

# Estrutura dos Ácidos Nucleicos e Expressão Gênica

# 1

## CONCEITOS PRINCIPAIS

- A maior parte da informação genética eucariótica está armazenada no DNA encontrado no núcleo. Uma pequena quantidade está também estocada no DNA mitocondrial e no cloroplasto.
- Moléculas de DNA são polímeros de unidades repetidas de nucleotídeos, as quais são compostas por um dentre quatro tipos de bases nitrogenadas, mais um açúcar e um fosfato.
- A cadeia principal de qualquer molécula de DNA é um polímero de açúcar-fosfato, mas é a sequência de bases ligadas aos açúcares que determina a identidade e a função genética de uma determinada sequência de DNA.
- O DNA normalmente ocorre sob a forma de uma dupla-hélice, compreendendo duas fitas unidas por ligações de hidrogênio estabelecidas entre pares de bases nitrogenadas complementares.
- A transmissão da informação genética de uma célula para outra ocorre normalmente pela cópia das moléculas de DNA complementar que são então igualmente distribuídas para as duas células-filhas.
- Genes são segmentos de DNA utilizados como molde para sintetizar uma molécula de RNA complementar funcional.
- A maioria dos genes gera um RNA que servirá como molde para a produção de um polipeptídeo.
- Vários genes produzem moléculas de RNA que não codificam polipeptídeos. Tais RNAs não codificantes frequentemente auxiliam a regulação da expressão de outros genes.
- Assim como o DNA, moléculas de RNA são polímeros de unidades repetidas de nucleotídeos, as quais são compostas por quatro tipos de bases nitrogenadas (sendo três idênticas no DNA), mais um açúcar levemente diferente e um fosfato.
- Diferentemente do DNA, moléculas de RNA são normalmente fita simples.
- Para se tornar funcionais, RNAs recém-sintetizados devem passar por uma série de etapas de maturação, como excisão de sequências intervenientes não desejadas e alterações químicas de determinadas bases.
- A síntese de polipeptídeos ocorre nos ribossomos, tanto no citoplasma como no interior de mitocôndrias e cloroplastos.
- A informação sequencial codificada no RNA é interpretada nos ribossomos através de um código genético em trincas (*triplets*), determinando a estrutura básica do polipeptídeo.
- Peptídeos com frequência sofrem uma ampla variedade de alterações químicas.
- Proteínas apresentam extraordinárias diversidades estrutural e funcional.

## 1.1 DNA, RNA E POLIPEPTÍDEOS

A genética molecular foca primariamente na inter-relação entre dois ácidos nucleicos, DNA e RNA, e na maneira como eles são utilizados para a síntese de polipeptídeos, os componentes básicos de todas as proteínas. O RNA deve ter sido o material hereditário em um estágio inicial da evolução, no entanto, excetuando-se certos vírus, ele agora não desempenha mais esse papel. A informação genética é alternativamente armazenada em moléculas de DNA, quimicamente mais estáveis, as quais podem ser copiadas de maneira fiel e transmitidas às células-filhas.

Os ácidos nucleicos foram originalmente isolados do núcleo de células brancas sanguíneas, mas podem ser encontrados em todas as células e nos vírus. Em eucariotos, as moléculas de DNA são encontradas principalmente nos cromossomos do núcleo, mas cada mitocôndria também possui uma pequena molécula de DNA, do mesmo modo que os cloroplastos das células vegetais.

Um **gene** é uma parte de uma determinada molécula de DNA, que serve como molde para fazer uma molécula de RNA funcionalmente importante. Em organismos simples, como as bactérias, o DNA é condensado com genes (entre várias centenas e até alguns milhares de genes diferentes). Em eucariotos, as pequenas moléculas de DNA das mitocôndrias ou cloroplastos contêm poucos genes (de dezenas a centenas). Contudo, o núcleo muitas vezes contém milhares de genes, e eucariotos complexos normalmente apresentam dezenas de milhares. Neste último caso, todavia, muito do DNA consiste em sequências repetitivas, cujas funções não são facilmente identificadas. Parte das sequências de DNA repetitiva mantém funções cromossômicas essenciais, mas também existem várias cópias imperfeitas de genes funcionais.

Existem vários tipos distintos de moléculas de RNA, mas eles podem ser divididos em duas classes principais. Em uma classe, cada molécula de RNA contém uma sequência de **RNA codificante** que pode ser decodificado para gerar uma sequência polipeptídica correspondente. Como esta classe de RNA carrega informação genética do DNA para a maquinaria de síntese proteica, ela é descrita como **RNA mensageiro (mRNA)**. O RNA mensageiro produzido no núcleo precisa ser exportado para o citoplasma para ser utilizado para a produção de proteínas, mas o RNA mensageiro sintetizado na mitocôndria e nos cloroplastos é utilizado para fabricar proteínas no interior dessas organelas. A expressão da maioria dos genes é dedicada à produção de polipeptídeos e, assim, as proteínas representam o principal objetivo funcional da informação armazenada no DNA.

A outra classe de RNA é denominada **RNA não codificante**. Essas moléculas não servem como molde para a produção de polipeptídeos. Alternativamente, elas estão envolvidas no controle da expressão de outros genes, às vezes atuando de modo mais geral, às vezes regulando a expressão de um pequeno conjunto de genes-alvo. Esses processos regulatórios podem envolver moléculas de RNA catalítico (ribozimas).

### A maior parte da informação genética ocorre no sentido

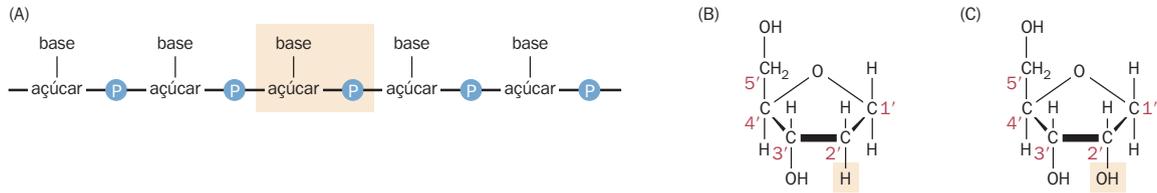
#### DNA → RNA → polipeptídeo

O fluxo de informação genética normalmente ocorre em sentido único: o DNA é transcrito para a produção de RNA, e então o RNA é utilizado para a produção de polipeptídeos que subsequentemente formam as proteínas. Em função dessa universalidade, esta sequência de transmissão da informação genética foi descrita como o dogma central da biologia molecular. Dois processos são essenciais em todos os organismos celulares:

- **transcrição**, pela qual o DNA é utilizado por uma RNA-polimerase como molde para sintetizar um dentre vários tipos diferentes de RNA.
- **tradução**, pela qual o mRNA é decodificado para produzir polipeptídeos nos ribossomos, os quais são grandes complexos de RNA-proteína encontrados no citoplasma, bem como em mitocôndrias e cloroplastos.

A informação genética é codificada em uma sequência linear de nucleotídeos do DNA e decodificada em grupos de três nucleotídeos por vez (*triplets*) para gerar uma sequência linear de nucleotídeos no RNA. Essa sequência é então decodificada em grupos de três nucleotídeos (códon) para gerar uma sequência linear de aminoácidos no produto polipeptídico.

Células eucarióticas, incluindo as células de mamíferos, contêm sequências de DNA cromossômico não viral, como os membros da família de DNA repetitivo LINE-1. Esta família gênica codifica transcriptases reversas celulares em mamíferos, as quais podem pro-



duzir seqüências de DNA a partir de um molde de RNA. O dogma central do sentido único no fluxo de informação genética nas células não é, portanto, estritamente válido.

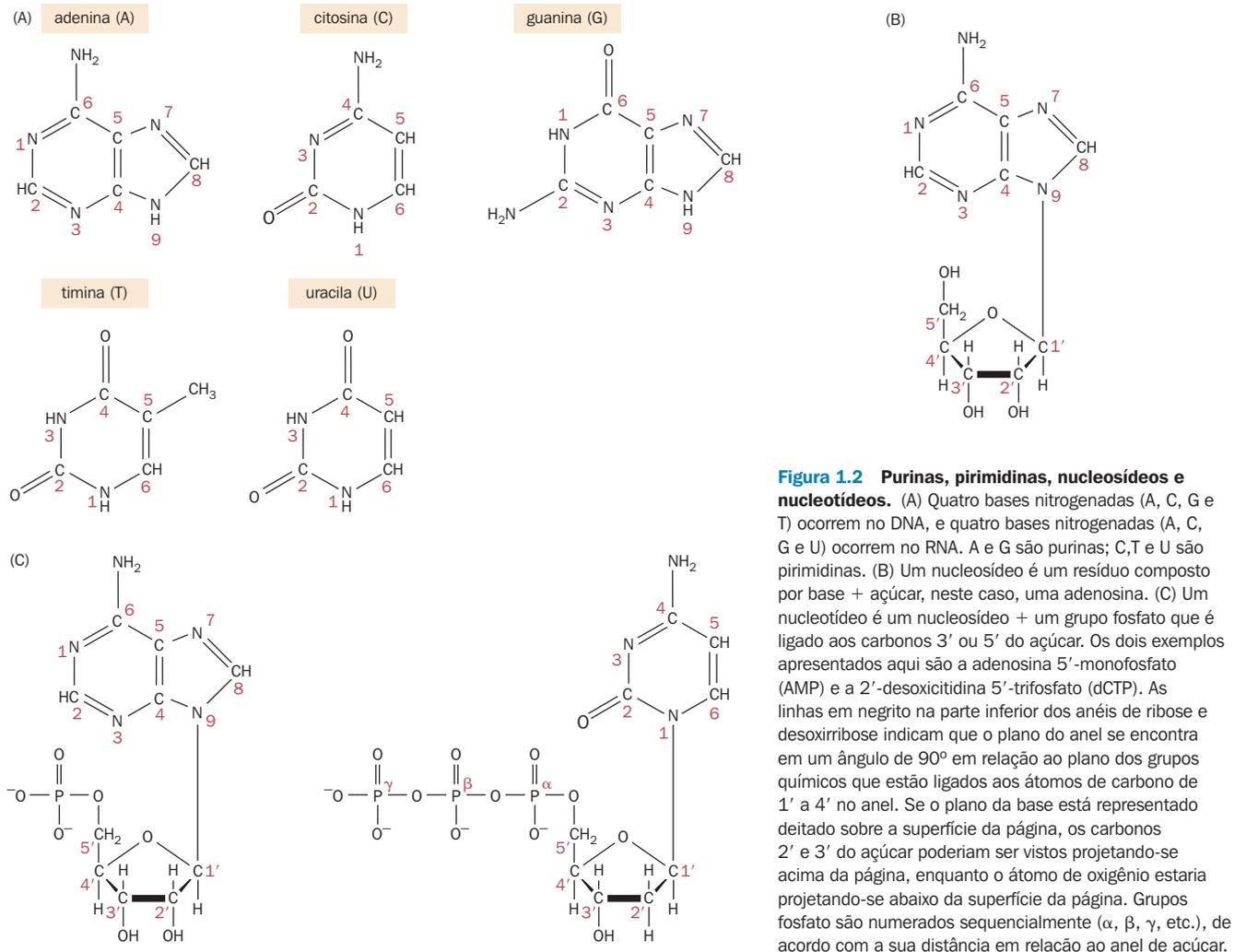
## Ácidos nucleicos e polipeptídeos são seqüências lineares de unidades repetitivas simples

### Ácidos nucleicos

DNA e RNA possuem estruturas muito similares. Ambos são grandes polímeros com longas cadeias principais de resíduos alternados de um fosfato e cinco açúcares de carbono. Preso a cada resíduo de açúcar se encontra uma base nitrogenada (Figura 1.1A). Os açúcares no DNA e no RNA diferem quanto à falta ou à presença, respectivamente, de um grupo  $-OH$  em seu carbono 2' (Figura 1.1B, C). O açúcar é uma desoxirribose no ácido desoxirribonucleico (DNA), enquanto no ácido ribonucleico (RNA), uma ribose.

Diferentemente dos resíduos de açúcar e fosfato, as bases de uma molécula de ácido nucleico variam. A seqüência de bases identifica o ácido nucleico e determina a sua função. Quatro tipos de bases são comumente encontrados no DNA: adenina (A), citosina

**Figura 1.1** Unidades de repetição nos ácidos nucleicos. (A) A cadeia principal dos ácidos nucleicos é composta por resíduos alternados de fosfato e açúcar. Ligada a cada açúcar se encontra uma base. A unidade básica de repetição (sombreado em pêssego claro) consiste em uma base + açúcar + fosfato = um nucleotídeo. O açúcar apresenta 5 átomos de carbono numerados de 1' a 5'. (B) No DNA, o açúcar é a desoxirribose. (C) No RNA, o açúcar é a ribose, a qual difere da desoxirribose por apresentar um grupo hidroxila (OH) ligado ao carbono 2'.



**Figura 1.2** Purinas, pirimidinas, nucleosídeos e nucleotídeos. (A) Quatro bases nitrogenadas (A, C, G e T) ocorrem no DNA, e quatro bases nitrogenadas (A, C, G e U) ocorrem no RNA. A e G são purinas; C, T e U são pirimidinas. (B) Um nucleosídeo é um resíduo composto por base + açúcar, neste caso, uma adenosina. (C) Um nucleotídeo é um nucleosídeo + um grupo fosfato que é ligado aos carbonos 3' ou 5' do açúcar. Os dois exemplos apresentados aqui são a adenosina 5'-monofosfato (AMP) e a 2'-desoxicitidina 5'-trifosfato (dCTP). As linhas em negrito na parte inferior dos anéis de ribose e desoxirribose indicam que o plano do anel se encontra em um ângulo de  $90^\circ$  em relação ao plano dos grupos químicos que estão ligados aos átomos de carbono de 1' a 4' no anel. Se o plano da base está representado deitado sobre a superfície da página, os carbonos 2' e 3' do açúcar poderiam ser vistos projetando-se acima da página, enquanto o átomo de oxigênio estaria projetando-se abaixo da superfície da página. Grupos fosfato são numerados sequencialmente ( $\alpha$ ,  $\beta$ ,  $\gamma$ , etc.), de acordo com a sua distância em relação ao anel de açúcar.

TABELA 1.1 Nomenclatura para bases, nucleosídeos e nucleotídeos

Base	Nucleosídeo = base + açúcar		Nucleotídeo = nucleosídeo + fosfato(s)		
	Ribose	Desoxirribose	Monofosfato	Difosfato	Trifosfato
<b>Purinas</b>					
Adenina	adenosina		adenosina monofosfato (AMP) <sup>a</sup>	adenosina difosfato (ADP)	adenosina trifosfato (ATP)
		desoxiadenosina	desoxiadenosina monofosfato (dAMP) <sup>b</sup>	desoxiadenosina difosfato (dADP)	desoxiadenosina trifosfato (dATP)
Guanina	guanosina		guanosina monofosfato (GMP) <sup>a</sup>	guanosina difosfato (GDP)	guanosina trifosfato (GTP)
		desoxiguanosina	desoxiguanosina monofosfato (dGMP)	desoxiguanosina difosfato (dGDP)	desoxiguanosina trifosfato (dGTP)
<b>Pirimidinas</b>					
Citosina	citidina		citidina monofosfato (CMP) <sup>a</sup>	citidina difosfato (CDP)	citidina trifosfato (CTP)
		desoxicitidina	desoxicitidina monofosfato (dCMP)	desoxicitidina difosfato (dCDP)	desoxicitidina trifosfato (dCTP)
Timina	timidina		timidina monofosfato (TMP) <sup>a</sup>	timidina difosfato (TDP)	timidina trifosfato (TTP)
		desoxitimidina	desoxitimidina monofosfato (dTMP)	desoxitimidina difosfato (dTDP)	desoxitimidina trifosfato (dTTP)
Uracila	uridina		uridina monofosfato (UMP) <sup>a</sup>	uridina difosfato (UDP)	uridina trifosfato (UTP)
		desoxiuridina	desoxiuridina monofosfato (dUMP)	desoxiuridina difosfato (dUDP)	desoxiuridina trifosfato (dUTP)

<sup>a</sup> Monofosfatos de nucleosídeo são alternativamente denominados como segue: AMP, adenilato; GMP, guanilato; CMP, citidilato; TMP, timidilato; UMP, uridilato.

<sup>b</sup> Quando o açúcar é uma ribose, o nucleotídeo é AMP; quando o açúcar é uma desoxirribose, o nucleotídeo é dAMP. Este padrão se aplica para toda a tabela. Note que TMP, TDP e TTP não são normalmente encontrados nas células.

(C), guanina (G) e timina (T). O RNA também apresenta quatro tipos principais de bases. Três deles (adenina, citosina, guanina) também ocorrem no DNA, mas no RNA a uracila (U) substitui a timina (Figura 1.2A).

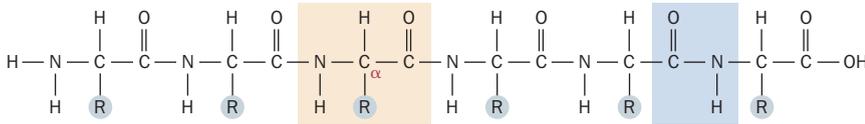
As bases são constituídas por anéis heterocíclicos de carbono e por átomos de nitrogênio, podendo ser divididas em duas classes: **purinas** (A e G), as quais possuem dois anéis interligados, e **pirimidinas** (C, T e U), as quais possuem um anel simples. Nos ácidos nucleicos, cada base é ligada ao carbono 1' (um linha) do açúcar, sendo que um açúcar contendo uma base ligada é chamado de nucleosídeo (Figura 1.2B). Um nucleosídeo contendo um grupo fosfato ligado aos carbonos 5' ou 3' do açúcar constitui a unidade básica de repetição de uma fita de DNA, sendo denominada de **nucleotídeo** (Figura 1.2C e Tabela 1.1).

### Polipeptídeos

As proteínas são compostas por um ou mais **polipeptídeos** que podem ser modificados pela adição de cadeias laterais de carboidratos ou outros grupos químicos. Assim como o DNA e o RNA, polipeptídeos são polímeros constituídos por uma sequência linear de unidades repetitivas. A unidade básica de repetição é denominada **aminoácido**. Um aminoácido possui um grupo amino positivamente carregado ( $-NH_2$ ) e um grupo ácido carboxílico (carboxila) negativamente carregado ( $-COOH$ ). Esses grupos são conectados por um átomo central de carbono (carbono  $\alpha$ ) que também abriga uma cadeia lateral identificadora, a qual determina a natureza química do aminoácido. Polipeptídeos são formados por uma reação de condensação entre o grupo amino de um aminoácido e o grupo carboxila do próximo, formando uma cadeia principal repetitiva, com cadeias laterais (chamadas de grupos R) que podem diferir de um aminoácido para o outro (Figura 1.3).

Os 20 aminoácidos comuns podem ser categorizados de acordo com suas cadeias laterais:

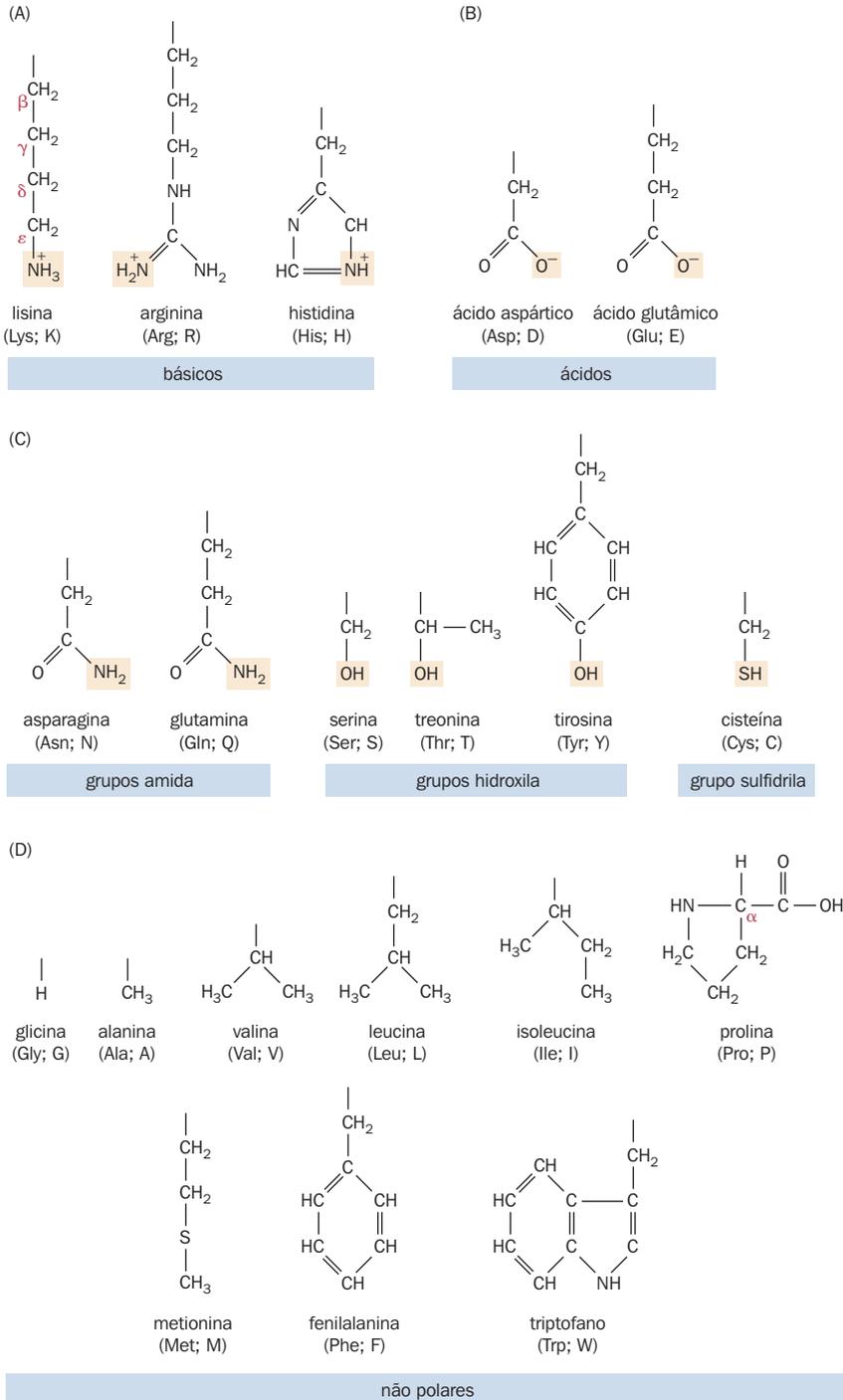
- aminoácidos básicos (Figura 1.4A) carregam uma cadeia lateral com carga elétrica líquida positiva em pH fisiológico;
- aminoácidos com grupo ácido (Figura 1.4B) carregam uma cadeia lateral com carga elétrica líquida negativa em pH fisiológico;
- aminoácidos polares não carregados (Figura 1.4C) apresentam carga elétrica global neutra, embora suas cadeias laterais carreguem grupos elétricos polares com cargas elétricas parciais (denominadas  $\delta+$  ou  $\delta-$ );



**Figura 1.3 A estrutura de repetição básica dos polipeptídeos.** Um polipeptídeo é um polímero que consiste em unidades repetidas de aminoácidos (sombreado em pêssego claro). Aminoácidos apresentam a fórmula geral  $H_2N-CH(R)-COOH$ , onde R representa a cadeia lateral,  $H_2N-$ , o grupo amino, e  $-COOH$ , o grupo carboxila. Em cada aminoácido, o carbono  $\alpha$  central carrega todos os três grupos. O sombreado em azul representa uma das ligações peptídicas que conecta aminoácidos adjacentes.

- aminoácidos neutros não polares (**Figura 1.4D**) são hidrofóbicos (repelem a água), geralmente interagem entre si e com outros grupos hidrofóbicos.

Em geral, aminoácidos polares são hidrofílicos, e aminoácidos não polares são hidrofóbicos. A glicina, que apresenta uma cadeia lateral muito pequena, e a cisteína (cujo



**Figura 1.4 Grupos R para os 20 aminoácidos comuns, reunidos de acordo com o grupo químico.** Existem 11 aminoácidos polares, divididos em três classes: (A) aminoácidos básicos (positivamente carregados); (B) aminoácidos ácidos (negativamente carregados) e (C) aminoácidos polares não carregados. Grupos químicos polares são destacados. (D) Além disso, uma quarta classe é composta por nove aminoácidos neutros não polares. Os aminoácidos pertencentes a cada grupo são quimicamente muito semelhantes. Os átomos de carbono das cadeias laterais são numerados a partir do carbono  $\alpha$  central (ver cadeia lateral da lisina). Na prolina, o grupo R se conecta ao grupo  $-NH_2$  do aminoácido, bem como ao carbono  $\alpha$  central.

**TABELA 1.2** Forças e ligações não covalentes fracas

Tipo de ligação	Natureza da ligação
Hidrogênio	Ligações de hidrogênio se formam quando um átomo de hidrogênio interage com átomos eletronegativos, normalmente átomos de oxigênio e nitrogênio
Iônica	Interações iônicas ocorrem entre grupos carregados. Elas podem ser muito fortes em cristais, mas em ambiente aquoso os grupos carregados estão blindados tanto por moléculas de água como por íons em solução, tornando-se relativamente fracas. No entanto, elas podem ser muito importantes nas funções biológicas, como no reconhecimento enzima-substrato
Forças de van der Waals	Qualquer par de átomos posicionados próximos um ao outro apresenta uma força de interação fraca (atração de van der Waals) como resultado da flutuação de suas cargas elétricas. Quando os átomos ficam extremamente próximos, se repelem com grande intensidade (repulsão de van der Waals). Embora atrações de van der Waals sejam muito fracas, o efeito cumulativo de várias dessas interações pode ser importante quando existe uma boa sobreposição entre as superfícies de duas macromoléculas
Forças hidrofóbicas	A água é uma molécula polar. Moléculas hidrofóbicas ou grupos químicos em ambiente aquoso tendem a se agrupar. Isso minimiza o seu efeito disruptivo na complexa rede de ligações de hidrogênio entre as moléculas de água. Grupos hidrofóbicos são mantidos unidos pelas chamadas ligações hidrofóbicas, embora a base para essa interação seja na verdade a repulsão em comum destes grupos às moléculas de água.

grupo -SH não é tão polar quanto um grupo -OH) ocupam posições intermediárias na escala de hidrofobicidade.

Conforme descrito a seguir, as cadeias laterais podem ser modificadas pela adição de vários grupos químicos ou cadeias de açúcar.

### O tipo de ligação química determina a estabilidade e a função

A estabilidade dos ácidos nucleicos e das proteínas é primariamente dependente de ligações covalentes fortes entre os átomos da cadeia principal. Adicionalmente às ligações covalentes, ligações não covalentes fracas (**Tabela 1.2**) são importantes tanto nas moléculas de ácidos nucleicos e proteínas como entre essas moléculas (**Quadro 1.1**). Ligações não covalentes individuais são normalmente cerca de 10 vezes mais fracas do que ligações covalentes individuais.

A estrutura da água é particularmente complexa, com uma rede oscilante de ligações não covalentes ocorrendo entre as moléculas de água. A força predominante nessa estrutura é a **ligação de hidrogênio**, uma ligação eletrostática fraca entre átomos de hidrogênio parcialmente positivos e átomos parcialmente negativos (átomos de oxigênio, no caso das moléculas de água).

**QUADRO 1.1** A importância das ligações de hidrogênio em ácidos nucleicos e proteínas

#### Ligações de hidrogênio intermoleculares em ácidos nucleicos

São importantes para a formação dos seguintes ácidos nucleicos de dupla-fita:

- **DNA dupla-fita.** A estabilidade da dupla-hélice é mantida pelas ligações de hidrogênio entre os pares de bases A-T e C-G. As ligações de hidrogênio individuais são fracas, mas em células eucarióticas as duas fitas de DNA são mantidas unidas por dezenas de milhares a centenas de milhões de ligações de hidrogênio.
- **Dúplices DNA-RNA.** Ligações de hidrogênio se formam naturalmente entre DNA e RNA durante a transcrição, mas o pareamento de bases é transiente, pois o RNA se afasta do DNA à medida que amadurece.
- **RNA dupla-fita.** Ocorre de maneira estável no genoma de alguns vírus, assim como de maneira transiente em células durante a expressão gênica. Por exemplo, durante o *splicing* do RNA, pequenas moléculas de RNA nuclear se ligam a seqüências complementares no pré-

-mRNA, e códons de mRNA se ligam ao tRNA durante a tradução. Muitos RNAs regulatórios, como micro RNAs, controlam a expressão de genes-alvo pelo pareamento de bases com seqüências complementares em nível de RNA.

#### Ligações de hidrogênio intramoleculares em ácidos nucleicos

São particularmente prevalentes em moléculas de RNA. Pareamento intramolecular de bases proporciona a formação de estruturas secundárias (*hairpins*) que podem ser imprescindíveis para a estrutura de alguns RNAs, como rRNAs e tRNAs (ver Figura 1.9), e como alvo para a regulação gênica.

#### Ligações de hidrogênio intramoleculares em proteínas

Vários elementos característicos da estrutura secundária das proteínas, como  $\alpha$ -hélices e folhas  $\beta$ -pregueadas, surgem em função das ligações de hidrogênio entre as cadeias laterais de diferentes aminoácidos na mesma cadeia polipeptídica.

Moléculas carregadas são altamente solúveis em água. Em função dos grupos fosfato presentes em seus nucleotídeos, tanto DNA como RNA são poliânions negativamente carregados. Dependendo da composição de aminoácidos, as proteínas podem ser eletricamente neutras, apresentar carga líquida positiva (proteínas básicas), ou carga líquida negativa (proteínas ácidas). Todas essas moléculas podem formar múltiplas interações com a água durante sua solubilização. Mesmo proteínas eletricamente neutras são facilmente solúveis se contiverem uma quantidade suficiente de aminoácidos carregados ou polares neutros. Em contrapartida, proteínas de membrana com muitos aminoácidos hidrofóbicos são termodinamicamente mais estáveis em um ambiente hidrofóbico.

Embora individualmente fracas, as numerosas ligações não covalentes realizam em conjunto uma grande contribuição à estabilidade da **conformação** (estrutura) dessas moléculas, sendo importantes para especificar o formato de uma macromolécula. Ligações covalentes são comparativamente estáveis, de modo que um grande ganho de energia é necessário para rompê-las. Ligações não covalentes, entretanto, são constantemente estabelecidas e rompidas em temperaturas fisiológicas (ver Quadro 1.1).

## 1.2 ESTRUTURA DOS ÁCIDOS NUCLEICOS E REPLICAÇÃO DO DNA

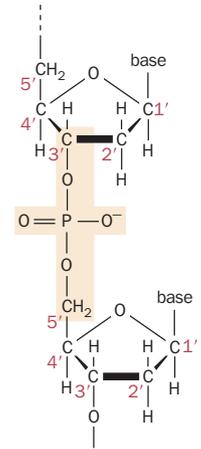
### Estrutura de DNA e RNA

Moléculas de DNA e RNA apresentam cadeias principais lineares, compostas por resíduos alternados de açúcares e grupos fosfato. Os resíduos de açúcar são unidos por **ligações 3', 5'-fosfodiéster**, nas quais um grupo fosfato conecta o carbono 3' de um açúcar ao carbono 5' do próximo açúcar na cadeia de açúcar-fosfato (Figura 1.5).

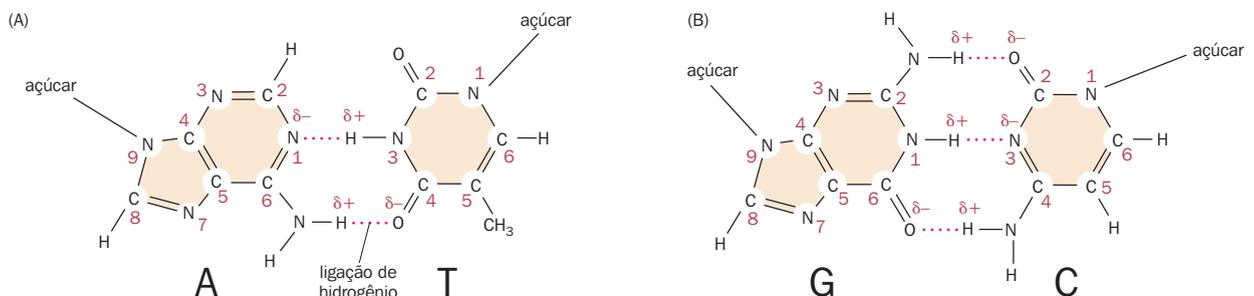
Embora certos genomas virais sejam compostos por DNA de fita simples, o DNA celular forma uma **dupla-hélice**: duas fitas de DNA são mantidas unidas por ligações de hidrogênio para formar um dúplice. As ligações de hidrogênio ocorrem entre as cadeias laterais de **pares de bases complementares**, lateralmente opostos nas duas fitas do dúplice de DNA. Tais pares de bases ocorrem de acordo com as regras de Watson-Crick: A pareja com T, enquanto G pareja com C (Figura 1.6).

Em função do pareamento de bases, a composição de bases do DNA não é aleatória: a quantidade de A equivale à de T, e a quantidade de G equivale à de C. A composição de bases do DNA pode, portanto, ser especificada pela estimativa da porcentagem de GC (= porcentagem de G + porcentagem de C) em sua composição. Por exemplo, uma sequência de DNA com 42% de GC apresenta a seguinte composição de bases: G, 21%; C, 21%; A, 29% e T, 29%.

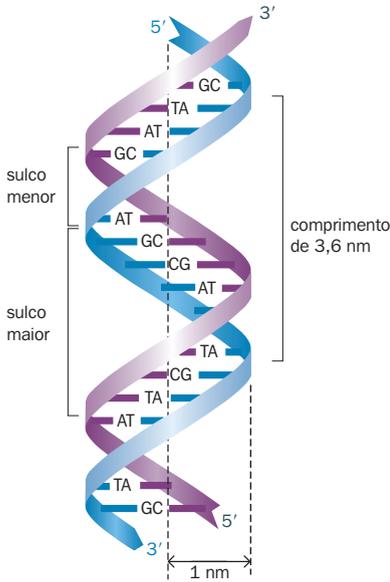
As duas fitas do DNA dupla-hélice se curvam uma em torno da outra para produzir um sulco menor e um sulco maior na dupla-hélice, onde a distância ocupada por um único giro completo da hélice (seu comprimento) mede 3,6 nm (Figura 1.7). O DNA pode adotar diferentes tipos de estrutura helicoidal. Em condições fisiológicas, a maior parte do DNA em bactérias ou células eucarióticas adota a forma B, ou seja, uma hélice com giro para a direita (forma uma espiral com sentido horário se afastando do observador) que apresenta 10 pares de bases por giro. As formas mais raras são o DNA A (hélice com giro para a direita apresentando 11 pares de bases por giro) e o DNA Z (hélice com giro para a esquerda apresentando 12 pares de bases por giro).



**Figura 1.5** Ligação 3',5'-fosfodiéster. A ligação fosfodiéster (sombreado em pêssego claro) reúne o carbono 3' de um açúcar ao carbono 5' do próximo açúcar, na cadeia de açúcar-fosfato de um ácido nucleico.



**Figura 1.6** Pares de bases AT e CG. (A) O par de bases AT apresenta duas ligações de hidrogênio (linha vermelha tracejada); (B) O par de bases CG apresenta três. Cargas positivas e negativas parciais são representadas por  $\delta+$  e  $\delta-$ , respectivamente.



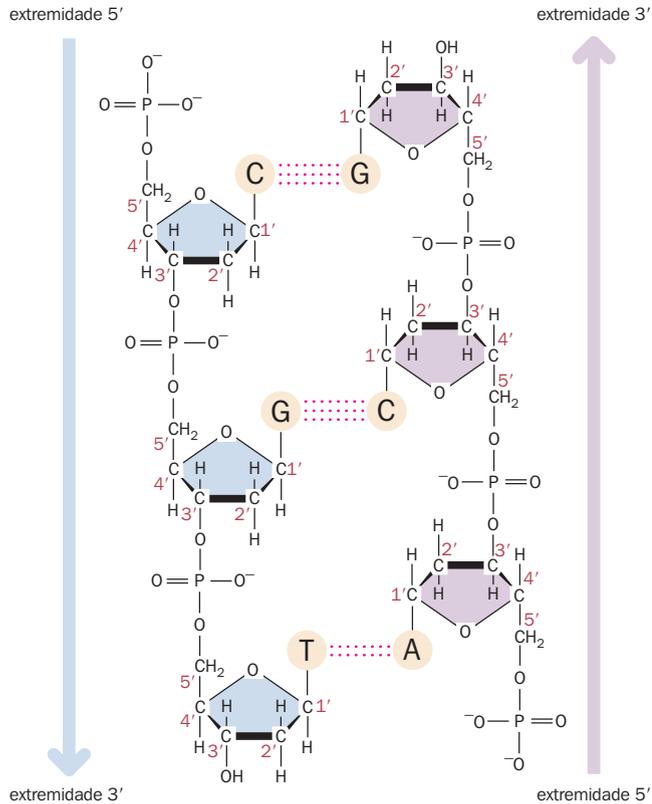
**Figura 1.7 Características da dupla-hélice de DNA.** As duas fitas de DNA se enrolam uma na outra, produzindo um sulco menor e um sulco maior na dupla-hélice. A dupla-hélice apresenta um comprimento de 3,6 nm e um raio de 1 nm por giro.

Como as ligações fosfodiéster conectam os carbonos 3' e 5' de resíduos de açúcar sucessivos, as duas extremidades de uma fita simples de DNA são diferentes. A **extremidade 5'** apresenta um resíduo de açúcar no qual o carbono 5' não está conectado a outro resíduo de açúcar. A **extremidade 3'** apresenta um resíduo de açúcar cujo carbono 3' não está envolvido em uma ligação fosfodiéster. As duas fitas do dúplice de DNA são descritas como antiparalelas uma à outra, visto que a direção 5' → 3' de uma fita de DNA é a oposta da direção de sua complementar, de acordo com as regras de Watson-Crick para o pareamento de bases (**Figura 1.8**).

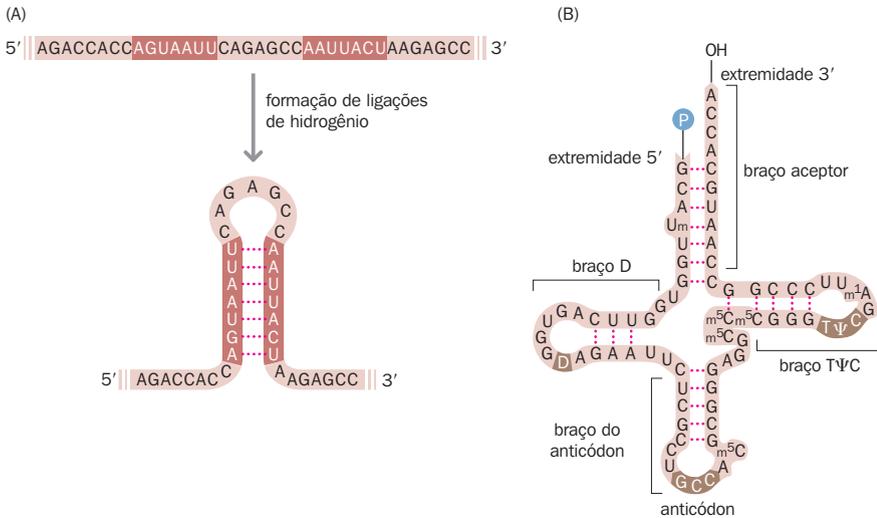
A informação genética é codificada pela sequência linear de bases nas fitas de DNA. As duas fitas do dúplice de DNA apresentam sequências complementares, portanto a sequência de bases de uma fita de DNA pode ser prontamente inferida a partir da sequência da outra fita. É comum traçar o DNA escrevendo a sequência de bases de apenas uma fita, na direção 5' → 3', a qual é a direção de síntese de uma nova fita de DNA ou RNA a partir de um DNA molde. Ao descrever a sequência de uma região do DNA compreendendo duas bases vizinhas (um dinucleotídeo) em uma fita de DNA, é frequente a inserção de um 'p' para indicar a ligação fosfodiéster entre eles. Deste modo, um par de bases CG indica que um C em uma fita de DNA apresenta uma ligação de hidrogênio com um G na fita complementar. Por outro lado, a indicação CpG representa uma desoxicitidina ligada covalentemente a uma desoxiguanosina na mesma fita de DNA (ver Figura 1.8).

Diferente do DNA, o RNA apresenta-se normalmente como uma fita simples, excetuando-se certos vírus que apresentam genomas de RNA dupla-fita. No entanto, para realizar algumas funções celulares, pode ser necessário que duas moléculas de RNA se associem transientemente formando pares de bases. Além disso, ligações de hidrogênio intermoleculares também permitem a formação de dúplices de RNA-DNA (ver Quadro 1.1).

Além disso, ligações de hidrogênio podem ocorrer entre bases de uma mesma fita de RNA (ou DNA), produzindo trechos de sequência dupla-fita que são estrutural e funcionalmente importantes. Podem ser constituídas estruturas em forma de grampo (*hairpin*), cujas hastes são estabilizadas por ligações de hidrogênio entre as bases (**Figura 1.9A**). O pareamento de bases intracadeia permite que certas moléculas de RNA apresentem estruturas complexas (**Figura 1.9B**).



**Figura 1.8 Natureza antiparalela da dupla-hélice de DNA.** As duas fitas antiparalelas do DNA correm em direções opostas ligando átomos de carbono 3', dos resíduos de açúcar, a átomos de carbono 5'. Este trinucleotídeo dupla-fita apresenta a sequência 5' pCpGpT-OH 3'/5' pApCpG-OH 3', onde p representa o grupo fosfato e -OH 3', o grupo hidroxila terminal. Isso é normalmente abreviado de modo a apresentar a sequência de nucleotídeos de 5' → 3' em apenas uma fita, seja como 5'-CGT-3' (fita azul), seja como 5'-AGC-3' (fita roxa).



Na molécula de RNA dupla-fita, A pareia com U, em vez de T. Embora G pareie com C, às vezes pares de bases G-U são formados (ver exemplo na Figura 1.9B). Embora não sejam particularmente estáveis, pares de bases G-U não alteram de modo significativo a hélice RNA-RNA.

**A replicação é semiconservativa e semidescontínua**

Para que a síntese de uma nova fita de DNA (**replicação**) se inicie, as duas fitas de DNA de uma hélice precisam ser desenroladas pela enzima helicase. Cada uma das fitas desenroladas de DNA serve então como molde para a DNA-polimerase produzir fitas de DNA complementar, utilizando os quatro desoxinucleosídeos trifosfato (dATP, dCTP, dGTP e dTTP). Duas hélices “filhas” de DNA são formadas, cada uma idêntica à molécula parental (**Figura 1.10**). Cada hélice filha contém uma fita da molécula parental e uma fita de DNA recém-sintetizada, portanto o processo de replicação é dito **semiconservativo**.

A replicação do DNA é iniciada em pontos específicos, chamados **origens de replicação**, gerando forquilha de replicação em formato de Y, onde a dupla-hélice de DNA está aberta. As fitas de DNA antiparalelas servem como molde para a síntese de fitas filhas complementares, que correm na direção oposta.

A direção do crescimento da cadeia é 5' → 3' para uma fita filha, a **fita líder** (*leading strand*); mas é 3' → 5' para a outra fita filha, a **fita retardada** (*lagging strand*) (**Figura 1.11**). As reações catalisadas pela DNA-polimerase envolvem a adição de um resíduo de desoxinucleosídeo monofosfato (dNMP) ao grupo hidroxila livre na extremidade 3' da cadeia nascente de DNA. No entanto, apenas a fita líder possui sempre um grupo hidroxila livre que permite alongação contínua na mesma direção em que a forquilha de replicação se move.

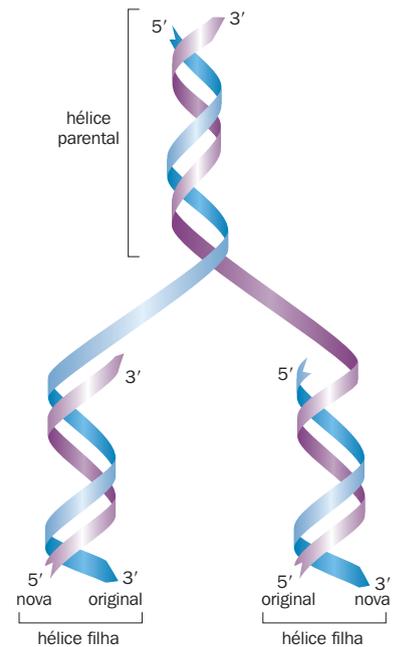
A direção da síntese da fita retardada é oposta àquela na qual se move a forquilha de replicação. Como resultado, a síntese da nova fita precisa ser realizada numa série progressiva de passos, produzindo segmentos de DNA que normalmente apresentam de 100 a 1.000 nucleotídeos em extensão (fragmentos de Okazaki). Fragmentos sucessivos são posteriormente ligados de forma covalente pela enzima DNA-ligase, garantindo a criação de duas hélices de DNA filhas. Apenas a fita líder é sintetizada continuamente, portanto a síntese de DNA é dita **semidescontínua**.

**A DNA-polimerase atua, por vezes, no reparo e na recombinação do DNA**

A maquinaria para a replicação do DNA se baseia em uma variedade de proteínas (**Quadro 1.2**) e *primers* (iniciadores) de RNA e tem sido altamente conservada ao longo da evolução. No entanto, a complexidade do processo é maior em células de mamíferos, em termos do número de diferentes DNA-polimerases (**Tabela 1.3**) e de suas proteínas e subunidades constituintes.

A maioria das DNA-polimerases em células de mamíferos utilizam uma fita individual de DNA como molde para sintetizar uma fita complementar de DNA, sendo, por-

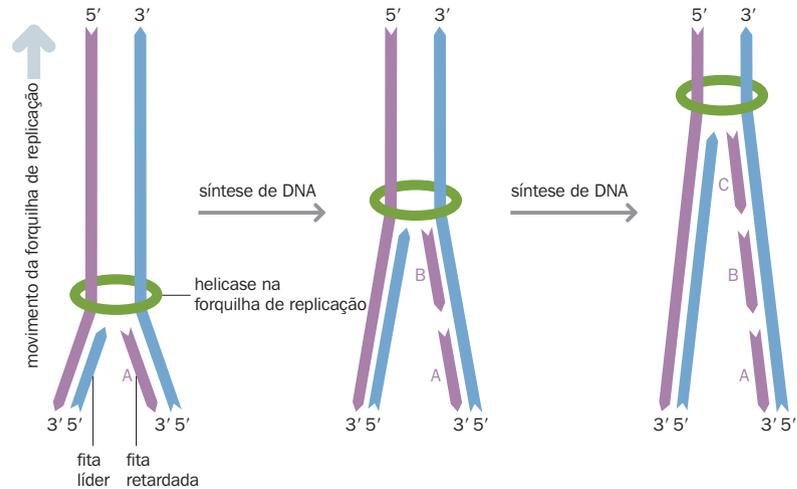
**Figura 1.9 Pareamento de bases em ácidos nucleicos fita simples.** (A) Formação de grampo (*hairpin*) mediada por ligações de hidrogênio intramoleculares. Ligações de hidrogênio entre as sequências destacadas pelo sombreamento em rosa escuro neste ácido nucleico fita simples (no caso um RNA) podem estabilizar o dobramento que leva à formação de um grampo com a haste em dupla-fita. (B) Intenso pareamento de bases intramolecular no RNA transportador. O tRNA<sup>Gly</sup>, apresentado aqui como exemplo, ilustra a estrutura clássica em forma de trevo apresentada por um tRNA. Existem três grampos (braço D, braço do anticódon e braço TΨC) além de um trecho de pareamento de bases entre as sequências terminais 5' e 3' (chamada de braço aceptor, pois a extremidade 3' é utilizada para ligar um aminoácido). Note que os tRNAs sempre apresentam o mesmo número de pares de bases nas hastes dos diferentes braços de sua estrutura em forma de trevo, e que o anticódon no centro da alça do meio identifica o tRNA de acordo com o aminoácido que ele irá carregar. Os nucleotídeos minoritários apresentados são: D, 5,6-dihidrouridina; Ψ, pseudouridina (5-ribosiluracila); m<sup>5</sup>C, 5-metilcitidina; m<sup>1</sup>A, 1-metiladenosina; Um, 2'-O-metiluridina.



**Figura 1.10 Replicação semiconservativa do DNA.** A hélice de DNA parental consiste em duas fitas complementares de DNA, antiparalelas, que se desenrolam para servir como molde para a síntese de novas fitas complementares de DNA. Cada hélice filha completa contém uma das duas fitas de DNA parental mais uma fita de DNA recém-sintetizada, sendo estruturalmente idêntica à hélice de DNA original.

**Figura 1.11 Replicação**

**semidescontínua do DNA.** A enzima helicase abre a forquilha de replicação para que a síntese de uma nova fita de DNA possa ser iniciada. A direção do movimento da forquilha de replicação acompanha a síntese contínua da fita líder, que se estende no sentido 5'→3'. A replicação é dita semidescontínua em função da fita retardada, a qual é sintetizada na direção oposta e em pequenos blocos (fragmentos de Okazaki, apresentados aqui como fragmentos A, B e C). Estes blocos serão posteriormente unidos pela DNA-ligase.



tanto, as DNA-polimerases dependentes de DNA. Diferentemente das RNA-polimerases, DNA-polimerases normalmente exigem como substrato o terminal 3'-hidroxila de uma fita *primer* complementar. Portanto, um *primer* de RNA, sintetizado por uma primase, é necessário para fornecer um grupo 3' OH livre para a DNA-polimerase iniciar a síntese de DNA.

Existem aproximadamente 20 tipos diferentes de DNA-polimerases em células de mamíferos. A maioria utiliza DNA como molde para sintetizar DNA, e foram classificadas em quatro famílias – A, B, X e Y – com base na comparação de suas sequências (ver Tabela 1.3).

Os membros da família B são as DNA-polimerases clássicas (alta fidelidade) e incluem as enzimas responsáveis pela replicação do DNA nuclear. A maioria apresenta atividade 3' – 5' exonuclease, a qual é importante para a **revisão**: se uma base incorreta é adicionada à extremidade 3' OH livre da cadeia nascente de DNA, a exonuclease 3' – 5' a remove. Isso resulta em uma replicação de alta fidelidade, visto que erros na incorporação de bases passam a ser extremamente infreqüentes. A DNA-polimerase  $\alpha$  é um complexo formado por uma polimerase e por uma primase, sendo responsável pela iniciação da síntese de DNA e pela iniciação dos fragmentos de Okazaki. As DNA-polimerases  $\delta$  e  $\epsilon$  são responsáveis pela maior parte da síntese de DNA e são fita-específicas (ver Tabela 1.3).

Muitas DNA-polimerases atuam no reparo ou na recombinação de DNA. Elas incluem as clássicas de alta fidelidade, que também estão envolvidas com a replicação (DNA-polimerases  $\delta$  e  $\epsilon$ ), e também outras que estão envolvidas apenas com reparo e recombinação. Dentre estas últimas, algumas são polimerases de alta fidelidade, mas muitas delas são comparativamente mais propensas à incorporação de bases incorretas, com destaque

**QUADRO 1.2 Principais classes de enzimas envolvidas na replicação do DNA**

- Topoisomerases – iniciam o processo de desenrolamento do DNA pela quebra de uma das fitas de DNA, aliviando a tensão que mantém a hélice em sua forma enrolada e superenrolada.
- Helicases – desenrolam a dupla-hélice na forquilha de replicação, uma vez que o superenrolamento foi desfeito pela topoisomerase.
- Proteínas de ligação à fita simples – mantêm a estabilidade da forquilha de replicação. O DNA fita simples é bastante vulnerável ao ataque enzimático e as proteínas ligadas o protegem da degradação.
- Primases – enzimas que adicionam uma pequena sequência de RNA complementar (um *primer*) ao DNA fita simples na forquilha de replicação. O *primer* de RNA fornece a extremidade 3' OH necessária para que a DNA-polimerase inicie a síntese (diferentemente das RNA-polimerases, DNA-polimerases não podem iniciar a síntese de novas fitas a partir de moldes livres de fita simples, necessitando de uma molécula de iniciação com um grupo 3' OH livre, no qual desoxinucleosídeos trifosfato podem ser adicionados para formar a fita complementar).
- DNA-polimerases – sintetizam novas fitas de DNA. Nova síntese de DNA celular depende normalmente da existência de uma fita molde de DNA, a qual é lida por uma polimerase dependente de DNA. Este complexo agregado de subunidades proteicas também fornece atividades de revisão e reparo do DNA (ver Tabela 1.3). Isto significa que qualquer base que seja incorporada incorretamente pode ser identificada, removida e reparada. O DNA também pode ser sintetizado a partir de um molde de RNA, utilizando uma DNA-polimerase dependente de RNA (uma transcriptase reversa). As extremidades de cromossomos lineares são copiadas utilizando-se uma transcriptase reversa (telomerase).
- DNA-ligasas – necessárias para fechar falhas deixadas na fita de DNA recém-sintetizada, após a remoção dos *primers* de RNA e o preenchimento destas lacunas pela DNA-polimerase. As DNA-ligasas catalisam a formação de uma ligação fosfodiéster entre grupos 3' hidroxila e 5' fosfato que estejam adjacentes e não ligados.

TABELA 1.3 DNA-polimerases de mamíferos

DNA-polimerases dependentes de DNA			
Polimerase	Família	Papel principal na replicação de DNA	Atividades adicionais ou alternativas no reparo de DNA, recombinação, etc.
$\alpha$ (alpha)	B	inicia a síntese nas origens de replicação e também inicia a síntese dos fragmentos de Okazaki na fita retardada	
$\beta$ (beta)	X		reparo por excisão de base <sup>b</sup>
$\gamma$ (gama)	A	síntese de DNA mitocondrial	reparo de DNA mitocondrial
$\delta$ (delta)	B	principal polimerase que sintetiza a fita retardada	múltiplas atividades no reparo de DNA
$\epsilon$ (épsilon)	B	sintetiza a fita líder	múltiplas atividades no reparo de DNA
$\zeta$ (zeta)	B		síntese translesão <sup>c</sup>
$\eta$ (eta)	Y		síntese translesão <sup>c</sup>
$\theta$ (theta)	A		possível atividade no reparo de DNA cruzado interfitas <sup>d</sup> ; reparo por excisão de base <sup>b</sup> ; síntese translesão <sup>c</sup> ; hipermutação somática <sup>g</sup>
$\iota$ (iota)	Y		síntese translesão <sup>c</sup> ; possível atividade no reparo por excisão de base <sup>b</sup> ; e reparo de mau pareamento <sup>e</sup>
$\kappa$ (kappa)	Y		síntese translesão <sup>c</sup> ; reparo por excisão de nucleotídeo <sup>f</sup>
$\lambda$ (lambda)	X		reparo de quebras duplas de DNA; recombinação VDJ <sup>g</sup> ; reparo por excisão de base <sup>b</sup>
$\mu$ (mu)	X		
$\nu$ (nu)	A		possível atividade no reparo de DNA cruzado interfitas <sup>d</sup>
Rev1	Y		síntese translesão <sup>c</sup>
TdT <sup>a</sup>	X		recombinação VDJ <sup>g</sup>
DNA-polimerases dependentes de RNA (transcriptases reversas)			
Transcriptases reversas de repetições intercaladas (LINE-1 ou elementos retrovirais endógenos)		ocasionalmente converte mRNA e outros RNAs em cDNA, o qual pode se integrar em outra parte do genoma	
Transcriptase reversa da telomerase (Tert)		replica o DNA na extremidade de cromossomos lineares	

<sup>a</sup> Desoxinucleotídeo terminal transferase. <sup>b</sup> Reparo por excisão de bases identifica e remove bases impróprias ou indevidamente modificadas. <sup>c</sup> Síntese translesão envolve a replicação de DNA a partir de DNA danificado (lesões) na fita molde. <sup>d</sup> Reparo de DNA cruzado interfitas diz respeito ao reparo de lesões altamente citotóxicas, onde ligações covalentes de DNA são formadas entre as fitas de DNA. <sup>e</sup> Reparo de mau pareamento diz respeito a uma forma de reparo de DNA que corrige erros ocasionados pela formação de um par de bases com nucleotídeos não complementares. <sup>f</sup> Reparo por excisão de nucleotídeo é utilizada para corrigir lesões que distorcem a hélice. <sup>g</sup> Hipermutação somática e recombinação VDJ são mecanismos utilizados em células B para diversificar as sequências de imunoglobulinas.

para os membros das famílias X e Y. A DNA-polimerase  $\iota$  (iota), por exemplo, pode apresentar uma taxa de erro mais de 20 mil vezes superior à apresentada pela polimerase  $\epsilon$ .

A alta taxa de erro apresentada por algumas polimerases é tolerada em função de sua atuação no processo de reparo de DNA, sendo utilizadas apenas para sintetizar pequenos fragmentos deste. Em outros casos, elevadas taxas de erro podem ser vantajosas. Por exemplo, DNA-polimerases de baixa fidelidade podem continuar a síntese de uma nova fita de DNA mesmo quando o molde está danificado (síntese translesão), bem como contribuir para a geração de variabilidade em imunoglobulinas (p. ex., pela introdução de várias trocas de bases nas sequências codificantes) e, deste modo, auxiliar no reconhecimento de diversos antígenos exógenos pelo sistema imune.

## Muitos vírus apresentam genomas de RNA

O DNA é o material hereditário em todas as células da atualidade e normalmente pensa-se no **genoma** como o termo coletivo para as diferentes moléculas hereditárias de DNA em um organismo ou célula. No entanto, muitos vírus apresentam um genoma de RNA. Essas moléculas de RNA podem sofrer autorreplicação, embora o grupo 2' OH em seus resíduos de ribose tornem as ligações açúcar-fosfato quimicamente mais instáveis. Em contrapartida, no DNA, os resíduos de desoxirribose carregam apenas átomos de hidrogênio na posição 2', tornando o DNA um portador mais estável para a informação genética.

Os vírus desenvolveram várias estratégias diferentes para infectar e subverter as células, de modo que seus genomas apresentam uma extraordinária diversidade quando

TABELA 1.4 Diferentes classes de genoma

	Simple linear	Múltiplo linear	Simple circular	Múltiplo circular	Misto (linear + circular)
<b>Genomas de DNA</b>					
ssDNA (fita simples)	alguns vírus	vírus com ssDNA segmentado	alguns vírus	–	–
dsDNA (dupla-fita)	alguns vírus; raras bactérias (p. ex., <i>Borrelia</i> )	vírus com dsDNA segmentado; núcleos eucarióticos	mitocôndria; cloroplastos; muitas bactérias e arqueias	vírus multipartidos; algumas bactérias	raras bactérias, (p. ex., <i>Agrobacterium tumefaciens</i> )
<b>Genomas de RNA</b>					
ssRNA (fita simples)	alguns vírus	vírus com ssRNA segmentado	raros vírus	–	–
dsRNA (dupla-fita)	poucos vírus	vírus com dsRNA segmentado	–	–	–

Ver Figura 1.12 para exemplos de genomas virais e explicações sobre vírus segmentados e multipartidos.

comparados aos genomas celulares (Tabela 1.4 e Figura 1.12). Como a replicação do RNA apresenta uma taxa de erro muito maior do que a replicação do DNA, genomas de RNA viral apresentam um maior acúmulo mutacional do que genomas de DNA. Embora genomas de RNA viral sejam normalmente pequenos, a elevada taxa de mutação permite uma adaptação mais rápida a mudanças nas condições ambientais. Vírus de RNA geralmente replicam no citoplasma, enquanto vírus de DNA geralmente replicam no núcleo.

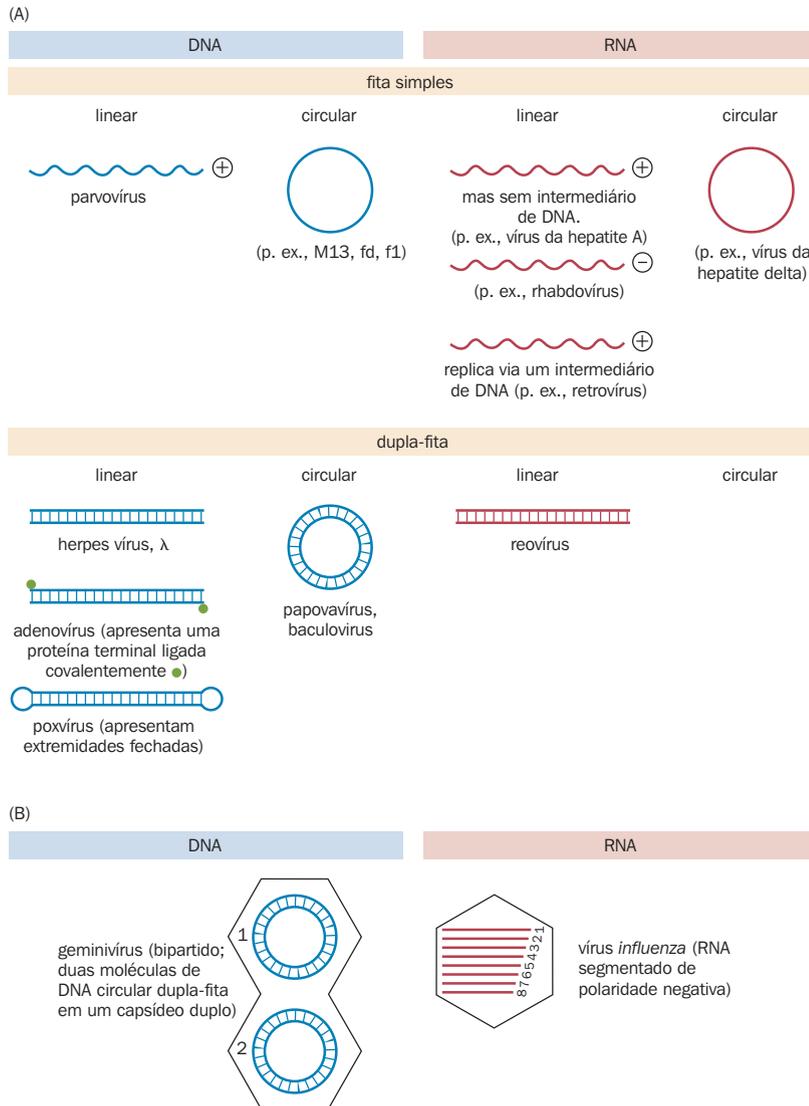
Os retrovírus são vírus de RNA incomuns, tanto por replicarem no núcleo como por utilizarem um intermediário de DNA. O genoma de RNA fita simples é convertido em um cDNA fita simples utilizando a transcriptase reversa viral. A DNA-polimerase da própria célula hospedeira é então utilizada para converter o cDNA viral em um DNA dupla-fita, o qual é posteriormente inserido no DNA cromossomal da célula hospedeira com auxílio de outras proteínas virais. Este DNA viral pode permanecer inativo por longos períodos ou ser utilizado para sintetizar novos genomas de RNA viral, os quais são empacotados em novas partículas infecciosas.

### 1.3 TRANSCRIÇÃO DO RNA E EXPRESSÃO GÊNICA

Além de seu envolvimento geral no armazenamento e na transmissão da informação genética, bem como no suporte às funções cromossômicas, o DNA ainda pode apresentar funções específicas em cada célula. Isso é possível porque o DNA contém sequências que podem ser utilizadas para produzir RNA e polipeptídeos de formas distintas de uma célula para outra. **Genes** são segmentos de DNA distribuídos em intervalos irregulares ao longo da sequência de DNA e que servem como molde para a produção de sequências de RNA complementar (transcrição). O transcrito inicial de RNA deve ser submetido a uma série de etapas de maturação que resultam em um RNA maduro. Este pode ser um RNA não codificante funcional ou um RNA mensageiro, que por sua vez servirá como molde para a produção de polipeptídeos. Alguns dos produtos gênicos são necessários para essencialmente todas as células, envolvidos em uma variedade de processos celulares vitais (tais como replicação do DNA e síntese de proteínas). Contudo, outros RNAs e proteínas são produzidos em apenas algumas células, podendo até mesmo ser específicos de uma única célula (em alguns casos excepcionais, como em linfócitos B e T).

A composição de DNA dos diferentes tipos celulares em organismos multicelulares é essencialmente idêntica. A variação entre células ocorre em função de diferenças na expressão gênica, primariamente no nível da transcrição: diferentes genes são transcritos em diferentes células, de acordo com as necessidades de cada célula. Alguns genes, conhecidos como genes “domésticos” (*housekeeping genes*) precisam ser expressos essencialmente em todas as células, enquanto outros genes apresentam um perfil de expressão tecido-específico ou são expressos em períodos específicos (p. ex., em estágios específicos do desenvolvimento ou do ciclo celular).

Normalmente, apenas uma das duas fitas do dúplice de DNA serve como molde para a síntese de RNA. Durante a transcrição, a RNA-polimerase se liga ao DNA dupla-fita, o qual é então desenrolado. Isso permite que a fita de DNA que servirá como molde

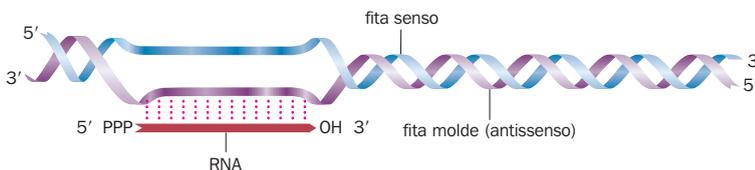


**Figura 1.12 A extraordinária variedade dos genomas virais.** (A) Número de fitas e topologia. Em genomas virais de fita simples, o RNA utilizado para produzir produtos proteicos pode apresentar a mesma orientação apresentada pelo genoma, sendo dito *genoma de polaridade positiva (+)*, ou apresentar orientação oposta ao genoma (antissenso), sendo, neste caso, referido como um *genoma de polaridade negativa (-)*. Alguns vírus de RNA fita simples (+) (p. ex., retrovírus) utilizam um intermediário de DNA e alguns vírus de DNA dupla-fita utilizam uma forma de RNA replicativo (p. ex., vírus da hepatite B). (B) Genomas segmentados e multipartidos. *Genomas segmentados* são aqueles nos quais o genoma se apresenta dividido em múltiplas moléculas de ácido nucleico, cada uma especificando um RNA capaz de produzir um único polipeptídeo. Por exemplo, o genoma de um vírus *influenza* apresenta-se dividido em oito moléculas diferentes de RNA fita simples com polaridade negativa. Em alguns genomas segmentados, cada uma das diferentes moléculas é empacotada em um compartimento separado (capsídeo). Tais genomas são descritos como *genomas multipartidos*, conforme ilustrado aqui pelo genoma bipartido do geminivírus.

para a síntese de RNA forme temporariamente um híbrido de RNA-DNA com a cadeia nascente de RNA.

O transcrito de RNA é complementar à fita molde do DNA, apresentando a mesma orientação 5' → 3' e a mesma sequência de bases (exceto que U substitui o T) da fita de DNA oposta, que não serviu como molde. Esta fita que não serviu como molde é frequentemente chamada de **fita senso**, e a fita molde é frequentemente chamada de **fita antis-senso** (Figura 1.13).

Quando se armazenam sequências gênicas, é comum apresentar apenas a sequência de DNA da fita senso. A orientação das sequências relativas a um dado gene também se refere normalmente à fita senso. Por exemplo, a extremidade 5' de um gene se refere às sequências na extremidade 5' da fita senso, e sequências a montante (*upstream*) ou a jusante (*downstream*) flanqueiam o gene em suas extremidades 5' e 3', respectivamente, com relação a fita *senso*. Para que a transcrição proceda de forma eficiente, várias proteí-



**Figura 1.13 O RNA transcrito é complementar a uma das fitas de DNA.** A sequência de nucleotídeos do RNA transcrito é normalmente idêntica àquela da fita senso, exceto que U substitui T, e é complementar à sequência da fita molde. O nucleotídeo da extremidade 5' do transcrito de RNA primário carrega um grupo 5' trifosfato que pode sofrer modificações posteriormente; a extremidade 3' apresenta um grupo hidroxila livre.

nas (fatores de transcrição) devem se ligar a elementos particulares da sequência de DNA (coletivamente chamados de **promotores**), os quais são com frequência localizados na vizinhança a montante dos genes. Os fatores de transcrição ligados a um promotor servem para posicionar e guiar a RNA-polimerase.

As RNA-polimerases sintetizam RNA a partir de quatro precursores: ATP, CTP, GTP e UTP. A elongação envolve a adição do resíduo apropriado de ribonucleosídeo monofosfato (AMP, CMP, GMP ou UMP) ao grupo 3' OH livre na extremidade da cadeia nascente de RNA. Esses nucleotídeos são gerados pela remoção de um resíduo de pirofosfato (PP<sub>i</sub>) do respectivo ribonucleosídeo trifosfato (rNTP) precursor. Apenas o nucleotídeo iniciador, na extremidade 5' do transcrito primário, carrega um grupo 5' trifosfato.

### A maioria dos genes é expresso para produzir polipeptídeos

A maioria dos genes eucarióticos é expressa para produzir polipeptídeos utilizando a RNA-polimerase II, uma das três RNA-polimerases (**Tabela 1.5**). Nenhuma das três RNA-polimerases pode iniciar a transcrição sozinha: elas precisam ser auxiliadas por fatores regulatórios. Um elemento regulatório crucial é o **promotor**, uma coleção de pequenos elementos na sequência de DNA, que se encontram agrupados na vizinhança do gene. Promotores são reconhecidos e ligados por fatores de transcrição, os quais irão guiar e ativar a polimerase. Fatores de transcrição são ditos **ativos em trans** (*trans-acting*) por serem produzidos por genes remotos e precisarem migrar até seus sítios de ação. Por outro lado, sequências promotoras são **ativas em cis** (*cis-acting*) por estarem localizadas na mesma molécula de DNA dos genes que elas regulam.

Promotores reconhecidos pela RNA-polimerase II em geral incluem os elementos a seguir:

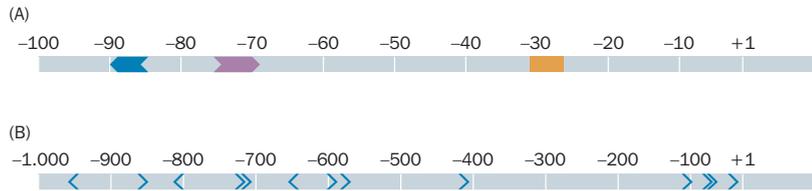
- **TATA box.** Este elemento frequentemente apresenta a sequência TATAAA, ou uma variação desta, sendo encontrado a cerca de 25 pares de bases (pb) a montante (*upstream*) do sítio de transcrição inicial (referido como -25; **Figura 1.14A**). Normalmente ocorre em genes que são ativamente transcritos pela RNA-polimerase II apenas em um estágio específico do ciclo celular (p. ex., genes de histona) ou em tipos celulares específicos (p. ex., gene da β-globina). Uma mutação no sítio TATA *box* não impede a iniciação da transcrição, mas provoca que a transcrição inicie em uma localização incorreta.
- **GC box.** Apresenta uma variante da sequência GGGCGG e ocorre em uma variedade de genes, muitos dos quais não apresentam o sítio TATA *box*. Este é o caso dos genes “domésticos” que realizam a mesma função em todas as células (como aqueles que codificam as polimerases, histonas e proteínas ribossômicas). Embora a sequência do sítio GC *box* seja assimétrica, ela parece funcionar em ambas as direções (**Figura 1.14B**).
- **CAAT box.** Este elemento frequentemente se localiza na posição -80, sendo muitas vezes o mais forte determinante da eficiência do promotor. Assim como o sítio GC *box*, ele funciona em ambas as direções.

Para um gene ser transcrito pela RNA-polimerase II ele precisa primeiro ser ligado por fatores de transcrição gerais para formar um complexo de pré-iniciação. Os fatores de transcrição gerais requeridos pela DNA-polimerase II incluem TFIIA, TFIIB, TFIID, TFIIE,

**TABELA 1.5** As três classes de RNA-polimerases eucarióticas

RNA-polimerase	RNA sintetizado	Notas
I	28S rRNA <sup>a</sup> , 18S rRNA <sup>a</sup> , 5,8S rRNA <sup>a</sup>	localizado no núcleo; a RNA-polimerase I produz um único transcrito primário (45S rRNA), o qual é clivado para gerar as três classes de rRNA listadas aqui
II	mRNA <sup>b</sup> , miRNA <sup>c</sup> , a maioria dos snRNAs <sup>d</sup> e snoRNAs <sup>e</sup>	os transcritos da RNA-polimerase II são especiais por serem sujeitos a incorporação do <i>cap</i> ( <i>capping</i> ) e da poliadenilação
III	5S rRNA <sup>a</sup> , tRNA <sup>f</sup> , U6 snRNA <sup>g</sup> , 7SL RNA <sup>h</sup> , vários outros pequenos RNAs não codificantes	para alguns genes transcritos pela RNA-polimerase III o promotor é interno ao gene (p. ex., 5S rRNA, tRNA, 7SL RNA); para outros, é localizado a montante do gene (ver Figura 1.15)

<sup>a</sup>RNA ribossômico. <sup>b</sup>RNA mensageiro. <sup>c</sup>Micro RNA. <sup>d</sup>Pequenos RNAs nucleares. <sup>e</sup>Pequenos RNAs nucleolares. <sup>f</sup>RNA transportador. <sup>g</sup>U6 snRNA é um componente do spliceossomo, um complexo RNA-proteína que remove sequências não codificantes indesejadas de transcritos de RNA recém-formados. <sup>h</sup>7SL RNA forma parte da partícula de reconhecimento de sinal, a qual apresenta um importante papel no transporte de proteínas recém-sintetizadas.



**Figura 1.14 Promotores para dois genes eucarióticos que codificam polipeptídeos.** Genes que codificam polipeptídeos são transcritos pela RNA-polimerase II. Os promotores são definidos por sequências curtas localizadas em sítios a montante do sítio de início da transcrição (+1). (A) O promotor do gene da  $\beta$ -globina inclui elementos TATA box (laranja), CAAT box (roxo) e GC box (azul). (B) O gene do receptor de glicocorticoide é incomum, apresentando 13 elementos GC box a montante de sua sequência: 10 na orientação normal e 3 em orientação inversa (orientações alternativas para os elementos GC box são indicadas pela direção das pontas de seta).

TFIIIF e TFIIF. Estes fatores de transcrição podem ainda ser formados por vários subcomponentes. Por exemplo, TFIID é constituído pela proteína de ligação ao TATA box, a TBP (do inglês *TATA-box-binding protein*, também encontrada em associação com as RNA-polimerases I e III), mais uma série de fatores associados à TBP (proteínas TAF). O complexo necessário para que uma RNA-polimerase inicie a transcrição é conhecido como **aparato basal de transcrição**, sendo constituído pela polimerase em associação com todos os seus fatores de transcrição gerais.

Além desses fatores de transcrição requeridos pela DNA-polimerase II, elementos específicos são reconhecidos por fatores de transcrição tecido-específicos. Por exemplo, um **acentuador** (*enhancer*) é um conjunto de elementos curtos, ativos em *cis*, que podem aumentar a atividade transcricional de um gene eucariótico específico. Diferente do promotor, que apresenta uma posição relativamente constante com relação ao sítio de início da transcrição, acentuadores são localizados a distâncias variáveis de seus sítios transcricionais. Além disso, sua função independe de sua orientação. No entanto, acentuadores também se ligam a proteínas regulatórias. O DNA entre o promotor e os sítios acentuadores se afasta formando uma alça, o que aproxima essas duas sequências, permitindo que proteínas ligadas ao acentuador interajam com os fatores de transcrição ligados ao promotor (ou à RNA-polimerase).

Um silenciador apresenta propriedades similares ao acentuador, mas em vez de aumentar, inibe a atividade transcricional de um gene específico.

### Diferentes conjuntos de RNA são transcritos pelas três RNA-polimerases eucarióticas

Genes que codificam polipeptídeos são sempre transcritos pela RNA-polimerase II. No entanto, genes que produzem RNAs não codificantes podem ser transcritos pelas polimerases I, II ou III, dependendo do tipo de RNA (ver Tabela 1.5). A RNA-polimerase I apresenta um comportamento incomum, visto que se dedica à transcrição de uma única unidade transcricional, gerando um transcrito grande que é posteriormente processado para originar três tipos de RNAs ribossomais (ver a seguir).

A RNA-polimerase II sintetiza vários tipos de pequenos RNAs não codificantes, além dos mRNAs. Eles incluem muitos tipos de pequenos RNAs nucleares (snRNA) e de pequenos RNAs nucleolares (snoRNA), os quais estão envolvidos em diferentes eventos de processamento de RNAs. Além disso, a RNA-polimerase II sintetiza muitos micro RNAs (miRNAs), os quais podem apresentar padrão de expressão tecido-específico e regular a expressão de importantes conjuntos de genes-alvo.

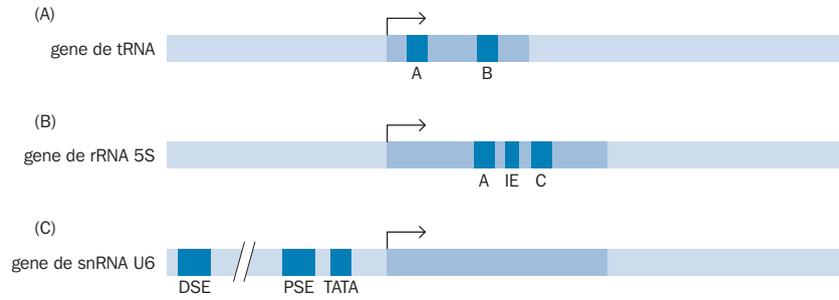
A RNA-polimerase III transcreve uma variedade de pequenos RNAs não codificantes que são tipicamente expressos em todas as células, incluindo as diferentes espécies de RNA transportador (tRNA), o RNA ribossomal (rRNA) 5S e alguns snRNAs. Os genes para tRNAs e rRNAs 5S apresentam uma organização incomum, onde o promotor se localiza no interior da sequência transcrita (Figura 1.15).

Promotores internos são possíveis uma vez que a função de um promotor é simplesmente atrair fatores de transcrição que irão guiar a RNA-polimerase ao sítio correto de início da transcrição. Uma vez que a polimerase esteja posicionada e pronta para iniciar a transcrição, quaisquer fatores de transcrição ligados a jusante do promotor serão removidos da fita molde. Como exemplo, a transcrição de um gene de tRNA inicia com os seguintes passos:

- TFIIC (inglês para “fator de transcrição para a polimerase IIIC”) se liga aos quadros A e B do promotor interno de um gene de tRNA (ver Figura 1.15).
- A ligação do TFIIC direciona a ligação de outro fator de transcrição, TFIIB, para uma posição a montante do sítio de início da transcrição; TFIIC não é mais necessário, tendo suas unidades removidas do promotor interno.
- TFIIB guia a RNA-polimerase III para se ligar ao sítio de início da transcrição.

**Figura 1.15 Elementos promotores em três genes transcritos pela RNA-polimerase III.**

(A) Genes de tRNA apresentam um promotor interno que consiste em um quadro A (localizado no braço D do tRNA; ver Figura 1.9B) e um quadro B que é comumente encontrado no braço TΨC. (B) O promotor do gene de rRNA 5S de *Xenopus* apresenta três componentes: um quadro A (+50 a +60), um elemento intermediário (IE; +67 a +72), e o quadro C (+80 a +90). (C) O gene humano de snRNA U6 apresenta um promotor externo, constituído por três componentes. Um elemento distal (DSE, do inglês *distal sequence element*; -240 a -215) acentua a transcrição e atua em conjunto com um promotor composto de um elemento proximal (PSE, do inglês *proximal sequence element*; -65 a -48) e uma TATA box (-32 a -25). As flechas indicam a posição +1.



**1.4 PROCESSAMENTO DE RNA**

Os transcritos de RNA da maioria dos genes eucarióticos são submetidos a uma série de reações de processamento para produzir um mRNA maduro ou um RNA não codificante.

**O *splicing* do RNA remove seqüências indesejadas dos transcritos primários**

Para a maioria dos genes de vertebrados – quase todos os genes codificadores de proteína e alguns genes de RNA –, apenas uma pequena porção da seqüência gênica é por fim decodificada para gerar um produto final. Nestes casos, as instruções genéticas para produzir um mRNA ou um RNA maduro não codificante ocorrem em segmentos denominados **éxons**. Esses segmentos são separados por seqüências intervenientes denominadas **íntrons**, que não contribuem com informação genética para o produto final.

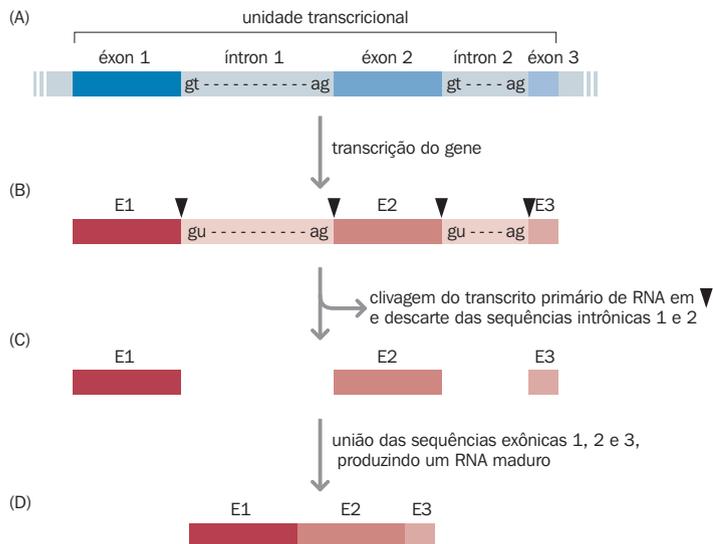
A transcrição de um gene produz inicialmente um transcrito primário de RNA que é complementar ao comprimento total do gene, incluindo tanto íntrons como éxons. Esse transcrito primário sofre então o processo denominado ***splicing* do RNA**, que se caracteriza por uma série de reações, por meio das quais os segmentos intrônicos são removidos e descartados, enquanto os segmentos de éxons remanescentes são encadeados em uma única seqüência curta de RNA (Figura 1.16).

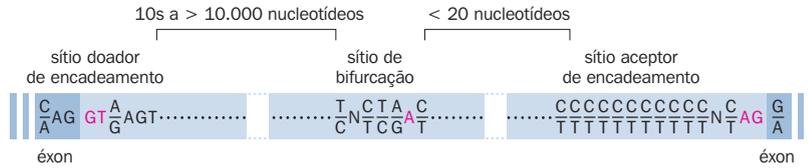
O *splicing* do RNA requer o reconhecimento de seqüências nucleotídicas no limite dos éxons e íntrons (junções de *splicing*). Os dinucleotídeos nas extremidades dos íntrons são altamente conservados: a grande maioria dos íntrons inicia com GT (que se torna GU no RNA intrônico) e termina com AG (**regra GT-AG**).

Embora os dinucleotídeos conservados GT e AG sejam cruciais para o *splicing*, eles não são suficientes para marcar os limites de um íntron. As seqüências nucleotídicas imediatamente adjacentes a eles também são altamente conservadas, constituindo as seqüências consenso das junções de *splicing* (Figura 1.17). Uma terceira seqüência intrônica conservada que também é importante para o *splicing* é conhecida como sítio de bifurcação (*branch site*) e se localiza normalmente a até 40 nucleotídeos a montante do

**Figura 1.16 O processo de *splicing* do RNA.**

Neste exemplo, o gene contém três éxons e dois íntrons. (B) O transcrito primário de RNA é uma cópia contínua do gene e contém seqüências transcritas a partir dos éxons (E1, E2 e E3) e dos íntrons. (C) O transcrito primário é clivado nas regiões correspondentes aos limites éxon-íntron (junções de *splicing*). As porções referentes aos íntrons, na cópia de RNA, são removidas e descartadas. (D) As cópias de RNA dos éxons são mantidas e finalmente unidas, mantendo a mesma ordem linear da seqüência de DNA genômico.





terminal 5' AG dos íntrons (ver Figura 1.17). Outras sequências exônicas e intrônicas podem promover o *splicing* (sequências acentuadoras de *splicing*) ou inibi-lo (sequências silenciadoras de *splicing*), e mutações nessas sequências podem causar doenças.

Os passos essenciais ao *splicing* são os seguintes:

- Ataque nucleofílico do nucleotídeo terminal 5' G pelo A invariante da sequência consenso do sítio de bifurcação, formando uma estrutura em forma de laço.
- Clivagem da junção de *splicing* no sítio doador de encadeamento.
- Ataque nucleofílico pela extremidade 3' do éxon a montante do sítio receptor de *splicing*, levando à clivagem e à liberação de um RNA intrônico em forma de laço, e o *splicing* de dois segmentos de RNA (Figura 1.18).

Para genes que residem no núcleo eucariótico, o *splicing* do RNA é mediado por um grande complexo RNA-proteína, denominado spliceossomo. Spliceossomos apresentam cinco tipos de snRNA (pequenos RNAs nucleares) e mais de 50 proteínas. As moléculas de snRNA associam-se a proteínas para formar pequenas partículas de ribonucleoproteínas nucleares (snRNP ou *snurp*). A especificidade da reação de encadeamento é estabelecida pelo pareamento de bases entre o transcrito de RNA a ser encadeado e as moléculas de snRNA do spliceossomo. Existem dois tipos de spliceossomo:

- O *spliceossomo principal* (GU-AG) processa transcritos correspondentes aos íntrons clássicos GT-AG. Ele contém cinco tipos de snRNA. Os snRNAs U1 e U2 reconhecem e se ligam ao sítio doador de *splicing* e ao sítio de bifurcação, respectivamente. Os snRNAs U4, U5 e U6 ligam-se posteriormente, induzindo o RNA intrônico a se protraírem na forma de uma alça.
- O *spliceossomo secundário* (AU-AC) processa transcritos correspondentes a raros íntrons AU-AC. Ele também apresenta cinco snRNAs mas utiliza U11 e U12 em vez de U1 e U2, além de apresentar variantes dos snRNAs U4 e U6.

Uma vez que o spliceossomo reconhece um sítio doador, ele percorre a sequência de RNA até encontrar o próximo sítio receptor de *splicing* (sinalizado como um alvo pela presença a montante da sequência consenso do sítio de bifurcação).

## Nucleotídeos especializados são adicionados ao final da maioria dos transcritos de RNA-polimerase II

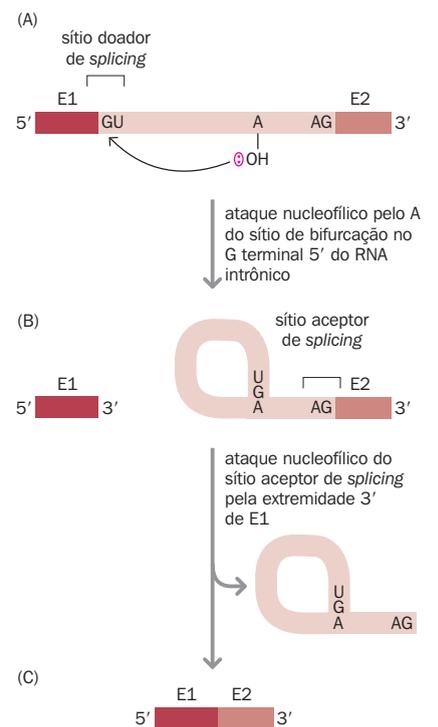
Adicionalmente ao *splicing* do RNA, as extremidades dos transcritos de RNA-polimerase II sofrem modificações. A extremidade 5' é protegida pela adição de uma estrutura referida como *cap*, tratando-se basicamente de uma variante de guanina com uma ligação fosfodiéster incomum. Além disso, uma longa sequência de adeninas é adicionada à extremidade de 3' dos transcritos. Essas modificações servem tanto para proteger essas extremidades frente a ação de exonucleases celulares como para auxiliar no correto funcionamento dos transcritos de RNA.

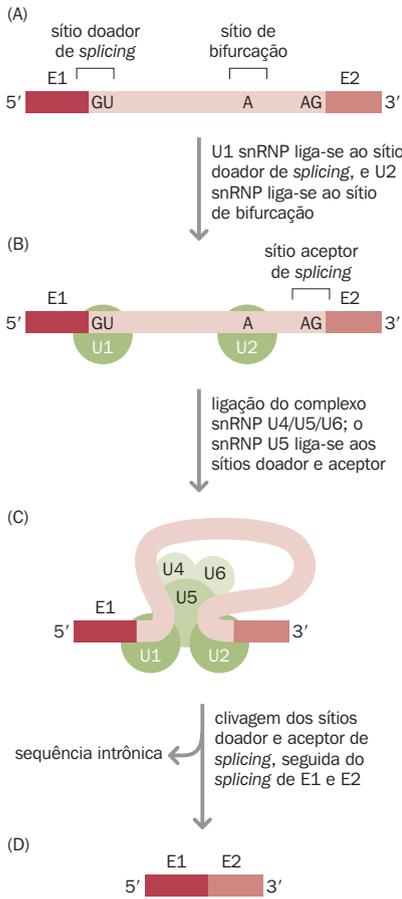
### Capecamento 5'

Logo após o início da síntese do transcrito primário de RNA que se tornará um mRNA, um nucleosídeo metilado (7-metilguanósina, m<sup>7</sup>G) é ligado por uma ligação 5'-5' fosfodiéster

**Figura 1.18 Os mecanismos do encadeamento de RNA.** (A) Transcrito primário de RNA não processado, contendo RNA intrônico entre as sequências E1 e E2 que correspondem aos éxons do DNA. O mecanismo de *splicing* envolve um ataque nucleofílico no G do dinucleotídeo 5' GU. Isso é realizado pelo grupo 2' OH da alanina conservada no sítio de bifurcação, resultando na formação de uma estrutura em forma de laço (B) e na clivagem do sítio doador de *splicing*. O grupo 3' OH na extremidade 3' da sequência E1 realiza um ataque nucleofílico no sítio receptor, causando a liberação do RNA intrônico (como uma estrutura em forma de laço) e (C) a fusão (*splicing*) de E1 e E2.

**Figura 1.17 Três sequências consenso apresentadas por íntrons de eucariotos complexos.** A maioria dos íntrons em genes eucarióticos contém sequências conservadas que correspondem a três regiões funcionalmente importantes. Duas destas regiões, sítio doador de encadeamento e sítio receptor de encadeamento, demarcam os limites 3' e 5' dos íntrons. O sítio de bifurcação (*branch site*) é uma região adicional importante que ocorre tipicamente a menos de 20 nucleotídeos a montante do sítio receptor de encadeamento. Os nucleotídeos apresentados em vermelho nestas três sequências consenso são praticamente invariantes. Os demais nucleotídeos apresentados tanto no íntron como nos éxons são os mais comumente encontrados em cada posição. Em alguns casos, dois nucleotídeos podem ser igualmente frequentes, como no caso do C e do T próximos a extremidade 3' do íntron. O "N" indica a ocorrência de qualquer um dos quatro nucleotídeos.





**Figura 1.19 Papel das pequenas ribonucleoproteínas nucleares (snRNPs) no splicing do RNA.** (A) Transcrito primário de RNA não processado, como na Figura 1.18. (B) No spliceossomo, parte do snRNA U1 apresenta sequência complementar à sequência consenso do sítio doador de *splicing*. Como resultado, o complexo snRNA U1-proteína (snRNP U1) liga-se à junção de *splicing* pelo pareamento de bases RNA-RNA. O complexo snRNP U2, de modo similar, liga-se ao sítio de bifurcação pelo pareamento de bases RNA-RNA. A interação entre os sítios doador e receptor de *splicing* é estabilizada pela (C) ligação de uma partícula multi-snRNP, a qual contém os snRNAs U4, U5 e U6. O snRNP U5 liga-se simultaneamente a ambos os sítios de *splicing* (doador e receptor). A subsequente clivagem do complexo gerado libera a sequência intrônica e permite (D) que as sequências E1 e E2 sofram o *splicing*.

ao primeiro nucleotídeo 5'. Esse processo é referido como **capeamento** (*capping*) da extremidade 5' do transcrito (**Figura 1.20**); os *caps* dos transcritos de snRNA podem sofrer uma modificação adicional. O *cap* 5' pode apresentar várias funções:

- proteger o transcrito do ataque de exonucleases 5' → 3' (moléculas de mRNA sem o *cap* são rapidamente degradadas);
- facilitar o transporte dos mRNAs do núcleo para o citoplasma;
- facilitar o encadeamento do RNA;
- facilitar a ligação da subunidade 40S dos ribossomos citoplasmáticos ao mRNA durante a tradução.

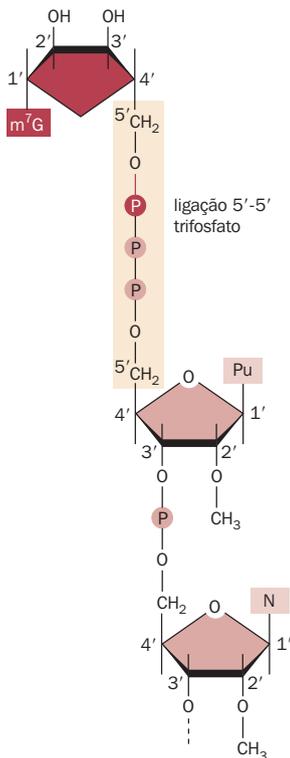
### Poliadenilação 3'

A transcrição por ambas as RNA-polimerases I e III é interrompida após o reconhecimento de um sítio específico de terminação da transcrição. No entanto, as extremidades 3' das moléculas de mRNA são determinadas por uma reação de clivagem pós-transcricional. A sequência AAUAAA (às vezes AUUAAA) sinaliza a clivagem 3' para a maioria dos transcritos da polimerase II.

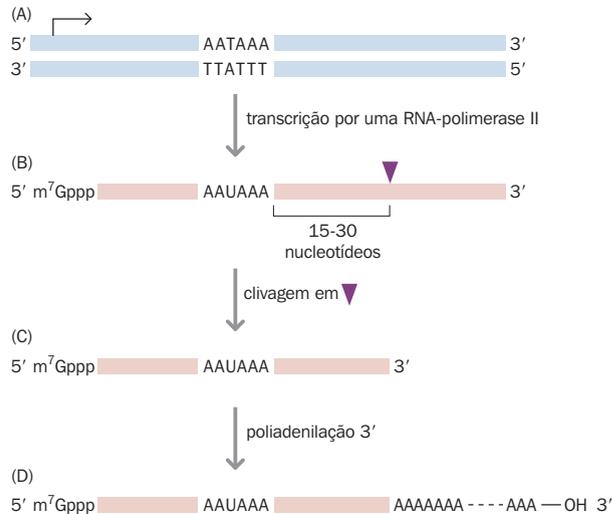
A clivagem ocorre em um sítio específico, de 15 a 30 nucleotídeos a jusante da sequência AAUAAA, embora o transcrito primário possa continuar por centenas e até mesmo por milhares de nucleotídeos após o ponto de clivagem. Subsequente à clivagem, a enzima poli(A) polimerase adiciona sequencialmente resíduos de adenilato (AMP) à extremidade 3' (cerca de 200 no caso dos mRNAs de mamíferos). Essa reação de poliadenilação (**Figura 1.21**) produz uma **cauda poli(A)**, a qual se acredita estar envolvida com as seguintes atividades:

- Auxílio no transporte do mRNA para o citoplasma.
- Estabilização de pelo menos algumas moléculas de mRNA no citoplasma.
- Aumento do reconhecimento do mRNA pela maquinaria ribossômica.

Genes de histona são os únicos a produzirem um mRNA que não se torna poliadenilado; a terminação de sua transcrição, todavia, também envolve a clivagem 3' do transcrito primário.



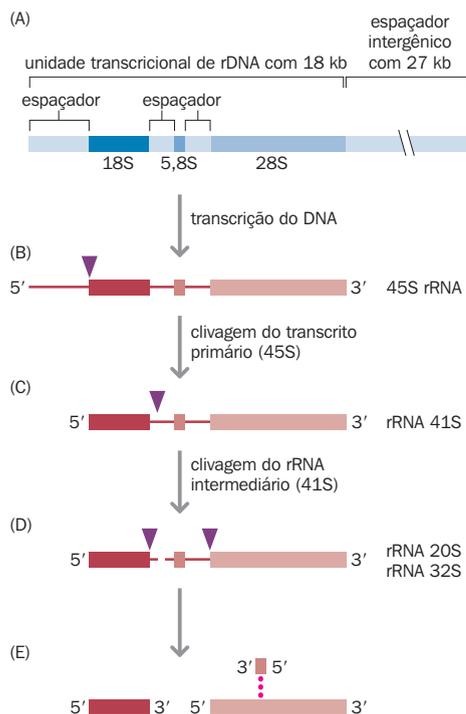
**Figura 1.20 O cap 5' do mRNA eucariótico.** Os componentes nucleotídicos em rosa representam o resíduo da extremidade 5' original de um pré-mRNA eucariótico. O transcrito primário de pré-mRNA inicia com um nucleotídeo que contém uma purina (Pu) e um grupo 5' trifosfato. No entanto, como o pré-mRNA sofre processamento, o grupo fosfato terminal da extremidade 5' é excisado com uma fosfatase, gerando um grupo 5' difosfato ao qual um nucleotídeo especializado é ligado de forma covalente, formando uma estrutura referida como *cap* 5'. Esta estrutura irá proteger o mRNA frente ao ataque de exonucleases e auxiliar no início da tradução. O nucleotídeo do *cap* (com a base representada em vermelho) é inicialmente formado quando um resíduo GTP é clivado para gerar uma guanosina monofosfato, que é então adicionada ao grupo difosfato da purina terminal original, por meio de uma *ligação 5'-5' trifosfato* (sombreado em pêssego claro). Subsequentemente, o átomo de nitrogênio número 7 da nova extremidade 5' G é metilado. Nos mRNAs sintetizados em células de vertebrados, o carbono 2' da ribose de cada par de nucleotídeos adjacentes (a purina terminal original e o nucleotídeo vizinho) também são metilados, conforme ilustrado neste exemplo. m<sup>7</sup>G, 7-metilguanossina; N, qualquer nucleotídeo.



## Os transcritos de rRNA e tRNA sofrem um extenso processamento

Quatro classes de rRNA eucariótico foram identificadas: 28S, 18S, 5,8S e 5S (S é o coeficiente de sedimentação de Svedberg, uma medida da velocidade de sedimentação de grandes estruturas moleculares mediante ultracentrifugação, correspondendo diretamente ao tamanho e à forma). O rRNA 18S é encontrado nas subunidades menores dos ribossomos, enquanto os outros três são componentes das subunidades maiores. As células exigem grandes quantidades de rRNA para realizar a síntese de proteínas e, assim sendo, vários genes são dedicados a produzir rRNA no nucléolo (um compartimento visivelmente distinto do núcleo).

Em células humanas um conjunto de aproximadamente 250 genes sintetizam rRNA 5S utilizando a RNA-polimerase III, a qual também transcreve algumas outras espécies de pequenos RNAs. Os rRNAs 28S, 18S e 5,8S são codificados por genes consecutivos em uma unidade transcricional única de 13 kb (**Figura 1.22**) que é transcrita pela RNA-polimerase I. Uma unidade composta de um transcrito de 13 kb e um espaçador adjacente não transcrito, com 27 kb, é repetida sequencialmente entre 30 a 40 vezes nas regiões nucleolares



**Figura 1.21 Poliadenilação da extremidade 3' dos mRNAs eucarióticos.** (A, B) Conforme avança na transcrição de um gene, a RNA-polimerase carrega consigo dois complexos multiproteicos necessários para a poliadenilação: o fator de especificidade de clivagem e poliadenilação (CPSF, do inglês *cleavage and polyadenylation specificity factor*) e o fator de estimulação à clivagem (CStF, do inglês *cleavage and stimulation factor*). Esses fatores cooperam para identificar o sinal de poliadenilação, a jusante do códon de terminação no transcrito de RNA, bem como para cortar o transcrito. O sinal de poliadenilação inclui uma sequência AAUAAA, ou uma variante desta, além de alguns sinais a jusante que ainda permanecem pouco compreendidos. (C) A clivagem normalmente ocorre cerca de 15 a 30 nucleotídeos a jusante do elemento AAUAAA. (D) Resíduos de AMP são subsequentemente adicionados por uma poli(A) polimerase para formar uma cauda poli(A).

**Figura 1.22 As principais espécies de rRNA são sintetizadas por meio da clivagem de um mesmo transcrito primário.** (A) Em células humanas, os rRNAs 18S, 5,8S e 28S são codificados por uma única unidade transcricional com 13 kb de comprimento. Ela ocorre no interior de repetições adjacentes (em *tandem*) com cerca de 40 kb, as quais também incluem um espaçador não transcrito (intergênico) com aproximadamente 27 kb. (B) A transcrição pela RNA-polimerase I produz um transcrito primário com 13 kb (rRNA 45S), o qual posteriormente sofre uma complexa série de clivagens pós-transcricionais. (C-E) Finalmente, as moléculas individuais de rRNAs 18S, 28S e 5,8S são liberadas. O rRNA 18S fará parte da subunidade menor do ribossomo. O rRNA 5,8S se liga a um segmento complementar do rRNA 28S, e o complexo resultante fará parte da subunidade maior do ribossomo. Essa última também contém o rRNA 5S, o qual é codificado separadamente por genes específicos transcritos pela RNA-polimerase III.

organizadoras dos braços curtos de cada um dos cinco cromossomos humanos acrocêntricos (13, 14, 15, 21 e 22). Estes cinco conjuntos de genes rRNA, cada um com cerca de 1,5 milhões de bases (Mb), são algumas vezes referidos como **DNA ribossomal (rDNA)**.

Além da sequência de reações de clivagem (ver Figura 1.22), o transcrito primário de rRNA também sofre uma variedade de modificações base-específicas. Este extenso processamento de RNA é conduzido por vários pequenos RNAs nucleolares distintos, os quais são codificados por cerca de 200 genes diferentes no genoma humano.

As moléculas maduras de tRNA também sofrem extensas modificações nas bases, e cerca de 10% das bases de qualquer tRNA são versões modificadas de A, C, G ou U. Exemplos comuns de nucleosídeos modificados incluem dihidrouridina, a qual apresenta hidrogênios extras nos carbonos 5 e 6; pseudouridina, um isômero da uridina; inosina (guanina deaminada); e *N,N'*-dimetilguanina.

## 1.5 TRADUÇÃO, PROCESSAMENTO PÓS-TRADUCIONAL E ESTRUTURA PROTEICA

O mRNA produzido pelos genes no núcleo migra para o citoplasma, onde se liga aos ribossomos e a outros componentes para iniciar a tradução e a síntese polipeptídica. O RNA mensageiro transcrito a partir de genes das mitocôndrias e cloroplastos é traduzido em ribossomos específicos dentro dessas organelas.

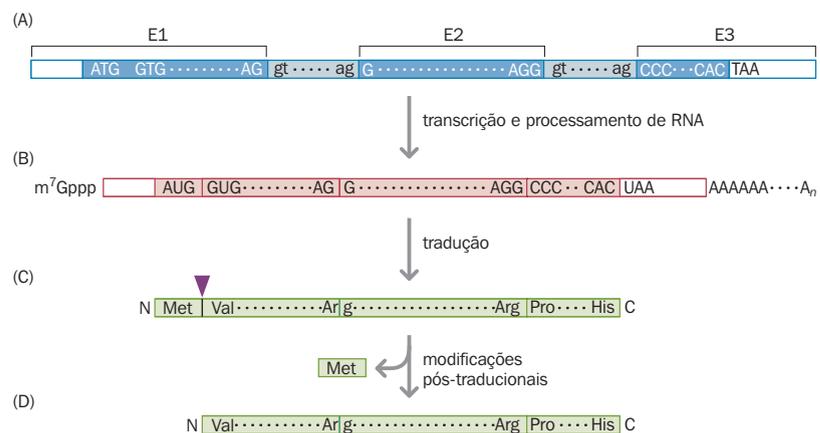
Apenas um segmento central da molécula de mRNA eucariótica é traduzido para formar um polipeptídeo. As **regiões não traduzidas** flangeadoras (5' UTR e 3' UTR, do inglês *untranslated region*) são transcritas de sequências de éxons presentes nas extremidades 5' e 3' dos genes. Eles ajudam na ligação e na estabilização do mRNA nos ribossomos, e promovem traduções eficientes (Figura 1.23).

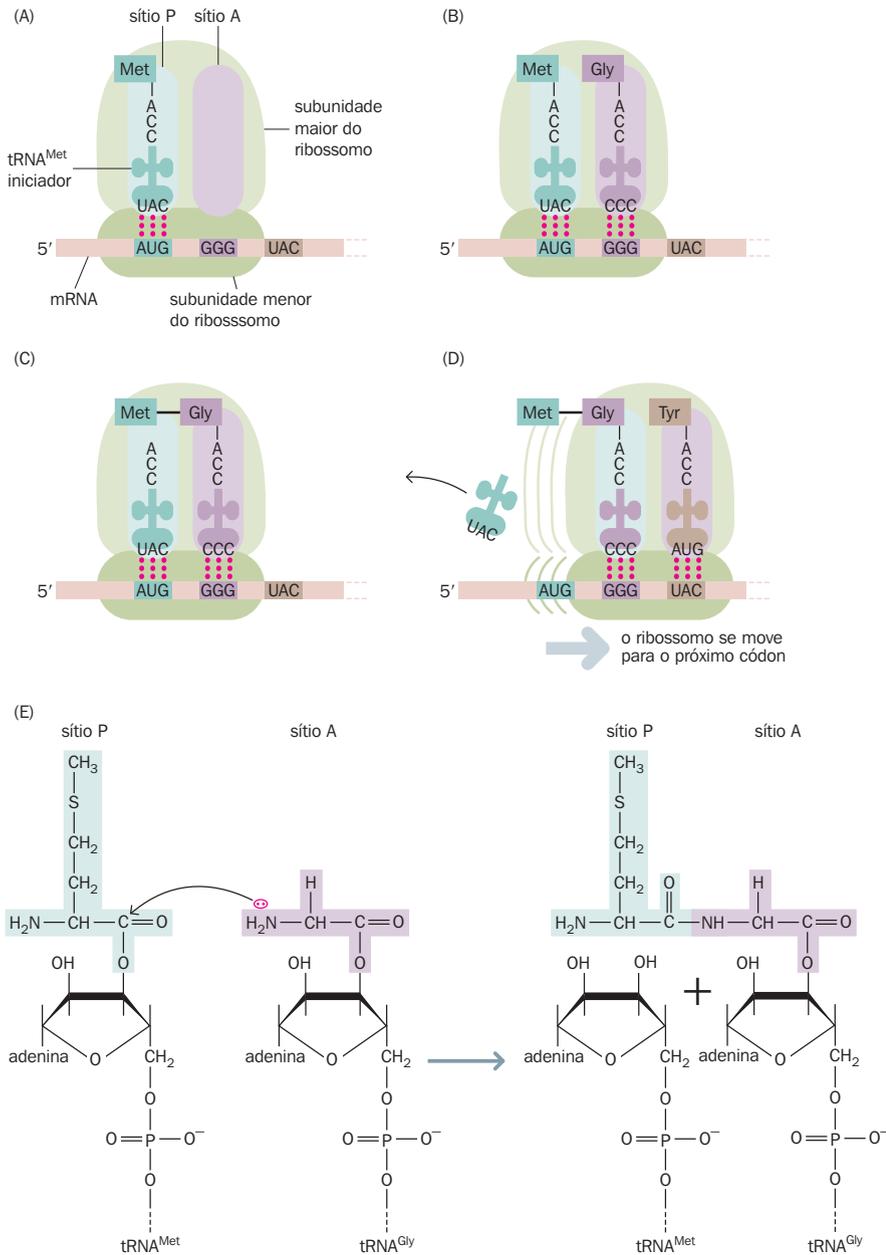
Os ribossomos são grandes complexos RNA-proteína compostos por duas subunidades. Nos eucariotos, ribossomos citoplasmáticos têm uma grande subunidade 60S e uma menor subunidade 40S. A subunidade 60S contém três tipos de moléculas de rRNA: rRNA 28S, rRNA 5,8 e rRNA 5S, bem como cerca de 50 proteínas ribossomais. A subunidade 40S contém um único rRNA 18S e mais de 30 proteínas ribossomais. Os ribossomos fornecem a maquinaria necessária para a síntese polipeptídica. Os RNAs que compõem esse complexo são predominantemente responsáveis pela função catalítica dos ribossomos; as proteínas que compõem esse complexo parecem aumentar a função das moléculas de rRNA, embora um surpreendente número delas não pareça ser essencial para a função do ribossomo.

### O mRNA é decodificado para polipeptídeos específicos

A montagem de um novo polipeptídeo a partir de seus aminoácidos constituintes é governada por um código genético em trincas (ou *triplets*). Em um mRNA no ribossomo, a sequência nucleotídica central utilizada para produzir o polipeptídeo é percorrida no sentido 5' → 3' em grupos de 3 nucleotídeos (**códons**). Cada códon especifica um aminoácido, e o processo de decodificação utiliza uma coleção de diferentes moléculas de tRNA, cada uma ligando-se a um tipo de aminoácido. Um complexo tRNA-aminoácido é conhecido como um aminoacil tRNA, sendo formado quando uma aminoacil tRNA sintetase especí-

**Figura 1.23** Transcrição e tradução do gene da  $\beta$ -globina humana. (A) O gene da  $\beta$ -globina humana possui três éxons (E1-E3) e dois íntrons. A sequência 5' terminal de E1 e a sequência 3' terminal de E3 são não codificantes (seções não sombreadas). (B) Estas sequências são transcritas e, portanto, ocorrem nas extremidades 5' e 3' (seções não sombreadas) do mRNA de  $\beta$ -globina que resulta do processamento de RNA. No entanto, elas não são traduzidas e não especificam nenhuma parte do polipeptídeo precursor (C). Esta figura também ilustra que alguns códons podem ser especificados por bases que estão separadas por um íntron. A arginina na posição 104 no polipeptídeo de  $\beta$ -globina é codificada pelos últimos três nucleotídeos do éxon 2 (AGG). A arginina na posição 30 é codificada por um códon AGG cujas duas primeiras bases se encontram nas últimas duas posições do éxon 1 e cuja terceira base se encontra na primeira posição do éxon 2. (D) Durante as modificações pós-traducionais, o polipeptídeo precursor (com 147 aminoácidos) sofre clivagem para remover seu resíduo de metionina N-terminal, gerando a proteína madura de  $\beta$ -globina (com 146 resíduos). Os símbolos N e C representados nas imagens (C) e (D) indicam, respectivamente, as extremidades N-terminal e C-terminal.





**Figura 1.24** Durante a tradução, o código genético é decifrado nos ribossomos pelo reconhecimento códon-anticódon. (A) A subunidade maior do ribossomo (60S em eucariotos) apresenta dois sítios para a ligação de um aminoacil tRNA (um RNA transportador com um aminoácido anexado): o sítio P (peptidil) e o sítio A (aminoacil). A subunidade menor do ribossomo (40S nos eucariotos) se liga ao mRNA, o qual é percorrido ao longo da região 5' UTR em sentido 5' → 3' até que o códon de iniciação seja identificado, um AUG localizado no interior de uma grande sequência conservada (ver texto). Um tRNA<sup>Met</sup> carregando um resíduo de metionina liga-se ao sítio P com seu anticódon "em registro" com o códon iniciador AUG. (B) O aminoacil tRNA apropriado se liga ao sítio A por meio do pareamento de bases entre seu anticódon e o próximo códon (neste caso, GGG, especificando uma glicina). (C) O rRNA na subunidade maior catalisa a formação da ligação peptídica, resultando na liberação do tRNA que carregava a metionina e na ligação desta última à glicina anexada ao tRNA que se encontra no sítio A. (D) O ribossomo se desloca ao longo do mRNA de modo que o tRNA carregando o dipeptídeo Met-Gly passa a ocupar o sítio P. O próximo aminoacil tRNA (neste caso, carregando uma tirosina) se liga ao sítio A, em preparação para a formação de uma nova ligação peptídica. O átomo de nitrogênio (N) do grupo amino, pertencente ao aminoácido ligado ao tRNA no sítio A, realiza um ataque nucleofílico ao átomo de carbono do grupo carboxila, pertencente ao aminoácido anexado ao tRNA, que se encontra no sítio P.

fica liga covalentemente o aminoácido requerido ao terminal adenosina do trinucleotídeo consensual CCA, na extremidade 3' do tRNA.

Cada tRNA possui seu próprio **anticódon**, um trinucleotídeo no centro do braço do anticódon (ver Figura 1.9B), que fornece a especificidade necessária para interpretar o código genético. Para que um aminoácido seja adicionado a um polipeptídeo crescente, o códon relevante da molécula mRNA precisa ser reconhecido pelo pareamento de bases com o anticódon complementar na molécula apropriada de aminoacil tRNA. Esse processo acontece no ribossomo. A subunidade menor do ribossomo liga-se ao mRNA, enquanto a subunidade maior apresenta dois sítios para a ligação de aminoacil tRNAs, referidos como sítio P (peptidil) e sítio A (aminoacil) (Figura 1.24).

O *cap* na extremidade 5' da molécula de RNA mensageiro é importante para iniciar a tradução. Ele é reconhecido por certas proteínas-chave que se ligam à subunidade menor do ribossomo, de modo que esses fatores de iniciação mantêm o RNA na posição. Na iniciação da tradução dependente de *cap*, o ribossomo percorre a extremidade 5' UTR do mRNA na direção 5' → 3' a fim de encontrar um códon de iniciação compatível, um AUG que é encontrado dentro da sequência consenso de Kozak 5'-GCCPuCCAUGG-3' (onde

*Pu* representa uma purina). Os determinantes mais importantes são o G na posição +4 (imediatamente após o AUG) e a purina (preferencialmente A) na posição -3 (3 nucleotídeos a montante do códon AUG).

Quando um códon de iniciação compatível é identificado, um tRNA<sup>Met</sup> iniciador carregando uma metionina se liga ao sítio P na subunidade maior do ribossomo. Isso ocorre de modo que o anticódon do tRNA<sup>Met</sup> realize o pareamento de bases com o códon iniciador AUG no mRNA (ver Figura 1.24). Uma vez que isso tenha acontecido, a fase de leitura transcricional é estabelecida e os códons são interpretados como grupos sucessivos de três nucleotídeos, continuando na direção 5' → 3' a jusante do códon de iniciação AUG. Um aminoacil tRNA para o segundo códon (um tRNA<sup>Gly</sup> para reconhecer o códon GGG, no exemplo da Figura 1.24) se liga ao sítio vizinho A na subunidade maior.

Uma vez que os sítios P e A estejam ocupados por aminoacil tRNAs, se acredita que o maior rRNA presente na subunidade maior do ribossomo atue como uma peptidil transferase. Ela catalisa a formação da ligação peptídica por uma reação de condensação entre o grupo amino do aminoácido preso pelo tRNA no sítio A e o grupo carboxila da metionina presa pelo tRNA<sup>Met</sup>. Essa atividade resulta na liberação da metionina iniciadora de seu tRNA e na sua ligação ao segundo aminoácido, formando um dipeptídeo (ver Figura 1.24). Uma vez liberado do aminoácido, o tRNA<sup>Met</sup> migra para fora do sítio P, enquanto seu lugar é tomado pelo tRNA ligado ao dipeptídeo que inicialmente ocupava o sítio A. O sítio A liberado é agora preenchido por um aminoacil tRNA carregando um anticódon que é complementar ao terceiro códon. Uma nova ligação peptídica é formada para fazer um tripeptídeo, e assim sucessivamente.

Após a iniciação da tradução de um mRNA por um ribossomo, e sua movimentação ao longo do mRNA, outros ribossomos podem se ligar ao mesmo mRNA. As estruturas polirribossômicas resultantes (polissomos) realizam múltiplas cópias de um polipeptídeo a partir da mesma molécula de mRNA. A elongação da cadeia polipeptídica ocorre até que um códon de terminação seja encontrado. Para mRNAs transcritos a partir de genes nucleares, códons de terminação existem em três variedades: UUA (ocre), UAG (âmbar) e UGA (opala), mas existem algumas diferenças para mRNA mitocondrial conforme descrito na próxima seção.

Em resposta a um códon de terminação, um fator de liberação da proteína entra no sítio A em vez de um aminoacil tRNA, sinalizando que o polipeptídeo deve desligar-se do ribossomo. O polipeptídeo completo irá então sofrer um processamento que pode incluir clivagem e modificação das cadeias laterais. Sua cadeia principal apresentará um grupo amino livre em uma extremidade (extremidade N-terminal) e um grupo carboxila livre na outra extremidade (extremidade C-terminal).

### O código genético é degenerado e não completamente universal

O código genético é um código de trincas, havendo quatro bases possíveis para ocupar cada uma das três posições em um códon. Existem, portanto,  $4^3 = 64$  códons possíveis, o que é mais do que suficiente para codificar os 20 principais tipos de aminoácido. O código genético é degenerado porque, em média, cada aminoácido é especificado por cerca de três códons diferentes. Alguns aminoácidos (como leucina, serina e arginina) chegam a

**Figura 1.25 O código genético.** Todos os 64 possíveis códons do código genético e o aminoácido especificado por cada um, conforme leitura no sentido 5' → 3' da sequência de mRNA. Interpretações dos 64 códons no código genético "universal" são apresentadas em branco, imediatamente à direita dos códons. Sessenta e um códons especificam aminoácidos. Três códons de parada (UAA, UAG e UGA), também referidos como códons de parada, não codificam aminoácido algum. O código genético para o DNA mitocondrial (mtDNA) obedece o código universal, com exceção de algumas poucas variações. Por exemplo, no código genético mitocondrial (em humanos e várias outras espécies) quatro códons são utilizados de forma distinta: UGA codifica triptofano em vez de ser um códon de parada, AUA codifica uma metionina e, em vez de codificar uma arginina, AGA e AGG são códons de parada.

variantes no mtDNA						variantes no mtDNA	
AAA	Lys	CAA	Gln	GAA	Glu	UAA	Parada
AAG		CAG		GAG		UAG	
AAC	Asn	CAC	His	GAC	Asp	UAC	Tyr
AAU		CAU		GAU		UAU	
ACA	Thr	CCA	Pro	GCA	Ala	UCA	Ser
ACG		CCG		GCG		UCG	
ACC		CCC		GCC		UCC	
ACU		CCU		GCU		UCU	
Parada	AGA	CGA	GGA	UGA	Parada	Trp	
	AGG	CGG	GGG	UGG	Trp		
	AGC	CGC	GGC	UGC	Cys		
	AGU	CGU	GGU	UGU			
Met	AUA	CUA	GUA	UUA	Leu		
	AUG	CUG	GUG	UUG			
	AUC	CUC	GUC	UUC	Phe		
	AUU	CUU	GUU	UUU			

ser especificados por até seis códons, enquanto outros são menos representados (Figura 1.25). A degeneração do código genético geralmente envolve a terceira base do códon.

Embora mais de 60 códons possam especificar um aminoácido, o número de diferentes moléculas citoplasmáticas de tRNA é um tanto quanto menor, e apenas 22 tipos de tRNA mitocondrial são produzidos. A interpretação para a existência de mais de 60 códons e um número muito menor de tRNAs é possível pois o pareamento de bases no RNA é mais flexível do que no DNA. O pareamento de códon e anticódon segue a regra normal A-U e G-C para as duas primeiras posições do códon. No entanto, na terceira posição existe certa flexibilidade (*base wobble*), e o pareamento de bases G-U é permitido (Tabela 1.6).

O código genético é o mesmo por meio de praticamente todas as formas de vida. No entanto, mitocôndrias e cloroplastos apresentam uma capacidade limitada para a síntese proteica, e durante a evolução seu código genético divergiu levemente daquele utilizado nos ribossomos citoplasmáticos. A tradução de mRNAs codificados pelo núcleo continua até que um dos três códons de parada (*stop codons*) seja encontrado (UAA, UAG ou UGA), mas nas mitocôndrias de mamíferos existem quatro possibilidades (UAA, UAG, AGA e AGG).

O significado de um códon também pode depender do contexto da sequência, ou seja, da natureza da sequência de nucleotídeos na qual ela está contida. Dependendo das sequências vizinhas, alguns códons de alguns poucos tipos de mRNAs codificados pelo núcleo podem ser interpretados de forma não convencional. Por exemplo, em uma ampla variedade de células o códon de parada UGA é alternativamente interpretado como codificando uma seleno-cisteína em alguns mRNAs de genes nucleares, enquanto o UAG pode, em alguns casos, ser interpretado como codificando uma glutamina.

## Processamento pós-traducional: modificações químicas de aminoácidos e clivagem polipeptídica

Os produtos primários da tradução muitas vezes sofrem uma variedade de modificações durante ou após a tradução. Grupos químicos simples ou complexos são frequentemente adicionados covalentemente às cadeias laterais de certos aminoácidos (Tabela 1.7). Além disso, polipeptídeos podem ocasionalmente ser clivados para gerar um ou mais produtos polipeptídicos.

### Adição de carboidratos

Glicoproteínas possuem oligossacarídeos covalentemente ligados às cadeias laterais de certos aminoácidos. Poucas proteínas no citosol são glicosiladas (carregam um carboidrato). Caso sejam, elas apresentam um único resíduo de açúcar, *N*-acetil-glicosamina, ligado a uma serina ou a um resíduo de treonina. No entanto, proteínas que são secretadas pelas células, bem como aquelas transportadas para os lisossomos, para o aparelho de Golgi, ou

**TABELA 1.6** Regras para o pareamento de bases podem ser relaxadas (*wobble*) na posição 3 de um códon

Base na extremidade 5' do anticódon no tRNA	Base reconhecida na extremidade 3' do códon no mRNA
A	apenas U
C	apenas G
G (ou I) <sup>a</sup>	C ou U
U	A ou G

<sup>a</sup> Inosina (I) é uma forma deaminada de guanosina.

**TABELA 1.7** Principais tipos de modificação dos polipeptídeos

Tipo de modificação (grupo adicionado)	Aminoácido(s) alvo	Notas
Fosforilação (PO <sub>4</sub> <sup>-</sup> )	Tyr, Ser, Thr	realizada por cinases específicas; pode ser revertida por fosfatases
Metilação (CH <sub>3</sub> )	Lys	realizada por metilases; revertida por demetilases
Hidroxilação (OH)	Pro, Lys, Asp	hidroxiprolina (Hyp) e hidroxilisina (Hyl) são particularmente comuns no colágeno
Acetilação (CH <sub>3</sub> CO)	Lys	realizada por uma acetilase; revertida por uma desacetilase
Carboxilação (COOH)	Glu	realizada por uma $\gamma$ -carboxilase
<i>N</i> -glicosilação (carboidratos complexos)	Asn <sup>a</sup>	ocorre inicialmente no retículo endoplasmático, com modificações posteriores sendo realizadas no aparelho de Golgi
<i>O</i> -glicosilação (carboidratos complexos)	Ser, Thr, Hyl <sup>b</sup>	ocorre no aparelho de Golgi; menos comum que a <i>N</i> -glicosilação
Glicosilfosfatidilinositol (glicolípido)	Asp <sup>c</sup>	serve para ancorar proteínas à camada externa da membrana plasmática
Miristoilação (ácido graxo C <sub>14</sub> )	Gly <sup>d</sup>	serve como âncora à membrana
Palmitoilação (ácido graxo C <sub>16</sub> )	Cys <sup>e</sup>	serve como âncora à membrana
Farnesilação (grupo prenil C <sub>15</sub> )	Cys <sup>c</sup>	serve como âncora à membrana
Geranilgeranilação (grupo prenil C <sub>20</sub> )	Cys <sup>c</sup>	serve como âncora à membrana

<sup>a</sup> Isso é especialmente comum quando asparagina (Asn) está presente na sequência Asn-X-(Ser/Thr), em que X é um aminoácido diferente de Pro. <sup>b</sup> Hidroxilisina. <sup>c</sup> Na extremidade C-terminal do polipeptídeo. <sup>d</sup> Na extremidade N-terminal do polipeptídeo. <sup>e</sup> Para formar uma ligação S-palmitoil.

para a membrana plasmática, são normalmente glicosiladas. Nestes casos, os resíduos de açúcar são montados como oligossacarídeos antes de ser ligados a uma proteína.

Dois tipos principais de glicosilação ocorrem. A *N*-glicosilação de carboidratos ocorre pela ligação de um grupo carboidrato ao átomo de nitrogênio de uma cadeia lateral de uma asparagina, enquanto a *O*-glicosilação diz respeito à adição de um carboidrato ao átomo de oxigênio de um grupo OH da cadeia lateral de alguns aminoácidos (ver Tabela 1.7).

Proteoglicanos são proteínas com glicosaminoglicanos anexos (polissacarídeos), os quais incluem unidades repetidas de dissacarídeos contendo glucosamina ou galactosamina. Os proteoglicanos mais bem caracterizados são componentes da matriz extracelular, uma rede complexa de macromoléculas secretada por células (e rodeada por essas) em tecidos ou em sistemas de cultivo.

### Adição de grupos lipídicos

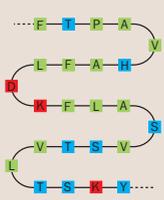
Algumas proteínas, sobretudo proteínas de membrana, são modificadas pela adição de grupos lipídicos acil ou prenil. Estes grupos adicionados servem como âncoras de membrana, sequências hidrofóbicas de aminoácidos que mantêm a proteína recém-sintetizada embebida na membrana plasmática ou na membrana do retículo endoplasmático (Tabela 1.8).

O ancoramento de uma proteína à camada externa da membrana plasmática envolve a ligação de um grupo glicosilfosfatidilinositol (GPI). Este grupo glicolipídico contém ácidos graxos que servem como âncoras de membrana; ele está ligado sucessivamente a uma unidade de glicerofosfato, uma unidade de oligossacarídeo e, finalmente – por meio de uma unidade de fosfoetanolamina – à extremidade C-terminal da proteína. A proteína inteira, com exceção da âncora GPI, se localiza no espaço extracelular.

### Clivagem pós-traducional

O produto primário da tradução pode sofrer clivagens internas para gerar um produto maduro com tamanho menor. A metionina inicial é ocasionalmente clivada do produto primário da tradução, conforme mecanismo descrito para a síntese da  $\beta$ -globina (ver Figura 1.23 C, D). Clivagens ainda mais substanciais são observadas durante a maturação de várias proteínas, incluindo proteínas plasmáticas, hormônios polipeptídicos, neuropeptí-

**TABELA 1.8** Níveis de estrutura proteica

Nível	Definição	Notas
Primária 	a sequência linear de aminoácidos em um polipeptídeo	pode variar enormemente em comprimento, desde alguns poucos até milhares de aminoácidos
Secundária 	o caminho que a cadeia principal de um polipeptídeo segue em regiões locais da estrutura primária	varia ao longo do comprimento do polipeptídeo; elementos comuns da estrutura secundária incluem as $\alpha$ -hélices e as folhas $\beta$ -pregueadas
Terciária 	a estrutura tridimensional geral de um polipeptídeo gerada pela combinação de todas as estruturas secundárias	pode assumir várias formas (p. ex., globular, em forma de bastão, tubo, alça, folha)
Quaternária 	a estrutura agregada de uma proteína multimérica (contendo mais de uma subunidade, as quais podem ser de mais de um tipo)	pode ser estabilizada por pontes dissulfeto entre as subunidades ou pela ligação de ligantes e outros fatores

deos e fatores de crescimento. Sequências sinalizadoras de clivagem são frequentemente utilizadas para marcar proteínas tanto para a exportação como para o transporte para uma localização intracelular específica. Uma molécula de mRNA pode às vezes especificar mais de um polipeptídeo funcional, como resultado de clivagens pós-traducionais de um precursor polipeptídico grande (Figura 1.26).

## A complexa relação entre a sequência de aminoácidos e a estrutura da proteína

As proteínas podem ser compostas por um ou mais polipeptídeos, cada um dos quais pode ser submetido a modificações pós-traducionais. Interações entre a proteína e cada um dos seguintes fatores podem alterar substancialmente a conformação da proteína:

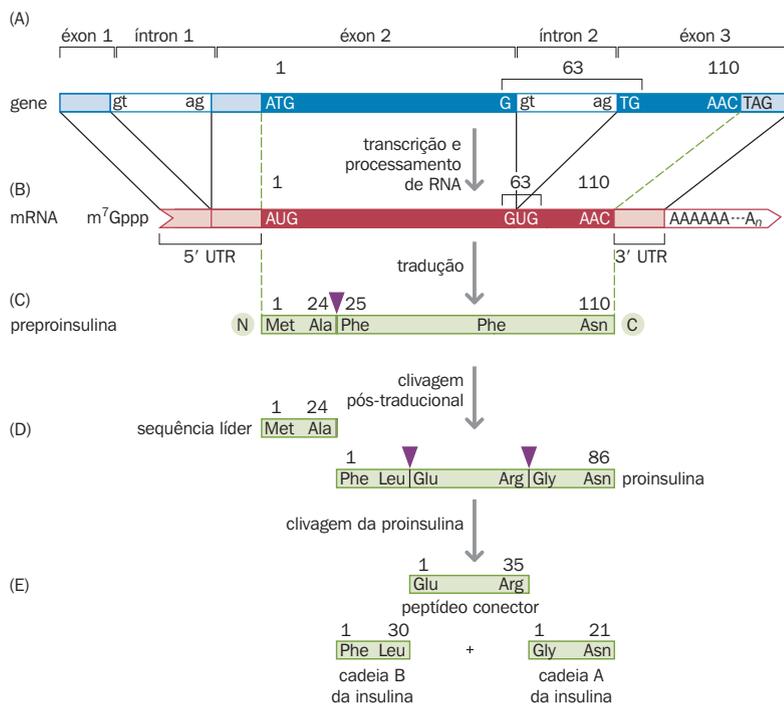
- Um **cofator**, tal como um cátion bivalente (como  $\text{Ca}^{2+}$ ,  $\text{Fe}^{2+}$ ,  $\text{Cu}^{2+}$  ou  $\text{Zn}^{2+}$ ) ou uma pequena molécula que é requerida para uma atividade enzimática funcional (como  $\text{NAD}^+$ ).
- Um **ligante** (qualquer molécula que a proteína ligue especificamente).

Quatro níveis distintos de organização estrutural nas proteínas têm sido identificados e definidos (ver Tabela 1.8).

Mesmo dentro de um único polipeptídeo, existe um amplo escopo para as ligações de hidrogênio entre os diferentes resíduos de aminoácidos. Isso estabiliza as cargas polares parciais ao longo da cadeia principal do polipeptídeo, com profundos efeitos na forma geral dessa proteína. No que diz respeito à conformação da proteína, as ligações de hidrogênio mais significativas são aquelas que ocorrem entre o oxigênio do grupo carbonila de uma ligação peptídica ( $\text{C}=\text{O}$ ) e o hidrogênio de um grupo amino ( $\text{NH}$ ) de outra ligação peptídica. Diversos padrões estruturais fundamentais (motivos), estabilizados por ligações de hidrogênio em um mesmo polipeptídeo, têm sido identificados. Dentre estes, os mais básicos são descritos a seguir.

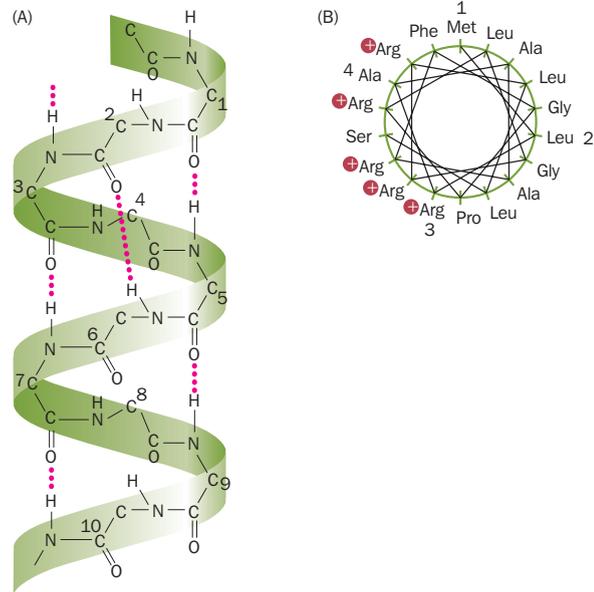
### A $\alpha$ -hélice

Refere-se à formação de um cilindro rígido que é estabilizado por ligações de hidrogênio entre o oxigênio do grupo carbonila de uma ligação peptídica e o átomo de hidrogênio do grupo amino de uma ligação peptídica localizada a quatro aminoácidos de distância (Figura 1.27). As  $\alpha$ -hélices frequentemente ocorrem em proteínas que executam funções celulares chaves (como fatores de transcrição, onde são representados nos domínios de



**Figura 1.26** A síntese da insulina envolve múltiplas clivagens pós-traducionais de polipeptídeos precursores. (A) O gene da insulina humana contém três éxons e dois íntrons. A sequência codificadora (a parte que será utilizada para fazer o polipeptídeo) é apresentada em azul escuro. Ela é confinada à sequência 3' do éxon 2 e à sequência 5' do éxon 3. (B) O éxon 1 e a parte 5' do éxon 2 especificam a região 5' não traduzida (5' UTR), e a extremidade 3' do éxon 3 especifica a região 3' UTR. As UTRs são transcritas e, portanto, são apresentadas nas extremidades do mRNA. (C) O produto primário da tradução, a preproinsulina, contém 110 resíduos e é clivado para gerar (D) uma sequência líder N-terminal com 24 resíduos (que é necessária para a proteína atravessar a membrana celular, sendo posteriormente descartada) mais um precursor proinsulina com 86 resíduos. (E) A proinsulina é clivada para gerar um segmento central (o peptídeo conector), o qual pode manter a conformação das cadeias A e B da insulina, antes da formação das pontes dissulfeto que conectam essas cadeias (ver Figura 1.29).

**Figura 1.27 A estrutura de uma  $\alpha$ -hélice padrão e de uma  $\alpha$ -hélice anfipática.** (A) A estrutura de uma  $\alpha$ -hélice é estabilizada por ligações de hidrogênio entre o oxigênio do grupo carbonila ( $C=O$ ) de cada ligação peptídica e o hidrogênio do grupo amino da ligação peptídica ( $NH$ ) do quarto aminoácido adiante, resultando em 3,6 resíduos de aminoácido por volta da hélice. As cadeias laterais de cada aminoácido são localizadas no lado externo da hélice; praticamente não existe espaço livre dentro dela. Note que apenas a cadeia principal do polipeptídeo é representada e que algumas ligações foram omitidas por questões de clareza. (B) Uma  $\alpha$ -hélice anfipática apresenta um empacotamento mais intenso, com aminoácidos carregados e aminoácidos hidrofóbicos localizados em superfícies opostas. Aqui é representada uma visão final desta hélice: cinco resíduos de arginina carregados positivamente são agrupados em um lado da hélice, enquanto o lado oposto apresenta uma série de aminoácidos hidrofóbicos (principalmente Ala, Leu e Gly). As linhas dentro do círculo indicam resíduos vizinhos – a metionina iniciadora (posição 1) é conectada a uma leucina (2), a qual é conectada a uma arginina (3), que é adjacente a uma alanina (4), e assim por diante.



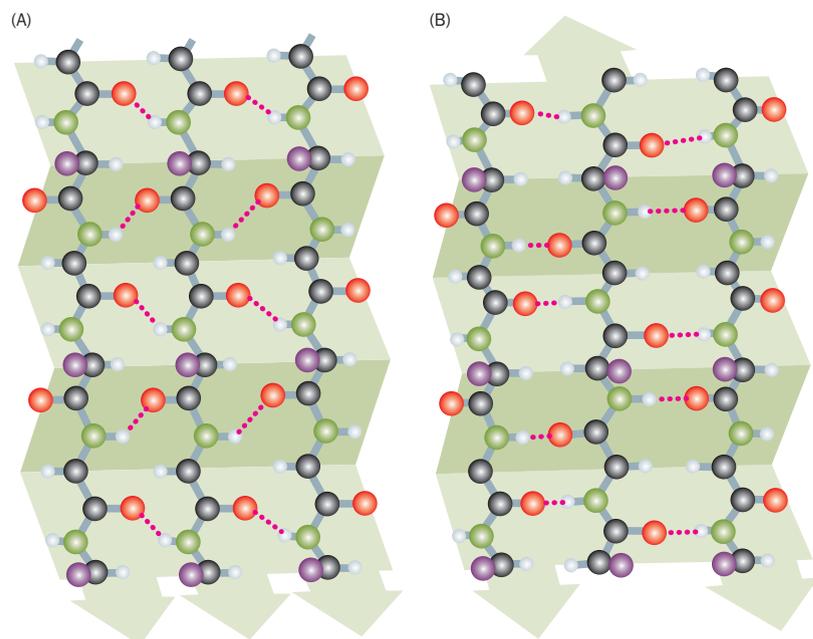
ligação ao DNA). As  $\alpha$ -hélices idênticas, com um arranjo repetido de cadeias laterais não polares, podem se enrolar para formar uma estrutura estável referida como super-hélice. As super-hélices ocorrem em muitas proteínas fibrosas, como o colágeno da matriz extracelular, a proteína muscular tropomiosina, a  $\alpha$ -queratina no cabelo e o fibrinogênio nos coágulos sanguíneos.

### As folhas $\beta$ -pregueadas

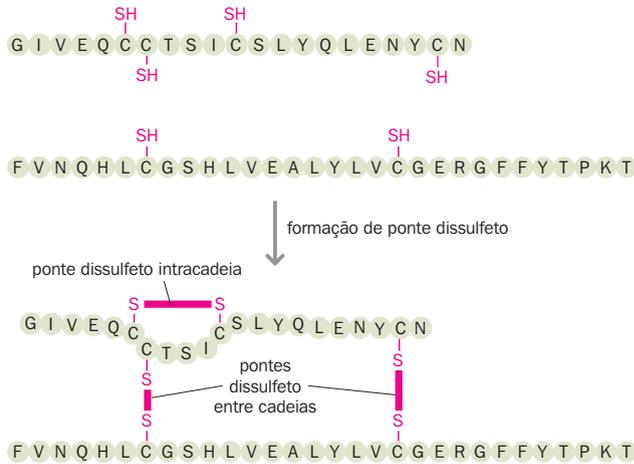
As folhas  $\beta$ -pregueadas também são estabilizadas por ligações de hidrogênio, mas, neste caso, ocorrem entre ligações peptídicas opostas em segmentos paralelos ou antiparalelos da mesma cadeia polipeptídica (Figura 1.28). As folhas  $\beta$ -pregueadas ocorrem – geralmente em conjunto com  $\alpha$ -hélices – no centro da maioria das proteínas globulares.

### A volta- $\beta$

Ligações de hidrogênio podem ocorrer entre aminoácidos que estão ainda mais próximos uns dos outros em um polipeptídeo. Quando isso acontece entre as ligações peptídicas do



**Figura 1.28 A estrutura da folha  $\beta$ -pregueada.** Neste caso, ligações de hidrogênio ocorrem entre o oxigênio da carbonila ( $C=O$ ) e o hidrogênio do grupo amino em segmentos adjacentes de folhas  $\beta$ -pregueadas (A) paralelas e (B) antiparalelas. [Adaptada de Lehninger AL, Nelson DL & Cox MM (1993) *Principles of Biochemistry*, 2nd ed. Com permissão de WH Freeman and Company.]



**Figura 1.29 Pontes dissulfeto intracadeias e intercadeias na insulina humana.** Pontes dissulfeto (–S–S–) formam, por uma reação de condensação entre os grupos sulfidril (–SH), grupos nas cadeias laterais dos resíduos de cisteína. Eles podem ser formados entre cadeias laterais de cisteínas dentro do mesmo polipeptídeo (como entre as posições 6 e 11 na cadeia A da insulina) e também entre cadeias laterais de cisteínas em diferentes polipeptídeos que interagem, tais como as cadeias A e B da insulina.

grupo CO de um resíduo de aminoácido e o grupo amino da ligação peptídica de um resíduo de aminoácido três posições adiante, resulta em uma volta- $\beta$  em forma de grampo. Mudanças abruptas na orientação de um polipeptídeo impedem a formação de formas globulares compactas. Essas voltas- $\beta$  podem conectar fitas paralelas ou antiparalelas nas folhas  $\beta$ -pregueadas.

### Estruturas de ordem superior

Muitos outros motivos estruturais complexos, que consistem em combinações dos módulos estruturais anteriores, formam **domínios de proteínas**. Estes domínios são cruciais para a forma e para a estabilidade geral da proteína, representando unidades funcionais envolvidas na ligação a outras moléculas. Outra determinante fundamental da estrutura (e função) de uma proteína é a **ponte dissulfeto**. Elas podem se formar entre os átomos de enxofre de um grupo sulfidril (–SH) em dois aminoácidos que podem residir em uma mesma cadeia polipeptídica ou em cadeias polipeptídicas distintas (Figura 1.29).

Em geral, a estrutura primária de uma proteína determina o conjunto de estruturas secundárias que, combinadas, geram a estrutura terciária da proteína. Motivos estruturais secundários podem ser preditos a partir de uma análise da estrutura primária, mas a estrutura terciária completa não pode ser facilmente predita de forma acurada. Finalmente, algumas proteínas formam complexos agregados de subunidades polipeptídicas, resultando em um arranjo conhecido como estrutura quaternária.

## LEITURAS ADICIONAIS

- Agris PF, Vendeix FA & Graham WD (2007) tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* 366, 1-13.
- Calvo O & Manley JL (2003) Strange bedfellows: polyadenylation factors at the promoter. *Genes Dev.* 17, 1321-1327.
- Carradine CR, Drew HR, Luisi B & Travers AA (2004) Understanding DNA. The Molecule And How It Works, 3rd ed. Academic Press.
- Crain PF, Rozenski J & McCloskey JA. RNA Modification Database. <http://library.med.utah.edu/RNAmods>
- Fedorova O & Zingler N (2007) Group II introns: structure, folding and splicing mechanisms. *Biol. Chem.* 388, 665-678.
- Garcia-Diaz M & Bebenek K (2007) Multiple functions of DNA polymerases. *Crit. Rev. Plant Sci.* 26, 105-122.
- Preiss T & Hentze MW (2003) Starting the protein synthesis machine: eukaryotic translation initiation. *BioEssays* 25, 1201-1211.
- Sander DM. Big Picture Book of Viruses. [http://www.virology.net/Big\\_Virology/BVHomePage.html](http://www.virology.net/Big_Virology/BVHomePage.html)
- Wear MA & Cooper JA (2004) Capping protein: new insights into mechanism and regulation. *Trends Biochem. Sci.* 29, 418-428.
- Whitford D (2005) Protein Structure And Function. John Wiley.

