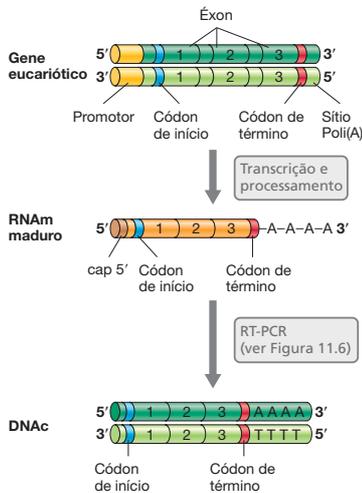
situações, um único RNA é predominante em um tipo de tecido, e a extração do RNAm a partir desse tecido consiste em um ponto de partida útil para a clonagem gênica.

Em uma célula de mamífero típica, cerca de 80 a 85% do RNA correspondem a RNA ribossomais, 10 a 15% o RNA transportador e somente 1 a 5% correspondem aos RNAm. Todavia, o RNAm eucariótico é único, devido à presença das caudas poli(A), situadas na extremidade 3' (↔ Seção 4.9), facilitando a sua obtenção, mesmo que ele seja raro. Quando um extrato celular é aplicado em uma coluna cromatográfica contendo fitas de poli(T) ligadas a um suporte de celulose, a maioria do RNAm é separada dos demais RNA pelo pareamento específico das bases A e T. O RNA é liberado da coluna com um tampão com baixa concentração de sal, originando uma preparação altamente enriquecida em RNAm.

Uma vez isolado o RNAm, é necessária a conversão da sua informação em DNA complementar (DNAc). Esse procedimento é realizado empregando-se a enzima transcriptase reversa, como ilustrado na Figura 11.6. Esse DNAc de dupla-fita contém a sequência codificadora de interesse, sendo desprovida de íntrons (Figura 11.23). Ele pode ser inserido em um plasmídeo ou outro vetor de clonagem. Entretanto, como o DNAc corresponde ao RNAm, não há a presença de um promotor e de outras sequências reguladoras a montante, as quais não são transcritas em RNA. Vetores especializados de expressão contendo promotores bacterianos e sítios de ligação aos ribossomos são utilizados para a obtenção de altos níveis de expressão de genes clonados dessa maneira (↔ Seção 11.9).

#### Encontrando o gene via proteína

Conhecendo-se a sequência de um RNAm é possível a produção do DNAc para clonagem. Em alguns casos, no entanto, apenas a sequência de aminoácidos da proteína desejada é conhecida. A sequência de aminoácidos de uma proteína pode ser usada para desenhar e sintetizar uma sonda de oligonucleotídeo que a codifica. Esse procedimento é ilustrado na Figura 11.24. Infelizmente, a degeneração do código genético dificulta essa abordagem. A maioria dos aminoácidos é codifica-

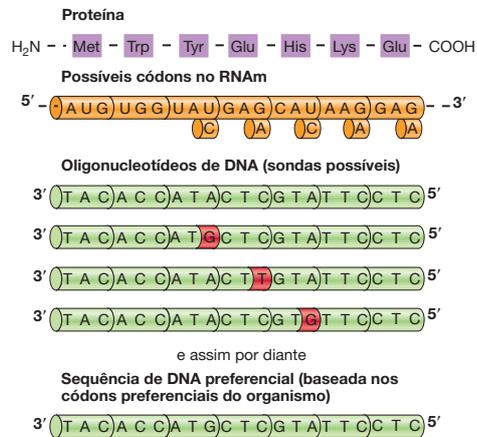


**Figura 11.23** DNA complementar (DNAc). Ilustração das etapas de síntese de um DNAc sem íntrons, correspondente a um gene eucariótico, utilizando a transcrição reversa – PCR (RT-PCR).

da por mais de um códon (↔ Tabela 4.5), e o uso preferencial de códons varia de organismo para organismo. Dessa forma, a melhor região de um gene para a síntese de uma sonda corresponde à região da proteína rica em aminoácidos codificados por um único códon (p. ex., metionina, AUG; triptofano, UGG) ou, no máximo, por dois códons (p. ex., fenilalanina, UUU, UUC; tirosina, CAU, CAC). Essa estratégia aumenta as possibilidades da sonda de DNA ser praticamente complementar ao RNAm ou ao gene de interesse. Se a sequência completa de aminoácidos da proteína não for conhecida, dados parciais de sequência podem ser utilizados.

No caso de certas proteínas pequenas, a síntese do gene completo pode ser uma estratégia interessante (Seção 11.5). Muitas proteínas de mamíferos (incluindo hormônios peptídicos de alto valor comercial) são produzidas pela clivagem de precursores maiores. Assim, para a produção de um hormônio peptídico pequeno, como a insulina, a construção de um gene artificial que codifica apenas o hormônio final, em lugar da proteína precursora maior do qual é naturalmente derivado, consiste em uma abordagem mais eficiente. A síntese química também permite a obtenção de genes modificados, os quais podem produzir novas proteínas úteis. Atualmente, as técnicas de síntese de moléculas de DNA estão bem estabelecidas, o que possibilita a síntese de genes que codificam proteínas com mais de 200 resíduos de aminoácidos de extensão (600 nucleotídeos). A abordagem sintética foi primeiramente utilizada de forma significativa, na produção do hormônio insulina humana, em bactérias. Também deve-se levar em consideração que um gene é desprovido de íntrons e, assim, o RNAm não requer qualquer processamento. Além disso, promotores e outras sequências reguladoras podem ser facilmente inseridos no gene, a montante à região codificadora, e a utilização preferencial de códons (↔ Seções 4.11 e 6.3) pode ser considerada.

Pela utilização dessas técnicas, muitas proteínas humanas e virais foram expressas com alto rendimento sob o controle de sistemas reguladores bacterianos. Elas incluem insulina, somatostatina, proteínas de capsídeo viral e interferon.



**Figura 11.24** Dedução da melhor sequência de uma sonda oligonucleotídica, a partir da sequência de aminoácidos da proteína. Como muitos aminoácidos são codificados por múltiplos códons, muitas sondas de ácidos nucleicos são possíveis para uma determinada sequência polipeptídica. Se o uso preferencial de códons pelo organismo estudado for conhecido, uma sequência preferencial pode ser selecionada. Não é essencial obter-se exatamente total, já que alguns pareamentos incorretos podem ser tolerados, principalmente com sondas mais longas.

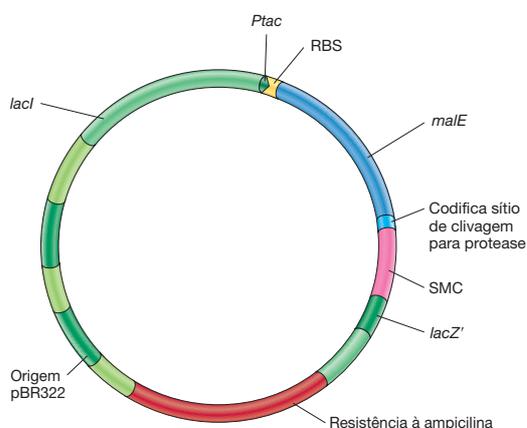
**Dobramento e estabilidade de proteínas**

A síntese de uma proteína em um novo hospedeiro é, algumas vezes, acompanhada por problemas adicionais. Por exemplo, algumas proteínas são suscetíveis à degradação por proteases intracelulares e podem ser destruídas antes de serem isoladas. Além disso, algumas proteínas eucarióticas são tóxicas para os hospedeiros procarióticos e, desse modo, o hospedeiro do vetor de clonagem pode ser morto antes que uma quantidade suficiente do produto seja sintetizada. Modificações adicionais, tanto no hospedeiro quanto no vetor, podem eliminar esses problemas.

Ocasionalmente, quando proteínas exógenas são produzidas em grandes quantidades, podem formar corpos de inclusão no interior do hospedeiro. Os corpos de inclusão consistem em agregados insolúveis de proteínas que frequentemente não se dobram corretamente ou são parcialmente desnaturadas, sendo muitas vezes tóxicas para a célula hospedeira. Embora os corpos de inclusão sejam relativamente fáceis de purificar devido ao seu tamanho, as proteínas que contêm são de difícil solubilização e podem estar inativas. Uma solução possível para esse problema consiste na utilização de um hospedeiro que produza chaperonas moleculares em grandes quantidades, as quais auxiliarão no dobramento (↔ Seção 4.14).

**Proteínas de fusão para a purificação otimizada**

A purificação da proteína pode ser frequentemente simplificada se a proteína-alvo for produzida como uma *proteína de fusão*, carregando uma proteína codificada pelo vetor. Para tanto, os dois genes são fusionados, originando uma única sequência codificadora. Um pequeno segmento reconhecido e clivado por uma protease comercialmente disponível é inserido entre os dois genes. Após a transcrição e tradução, uma única proteína é sintetizada. Ela é purificada pelos métodos desenvolvidos para a proteína carreadora. A proteína de fusão



**Figura 11.25** Um vetor de expressão para proteínas fusionadas. O gene a ser clonado é inserido no SMC, de modo a ficar em fase com o gene *malE*, que codifica a proteína de ligação à maltose. Essa inserção inativa o fragmento alfa do gene *lacZ*, que codifica a  $\beta$ -galactosidase. O gene fusionado é controlado pelo promotor híbrido *tac* (*P<sub>tac</sub>*) e em um sítio de ligação de ribossomo (SLR). O plasmídeo também contém o gene *lacI*, que codifica o repressor *lac*. Portanto, um indutor deve ser adicionado às células, a fim de ativar o promotor *tac*. O plasmídeo contém um gene que confere ao hospedeiro a resistência à ampicilina. Esse vetor foi desenvolvido pela New England Biolabs.

é então clivada pela protease, para a liberação da proteína-alvo da proteína carreadora. Proteínas de fusão simplificam a purificação da proteína-alvo, porque proteínas carreadoras podem ser escolhidas com base nas propriedades ideais para a purificação.

Vários vetores de fusão são comercializados, visando a geração de proteínas de fusão. A **Figura 11.25** ilustra um exemplo de um vetor de fusão que é também um vetor de expressão. Nesse exemplo, a proteína carreadora é a proteína de ligação à maltose de *Escherichia coli*, sendo facilmente purificada por métodos baseados em sua afinidade por maltose. Uma vez purificada, as duas porções da proteína de fusão são separadas por uma protease específica. Em alguns casos, a proteína-alvo clonada é liberada da proteína carreadora por tratamentos químicos específicos, em vez de clivagem proteolítica.

Sistemas de fusão são também utilizados com outras finalidades, além do aumento da estabilidade proteica. Uma vantagem na produção de proteínas de fusão está no fato de a proteína carreadora poder ser escolhida de modo a conter a sequência bacteriana que codifica o *peptídeo-sinal*, um peptídeo rico em aminoácidos hidrofóbicos, que possibilita o transporte da proteína através da membrana citoplasmática (↔ Seção 4.14). Isso torna possível que um sistema de expressão bacteriana, além de produzir proteínas de mamíferos, também as secrete. Utilizando as linhagens e os vetores adequados, a proteína desejada pode corresponder a 40% das moléculas proteicas em uma célula.

#### MINIQUESTIONÁRIO

- Qual a principal vantagem da clonagem de genes de mamíferos a partir do RNAm, ou utilizando genes sintéticos, apresenta em relação à amplificação por PCR e clonagem do gene nativo?
- Como uma proteína de fusão é produzida?

## 11.12 Somatotrofina e outras proteínas de mamíferos

Atualmente, as áreas da biotecnologia mais robustas economicamente são a produção de proteínas humanas e o uso de organismos geneticamente modificados na agricultura. Muitas proteínas de mamíferos apresentam alto valor farmacêutico, mas geralmente são encontradas em quantidades muito pequenas no tecido normal, tornando sua purificação extremamente dispendiosa. Mesmo que a proteína possa ser produzida em cultura de células, isso é muito mais dispendioso e difícil que em culturas microbianas, que a produzem em altas quantidades. Por essa razão, a indústria biotecnológica dispõe de microrganismos geneticamente modificados para a produção de muitas proteínas diferentes de mamíferos.

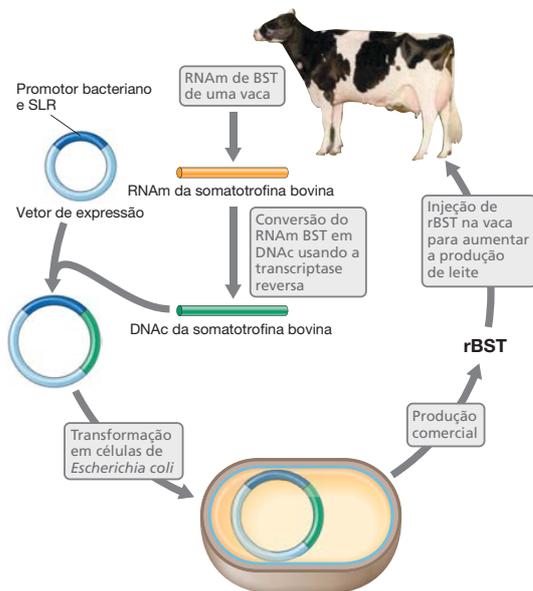
### Somatotrofina modificada por engenharia genética

Apesar de a insulina ter sido a primeira proteína humana a ser produzida dessa forma, o procedimento apresenta várias complicações não usuais, porque a insulina consiste em dois polipeptídeos curtos, conectados por ligações dissulfeto. Um exemplo mais típico na *somatotrofina* (hormônio de crescimento), que abordaremos a partir de agora.

O hormônio de crescimento, ou somatotrofina, consiste em um polipeptídeo codificado por um único gene. A ausência de somatotrofina resulta em nanismo hereditário. Pelo fato de o gene da somatotrofina humana ter sido clonado e expresso com sucesso em bactérias, crianças apresentando crescimento retardado podem ser tratadas com a *somatotrofina humana recombinante* (HSTr). Contudo, o nanismo pode também ser causado pela ausência do receptor de somatotrofina. Nesse caso, a administração de somatotrofina não exibe qualquer efeito. (Indivíduos das tribos de pigmeus africanos apresentam concentrações normais de somatotrofina humana, porém raramente apresentam estatura maior que 1,47 metro, porque são deficientes nos receptores do hormônio de crescimento.)

O gene de somatotrofina foi clonado na forma DNAC, a partir do RNAm, conforme descrito na Seção 11.11 (**Figura 11.26**). O DNAC foi então expresso em um vetor de expressão bacteriano. O principal problema na produção de hormônios polipeptídicos relativamente curtos, como a somatotrofina, refere-se a sua suscetibilidade à digestão por proteases. Tal problema pode ser solucionado pelo uso de linhagens bacterianas deficientes em várias proteases.

A *somatotrofina bovina recombinante* (rBST) é utilizada na indústria de laticínios (**Figura 11.26**). A injeção de rBST em vacas não resulta em um maior crescimento; e, em vez disso, estimula a produção de leite. Isso ocorre porque a somatotrofina possui dois sítios de ligação. Um deles liga-se ao receptor de somatotrofina e estimula o crescimento, o outro liga-se ao receptor de prolactina e promove a produção de leite. A produção excessiva de leite pelas vacas provoca alguns problemas de saúde nos animais, incluindo uma maior frequência de infecções do úbere e a diminuição da capacidade reprodutiva. Quando a somatotrofina é utilizada no tratamento de distúrbios no crescimento humano, é desejável evitarem-se os efeitos colaterais da atividade prolactina (a prolactina estimula a lactação) decorrentes do hormônio. A mutagênese sítio-dirigida (Seção 11.5) do gene de somatotrofina foi utilizada para modificá-la geneticamente, impedindo sua ligação ao re-



**Figura 11.26** Clonagem e expressão da somatotrofina bovina. O RNAm da somatotrofina bovina (BST) é obtido de um animal. O RNAm é convertido em DNAc pela transcriptase reversa. A versão de DNAc do gene de somatotrofina é então clonado em um vetor de expressão bacteriano que possui um promotor e um sítio de ligação de ribossomo (SLR). A construção é transformada em células de *Escherichia coli*, e a somatotrofina bovina recombinante (rBST) é produzida. Vacas tratadas com rBST apresentam maior produção de leite.

ceptor de prolactina. Para realizar-se isso, vários aminoácidos necessários à ligação ao receptor da prolactina foram alterados por mutação da sequência codificadora. Assim, é possível não apenas produzir hormônios humanos genuínos, mas também alterar suas especificidade e atividade, a fim de torná-los produtos farmacêuticos melhores.

### Outras proteínas e produtos de mamíferos

Muitas outras proteínas de mamíferos são produzidas por engenharia genética (Tabela 11.2). Elas incluem, em particular, uma variedade de hormônios e proteínas envolvidos na coagulação sanguínea e em outros processos sanguíneos. Por exemplo, o *ativador de plasminogênio tecidual* (TPA, *tissue plasminogen activator*) é uma proteína sanguínea que remove e dissolve coágulos que podem ser formados nos estágios finais do processo de cura. O TPA é principalmente útil para pacientes cardíacos ou outros que sofrem de problemas de insuficiência circulatória devido à excessiva formação de coágulos. TPA é administrado após ataques cardíacos, pontes cardíacas, transplantes e outras cirurgias cardíacas, para impedir o desenvolvimento de coágulos que podem trazer risco à vida. As doenças cardíacas são uma das principais causas de mortes em vários países desenvolvidos, especialmente nos Estados Unidos, de modo que a produção microbiana de TPA tem uma alta demanda.

Os fatores de coagulação sanguínea VII, VIII e IX são produtos importantes da engenharia genética. Contrariamente ao TPA, tais proteínas têm importância crítica na *formação* de

**Tabela 11.2** Alguns produtos terapêuticos obtidos por engenharia genética

Produto	Função
<b>Proteínas sanguíneas</b>	
Eritropoietina	Tratamento de certos tipos de anemia
Fatores VII, VIII e IX	Promove a coagulação
Ativador de plasminogênio tecidual	Dissolução de coágulos
Urocinase	Promove a coagulação sanguínea
<b>Hormônios humanos</b>	
Fator de crescimento epidérmico	Cicatrização de ferimentos
Hormônio foliculo-estimulante	Tratamento de distúrbios reprodutivos
Insulina	Tratamento de diabetes
Fator de crescimento neural	Tratamento de distúrbios neurológicos degenerativos e derrame
Relaxina	Facilitação do parto
Somatotrofina (hormônio de crescimento)	Tratamento de algumas anormalidades de crescimento
<b>Imunomoduladores</b>	
Interferon $\alpha$	Agente antiviral, antitumoral
Interferon $\beta$	Tratamento de esclerose múltipla
Fator estimulador de colônias	Tratamento de infecções e câncer
Interleucina 2	Tratamento de certos tipos de câncer
Lisozima	Anti-inflamatório
Fator de necrose tumoral	Agente antitumoral, tratamento potencial da artrite
<b>Enzimas de reposição</b>	
$\beta$ -glicosidase	Tratamento da doença de Gaucher, uma doença neurológica hereditária
<b>Enzimas terapêuticas</b>	
Dnase I humana	Tratamento de fibrose cística
Alginato liase	Tratamento de fibrose cística

coágulos sanguíneos. Os hemofílicos sofrem de uma deficiência em um ou mais fatores de coagulação, podendo, portanto, ser tratados com fatores de coagulação produzidos por microrganismos. Fatores de coagulação recombinantes exibem maior importância quando se considera que, no passado, os hemofílicos eram tratados com fatores de coagulação concentrados oriundos de um conjunto de amostras sanguíneas humanas, obtidas de vários doadores, algumas das quais encontravam-se contaminadas por vírus como o HIV e da hepatite C, sujeitando os hemofílicos a um alto risco de contrair essas doenças. Os fatores de coagulação recombinantes eliminaram esse risco à saúde.

Algumas proteínas de mamíferos produzidas por engenharia genética consistem em enzimas, em vez de hormônios (Tabela 11.2). Por exemplo, a *Dnase I humana* é produzida e utilizada no tratamento do acúmulo de muco contendo DNA, em pacientes com fibrose cística. O muco é formado porque a fibrose cística é acompanhada por infecções

pulmonares graves, causadas pela bactéria *Pseudomonas aeruginosa*. As células bacterianas formam biofilmes (↔ Seções 7.9 e 19.4) no interior dos pulmões, dificultando o tratamento com fármacos. DNA é liberado quando as células bacterianas sofrem lise e isso contribui para a formação do muco. A DNase digere o DNA e reduz acentuadamente a viscosidade do muco.

#### MINIQUESTIONÁRIO

- Qual é a vantagem de utilizar a engenharia genética para fazer insulina?
- Quais são os principais problemas na produção de proteínas em bactérias?
- Explique como uma enzima que degrada DNA pode ser útil no tratamento de uma infecção bacteriana, como a que ocorre no caso da fibrose cística.

### 11.13 Organismos transgênicos na agricultura e na aquicultura

O melhoramento genético de plantas por métodos tradicionais de seleção e cruzamento tem uma longa história, mas a tecnologia do DNA recombinante promoveu alterações revolucionárias. Por um lado, a engenharia genética de organismos superiores não é, verdadeiramente, microbiologia. Por outro lado, muitas das manipulações de DNA são realizadas utilizando-se bactérias e seus genes e plasmídeos (ver a seguir, herbicidas e plantas resistentes a insetos) muito antes do transgene modificado ser finalmente inserido em uma planta ou um animal. Portanto, enfatizaremos os sistemas microbianos que contribuíram para a manipulação genética de plantas e animais.

Devido às plantas e animais produtos da engenharia genética possuírem um gene de outro organismo – denominado *transgene* –, eles são **organismos transgênicos**. A população conhece esses organismos como **organismos geneticamente modificados (OGMs)**. De forma estrita, esse termo, *geneticamente modificado*, refere-se a qualquer organismo modificado por engenharia genética, contendo ou não DNA exógeno. Nesta seção, discutiremos como genes exógenos são inseridos

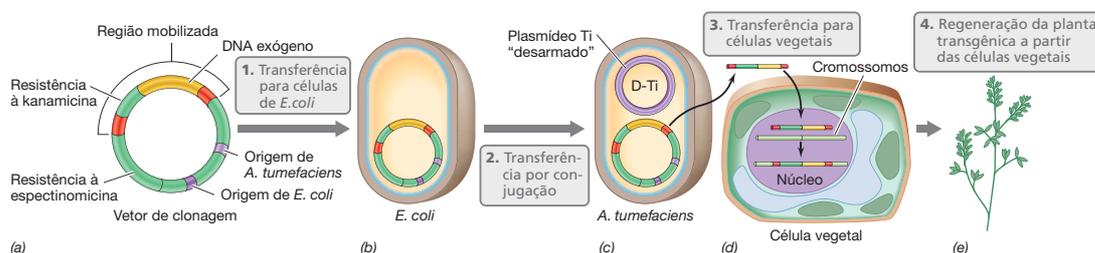
no genoma de plantas e como essas plantas transgênicas podem ser utilizadas.

#### O plasmídeo Ti e as plantas transgênicas

A engenharia genética pode modificar o DNA da planta e então utilizá-lo na transformação de células vegetais por métodos de eletroporação ou transfecção (ver Figura 11.28). Alternativamente, podem-se utilizar plasmídeos do patógeno gram-negativo de plantas, *Agrobacterium tumefaciens*, contém um grande plasmídeo, denominado **plasmídeo Ti**, que naturalmente transfere DNA diretamente para as células de certos tipos de plantas e é responsável por sua virulência. Esse plasmídeo contém genes que mobilizam o DNA, que será transferido para a planta, a qual desenvolverá, como consequência, a doença galha-da-coroa (↔ Seção 22.4). O segmento de DNA do plasmídeo Ti que é, de fato, transferido para a planta, é denominado **T-DNA**. As sequências nas extremidades do T-DNA são essenciais para a transferência, e o DNA a ser transferido deve estar situado entre essas extremidades.

Um sistema vetor-Ti comum utilizado para a transferência de genes para plantas consiste no sistema com dois plasmídeos denominado *vetor binário*, que compreende um vetor de clonagem e um plasmídeo auxiliar. O vetor de clonagem contém as duas extremidades do T-DNA flanqueando um sítio múltiplo de clonagem e um marcador de resistência a antibióticos, que pode ser utilizado em plantas. O plasmídeo também possui duas origens de replicação, de modo que pode se replicar em *A. tumefaciens* e *Escherichia coli* (essa última é a hospedeira para os procedimentos de clonagem) e outro marcador de resistência para a seleção em bactérias. O DNA exógeno é inserido no vetor, que é então transformado em *E. coli* e transferido para *A. tumefaciens* por conjugação (Figura 11.27).

Esse vetor de clonagem não possui os genes necessários à transferência do T-DNA para a planta. Entretanto, quando inserido em uma célula de *Agrobacterium* contendo um plasmídeo auxiliar adequado, pode ocorrer a transferência do T-DNA para a planta. O plasmídeo auxiliar “desarmado”, denominado *D-Ti*, contém a região de virulência (*vir*) do plasmídeo Ti, mas não apresenta o T-DNA. Portanto, pode direcionar a transferência de DNA para a planta, mas é desprovido do T-DNA. Assim, este pode dirigir a transferência do DNA para



**Figura 11.27** Produção de plantas transgênicas, utilizando-se um sistema de vetor binário em *Agrobacterium tumefaciens*. (a) Vetor geral de clonagem em plantas, contendo as extremidades do T-DNA (em vermelho), DNA exógeno, origens de replicação e marcadores de resistência. (b) O vetor é introduzido em células de *E. coli* com finalidade de clonagem, sendo posteriormente transferido para *A. tumefaciens*, por conjugação. (c) O plasmídeo Ti residente (D-Ti), utilizado para a transferência do vetor à planta, foi também

modificado geneticamente para a remoção dos genes essenciais associados à patogenicidade. (d) Todavia, o D-Ti pode mobilizar a região do T-DNA do vetor, transferindo-o às células vegetais que se desenvolvem em culturas de tecidos. (e) A partir de uma célula recombinante, plantas completas podem ser regeneradas. Detalhes da transferência do plasmídeo Ti de uma bactéria para uma planta são mostrados na Figura 22.21.

uma planta, sem possuir os genes responsáveis pela doença. Desse modo, o plasmídeo auxiliar fornece todas as funções necessárias à transferência do T-DNA a partir do vetor de clonagem. O DNA clonado e o marcador de resistência à kanamicina são mobilizados pelo D-Ti e transferidos para uma célula vegetal (Figura 11.27d). Após a integração no cromossomo da planta, o DNA exógeno pode ser expresso, conferindo novas propriedades a esta.

Várias plantas transgênicas foram produzidas utilizando-se o plasmídeo Ti de *A. tumefaciens*. O sistema Ti funciona bem em plantas de folhas largas (dicotiledôneas), incluindo culturas de tomate, tabaco, soja, alfafa e algodão. Este foi também utilizado na produção de árvores transgênicas, como nogueira e macieira. O sistema não funciona com plantas da família das gramíneas (monocotiledôneas, incluindo o milho, uma importante planta cultivável), porém outros métodos de introdução de DNA, como a transfecção por bombardeamento de microprojéteis com uma pistola de partículas (biobalística) (Figura 11.28), foram utilizados com sucesso nesses vegetais.

### Plantas resistentes a herbicidas e insetos

As principais áreas voltadas ao melhoramento genético de plantas incluem a resistência a herbicidas, insetos e doenças microbianas, assim como produtos com melhor qualidade. Atualmente, as principais culturas geneticamente modificadas (GM) são a soja, o milho, o algodão e a canola. Praticamente toda a soja e a canola cultivadas são resistentes a herbicidas, enquanto o milho e o algodão são resistentes a herbicidas, a insetos, ou a ambos.

A resistência a herbicidas é inserida por modificações genéticas de uma planta de interesse agrícola, a fim de protegê-la dos herbicidas aplicados para matar ervas daninhas. Muitos herbicidas inibem uma enzima ou proteína vegetal essencial, necessária ao crescimento. Por exemplo, o herbicida *glifosato* (Roundup™) mata plantas pela inibição de uma enzima necessária à síntese de aminoácidos aromáticos. Algumas bactérias



**Figura 11.29** Plantas transgênicas: resistência a herbicidas. A fotografia mostra parte de um campo de soja tratada com Roundup™, um herbicida à base de glifosato, produzido pela Monsanto Company (St. Louis, Missouri, EUA). As plantas, à direita, são a soja normal; aquelas à esquerda foram modificadas geneticamente, expressando a resistência ao glifosato.

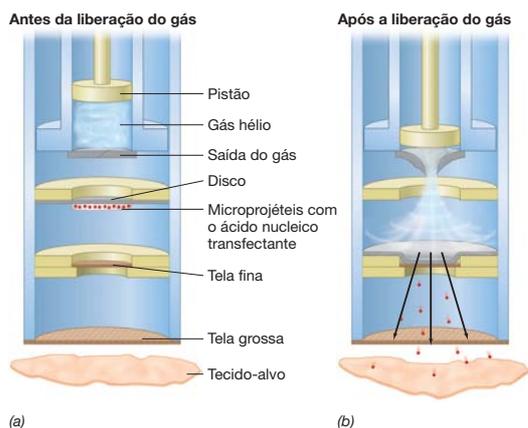
contêm uma enzima equivalente, sendo também mortas pelo glifosato. Contudo, bactérias mutantes foram selecionadas por apresentarem resistência ao glifosato, as quais continham uma forma resistente da enzima. O gene que codifica essa enzima resistente de *A. tumefaciens* foi clonado, modificado de forma a ser expresso em plantas, e transferido para importantes culturas vegetais, como a soja. Quando o glifosato é aplicado, as plantas que contêm o gene bacteriano não são mortas (Figura 11.29). Assim, o glifosato é utilizado para matar as ervas daninhas que competem pela água e por nutrientes com as plantas de interesse comercial que estão sendo cultivadas. Hoje, a soja resistente a herbicidas é amplamente cultivada nos Estados Unidos.

### Resistência a insetos: a toxina Bt

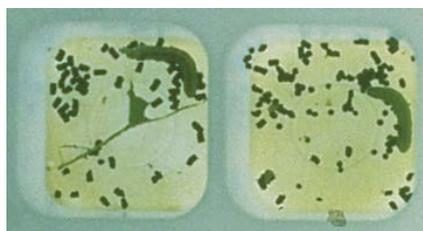
A resistência a insetos foi também introduzida geneticamente em plantas (Figura 11.30). Uma abordagem amplamente utilizada baseia-se na introdução de genes que codificam a proteína tóxica de *Bacillus thuringiensis* em plantas. *B. thuringiensis* produz uma proteína cristalina, denominada *toxina Bt* (⇨ Seção 15.8), que é tóxica para larvas de borboletas e mariposas. Existem muitas variantes da toxina Bt, específicos para diferentes insetos. Algumas linhagens de *B. thuringiensis* produzem proteínas adicionais, as quais são tóxicas para larvas de besouro e moscas e para mosquitos.

Várias abordagens diferentes foram utilizadas para aumentar a eficácia da toxina Bt no controle de pragas em plantas. Uma abordagem foi o desenvolvimento de um conjunto único de toxinas Bt, eficaz contra vários insetos diferentes. Uma abordagem efetiva para a obtenção da expressão e estabilidade do transgene Bt foi sua transferência diretamente para o genoma da planta. Por exemplo, um gene natural da toxina Bt foi clonado em um vetor plasmidial sob o controle do promotor de RNAr de cloroplasto, o qual foi transferido para cloroplastos de tabaco por bombardeamento de microprojéteis (Figura 11.28). Empregando-se tal metodologia, foram obtidas plantas transgênicas que expressavam essa proteína em níveis extremamente tóxicos às larvas de várias espécies de insetos.

A toxina Bt é inócua para mamíferos, incluindo o homem, por diversas razões. Primeiro, a cocção e o processamento dos alimentos a destroem. Segundo, nenhuma toxina ingerida é



**Figura 11.28** Pistola disparadora de DNA para a transfecção de células eucarióticas. Os mecanismos internos da pistola ilustram como as esferas metálicas revestidas com ácidos nucleicos (microprojéteis) são disparadas em células-alvo. (a) Antes do disparo e (b) depois do disparo. Uma onda de choque devido à liberação do gás lança o disco carregando os microprojéteis contra a tela fina. Os microprojéteis prosseguem até o tecido-alvo.



(a)



(b)

**Figura 11.30 Plantas transgênicas: resistência a insetos.** (a) Resultados de dois ensaios distintos de determinação do efeito das larvas da lagarta da beterraba em folhas de tabaco oriundas de plantas normais. (b) Resultados de ensaios similares, utilizando folhas de tabaco de plantas transgênicas, que expressam a toxina Bt em seus cloroplastos.

digerida no trato gastrointestinal de mamíferos. Terceiro, a toxina Bt atua ligando-se a receptores específicos no intestino dos insetos, os quais não são encontrados no intestino de outros grupos de organismos. A ligação promove uma alteração conformacional da toxina, que passa a formar poros no revestimento intestinal do inseto, os quais comprometem o seu sistema digestivo, matando-o.

### Peixes transgênicos

Muitos genes exógenos já foram incorporados e expressos em animais de laboratório de pesquisa e também em animais com importância comercial. A engenharia genética utiliza técnicas como a microinjeção para introduzir genes clonados em ovos fertilizados; em seguida, por recombinação genética, o DNA exógeno é incorporado ao genoma dos ovos. Mais recentemente, animais domésticos, incluindo os peixes, foram geneticamente modificados para melhorar o rendimento.

Um exemplo prático da transgenia animal muito interessante é o salmão transgênico, modificado para um crescimento rápido (do inglês, *fast-growing salmon*) (Figura 11.31). Esse salmão transgênico não ficará maior do que os animais comuns, mas atinge o tamanho de mercado muito mais rápido. O gene do hormônio de crescimento no salmão selvagem é ativado pela luz. Consequentemente, o salmão tem um crescimento acelerado somente durante os meses do verão. Nos salmões geneticamente modificados, o promotor do gene do hormônio de crescimento foi substituído pelo de outra espécie desse peixe que cresce a uma taxa mais ou menos constante durante todo o ano. O resultado é um salmão com o crescimento constante e, conseqüentemente, mais rápido. Esses salmões podem ter um maior rendimento comercial em operações de aquicultura e atingem o tamanho de venda muito mais rápido que os salmões que não são OGM quando criados em cativeiro.



**Figura 11.31 Salmão transgênico de crescimento rápido.** O salmão *AquAdvantage™* (acima) foi modificado geneticamente pela empresa Aqua Bounty Technologies (St Johns, Newfoundland, Canadá). Ambos, o peixe transgênico e o controle, têm 18 meses de vida e pesam 4,5 kg e 1,2 kg, respectivamente.

### MINIQUESTIONÁRIO

- O que é uma planta transgênica?
- Dê um exemplo de uma planta geneticamente modificada e descreva como essa modificação genética beneficia a agricultura.
- Como o salmão transgênico foi modificado para atingir o tamanho de mercado mais rapidamente?

### 11.14 Vacinas produzidas por engenharia genética

Vacinas são substâncias que induzem imunidade contra uma determinada doença, quando injetadas em um animal (↔ Seção 24.6). Normalmente, vacinas são suspensões de microrganismos patogênicos ou vírus mortos ou modificados (ou de componentes específicos isolados deles). Frequentemente, o componente que desencadeia a resposta imune é uma proteína de superfície, como, por exemplo, uma proteína do capsídeo ou do envelope viral. A engenharia genética pode ser aplicada de muitas formas diferentes na produção de vacinas.

#### Vacinas recombinantes

Técnicas de DNA recombinante podem ser utilizadas para modificar o próprio patógeno. Por exemplo, pode-se deletar genes do patógeno que codificam fatores de virulência, mantendo aqueles cujos produtos induzem uma resposta imune. Isso gera uma vacina recombinante atenuada viva. Por outro lado, podem-se adicionar genes de um vírus patogênico em outro vírus relativamente inofensivo, denominado *vírus carreador*. Tais vacinas são denominadas **vacinas de vetores** e induzem imunidade à doença causada pelo vírus patogênico. De fato, pode-se até mesmo associar as duas abordagens. Por exemplo, uma vacina recombinante é utilizada para proteger aves domésticas contra o epitélioma contagioso (uma doença que reduz o ganho de peso e a produção de ovos) e a doença de Newcastle (uma doença viral, frequentemente fatal). O vírus do epitélioma contagioso (um poxvírus típico; ↔ Seção 9.6) foi inicialmente modificado pela deleção dos genes que causavam a doença, mas não daqueles que induziam imunidade. Em seguida, os genes indutores de imunidade do vírus Newcastle foram introduzidos. Isso resultou em uma **vacina polivalente**, uma vacina única que imuniza contra duas doenças diferentes.

O vírus vaccínia é amplamente utilizado na preparação de vacinas recombinantes vivas de uso humano (↔ Seção 9.6). O vírus vaccínia geralmente não é patogênico para seres humanos, e tem sido utilizado há mais de 100 anos como uma vacina contra o vírus da varíola relacionado. Entretanto, a clonagem de genes no vírus vaccínia requer um marcador seletivo. Ele é o gene da timidina-cinase. O vírus vaccínia, contrariamente a vários outros vírus, possui sua própria timidina-cinase, uma enzima que converte o análogo de base 5-bromodesoxiuridina em um nucleotídeo que é incorporado no DNA. Entretanto, essa é uma reação letal. Portanto, células que expressam timidina-cinase (seja a partir do genoma da célula hospedeira, seja de um genoma viral) são mortas pelo 5-bromodesoxiuridina.

Genes a serem inseridos no vírus vaccínia são inicialmente clonados em um plasmídeo de *Escherichia coli* que contém um fragmento do gene da timidina-cinase (*tk*) do vírus vaccínia (Figura 11.32). O DNA exógeno é inserido no gene *tk*, o qual se torna, portanto, inativo. Esse plasmídeo recombinante é então transformado em células animais cujo gene de timidina-cinase encontra-se inativado. As mesmas células são também infectadas com o vírus vaccínia selvagem. Ocorre a recombinação homóloga entre as duas versões do gene *tk* – uma do plasmídeo e outra do vírus. Portanto, alguns vírus adquirem um gene *tk* inativo, contendo o inserto exógeno (Figura 11.32). As células infectadas pelo vírus selvagem, contendo a timidina-cinase ativa, são mortas pela 5-bromodesoxiuridina. As células infectadas pelo vírus vaccínia recombinante, contendo o gene *tk* inativo, crescem o suficiente para originar uma nova geração de partículas virais (Figura 11.32). Como resultado, os vírus cujo gene *tk* contém um inserto de DNA exógeno clonado são selecionados.

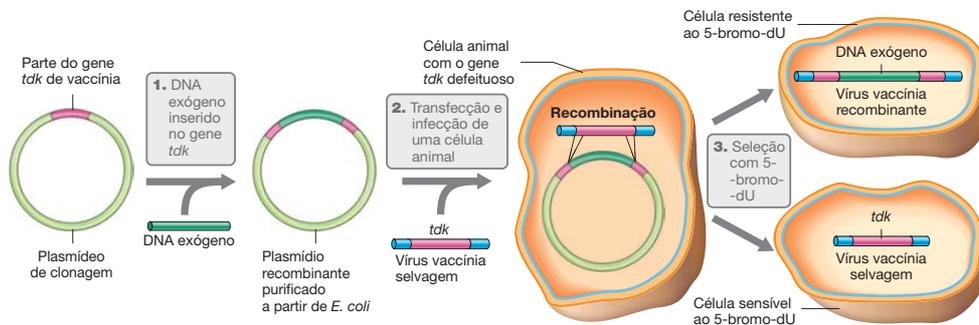
O vírus vaccínia, na realidade, não requer a timidina-cinase para a sua sobrevivência. Consequentemente, os vírus vaccínia recombinantes podem, ainda, infectar células humanas e expressar os genes exógenos que carregam. De fato, podem ser construídos de modo a carregarem genes de múltiplos vírus (ou seja, correspondem a vacinas polivalentes). Atualmente, várias vacinas de vetor de vaccínia foram desenvolvidas e licenciadas para o uso veterinário, incluindo uma contra a raiva. Muitas outras vacinas de vaccínia encontram-se em fase de testes clínicos. As vacinas de vaccínia são relativamente benignas, altamente imunogênicas em seres humanos e, provavelmente, apresentarão uso crescente nos próximos anos.

### Vacinas de subunidades

As vacinas recombinantes não precisam conter o conjunto completo de proteínas do organismo patogênico. Vacinas de subunidades podem conter somente uma ou mais proteínas específicas de um organismo patogênico. No caso dos vírus, essa proteína frequentemente corresponde à(s) proteína(s) do capsídeo, uma vez que são, em geral, fortes imunógenos. As proteínas do capsídeo são purificadas e utilizadas em alta dose, para induzir uma imunidade rápida e em altos níveis. As vacinas de subunidade são atualmente muito populares porque podem ser utilizadas na produção de grandes quantidades de proteínas imunogênicas, descartando-se a possibilidade dos produtos purificados conterem o organismo patogênico inteiro, mesmo em quantidades diminutas. Em alguns casos, somente uma porção das proteínas virais são expressas em vez da proteína inteira, porque células imunes e anticorpos normalmente reagem apenas com porções pequenas das proteínas.

As etapas na preparação de uma vacina de subunidade viral são as seguintes: fragmentação do DNA viral por enzimas de restrição; clonagem dos genes de proteínas do capsídeo viral em um vetor adequado; uso de promotores, fase de leitura e sítio de ligação a ribossomos adequados; e reinserção e expressão dos genes virais em um microrganismo. Algumas vezes, somente determinadas porções da proteína são expressas, em lugar da proteína completa, visto que as células imunes e os anticorpos normalmente reagem somente com pequenas porções da proteína. (Quando essa metodologia é utilizada com um vírus de RNA, o genoma viral deve ser primeiramente em uma cópia de DNAc.)

Quando bactérias são utilizadas como hospedeiro de expressão, as vacinas de subunidade virais frequentemente são pouco imunogênicas e não conferem proteção em testes experimentais de infecção pelo vírus. O problema é que muitas proteínas do envoltório viral são modificadas pós-traducionalmente, em geral pela adição de resíduos de açúcar (glicosilação), pelas células hospedeiras animais, quando o vírus se replica. Contrariamente, as proteínas recombinantes produzidas por bactérias não são glicosiladas, sendo esse processo necessário para que as proteínas se tornem imunologicamente ativas. Para solucionar esse problema, um hospedeiro eucariótico é utilizado. Por exemplo, a primeira vacina de subunidades recombinante aprovada para uso em seres humanos (contra hepatite B) foi produzida em levedura. O gene que codifica



**Figura 11.32** Produção de vírus vaccínia recombinantes. O DNA exógeno é inserido em um plasmídeo contendo uma pequena porção do gene de timidina-cinase (*tk*) do vírus vaccínia. Em seguida, o plasmídeo com o inserto e o vírus vaccínia selvagem são introduzidos na mesma célula hospedeira,

onde recombinam. As células são tratadas com 5-bromodesoxiuridina (5-bromo-dU), um composto que mata as células que apresentam a forma ativa da timidina-cinase. Somente os vírus vaccínia recombinantes, cujo gene *tk* está inativado pela inserção do DNA exógeno, sobrevivem.

uma proteína de superfície do vírus da hepatite B foi clonado e expresso em levedura. A proteína foi produzida e formou agregados muito similares àqueles encontrados em pacientes infectados pelo vírus. Esses agregados foram purificados e utilizados na vacinação efetiva de seres humanos contra a infecção pelo vírus da hepatite B.

#### MINIQUESTIONÁRIO

- Explique por que as vacinas recombinantes poderiam ser mais seguras que algumas vacinas produzidas por métodos tradicionais.
- Quais as diferenças importantes entre uma vacina recombinante, uma vacina viva atenuada, uma vacina de vetor e uma vacina de subunidade?

### 11.15 Mineração genômica (garimpagem de genomas)

Assim como o conteúdo genético total de um organismo é o seu *genoma*, os genomas coletivos de um ambiente são conhecidos como o seu *metagenoma* (↔ Seções 6.10 e 18.7). Ambientes complexos, como solos férteis, contêm um enorme número de bactérias e outros microrganismos não cultivados, juntamente com os vírus que os parasitam (↔ Seção 6.10). Conjuntamente, esses organismos contêm vastos números de genes novos. De fato, a maior parte da informação genética na Terra é encontrada em microrganismos e seus vírus, os quais ainda não foram cultivados. Como a engenharia genética pode impulsionar esse recurso?

#### Garimpagem de genes ambientais

A *garimpagem de genes* consiste no processo de isolamento de novos genes potencialmente úteis a partir do ambiente, sem o cultivo dos organismos que os carregam. Em vez do cultivo, o DNA (ou RNA) é isolado diretamente de amostras ambientais e clonados em vetores adequados para a construção de uma biblioteca metagenômica (Figura 11.33). O ácido nucleico inclui genes oriundos de organismos não cultivados, assim como DNA de organismos mortos, o qual foi liberado no ambiente e que ainda não sofreu degradação. Quando o RNA é isolado, ele deve ser convertido em uma cópia de DNA por meio da transcriptase reversa (Figura 11.6). Todavia, o isolamento de RNA tem a desvantagem de requerer maior tempo e limitar a

biblioteca metagenômica àqueles genes que foram transcritos, e, portanto, ativos, na amostra ambiental.

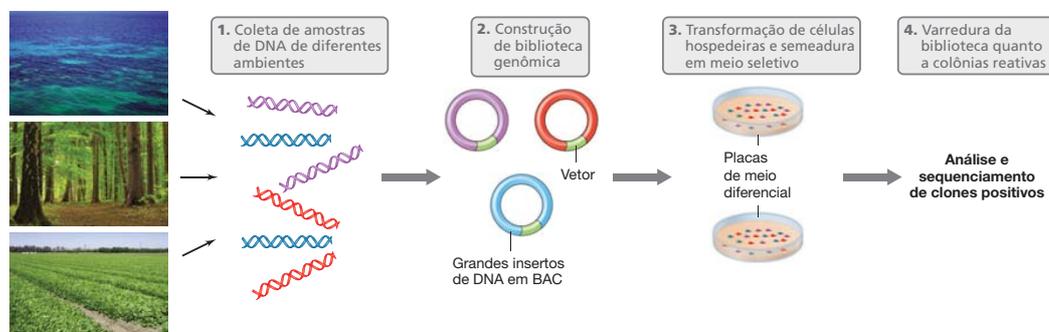
A biblioteca metagenômica é em seguida utilizada em varreduras pelas mesmas técnicas utilizadas em qualquer outra biblioteca de clones (Figura 11.8). A metagenômica identificou novos genes ambientais codificadores de enzimas que degradam poluentes e enzimas que produzem novos antibióticos. Até o momento, várias lipases, quitinases, esterases e outras enzimas que degradam novos substratos e exibem outras propriedades foram isoladas por essa abordagem. Essas enzimas são utilizadas em processos industriais, com várias finalidades. Enzimas com maior resistência às condições industriais, como alta temperatura, pH alto ou baixo e condições oxidantes, são especialmente valiosas e muito procuradas.

A descoberta de genes que codificam vias metabólicas inteiras, como a da síntese de antibióticos, contrariamente ao que ocorre no caso de genes únicos, requer vetores capazes de carrear grandes insertos de DNA, como os cromossomos artificiais bacterianos (BAC) (Seção 11.10). Os BACs são especialmente úteis à varredura de amostras provenientes de ambientes ricos, como solos, em que um grande número de genomas desconhecidos está presente e, provavelmente, há um grande número de genes a serem analisados.

#### Garimpagem direcionada de genes

A metagenômica pode ser utilizada na varredura direta de enzimas com determinadas propriedades. Suponha que se deseje uma enzima, ou uma via completa, capaz de degradar um determinado poluente em uma temperatura elevada. A primeira etapa consiste na descoberta de um ambiente quente e poluído com o composto-alvo. Presumindo-se que microrganismos capazes de degradar tal composto encontrem-se presentes nesse ambiente, uma hipótese razoável, o DNA desse ambiente poderia, então, ser isolado e clonado. Bactérias hospedeiras contendo os clones podem ser analisadas quanto ao crescimento na presença do composto-alvo. Por uma questão de facilidade, essa etapa geralmente é realizada em uma *E. coli* hospedeira, na presunção de que as enzimas termoestáveis ainda apresentarão alguma atividade a 40°C (esse é geralmente o caso). Uma vez que supostos clones tenham sido identificados, extratos enzimáticos podem ser testados *in vitro*, em temperaturas elevadas.

A estratégia de garimpagem de genes foi utilizada para isolar uma enzima lipase termoestável para aplicações comerciais. Lipases catalisam a hidrólise de triglicerídeos (gordura) e, por



**Figura 11.33** Busca metagenômica de genes ambientais úteis. Amostras de DNA são obtidas de sítios diferentes, como solo agrícola, água de

mar ou solo de florestas. Uma biblioteca de clones é construída e varrida quanto a genes de interesse. Possíveis clones úteis são analisados mais detalhadamente.

causa disso, elas algumas vezes são incorporadas em formulações cosméticas e farmacêuticas. Mas a produção industrial dessas enzimas requer que elas mantenham a atividade em temperaturas elevadas. Usando DNA isolado de microrganismos termófilos de fontes termais, biólogos moleculares criaram uma biblioteca metagenômica; a biblioteca foi usada para transformar células de *E. coli*, e colônias recombinantes expressando a atividade de lipase foram selecionadas usando um meio de cultura especial. Análises do extrato de enzimas a partir desses isolados produtores de lipase indicaram um isolado com atividade enzimática em temperaturas acima de 90°C. O gene codificando essa lipase termoestável foi identificado por análise do DNA do vetor recombinante a partir clone isolado e foi clonado em um vetor de expressão para a produção comercial.

#### MINIQUESTIONÁRIO

- Explique por que a clonagem metagenômica resulta em grandes números de novos genes.
- Quais são as vantagens e desvantagens do isolamento de RNA ambiental, comparado ao DNA?

### 11.16 Modificação de vias metabólicas por engenharia genética

Embora proteínas sejam moléculas grandes, a expressão de uma única proteína codificada por um único gene em grandes quantidades é relativamente simples. Por outro lado, pequenos metabólitos normalmente são produzidos por vias bioquímicas que contêm várias enzimas. Nesses casos, não são necessários apenas genes múltiplos, mas a sua expressão deve ser regulada de maneira coordenada.

A **engenharia de vias** é o processo de montagem de uma via bioquímica nova ou otimizada, utilizando-se genes de um ou mais organismos. A maioria dos esforços realizados até o momento modificou e otimizou vias existentes, em vez de criar vias totalmente novas. Uma vez que a engenharia genética de bactérias é mais simples do que a de organismos superiores, a maioria das engenharias de via foi realizada com bactérias. Microrganismos modificados por engenharia genética são utilizados na síntese de produtos, incluindo álcool, solventes, aditivos alimentares, corantes e antibióticos. Eles podem também ser utilizados para a degradação de resíduos agrícolas, poluentes, herbicidas e outros materiais tóxicos ou indesejáveis.

Um exemplo de engenharia de vias é a produção de índigo por *Escherichia coli* (Figura 11.34). Índigo é um corante importante utilizado no tratamento de lã e algodão. Calças jeans azuis, por exemplo, são produzidas de algodão corado com índigo. Em tempos antigos, o índigo e os corantes relacionados eram extraídos de caracóis. Em épocas mais recen-

tes, o índigo era extraído de plantas, porém, hoje, é sintetizado quimicamente. Entretanto, para atender a demanda por índigo pela indústria têxtil, novas abordagens foram adotadas em sua síntese, incluindo uma biotecnológica.

O índigo é baseado no sistema de anéis do indol, cuja estrutura global assemelha-se à do naftaleno. Consequentemente, enzimas oxigenases que oxigenam naftaleno também oxidam o indol em seu derivado di-hidroxi, o qual é espontaneamente oxidado no ar, gerando índigo, um pigmento azul-brilhante. As enzimas que oxigenam o naftaleno estão presentes em vários plasmídeos encontrados em *Pseudomonas* e outras bactérias de solo. Quando genes de tais plasmídeos foram clonados em *E. coli*, as células tornaram-se azuis devido à produção de índigo; as células azuis adquiriram os genes da enzima naftaleno oxigenase.

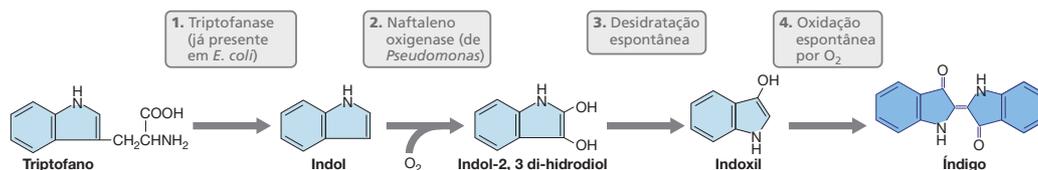
Embora somente o gene da naftaleno oxigenase tenha sido, de fato, clonado durante a engenharia de via do índigo, que consiste em quatro etapas, duas enzimáticas e duas espontâneas (Figura 11.34). A enzima que catalisa a primeira etapa, a conversão de triptofano a indol (triptofanase), ocorre naturalmente em *E. coli*. Portanto, para a produção de índigo, é necessário o fornecimento de triptofano. Isso pode ser realizado imobilizando-se as células da *E. coli* recombinante em um suporte sólido em um biorreator, e então gotejando-se uma solução de triptofano, a partir proteínas de rejeitos ou outras origens, sobre as células suspensas. A recirculação do material sobre as células por várias vezes, como é normalmente realizado nesses tipos de processos industriais de células imobilizadas, paulatinamente aumenta os níveis de índigo até o corante ser coletado.

#### MINIQUESTIONÁRIO

- Por que a engenharia de vias é mais difícil do que a clonagem e a expressão de um hormônio humano?
- Como *Escherichia coli* foi modificada a fim de produzir índigo?

### 11.17 Biologia sintética

Discutimos ao longo deste capítulo como a engenharia genética é usada para modificar genes e organismos. Contudo, a biologia, hoje, pode ir muito mais além. O termo "*biologia sintética*" refere-se ao uso da engenharia genética para criar novos sistemas biológicos a partir de partes biológicas disponíveis, frequentemente oriundas de vários organismos diferentes. Essas partes biológicas (promotores, acentuadores, operadores, *riboswitches*, proteínas reguladoras, domínios enzimáticos, sensores, etc.) foram denominadas *tijolos biológicos*. Um dos objetivos principais da biologia sintética consiste na síntese de uma célula viável a partir desses tijolos biológicos.



**Figura 11.34** Via modificada por engenharia genética, para a produção de índigo. *Escherichia coli* naturalmente expressa a triptofanase, que converte triptofano em indol. A naftaleno oxigenase (originalmente de

*Pseudomonas*) converte indol em di-hidroxi-indol que, espontaneamente, desidrata-o a indoxil. Quando exposto ao ar, o indoxil forma índigo, que é azul.

Um importante ponto de partida nessa direção ocorreu no Instituto J. Craig Venter na Califórnia (Estados Unidos), quando uma equipe de biólogos sintéticos substituiu o cromossomo inteiro da bactéria *Mycoplasma capricolum* por um genoma artificialmente sintetizado contendo 1,08 milhões de pares de bases (Mpb) baseado na sequência do genoma da bactéria *Mycoplasma mycoides*. As células de *M. capricolum* adquiriram então todas as propriedades da espécie cujo genoma passou a conter.

Um exemplo interessante da biologia sintética em menor escala é o uso de células de *Escherichia coli* modificadas geneticamente para a produção de fotografias. As bactérias modificadas são cultivadas de maneira confluyente em placas de meio sólido. Quando uma imagem é projetada no tapete de células, as bactérias que estão no escuro produzem um pigmento escuro, enquanto as bactérias que estão sob a luz não o produzem. O resultado é uma fotografia em preto e branco primitiva da imagem projetada (Figura 11.35).

A construção da *E. coli* fotográfica requereu a modificação por engenharia genética e a inserção de três módulos genéticos: (1) um detector de luz e módulo de sinalização; (2) uma via que converte heme (já presente em *E. coli*) em ficocianobilina, um pigmento fotorreceptor; e (3) uma enzima codificada por um gene, cuja transcrição pode ser ativada ou não para a síntese do pigmento escuro (Figura 11.35a). O detector de luz é uma proteína de fusão. A metade externa é a parte detectora de luz da proteína fitocromo da cianobactéria *Synechocystis*. Ela requer um pigmento especial que absorve luz, ficocianobilina (pigmento acessório para absorção de luz das cianobactérias, ⇨ Seção 13.2), que não é produzido por *E. coli*, por isso a necessidade de instalar a via de produção da ficocianobilina.

A metade interna do detector de luz é o domínio que transmite o sinal da proteína sensora EnvZ de *E. coli*. EnvZ é parte de um sistema regulador de dois componentes, sendo OmpR sua parceira (⇨ Seção 7.7). Normalmente, EnvZ ativa a proteína de ligação ao DNA, OmpR. A OmpR ativada, por sua vez, ativa genes-alvos ligando-se aos promotores. No caso que estamos apresentando, a proteína híbrida foi projetada para ativar OmpR no escuro, mas não na luz. Isso porque a

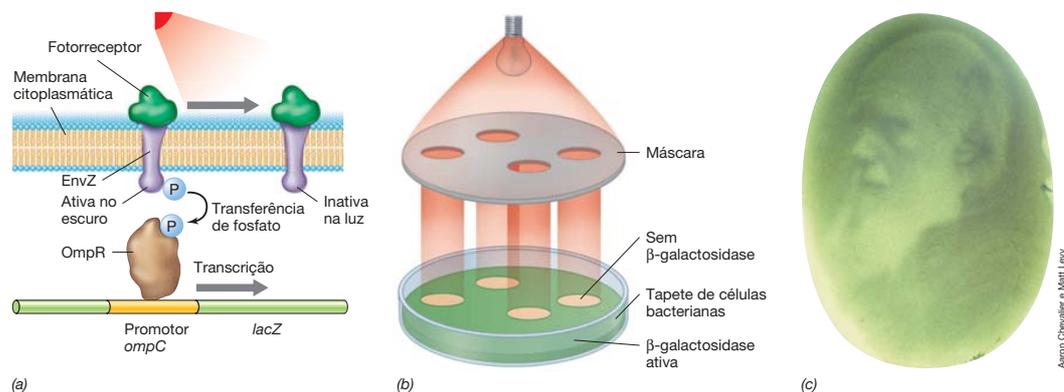
fosforilação de OmpR é necessária à sua ativação, com a luz vermelha convertendo o sensor a um estado no qual a fosforilação é inibida. Consequentemente, o gene-alvo encontra-se inativo na luz e ativo no escuro. Quando uma máscara é depositada sobre a placa de Petri contendo um tapete de células de *E. coli* modificadas por engenharia genética (Figura 11.35b), as células no escuro produzem um pigmento que não é produzido pelas células que encontram-se na área iluminada; dessa forma, uma “fotografia” da imagem da máscara é produzida (Figura 11.35c).

O pigmento sintetizado pelas células de *E. coli* é o resultado da atividade de uma enzima naturalmente encontrada nesse organismo, a qual atua no metabolismo de lactose, a  $\beta$ -galactosidase. O gene-alvo, *lacZ*, codifica essa enzima. No escuro, o gene *lacZ* é expresso e a  $\beta$ -galactosidase é produzida. A enzima cliva o análogo de lactose, denominado X-gal (Seção 11.6), presente no meio de cultura, liberando galactose e um composto colorido. Na luz, o gene *lacZ* não é expresso, a  $\beta$ -galactosidase não é produzida e o corante negro não é liberado. A diferença no contraste entre as células produtoras do corante e as células não produtoras gera a fotografia bacteriana (Figura 11.35c).

Embora a *M. capricolum* “sintética” não fosse uma célula onde todos os seus componentes – citoplasma, membranas, ribossomos e demais componentes celulares – foram feitos a partir do zero e as culturas de *E. coli* jamais substituírem a fotografia digital, o conhecimento obtido em cada um dos casos de montagem das partes necessárias por meio da biologia sintética ajuda a construir o entendimento de como a bioengenharia dos componentes funciona *in vivo*. Isso, por sua vez, vai permitir uma biologia sintética ainda mais complexa e pode algum dia levar a aplicações dessa ciência na solução de problemas urgentes nas áreas da saúde, agricultura e meio ambiente.

#### MINIQUESTIONÁRIO

- O que são tijolos biológicos?
- Como *Escherichia coli* foi modificada a fim de produzir fotografias?



**Figura 11.35 Fotografia bacteriana.** (a) Células de *Escherichia coli* detectoras de luz foram modificadas por engenharia genética, utilizando componentes de cianobactérias e da própria *E. coli*. A luz vermelha inibe a transferência de fosfato (P) para a proteína de ligação ao DNA, OmpR; a OmpR fosforilada é necessária para ativar a transcrição de *lacZ* (*lacZ* codifica a  $\beta$ -galactosidase).

(b) Sistema para a produção de uma fotografia bacteriana. As porções opacas da máscara correspondem às zonas onde a  $\beta$ -galactosidase está ativa e, portanto, às regiões escuras da imagem final. (c) Uma fotografia bacteriana de um retrato de Charles Darwin.

## CONCEITOS IMPORTANTES

**11.1 •** Enzimas de restrição reconhecem pequenas sequências específicas no DNA, clivando-o. Os produtos da digestão com enzimas de restrição podem ser separados utilizando-se eletroforese em gel.

**11.2 •** Sequências complementares de ácidos nucleicos podem ser detectadas por hibridização. Sondas compostas por DNA ou RNA de fita simples marcadas radioativamente, ou com um corante fluorescente, hibridizam-se às sequências de DNA ou RNA-alvo.

**11.3 •** A reação de polimerização em cadeia é um procedimento de amplificação de DNA *in vitro* que emprega DNA-polimerases termoestáveis. O calor é utilizado para desnaturar o DNA em duas moléculas de fitas simples, sendo cada uma copiada pela polimerase. Após cada ciclo, o DNA recém-formado é desnaturado, iniciando-se um novo ciclo de síntese. Após cada ciclo, a quantidade do DNA-alvo é duplicada.

**11.4 •** O isolamento de um gene específico, ou de uma região cromossômica, pela clonagem molecular é realizado utilizando-se um plasmídeo ou vírus como vetor de clonagem. Enzimas de restrição e DNA-ligase são utilizadas *in vitro* para produzir uma molécula quimérica de DNA, composta por DNA de uma ou mais origens. Uma vez introduzido em um hospedeiro adequado, o DNA clonado pode ser produzido em grandes quantidades, sob o controle do vetor de clonagem. A identificação dos genes clonados é conduzida por uma variedade de técnicas moleculares.

**11.5 •** Moléculas de DNA sintético, contendo sequências de interesse, podem ser sintetizadas *in vitro* e utilizadas na construção direta de genes mutados, ou na substituição de pares de bases específicos no interior de um gene, por meio de mutagênese sítio-dirigida. Genes podem também ser interrompidos pela inserção de fragmentos de DNA, denominados cassetes, que geram mutantes de nocaute.

**11.6 •** Genes repórteres são genes cujos produtos, como a  $\beta$ -galactosidase ou GFP, são de fácil análise ou detecção. Eles são utilizados para simplificar e aumentar a velocidade das análises genéticas. Em fusões gênicas, segmentos de dois genes diferentes, um dos quais normalmente corresponde a um gene repórter, são unidos.

**11.7 •** Plasmídeos são vetores de clonagem úteis porque são facilmente isolados e purificados e frequentemente capazes de multiplicarem-se em um número elevado de cópias nas células bacterianas. Genes de resistência a antibióticos nos plasmídeos são usados para selecionar as células bacterianas contendo o plasmídeo, enquanto os sistemas de seleção com base na cor são utilizados para identificar as colônias contendo DNA clonado.

**11.8 •** A escolha do hospedeiro para a clonagem depende da aplicação final. Em muitos casos, o hospedeiro pode ser um procarionte, enquanto, em outros, é essencial que

o hospedeiro seja eucariótico. Os hospedeiros devem ser capazes de captar DNA, e existe uma variedade de técnicas pelas quais tal processo pode ser realizado, tanto natural quanto artificialmente.

**11.9 •** Muitos genes clonados não são expressos de maneira eficiente em um hospedeiro exógeno. Vetores de expressão foram desenvolvidos tanto para hospedeiros procariotos quanto para eucariotos, os quais contêm genes ou sequências reguladoras que aumentam a transcrição do gene clonado e controlam o nível da transcrição. Sinais para melhorar a eficiência de tradução também devem estar presentes no vetor de expressão.

**11.10 •** Vetores especializados de clonagem, como bacteriófagos, cosmídeos e cromossomos artificiais, foram desenvolvidos para a clonagem de grandes fragmentos de DNA. Bacteriófagos contendo DNA-recombinante podem ser empacotados *in vitro*, sendo transferidos de maneira eficiente para uma célula hospedeira, enquanto os cosmídeos são vetores plasmidiais contendo o sítio *cos* de lambda. Cromossomos artificiais são úteis para a clonagem de fragmentos de DNA contendo cerca de até um megabase.

**11.11 •** É possível obter-se níveis de expressão muito altos de genes eucarióticos em procariotos. Entretanto, o gene expresso não deve conter introns. Isso pode ser realizado pela utilização da transcriptase reversa para a síntese de DNAc, a partir do RNAm maduro que codifica a proteína de interesse. Isso também pode ocorrer pela síntese de um gene totalmente sintético, a partir do conhecimento da sequência de aminoácidos da proteína de interesse. Proteínas de fusão são frequentemente utilizadas para estabilizar ou solubilizar a proteína clonada.

**11.12 •** A primeira proteína humana produzida comercialmente, utilizando bactérias modificadas por engenharia genética, foi a insulina humana, embora vários outros hormônios e proteínas humanas estejam sendo produzidos atualmente. A somatotrofina bovina recombinante é amplamente usada nos Estados Unidos para aumentar a produção de leite no gado leiteiro.

**11.13 •** A engenharia genética pode criar plantas resistentes a doenças, aumentar a qualidade do produto e tornar culturas de interesse agrícola uma fonte de proteínas recombinantes e, até mesmo, vacinas. Um vetor de clonagem utilizado comumente em plantas é o plasmídeo Ti da bactéria *Agrobacterium tumefaciens*. Este plasmídeo pode transferir DNA para as células vegetais. Plantas comerciais, cujos genomas foram modificados utilizando-se técnicas genéticas *in vitro*, são denominadas organismos geneticamente modificados ou OGMs.

**11.14 •** Muitas vacinas recombinantes foram produzidas ou estão sendo desenvolvidas. Elas incluem vacinas recombinantes vivas, de vetores e de subunidade.

**11.15** • Usando metagenômica, genes de produtos úteis são clonados diretamente de DNA ou RNA em amostras ambientais, sem o prévio isolamento dos organismos que os contêm.

**11.16** • Na engenharia de vias, múltiplos genes que codificam enzimas da via metabólica são organizados. Esses genes podem ser oriundos de um ou mais organismos, mas devem ser clonados e expressos de maneira sincronizada.

**11.17** • Em vez de modificar ou melhorar uma única via existente, a biologia sintética enfoca na engenharia de novos sistemas biológicos por meio da junção de conhecidos componentes biológicos em várias combinações, formando módulos capazes de gerar comportamentos complexos.

## REVISÃO DOS TERMOS-CHAVE

**Biblioteca de DNA** (também denominada biblioteca genômica) uma coleção de segmentos de DNA clonado, grande o suficiente para conter pelo menos uma cópia de cada gene de um organismo em particular; o mesmo que biblioteca genômica.

**Biotecnologia** utilização de organismos, geralmente contendo modificações genéticas, com finalidades industriais, médicas ou agrícolas.

**Cassete de DNA** segmento de DNA artificialmente desenvolvido, que geralmente carrega um gene de resistência a um antibiótico ou algum outro marcador conveniente e que é flanqueado por sítios de restrição adequados.

**Clonagem molecular** isolamento e incorporação de um fragmento de DNA em um vetor no qual pode ser replicado.

**Clonagem shotgun** construção de uma biblioteca gênica por meio da clonagem randômica de fragmentos de DNA.

**Códons preferenciais** proporções relativas de diferentes códons que codificam o mesmo aminoácido; elas variam em diferentes organismos. O mesmo que utilização preferencial de códons.

**Cromossomo artificial** vetor de cópia única, capaz de carrear insertos extremamente longos de DNA, muito utilizado para a clonagem de segmentos de grandes genomas.

**Cromossomo artificial bacteriano (BAC)** cromossomo circular artificial, contendo uma origem de replicação bacteriana.

**Cromossomo artificial de levedura (YAC)** cromossomo artificial contendo uma origem de replicação e a sequência CEN de levedura.

**Cromossomo artificial humano (HAC)** cromossomo artificial, contendo um arranjo da sequência centromérica humana.

**DNA recombinante** molécula de DNA que contém DNA de duas ou mais fontes de origem.

**Eletroforese em gel** técnica de separação de moléculas de ácidos nucleicos, pela aplicação de uma corrente elétrica por meio de um gel de agarose ou de poliacrilamida.

**Engenharia genética** uso de técnicas de isolamento, manipulação, recombinação e expressão de DNA (ou RNA) *in vitro*, bem como o desenvolvimento de organismos geneticamente modificados.

**Engenharia genética** uso de técnicas *in vitro* para o isolamento, alteração e expressão de DNA, e no desenvolvimento de organismos geneticamente modificados.

**Engenharia de via** montagem de uma via metabólica nova ou otimizada, utilizando-se genes de um ou mais organismos.

**Enzima de modificação** enzima que modifica quimicamente as bases em um sítio de reconhecimento de uma enzima de restrição, impedindo, dessa forma, a clivagem do sítio.

**Enzima de restrição** enzima que reconhece uma sequência específica de DNA, clivando-o; também conhecida como endonuclease de restrição.

**Fingerprinting de DNA** uso da tecnologia de DNA para a determinação da origem do DNA em uma amostra de tecido.

**Fusão gênica** estrutura criada pela união de segmentos de dois genes distintos, especialmente quando a região reguladora de um gene é ligada à região codificadora de um gene repórter.

**Fusão de operon** é uma fusão de genes em que a sequência codificadora que mantém o seu próprio sinal de tradução é fusionada com os sinais de transcrição de outro gene.

**Fusão proteica** uma fusão de genes em que duas sequências codificadoras

são fusionadas de modo que elas compartilham os mesmos sítios de início da transcrição e da tradução.

**GFP** uma proteína que brilha verde e é amplamente utilizada na análise genética.

**Gene repórter** gene utilizado em análises genéticas porque seu produto gênico é de fácil detecção.

**Hibridização** pareamento de bases de DNA ou RNA de fitas simples de duas fontes diferentes (porém relacionadas), para a obtenção de uma dupla-hélice híbrida.

**Iniciador** pequeno DNA ou RNA utilizado para iniciar a síntese de uma nova fita de DNA.

**Mapa de restrição** mapa que indica a localização de sítios de clivagem de enzimas de restrição em um segmento de DNA.

**Mutagenese por inserção de cassete** criação de mutações por inserção de um cassete de DNA.

**Mutagenese sítio-dirigida** técnica pela qual um gene contendo uma mutação específica pode ser construído *in vitro*.

**Northern blot** procedimento de hibridização no qual o RNA está presente no gel e DNA ou RNA correspondem à sonda.

**Organismo geneticamente modificado (OGM)** organismo cujo genoma foi alterado por engenharia genética. A abreviação GM é também utilizada em expressões como culturas GM e alimentos GM.

**Organismo transgênico** planta ou animal que contém um DNA exógeno inserido em seu genoma.

**Plasmídeo Ti** plasmídeo de *Agrobacterium tumefaciens*, capaz de transferir genes bacterianos para as plantas.

**Proteína de fusão** proteína resultante de uma fusão de duas proteínas diferentes, por meio da união de suas sequências codificadoras em um único gene.

**Proteína fluorescente verde** uma proteína que fluoresce em verde, amplamente utilizada em análises genéticas.

**Reação em cadeia de polimerase (PCR)** amplificação artificial de uma sequência de DNA por meio de ciclos repetidos de separação e replicação das fitas.

**Ruptura gênica** (também denominada mutagenese por cassete ou nocaute gênico) inativação de um gene pela inserção de um fragmento de DNA contendo um marcador selecionável. O fragmento inserido é denominado cassete de DNA.

**Sequenciamento** dedução da sequência de uma molécula de DNA ou RNA, por meio de uma série de reações químicas.

**Sonda de ácido nucleico** fita de ácido nucleico que pode ser marcada e utilizada para hibridizar-se com uma molécula complementar, presente em uma mistura de outros ácidos nucleicos.

**Southern blot** procedimento de hibridização em que o DNA está no gel e o RNA ou DNA é a sonda.

**T-DNA** segmento do plasmídeo Ti de *Agrobacterium tumefaciens* que é transferido para células vegetais.

**Transcrição reversa** a conversão de uma sequência de RNA em uma sequência de DNA correspondente.

**Vacina de DNA** vacina que utiliza o DNA de um patógeno para eliciar uma resposta imune.

**Vacina de vetor** vacina produzida pela inserção de genes de um vírus patogênico em um vírus carreador relativamente inócuo.

**Vacina polivalente** vacina que imuniza contra mais de uma doença.

**Vetor** (como em vetor de clonagem) molécula de DNA autorreplicativa, utilizada para carrear genes clonados ou outros segmentos de DNA úteis à engenharia genética.

**Vetor bifuncional** vetor de clonagem que pode se replicar em dois ou mais hospedeiros não relacionados.

**Vetor de expressão** vetor de clonagem que apresenta as sequências reguladoras necessárias à transcrição e tradução dos genes clonados.

## QUESTÕES PARA REVISÃO

- O que são enzimas de restrição? Qual a provável função de uma enzima de restrição na célula que a produz? Por que a presença de uma enzima de restrição em uma célula não provoca a degradação do DNA desta célula? (Seção 11.1)
- Como você poderia detectar uma colônia contendo um gene clonado, caso já conhecesse a sequência do gene? (Seção 11.2)
- Descreva os princípios básicos da amplificação gênica utilizando reação em cadeia de polimerase (PCR). Como os procariontos termófilos e hipertermófilos simplificaram o uso da PCR? (Seção 11.3)
- A engenharia genética é dependente de vetores. Descreva as propriedades necessárias a um vetor plasmidial de clonagem eficiente. (Seção 11.4)
- Como você detectaria uma colônia contendo um gene clonado, caso não conhecesse a sequência do gene, mas dispusesse da proteína purificada codificada por esse gene? (Seção 11.4)
- Quais são as principais aplicações de DNA sintetizados artificialmente? (Seção 11.5)
- O que a mutagênese sítio-dirigida permite realizar, que não é permitido pela mutagênese normal? (Seção 11.5)
- O que é um gene repórter? Descreva dois genes repórter amplamente utilizados. (Seção 11.6)
- Como as fusões gênicas são utilizadas no estudo da regulação gênica? (Seção 11.6)
- Como a inativação por inserção da  $\beta$ -galactosidase permite a detecção da presença de DNA exógeno em um vetor plasmidial, como o pUC19? (Seção 11.7)
- Descreva dois hospedeiros procarionóticos e as características vantajosas e desvantajosas de cada um. (Seção 11.8)
- Descreva as semelhanças e diferenças entre os vetores de expressão e os vetores bifuncionais. (Seção 11.9)
- Como o bacteriófago T7 foi utilizado para a expressão de genes exógenos em *Escherichia coli* e que características desejáveis esse sistema regulador possui? (Seção 11.9)
- Quais as vantagens da utilização de um vetor de clonagem baseado no fago lambda, quando comparado a um vetor plasmidial? (Seção 11.10)
- Quais as características essenciais de um cromossomo artificial? Qual a diferença entre um BAC e um YAC? Que características do plasmídeo F o tornam menos útil para o uso *in vitro*? (Seção 11.10)
- Qual a importância da transcriptase reversa quando genes animais são clonados, visando sua expressão em bactérias? (Seção 11.11)
- Que classes de proteínas são produzidas por biotecnologia? Como os genes dessas proteínas são obtidos? (Seção 11.12)
- O que é o plasmídeo Ti e como este foi utilizado em engenharia genética? (Seção 11.13)
- O que é uma vacina de subunidades e por que tais vacinas são consideradas uma forma mais segura de conferir imunidade contra patógenos virais, quando comparadas às vacinas virais atenuadas? (Seção 11.14)
- Como a metagenômica tem sido utilizada para encontrar novos produtos úteis? (Seção 11.15)
- O que é engenharia de via? Por que produzir um antibiótico é mais difícil que uma única enzima por meio da engenharia genética? (Seção 11.16)
- Como a biologia sintética difere da engenharia de via? (Seção 11.17)

## QUESTÕES APLICADAS

- Suponha que você tem a tarefa de construir um vetor plasmidial de expressão adequado à clonagem molecular em um organismo de interesse industrial. Cite as características que tal plasmídeo deve apresentar. Liste as etapas que você utilizaria para construir tal plasmídeo.
- Suponha que você determinou recentemente a sequência de bases de DNA de um promotor especialmente forte em *Escherichia coli* e que você está interessado na incorporação dessa sequência em um vetor de expressão. Descreva as etapas que utilizaria para isso. Que precauções devem ser tomadas para se garantir que esse promotor atuará, de fato, conforme o esperado neste novo local?
- Muitos sistemas genéticos usam o gene *lacZ*, que codifica  $\beta$ -galactosidase, como um repórter. Quais as vantagens e os problemas que podem ocorrer quando (a) a luciferase ou (b) a proteína verde fluorescente é utilizada como repórteres, em vez da  $\beta$ -galactosidase?
- Você acabou de descobrir uma proteína de camundongos que pode trazer a cura efetiva do câncer, no entanto, ela é encontrada somente em pequenas quantidades. Descreva as etapas que você realizaria para produzir essa proteína em quantidades terapêuticas. Que hospedeiro você utilizaria para a clonagem do gene e por quê? Qual hospedeiro você utilizaria para expressar tal proteína e por quê?