

# Como as células leem o genoma: do DNA à proteína

CAPÍTULO

# 6

Desde a descoberta da estrutura do DNA no início de 1950, os progressos em biologia celular e molecular têm sido surpreendentes. Atualmente, são conhecidas as sequências genômicas completas de milhares de organismos diferentes, o que nos revelou detalhes fascinantes da sua bioquímica, bem como indícios importantes sobre a forma como esses organismos evoluíram. Também foram determinadas as sequências completas do genoma de milhares de seres humanos individuais, bem como de alguns de nossos parentes agora extintos, como os neandertais. O conhecimento da quantidade máxima de informação necessária para que se produza um organismo complexo, como um ser humano, nos permite identificar os limites das características bioquímicas e estruturais das células e deixa claro que a biologia não é infinitamente complexa.

Como discutido no Capítulo 1, o DNA genômico não controla a síntese de proteínas diretamente, mas utiliza o RNA como intermediário. Quando a célula requer uma proteína específica, a sequência de nucleotídeos da região apropriada de uma molécula de DNA extremamente longa em um cromossomo é inicialmente copiada sob a forma de RNA (através de um processo denominado *transcrição*). São essas cópias de RNA de segmentos de DNA que são utilizadas diretamente como moldes para promover a síntese da proteína (em um processo denominado *tradução*). O fluxo da informação genética nas células é, portanto, de DNA para RNA e deste para proteína (**Figura 6-1**). Todas as células, desde a bactéria até os seres humanos, expressam sua informação genética dessa maneira – um princípio tão fundamental que é denominado *dogma central* da biologia molecular. Apesar da universalidade do dogma central da biologia molecular, existem variações importantes em como a informação flui do DNA para a proteína em certos organismos. A principal delas é que os transcritos de RNA em células eucarióticas são submetidos a uma série de etapas de processamento no núcleo, incluindo o *splicing do RNA*, antes que se permita sua saída do núcleo e sua tradução em proteína. Como discutiremos neste capítulo, tais etapas de processamento podem modificar substancialmente o “significado” de uma molécula de RNA e são, portanto, cruciais para a compreensão de como as células eucarióticas leem o genoma.

Embora nosso enfoque neste capítulo seja a produção das proteínas codificadas pelo genoma, veremos que, para diversos genes, o RNA é o produto final. Assim como as proteínas, algumas dessas moléculas de RNA se enovelam em estruturas tridimensionais precisas que desempenham funções estruturais e catalíticas na célula. Outras moléculas de RNA, como discutiremos no próximo capítulo, atuam principalmente como reguladores da expressão gênica. Mas as funções de diversas moléculas de RNA não codificadas ainda não são conhecidas.

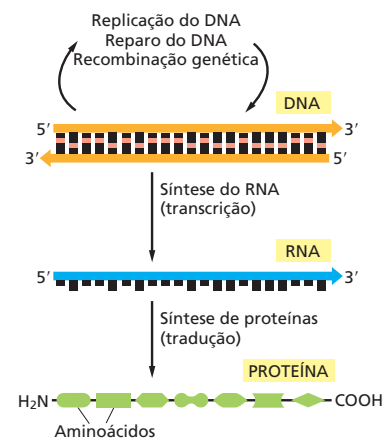
Apesar de termos imaginado originalmente que a informação presente nos genomas estaria organizada de forma ordenada, de forma semelhante a um dicionário ou uma lista telefônica, o que se viu é que os genomas da maioria dos organismos multicelulares são surpreendentemente desordenados, o que reflete as suas histórias evolutivas caóticas. Os genes nesses organismos consistem principalmente de uma longa sequência de éxons curtos e íntrons longos que se alternam, como discutido no Capítulo 4 (ver Figura 4-15D). Além disso, os pequenos segmentos da sequência de DNA que codificam proteínas estão intercalados com grandes blocos de DNA aparentemente sem função. Algumas regiões do genoma contêm vários genes, e outras carecem inteiramente de genes. Mesmo as proteínas que funcionam intimamente unidas na célula frequentemente são codificadas por genes localizados em diferentes cromossomos, e genes adjacentes codificam proteínas que apresentam pouca relação na célula. Portanto, decodificar genomas não é uma tarefa simples; mesmo com o auxílio de computadores potentes, é difícil para os pesquisadores definir categoricamente o início e o fim dos genes, e ainda mais difícil é decifrar quando e onde cada gene será expresso na vida do organismo. Ainda assim, as células no nosso organismo executam essas funções automaticamente milhares de vezes a cada segundo.

## NESTE CAPÍTULO

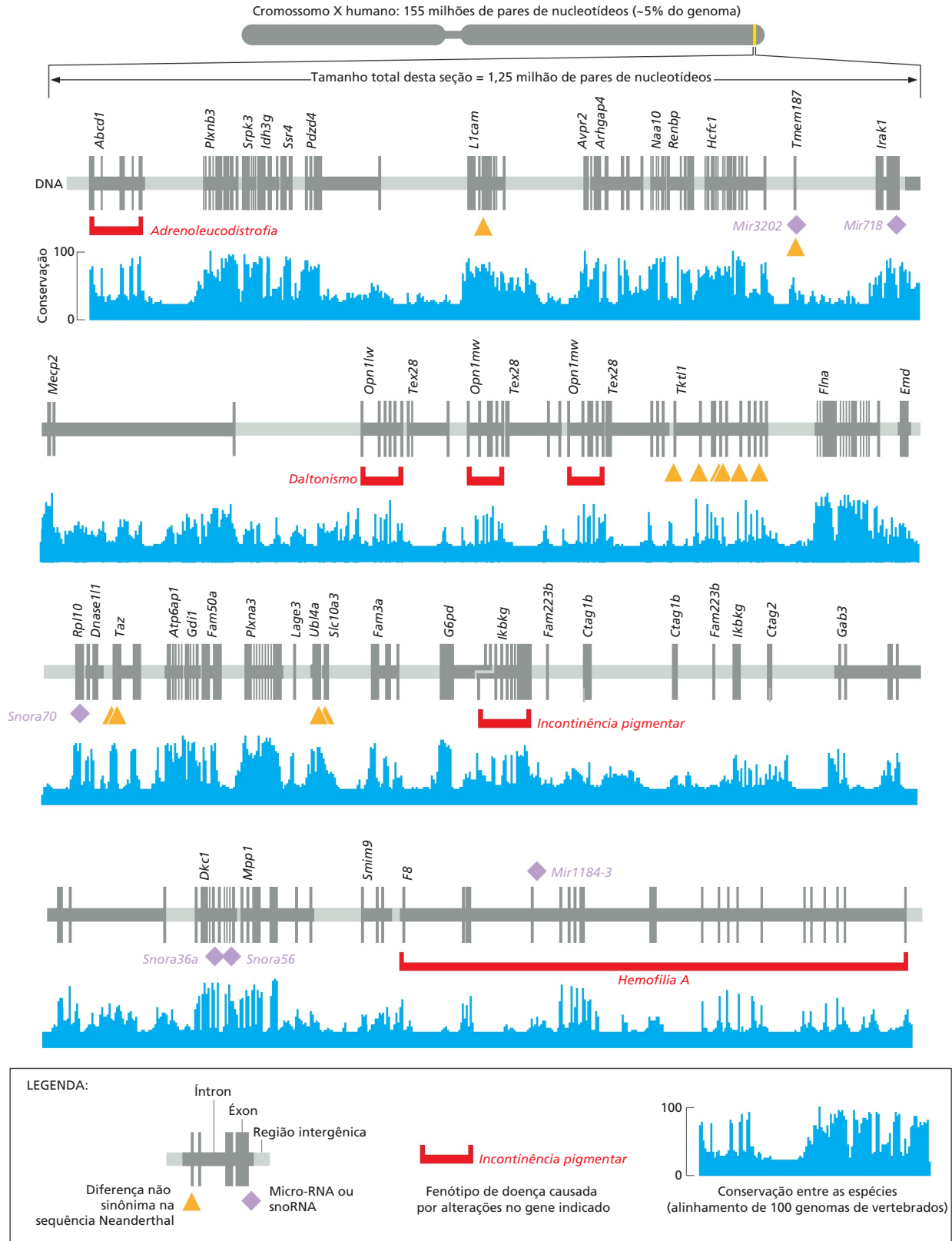
DO DNA AO RNA

DO RNA À PROTEÍNA

O MUNDO DE RNA E A ORIGEM DA VIDA



**Figura 6-1** A via do DNA à proteína. O fluxo da informação genética do DNA ao RNA (transcrição) e do RNA à proteína (tradução) ocorre em todas as células vivas.



**Figura 6-2 Representação esquemática de uma pequena porção do cromossomo X humano.** Como resumido na legenda, os genes codificadores de proteínas conhecidos (começando no *Abcd1* e terminando no *F8*) são mostrados em *cinza escuro*, com as regiões codificadoras (éxons) indicadas por barras que se estendem acima e abaixo da linha central. RNAs não codificadores, com funções conhecidas, são indicados por *losangos roxos*. *Triângulos amarelos* indicam as posições dentro das regiões codificadoras de proteínas, onde as sequências do genoma Neanderthal codificam um aminoácido diferente daquele encontrado no genoma humano. O trecho de *triângulos amarelos* no gene *Tktl1* parece ter sido positivamente selecionado desde a divergência entre *Homo sapiens* e neandertais, cerca de 200 mil anos atrás. Observe que a maioria das proteínas é idêntica entre nós e nosso parente extinto. O histograma *azul* indica o grau de conservação de porções do genoma humano em relação a outras espécies de vertebrados. É provável que genes adicionais, atualmente não identificados, também se encontrem nessa porção do genoma humano.

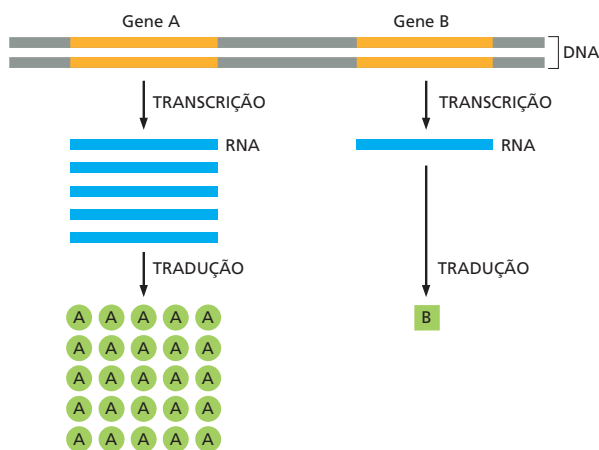
Os genes nos quais uma mutação provoca uma doença humana hereditária são indicados por *colchetes vermelhos*. O gene *Abcd1*, codifica uma proteína que importa ácidos graxos para o peroxissomo; mutações nesse gene causam desmielinização dos nervos que pode resultar em perturbações cognitivas e distúrbios do movimento. A *incontinência pigmentar* é uma doença que afeta a pele, os cabelos, as unhas, os dentes e os olhos. A *hemofilia A* é uma doença hemorrágica causada por mutações no gene do Fator VIII, que codifica uma proteína da cascata de coagulação do sangue. Visto que os homens têm uma única cópia do cromossomo X, a maioria das condições aqui apresentadas afeta somente os homens; as mulheres que herdam um desses genes defeituosos são frequentemente assintomáticas, pois uma proteína funcional é produzida a partir de seu outro cromossomo X. (Cortesia de Alex Williams, obtida da University of California, Genome Browser, <http://genome.ucsc.edu>.)

Os problemas que as células enfrentam na decodificação dos genomas podem ser avaliados ao se considerar uma porção bastante pequena do genoma humano (**Figura 6-2**). A região ilustrada representa menos de 1/2.000 do nosso genoma e inclui pelo menos 48 genes que codificam proteínas e seis genes para moléculas de RNA não codificador. Quando consideramos todo o genoma humano, só podemos maravilhar-nos com a capacidade das nossas células em lidar com tamanha presteza e precisão com essa enorme quantidade de informação.

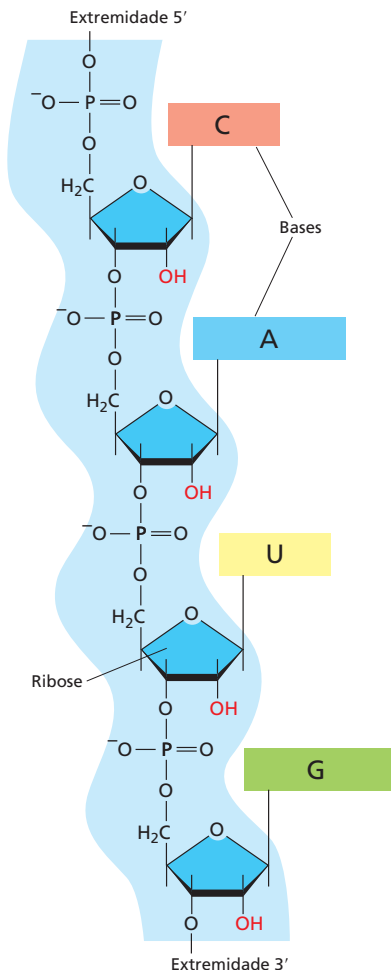
Neste capítulo, explicaremos como as células decodificam e usam a informação contida em seus genomas. Veremos que muito foi descoberto sobre como as instruções genéticas escritas em um alfabeto de apenas quatro “letras” – os quatro diferentes nucleotídeos do DNA – determinam a formação de uma bactéria, uma mosca-das-frutas ou um ser humano. No entanto, se ainda temos muito a descobrir sobre como a informação armazenada no genoma de um organismo é capaz de produzir mesmo a mais simples bactéria unicelular, a qual contém 500 genes, o que não dizer do desenvolvimento de um ser humano com aproximadamente 30 mil genes. Ainda desconhecemos uma enorme quantidade de informações; portanto muitos desafios fascinantes aguardam as próximas gerações de biólogos celulares.

## DO DNA AO RNA

A transcrição e a tradução são os meios pelos quais as células leem, ou expressam, as instruções genéticas de seus genes. Como muitas cópias idênticas de RNA podem ser produzidas a partir do mesmo gene, e como cada molécula de RNA pode promover a síntese de várias moléculas idênticas de proteína, as células podem, quando necessário, sintetizar uma grande quantidade de proteína a partir de um simples gene. No entanto, genes podem ser transcritos e traduzidos em taxas diferentes, permitindo que a célula sintetize enormes quantidades de certas proteínas e mínimas quantidades de outras (**Fi-**



**Figura 6-3 Os genes podem ser expressos em diferentes graus de eficiência.** Neste exemplo, o gene A é transcrito de maneira mais eficiente do que o gene B e cada molécula de RNA que ele produz também é traduzida mais frequentemente. Isso torna a quantidade da proteína A na célula muito maior do que a quantidade da proteína B.



**Figura 6-4** Um pequeno fragmento de RNA. A ligação química fosfodiéster entre nucleotídeos no RNA é a mesma que ocorre no DNA.

**Figura 6-3**). Além disso, como veremos no próximo capítulo, uma célula pode alterar (ou regular) a expressão de cada um de seus genes de acordo com suas necessidades – na maioria das vezes, pelo controle da produção das moléculas de RNA.

## As moléculas de RNA são fitas simples

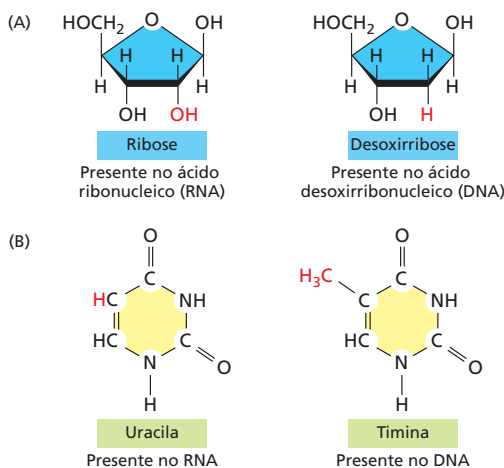
A primeira etapa executada pela célula para ler a informação necessária a partir de suas instruções genéticas é a cópia de um segmento específico da sequência de nucleotídeos do DNA – um gene – sob a forma de uma sequência de nucleotídeos de RNA (**Figura 6-4**). A informação na forma de RNA, embora copiada em uma forma química distinta, ainda é escrita essencialmente na mesma linguagem do DNA – a linguagem de uma sequência de nucleotídeos. Por isso, o nome dado para a produção de moléculas de RNA a partir do DNA é *transcrição*.

Assim como o DNA, o RNA é um polímero linear composto por quatro tipos diferentes de subunidades nucleotídicas unidas entre si por ligações fosfodiéster (ver Figura 6-4). O RNA difere quimicamente do DNA em dois aspectos: (1) os nucleotídeos do RNA são *ribonucleotídeos* – isto é, eles contêm o açúcar ribose (de onde vem o nome ácido ribonucleico) em vez de desoxirribose; (2) embora, assim como o DNA, o RNA contenha as bases adenina (A), guanina (G) e citosina (C), ele contém a base uracila (U), em vez da timina (T), que ocorre no DNA (**Figura 6-5**). Uma vez que U, assim como T, pode formar pares pelo estabelecimento de ligações de hidrogênio com A (**Figura 6-6**), as propriedades de complementaridade por pareamento de bases descritas para o DNA nos Capítulos 4 e 5 também se aplicam ao RNA (no RNA, G forma par com C, e A forma par com U). No entanto, é possível encontrar outros tipos de pareamento de bases no RNA: por exemplo, G ocasionalmente forma par com U.

Apesar de essas diferenças químicas serem pequenas, o DNA e o RNA diferem drasticamente em termos de sua estrutura geral. Enquanto o DNA sempre ocorre nas células sob a forma de uma hélice de fita dupla, o RNA se apresenta como fita simples. Assim, as cadeias de RNA podem se enovelar sob diversas formas, similarmente ao que ocorre com uma cadeia de polipeptídeos que se enovela em uma conformação determinada dando uma forma final à proteína (**Figura 6-7**). Como veremos posteriormente neste capítulo, a capacidade de se enovelar em formas tridimensionais complexas permite que algumas moléculas de RNA desempenhem funções estruturais e catalíticas.

## A transcrição produz uma molécula de RNA complementar a uma das fitas do DNA

Todo o RNA de uma célula é produzido a partir da **transcrição do DNA**, em um processo que apresenta certas similaridades em relação ao processo de replicação do DNA, discutido no Capítulo 5. A transcrição inicia com a abertura e a desespiralização de uma pequena porção da dupla-hélice de DNA, o que expõe as bases em cada fita do DNA. Uma



**Figura 6-5** Estrutura química do RNA. (A) O RNA contém o açúcar ribose, que difere da desoxirribose, o açúcar presente no DNA, pela presença de um agrupamento -OH adicional. (B) O RNA contém a base uracila, que difere da timina, a base equivalente no DNA, pela ausência de um grupo -CH<sub>3</sub>.

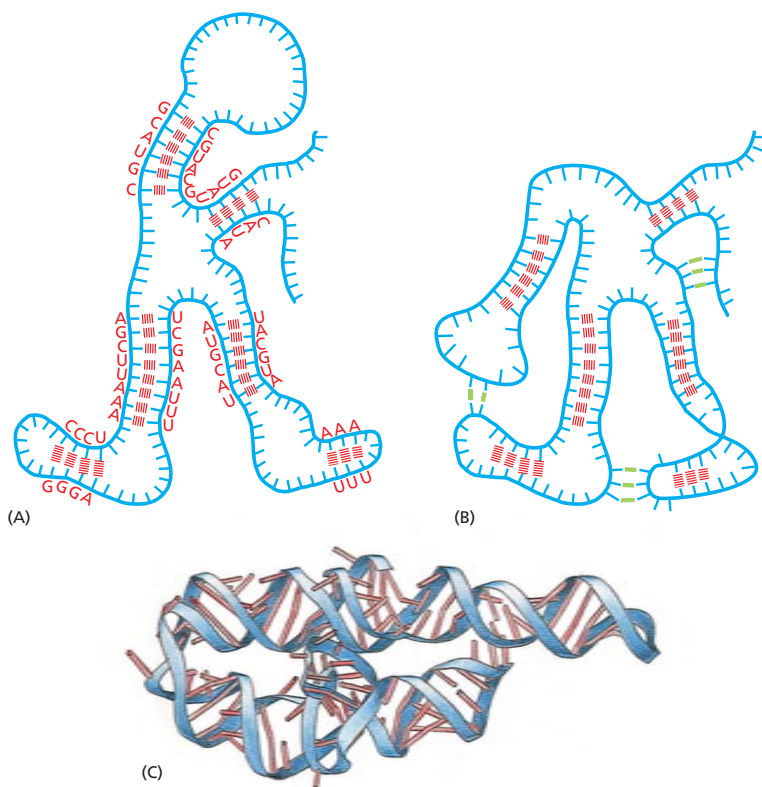
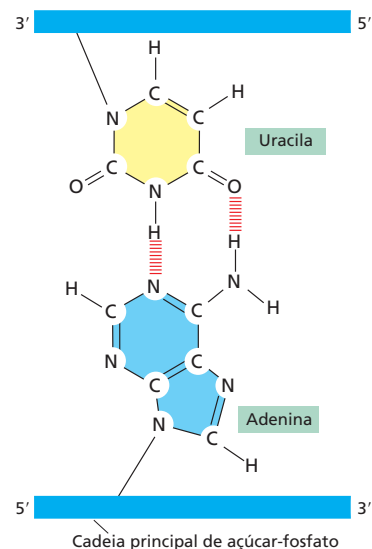
**Figura 6-6 A uracila forma pares de bases com a adenina.** A ausência de um grupo metila em U não tem efeito no pareamento de bases; assim, os pares de bases U-A assemelham-se muito aos pares de bases T-A (ver Figura 4-4).

das duas fitas da dupla-hélice de DNA, então, serve como um molde para a síntese de uma molécula de RNA. Assim como na replicação de DNA, a sequência de nucleotídeos da cadeia de RNA é determinada pelo pareamento de bases complementares entre os nucleotídeos a serem incorporados e o DNA-molde. Quando um pareamento adequado é estabelecido (A com T, U com A, G com C e C com G), o ribonucleotídeo a ser incorporado é covalentemente ligado à cadeia de RNA em formação, através de uma reação catalisada enzimaticamente. A cadeia de RNA produzida por transcrição – o *transcrito* – é, a seguir, aumentada 1 nucleotídeo por vez e possui uma sequência de nucleotídeos exatamente complementar à fita de DNA utilizada como molde (**Figura 6-8**).

No entanto, a transcrição difere da replicação de DNA em vários aspectos importantes. Diferentemente de uma fita de DNA recém-formada, a fita de RNA não permanece ligada à fita de DNA-molde por ligações de hidrogênio. Em um ponto situado imediatamente após a região onde os ribonucleotídeos foram adicionados, a cadeia de RNA é deslocada, e a hélice de DNA se reassocia. Assim, as moléculas de RNA produzidas pela transcrição são liberadas do DNA-molde sob a forma de fita simples. Além disso, como essas moléculas de RNA são copiadas apenas de uma região definida do DNA, as moléculas de RNA são muito menores que as moléculas de DNA. Uma molécula de DNA em um cromossomo humano pode chegar a 250 milhões de pares de nucleotídeos de comprimento, enquanto a maioria das moléculas de RNA não tem comprimento superior a alguns milhares de nucleotídeos, e muitas moléculas de RNA são mesmo consideravelmente menores.

### RNA-polimerases realizam a transcrição

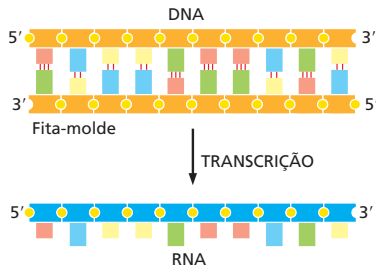
As enzimas que realizam a transcrição são denominadas **RNA-polimerases**. Assim como a DNA-polimerase catalisa a replicação do DNA (discutida no Capítulo 5), as RNA-polime-



**Figura 6-7 O RNA pode se enovelar formando estruturas específicas.**

O RNA é principalmente de fita simples, mas com frequência contém pequenos segmentos de nucleotídeos que podem formar um pareamento convencional com sequências complementares encontradas em outras regiões da mesma molécula. Tais interações, juntamente a pareamentos “não convencionais” adicionais, permitem a uma molécula de RNA se enovelar sobre si mesma, formando uma estrutura tridimensional determinada por sua sequência de nucleotídeos (**Animação 6.1**). (A) Diagrama de uma estrutura enovelada de RNA mostrando apenas interações convencionais entre pares de bases. (B) Estrutura apresentando tanto interações convencionais (vermelho) quanto não convencionais (verde) entre pares de bases. (C) A estrutura de um RNA real, que catalisa seu próprio *splicing* (ver p. 324). Cada pareamento convencional é indicado por um “traço” na dupla-hélice. As bases em outras configurações estão indicadas por traços descontinuos.





**Figura 6-8** A transcrição do DNA produz uma molécula de RNA de fita simples que é complementar a uma das fitas da dupla-hélice de DNA. Observe que a sequência de bases na molécula de RNA produzida é a mesma que a sequência de bases da cadeia de DNA não molde, exceto que uma base U substitui cada base T do DNA.

**Figura 6-9** O DNA é transcrito pela enzima RNA-polimerase. A RNA-polimerase (azul-clara) move-se passo a passo ao longo do DNA, desenrolando a hélice do DNA no seu sítio ativo indicado pelo  $Mg^{2+}$  (vermelho), que é necessário para a catálise. Conforme avança, a polimerase adiciona nucleotídeos, um a um, à cadeia de RNA no sítio de polimerização, usando uma fita de DNA exposta como molde. Conseqüentemente, o RNA transcrito é uma cópia complementar de fita simples de uma das duas fitas do DNA. Uma região curta de hélice DNA/RNA (cerca de nove pares de nucleotídeos em comprimento) é formada apenas transitoriamente, e uma "janela" dessa hélice DNA/RNA move-se ao longo do DNA com a polimerase, e a dupla hélice do DNA é reestruturada após sua passagem. Os nucleotídeos a serem incorporados estão na forma de ribonucleosídeos trifosfato (ATP, UTP, CTP e GTP), e a energia estocada em suas ligações fosfato-fosfato fornece a força necessária para a reação de polimerização (ver Figura 5-4). A figura, baseada em uma estrutura cristalográfica de raios X, mostra uma vista em corte da polimerase: a porção virada para o leitor foi cortada para mostrar o seu interior (Animação 6.3). (Adaptada de P. Cramer et al., *Science* 288:640–649, 2000; Código PDB: 1HQM.)

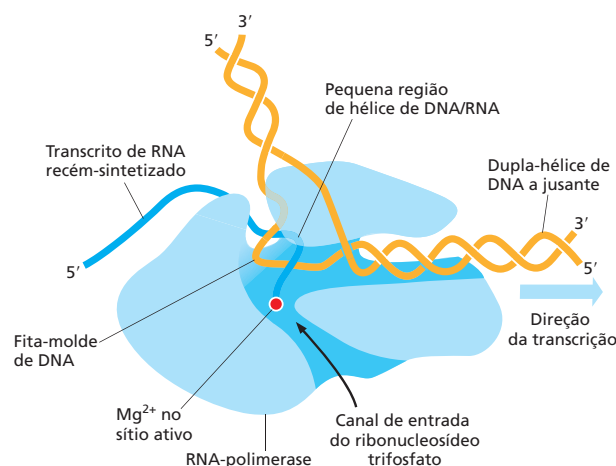
rases catalisam a formação de ligações fosfodiéster que conectam os nucleotídeos entre si formando uma cadeia linear. A RNA-polimerase se desloca passo a passo sobre o DNA, desespiralizando a dupla-hélice à frente do sítio ativo de polimerização e expondo, dessa forma, uma nova região da fita-molde para o pareamento de bases complementares. Dessa maneira, a cadeia de RNA em formação é estendida, nucleotídeo a nucleotídeo, na direção de 5' para 3' (Figura 6-9). Os substratos são ribonucleosídeos trifosfato (ATP, CTP, UTP e GTP); assim como na replicação do DNA, a hidrólise de ligações altamente energéticas fornece a energia necessária para promover a reação (ver Figura 5-4 e Animação 6.2).

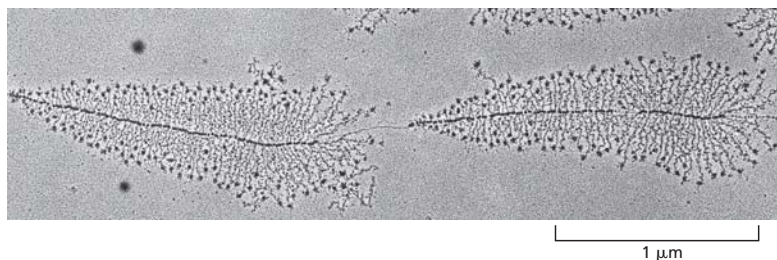
A liberação quase imediata da fita de RNA do DNA, à medida que a primeira está sendo sintetizada, significa que muitas cópias de RNA podem ser produzidas a partir do mesmo gene em um período de tempo relativamente pequeno; a síntese de moléculas de RNA adicionais pode ser iniciada antes que as moléculas anteriores de RNA tenham sido finalizadas (Figura 6-10). Quando várias moléculas de RNA-polimerase usam a mesma região como molde, deixando um pequeno intervalo entre si, cada uma sintetizando aproximadamente 50 nucleotídeos/segundo, mais de mil transcritos podem ser sintetizados em 1 hora, a partir de um único gene.

Apesar de a RNA-polimerase catalisar essencialmente a mesma reação química que a DNA-polimerase, existem algumas diferenças importantes entre essas duas enzimas. Primeiramente, e mais óbvio, a RNA-polimerase catalisa a ligação de ribonucleotídeos, e não de desoxirribonucleotídeos. Segundo, ao contrário das DNA-polimerases envolvidas na replicação de DNA, as RNA-polimerases podem começar a síntese de uma cadeia de RNA sem um iniciador. Acredita-se que essa diferença seja possível porque a transcrição não necessita ser tão exata quanto a replicação do DNA (ver Tabela 5-1, p. 244). As RNA-polimerases cometem aproximadamente 1 erro a cada  $10^4$  nucleotídeos copiados em RNA (em comparação com uma taxa de erro de cópia direta da DNA-polimerase de cerca de 1 em cada  $10^7$  nucleotídeos); e as conseqüências de um erro na transcrição do RNA são muito menos significativas, pois o RNA não armazena de modo permanente a informação genética nas células. Por fim, diferentemente das DNA-polimerases, que fazem seus produtos em segmentos posteriormente unidos, as RNA-polimerases são absolutamente processuais; isto é, a mesma RNA-polimerase que inicia uma molécula de RNA deve terminar sua síntese sem dissociação do molde de DNA.

Embora não sejam tão exatas quanto as DNA-polimerases que replicam o DNA, as RNA-polimerases possuem um modesto mecanismo de correção. Se um ribonucleotídeo incorreto for adicionado à cadeia de RNA em formação, a polimerase pode retroceder e o sítio ativo da enzima pode realizar uma reação de excisão que é semelhante ao procedimento reverso da reação de polimerização, exceto que uma molécula de água substitui o pirofosfato e um nucleosídeo monofosfato é liberado.

Considerando-se que tanto as DNA-polimerases quanto as RNA-polimerases realizam polimerização dependente de molde, seria de esperar que esses dois tipos de enzimas fossem estruturalmente relacionados. No entanto, estudos de cristalografia de raios X revelaram que, apesar de ambas conterem um íon  $Mg^{2+}$  essencial no seu sítio catalítico, essas duas enzimas são bastante diferentes. As enzimas de polimerização de nucleotídeos





**Figura 6-10 Transcrição de dois genes visualizada sob microscópio eletrônico.**

A fotomicrografia mostra diversas moléculas de RNA-polimerase transcrevendo simultaneamente dois genes adjacentes. As moléculas de RNA-polimerase são visíveis como uma série de pontos ao longo do DNA, com os transcritos recentemente sintetizados (filamentos finos) ligados a elas. As moléculas de RNA (RNAs ribossômicos) mostradas neste exemplo não são traduzidas em proteína, sendo utilizadas diretamente como componentes dos ribossomos, as máquinas em que a tradução ocorre. Acredita-se que as partícula na extremidade 5' (a extremidade livre) de cada rRNA transcrito representem o início da associação de ribossomos. Considerando-se o comprimento relativo dos transcritos recentemente sintetizados, pode-se deduzir que as moléculas de RNA-polimerase estão transcrevendo da esquerda para a direita. (Cortesia de Ulrich Scheer.)

dependentes de molde parecem ter surgido pelo menos duas vezes durante a evolução inicial das células. Uma linhagem deu origem às DNA-polimerases modernas e às transcriptases reversas discutidas no Capítulo 5, bem como a umas poucas RNA-polimerases virais. A outra linhagem resultou em todas as RNA-polimerases celulares que são discutidas neste capítulo.

### As células produzem diferentes categorias de moléculas de RNA

A maioria dos genes presentes no DNA das células especifica a sequência de aminoácidos de proteínas; as moléculas de RNA que são copiadas a partir desses genes (e que conseqüentemente promovem a síntese de proteínas) são chamadas de moléculas de **RNA mensageiro (mRNA)**. O produto final de outros genes, entretanto, é a própria molécula de RNA. Esses RNAs são conhecidos como **RNAs não codificadores**, pois eles não codificam proteínas. Na levedura *Saccharomyces cerevisiae*, um eucarioto unicelular bastante estudado, mais de 1.200 genes (mais de 15% do total) têm o RNA como seus produtos finais. Os seres humanos provavelmente produzem em torno de 10 mil moléculas de RNA não codificador. Tais moléculas de RNA, assim como as proteínas, servem como componentes estruturais, reguladores e enzimáticos para uma ampla gama de processos na célula. No Capítulo 5 encontramos um deles atuando como o molde carregado pela enzima telomerase. Apesar das funções de várias dessas moléculas de RNA não codificador ainda estarem cobertas por um certo mistério, veremos neste capítulo que algumas moléculas de *pequenos RNAs nucleares (snRNA, small nuclear RNA)* promovem o *splicing* (excisão de íntrons) do pré-mRNA para formar o mRNA, que moléculas de *RNA ribossômico (rRNA)* formam a porção central dos ribossomos e que moléculas de *RNA transportador (tRNA)* formam os adaptadores que selecionam aminoácidos e os colocam no local adequado nos ribossomos para serem incorporados em proteínas. No Capítulo 7, veremos que moléculas de *micro-RNA (miRNA)* e moléculas de *pequenos RNAs de interferência (siRNA)* atuam como importantes reguladores na expressão gênica em eucariotos, e que os *RNAs que interagem com piwi (piRNAs)* protegem linhagens germinativas animais da ação dos elementos de transposição; nós também discutiremos os *RNAs não codificadores longos (lncRNAs)*, um subgrupo diverso de moléculas de RNA cujas funções ainda estão sendo descobertas (**Tabela 6-1**).

**TABELA 6-1 Principais tipos de RNA produzidos nas células**

Tipo de RNA	Função
mRNAs	RNAs mensageiros; codificam proteínas.
rRNAs	RNAs ribossômicos; formam a estrutura básica do ribossomo e catalisam a síntese proteica.
tRNAs	RNAs transportadores; elementos essenciais para a síntese proteica, atuando como adaptadores entre o mRNA e os aminoácidos.
snRNAs	Pequenos RNAs nucleares; atuam em uma série de processos nucleares, incluindo o <i>splicing</i> do pré-mRNA.
snoRNAs	Pequenos RNAs nucleares; ajudam a processar e modificar quimicamente os rRNAs.
miRNAs	Micro-RNAs; regulam a expressão gênica pelo bloqueio da tradução de mRNAs específicos e provocam a sua degradação.
siRNAs	Pequenos RNAs de interferência; desligam a expressão de genes pela degradação direta de mRNAs selecionados e pelo estabelecimento de estruturas de cromatina compacta.
piRNAs	RNAs que interagem com piwi; ligam-se a proteínas piwi e protegem a linhagem germinativa da ação de elementos transponíveis.
lncRNAs	RNAs não codificadores longos; muitos têm função de suporte estrutural; eles regulam diversos processos celulares, inclusive a inativação do cromossomo X.

Cada segmento de DNA transcrito é denominado *unidade de transcrição*. Nos eucariotos, uma unidade de transcrição normalmente carrega a informação de apenas um gene e, portanto, codifica uma única molécula de RNA ou uma única proteína (ou grupo de proteínas relacionadas, se o transcrito de RNA inicial for processado de diferentes maneiras para produzir diferentes mRNAs). Em bactérias, um conjunto de genes adjacentes é frequentemente transcrito como uma unidade e a molécula de mRNA resultante carrega, dessa forma, a informação para várias proteínas distintas.

Em geral, o RNA representa uma pequena porcentagem do peso seco de uma célula, enquanto as proteínas constituem cerca de 50% desse valor. A maioria do RNA nas células é rRNA; o mRNA representa somente 3 a 5% do RNA total em uma célula típica de mamíferos. A população de mRNAs é composta por dezenas de milhares de diferentes tipos, existindo em média apenas 10 a 15 moléculas de cada tipo de mRNA em cada célula.

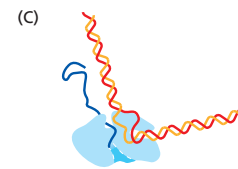
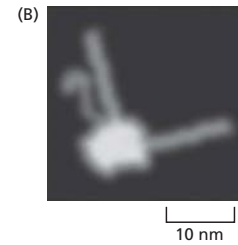
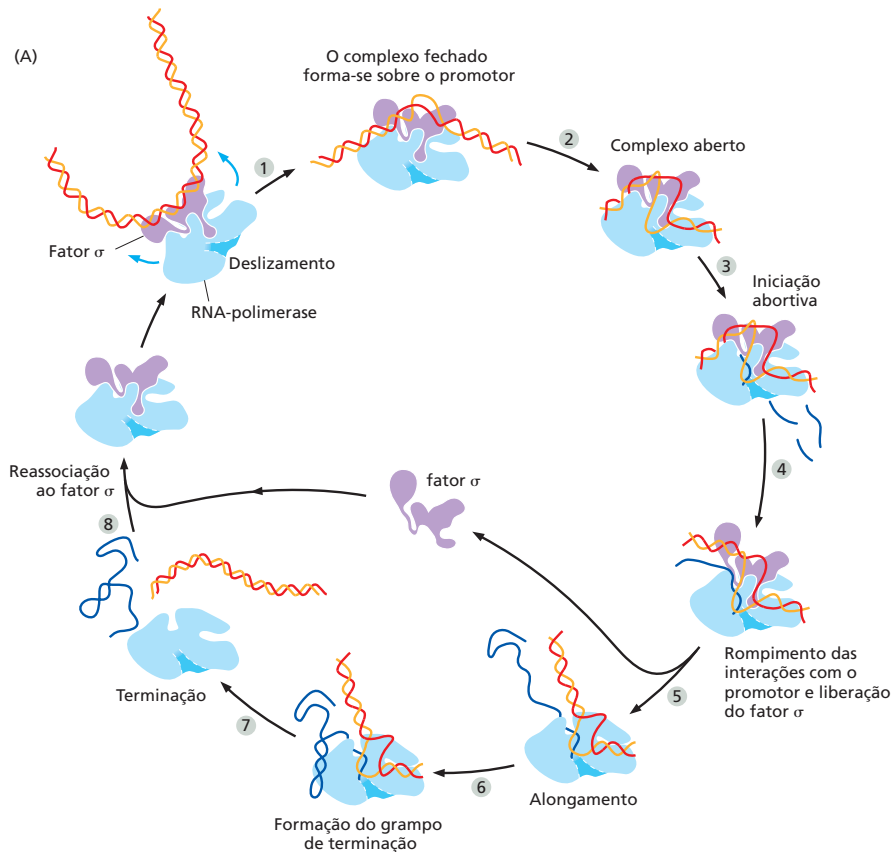
### Sinais codificados no DNA indicam à RNA-polimerase onde iniciar e onde terminar a transcrição

Para transcrever um gene com precisão, a RNA-polimerase deve reconhecer seu início e término no genoma. A maneira pela qual as RNA-polimerases desempenham essa tarefa difere entre bactérias e eucariotos. Como o processo em bactérias é mais simples, ele será discutido primeiro.

A iniciação, da transcrição é uma etapa extremamente importante na expressão de um gene, pois esse é o ponto principal onde a célula regula quais proteínas serão produzidas e a frequência dessa produção. A enzima central da RNA-polimerase bacteriana é um complexo de múltiplas subunidades que sintetiza RNA a partir de um molde de DNA. Uma subunidade adicional, denominada *fator sigma* ( $\sigma$ ), associa-se a essa enzima e auxilia a leitura dos sinais no DNA que indicam onde iniciar a transcrição (**Figura 6-11**). Em conjunto, a enzima-base e o fator  $\sigma$  são denominados *holoenzima RNA-polimerase*; esse complexo se liga fracamente ao DNA bacteriano quando colide com ele, e uma holoenzima desliza rapidamente ao longo da molécula de DNA até dissociar-se. No entanto, quando a holoenzima polimerase desliza sobre uma sequência especial de nucleotídeos que indica o ponto de início para a síntese de RNA, chamada **promotor**, a polimerase liga-se fortemente, pois o seu factor  $\sigma$  faz contatos específicos com a região das bases expostas do lado externo da dupla-hélice do DNA (etapa 1 na Figura 6-11A).

A holoenzima RNA-polimerase fortemente ligada a um promotor abre a dupla-hélice para expor um pequeno trecho de nucleotídeos em cada fita (etapa 2 na Figura 6-11A). A região de DNA não pareada (cerca de 10 nucleotídeos) é chamada de *bolha de transcrição* e é estabilizada pela ligação do factor  $\sigma$  às bases não pareadas de uma das fitas expostas. A outra fita de DNA exposta atua como um molde para o pareamento de bases complementares com os ribonucleotídeos, dois dos quais são unidos pela polimerase para dar início a uma cadeia de RNA (etapa 3 na Figura 6-11A). Os primeiros 10 ou mais nucleotídeos de RNA são sintetizados usando um mecanismo de “arraste”, durante o qual a RNA-polimerase permanece ligada ao promotor e puxa o DNA a montante para o seu sítio ativo, expandindo assim a bolha de transcrição. Esse processo cria um estresse considerável, e frequentemente os RNAs curtos são liberados, aliviando a tensão e forçando a polimerase, que permanece no mesmo lugar, para reiniciar a síntese. Por fim, esse processo de *iniciação abortiva* é superado e o estresse gerado pelo arraste ajuda a enzima central a se dissociar do DNA promotor (etapa 4 na Figura 6-11A) e descartar o fator  $\sigma$  (etapa 5 na Figura 6-11A). Nesse momento, a polimerase começa a mover-se sobre o DNA, sintetizando o RNA, de uma forma gradativa: a polimerase se desloca para frente um par de bases para cada nucleotídeo adicionado. Durante esse processo, a bolha de transcrição expande-se continuamente na parte da frente da polimerase e contrai-se na sua retaguarda. O alongamento da cadeia continua (a uma velocidade de cerca de 50 nucleotídeos/segundo no caso de RNA-polimerases bacterianas) até que a enzima encontre um segundo sinal, o **terminador** (etapa 6 na figura 6-11A), onde a polimerase para e libera tanto a molécula de RNA recém-sintetizada quanto o molde de DNA (etapa 7 na Figura 6-11A). Em seguida, a enzima polimerase livre se reassocia a um factor  $\sigma$  livre para formar uma holoenzima que pode novamente dar início ao processo de transcrição (etapa 8 na Figura 6-11A).





**Figura 6-11** Ciclo de transcrição da

**RNA-polimerase bacteriana.** (A) Na etapa 1, a holoenzima RNA-polimerase (a enzima central da polimerase mais o fator  $\sigma$ ) forma-se e, então, localiza uma sequência de DNA promotor (ver Figura 6-12). A polimerase abre (desespiraliza) o DNA na posição em que a transcrição está para começar (etapa 2) e começa a transcrever (etapa 3). Essa síntese de RNA inicial (iniciação abortiva) é relativamente ineficiente, pois transcritos curtos e improdutivos são frequentemente liberados. No entanto, uma vez que a RNA-polimerase consegue sintetizar cerca de 10 nucleotídeos de RNA, suas interações com o promotor do DNA são rompidas (etapa 4) e o fator  $\sigma$  é finalmente liberado – ao mesmo tempo em que a polimerase envolve o DNA alterando o modo da síntese do RNA para alongamento, movendo-se sobre o DNA (etapa 5). Durante o modo de alongamento, a transcrição é altamente eficiente, e a polimerase só irá se dissociar do DNA-molde e liberará o RNA recentemente transcrito quando encontrar um sinal de terminação (etapas 6 e 7). Os sinais de terminação são normalmente codificados no DNA, e muitos funcionam por meio da formação de uma estrutura de RNA em grampo que desestabiliza a ligação da polimerase ao RNA.

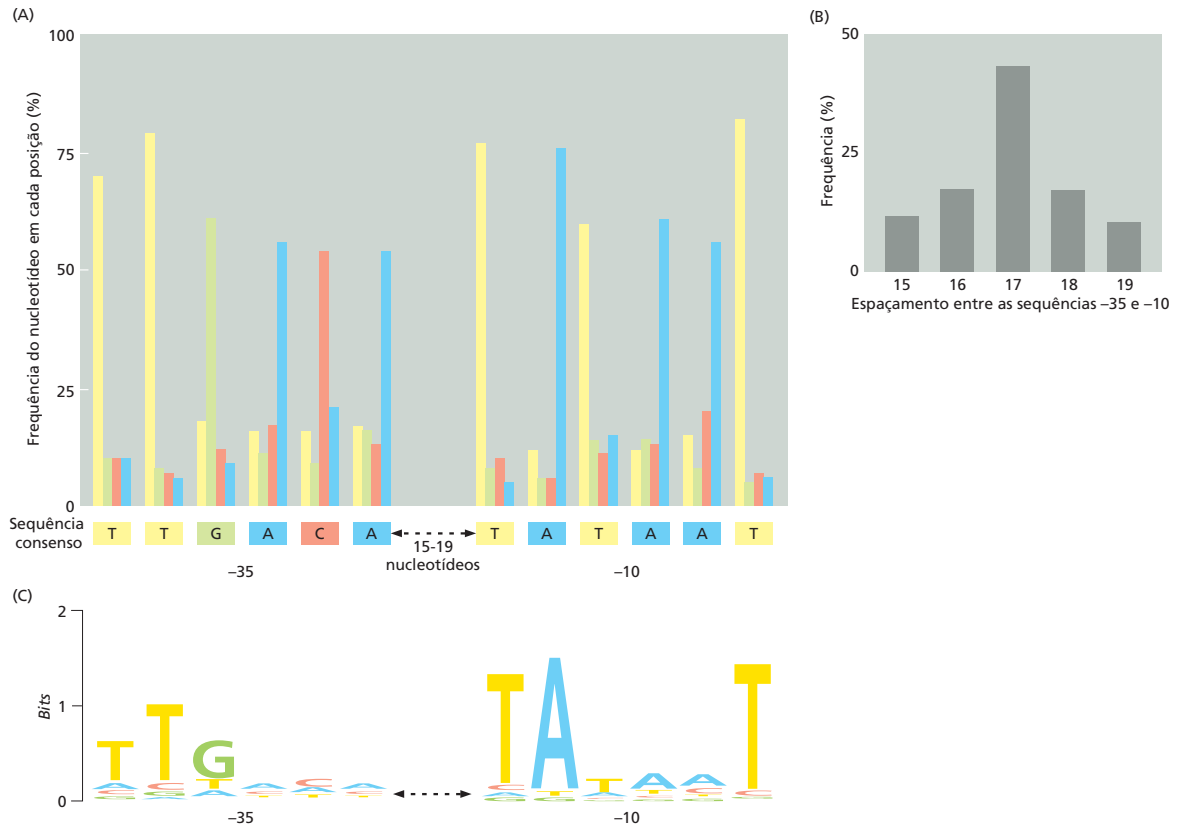
Nas bactérias, todas as moléculas de RNA são sintetizadas por um único tipo de RNA-polimerase, e o ciclo apresentado na figura se aplica tanto à produção de mRNAs quanto à produção de RNAs estruturais e catalíticos. (B) Imagem bidimensional de uma RNA-polimerase bacteriana em alongamento, determinada por microscopia de força atômica (ver Figura 9-33). (C) Interpretação da imagem mostrada em (B). (Adaptada de K.M. Herbert et al., *Annu. Rev. Biochem.* 77:149–176, 2008.)

O processo de iniciação da transcrição é complicado e requer que tanto a holoenzima RNA-polimerase quanto o DNA sofram uma série de alterações conformacionais. Podemos imaginar essas alterações como a abertura e o posicionamento do DNA no sítio ativo seguidos pela ligação com maior afinidade da enzima ao redor do DNA e do RNA para assegurar que estes não se dissociem antes do final da transcrição de um gene. Se uma RNA-polimerase se dissociar prematuramente, ela deverá recomençar o processo sobre o promotor.

Como os sinais de terminação no DNA finalizam o alongamento da polimerase? No caso da maioria dos genes bacterianos, o sinal de terminação consiste em uma sequência de pares de nucleotídeos A-T, precedida por uma sequência de DNA duplamente simétrica, a qual, quando transcrita em RNA, enovela-se em uma estrutura em “grampo” pelo pareamento de bases tipo Watson-Crick (ver Figura 6-11A). Conforme a polimerase transcreve esse terminador, a formação do grampo ajuda a desligar o transcrito de RNA do sítio ativo (etapa 7 na Figura 6-11A). O processo de terminação fornece um exemplo de um tema comum deste capítulo: a capacidade de enovelamento do RNA em estruturas específicas afeta diferentes etapas da decodificação do genoma.

## Os sinais de início e término da transcrição na sequência nucleotídica são heterogêneos

Como acabamos de ver, os processos de iniciação e de terminação da transcrição envolvem uma complicada série de transições estruturais nas moléculas de proteínas, de DNA e de RNA. Os sinais codificados no DNA que especificam essas transições são frequentemente difíceis de serem reconhecidos pelos pesquisadores. De fato, uma comparação de vários promotores bacterianos diferentes revela um extraordinário nível de diversidade. Apesar disso, todos contêm sequências relacionadas, refletindo os aspectos do DNA que são reconhecidos diretamente pelo fator  $\sigma$ . Essas características comuns são frequentemente resumidas sob a forma de uma *sequência consenso* (Figura 6-12). Uma *sequência*



**Figura 6-12** Sequência nucleotídica consenso e logotipo da sequência das principais classes de promotores de *E. coli*.

(A) Com base em uma comparação de 300 promotores, são dadas as frequências de cada um dos quatro nucleotídeos em cada posição no promotor.

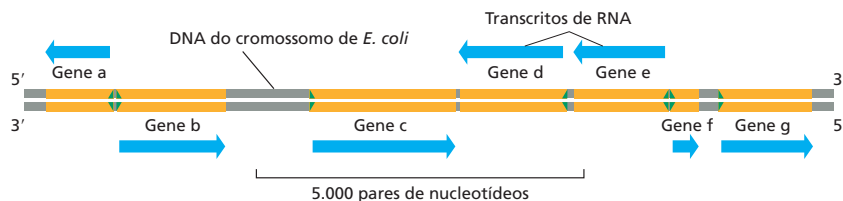
A sequência consenso, ilustrada abaixo do gráfico, reflete os nucleotídeos mais comuns encontrados em cada posição no conjunto de promotores. Esses promotores são caracterizados por duas sequências hexaméricas de DNA, a sequência -35 e a sequência -10, assim denominadas por sua localização aproximada com relação ao ponto de início da transcrição (designado como +1). A sequência de nucleotídeos entre os hexâmeros -35 e -10 não mostra similaridades significativas entre os promotores. Por conveniência, é mostrada a sequência nucleotídica de um DNA de fita simples; na realidade, os promotores são DNA de fita dupla. Os nucleotídeos mostrados na figura são reconhecidos pelo factor  $\sigma$ , uma subunidade da holoenzima RNA-polimerase. (B) Distribuição de espaçamento entre os hexâmeros -35 e -10, encontrados nos promotores de *E. coli*. (C) Um logotipo da sequência exibindo as mesmas informações do painel (A). Aqui, a altura de cada letra é proporcional à frequência na qual a base ocorre nessa posição em uma ampla gama de sequências promotoras. A altura total de todas as letras em cada posição é proporcional ao conteúdo da informação (expresso em bits) nessa posição. Por exemplo, o conteúdo total de informação de uma posição que pode tolerar diversas bases diferentes é pequeno (ver as três últimas bases das sequências -35), mas estatisticamente maior do que ao acaso.

nucleotídica consenso é derivada pela comparação de muitas sequências que apresentam a mesma função básica e pelo alinhamento dos nucleotídeos mais comuns encontrados em cada posição. Isso serve, portanto, como um resumo ou uma “média” de um grande número de sequências nucleotídicas individuais. Uma maneira mais exata de ilustrar a gama de sequências de DNA reconhecidas por uma proteína é pelo uso de um logotipo da sequência, que revela as frequências relativas de cada nucleotídeo em cada posição (Figura 6-12C).

As diferenças entre as sequências de DNA dos promotores bacterianos individuais determinam a sua força (ou o número de eventos de iniciação por unidade de tempo para cada promotor). Os processos evolutivos sintonizaram cada promotor para iniciar com a frequência necessária e criaram, assim, um amplo espectro de força para os promotores. Os promotores de genes que codificam as proteínas abundantes são muito mais fortes do que aqueles associados a genes que codificam proteínas raras, e as sequências nucleotídicas dos seus promotores são as responsáveis por essas diferenças.

Assim como os promotores bacterianos, os terminadores de transcrição também apresentam um amplo espectro de sequências, e o potencial de formar uma estrutura de RNA em grampo é a característica comum mais importante desses promotores. Uma vez que um número quase ilimitado de sequências nucleotídicas tem esse potencial, as sequências de terminadores são muito mais heterogêneas do que as dos promotores.

Nós apresentamos os promotores e terminadores bacterianos e alguns de seus detalhes para ilustrar um ponto importante no que diz respeito à análise das sequências genômicas. Apesar de conhecermos muito sobre promotores e terminadores bacterianos e podermos estabelecer sequências consenso que resumem suas características mais óbvias, sua identificação exata e definitiva no genoma pela análise da sequência nucleotídica é bastante dificultada devido à sua diversidade em termos de sequência de nucleotídeos. É ainda mais difícil posicionar sequências análogas em genomas de eucariotos, parcialmente devido ao excesso de DNA presente nesses genomas. Com



**Figura 6-13** Direção da transcrição em um segmento curto de um cromossomo bacteriano. Alguns genes são transcritos utilizando uma das fitas de DNA como molde, enquanto outros são transcritos utilizando a outra fita de DNA. A direção de transcrição é determinada pelo promotor no início de cada gene (setas verdes). Aproximadamente 0,2% (9 mil pares de bases) do cromossomo de *E. coli* está ilustrado aqui. Os genes transcritos da esquerda para a direita usam a fita inferior de DNA como molde; aqueles transcritos da direita para a esquerda usam a fita superior como molde.

freqüência, são necessárias informações adicionais, muitas vezes obtidas por experimentação direta, para posicionar e identificar com exatidão curtos sinais de DNA nos genomas.

Como ilustrado na Figura 6-11, as sequências promotoras são assimétricas, assegurando que a RNA-polimerase possa se ligar em apenas uma orientação. Visto que a polimerase pode sintetizar RNA apenas na direção de 5' para 3', a orientação do promotor determina a fita que será utilizada como molde. As sequências genômicas revelam que a fita de DNA utilizada como molde para a síntese de RNA varia de gene para gene, dependendo da orientação do promotor (Figura 6-13).

Tendo considerado a transcrição em bactérias, veremos agora a situação nos eucariotos, nos quais a síntese de moléculas de RNA é uma tarefa muito mais elaborada.

### A iniciação da transcrição nos eucariotos requer várias proteínas

Em contraste com as bactérias, que contêm um único tipo de RNA-polimerase, os núcleos eucarióticos têm três: *RNA-polimerase I*, *RNA-polimerase II* e *RNA-polimerase III*. As três polimerases são estruturalmente similares entre si e compartilham algumas subunidades, mas transcrevem diferentes categorias de genes (Tabela 6-2). As RNA-polimerases I e III transcrevem os genes que codificam RNA de transferência, RNA ribossômico e vários pequenos RNAs. A RNA-polimerase II transcreve a grande maioria dos genes, inclusive todos aqueles que codificam proteínas; assim, nossa discussão subsequente será focada nessa enzima.

A RNA-polimerase II eucariótica tem muitas semelhanças estruturais com a RNA-polimerase bacteriana (Figura 6-14). No entanto, há algumas diferenças importantes na maneira em que as enzimas bacterianas e eucarióticas funcionam, e duas dessas diferenças nos interessam diretamente.

1. Enquanto a RNA-polimerase bacteriana requer apenas um único fator de iniciação da transcrição ( $\sigma$ ) para começar a transcrição, as RNA-polimerases eucarióticas exigem muitos desses fatores, chamados coletivamente de *fatores gerais de transcrição*.
2. A iniciação da transcrição eucariótica deve ocorrer no DNA que está empacotado nos nucleossomos e em estruturas de cromatina de ordem superior (descrito no Capítulo 4), estruturas essas que estão ausentes dos cromossomos bacterianos.

**TABELA 6-2** As três RNA-polimerases de células eucarióticas

Tipo de polimerase	Genes transcritos
RNA-polimerase I	Genes do rRNA 5,8S, 18S e 28S
RNA-polimerase II	Todos os genes que codificam proteínas, além dos genes que codificam snoRNA, miRNA, siRNA, lncRNA e a maioria dos genes de snRNA
RNA-polimerase III	Genes de tRNA, rRNA 5S, alguns snRNA e genes de outros pequenos RNAs

Os rRNAs são denominados de acordo com seus valores "S", os quais se referem às suas taxas de sedimentação em ultracentrifugação. Quanto maior o valor de S, maior é o rRNA.

**Figura 6-14 Similaridade estrutural entre uma RNA-polimerase bacteriana e uma RNA-polimerase II eucariótica.**

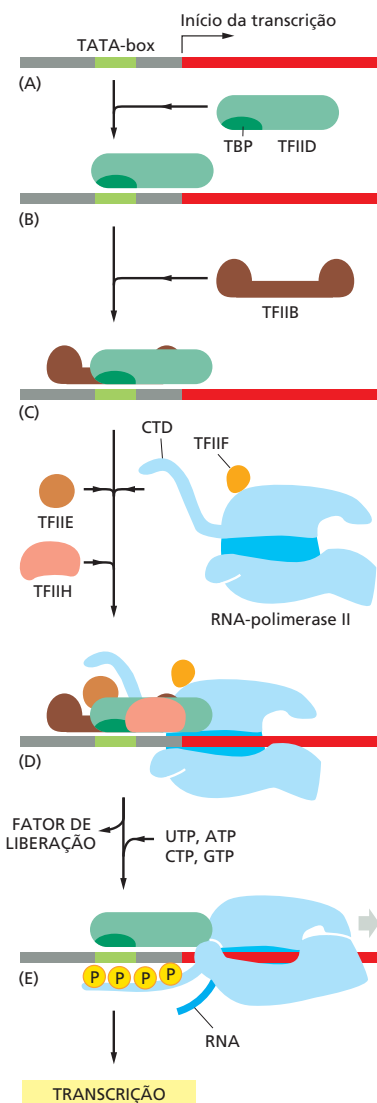
As regiões das duas RNA-polimerases que têm similaridade estrutural estão indicadas em verde. A polimerase eucariótica é maior do que a enzima bacteriana (12 subunidades em vez de cinco), e algumas das regiões adicionais estão ilustradas em cinza. As esferas azuis representam átomos de Zn que atuam como componentes estruturais das polimerases, e a esfera vermelha representa o átomo de Mg presente no sítio ativo, onde a polimerização ocorre. As RNA-polimerases de todas as células atuais (bactérias, arqueobactérias e eucariotos) são intimamente relacionadas, indicando que as características básicas da enzima existiam anteriormente à divergência dos três principais ramos da vida. (Cortesia de P. Cramer e R. Kornberg.)



## A RNA-polimerase II requer um conjunto de fatores gerais de transcrição

Os **fatores gerais de transcrição** ajudam a posicionar corretamente a RNA-polimerase eucariótica sobre o promotor, auxiliando a separação das duas cadeias de DNA para permitir o início da transcrição e liberando a RNA-polimerase do promotor para dar início ao seu modo de alongamento. As proteínas são “gerais”, porque elas são necessárias para praticamente todos os promotores utilizados pela RNA-polimerase II. Elas consistem em um conjunto de proteínas de interação denominadas arbitrariamente como TFIIA, TFIIB, TFIIC, TFIID, e assim por diante (TFII significando “fator de transcrição para a polimerase II”, do inglês *transcription factor for polymerase II*). Em um sentido amplo, os fatores gerais de transcrição eucarióticos desempenham funções equivalentes àquelas do fator  $\sigma$  em bactérias; de fato, determinadas regiões de TFIIF apresentam a mesma estrutura tridimensional que as regiões equivalentes do fator  $\sigma$ .

A **Figura 6-15** ilustra como os fatores gerais de transcrição se associam aos promotores utilizados pela RNA-polimerase II, e a **Tabela 6-3** resume suas atividades. O processo de associação começa quando o TFIID se liga a uma curta sequência de DNA de dupla-hélice principalmente composta por nucleotídeos T e A. Por essa razão, essa sequência é conhecida como a sequência TATA ou **TATA-box**, e a subunidade de TFIID que a reconhece é chamada de TBP (proteína de ligação a TATA, do inglês *TATA-binding protein*). A sequência TATA-box normalmente está localizada 25 nucleotídeos antes do sítio de início da transcrição. Essa não é a única sequência de DNA que sinaliza o início da transcrição (**Figura 6-16**), mas, para a maioria dos promotores de polimerase II, ela é a mais importante. A ligação de TFIID provoca uma grande distorção no DNA do TATA-box (**Figura 6-17**). Acredita-se que essa distorção sirva como um marco físico para a



**Figura 6-15 Iniciação da transcrição de um gene eucariótico pela RNA-polimerase II.**

Para iniciar a transcrição, a RNA-polimerase requer vários fatores gerais de transcrição. (A) O promotor contém uma sequência de DNA denominada TATA-box, localizada a 25 nucleotídeos do sítio no qual a transcrição é iniciada. (B) Por meio de sua subunidade TBP, o TFIID reconhece e se liga ao TATA-box, o que permite a ligação adjacente de TFIIB. (C) Para simplificar, a distorção do DNA produzida pela ligação de TFIID (ver Figura 6-17) não está ilustrada. (D) Os demais fatores gerais de transcrição, assim como a própria RNA-polimerase, associam-se no promotor. (E) Então, o TFIIH usa a energia da hidrólise do ATP para separar a dupla fita do DNA no ponto de início da transcrição, expondo localmente a fita-molde. O TFIIF também fosforila a RNA-polimerase II, modificando sua conformação de tal modo que a polimerase se dissocia dos fatores gerais e pode iniciar a fase de extensão da transcrição. Como ilustrado, o sítio de fosforilação é uma longa cauda polipeptídica C-terminal, também denominado domínio C-terminal (CTD), que se estende a partir da molécula de polimerase. O esquema de associação mostrado nesta figura foi deduzido a partir de experimentos realizados *in vitro*, e a ordem exata na qual os fatores gerais de transcrição se associam nos promotores, *in vivo*, provavelmente varia de acordo com o gene. Os fatores gerais de transcrição são altamente conservados; alguns dos fatores de células humanas podem ser substituídos em experimentos bioquímicos pelos fatores correspondentes de simples leveduras.

**TABELA 6-3** Os fatores gerais de transcrição necessários à iniciação da transcrição pela RNA-polimerase II eucariótica

Nome	Número de subunidades	Funções na iniciação da transcrição
TFIID Subunidade TBP Subunidades TAF	1 ~11	Reconhece o TATA-box. Reconhece outras sequências de DNA próximas ao ponto de início da transcrição; regula a ligação ao DNA pela TBP.
TFIIB	1	Reconhece o elemento BRE nos promotores; posiciona com exatidão a RNA-polimerase no sítio de início da transcrição.
TFIIF	3	Estabiliza a interação da RNA-polimerase com TBP e TFIIB; auxilia a atrair TFIIE e TFIIH.
TFIIE	2	Atrai e regula TFIIH.
TFIIH	9	Desespiraliza o DNA no sítio de início da transcrição, fosforila a Ser5 do CTD da RNA-polimerase; libera a RNA-polimerase do promotor.

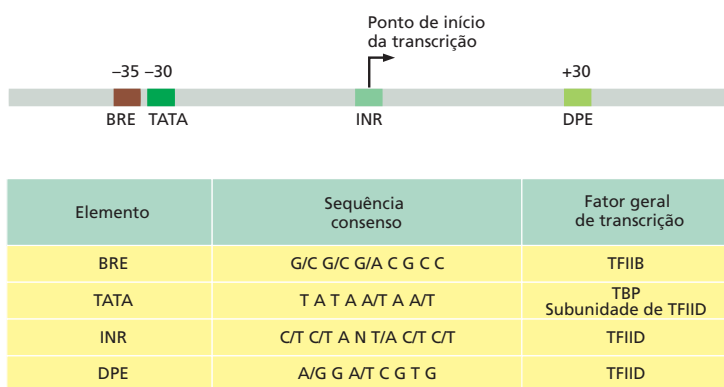
TFIID é composto por TBP e ~11 subunidades adicionais denominadas TAFs (fatores associados à TBP); CTD, domínio C-terminal.

localização de um promotor ativo no interior de um genoma extremamente grande e que mantenha as sequências de DNA de ambos os lados da distorção unidas para permitir as etapas subsequentes de associação das proteínas do complexo. Outros fatores são, então, reunidos, junto à RNA-polimerase II, para formar um *complexo de iniciação da transcrição* completo (ver Figura 6-15). O mais complexo dos fatores gerais de transcrição é TFIIH. Composto por nove subunidades, ele é praticamente tão grande quanto a própria RNA-polimerase II, sendo, como veremos em breve, o responsável pela realização de diferentes etapas necessárias à iniciação da transcrição.

Após a formação de um complexo de iniciação de transcrição sobre o DNA, a RNA-polimerase II deverá ter acesso à fita-molde no ponto de início da transcrição. O TFIIH, que contém uma DNA-helicase como uma de suas subunidades, torna possível essa etapa com a hidrólise de ATP, desespiralização do DNA e conseqüente exposição da fita-molde. A seguir, a RNA-polimerase II, da mesma forma que a polimerase bacteriana, se liga ao promotor, sintetizando pequenos fragmentos de RNA até sofrer uma série de alterações estruturais que permitem sua dissociação ao promotor e início da fase de extensão (ou alongamento) da transcrição. Uma etapa-chave para essa transição é a adição de grupos fosfato à “cauda” da RNA-polimerase (conhecida como CTD, ou domínio C-terminal, do inglês *C-terminal domain*). Em seres humanos, o CTD consiste em 52 repetições adjacentes de uma sequência de sete aminoácidos, que se estende a partir da estrutura central da RNA-polimerase. Durante a iniciação da transcrição, a serina localizada na quinta posição da sequência repetida (Ser5) é fosforilada por TFIIH, que contém uma proteína-cinase como uma de suas subunidades (ver Figura 6-15D e E). A polimerase pode, então, separar-se do agrupamento de fatores

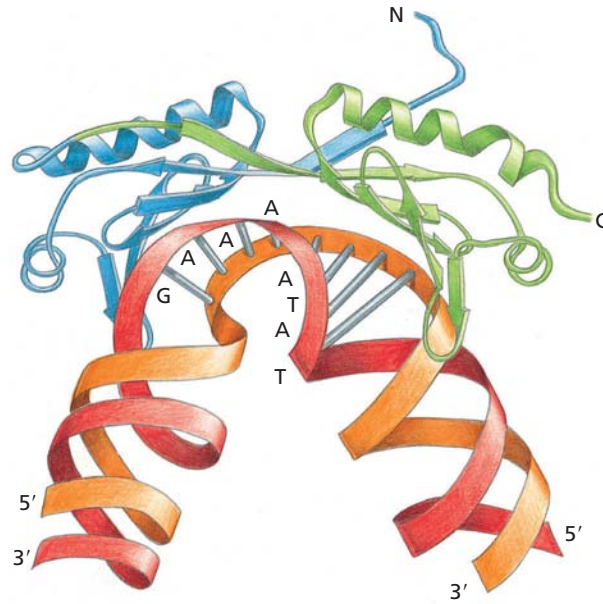
**Figura 6-16** Sequências consenso adjacentes aos pontos de iniciação da RNA-polimerase II eucariótica. Estão indicados o nome dado a cada sequência consenso (*primeira coluna*) e o fator geral de transcrição que a reconhece (*última coluna*). N indica qualquer nucleotídeo, e dois nucleotídeos separados por uma barra indicam uma probabilidade igual de qualquer um deles ocorrer na posição indicada. Na realidade, cada sequência consenso é uma representação resumida de um histograma similar ao da Figura 6-12.

Na maioria dos pontos de iniciação da transcrição da RNA-polimerase II, apenas duas ou três das quatro sequências estão presentes. Por exemplo, a maioria dos promotores da polimerase II tem uma sequência TATA-box, e aqueles que não a possuem normalmente apresentam uma sequência INR “forte”. Embora a maioria das sequências de DNA que influenciam o início da transcrição esteja localizada acima do ponto de iniciação da transcrição, algumas poucas, como o elemento DPE mostrado na figura, estão localizadas na região transcrita.





**Figura 6-17 Estrutura tridimensional da TBP (proteína de ligação ao TATA) ligada ao DNA.** A TBP é a subunidade do fator geral de transcrição TFIID responsável pelo reconhecimento e pela ligação à sequência TATA-box no DNA (*vermelho*). Acredita-se que a curvatura característica do DNA provocada por TBP – duas porções curvadas na dupla hélice separadas por DNA parcialmente desenovelado – sirva como um sinal que ajuda a atrair os demais fatores gerais de transcrição (*Animação 6.4*). A TBP é uma cadeia polipeptídica simples enovelada em dois domínios muito semelhantes (*azul e verde*). (Adaptada de J.L. Kim et al., *Nature* 365:520–527, 1993. Com permissão de Macmillan Publishers Ltd.)

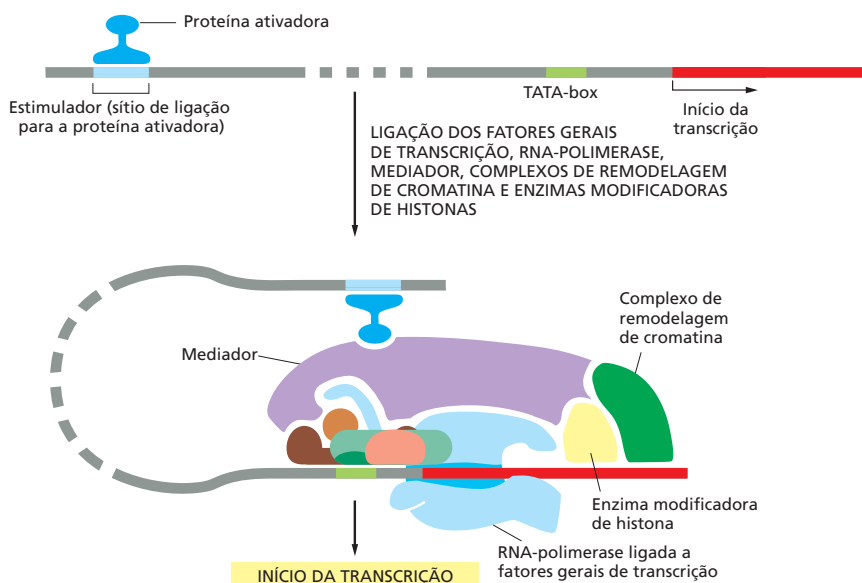


gerais de transcrição. Durante esse processo, ela sofre uma série de modificações conformacionais que fortalecem sua interação com o DNA e adquire novas proteínas que lhe permitem transcrever por longas distâncias e, em muitos casos, por várias horas, sem se dissociar do DNA.

Uma vez que a polimerase II tenha iniciado a extensão do transcrito de RNA, a maioria dos fatores gerais de transcrição é liberada do DNA de forma que eles estarão disponíveis para iniciar outro ciclo de transcrição, com uma nova molécula de RNA-polimerase. Como vimos resumidamente, a fosforilação da cauda da RNA-polimerase II tem uma função adicional: ela também faz os componentes da maquinaria do processamento do RNA se associarem à polimerase e, dessa forma, estarem em posição para modificar o RNA recém-transcrito assim que ele emergir da polimerase.

### A polimerase II também requer proteínas ativadoras, mediadoras e modificadoras de cromatina

Estudos da RNA-polimerase II e de seus fatores gerais de transcrição atuando sobre moléculas de DNA em sistemas *in vitro* purificados estabeleceram o modelo para a iniciação da transcrição que acabamos de descrever. Entretanto, como discutido no Capítulo 4, o DNA das células eucarióticas está empacotado em nucleossomos, os quais ainda são organizados em estruturas de cromatina de maior complexidade. Como resultado, a iniciação da transcrição nas células eucarióticas é mais complexa e requer mais proteínas do que a iniciação da transcrição em DNA purificado. Primeiramente, as proteínas reguladoras de genes conhecidas como *ativadoras transcricionais* devem se ligar a sequências específicas sobre o DNA (denominadas *enhancers* ou *estimuladores*) e auxiliar a atrair a RNA-polimerase II para o ponto de iniciação da transcrição (**Figura 6-18**). Discutiremos o papel desses ativadores no Capítulo 7, pois representam uma das principais formas de regulação da expressão gênica nas células. Aqui, apenas chamamos a atenção para o fato de que sua presença no DNA é necessária para a iniciação da transcrição em uma célula eucariótica. Em segundo lugar, a iniciação da transcrição eucariótica *in vivo* necessita da presença de um grande complexo proteico conhecido como *Mediador*, o qual permite que as proteínas ativadoras se comuniquem adequadamente com a polimerase II e com os fatores gerais de transcrição. Por fim, a iniciação da transcrição nas células eucarióticas normalmente requer o recrutamento de enzimas modificadoras da cromatina, como complexos remodeladores de cromatina e enzimas modificadoras de histonas. Como discutimos no Capítulo 4, ambos os tipos de enzimas podem aumentar o acesso ao DNA



**Figura 6-18 Iniciação da transcrição pela RNA-polimerase II em uma célula eucariótica.** O início da transcrição *in vivo* requer a presença de proteínas ativadoras de transcrição. Como descrito no Capítulo 7, essas proteínas se ligam a pequenas sequências específicas no DNA. Embora somente uma seja aqui apresentada, um gene eucariótico típico utiliza várias proteínas ativadoras de transcrição que, combinadas, determinam sua taxa e seu padrão de transcrição. Às vezes agindo a uma distância de vários milhares de pares de nucleotídeos (indicados pela linha tracejada na molécula de DNA), essas proteínas auxiliam a RNA-polimerase, os fatores gerais e o Mediador a associarem-se no promotor. Além disso, ativadores atraem complexos de remodelagem da cromatina dependentes de ATP e enzimas modificadoras de histonas. Um dos principais papéis do Mediador é coordenar a associação de todas essas proteínas sobre o promotor de tal forma que a transcrição possa começar. Como discutido no Capítulo 4, o estado “padrão” da cromatina é o de uma fibra condensada (ver Figura 4-28), e essa é provavelmente a forma de DNA sobre a qual a maior parte da transcrição é iniciada. Para simplificar, a cromatina não foi ilustrada na figura.

da cromatina e, assim, facilitam a montagem da maquinaria de iniciação da transcrição sobre o DNA.

Como ilustrado na Figura 6-18, várias proteínas (bem mais de uma centena de subunidades individuais) devem se associar no ponto de início da transcrição para promover a iniciação da transcrição em uma célula eucariótica. A ordem de associação dessas proteínas não parece seguir uma rota preestabelecida; de fato, ela varia entre diferentes genes. Na verdade, alguns desses diferentes complexos proteicos podem ser transportados para o DNA sob a forma de subarranjos pré-formados.

Para iniciar a transcrição, a RNA-polimerase II deve ser liberada desse grande complexo de proteínas. Além das etapas descritas na Figura 6-14, essa liberação muitas vezes requer a proteólise *in situ* da proteína ativadora. Voltaremos a algumas dessas questões, incluindo o papel dos complexos de remodelagem da cromatina e das enzimas modificadoras de histonas, no Capítulo 7, no qual discutiremos como as células eucarióticas regulam o processo de iniciação da transcrição.

## O alongamento da transcrição nos eucariotos requer proteínas acessórias

Uma vez que a RNA-polimerase tenha iniciado a transcrição, ela move-se de forma irregular, parando em algumas sequências de DNA e transcrevendo rapidamente outras. As RNA-polimerases em funcionamento, tanto em bactérias quanto em eucariotos, estão associadas a uma série de *fatores de alongamento* (ou *fatores de extensão*), proteínas que diminuem a probabilidade de dissociação da RNA-polimerase antes que esta chegue ao término de um gene. Esses fatores caracteristicamente associam-se à RNA-polimerase logo após a iniciação ter ocorrido e ajudam a polimerase a se mover sobre a ampla variedade de sequências de DNA encontradas nos genes. As RNA-polimerases eucarióticas também devem lidar com a estrutura da cromatina conforme elas se movem sobre o molde de DNA e, para isso, geralmente são auxiliadas por complexos de remodelagem da cromatina dependentes de ATP que podem mover-se com a polimerase ou simplesmente podem procurar e resgatar uma polimerase que eventualmente esteja paralisada. Além disso, as chaperonas de histonas ajudam a dissociar parcialmente os nucleossomos a frente de uma RNA-polimerase em movimento e a associá-los após sua passagem.

À medida que a RNA-polimerase move-se sobre um gene, algumas das enzimas ligadas a ela modificam as histonas, deixando para trás um registro da passagem da polimerase. Embora não esteja exatamente claro como a célula usa essa informação, isso

talvez possa ajudar a transcrever novamente um gene, uma vez que ele tenha se tornado ativo anteriormente. Talvez essa informação também seja útil para acoplar o alongamento da transcrição ao processamento do RNA, conforme este emerge da RNA-polimerase, um tópico que discutiremos mais adiante neste capítulo.

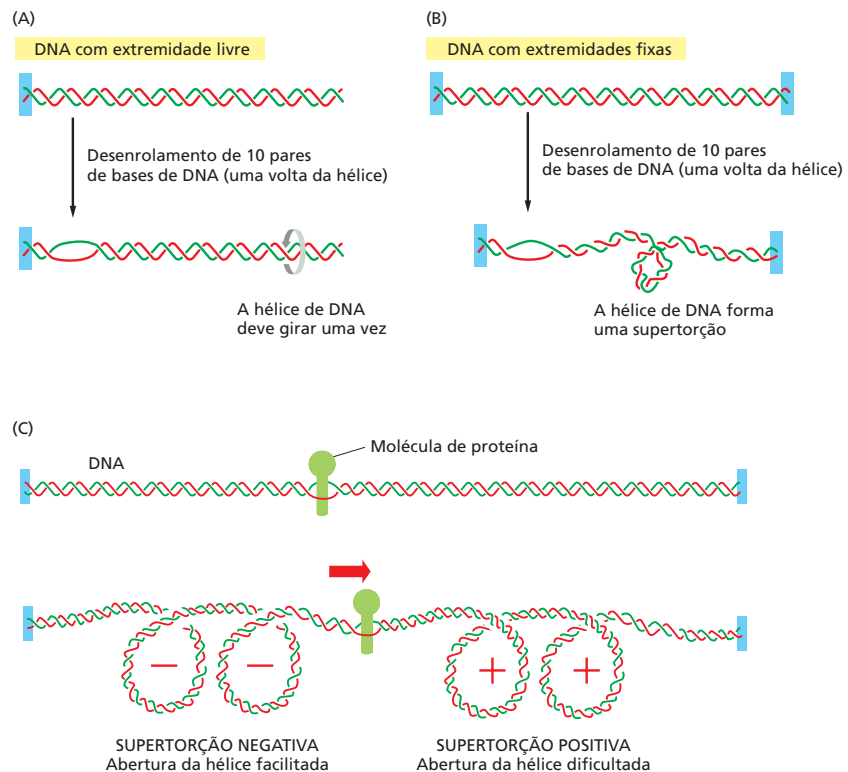
### A transcrição cria tensão super-helicoidal

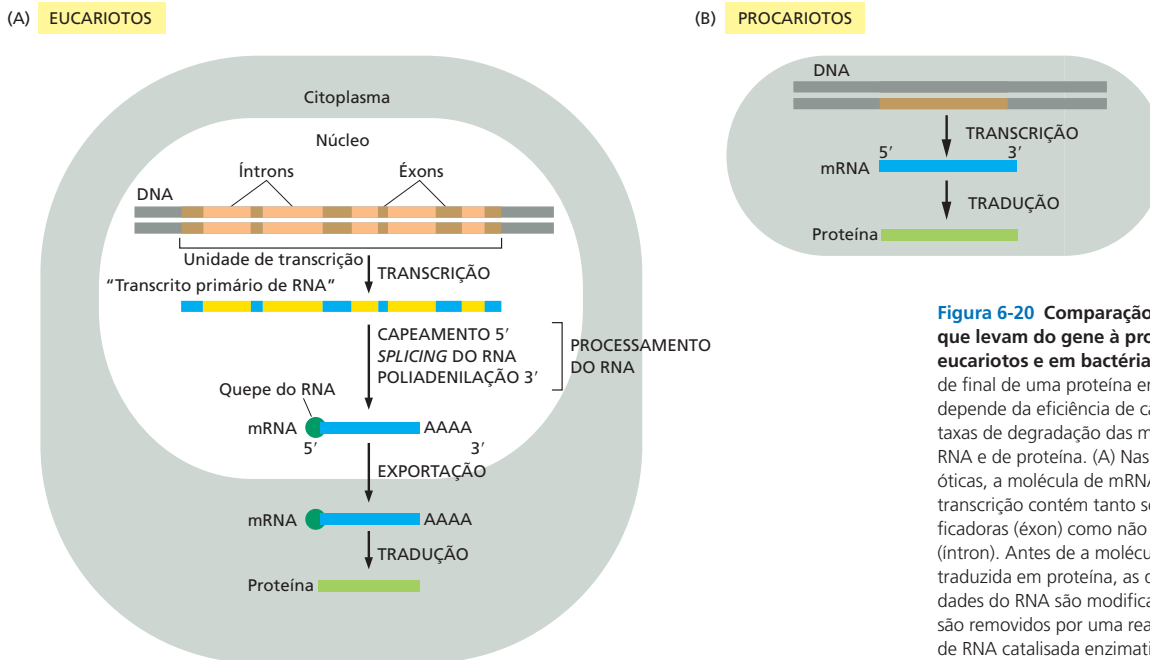
Existe ainda outra barreira para o alongamento pelas RNA-polimerases, sejam elas bacterianas ou eucarióticas, que também se aplica às DNA-polimerases, como discutido no Capítulo 5 (ver Figura 5-20). Para descrever esse assunto em mais detalhes, primeiramente devemos considerar uma propriedade sutil inerente ao DNA de dupla-hélice denominada **supertorção do DNA**. A supertorção do DNA é o nome dado a uma conformação que o DNA adota em resposta à tensão super-helicoidal; alternativamente, a criação de alças ou dobras em uma molécula de DNA dupla hélice pode criar tal tensão. A **Figura 6-19** ilustra a razão dessa situação. Existem aproximadamente 10 pares de nucleotídeos para cada giro da hélice em uma dupla-hélice de DNA. Se imaginarmos uma hélice, cujas duas extremidades estão fixas uma em relação à outra (como ocorre em um DNA circular, como um cromossomo bacteriano, ou em alças firmemente apertadas, como se acredita estarem dispostos os cromossomos eucarióticos), uma grande supertorção se formará para compensar cada 10 pares de nucleotídeos que são abertos (desenrolados). A formação dessa supertorção é energeticamente favorável, pois restaura o enrolamento helicoidal normal das regiões que permanecem pareadas, que, caso contrário, sofreriam uma superespiralização devido às suas extremidades fixas.

A tensão super-helicoidal é criada conforme a RNA-polimerase se move ao longo da fita de DNA que possui extremidades fixas (ver Figura 6-19C). Considerando que a polimerase não é livre para girar rapidamente (e que tal rotação é pouco provável devido ao tamanho das RNA-polimerases e de seus transcritos acoplados), uma polimerase em movimento gera tensão positiva da super-hélice no DNA à sua frente e tensão helicoidal negativa atrás de si. Para eucariotos, acredita-se que essa situação represente um bônus: embora a tensão super-helicoidal positiva à frente da polimerase torne a hélice de DNA mais difícil de abrir, a tensão deve facilitar o desenrolamento parcial do DNA nos

**Figura 6-19** A tensão super-helicoidal no DNA causa supertorção do DNA.

(A) Para uma molécula de DNA com uma extremidade livre (ou com uma quebra em uma das fitas que serve como ponto de torção), a dupla-hélice de DNA gira uma volta a cada 10 pares de nucleotídeos que são abertos. (B) Se a rotação é impedida, ocorre introdução de tensão super-helicoidal no DNA quando a hélice é aberta. No exemplo mostrado, a hélice de DNA contém 10 voltas helicoidais, uma das quais está aberta. Uma forma de acomodar a tensão criada seria aumentar a torção helicoidal de 10 para 11 pares de nucleotídeos por volta na dupla hélice remanescente. A hélice do DNA, no entanto, resiste a tal deformação como uma mola, preferindo aliviar a tensão super-helicoidal pela formação de alças supertorcidas. Como resultado, uma supertorção de DNA forma-se na dupla-hélice do DNA a cada 10 pares de nucleotídeos abertos. A supertorção formada nesse caso é uma supertorção positiva. (C) A supertorção do DNA é induzida por uma proteína que trafega sobre a dupla-hélice de DNA. As duas extremidades do DNA ilustradas aqui não são capazes de girar livremente uma em relação à outra, e acredita-se que a molécula proteica também seja impedida de rotação livre conforme se move. Sob essas condições, o movimento da proteína provoca um excesso de torção que se acumula na hélice de DNA à sua frente e um déficit de torção no DNA atrás da proteína, conforme ilustrado.





**Figura 6-20 Comparação das etapas que levam do gene à proteína em eucariotos e em bactérias.** A quantidade final de uma proteína em uma célula depende da eficiência de cada etapa e das taxas de degradação das moléculas de RNA e de proteína. (A) Nas células eucarióticas, a molécula de mRNA produzida por transcrição contém tanto sequências codificadoras (éxon) como não codificadoras (íntron). Antes de a molécula de RNA ser traduzida em proteína, as duas extremidades do RNA são modificadas, os íntrons são removidos por uma reação de *splicing* de RNA catalisada enzimaticamente e o mRNA resultante é transportado do núcleo para o citoplasma. Por conveniência, as etapas nesta figura estão representadas como se ocorressem uma de cada vez; na realidade, várias etapas ocorrem simultaneamente. Por exemplo, o quepe de RNA é adicionado, e o *splicing* inicia antes que a transcrição tenha sido completada. Devido ao acoplamento entre transcrição e processamento de RNA, os transcritos primários intactos – os RNAs completos que seriam, em teoria, produzidos antes de o processamento ocorrer – raramente são encontrados. (B) Nos procariotos, a produção de moléculas de mRNA é muito mais simples. A extremidade 5' de uma molécula de mRNA é produzida por meio da iniciação da transcrição, e a extremidade 3' é produzida pela terminação da transcrição. Visto que as células procarióticas não possuem um núcleo, a transcrição e a tradução acontecem em um compartimento comum, e a tradução de um mRNA bacteriano frequentemente tem início antes da sua síntese ter sido concluída.

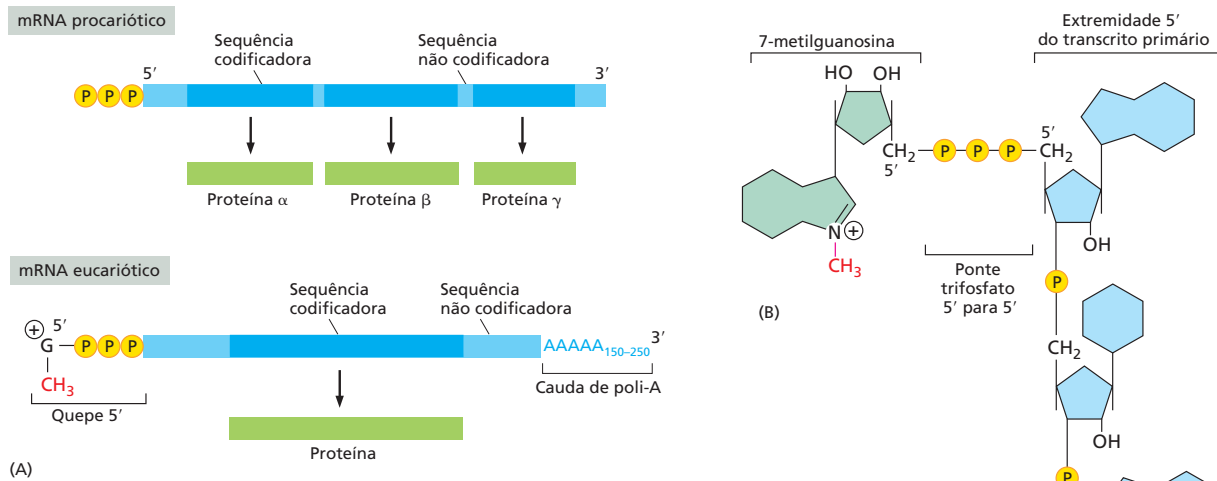
nucleossomos, conforme a separação do DNA dos grupos de histonas ajuda a relaxar essa tensão.

Qualquer proteína que se autopropulsiona sobre uma fita de DNA dupla-hélice, como uma DNA-helicase ou uma RNA-polimerase, tende a gerar tensão super-helicoidal. Nos eucariotos, as enzimas DNA-topoisomerases removem rapidamente essa tensão da super-hélice (ver p. 251-253). Nas bactérias, porém, uma topoisomerase especializada, denominada *DNA-girase*, usa a energia de hidrólise de ATP para continuamente introduzir supertorções no DNA, mantendo, dessa forma, o DNA sob tensão constante. Elas são *supertorções negativas*, possuindo um direcionamento oposto ao das *supertorções positivas* que se formam quando uma região da hélice de DNA se abre (ver Figura 6-19B). Essas supertorções negativas são removidas do DNA bacteriano toda vez que uma região da hélice se abre, reduzindo a tensão da super-hélice. A DNA-girase, portanto, torna a abertura da hélice de DNA das bactérias energeticamente favorável comparada à abertura da hélice no DNA que não está supertorcido. Por essa razão, a supertorção facilita os processos genéticos, como a iniciação da transcrição pela RNA-polimerase bacteriana, que requer a abertura da hélice em bactérias (ver Figura 6-11).

### O alongamento da transcrição em eucariotos está fortemente associado ao processamento de RNA

Vimos que os mRNAs bacterianos são sintetizados pela RNA-polimerase, iniciando e terminando em regiões específicas do genoma. A situação em eucariotos é consideravelmente diferente. Particularmente, a transcrição é apenas a primeira de diferentes etapas necessárias para a produção de uma molécula de mRNA madura. Outras etapas essenciais incluem a modificação covalente de ambas as extremidades do RNA e a remoção de *sequências de íntrons* que são retiradas do transcrito de RNA pelo processo de *splicing do RNA* (Figura 6-20).

Ambas as extremidades do mRNA eucariótico são modificadas: pelo *capeamento* na extremidade 5' e pela *poliadenilação* na extremidade 3' (Figura 6-21). Essas extremidades especiais permitem que a célula verifique se ambas as extremidades de uma molécula de mRNA estão presentes (e, como consequência, se a mensagem está intacta), antes de exportar o RNA do núcleo para ser traduzido em proteína. A retirada de íntrons do RNA une as diferentes porções de uma sequência codificadora de proteínas e permite



**Figura 6-21** Uma comparação entre estruturas de moléculas de mRNA procariótico e eucariótico. (A) As extremidades 5' e 3' de um mRNA bacteriano são as extremidades não modificadas da cadeia sintetizada pela RNA-polimerase, a qual inicia e termina a transcrição naqueles pontos, respectivamente. As extremidades correspondentes de um mRNA eucariótico são formadas pela adição de um quepe na extremidade 5' e pela clivagem do transcrito pré-mRNA próximo à extremidade 3' e adição de uma sequência poli-A, respectivamente. A figura também ilustra outra diferença entre os mRNAs procarióticos e eucarióticos: os mRNAs bacterianos podem conter as instruções para várias proteínas diferentes, enquanto os mRNAs eucarióticos praticamente sempre contêm a informação para uma única proteína. (B) A estrutura do quepe na extremidade 5' de moléculas de mRNA eucariótico. Observe a ligação incomum 5' para 5' da 7-metil G ao restante do RNA. Muitos mRNAs eucarióticos possuem uma modificação adicional: a metilação do grupo 2'-hidroxila da ribose na extremidade 5' do transcrito primário (ver Figura 6-23).

que os eucariotos superiores tenham a capacidade de sintetizar várias proteínas diferentes a partir do mesmo gene.

Uma estratégia simples evoluiu para acoplar todas as etapas do processamento do RNA supradescritas ao alongamento da transcrição. Como discutido anteriormente, uma etapa-chave da iniciação da transcrição pela RNA-polimerase II é a fosforilação da cauda da RNA-polimerase II, também denominada CTD (domínio C-terminal). Essa fosforilação, que ocorre gradualmente à medida que a RNA-polimerase inicia a transcrição e move-se ao longo do DNA, não apenas ajuda a dissociar a RNA-polimerase II das outras proteínas presentes no ponto de início da transcrição, como também permite que um novo conjunto de proteínas que atuam no alongamento da transcrição e no processamento do RNA se associe à cauda da RNA-polimerase. Como discutiremos a seguir, algumas dessas proteínas de processamento parecem “saltar” da cauda da polimerase sobre a molécula de RNA em formação para começar seu processamento, conforme ela emerge da RNA-polimerase. Assim, podemos ver a RNA polimerase II no seu modo de alongamento, como uma fábrica de RNA que não apenas se move sobre o DNA sintetizando uma molécula de RNA, mas que também processa o RNA produzido (**Figura 6-22**). A CTD totalmente estendida é quase 10 vezes mais longa que o restante da RNA-polimerase. Por ser um domínio proteico flexível, a CTD atua como um suporte ou como uma corda, mantendo próximas uma ampla variedade de proteínas que poderão atuar de forma rápida, quando necessário. Essa estratégia, que acelera significativamente a taxa global de uma série de reações consecutivas, é amplamente utilizada nas células (ver Figuras 4-58 e 16-18).

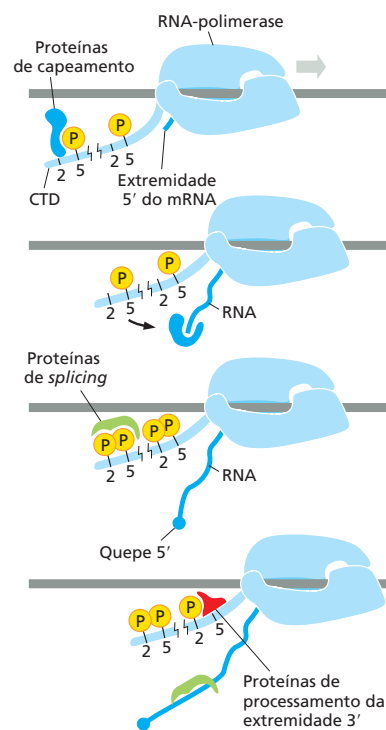
### O capeamento do RNA é a primeira modificação dos pré-mRNAs eucarióticos

Assim que a RNA-polimerase II tenha produzido aproximadamente 25 nucleotídeos de RNA, a extremidade 5' da nova molécula de RNA é modificada pela adição de um “quepe” (ou capa) que consiste em um nucleotídeo guanina modificado (ver Figura 6-21B). A reação de capeamento é realizada por três enzimas que agem sucessivamente: uma remove um fosfato da extremidade 5' do RNA nascente (uma fosfatase), outra (uma guaniltransferase) adiciona um GMP em uma ligação reversa (5' para 5' em vez de 5' para



**Figura 6-22** O conceito de “fábrica de RNA” para a RNA-polimerase II eucariótica. A polimerase não somente transcreve DNA em RNA, mas também transporta proteínas processadoras de RNA em sua cauda, as quais são transferidas para o RNA em formação no momento adequado. A cauda contém 52 repetições adjacentes de uma sequência de sete aminoácidos contendo duas serinas em cada repetição. As proteínas de capeamento se ligam à cauda da RNA-polimerase quando esta é fosforilada na Ser5 do heptâmero repetido, no final do processo de iniciação da transcrição (ver Figura 6-15). Essa estratégia assegura que a molécula de RNA seja capeada de maneira eficiente assim que sua extremidade 5’ emergir da RNA-polimerase. Conforme a polimerase segue a transcrição, sua cauda é extensivamente fosforilada nas posições de Ser2 por uma cinase associada à polimerase de extensão, sendo eventualmente desfosforilada nas posições de Ser5. Essas novas alterações atraem proteínas de *splicing* e de processamento da extremidade 3’ para a polimerase em movimento, posicionando-as para agir sobre o RNA recém-sintetizado assim que ele venha a emergir da RNA-polimerase. Existem muitas enzimas de processamento do RNA e nem todas se associam à polimerase. No caso do *splicing* do RNA, por exemplo, a cauda carrega apenas uns poucos componentes essenciais; uma vez transferidos para a molécula de RNA, eles atuam como nucleadores para os demais componentes do processo.

Uma vez que a RNA-polimerase II termina de transcreever um gene, ela é liberada do DNA, os fosfatos da sua cauda são removidos por fosfatases solúveis, e ela pode reiniciar a transcrição. Apenas essa forma completamente desfosforilada da RNA-polimerase II é capaz de iniciar a síntese de RNA em um promotor.



3’), e uma terceira (uma metiltransferase) adiciona um grupo metila à guanosina (**Figura 6-23**). Visto que todas as três enzimas ligam-se à cauda fosforilada da RNA-polimerase na posição Ser5 – a modificação adicionada pelo TFIIF durante a iniciação da transcrição – elas estão prontas para modificar a extremidade 5’ do transcrito nascente assim que ela emerge da polimerase.

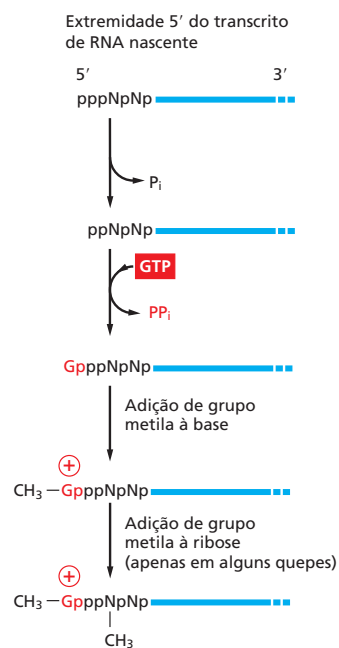
O quepe metil 5’ identifica a extremidade 5’ de mRNAs eucarióticos, e essa marca ajuda a célula a distinguir os mRNAs dos outros tipos de moléculas de RNA presentes na célula. Por exemplo, as RNA-polimerases I e III produzem RNAs não capeados durante a transcrição, em parte porque essas polimerases carecem de um CTD. No núcleo, o quepe se liga a um complexo proteico denominado complexo de ligação ao quepe (CBC, *cap-binding complex*), o qual, como discutiremos em seções subsequentes, ajuda o processamento e a exportação dos futuros mRNAs. O quepe metil 5’ também desempenha um importante papel na tradução dos mRNAs no citosol, como discutiremos adiante neste capítulo.

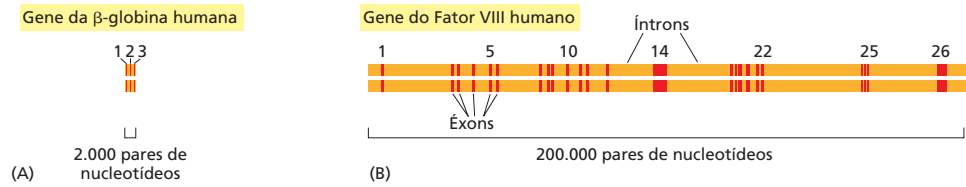
### O *splicing* do RNA remove as sequências de íntrons de pré-mRNAs recentemente transcritos

Como discutido no Capítulo 4, as sequências codificadoras de genes eucarióticos são caracteristicamente interrompidas por sequências intervenientes não codificadoras (íntrons). Descoberta em 1977, essa característica dos genes eucarióticos foi uma surpresa para os cientistas, que estavam familiarizados, até aquele momento, apenas com genes bacterianos, os quais, caracteristicamente, consistem em uma porção contínua de DNA codificador diretamente transcrita em mRNA. Contrastando fortemente, os genes eucarióticos são encontrados sob a forma de pequenos pedaços de sequências codificadoras (*sequências expressas* ou **éxons**) intercaladas por sequências muito mais longas, as *sequências intervenientes* ou **íntrons**; assim, a porção codificadora de um gene eucariótico é, em geral, apenas uma pequena fração do comprimento total do gene (**Figura 6-24**).

Tanto as sequências de íntrons quanto as de éxons são transcritas em RNA. As sequências dos íntrons são removidas do RNA recentemente sintetizado por meio de um processo denominado ***splicing* de RNA**. Grande parte do *splicing* de RNA que ocorre nas células atua na produção de mRNA, e nossa discussão sobre o *splicing* focaliza-se nesse tipo, denomina-

**Figura 6-23** Reações que adicionam o quepe na extremidade 5’ de cada molécula de RNA sintetizada pela RNA-polimerase II. O quepe final contém uma nova ligação 5’ para 5’ entre o resíduo 7-metil G positivamente carregado e a extremidade 5’ do transcrito de RNA (ver Figura 6-21B). A letra N representa qualquer um dos quatro ribonucleotídeos, embora o nucleotídeo inicial em uma cadeia de RNA geralmente seja uma purina (um A ou um G). (De A.J. Shatkin, *BioEssays* 7:275–277, 1987. Com permissão de Wiley-Liss, Inc., uma subsidiária de John Wiley & Sons, Inc.)





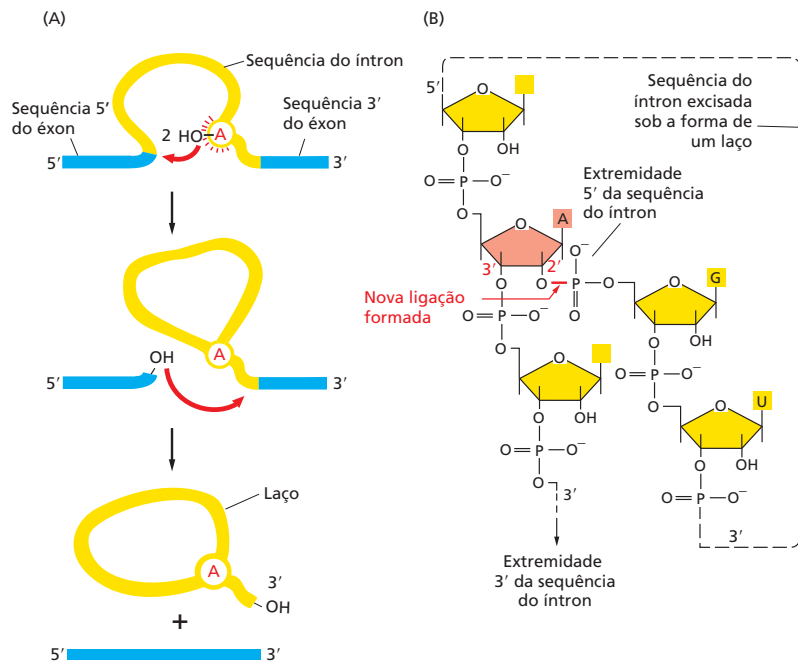
**Figura 6-24 Estrutura de dois genes humanos mostrando a organização de éxons e de íntrons.** (A) O gene relativamente pequeno da  $\beta$ -globina, que codifica uma subunidade da proteína carreadora de oxigênio hemoglobina, contém três éxons (ver também Figura 4-7). (B) O gene do Fator VIII, bem maior, contém 26 éxons; esse gene codifica uma proteína (Fator VIII) que atua na via de coagulação do sangue. A forma mais prevalente de hemofilia é resultado de mutações nesse gene.

do *splicing* do precursor de mRNA (ou pré-mRNA). Somente após ter ocorrido o *splicing* e o processamento das extremidades 5' e 3', esse RNA será denominado mRNA.

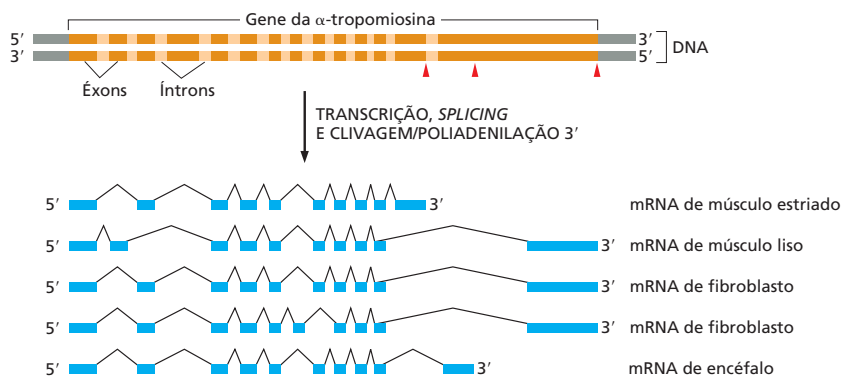
Cada evento de *splicing* remove um íntron, por meio de duas reações sequenciais de transferência de fosforil, conhecidas como transesterificações, as quais unem dois éxons, enquanto removem o íntron entre eles sob a forma de um “laço” (Figura 6-25). A maquinaria que catalisa o *splicing* do pré-mRNA é complexa, consistindo em cinco moléculas adicionais de RNA e várias centenas de proteínas, e muitas moléculas de ATP são hidrolisadas por evento de *splicing*. Essa complexidade é presumivelmente necessária para assegurar que o *splicing* seja exato, sendo ao mesmo tempo suficientemente flexível para lidar com a enorme diversidade de íntrons encontrada em uma célula eucariótica típica.

Podem parecer uma perda de tempo e energia remover um grande número de íntrons no *splicing* do RNA. Na tentativa de explicar as razões do *splicing*, os cientistas descobriram que o arranjo éxon-íntron parece facilitar o aparecimento de proteínas novas e úteis em uma escala de tempo evolutiva. Assim, a presença de numerosos íntrons no DNA permite que a recombinação genética facilmente combine éxons de diferentes genes, possibilitando que genes para novas proteínas evoluam mais facilmente devido à combinação de porções de genes preexistentes. Essa ideia é apoiada pela observação, descrita no Capítulo 3, de que muitas proteínas nas células atuais assemelham-se a uma “colcha de retalhos” composta a partir de um conjunto comum de *domínios* proteicos (ver p. 121-122).

Atualmente, o *splicing* de RNA também apresenta uma vantagem. Os transcritos de vários genes eucarióticos (estimam-se 95% dos genes em humanos) sofrem *splicing* de diferentes maneiras, permitindo que um mesmo gene produza um conjunto correspondente de diferentes proteínas (Figura 6-26). Assim, em vez de representar um processo desnecessário e de gasto excessivo, como aparentava ser, à primeira vista, o *splicing* de RNA permite aos eucariotos incrementar o potencial codificador de seus genomas. Retornaremos a essa ideia várias vezes neste capítulo e no próximo, mas primeiramente precisamos descrever a maquinaria celular que realiza essa fascinante tarefa.



**Figura 6-25 Reação de *splicing* do pré-mRNA.** (A) Na primeira etapa, um nucleotídeo de adenina específico na sequência do íntron (indicado em vermelho) ataca a região 5' de *splicing* e corta a cadeia principal de açúcar-fosfato do RNA nesse ponto. A extremidade 5' cortada do íntron torna-se covalentemente ligada ao nucleotídeo de adenina, como mostrado no detalhe em (B), criando, dessa forma, uma alça na molécula de RNA. A extremidade 3'-OH livre liberada da sequência do éxon reage com o início da sequência do éxon seguinte, unindo os dois éxons e liberando a sequência do íntron na forma de um laço. As duas sequências éxon são, desse modo, unidas em uma sequência codificadora contínua. A sequência do íntron liberada é eventualmente dissociada em nucleotídeos únicos, que são reciclados.



**Figura 6-26 Splicing alternativo do gene de α-tropomiosina de rato.** A α-tropomiosina é uma proteína de super-hélice (ver Figura 3-9) que desempenha várias tarefas, predominantemente na regulação da contração nas células musculares. O transcrito primário pode ter seus íntrons retirados de diferentes maneiras, como indicado na figura, para produzir mRNAs distintos que dão origem a variantes proteicas. Alguns desses padrões de *splicing* são específicos de certos tipos de células. Por exemplo, a α-tropomiosina produzida no músculo estriado é diferente daquela produzida pelo mesmo gene em um músculo liso. As pontas de seta na parte superior da figura marcam os sítios onde a clivagem e a adição de poli-A formam as extremidades 3' dos mRNAs maduros.

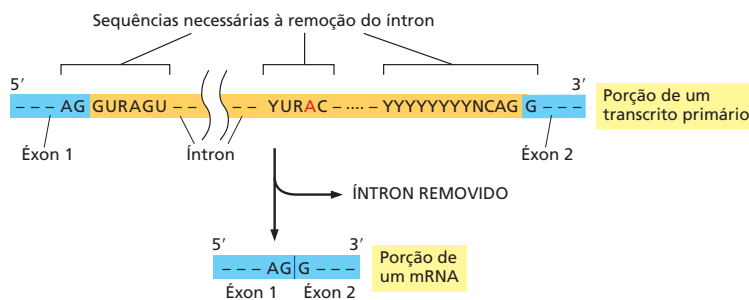
### As seqüências nucleotídicas sinalizam onde ocorre o *splicing*

O mecanismo de *splicing* de pré-mRNA ilustrado na Figura 6-24 implica que a maquinaria de *splicing* deve reconhecer três regiões na molécula RNA precursora: a região de *splicing* 5', a região de *splicing* 3' e o ponto da forquilha na sequência do íntron que forma a base do fragmento em laço a ser excisado. Como esperado, cada um desses três sítios tem uma seqüência nucleotídica consenso, que é similar entre diferentes íntrons e que fornece informações para a célula a respeito do local onde deve ocorrer o *splicing* (Figura 6-27). Entretanto, essas seqüências consenso são relativamente curtas e aceitam uma extensa variabilidade; como veremos em breve, a célula incorpora outros tipos de informações adicionais para definir exatamente onde, sobre cada molécula de RNA, deverá ocorrer o *splicing*.

A alta variabilidade das seqüências consenso de *splicing* é um desafio adicional para os cientistas determinados a decifrar as seqüências genômicas. O tamanho dos íntrons varia de aproximadamente 10 nucleotídeos a mais de 100 mil e a determinação dos limites exatos de cada íntron é uma tarefa árdua, mesmo com a ajuda de computadores de alto desempenho. A possibilidade da existência de *splicing* alternativo aumenta a dificuldade de previsão de seqüências proteicas unicamente a partir de uma seqüência genômica. Essa dificuldade é um dos principais obstáculos para a identificação de todos os genes em um genoma completo, sendo uma das principais razões de conhecermos somente o número aproximado de diferentes proteínas produzidas no genoma humano.

### O *splicing* do RNA é realizado pelo spliceossomo

Diferentemente das outras etapas de produção de mRNA já discutidas, as etapas-chave do *splicing* do RNA são realizadas por moléculas de RNA e não por proteínas. As moléculas especializadas de RNA reconhecem as seqüências nucleotídicas que especificam onde o *splicing* deve ocorrer e também catalisam a química do *splicing*. Essas moléculas de RNA são relativamente pequenas (possuem menos de 200 nucleotídeos cada uma), e existem cinco delas, U1, U2, U4, U5 e U6. Conhecidas como **snRNAs (pequenos RNAs nucleares, small nuclear RNAs)**, cada uma é complexada com pelo menos sete subunidades proteicas para formar uma snRNP (pequena ribonucleoproteína nuclear, de *small nuclear ribonucleoprotein*). As snRNPs formam o cerne do **spliceossomo**, o grande arranjo de moléculas de RNA e de proteínas que realiza o *splicing* do pré-mRNA na célula. Durante a reação de *splicing*, o reconhecimento da junção de processamento 5', do ponto



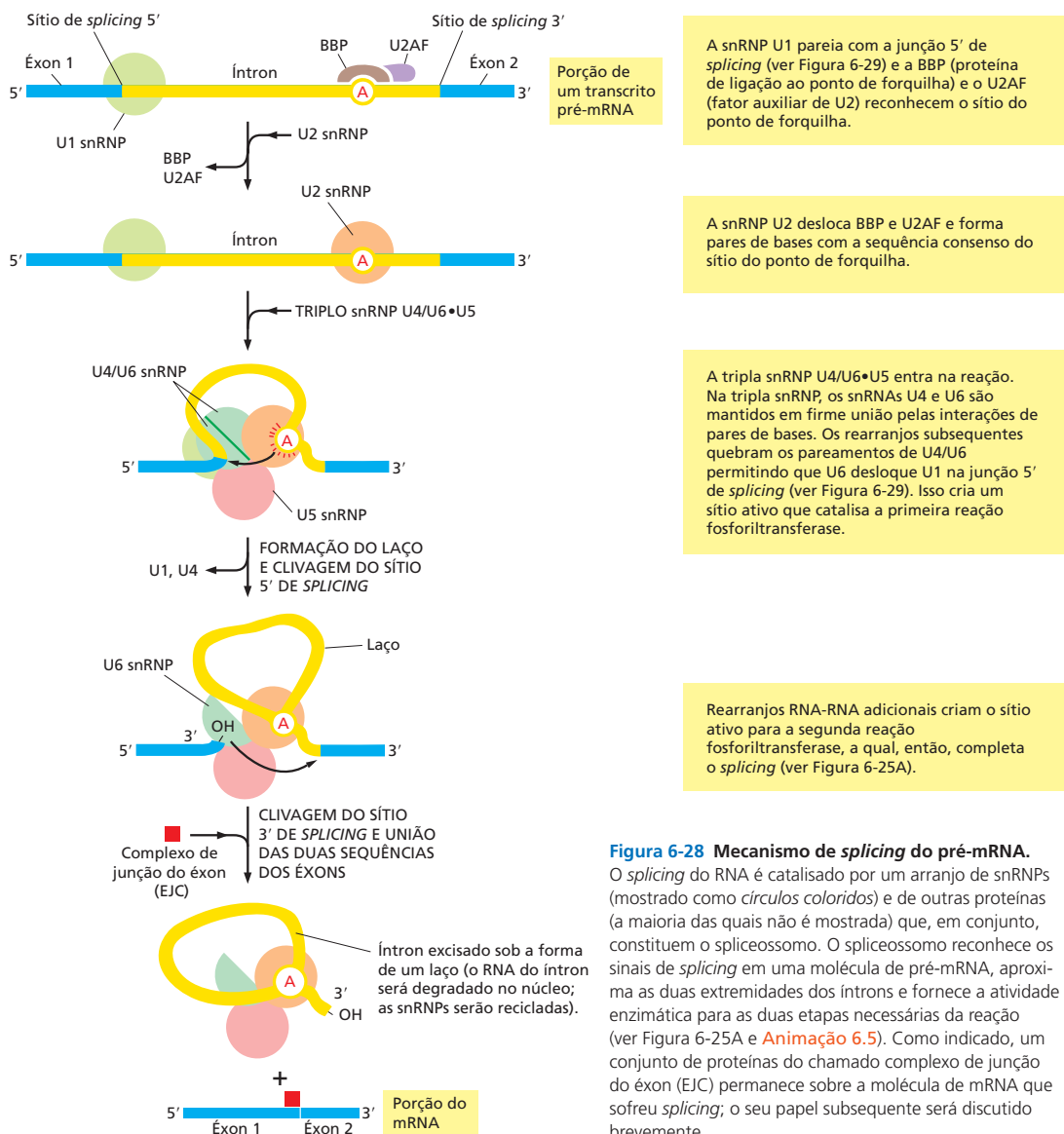
**Figura 6-27 Seqüências consenso de nucleotídeos em uma molécula de RNA que sinalizam o início e o final da maioria dos íntrons humanos.** Os três blocos de seqüências nucleotídicas ilustrados são necessários para a remoção do íntron. Aqui A, G, U e C são os nucleotídeos-padrão do RNA; R representa uma purina (A ou G); e Y representa uma pirimidina (C ou U). O A destacado em vermelho forma o ponto de forquilha do laço produzido pelo *splicing* (ver Figura 6-25). Somente o GU do início do íntron e o AG do seu final são nucleotídeos invariantes nas seqüências consenso do *splicing*. As posições restantes podem ser ocupadas por uma gama de nucleotídeos, embora os nucleotídeos indicados sejam preferenciais. As distâncias no RNA entre as três seqüências consenso de *splicing* são altamente variáveis; entretanto, a distância entre o ponto de forquilha e a junção 3' do *splicing* é caracteristicamente muito menor do que a distância entre a junção 5' do *splicing* e o ponto de forquilha.

da forquilha e da junção de processamento 3' é realizado principalmente por meio do pareamento de bases entre os snRNAs e as sequências consenso de RNA no pré-mRNA substrato.

O spliceossomo é uma máquina complexa e dinâmica. Quando estudado *in vitro* foi possível determinar que alguns poucos componentes do spliceossomo se organizam sobre o pré-mRNA e, conforme a reação de *splicing* ocorre, novos componentes são incorporados e substituem aqueles que já desempenharam suas funções e foram ejetados (Figura 6-28). No entanto, muitos cientistas acreditam que, no interior da célula, o spliceossomo já se encontra sob a forma de um agrupamento de baixa afinidade – capturando, realizando o *splicing* e liberando o RNA, sob a forma de uma unidade coordenada que sofre extensos rearranjos a cada momento em que realiza o *splicing*.

### O spliceossomo usa hidrólise de ATP para produzir uma série complexa de rearranjos RNA-RNA

A hidrólise de ATP não é, em si, necessária para a química do *splicing* do RNA, uma vez que as duas reações de transesterificação preservam as ligações fosfato de alta energia.



**Figura 6-28 Mecanismo de *splicing* do pré-mRNA.** O *splicing* do RNA é catalisado por um arranjo de snRNPs (mostrado como círculos coloridos) e de outras proteínas (a maioria das quais não é mostrada) que, em conjunto, constituem o spliceossomo. O spliceossomo reconhece os sinais de *splicing* em uma molécula de pré-mRNA, aproxima as duas extremidades dos introns e fornece a atividade enzimática para as duas etapas necessárias da reação (ver Figura 6-25A e Animação 6.5). Como indicado, um conjunto de proteínas do chamado complexo de junção do éxon (EJC) permanece sobre a molécula de mRNA que sofreu *splicing*; o seu papel subsequente será discutido brevemente.

No entanto, a hidrólise extensiva do ATP é necessária para a associação e os rearranjos do spliceossomo. Algumas das proteínas adicionais que fazem parte do spliceossomo utilizam a energia da hidrólise de ATP para romper interações RNA-RNA já existentes e permitir a formação de outras. Ao todo, em torno de 200 proteínas, incluindo aquelas que formam as snRNPs, são necessárias para cada evento de *splicing*.

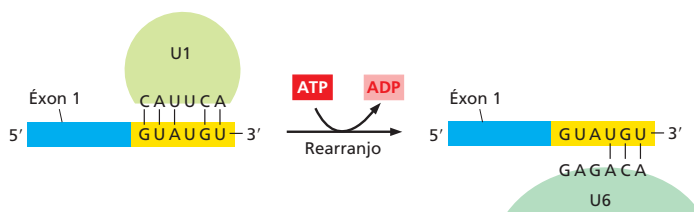
Qual é o propósito desses rearranjos? Em primeiro lugar, eles permitem que os sinais de *splicing* do pré-RNA sejam examinados pelos snRNPs diversas vezes durante o processo de *splicing*. Por exemplo, U1 snRNP inicialmente reconhece o sítio de *splicing* 5' por meio do pareamento de bases convencional; conforme o *splicing* procede, esses pares de bases são rompidos (pelo uso da energia da hidrólise do ATP) e U1 é substituído por U6 (Figura 6-29). Esse tipo de rearranjo de RNA-RNA (em que a formação de uma interação de RNA-RNA requer a ruptura de outra) ocorre várias vezes durante o *splicing* e permite que os spliceossomos verifiquem e reavaliem os sinais de *splicing*, aumentando, desse modo, a precisão global do *splicing*. Em segundo lugar, os rearranjos que ocorrem no spliceossomo criam os sítios ativos para as duas reações de transesterificação. Esses dois sítios ativos são criados um após o outro e somente após os sinais de *splicing* no pré-mRNA terem sido verificados várias vezes. Essa progressão ordenada garante que erros de *splicing* ocorrerão apenas raramente.

Uma das características mais surpreendentes do spliceossomo é a natureza dos sítios catalíticos: eles são formados por proteínas e moléculas de RNA, embora sejam as moléculas de RNA as reais catalisadoras químicas do *splicing*. Na última seção deste capítulo, discutiremos em termos gerais as propriedades químicas e estruturais das moléculas de RNA que lhes permitem atuar como catalisadoras.

Uma vez que a reação química do *splicing* foi completada, as snRNPs permanecem ligadas à alça em laço. A dissociação das snRNPs do laço (e umas das outras) requer outra série de rearranjos RNA-RNA que requer a hidrólise de ATP. Dessa forma, os snRNAs retornarão à sua configuração original, podendo ser reutilizados em uma nova reação. Ao final de uma reação de *splicing*, o spliceossomo promove a ligação de um conjunto de proteínas ao mRNA, próximo à região anteriormente ocupada pelo íntron. Denominadas *complexo de junção do éxon* (EJC, *exon junction complex*), essas proteínas marcam o sítio de um evento de *splicing* que ocorreu com sucesso e, como veremos adiante neste capítulo, influenciam o destino subsequente do mRNA.

## Outras propriedades do pré-mRNA e da sua síntese ajudam a explicar a escolha dos sítios adequados de *splicing*

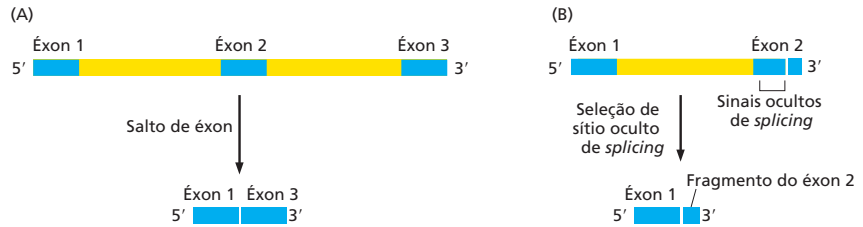
Como vimos, os íntrons variam incrivelmente em tamanho, alguns excedendo os 100 mil nucleotídeos. Se a seleção do sítio do *splicing* fosse determinada unicamente pela ação de snRNPs sobre uma molécula de RNA pré-formada livre de proteínas, poderíamos esperar erros frequentes de *splicing* – como a perda de éxons e o uso de regiões “ocultas” de *splicing* (Figura 6-30). Os mecanismos de fidelidade construídos em torno do spliceossomo para suprimir erros são suplementados por dois fatores adicionais que aumentam a chance de o *splicing* ocorrer de forma exata. O primeiro é uma simples consequência do acoplamento entre o *splicing* e a transcrição. Enquanto a transcrição ocorre, a cauda fosforilada da RNA-polimerase carrega vários componentes do spliceossomo (ver Figura 6-22), e esses componentes são diretamente transferidos da polimerase para o RNA à medida que o RNA emerge da polimerase. Essa estratégia ajuda a célula a identificar e controlar os íntrons e os éxons: por exemplo, as snRNPs que se organizam sobre um sítio de *splicing* 5' são inicialmente apresentados apenas ao sítio de *splicing* 3' que emerge imediatamente da polimerase; os



**Figura 6-29** Um dos vários rearranjos que ocorrem no spliceossomo durante o *splicing* do pré-mRNA. Este exemplo vem da levedura *Saccharomyces cerevisiae*, na qual as sequências nucleotídicas envolvidas são ligeiramente diferentes daquelas presentes em células humanas. A troca de snRNP U1 por snRNP U6 ocorre imediatamente antes da primeira reação de transferência de fosforil (ver Figura 6-28). Essa troca faz o sítio de *splicing* 5' ser lido por duas snRNPs diferentes, o que aumenta a exatidão da seleção do sítio de *splicing* 5' pelo spliceossomo.



**Figura 6-30** Dois tipos de erros de *splicing*. (A) Salto (ou perda) de éxon. (B) Seleção de sítio oculto de *splicing*. Sinais ocultos de *splicing* são sequências nucleotídicas de RNA com alta similaridade a sinais de *splicing* verdadeiros e são por vezes erroneamente utilizadas pelo spliceossomo.



sítios potenciais mais a jusante ainda não foram sintetizados. Essa coordenação entre transcrição e *splicing* é extremamente importante para evitar a retirada inadequada de exons.

Uma estratégia denominada “definição de éxon” também ajuda que as células escolham as regiões adequadas para *splicing*. O tamanho dos exons tende a ser muito mais uniforme que o tamanho dos íntrons, apresentando, em média, 150 pares de nucleotídeos em uma ampla variedade de organismos eucarióticos (Figura 6-31). Pela definição de um éxon, a maquinaria de *splicing* pode procurar sequências exônicas de tamanho relativamente homogêneo. Acredita-se que, à medida que ocorre a síntese de RNA, um grupo de componentes adicionais (principalmente as proteínas SR, assim chamadas por conterem um domínio rico em serinas e em argininas) se associa sobre os exons e marca cada sítio 3' e 5' do *splicing*, iniciando na extremidade 5' do RNA (Figura 6-32). Essas proteínas, por sua vez, recrutam o snRNA U1, que marca o limite 3' do éxon, e o snRNA U2, que ajuda a identificar o outro limite. Ao marcar dessa forma especificamente os exons e assim tirar partido do tamanho relativamente uniforme dos exons, a célula aumenta a precisão com que deposita os componentes iniciais de *splicing* sobre o RNA nascente e, portanto, evita “a perda” de sítios de *splicing*. Ainda não está claro como as proteínas SR discriminam as sequências exônicas das sequências intrônicas; entretanto, sabe-se que algumas das proteínas SR ligam-se preferencialmente a sequências específicas de RNA em exons, denominadas *estimuladores de splicing* (*splicing enhancers*). Em princípio, considerando-se a redundância do código genético, existe liberdade para a evolução da sequência nucleotídica dos exons de modo a formar um sítio de ligação para uma proteína SR, sem que a sequência de aminoácidos que o éxon codifica seja necessariamente afetada.

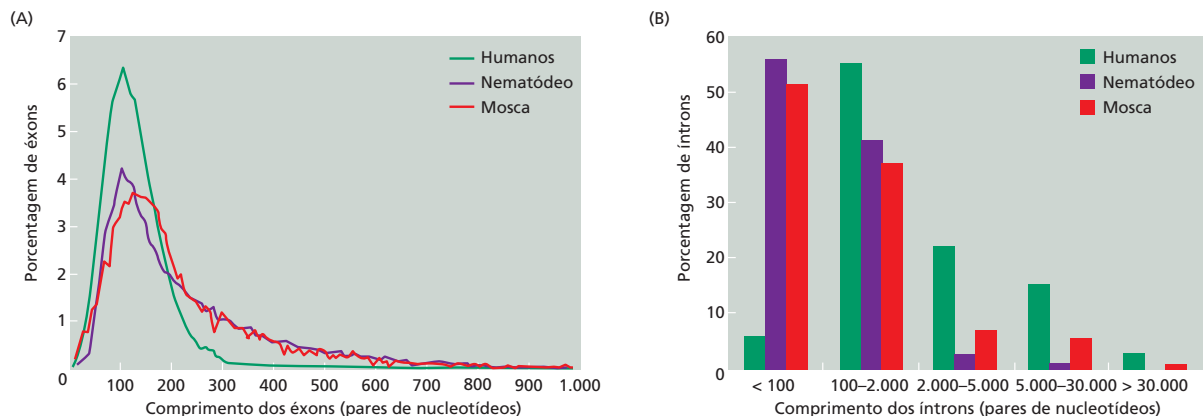
Tanto a marcação dos limites entre exons e íntrons quanto a organização do spliceossomo têm início sobre a molécula de RNA enquanto ela ainda está na fase de extensão de sua extremidade 3' pela RNA-polimerase. Entretanto, o verdadeiro processo químico do *splicing* pode ocorrer mais tarde. Esse atraso significa que os íntrons não são necessariamente removidos da molécula de pré-mRNA na ordem em que eles se encontram ao longo da cadeia de RNA.

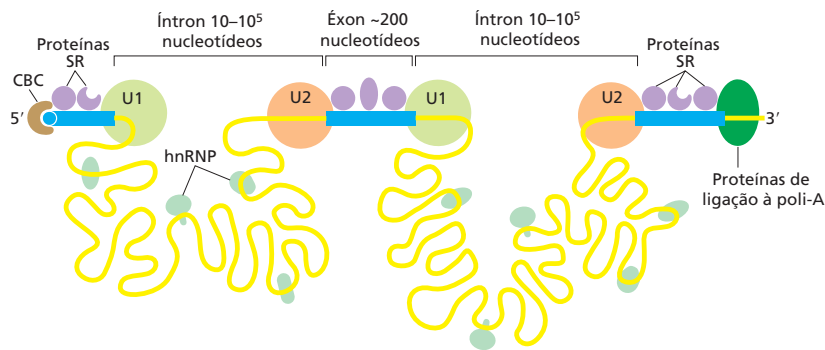
**Figura 6-31** Variação no tamanho de íntrons e de exons nos genomas humano, de nematódeo e de mosca.

(A) Distribuição do tamanho de exons. (B) Distribuição do tamanho de íntrons. Observe que o comprimento dos exons é muito mais uniforme do que o comprimento dos íntrons. (Adaptada de International Human Genome Sequencing Consortium, *Nature* 409:860–921, 2001. Com permissão de Macmillan Publishers Ltd.)

## A estrutura da cromatina afeta o *splicing* do RNA

Embora à primeira vista possa parecer contraditório, a forma como um gene é empacotado na cromatina pode afetar como o RNA transcrito desse gene será processado. Os nucleossomos tendem a ser posicionados sobre os exons (que têm, em média, o comprimento





**Figura 6-32 Hipótese da definição de éxon.** De acordo com essa ideia, proteínas SR ligam-se a cada sequência de éxon no pré-mRNA e, dessa forma, ajudam a guiar as snRNPs para os limites adequados entre íntron/éxon. Essa demarcação de exons pelas proteínas SR ocorre cotranscricionalmente, iniciando no CBC (complexo de ligação ao quepe) na extremidade 5'. Acredita-se que um grupo de proteínas conhecidas como ribonucleoproteínas nucleares heterogêneas (hnRNPs) possam associar-se preferencialmente aos íntrons ajudando ainda mais o spliceossomo a distinguir os íntrons dos exons. (Adaptada de R. Reed, *Curr. Opin. Cell Biol.* 12:340–345, 2000. Com permissão de Elsevier.)

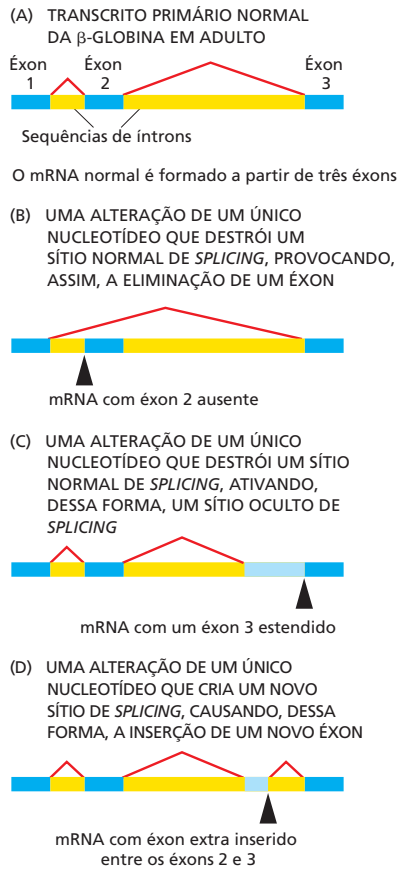
aproximado do DNA em um nucleossomo), e foi proposto que os nucleossomos atuam como “reduzidores de velocidade”, permitindo que as proteínas responsáveis pela identificação dos exons se associem sobre o RNA conforme ele emerge da polimerase. Além disso, alterações na estrutura da cromatina são usadas para alterar os padrões de *splicing* de duas maneiras. Primeiro, visto que o *splicing* e a transcrição são acoplados, a velocidade na qual a RNA-polimerase se move ao longo do DNA pode afetar o *splicing* do RNA. Por exemplo, se a polimerase se move lentamente, o salto de exons (ver Figura 6-30A) será minimizado: a organização do spliceossomo inicial poderá ser concluída antes mesmo que um sítio alternativo de *splicing* emergja da RNA-polimerase. Os nucleossomos na cromatina condensada podem pausar a polimerase; o padrão das pausas, por sua vez, afetará a quantidade de RNA disponível para a maquinaria do *splicing* em um dado momento.

Há uma segunda e mais direta forma da estrutura da cromatina afetar o *splicing* do RNA. Embora os detalhes ainda não estejam totalmente compreendidos, modificações específicas em histonas atraem componentes do spliceossomo, e, visto que a cromatina que está sendo transcrita está em forte associação com o RNA nascente, esses componentes do *splicing* podem facilmente ser transferidos para o RNA emergente. Dessa forma, certos tipos de modificações de histonas podem afetar o padrão final do *splicing*.

### O *splicing* de RNA possui uma plasticidade extraordinária

Vimos que a escolha dos sítios de *splicing* depende de características do transcrito pré-mRNA como a força dos três sinais no RNA (as junções 5' e 3' do *splicing* e o ponto de forquilha) pela maquinaria de *splicing*, a associação cotranscricional do spliceossomo, a estrutura da cromatina e a eficiência do “sistema de classificação” que define o éxon. Desconhecemos quão exato é o processo normal do *splicing*, pois, como veremos a seguir, existem vários sistemas de controle de qualidade que degradam rapidamente os mRNAs cujo *splicing* ocorreu de forma inadequada. No entanto, sabemos que, comparado a outras etapas da expressão gênica, o *splicing* geralmente é bastante flexível.

Assim, por exemplo, quando uma mutação ocorre em uma sequência nucleotídica crítica para o *splicing* de um íntron determinado, isso por si só não impedirá necessariamente o *splicing* deste íntron. Em vez disso, a mutação potencialmente criará um novo padrão de *splicing* (Figura 6-33). Com frequência, um éxon é simplesmente eliminado (ver Figura 6-33B). Em outros casos, a mutação leva ao uso eficiente de uma junção oculta (Figura 6-33C). Aparentemente, a maquinaria de *splicing* evoluiu para identificar o melhor padrão possível das junções de *splicing*, e, se este é alterado por uma mutação, ela procura o próximo melhor padrão, e assim por diante. Essa plasticidade inerente ao processo de *splicing* do RNA sugere que as modificações nos padrões de *splicing* provocadas por mutações ao acaso são importantes na evolução dos genes e dos organismos. Isso também significa que as mutações que afetam o *splicing* podem ser extremamente prejudiciais para o organismo: além da  $\beta$ -talassemia, o exemplo apresentado na Figura 6-33, o *splicing* aberrante desempenha papéis importantes no desenvolvimento da fibrose cística, da demência fronto-temporal, da doença de Parkinson, da retinite pigmentosa, da atrofia muscular espinal, da distrofia miotônica, do envelhecimento precoce e do câncer. Estima-se que, das muitas mutações pontuais que causam doenças hereditárias humanas, 10% produzam *splicing* aberrante do gene que contém a mutação.



**Figura 6-33** Processamento anormal do transcrito primário do RNA da  $\beta$ -globina em humanos com a doença  $\beta$ -talassemia. Nos exemplos ilustrados, a doença (uma anemia grave resultante da síntese anormal de hemoglobina) é causada por mutações em sítios de *splicing*, identificadas no genoma de pacientes afetados. Os segmentos azul-escuro representam as três sequências normais de éxons; as linhas vermelhas conectam os sítios de *splicing* 5' e 3' que são utilizados. Em (B), (C) e (D), os segmentos azul-claro indicam as novas sequências nucleotídicas inseridas na molécula de mRNA final como resultado de uma mutação indicada pela seta preta. Observe que quando uma mutação origina um sítio normal de *splicing* sem um acompanhante, um éxon é perdido (B), ou um ou mais sítios de *splicing* ocultos anormais adjacentes são usados (C). (Parcialmente adaptada de S.H. Orkin, em *The Molecular Basis of Blood Diseases* [G. Stamatoyannopoulos et al., eds.], pp. 106–126. Philadelphia: Saunders, 1987.)

A plasticidade do *splicing* de RNA também significa que a célula pode facilmente regular o padrão de *splicing* do RNA. Anteriormente, nesta seção, vimos que o *splicing* alternativo pode originar proteínas diferentes a partir do mesmo gene e que essa é uma estratégia comum para aumentar o potencial de codificação dos genomas. Alguns exemplos de *splicing* alternativo são constitutivos; isto é, os mRNAs que sofrem *splicing* alternativo são produzidos continuamente pelas células de um organismo. Entretanto, em muitos casos, os padrões de *splicing* são regulados pela célula de tal forma que diferentes proteínas são produzidas em diferentes momentos e em diferentes tecidos (ver Figura 6-26). No Capítulo 7, retornaremos a esse assunto para discutir alguns exemplos específicos da regulação de *splicing* do RNA.

### O *splicing* do RNA catalisado pelo spliceossomo provavelmente evoluiu a partir de mecanismos de auto-*splicing*

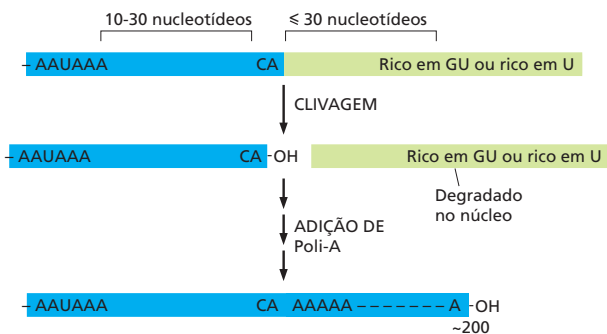
A descoberta do spliceossomo intrigou bastante os biólogos moleculares. Por que moléculas de RNA, em vez de proteínas, desempenhariam funções importantes no reconhecimento do sítio de *splicing* e nas reações químicas do *splicing*? Por que é utilizado um laço intermediário em vez de se utilizar uma alternativa aparentemente mais simples, que consistiria na aproximação dos sítios 5' e 3' do *splicing* em uma única etapa, seguida pela clivagem direta e união dos segmentos? As respostas a essas questões refletem a evolução do spliceossomo.

Como discutido brevemente no Capítulo 1 (e retomado com mais detalhes na seção final deste capítulo), é provável que as células primordiais tenham utilizado moléculas de RNA – em vez de proteínas – como seus principais catalisadores e que elas armazenassem sua informação genética sob a forma de sequências de RNA, em vez de DNA. As reações de *splicing* catalisadas por RNA possivelmente desempenharam funções essenciais nessas células primordiais. Como evidência, alguns íntrons de RNA de auto-*splicing* (ou seja, íntrons no RNA que sofrem *splicing* na ausência de proteínas ou de quaisquer outras moléculas de RNA) ainda existem – por exemplo, nos genes do rRNA nuclear do ciliado *Tetrahymena*, em uns poucos genes do bacteriófago T4 e em alguns genes mitocondriais e plastidiais. Nesses casos, a molécula de RNA adquire uma estrutura tridimensional enovelada específica que aproxima as junções íntron/éxon e catalisa as duas reações de transesterificação. Uma sequência de íntron que sofre auto-*splicing* pode ser identificada *in vitro* por meio da incubação de uma molécula de RNA purificada que contém a sequência do íntron e a observação da reação de *splicing*. Visto que a química básica de algumas reações de autoprocessamento é muito semelhante ao *splicing* do pré-mRNA, foi proposto que o complexo processo do *splicing* do pré-mRNA tenha evoluído de uma forma ancestral mais simples, a partir do auto-*splicing* do RNA.

### As enzimas de processamento do RNA geram a extremidade 3' dos mRNAs de eucariotos

Já vimos que a extremidade 5' de um pré-mRNA produzido pela RNA-polimerase II é capeada quase que imediatamente após a sua saída da RNA-polimerase. A seguir, à medida que a polimerase continua seu movimento ao longo de um gene, os componentes do spliceossomo se associam sobre o RNA e identificam os limites dos éxons e dos íntrons. A longa cauda C-terminal da RNA-polimerase coordena esses processos pela transferência de componentes do capeamento e do *splicing* diretamente para o RNA conforme ele emerge da enzima. Nesta seção, veremos que à medida que a RNA-polimerase II se aproxima do final de um gene, um mecanismo similar assegura que a extremidade 3' do pré-mRNA seja corretamente processada.

A posição da extremidade 3' de cada molécula de mRNA é, em última análise, especificada por sinais codificados no genoma (Figura 6-34). Esses sinais são transcritos em RNA, conforme a RNA-polimerase II se move ao longo deles, sendo, então, reconhecidos (como RNA) por uma série de proteínas de ligação ao RNA e por enzimas do processamento do RNA (Figura 6-35). Duas proteínas de subunidades múltiplas, denominadas fator de estimulação da clivagem (CstF, *cleavage stimulation factor*) e fator de especificidade de clivagem e poliadenilação (CPSE, *cleavage and polyadenylation specificity factor*), possuem importância especial. Ambas movimentam-se com a cauda da



RNA-polimerase e são transferidas à extremidade 3' da sequência em processamento de uma molécula de RNA logo que ela emerge da RNA-polimerase.

Após a ligação de CstF e CPSF às suas sequências de reconhecimento na molécula de RNA emergente, proteínas adicionais se ligam a eles para criar a extremidade 3' do mRNA. Inicialmente, o RNA é clivado da polimerase (ver Figura 6-35). Após, uma enzima denominada poli-A-polimerase (PAP) adiciona, um a um, aproximadamente 200 nucleotídeos A à extremidade 3' produzida pela clivagem. O nucleotídeo precursor dessas adições é o ATP, e o mesmo tipo de ligações 5' a 3' utilizado na síntese convencional de RNA é formado nessa situação. No entanto, diferentemente das outras RNA-polimerases comuns, a poli-A-polimerase não requer um molde e, como consequência, a cauda de poli-A dos mRNAs eucarióticos não está diretamente codificada no genoma. Conforme a cauda de poli-A é sintetizada, as denominadas proteínas de ligação à poli-A se ligam a ela e, por um mecanismo ainda pouco conhecido, ajudam a determinar o tamanho final da cauda.

Após a clivagem da extremidade 3' de uma molécula de pré-mRNA eucariótica, a RNA-polimerase II continua a transcrever, em alguns casos até várias centenas de nucleotídeos. Uma vez que a clivagem da extremidade 3' tenha ocorrido, o RNA recém-sintetizado que emerge das polimerases não apresenta um quepe 5'; esse RNA desprotegido é rapidamente degradado por uma exonuclease 5' → 3' presente na cauda da polimerase. Aparentemente, é essa degradação continuada do RNA que eventualmente permite à RNA polimerase se dissociar do molde e interromper a transcrição.

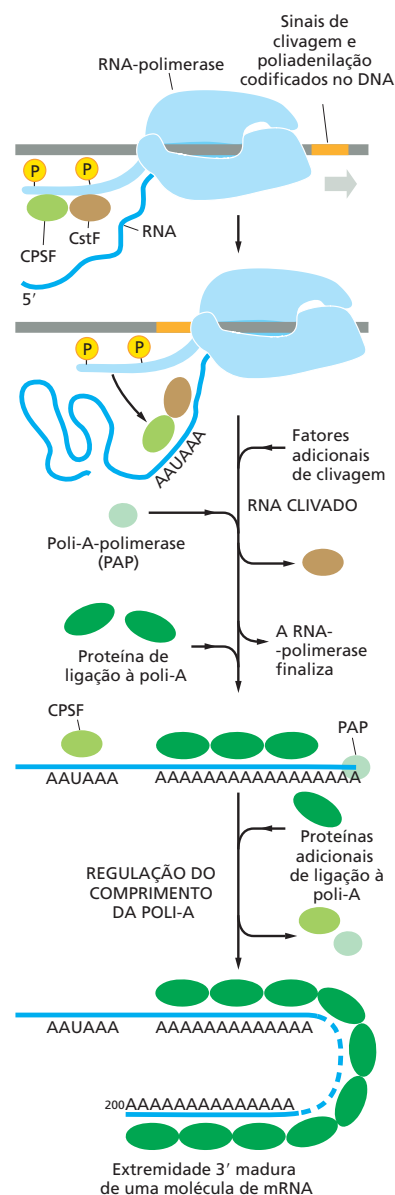
### mRNAs eucarióticos maduros são seletivamente exportados do núcleo

A síntese e o processamento do pré-mRNA eucariótico ocorrem de forma ordenada no interior do núcleo da célula. Porém, de todo o pré-mRNA que é sintetizado, somente uma pequena fração – o mRNA maduro – será utilizada posteriormente pela célula. A maior parte do material restante – íntrons excisados, RNAs clivados e formas de pré-mRNAs que sofreram *splicing* anormal – não somente não será utilizada como é potencialmente perigosa. Como a célula consegue distinguir as moléculas de mRNA maduras relativamente raras, que ela deseja manter, da esmagadora quantidade de detritos gerados pelo processamento de RNA?

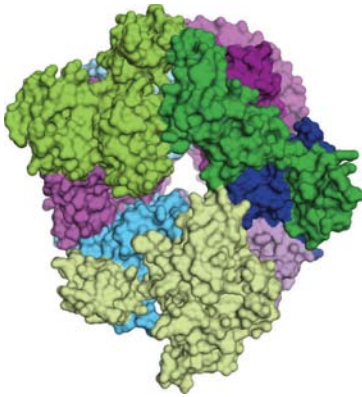
A resposta é que, quando uma molécula de RNA é processada, ela perde certas proteínas e adquire outras. Por exemplo, vimos que a ligação dos complexos de ligação ao quepe, dos complexos de junção do éxon e das proteínas de ligação à poli-A marcam a finalização respectiva do capeamento, do *splicing* e da adição da poli-A. Uma molécula de mRNA adequadamente completa também pode ser distinguida pela presença/ausência de determinadas proteínas. Por exemplo, a presença de uma proteína snRNP significará um *splicing* incompleto ou aberrante. Apenas quando as proteínas presentes sobre uma molécula de mRNA coletivamente indicarem que o processamento foi adequadamente finalizado é que o mRNA será exportado do núcleo rumo ao citosol, onde poderá ser traduzido em proteína.

**Figura 6-35** Algumas das principais etapas na geração da extremidade 3' de um mRNA eucariótico. Esse processo é muito mais complicado do que o processo análogo em bactérias, onde a RNA-polimerase simplesmente para no sinal de terminação e libera tanto a extremidade 3' do transcrito quanto o DNA-molde (ver Figura 6-11).

**Figura 6-34** Sequências nucleotídicas consenso que determinam a clivagem e a poliadenilação para formar a extremidade 3' de um mRNA eucariótico. Essas sequências são codificadas no genoma e reconhecidas – como RNA – por proteínas específicas após serem transcritas. Como mostrado na Figura 6-35, o hexâmero AAUAAA liga-se ao CPSF, e o elemento rico em GU à direita do sítio de clivagem liga-se ao CstF; a sequência CA é ligada a um terceiro fator proteico necessário para a clivagem. Como outras sequências consenso de nucleotídeos discutidas neste capítulo (ver Figura 6-12), as sequências ilustradas na figura representam uma variante de uma ampla gama de sinais de poliadenilação e de clivagem individuais.







**Figura 6-36** Estrutura da região central do exossomo de RNA humano. O RNA penetra por uma extremidade do poro central e é degradado por RNases que se associam com a outra extremidade. Nove diferentes subunidades proteicas (cada qual representada por uma cor diferente) compõem essa grande estrutura em anel. As células eucarióticas possuem tanto um exossomo nuclear quanto um exossomo citoplasmático; ambas as formas incluem a porca central do exossomo mostrado aqui e subunidades adicionais (incluindo RNases especializadas) que diferenciam as duas formas. O exossomo nuclear degrada RNAs aberrantes antes que eles sejam exportados para o citosol. Ele também processa certos tipos de RNA (p. ex., os RNAs ribossômicos) para produzir a sua forma final. A forma citoplasmática do exossomo é responsável pela degradação de mRNAs no citosol, e é, portanto, essencial para a determinação do tempo de vida de cada molécula de mRNA. (Código PDB: 2NN6.)

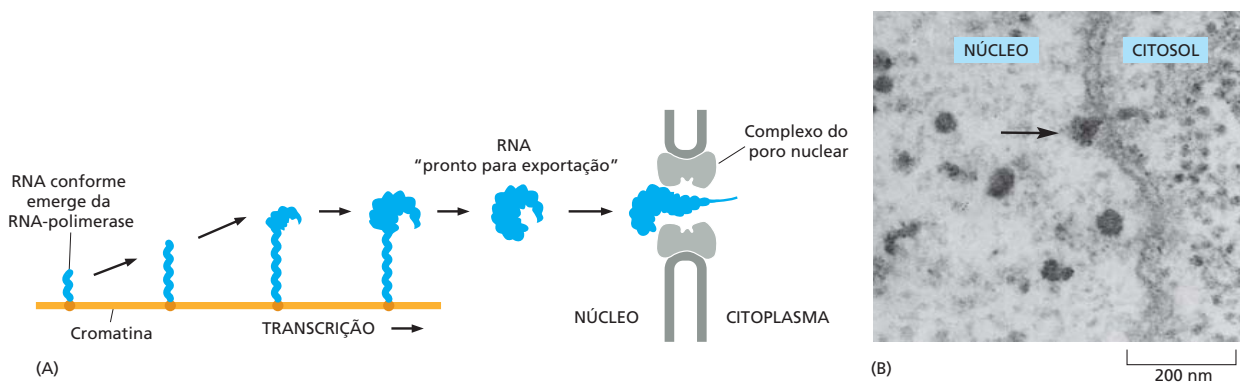
Os mRNAs processados de forma inadequada e outros resíduos de RNA (p. ex., seqüências intrônicas excisadas) são retidos no núcleo, onde eventualmente serão degradados pelo **exossomo nuclear**, um grande complexo proteico cujo interior é rico em exonucleases de RNA 3'-5' (**Figura 6-36**). Assim, as células eucarióticas exportam apenas moléculas de RNA úteis para o citoplasma, enquanto fragmentos de RNA são eliminados no núcleo.

Entre todas as proteínas que se agregam às moléculas de pré-mRNA conforme elas emergem das RNA-polimerases que estão transcrevendo, as mais abundantes são as proteínas ribonucleares nucleares heterogêneas (hnRNPs, *heterogeneous nuclear ribonuclear proteins*). Algumas dessas proteínas (existem aproximadamente 30 diferentes em humanos) desenrolam as hélices em grampo no RNA de tal forma que os sinais de *splicing* e outros sinais no RNA podem ser lidos mais facilmente. Outras empacotam preferencialmente o RNA contido nas seqüências de íntrons extremamente longos, típicos de organismos complexos (ver Figura 6-31) e podem desempenhar um papel importante na distinção entre mRNAs maduros e restos do processamento de RNA.

Os mRNAs adequadamente processados são transportados através dos **complexos do poro nuclear** (NPCs, *nuclear pore complexes*), canais aquosos da membrana nuclear, os quais conectam diretamente o nucleoplasma e o citosol (**Figura 6-37**). Pequenas moléculas (com menos de 60 mil daltons) podem difundir-se livremente através desses canais. No entanto, a maioria das macromoléculas celulares, inclusive os mRNAs complexados a proteínas, apresenta tamanho excessivo, o que as impossibilita de atravessar os canais sem o uso de processos especiais. A célula usa energia para o transporte ativo dessas macromoléculas em ambos os sentidos através dos complexos do poro nuclear.

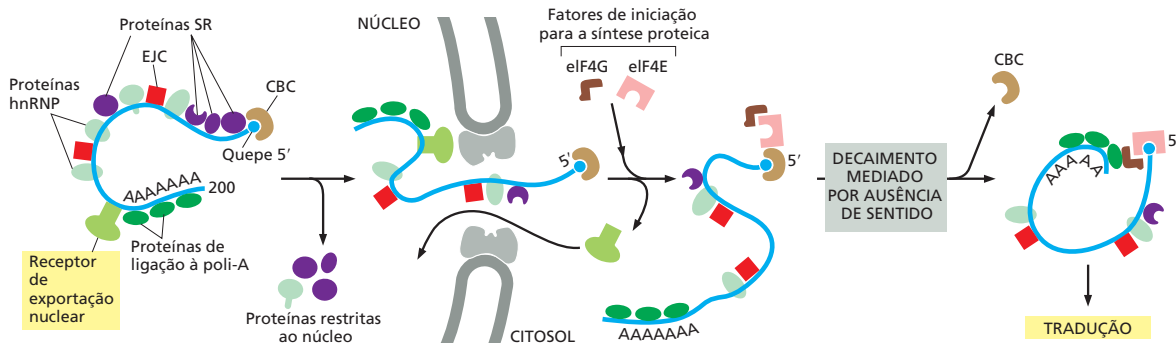
Como explicado em detalhes no Capítulo 12, as macromoléculas são transportadas através dos complexos do poro nuclear via *receptores de transporte nuclear*, os quais, dependendo da identidade da macromolécula, as escoltam do núcleo para o citoplasma ou vice-versa. Para que ocorra a exportação do mRNA, um receptor de transporte nuclear específico deve ser ligado ao mRNA, uma etapa que, em muitos organismos, ocorre simultaneamente à clivagem e poliadenilação 3'. Após ter ajudado a mover uma molécula de RNA através do complexo do poro nuclear, o receptor de transporte se dissocia do mRNA, penetra novamente no núcleo, e é utilizado para exportar uma nova molécula de mRNA.

A exportação dos complexos mRNA-proteína a partir do núcleo pode ser facilmente observada ao microscópio eletrônico para os incomumente abundantes mRNAs dos *genes do Anel de Balbiani* de insetos. Conforme esses genes são transcritos, pode-se observar o empacotamento do RNA recém-formado mediado por proteínas, como hnRNPs, proteínas SR e componentes do spliceossomo. Esse complexo proteína-RNA sofre uma série de transições estruturais, provavelmente refletindo eventos de processamento do RNA, culminando em uma fibra curva (**Figura 6-37**). Essa fibra curva move-se, então, através do nucleoplasma, penetra o complexo do poro nuclear (sendo seu quepe 5' a primeira porção a penetrar) e sofre outra série de transições estruturais, enquanto se move através do poro. Essas e outras observações revelaram que os complexos pré-mRNA-proteína e mRNA-pro-



**Figura 6-37** Transporte de uma grande molécula de mRNA pelo complexo do poro nuclear. (A) A maturação de uma molécula de mRNA conforme ela é sintetizada pela RNA-polimerase e empacotada pelas diversas proteínas nucleares. Esta ilustração de um RNA de inseto incomumente grande e abundante, chamado mRNA do anel de Balbiani, baseia-se em fotomicrografias de microscopia eletrônica como as mostradas em (B). (A, adaptada de B. Daneholt, *Cell* 88:585-588, 1997. Com permissão de Elsevier; B, de B.J. Stevens e H. Swift, *J. Cell Biol.* 31:55-77, 1966. Com permissão de The Rockefeller University Press.)





teína são estruturas dinâmicas que podem adquirir e perder numerosas proteínas específicas durante a síntese, o processamento e a exportação do RNA (Figura 6-38).

A análise anteriormente descrita foi complementada por novas metodologias que permitiram aos pesquisadores acompanhar o destino de moléculas de mRNA mais típicas, que podem ser marcadas com fluorescência e observadas individualmente. Uma típica molécula de RNA é liberada do seu sítio de transcrição e consome vários minutos difundindo rumo a um complexo do poro nuclear. Durante esse período, é provável que eventos de processamento do RNA continuem a ocorrer e que o RNA perca certas proteínas anteriormente a ele ligadas e adquira novas. Ao chegar à entrada do poro, o mRNA “pronto para exportação” permanece estacionário durante alguns segundos, tempo no qual pode ocorrer a conclusão do processamento, e, em seguida, é transportado através do poro muito rapidamente, em dezenas de milissegundos. Alguns complexos mRNA-proteínas são muito grandes, e os mecanismos que lhes permitem se mover tão rapidamente através dos complexos de poros nucleares permanecem ainda um mistério.

Algumas das proteínas depositadas sobre o mRNA enquanto ele ainda se encontra no núcleo podem afetar o destino do RNA após ele ter sido transportado para o citosol. Assim, a estabilidade de um mRNA no citosol, a eficiência sob a qual ele será traduzido em proteína e seu destino definitivo no citosol podem ser determinados pelas proteínas adquiridas no núcleo e que permanecem ligadas ao RNA após ele deixar o compartimento nuclear.

Antes de discutirmos o que acontece com os mRNAs no citosol, consideraremos brevemente como ocorre a síntese e o processamento de algumas moléculas de RNA não codificadoras. Há muitos tipos de RNAs não codificadores produzidos pelas células (ver Tabela 6-1, p. 305), mas aqui vamos nos concentrar nos rRNAs, que são extremamente importantes para a tradução dos mRNAs em proteína.

## RNAs não codificadores também são sintetizados e processados no núcleo

Apenas uma pequena porcentagem do peso seco de uma célula de mamífero é RNA, sendo que apenas cerca de 3 a 5% dessa porcentagem corresponde a mRNA. A maior parte do RNA nas células desempenha funções estruturais e catalíticas (ver Tabela 6-1). Os RNAs mais abundantes nas células são os RNAs ribossômicos (rRNAs), que constituem aproximadamente 80% do RNA em uma célula que se encontra em rápida divisão. Como discutido a seguir neste capítulo, esses RNAs formam o cerne do ribossomo. Diferentemente das bactérias – nas quais todos os RNAs celulares são sintetizados por uma única RNA-polimerase –, os eucariotos têm uma polimerase especializada, a RNA-polimerase I, que se dedica à produção dos rRNAs. A RNA-polimerase I é semelhante estruturalmente à RNA-polimerase II discutida anteriormente; entretanto, a ausência de uma cauda C-terminal na polimerase I ajuda a explicar por que seus transcritos não são capeados ou poliadenilados.

Visto que múltiplos ciclos de tradução de cada molécula de mRNA podem proporcionar uma amplificação enorme na produção das moléculas proteicas, muitas das proteínas abundantes na célula podem ser sintetizadas a partir de genes que apresentam uma única cópia por genoma haploide (ver Figura 6-3). Em contraste, os componentes RNA do ribossomo constituem produtos gênicos finais, e uma célula de mamífero em crescimento deve sintetizar cerca de 10 milhões de cópias de cada tipo de rRNA em cada geração da

**Figura 6-38 Ilustração esquemática de uma molécula de mRNA pronta para exportação e seu transporte através do poro nuclear.** Como indicado, algumas proteínas acompanham o mRNA à medida que ele atravessa o poro, enquanto outras permanecem no núcleo. O receptor de exportação nuclear de mRNA é um complexo de proteínas que se liga a uma molécula de mRNA, uma vez que ela tenha sido corretamente processada e poliadenilada. Depois que o mRNA é exportado para o citosol, esse receptor de exportação se dissocia do mRNA e é transportado de volta ao núcleo, onde pode ser novamente utilizado. A verificação final aqui indicada, chamada de *decaimento mediado por ausência de sentido*, será descrita mais adiante neste capítulo.

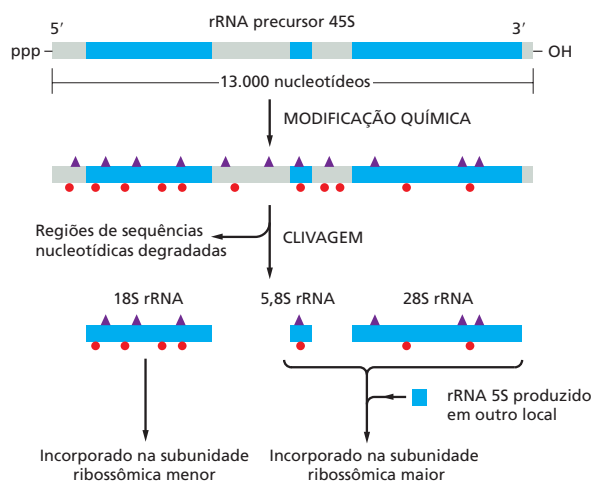
**Figura 6-39** Transcrição de genes de rRNA organizados em seqüência, como observado em microscopia eletrônica. O padrão de genes transcritos alternados a espaçadores não transcritos é facilmente observável. Uma visão de maior magnitude dos genes de rRNA é mostrada na Figura 6-10. (De V.E. Foe, *Cold Spring Harb. Symp. Quant. Biol.* 42:723–740, 1978. Com permissão de Cold Spring Harbor Laboratory Press.)



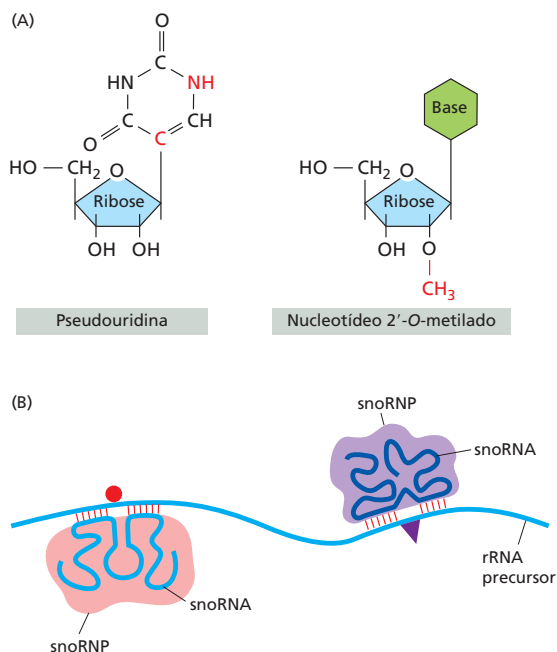
célula para construir seus 10 milhões de ribossomos. Quantidades adequadas de rRNAs só podem ser produzidas porque a célula contém múltiplas cópias de **genes de rRNA** que codificam **RNAs ribossômicas (rRNAs)**. Mesmo a *E. coli* necessita de sete cópias de seus genes de rRNA para suprir as necessidades da célula para a formação dos ribossomos. As células humanas contêm aproximadamente 200 cópias dos genes de rRNA por genoma haploide, dispersas em pequenos grupos em cinco cromossomos diferentes (ver Figura 4-11), enquanto as células da rã *Xenopus* contêm cerca de 600 cópias de genes de rRNA por genoma haploide, em um único agrupamento, em um cromossomo (**Figura 6-39**).

Existem quatro tipos de rRNAs eucarióticos, cada um representado no ribossomo por uma cópia. Três desses quatro rRNAs (18S, 5,8S e 28S) são sintetizados por meio de modificações químicas e clivagem de um único grande rRNA precursor (**Figura 6-40**); o quarto (5S RNA) é sintetizado a partir de um grupo separado de genes por uma polimerase diferente, a RNA-polimerase III, e não requer modificações químicas.

Grandes modificações químicas ocorrem no rRNA precursor de 13 mil nucleotídeos de comprimento antes que os rRNAs sejam clivados e organizados sob a forma de ribossomos. Estas incluem cerca de 100 metilações das posições 2'-OH nos açúcares dos nucleotídeos e 100 isomerizações de nucleotídeos uridina para pseudouridina (**Figura 6-41A**). As funções dessas modificações não são compreendidas em detalhe, mas elas provavelmente auxiliam no envelhecimento e na união dos rRNAs finais, ou alteram sensivelmente a função dos ribossomos. Cada modificação é feita em uma posição específica sobre o rRNA precursor, especificada por "RNAs-guia", que se posicionam no rRNA precursor pelo pareamento de bases e, assim, trazem uma enzima modificadora de RNA para a posição apropriada (**Figura 6-41B**). Outros RNAs-guia promovem a clivagem dos rRNAs precursores em rRNAs maduros, provavelmente por causarem modificações conformacionais no rRNA precursor, que expõem esses sítios para as nucleases. Todos esses RNAs-guia são membros de uma grande classe de RNAs chamada **pequenos RNAs nucleolares (snoRNAs, small**



**Figura 6-40** Modificações químicas e o processamento nucleolítico de uma molécula precursora de rRNA 45S eucariótica gerando três RNAs ribossômicos distintos. Dois tipos de modificações químicas (código de cores de acordo com a Figura 6-41) são realizados no rRNA precursor antes que ele seja clivado. Aproximadamente, metade das seqüências nucleotídicas desse rRNA precursor é descartada e degradada no núcleo pelo exossomo. Os rRNAs são denominados de acordo com seus valores "S", que se referem às suas taxas de sedimentação em ultracentrifugação. Quanto maior o valor de S, maior é o rRNA.



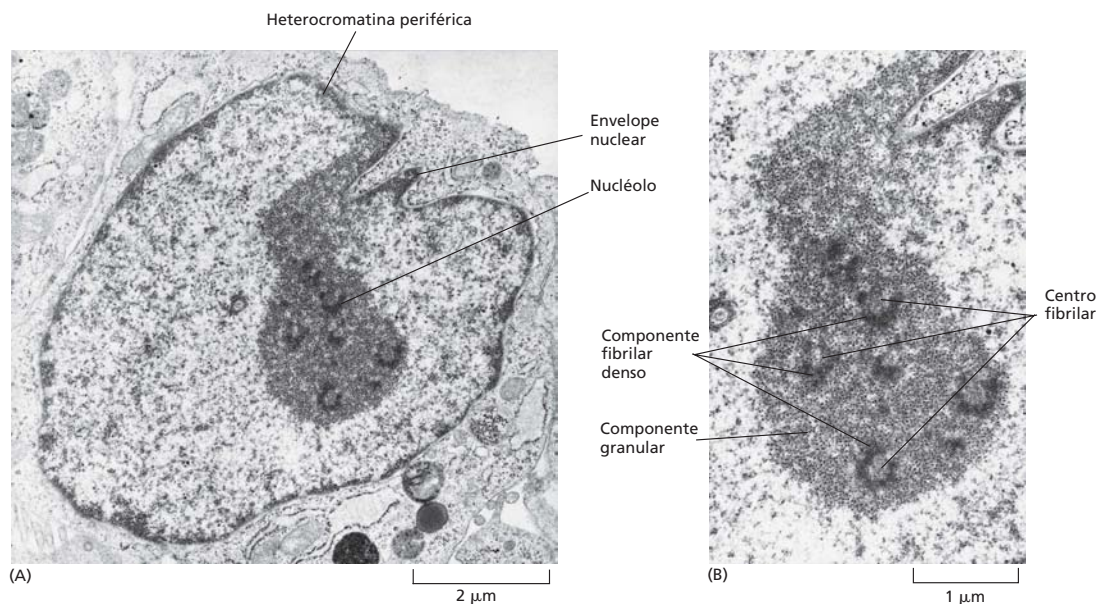
**Figura 6-41 Modificações do rRNA precursor por RNAs-guia.** (A) Duas importantes modificações covalentes feitas ao rRNA; as diferenças dos nucleotídeos inicialmente incorporados estão indicadas por átomos *vermelhos*. A pseudouridina é um isômero da uridina; a base foi “girada,” e está ligada ao C *vermelho* e não ao N *vermelho* do açúcar (compare à Figura 6-5B). (B) Conforme indicado, os snoRNAs localizam os sítios de modificação por meio do pareamento de bases com sequências complementares ao rRNA precursor. Os snoRNAs estão ligados a proteínas, e os complexos são denominados snoRNPs (pequenas ribonucleoproteínas nucleares). As snoRNPs contêm tanto as sequências-guia quanto as enzimas que modificam o rRNA.

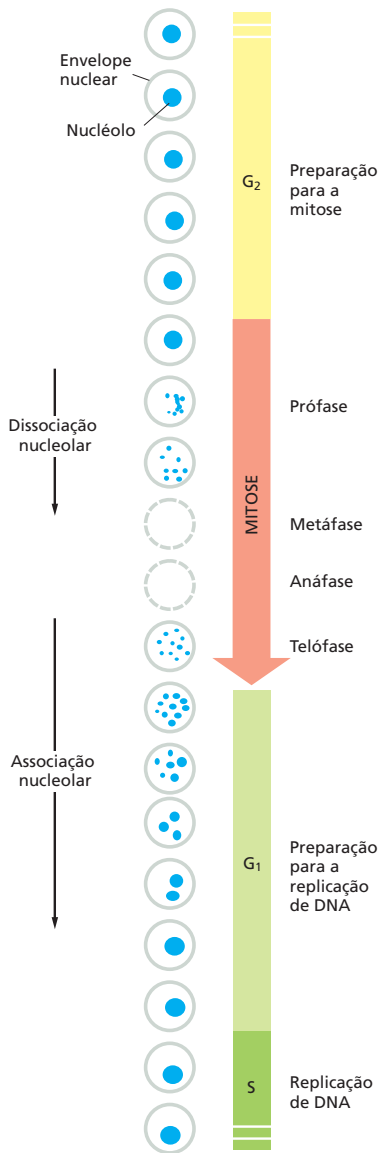
*nucleolar RNAs*), denominados dessa forma por desempenharem suas funções em um subcompartimento do núcleo – o nucléolo. Muitos snoRNAs são codificados nos íntrons de outros genes, especialmente aqueles que codificam proteínas ribossômicas. Eles são sintetizados pela RNA-polimerase II e processados a partir de sequências de íntrons excisados.

### O nucléolo é uma fábrica produtora de ribossomos

O nucléolo é a estrutura mais facilmente identificada no núcleo de uma célula eucariótica quando observada em microscopia óptica. Essa estrutura foi tão analisada pelos primeiros citologistas que uma revisão de 1898 chegou a listar aproximadamente 700 referências. Sabemos atualmente que o nucléolo é a região onde acontece o processamento de rRNAs e a sua organização sob a forma de subunidades ribossômicas. Ao contrário de muitas das principais organelas da célula, o nucléolo não está delimitado por uma membrana (**Figura 6-42**); ele é um enorme agregado de macromoléculas, incluindo os

**Figura 6-42 Fotomicrografia eletrônica de um corte fino do nucléolo de um fibroblasto humano, mostrando suas três zonas distintas.** (A) Imagem do núcleo completo. (B) Vista de maior magnitude do nucléolo. Acredita-se que a transcrição dos genes de rRNA ocorra entre o centro fibrilar e o componente denso fibrilar, e que o processamento dos rRNAs e sua organização nas duas subunidades do ribossomo ocorra fora do componente denso fibrilar, nos componentes granulares adjacentes. (Cortesia de E.G. Jordan e J. McGovern.)



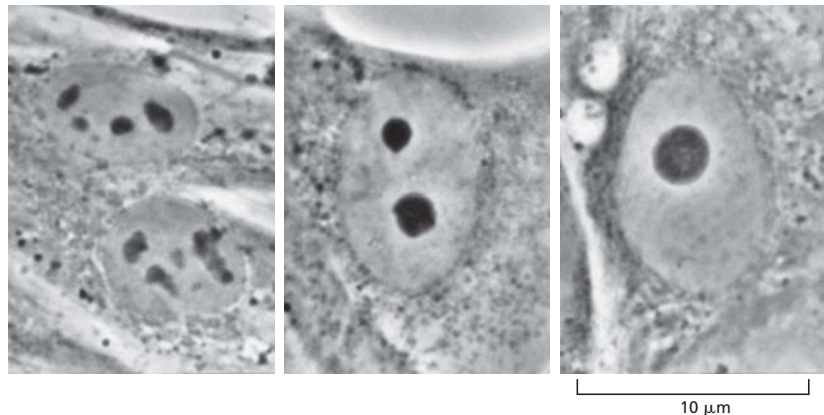


**Figura 6-43 Alterações na aparência de um nucléolo em uma célula humana durante o ciclo celular.** Apenas o núcleo da célula está representado neste diagrama. Na maioria das células eucarióticas, a membrana nuclear se desorganiza durante a mitose, como indicado pelos círculos tracejados.

próprios genes do rRNA, precursores dos rRNAs, rRNAs maduros, enzimas de processamento de rRNAs, snoRNPs, um grande conjunto de fatores de associação (incluindo ATPases, GTPases, proteínas-quinase e RNA-helicases), proteínas ribossômicas e ribossomos parcialmente formados. A íntima associação de todos esses componentes permite que a organização dos ribossomos ocorra rapidamente e sem perturbações.

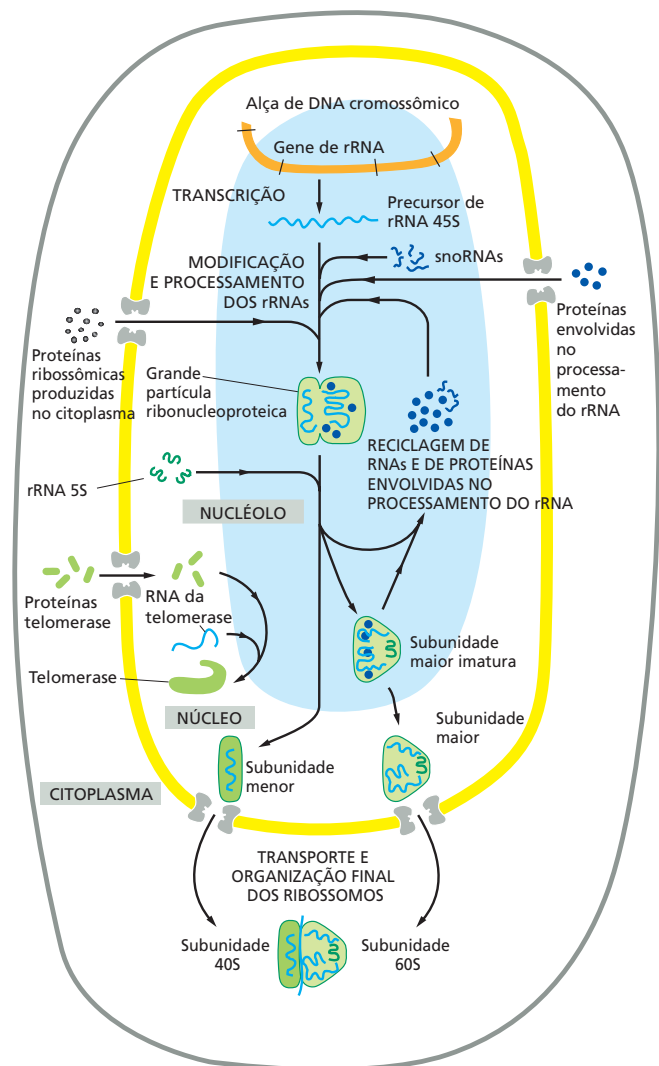
Vários tipos de moléculas de RNA desempenham um importante papel na química e na estruturação do nucléolo, sugerindo que ele possa ter evoluído a partir de uma estrutura ancestral presente em células dominadas pela catálise de RNA. Nas células atuais, os genes de rRNA desempenham uma função importante na formação dos nucléolos. Em uma célula humana diploide, os genes de rRNA estão distribuídos em dez grupos, localizados próximo à extremidade de uma das duas cópias de cinco pares cromossômicos diferentes (ver Figura 4-11). Durante a interfase, esses dez cromossomos contribuem com alças de DNA (contendo os genes rRNA) para o nucléolo; na fase M, quando os cromossomos condensam, os nucléolos se fragmentam e, então, desaparecem. Em seguida, na telófase da mitose, conforme os cromossomos retornam ao seu estado semidisperso, as extremidades dos dez cromossomos reorganizam pequenos nucléolos, que se aglutinam de modo progressivo em um único nucléolo (Figuras 6-43 e 6-44). Como deveria ser esperado, o tamanho dos nucléolos reflete o número de ribossomos que a célula está produzindo. Consequentemente, seu tamanho varia muito entre as diferentes células e pode sofrer alterações em uma única célula, ocupando 25% do volume total nuclear em células que estão produzindo quantidades anormalmente grandes de proteína.

A organização dos ribossomos é um processo complexo, e as principais características desse processo estão ilustradas na Figura 6-45. Além de seu papel central na biogênese dos ribossomos, o nucléolo é o sítio da produção de outros RNAs não codificadores e da organização de outros complexos RNA-proteína. Por exemplo, o snRNP U6, que atua no *splicing* do pré-mRNA (ver Figura 6-28), é composto por uma molécula de RNA e de, no mínimo, sete proteínas. O snRNA U6 é modificado quimicamente pelos snoRNAs no nucléolo antes de sua associação final em snRNP U6. Outros importantes complexos RNA-proteína, inclusive a telomerase (encontrada no Capítulo 5) e a partícula de reconhecimento de sinal (a qual discutiremos no Capítulo 12), também são formadas no nucléolo. Finalmente, os tRNAs (RNAs transportadores) que transportam os aminoácidos para a síntese proteica são processados no nucléolo; da mesma forma que os genes de rRNA, os genes que codificam tRNAs estão agrupados no nucléolo. Assim, o nucléolo pode ser visto como uma grande fábrica, na qual diferentes RNAs não codifica-



**Figura 6-44 Fusão nucleolar.** Estas micrografias ópticas de fibroblastos humanos crescidos em cultura mostram vários estágios de fusão nucleolar. Após a mitose, cada um dos dez cromossomos humanos que carrega um conjunto de genes de rRNA começa a formar pequenos nucléolos, os quais rapidamente coalescem à medida que crescem para formar um único grande nucléolo característico de muitas células interfásicas. (Cortesia de E.G. Jordan e J. McGovern.)





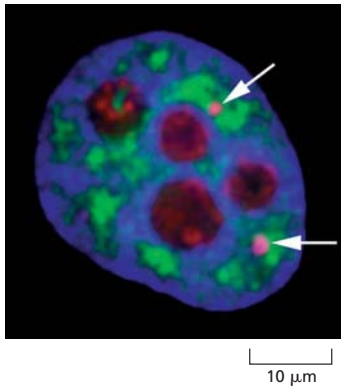
**Figura 6-45 A função do nucléolo na síntese do ribossomo e de outras ribonucleoproteínas.** O rRNA precursor 45S é empacotado em uma grande partícula ribonucleoproteica contendo várias proteínas ribossômicas importadas do citoplasma. Enquanto essa partícula permanece no nucléolo, componentes selecionados são adicionados e outros descartados conforme ela é processada em subunidades ribossômicas imaturas maiores e menores. As duas subunidades ribossômicas atingem sua forma funcional final apenas após serem transportadas individualmente através dos poros nucleares para o citoplasma. Outros complexos ribonucleoproteicos, incluindo a telomerase aqui mostrada, também são formados no nucléolo.

dores são transcritos, processados e ligados a proteínas para formar uma ampla gama de complexos ribonucleoproteicos.

### O núcleo contém uma variedade de agregados subnucleares

Apesar de o nucléolo ser a estrutura mais proeminente no núcleo, outros corpos nucleares foram visualizados e estudados (**Figura 6-46**). Esses corpos incluem os corpos de Cajal (assim nomeados em homenagem ao cientista que primeiro os descreveu em 1906) e grupos de grânulos de intercromatina (também denominados *speckles*, ou manchas). Como o nucléolo, essas outras estruturas nucleares não têm membranas e são altamente dinâmicas, dependendo das necessidades da célula. Sua formação é provavelmente mediada pela associação de domínios proteicos de baixa complexidade, como descrito no Capítulo 3 (ver Figura 3-36). A sua aparência é o resultado da estreita associação dos componentes proteicos e RNA envolvidos na síntese, associação e armazenamento das macromoléculas envolvidas na expressão gênica. Os corpos de Cajal são os locais onde os snRNPs e os snoRNPs passam pelas etapas finais de maturação, e onde os snRNPs são reciclados e seus RNAs são “reinicializados” após os rearranjos que ocorreram durante o *splicing* (ver p. 321.). Em contraste, foi sugerido que os grupos de grânulos de intercromatina correspondam a acúmulos de reserva de snRNPs totalmente maduros e de outros componentes do processamento de RNA, que estão prontos para ser utilizados na produção dos mRNAs.

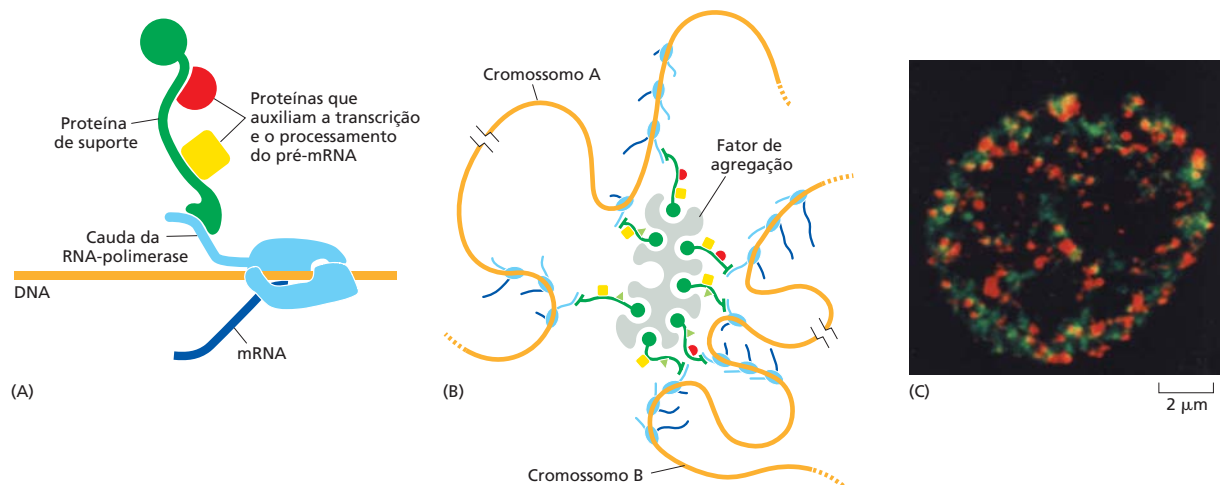




**Figura 6-46** Visualização de alguns corpos nucleares proeminentes. A proteína fibrilarina (*vermelho*), um componente de diversos snoRNPs, está presente tanto em nucléolos quanto em corpos de Cajal; (estes indicados por setas). Os corpos de Cajal (mas não os nucléolos) também estão destacados pela coloração de um dos seus principais componentes, a proteína coilina; a sobreposição do snRNP e da coilina aparece em rosa. Grupos de grânulos de intercromatina (*verde*) detectados com o uso de anticorpos contra uma proteína envolvida em *splicing* do pré-mRNA. O DNA está corado em azul pelo corante DAPI. (De J.R. Swedlow e A.I. Lomond, *Gen. Biol.* 2:1–7, 2001. Com permissão de BioMed Central. Fotomicrografia cortesia de Judith Sleeman.)

Os cientistas têm tido dificuldade na determinação da função das pequenas estruturas subnucleares descritas anteriormente, em parte devido às drásticas alterações na aparência que elas sofrem à medida que as células progridem no ciclo celular ou respondem a alterações no ambiente. Além disso, a disrupção de um tipo específico de corpo nuclear frequentemente não altera de forma significativa a viabilidade celular. Parece que a principal função destes agregados é aproximar os componentes mantendo-os sob uma concentração elevada de modo a acelerar sua associação. Por exemplo, estima-se que a organização do snRNP U4/U6 (ver Figura 6-28) ocorra dez vezes mais rapidamente nos corpos de Cajal do que seria o caso se o mesmo número de componentes estivesse disperso por todo o núcleo. Consequentemente, os corpos de Cajal parecem dispensáveis em muitos tipos de células, mas são absolutamente necessários em situações nas quais as células devem proliferar de forma rápida, como nos estágios iniciais do desenvolvimento dos vertebrados. Nesse momento, a síntese de proteínas (que depende do *splicing* do RNA) deve ser especialmente rápida, e atrasos podem ser letais.

Dada a importância dos corpos nucleares no processamento do RNA, poderíamos esperar que o *splicing* do pré-mRNA acontecesse em um local específico no núcleo, uma vez que ele requer numerosos componentes proteicos e de RNA. Entretanto, vimos que a associação dos componentes do *splicing* no pré-mRNA é cotranscricional; como consequência, o *splicing* deve ocorrer em muitas regiões sobre os cromossomos. Apesar de uma típica célula de mamífero poder expressar em torno de 15 mil genes simultaneamente, a transcrição e o *splicing* do RNA acontecem em apenas alguns milhares de regiões do núcleo. Esses sítios são altamente dinâmicos e provavelmente são o resultado da associação de componentes de transcrição e *splicing* para criar pequenas *fábricas*, o nome dado aos agregados específicos que contêm uma concentração local elevada de componentes selecionados que criam linhas de montagem bioquímicas (**Figura 6-47**).



**Figura 6-47** Um modelo para uma fábrica de produção de mRNA. A produção de mRNA é eficientemente realizada no núcleo pela agregação dos vários componentes necessários para a transcrição e para o processamento do pré-mRNA, originando uma fábrica bioquímica especializada. Em (A), uma proteína com provável função de suporte mantém vários componentes próximos a uma RNA-polimerase em transcrição. Outros componentes importantes estão ligados diretamente à cauda da RNA-polimerase, que também atua como um suporte complexo (ver Figura 6-22), mas não foram representados na ilustração por uma questão de simplificação. Em (B), diversos complexos de suporte foram reunidos para formar um agregado altamente enriquecido com os vários componentes necessários para a síntese e para o processamento do pré-mRNA. Esse modelo de suporte pode ser responsável pelos vários milhares de sítios ativos de transcrição e processamento de RNA normalmente observados no núcleo de uma célula de mamífero, cada um dos quais apresentando um diâmetro de cerca de 100 nm e contendo, de acordo com as estimativas, em média, cerca de 10 moléculas de RNA-polimerase II além de muitas outras proteínas. (C) Aqui, fábricas de produção de mRNA e fábricas de replicação de DNA foram visualizadas na mesma célula de mamífero pela incorporação rápida de nucleotídeos modificados diferentemente em cada ácido nucleico e pela detecção do RNA e do DNA produzido com o uso de anticorpos, um (*verde*) que detecta o DNA recém-sintetizado e o outro (*vermelho*) que detecta o RNA recém-sintetizado. (C, de D.G. Wansink et al., *J. Cell Sci.* 107:1449–1456, 1994. Com permissão de The Company of Biologists.)

Os grupos de grânulos de intercromatina – que contêm estoques de componentes envolvidos no processamento do RNA – são frequentemente observados nas proximidades dessas regiões de transcrição, e acredita-se que estejam envolvidos na reposição dos suprimentos utilizados. O núcleo pode ser considerado uma estrutura organizada em subdomínios, com snRNPs, snoRNPs e outros componentes nucleares movendo-se entre eles de forma ordenada, de acordo com as necessidades da célula.

## Resumo

*Antes de a síntese de uma determinada proteína poder ocorrer, a molécula de mRNA correspondente deve ser produzida por transcrição. As bactérias contêm um único tipo de RNA-polimerase (a enzima que realiza a transcrição de DNA em RNA). Uma molécula de mRNA é produzida depois que esta enzima inicia a transcrição em um promotor, sintetiza o RNA pela extensão da cadeia, finaliza a transcrição em um terminador e libera tanto o DNA-molde quanto a molécula de mRNA finalizada. Nas células eucarióticas, o processo de transcrição é muito mais complexo, e existem três RNA-polimerases – designadas como polimerase I, II e III –, evolutivamente relacionadas umas às outras e à polimerase bacteriana.*

*O mRNA dos eucariotos é sintetizado pela RNA-polimerase II. Essa enzima requer um conjunto de proteínas adicionais, os fatores gerais de transcrição, e proteínas específicas de ativação transcricional, para iniciar a transcrição em um molde de DNA. Ainda são necessárias mais proteínas (incluindo complexos de remodelagem da cromatina e enzimas modificadoras de histonas) para iniciar a transcrição nos moldes de cromatina no interior da célula.*

*Durante a fase de extensão ou alongamento da transcrição, o RNA em formação sofre três tipos de eventos de processamento: um nucleotídeo especial é adicionado à sua extremidade 5' (capeamento), os íntrons são removidos da molécula de RNA (splicing) e a extremidade 3' do RNA é gerada (por clivagem e poliadenilação). Cada um desses processos é iniciado por proteínas que acompanham a RNA-polimerase II por interação com sítios sobre sua longa cauda estendida C-terminal. O splicing difere dos demais pelo fato de muitas de suas etapas-chave serem mediadas por moléculas de RNA especializadas e não por proteínas. Apenas os mRNAs adequadamente processados são transportados através dos complexos do poro nuclear para o citosol, onde serão traduzidos em proteína.*

*No caso de diversos genes, o produto final é o RNA e não uma proteína. Nos eucariotos, esses genes são normalmente transcritos pela RNA-polimerase I ou pela RNA-polimerase III. A RNA-polimerase I produz os RNAs ribossômicos. Após sua síntese, sob a forma de um grande precursor, os rRNAs são modificados quimicamente, clivados e organizados sob a forma das duas subunidades ribossômicas no nucléolo – uma estrutura subnuclear distinta, que também ajuda a processar alguns complexos RNA-proteína menores na célula. As estruturas subnucleares adicionais (como os corpos de Cajal e os grupos de grânulos de intercromatina) são regiões onde os componentes envolvidos no processamento de RNA são organizados, estocados e reciclados. A concentração elevada de componentes em tais “fábricas” assegura que os processos serão catalisados de modo rápido e eficiente.*

## DO RNA À PROTEÍNA

Na seção anterior, vimos que o produto final de alguns genes é a própria molécula de RNA, como os RNAs presentes nos snRNPs e nos ribossomos. Entretanto, a maioria dos genes de uma célula produz moléculas de mRNA que são utilizadas como intermediárias na via de síntese de proteínas. Nesta seção, examinaremos como a célula converte a informação contida em uma molécula de mRNA em uma proteína. A tradução atraiu a atenção dos biólogos inicialmente no fim dos anos 1950, quando foi abordado o “problema da codificação”: como a informação em uma sequência linear de nucleotídeos no RNA é traduzida em uma sequência linear de um conjunto de subunidades quimicamente tão diferentes – os aminoácidos – em proteínas? Essa questão fascinante trouxe grande excitação. Havia um quebra-cabeças criado pela natureza que, após mais de 3 bilhões de anos de evolução, poderia finalmente ser resolvido por um dos produtos da evolução – os seres humanos. De fato, não somente o código foi finalmente decifrado passo a passo, como, no ano 2000, a elaborada maquinaria pela qual as células leem esse código – o ribossomo – foi finalmente revelada em seus detalhes atômicos.

