

# Expressão gênica humana

## 8.1 Uma visão geral da expressão gênica em células humanas

Os mecanismos de controle utilizados para regular a expressão gênica humana são fundamentalmente semelhantes àqueles encontrados em outros mamíferos e geralmente se assemelham aos dos eucariotos em geral. Embora muito mais complexos do que mecanismos equivalentes em organismos com genomas pequenos, muitos dos mesmos princípios básicos se aplicam e, como em outros eucariotos, o principal nível de controle da expressão gênica é a iniciação da transcrição. Os mamíferos são organismos multicelulares especialmente complexos e, por isso, talvez não seja surpreendente que existam alguns mecanismos de controle genético que não são utilizados em bactéria ou em alguns outros eucariotos. Vários mecanismos reguladores são necessários para manter as muitas faces diferentes da expressão gênica de mamíferos nos níveis espacial e temporal (Quadro 8.1). Embora simplista, é conveniente considerar-se três amplos níveis nos quais a regulação gênica pode operar.

### **Regulação da expressão gênica simultânea à transcrição**

Há muito tempo temos nos acostumado com a idéia de que um primeiro controle da regulação gênica em eucariotos ocorre em nível da iniciação da transcrição. A regulação da expressão gênica pode ocorrer por meio da região promotora central de um gene, em nível do recrutamento e processabilidade da RNA-polimerase relacionada. A expressão de genes é iniciada pela ligação de fatores de transcrição ao promotor. Níveis basais de transcrição podem ser modulados pela ligação de fatores proteicos a outras regiões reguladoras que ocorrem em seqüências que flanqueiam o gene, ou algumas vezes no interior de íntrons do gene.

### **Regulação da expressão gênica posterior à transcrição**

Essa categoria coincide com a seção anterior, uma vez que inclui mecanismos que operam em nível do processamento do RNA, tal como o encadeamento (*splicing*) do RNA, o qual pode ser considerado como simultâneo à transcrição em vez de posterior à transcrição (Steinmetz, 1997). Além do processamento do

RNA, outros níveis nos quais o controle da expressão gênica pode ser exercido incluem transporte do mRNA, tradução, estabilidade do mRNA, processamento da proteína, sinalização da proteína, estabilidade da proteína, etc.

Uma variedade surpreendente de mecanismos é empregada em nível de processamento do RNA, com alelos únicos em um indivíduo freqüentemente sendo capazes de gerar uma diversidade de produtos gênicos (isoformas). A ocorrência desses e de outros mecanismos necessitou de uma definição mais flexível para o termo *gene* do que a que foi primeiramente utilizada (ver adiante). Vários mecanismos estão envolvidos na regulação da expressão gênica em nível de tradução, e um número crescente de seqüências reguladoras vem sendo identificados nas regiões 5' e, especialmente, nas regiões 3' não-traduzidas do mRNA. O controle da expressão gênica em nível do processamento da proteína, sinalização e estabilidade, tem sido demonstrado em certos sistemas. Por exemplo, a ativação de alguns hormônios peptídicos, tais como a insulina, necessita de uma clivagem posterior à tradução, a partir de formas precursoras (Figura 1.23).

### **Mecanismos epigenéticos e controle da expressão gênica de longo alcance**

Além dos fatores genéticos, fatores adicionais, que podem ser transmitidos para a progênie de células após a divisão celular, *mas que não são diretamente atribuíveis às seqüências de DNA*, são **descritos como epigenéticos**. A metilação do DNA é um mecanismo epigenético que exerce um papel importante no controle genético em mamíferos, atuando como um método geral de manutenção da repressão da transcrição. Além disso, uma variedade de outros mecanismos, os quais afetam o ambiente da cromatina de um gene e, portanto, sua capacidade para expressão gênica, operam em células de mamíferos. Em alguns casos, os mecanismos garantem que, em uma célula, somente um dos dois alelos parentais herdados seja normalmente expresso, mesmo que a seqüência nucleotídica do alelo que não está sendo expresso possa ser idêntica à do que está sendo expresso.

A Tabela 8.1 fornece uma visão geral dos diferentes tipos de mecanismos conhecidos, que estão envolvidos na regulação da expressão de genes humanos.

## Quadro 8.1

### Restrição espacial e temporal da expressão gênica em células de mamíferos

A regulação da expressão gênica em células humanas e de mamíferos é exercida em diferentes níveis, a fim de obter-se expressão restrita em uma variedade de níveis espaciais e temporais.

#### Restrição espacial da expressão gênica

Alguns genes necessitam ser expressos em, essencialmente, todos os tipos de células nucleadas, porque eles codificam um produto-chave que é necessário para realizar uma função geral em todas as células, por exemplo, síntese protéica e produção de energia. Tais genes são freqüentemente denominados genes de manutenção (do inglês *housekeeping genes*). Muitos genes humanos e de mamíferos, no entanto, apresentam padrões de expressão gênica tecido-específica muito mais restrita. A restrição espacial da expressão gênica pode ocorrer em diferentes níveis:

- **Padrão múltiplo órgão/tecido.** Diferentes tipos de sistemas múltiplos de expressão são encontrados. Em alguns casos, um gene pode estar realizando um tipo semelhante de função em sistemas orgânicos diferentes. Uma variedade de genes que desempenham um papel importante no início do desenvolvimento pode estar envolvida na regulação de genes-alvo em vários sistemas orgânicos diferentes, por exemplo, o gene da corda vocal do ouriço, que é expresso em várias partes do sistema nervoso, em membros em desenvolvimento e em outros locais. Em outros casos, um gene pode codificar variantes diferentes (isoformas) em diferentes tecidos, pela utilização de promotores tecido-específicos ou de encadeamento tecido-específico alternativo. Em alguns casos, eles podem possuir funções diferentes (Seção 8.3.2).
- **Especificidade de tecido, linhagem celular ou tipo celular.** Alguns genes possuem uma função que é apropriada para um determinado tipo ou linhagem celular, como no caso do gene da  $\beta$ -globina, que é expresso em eritrócitos.
- **Células individuais.** Alguns genes produzem produtos diferentes em células individuais que são do mesmo tipo celular. Por exemplo, linfócitos B individuais expressam moléculas de anticorpos diferentes, os receptores de células T são diferentes em linfócitos T individuais, e neurônios olfativos individuais produzem re-

ceptores olfativos diferentes. Outros exemplos envolvem células em que um de dois alelos de um gene é normalmente reprimido. Ele pode consistentemente ser o alelo paterno ou o alelo materno (pré-programação genômica, ver Seção 8.5.4), ou um alelo pode ser reprimido independentemente da origem parental (Seção 8.5.3).

- **Distribuição intracelular.** As proteínas de genes diferentes são transportadas para diferentes localizações intracelulares (ou extracelulares). Em alguns casos, isoformas diferentes do mesmo gene podem ser enviadas para localizações intracelulares diferentes (Seção 8.3.2). Além disso, mecanismos de controle genético são necessários para enviar mRNA de alguns genes para localizações intracelulares diferentes (Seção 8.2.5).

#### Restrição temporal da expressão gênica

- **Estágios do desenvolvimento.** Nos estágios bem iniciais do desenvolvimento, a transcrição não ocorre; em vez disso, as células dependem do RNA previamente sintetizado. Mais tarde no desenvolvimento, alguns genes podem ser expressos transitóriamente em estágios específicos. Algumas famílias gênicas possuem membros que são expressos em diferentes estágios do desenvolvimento, como no caso dos genes da globina (Figura 8.23).
- **Estágios de diferenciação.** À medida que as células se desenvolvem e se diferenciam, seus genomas são modificados, resultando em padrões de expressão gênica alterados. Em algumas células terminalmente diferenciadas, a transcrição não ocorre. As modificações genômicas que resultam na progressão para uma célula somática adulta nucleada costumavam ser consideradas como irreversíveis, até a clonagem da ovelha Dolly (Seção 21.5.2).
- **Estágios do ciclo celular.** Alguns genes são expressos somente em períodos específicos do ciclo celular. Por exemplo, muitos genes de histonas são expressos somente na fase S (síntese de DNA).
- **Expressão induzível.** Alguns genes são ativados em resposta a sinais ambientais ou à sinalização extracelular de outras células (Seção 8.2.4). Tal expressão gênica é facilmente revertida se o fator de indução é removido.

## 8.2 Controle da expressão gênica pela ligação de fatores protéicos *trans*-atuantes em seqüências reguladoras *cis*-atuantes no DNA e RNA

Uma base molecular comum para muito do controle da expressão gênica (se ele ocorre em nível de iniciação da transcrição, processamento do RNA, tradução ou transporte do RNA) é a ligação de fatores protéicos a seqüências de ácidos nucléicos reguladoras. Estas últimas podem ser seqüências de DNA encontradas nas proximidades de um gene ou mesmo no seu interior, ou seqüências transcritas de RNA em nível de RNA-precursor ou mRNA. Como os fatores protéicos envolvidos na regulação da expressão gênica são, eles próprios, codificados por genes lo-

calizados a distância, são forçados a migrar para o seu local de ação e, por isso, são chamados de fatores *trans*-atuantes. Ao contrário, as seqüências reguladoras, às quais eles se ligam, estão sobre a mesma molécula de DNA ou RNA que o gene ou o transcrito de RNA que está sendo regulado. Tais seqüências são ditas *cis*-atuantes.

### Controle através da ligação DNA-proteína

Um importante controle da expressão gênica em células eucarióticas é exercido em nível de *iniciação* da transcrição, onde três tipos de RNA-polimerases são conhecidos para transcrever diferentes classes de genes (Tabela 1.3). Todas as três RNA-polimerases são enzimas grandes, consistindo de 8 a 14 subunidades e, em cada caso, a polimerase é recrutada para transcrever um gene após a ligação de proteínas (fatores de transcrição) em

*início (iniciação)*

**Tabela 8.1** Visão geral da regulação da expressão gênica em células humanas

Mecanismos de expressão seletiva	Exemplos
<b>De transcrição</b>	
Ligação de fatores de transcrição tecido-específicos a elementos <i>cis</i> -atuantes de um único gene	Ver Tabela 8.2
Ligação direta de hormônios, fatores de crescimento ou intermediários a elementos de resposta em elementos de transcrição induzíveis	Elementos de resposta a cAMP, elementos de resposta a hormônios esteróides, etc. (Tabela 8.4 e Figura 8.9)
Utilização de promotores alternativos em um único gene	Ver Figura 8.13 para o gene da distrofina e Figura 8.20 para o gene <i>Dnmt1</i> . Também aplica-se a muitos outros genes
<b>Posteriores à transcrição</b>	
Encadeamento alternativo	Seção 8.3.2 e Figura 8.14
Poliadenilação alternativa	Seção 8.3.2 e Figura 8.15
Edição de RNA tecido-específica	Seção 8.3.3 e Figura 8.16
Mecanismos de controle da tradução	Seção 8.2.5 e Figura 8.12
<b>Mecanismos epigenéticos e controle da expressão gênica de longo alcance pela estrutura da cromatina</b>	
Exclusão alélica	Rearranjos de DNA em linfócitos B e T, que produzem imunoglobulinas célula-específicas e receptores de células T (Seção 8.6.2) Pré-programação de certos genes (Seção 8.5.4) Inativação do cromossomo X (inativação por meio do produto do gene <i>XIST</i> de muitos genes no cromossomo X, o que é expresso em células de fêmeas (Seção 8.5.6) Exclusão alélica aleatória por mecanismos desconhecidos, p. ex., <i>IL-2</i> , <i>IL-4</i> , <i>PAX5</i> , etc. (Seção 8.5.3 e Quadro 8.6)
Controle de longo alcance pela estrutura da cromatina	Efeitos clássicos de posição (p. ex., possivelmente no caso da distrofia fâscio-escápulo-umeral (Seção 8.5.1) Supressão da expressão gênica por domínios de cromatina (p. ex., do gene <i>PAX6</i> na aniridia e do gene <i>SOX9</i> na displasia campomélica (Seção 8.5.1) Competição por reforçadores ou silenciadores (p. ex., na expressão da globina, ver Seção 8.5.2) e na pré-programação dos genes <i>H19</i> e <i>IGF2</i> (Seção 8.5.4)
Posição celular dependente, sinalização de curto alcance	Seção 8.4.1

seqüências reguladoras do DNA no interior do gene ou nas suas proximidades. A cromatina é uma estrutura altamente organizada e densamente empacotada, que não permite o acesso fácil das RNA-polimerases e, por isso, os fatores de transcrição são necessários para ajudar na sua ativação, originando uma estrutura aberta que possibilitará a ocorrência da transcrição.

#### Controle pela ligação RNA-proteína

Além dos fatores de transcrição, as proteínas de ligação a RNA são utilizadas para regular a expressão gênica. Os exemplos melhor estudados envolvem a ligação em seqüências reguladoras nas seqüências não-traduzidas do mRNA, permitindo o controle da expressão gênica em nível de tradução. Além disso, interações de ligações RNA-proteína também estão envolvidas no controle da expressão gênica em nível de processamento diferencial do RNA, como no caso da ligação das proteínas SR e HnRNP ao pré-mRNA, a fim de modular a escolha de éxons no encadeamento. Os últimos mecanismos são considerados separadamente na Seção 8.3, para ilustrar a enorme complexidade dos mecanismos de expressão que podem ser utilizados para deci-

frar genes únicos e a importância do grande número de isoformas que podem ser produzidas como resultado.

#### 8.2.1 Fatores de transcrição ubíquos são necessários para a transcrição pelas RNA-polimerases I e III

As RNA-polimerases I e III das células eucarióticas são destinadas a transcrever genes que dão origem a moléculas de RNA (rRNA, tRNA, etc.), as quais auxiliam na expressão dos genes que codificam polipeptídeos. Os genes transcritos são genes de manutenção (do inglês *housekeeping genes*), uma vez que rRNA e tRNA são necessários, essencialmente, em todas as células, para ajudar na síntese protéica. Como resultado, fatores de transcrição ubíquos são necessários para ajudar as RNA-polimerases I e III.

#### Transcrição pela RNA-polimerase I

A RNA-polimerase I está confinada ao nucléolo e está destinada à transcrição dos genes de RNA de 18S, 5,8S e 28S. Estes últimos estão organizados consecutivamente em uma unidade de transcrição comum de 13 kb (Figura 8.1). Uma unidade composta pe-

la unidade de transcrição de 13 kb e um espaçador não-transcrito adjacente de 27 kb é repetida em tandem cerca de 50 a 60 vezes, nos braços curtos de cada um dos cinco cromossomos acrocêntricos humanos, nas regiões organizadoras nucleolares (Figura 2.18). Os cinco agrupamentos de genes de rRNA resultantes, cada um com cerca de 2 Mb de extensão, são referidos como DNA ribossômico ou rDNA.

A transcrição dos genes de rRNA de 28S, 5,8S e 18S é iniciada após a ligação de dois fatores de transcrição a um elemento promotor central no sítio de iniciação e a um elemento controlador a montante, localizado a mais de cem nucleotídeos a montante. Um dos fatores de transcrição, UBF (do inglês *upstream binding factor*, fator de ligação a montante), é um homodímero, e suas subunidades idênticas podem ligar-se primeiramente ao elemento promotor central e ao elemento controlador a montante, fazendo com que os dois fiquem próximos, de maneira que eles possam ser ligados ao segundo fator, SL1 (do inglês *selectivity factor*, fator de seletividade; conhecido em camundongo como TFI-1B; Figura 8.2). A ligação dos fatores de transcrição subsequentemente recruta a RNA-polimerase I para formar o complexo de iniciação.

O transcrito primário expresso a partir de uma única unidade de transcrição é um rRNA precursor de 45S, que sofre uma variedade de reações de clivagem e modificações base-específicas (realizadas por um grande número de diferentes tipos de pequenos RNAs nucleolares [snoRNA]) para gerar as espécies maduras de rRNA de 28S, 5,8S e 18S (Figura 8.1). Portanto, esses genes diferem da grande maioria dos genes nucleares, os quais são individualmente transcritos. Em vez disso, a transcrição do rDNA assemelha-se à transcrição do mtDNA (Seção 7.1.1 e Figura 7.2): ambas resultam em transcritos multigênicos, que originam produtos funcionalmente relacionados. Essa utilização incomum de transcritos primários poligênicos não é diferente, em princípio, da maneira pela qual um único produto de tradu-

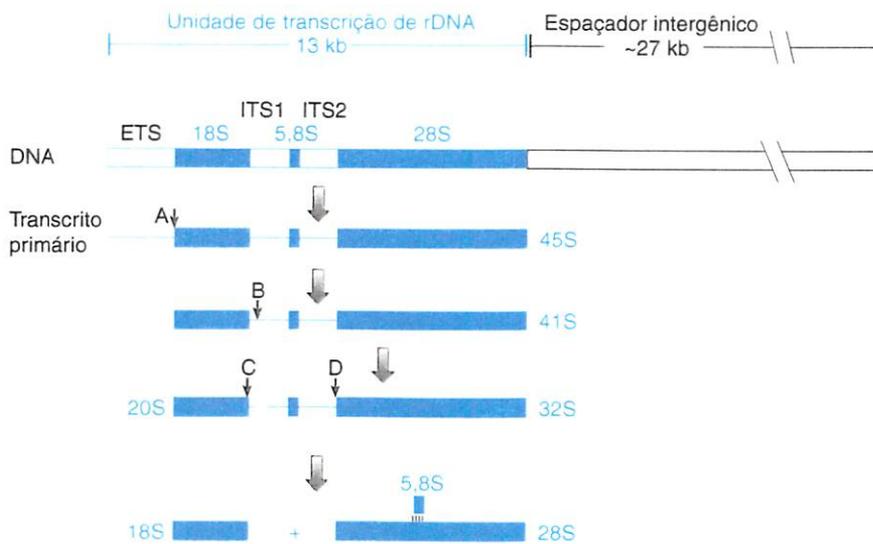
ção primário é ocasionalmente clivado para gerar dois ou mais polipeptídeos funcionalmente relacionados (ver o exemplo da insulina humana na Figura 1.23).

### Transcrição pela RNA-polimerase III

A RNA-polimerase III também está envolvida na transcrição de uma variedade de genes de manutenção, que codificam várias moléculas de pequenos RNAs estáveis, tais como rRNA de 5S, moléculas de tRNA, RNA 7SL e algumas das moléculas de snRNA (do inglês *small nuclear RNA*, pequeno RNA nuclear), necessárias para o encadeamento do RNA. Esses genes são caracterizados por promotores que se situam *no interior* da seqüência codificadora do gene, em vez de a montante dela. Nos genes de tRNA, o promotor é bipartido, consistindo de duas seqüências bem-conservadas, o *box A* e o *box B*, enquanto que no gene de rRNA de 5S, um único elemento promotor está presente, o *box C*. Em cada caso, imagina-se que a transcrição pela RNA-polimerase III seja precedida pela ligação de fatores de transcrição ubíquos aos elementos promotores, seguida da ligação subsequente de outros fatores e, finalmente, o recrutamento da polimerase (Figura 8.3).

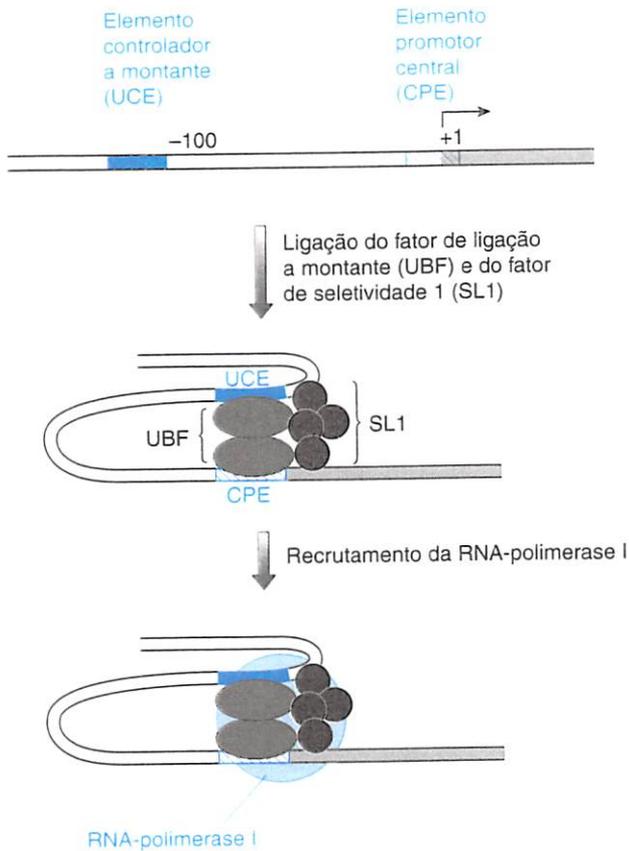
### 8.2.2 A transcrição de genes que codificam polipeptídeos freqüentemente necessita de conjuntos complexos de seqüências controladoras da transcrição cis-atuantes e de fatores de transcrição tecido-específicos

A RNA-polimerase II é responsável pela transcrição de todos os genes que codificam polipeptídeos e também de certas espécies de genes de snRNA. Como as RNA-polimerases I e III, a RNA-polimerase II depende do auxílio de fatores de transcrição gerais (a nomenclatura usual tem um prefixo comum TF para estimular fator de transcrição, seguido de um numeral romano para



**Figura 8.1** As principais espécies de rRNA humanas são sintetizadas por clivagem, a partir de uma unidade de transcrição precursora comum de 13 kb, que é parte de uma unidade de 40 kb repetida em tandem.

Setas pequenas indicadas pelas letras A-D significam posições de clivagem por endonucleases dos precursores de RNA. A clivagem do precursor de 41S em B gera dois produtos: 20S + 32S. Após a clivagem do precursor de 32S em D e excisão do pequeno rRNA de 5,8S, pontes de hidrogênio ocorrem entre o RNA de 5,8S e um segmento complementar central do rRNA de 28S. A seqüência de RNA de aproximadamente 6 kb, originada a partir das unidades espaçadoras transcritas externas e internas (ETS, ITS1 e ITS2), é degradada no núcleo. S é o coeficiente de sedimentação, uma medida de tamanho.

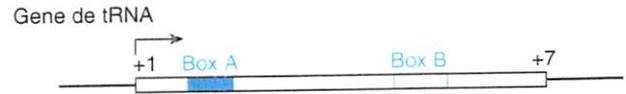


**Figura 8.2** Iniciação da transcrição pela RNA-polimerase I.

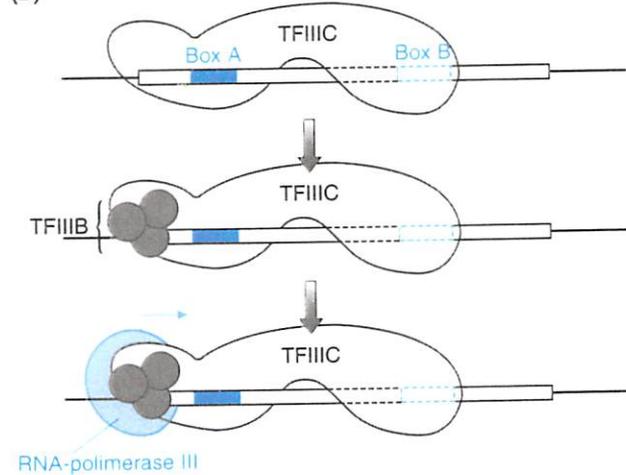
Um modelo possível pressupõe a ligação inicial de duas subunidades idênticas do fator de ligação a montante ao elemento controlador a montante e ao elemento promotor central, obrigando essas duas seqüências a ficarem bem próximas, possibilitando as suas ligações subseqüentes ao fator de seletividade 1 (SL1), que consiste de quatro subunidades. A estrutura estabilizada permite a ligação subseqüente de outros fatores (não-mostrado) e, posteriormente, da RNA-polimerase I.

estipular a RNA-polimerase associada). No caso da RNA-polimerase II, existe uma variedade de fatores de transcrição auxiliares, tais como TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, TFIIH, etc, os quais podem ser complexos em estrutura (Nikolov e Burley, 1997). Por exemplo, a proteína de ligação a TATA-box (TBP, do inglês *TATA box-binding protein*) é somente uma das múltiplas subunidades protéicas que fazem parte do TFIID, e as proteínas associadas são conhecidas como fatores associados à TBP ou proteínas TAF (do inglês *TBP associated factors*). O complexo da polimerase e dos fatores de transcrição gerais é conhecido como o aparato de transcrição basal; ele contém tudo que é necessário para iniciar a transcrição. Os genes são constitutivamente expressos em uma taxa mínima, determinada pela região promotora central (ver adiante) a menos que a taxa de transcrição seja elevada ou desligada por elementos reguladores positivos ou negativos adicionais (os quais podem estar localizados a alguma distância, em outro local, ou por componentes intrínsecos do próprio promotor).

(A)



(B)



**Figura 8.3** Genes de tRNA e rRNA de 5S possuem promotores localizados no interior da seqüência codificadora.

(A) Posições dos elementos promotores nos genes de tRNA e rRNA de 5S. Os elementos promotores A e B nos genes de tRNA estão localizados nas seqüências que especificam o braço D e os braços T $\psi$ CG, respectivamente (ver estrutura do tRNA na Figura 1.7B). (B) Iniciação da transcrição de um gene de tRNA. A ligação do fator de transcrição TFIIC nos elementos promotores permite a ligação subseqüente do fator trimérico TFIIB à seqüência imediatamente a montante do sítio de iniciação da transcrição. Em resposta à ligação do fator TFIIB, a RNA-polimerase III liga-se e inicia a transcrição. No caso dos genes de rRNA de 5S, um mecanismo semelhante ocorre mas, nesse caso, é necessária a ligação de um fator de transcrição adicional TFIIIA no box C; a ligação do fator TFIIIA permite a ligação subseqüente de TFIIB, seguida pelo recrutamento de TFIIC e da RNA-polimerase III, como no caso dos genes de tRNA.

Alguns dos genes que codificam polipeptídeos são genes de manutenção, mas, ao contrário dos produtos dos genes transcritos pelas RNA-polimerases I e III, a grande porcentagem dos genes transcritos pela RNA-polimerase II demonstra possuir modelos de expressão tecido-restritos ou tecido-específicos. Uma vez que o DNA de diferentes células nucleadas de um indivíduo é essencialmente idêntico, a identidade de uma célula, por exemplo se ela irá ser uma célula hepática ou um linfócito T, é definida pelas proteínas produzidas por ela. Além dos fatores de transcrição gerais ubíquos, portanto, os fatores de transcrição tecido-específicos ou tecido-restritos regulam a expressão de muitos genes que codificam polipeptídeos, pelo re-

conhecimento e pela ligação específica de elementos de seqüência *cis*-atuantes.

Em parte devido ao grande tamanho dos genomas nucleares de mamíferos, e também devido à necessidade geral de sistemas de controle mais sofisticados, imposta pelo fato de possuir um número muito grande de genes interagindo, os elementos controladores, em células eucarióticas, são um tanto elaborados. Frequentemente, a regulação da expressão de genes humanos individuais é controlada por vários conjuntos de elementos reguladores *cis*-atuantes. Enquanto os elementos re-

guladores individuais podem ser compostos de múltiplos elementos de seqüências curtas (tipicamente 4 a 8 nucleotídeos de extensão), distribuídos sobre umas poucas centenas de pares de bases, as diferentes classes de elementos reguladores que modulam a expressão de um único gene podem estar localizadas a distâncias consideráveis. Uma variedade de tipos de elementos *cis*-atuantes pode ser reconhecida, incluindo promotores, reforçadores (do inglês *enhancers*), silenciadores, elementos insuladores (ou isoladores) e elementos de resposta (ver Quadro 8.2).

## Quadro 8.2

### Classes de elementos de seqüência *cis*-atuantes envolvidas na regulação da transcrição de genes que codificam polipeptídeos

Promotores são combinações de curtos elementos de seqüência, normalmente localizados na região imediatamente a montante do gene, freqüentemente a 200 pb do sítio de iniciação da transcrição, e servem para iniciar a transcrição. Eles podem ser subdivididos em diferentes componentes.

- **A região promotora central** contém componentes que direcionam o complexo de transcrição basal para iniciar a transcrição do gene e que, na ausência de elementos reguladores adicionais, permite a expressão constitutiva do gene, mas em níveis muito baixos (basal). Elementos promotores centrais estão tipicamente localizados muito próximos do sítio de iniciação da transcrição, ao redor das posições -45 a +45. Eles incluem elementos tais como (i) o TATA-*box* (consenso TATA(A/T)A(A/T)); localizado na posição próximo a -25, circundado por seqüências ricas em GC e reconhecido pela *proteína de ligação ao TATA-*box**, subunidade de TFIID); (ii) a seqüência BRE (do inglês *TFIIB Recognition Element*, elemento de reconhecimento de TFIIB), que está localizada imediatamente a montante do elemento TATA e é reconhecida pelo componente TFIIB; (iii) a seqüência INR (iniciadora) localizada no sítio de iniciação da transcrição (freqüentemente uma variante da seqüência consenso PyPyCAPuPu onde Py = pirimidina e Pu = purina); e (iv) o DPE (do inglês *Downstream Promoter Element*, elemento promotor a jusante) localizado ao redor da posição +30, em relação à transcrição (ver Figura 8.4 e Burke e Kadonaga, 1997).

- **Elementos promotores não-centrais** podem ser encontrados na região promotora central, mas estão tipicamente localizados na seqüência a montante dessa, normalmente de -50 a -200 pb, em relação à iniciação da transcrição. Essa região é algumas vezes referida como região promotora proximal (em oposição à *região promotora distal*, que é algumas vezes utilizada para descrever regiões ainda mais distantes, a montante). Existem normalmente múltiplos sítios de reconhecimento para alguns fatores de transcrição ubíquos seqüência-específicos. Eles incluem: (i) GC-*box* (ou Sp1-*box*; consenso GGGCGG, freqüentemente encontrados a 100 pb do sítio de iniciação da transcrição e ligados pelo fator de transcrição Sp1); e (ii) CCAAT-*box* (seqüência consenso GGCCAATCT; tipicamente localizados na posição -75 e reconhecidos pelos fatores protéicos, CTF e CBF). CTF corresponde ao fator de transcrição de ligação a CCAAT (do inglês *CCAAT-binding Transcription Factor*) e é também conhecido como NF-1, fator nuclear 1 (do inglês *Nuclear Factor 1*). CBF corresponde a fator de ligação a CCAAT-

*box* (do inglês *CCAAT box-Binding Factor*) e é também conhecido como NF-Y para fator nuclear Y (do inglês *Nuclear Factor Y*).

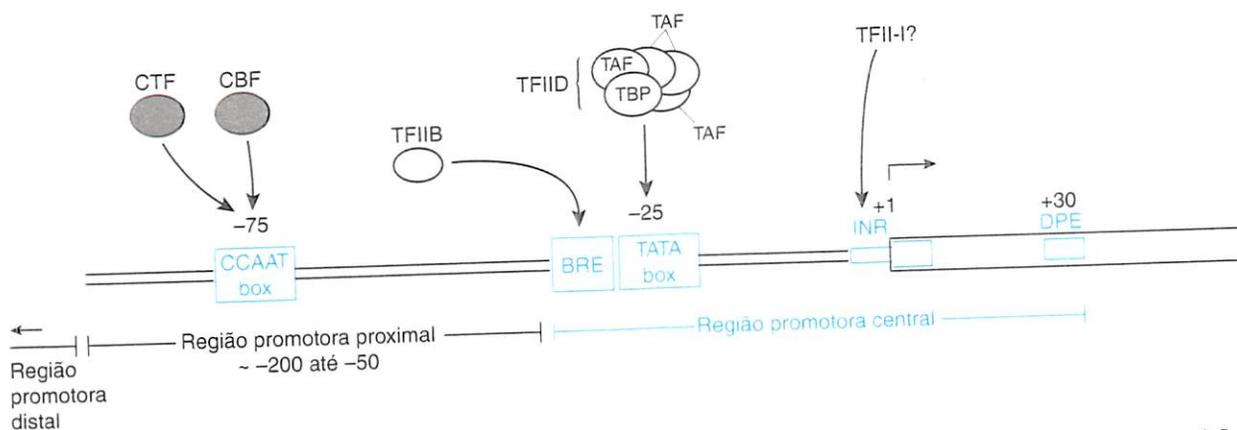
Note que os CCAAT-*box* e os GC-*box* servem para modular a transcrição basal da região promotora central e operam como *seqüências reforçadoras* (*enhancer sequences*) (ver próxima seção), enquanto que *elementos silenciadores* (ver adiante) podem também ser componentes integrais do promotor.

**Reforçadores** (*enhancers*) são elementos reguladores positivos, que servem para aumentar o nível de transcrição basal iniciado por meio dos elementos promotores centrais. Suas funções, ao contrário daquelas da região promotora central, são independentes de ambas as orientações e, em alguma extensão, de suas distâncias dos genes que regulam (Blackwood e Kadonaga, 1998). Elementos reforçadores podem estar localizados muito distantes dos genes que eles regulam (ver o exemplo das regiões controladoras de locus para os genes da globina – Seção 8.5.2) e freqüentemente contêm, com um intervalo de somente 200 a 300 pb, elementos reconhecidos por fatores de transcrição ubíquos e também por fatores de transcrição tecido-específicos, fornecendo, assim, um base para a expressão gênica tecido-específica (ver o exemplo do reforçador da  $\alpha$ -globina HS-40 na Figura 8.6). Além disso, alguns elementos reforçadores podem ser componentes integrais de promotores, como no caso dos CCAAT-*box* e dos GC-*box* (ver acima).

**Silenciadores** servem para reduzir os níveis de transcrição. Embora bem menos estudados, duas classes foram distinguidas: *silenciadores clássicos* (também chamados de *elementos silenciadores*) são elementos independentes de posição, que orientam um mecanismo ativo de repressão da transcrição, e *elementos reguladores negativos* são elementos dependentes de posição, que resultam em um mecanismo de repressão passivo (Ogbourne e Antalis, 1998). Quando estudados em genes humanos, elementos silenciadores foram relatados em várias posições: próximas ao promotor, a alguma distância a montante e também no interior de íntrons. No entanto, a evidência para tais seqüências baseia-se freqüentemente em estudos de ligação a DNA *in vitro* e sua importância *in vivo* ainda é incerta.

**Elementos de fronteira (insuladores)** são regiões de DNA freqüentemente estendendo-se de 0,5 a 3 kb, que funcionam para bloquear (ou *insular*) a propagação da influência de agentes que possuem um efeito positivo sobre a transcrição (reforçadores) ou negativo (silenciadores, efeitos repressivos semelhantes aos da heterocromatina; ver Geyer, 1997).

**Elementos de resposta** funcionam para modular a transcrição em resposta a um estímulo externo específico. Eles estão normalmente localizados a uma curta distância a montante dos elementos promotores (freqüentemente a 1 kb do sítio de iniciação da transcrição). Uma variedade de elementos de resposta foi definida; esses elementos permitem a ativação da transcrição de genes específicos, em resposta a hormônios específicos (p. ex., hormônios glicocorticóides ou esteróides, tais como o ácido retinóico) ou mensageiros secundários intracelulares como o AMP cíclico (ver Seção 8.2.4 e Tabela 8.4).



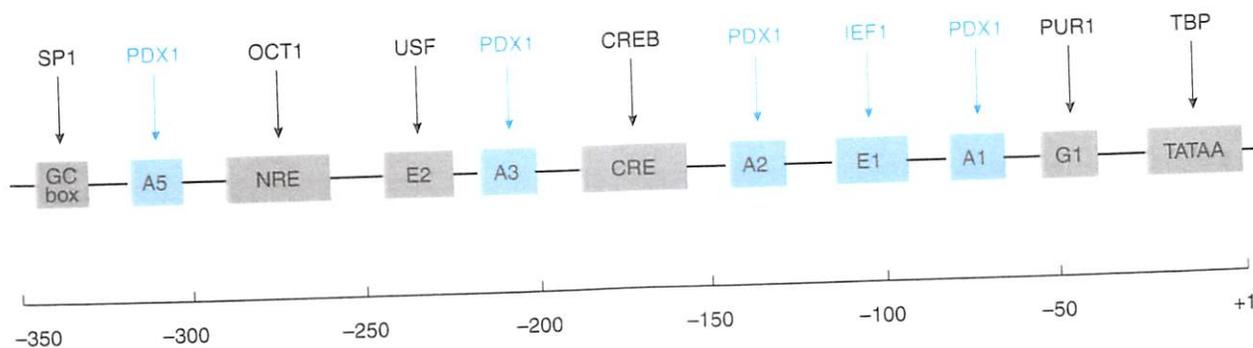
**Figura 8.4** Localizações conservadas em eucariotos complexos, de elementos promotores reguladores, ligados por fatores de transcrição ubíquos.

Note que a região promotora central de genes individuais não necessita conter todos os elementos. Por exemplo, muitos promotores não possuem um TATA-box, e utilizam o elemento iniciador (INR) funcionalmente análogo (ver Nikolov e Burley, 1997; Smale, 1997). Os GC-box também são normalmente encontrados em promotores, mas suas localizações são mais variáveis (ver Figuras 1.13 e 8.5 para alguns exemplos). Abreviaturas: BRE, elemento de reconhecimento de TFIIB; DPE, elemento promotor a jusante; CTF, fator de transcrição de ligação a CCAAT; CBF, fator de ligação a CCAAT-box; TBP, proteína de ligação a TATA-box; TAF, fatores associados à TBP. Ver texto.

A especificidade de tecido e a especificidade do estágio de desenvolvimento são, frequentemente, conferidas pelas seqüências reforçadoras e silenciadoras, e uma variedade de seqüências *cis*-atuantes foi identificada, as quais são especificamente reconhecidas por fatores de transcrição tecido-específicos. Por exemplo, a expressão específica, em eritrócitos, é seguidamente sinalizada por uma de duas seqüências: TGACTCAG (ou a sua complementar reversa CTGAGTCA), que são reconhecidas pelo fator de transcrição eritrócito-específico NF-E2, ou pela seqüência (A/T)GATA(A/G) ou a sua complementar reversa, que são reconhecidas pela série de fatores de transcrição eritrócito-espe-

cíficos GATA (ver Figura 8.6 para um exemplo). Alguns outros exemplos de elementos de seqüência *cis*-atuantes reconhecidos por fatores de transcrição tecido-específicos ou tecido-restritos estão listados na Tabela 8.2.

Além de promover ativamente a transcrição tecido-específica, alguns elementos silenciadores *cis*-atuantes conferem especificidade de tecido ou de estágio de desenvolvimento pelo bloqueio da expressão em todos os tecidos, exceto no tecido desejado. Por exemplo, o elemento silenciador neuro-restritivo (NRSE) reprime a expressão de vários genes em todos tecidos diferentes do tecido neural (Schoenherr *et al.*, 1996). Um fator de



**Figura 8.5** O promotor do gene da insulina humana contém uma variedade de elementos de seqüência reconhecidos por fatores de transcrição ubíquos e tecido-específicos.

As setas indicam a ligação de fatores de transcrição (linha superior) a elementos de seqüência reguladores presentes a montante do gene da insulina humana (linha inferior). Fatores de transcrição ubíquos ou amplamente expressos são mostrados em preto; aqueles mostrados em azul são específicos para as células pancreáticas beta. O fator de transcrição PDX1 (anteriormente chamado de IUF1) liga-se a quatro motivos de seqüência na forma de C(C/T)TAATG, que estão presentes no promotor da insulina (A1, A2, A3 e A5). Abreviaturas: CRE, elemento de resposta a cAMP; NRE, elemento regulador negativo.



**Figura 8.6** O sítio regulador da  $\alpha$ -globina HS-40 contém muitos elementos de reconhecimento para fatores de transcrição específicos de eritrócitos.

Note que o sítio HS-40 parece ser uma *região controladora de locus* para o agrupamento gênico da  $\alpha$ -globina (ver Seção 8.5.2).

transcrição que se liga ao NRSE e que é chamado de modo variado de fator silenciador neuro-restritivo (NRSF) ou de fator RE-1 de silenciamento da transcrição (REST) é expresso de forma ubíqua em tecidos não-neurais e em precursores neurais, durante o início do desenvolvimento, mas subsequente não é expresso nos neurônios mais maduros (pós-mitóticos).

### 8.2.3 Os fatores de transcrição contêm motivos estruturais conservados que permitem a ligação ao DNA

Os fatores de transcrição reconhecem e ligam-se a seqüências de nucleotídeos curtas, normalmente como resultado de uma complementaridade extensa entre a superfície da proteína e as características de superfície da hélice dupla na região de ligação. Embora as interações individuais entre os aminoácidos e os nucleotídeos sejam fracas (normalmente pontes de hidrogênio, pontes iônicas e interações hidrofóbicas), a região de ligação DNA-proteína é tipicamente caracterizada por cerca de 20 de tais contatos que coletivamente garantem que a ligação seja forte e específica. Em fatores de transcrição de humanos e de outros eucariotos, duas funções podem ser identificadas com frequência e localizadas em partes diferentes da proteína:

- **Um domínio de ativação.** Como o nome sugere, este tipo de domínio funciona na ativação da transcrição de genes-alvo, depois que o fator de transcrição ligou-se a ele. Imagine-se que os domínios de ativação estimulem a transcrição pela interação com fatores de transcrição basais, bem como auxiliem na formação do complexo de transcrição no promotor. Embora não tão bem-estudado como os domínios de ligação ao DNA, sabe-se que alguns são ricos em resíduos de aspartato e glutamato (*domínios de ativação ácidos*); outros são ricos em prolina ou glutamato.
- **Um domínio de ligação ao DNA.** Esse tipo de domínio é necessário para permitir a ligação específica do fator de transcrição aos seus genes-alvo. Ao contrário dos domínios de ativação, os domínios de ligação ao DNA dos fatores de transcrição foram bem-estudados. Um número de motivos estruturais conservados foi identificado; tais motivos são comuns a muitos fatores de transcrição diferentes com especificidades um tanto distintas, incluindo o zíper de leucina, hélice-alça-hélice, hélice-volta-hélice e dedos de zinco, descritos adiante. Cada um desses motivos utiliza hélices- $\alpha$  (ou ocasionalmente folhas- $\beta$ ; ver Figura 1.24) para ligar-se à curvatura maior do DNA. Evidentemente, embora os motivos em geral forneçam a base para a ligação ao DNA, a coleção exata de elementos de seqüência no domínio de ligação ao DNA irá fornecer a base para o reconhecimento seqüência-específico exigido. A maioria dos fatores de transcrição liga-se ao DNA como homodímeros, sendo a região

**Tabela 8.2** Exemplos de seqüências *cis*-atuantes reconhecidas por fatores de transcrição tecido-restritos e tecido-específicos

Seqüência de ligação consensual	Fator de transcrição	Padrões de expressão
(A/T)GATA(A/G)	GATA-1, -2, etc.	Eritrócitos
TGACTCAG	NF-E2	Eritrócitos
GTTAATNATTAAC (elemento PE)	HNF-1	Fígado diferenciado, rim, estômago, intestino, baço
T(G/A)TTTG(C/T)	HNF-5	Fígado
GCCTGCAGGC	Ker1	Queratinócitos
(C/T)TAAAATAA(C/T)3	MBF-1	Miócitos
(C/T)TA(A/T)AAATA(A/G)	MEF-2	Miócitos
CAACTGAC	MyoD	Mioblastos + miotúbulos
ATGCAAAT	OTF-2	Células linfóides
(C/A)A(C/A)AG	TCF-1	Células T

de ligação da proteína ao DNA normalmente diferente da região responsável pela formação dos dímeros.

### O motivo zíper de leucina

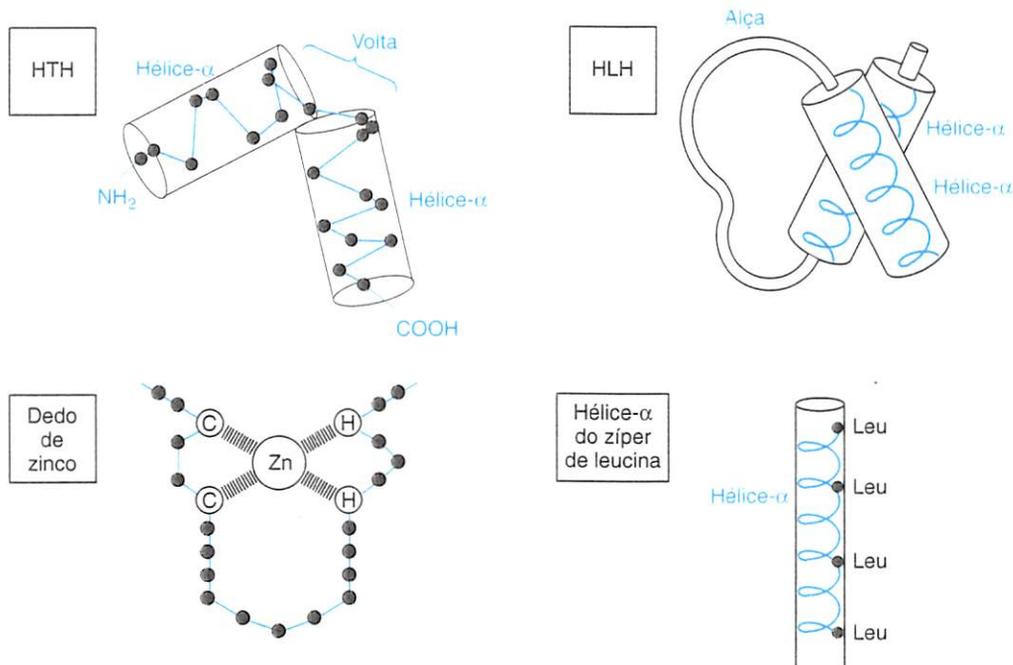
O zíper de leucina é uma seqüência helicoidal de aminoácidos, rica em resíduos de leucina (tipicamente ocorrendo uma vez a cada sete resíduos de aminoácidos, ou seja, uma vez a cada duas voltas da hélice; ver Figura 8.7), que rapidamente forma um dímero. Cada unidade monomérica consiste de uma hélice- $\alpha$  anfipática (os grupamentos laterais hidrofóbicos dos aminoácidos constituintes virados para um lado; os grupamentos polares virados para o outro lado; ver Figura 1.24). As duas hélices- $\alpha$  dos monômeros individuais combinam-se ao mesmo tempo a uma distância curta, para formar um anel enrolado (Seção 1.5.5), com as interações predominantes ocorrendo entre aminoácidos hidrofóbicos opostos dos monômeros individuais. Mais adiante dessa região, as duas hélices- $\alpha$  separam-se, de tal forma que o dímero resultante é uma estrutura em forma de Y. Imagina-se que o dímero se prenda na hélice dupla de maneira semelhante à dos prendedores em uma corda de roupas (Figura 8.8). Além de formar homodímeros, as proteínas com zíper de leucina podem ocasionalmente formar heterodímeros, dependendo da compatibilidade das superfícies hidrofóbicas dos dois monômeros diferentes. Tal formação de heterodímeros fornece um mecanismo de controle combinatório importante na regulação gênica.

### O motivo hélice-alça-hélice

O motivo hélice-alça-hélice (HLH, do inglês *helix-loop-helix*) é relacionado ao zíper de leucina e deveria ser distinguido do motivo hélice-volta-hélice (HTH, do inglês *helix-turn-helix*) descrito na próxima seção. Ele consiste de duas hélices- $\alpha$ , uma curta e uma longa, conectadas por uma alça flexível. Ao contrário da volta curta no motivo HTH, a alça no motivo HLH é flexível o bastante para permitir um dobramento para trás, de forma que as duas hélices possam se empacotar uma contra a outra; isto é, as duas hélices situam-se em planos que são paralelos um ao outro, ao contrário das duas hélices no motivo HTH (Figura 8.7). O motivo HLH medeia tanto a ligação ao DNA como a formação do dímero protéico (Figura 8.8) e permite uma ocasional formação de heterodímeros. Neste último caso, no entanto, os heterodímeros se formam entre uma proteína HLH completa e uma proteína HLH truncada, a qual não possui a extensão total da hélice- $\alpha$ , necessária para a ligação ao DNA. O heterodímero resultante é incapaz de ligar-se ao DNA firmemente. Como resultado, imagina-se que os dímeros HLH atuem como um mecanismo de controle, pela permissão da inativação de proteínas reguladoras gene-específicas.

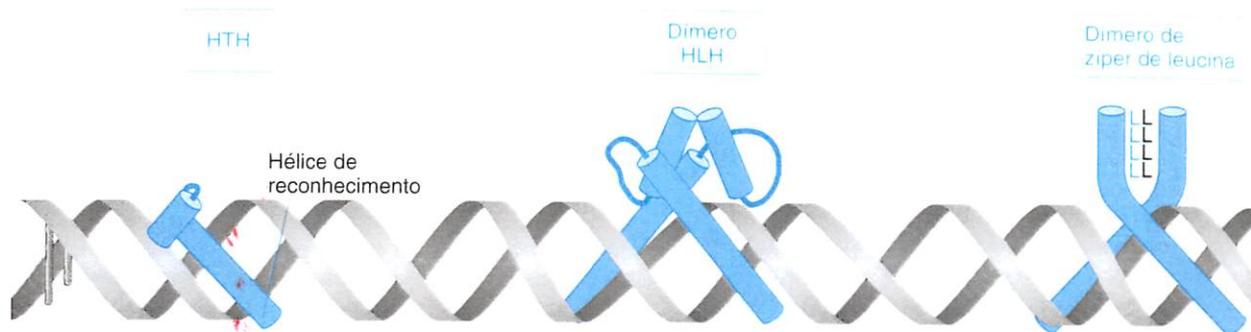
### O motivo hélice-volta-hélice

O motivo HTH é um motivo comum encontrado em homeoboxes e em um número de outros fatores de transcrição. Ele consiste de duas hélices- $\alpha$  curtas, separadas por uma curta seqüência de aminoácidos, que induz uma volta, de tal forma que as



**Figura 8.7** Motivos estruturais comumente encontrados em fatores de transcrição e em proteínas de ligação a DNA.

Abreviaturas: HTH, hélice-volta-hélice; HLH, hélice-alça-hélice. Note que o monômero do zíper de leucina é *anfipático* [i. e., possui resíduos hidrofóbicos (leucinas) consistentemente sobre uma face da hélice]. Duas de tais hélices podem alinhar-se com suas faces hidrofóbicas em oposição, para formar uma estrutura em forma de um "rolo-enrolado".



**Figura 8.8** Ligação de motivos estruturais conservados em fatores de transcrição à hélice dupla.

Note que os monômeros individuais do dímero da hélice-alça-hélice (HLH) e o dímero do zíper de leucina estão coloridos diferentemente para permitir distinção, mas podem ser idênticos (homodímeros). Os heterodímeros HLH e os heterodímeros do zíper de leucina podem fornecer um nível elevado de regulação (ver texto).

duas hélices- $\alpha$  são orientadas de maneira diferente (isto é, as duas hélices não se situam no mesmo plano, ao contrário daquelas no motivo HLH; Figura 8.7). A estrutura é muito semelhante à do motivo de ligação ao DNA de várias proteínas reguladoras de bacteriófagos, tal como a proteína cro de  $\lambda$ , cuja ligação ao DNA tem sido intensivamente estudada por cristalografia de raios X. Nos dois casos, da proteína cro de  $\lambda$  e dos motivos HTH eucarióticos, supõe-se que, enquanto o motivo HTH em geral medeia a ligação ao DNA, a hélice na região mais C-terminal atua como uma hélice de reconhecimento específica, porque ela se ajusta ao interior da curvatura maior do DNA (Figura 8.8), controlando a seqüência precisa de DNA que é reconhecida.

#### O motivo dedo de zinco

O motivo dedo de zinco envolve a ligação do íon zinco a quatro aminoácidos conservados, para formar uma alça (dedo), uma estrutura que é freqüentemente repetida em tandem. Embora existam várias formas diferentes, formas comuns envolvem a ligação de um íon de  $Zn^{2+}$  a dois resíduos conservados de cisteína e dois resíduos conservados de histidina, ou a quatro resíduos

conservados de cisteína. A estrutura resultante pode, então, consistir de uma hélice- $\alpha$  e uma folha- $\beta$ , mantidas juntas pela coordenação do íon de  $Zn^{2+}$ , ou duas hélices- $\alpha$ . Em um ou outro caso, o primeiro contato com o DNA é realizado pela ligação de uma hélice- $\alpha$  à curvatura maior. O assim chamado dedo  $Cys_2/His_2$  tipicamente compreende cerca de 23 aminoácidos com dedos vizinhos separados por uma seqüência de cerca de 7 ou 8 aminoácidos (Figura 8.7).

#### 8.2.4 Uma variedade de mecanismos permite a regulação da transcrição da expressão gênica em resposta a um estímulo externo

Nas células eucarióticas, a expressão gênica pode ser alterada de uma maneira semipermanente, à medida que as células se diferenciam, ou de uma maneira temporária, facilmente reversível, em resposta a sinais extracelulares (expressão gênica induzível). Sinais ambientais, tais como as concentrações extracelulares de certos íons e pequenas moléculas nutrientes, temperatura, choque, etc., podem resultar na alteração dramática do pa-

**Tabela 8.3** Maneiras diferentes de sinalização celular

Maneira	Características	Exemplos
Sinalização direta célula a célula	Um sinal sobre a superfície de uma célula é ligado por um receptor específico em outra célula	Fatores de crescimento ancorados à membrana e seus receptores
Sinalização endócrina	Hormônios são secretados por células endócrinas especializadas e transportados pela circulação, para se ligarem a receptores em células-alvo em localizações distantes, no corpo	Liberação de hormônios glicocorticóides pelas glândulas adrenais
Sinalização parácrina	Uma molécula liberada de uma célula atua localmente para afetar células-alvo próximas	Neurotransmissores e receptores; óxido nítrico no sistema imune, sistema nervoso, etc.
Sinalização autócrina	Uma célula produz uma molécula sinalizadora, à qual ela também responde	Linfócitos T podem responder à estimulação antigênica por meio da síntese de fatores que orientam a sua própria proliferação

Tabela 8.4 Exemplos de elementos de resposta na expressão gênica induzível

Elementos de resposta consensuais (RE)	Resposta a:	Fator protéico que reconhece o RE
(T/G)(T/A)CGTCA	cAMP	CREB (também chamado ATF)
CC(A/T)(A/T)(A/T)(A/T)(A/T)(A/T)GG	Fator de crescimento do soro	Fator de resposta do soro
TTNCNNNA	Gama-interferon	Stat-1
TGCGCCCGCC	Metais pesados	Mep-1
TGAGTCAG	Ésteres de forbol	AP1
CTNGAATNTTCTAGA	Choque térmico	HSP70, etc.

Ver também elementos de resposta para receptores de esteróides na Figura 8.9.

drão da expressão gênica nas células expostas a mudanças nesses parâmetros. Em animais multicelulares complexos também existem necessidades fundamentais para as células se comunicarem umas com as outras, e diferentes formas de sinalização celular são possíveis (Tabela 8.3). Em alguns casos, a alteração da expressão gênica é conduzida em nível de transcrição, o qual pode oferecer certas vantagens (Seção 8.2.5). Em outros casos, a expressão gênica é alterada pela modulação da transcrição.

A regulação da transcrição em resposta a um sinal celular pode acontecer de diferentes maneiras, mas o ponto final é sempre o mesmo: um fator de transcrição, previamente inativo, é ativado especificamente pela rota de sinalização e, depois, liga-se, subsequentemente, a seqüências reguladoras específicas, localizadas nos promotores dos genes-alvo, ativando, por meio disso, a transcrição desses genes. No caso da transcrição regulada por moléculas sinalizadoras ou seus intermediários, tais seqüências reguladoras são freqüentemente referidas como elementos de resposta (Tabela 8.4).

### Fatores de transcrição induzíveis por ligantes

Pequenos hormônios hidrofóbicos e morfógenos, tais como hormônios esteróides, tiroxina e ácido retinóico, são capazes de difundir-se pela membrana plasmática da célula-alvo e ligar-se a receptores intracelulares no citoplasma ou no núcleo. Esses receptores (freqüentemente chamados de receptores nucleares hormonais) são fatores induzíveis. Após a ligação do hormônio ligante, a proteína receptora associa-se a um elemento de resposta específico no DNA, localizado na região promotora de possivelmente 50 a 100 genes-alvo, e ativa a transcrição desses.

Embora tiroxina e ácido retinóico não sejam estrutural e biosinteticamente relacionados aos hormônios esteróides, seus receptores pertencem a uma superfamília de receptores nucleares comum. Dois domínios conservados caracterizam a família: um **domínio de ligação ao DNA**, localizado centralmente, com cerca de 68 aminoácidos, e um **domínio de ligação ao ligante**, com ~240 aminoácidos, localizado próximo à região C-terminal (Figura 8.9). O domínio de ligação ao DNA contém dedos de zinco e liga-se como um dímero; cada monômero reconhece um de dois hexa-

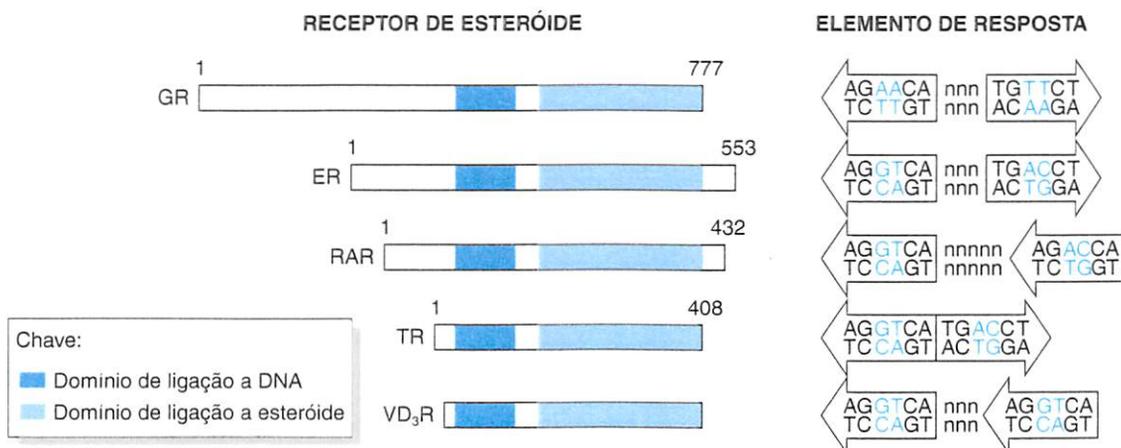


Figura 8.9 Receptores de esteróides e os respectivos elementos de resposta.

(A) Estrutura dos membros da superfamília de receptor nuclear. Os números referem-se ao tamanho da proteína em aminoácidos. Abreviaturas: ER, receptor de estrógeno; GR, receptor de glicocorticóide; PR, receptor de progesterona; RAR, receptor de ácido retinóico; TR, receptor de tiroxina; VDR, receptor de vitamina D. (B) Elementos de resposta. Note que os elementos de resposta são freqüentemente repetições de hexanucleotídeos invertidas, perfeitas, mas que os elementos de resposta para o ácido retinóico e a vitamina D<sub>3</sub> são repetições de hexanucleotídeos diretas, imperfeitas. Note também que todos os hexanucleotídeos possuem a seqüência geral AGNNCA, com os dois nucleotídeos centrais, mostrados em azul, conferindo especificidade e pertencendo a uma de três classes: GT, AA ou AC. Abreviatura: n, qualquer nucleotídeo.

nucleotídeos no elemento de resposta. Os dois hexanucleotídeos são repetições invertidas ou repetições diretas, as quais estão tipicamente separadas por três ou cinco nucleotídeos (Figura 8.9). Na ausência do ligante, o receptor é inativado por repressão direta da função do domínio de ligação ao DNA, exercida pelo domínio de ligação ao ligante, ou pela ligação de uma proteína inibitória, como no caso do receptor de glicocorticóide (Figura 8.10).

#### Ativação de fatores de transcrição pela transdução de sinal

Diferente dos hormônios lipídeo-solúveis ou morfógenos, as moléculas sinalizadoras hidrofílicas, tais como hormônios polipeptídicos, não podem difundir-se pela membrana plasmática. Em vez disso, elas se ligam a receptores específicos na superfície celular. Após a ligação da molécula ligante, o receptor sofre uma mudança conformacional e se torna ativado de uma forma tal que transmite o sinal por meio de outras moléculas da célula (transdução de sinal). Várias classes de receptores de superfície celular são conhecidas, mas muitas delas possuem uma atividade quinase, ou podem ativar quinases intracelulares (Tabela 8.5). Rotas de transdução de sinal são freqüentemente caracterizadas por complexos eventos reguladores recíprocos entre quinases e fosfatases, as quais podem ativar ou reprimir intermediários por fosforilação/defosforilação. Em muitos casos, a fosforilação ou a defosforilação induzem uma modificação conformacional. No caso da ativação de uma molécula sinalizadora, a conformação alterada freqüentemente significa que um fator sinalizador não está mais inibido por alguma seqüência repressora presente em uma proteína inibitória à qual ele está ligado, ou em um domínio ou motivo de seqüência de sua própria estrutura.

Em termos de ativação da transcrição, dois mecanismos gerais permitem a transmissão rápida de sinais para o núcleo a partir de receptores de superfície celulares, ambos envolvendo a fosforilação de proteínas:

- proteíno-quinases são ativadas e depois translocadas do citoplasma para o núcleo, onde elas fosforilam os fatores de transcrição alvos;
- fatores de transcrição inativos mantidos no citoplasma são ativados por fosforilação e translocados para dentro do núcleo.

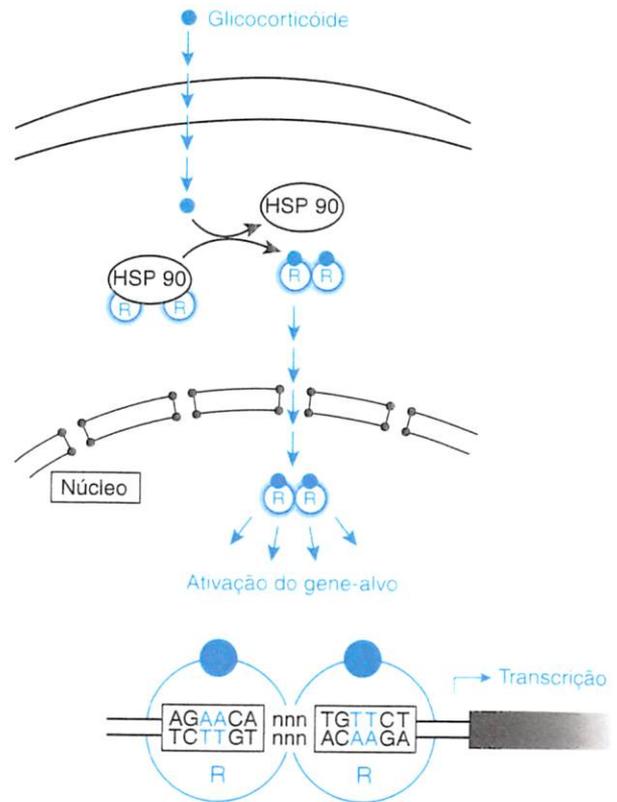


Figura 8.10 Regulação da transcrição por glicocorticóides.

O receptor de glicocorticóide é normalmente inativado por ligação a uma proteína inibidora, Hsp90. A ligação de glicocorticóides ao receptor de glicocorticóide libera Hsp90, o receptor forma dímeros e, então, ativa genes selecionados, os quais possuem um elemento de resposta a glicocorticóides em seus promotores (Figura 8.9).

Tabela 8.5 Principais classes de receptores de superfície celular

Classe de receptor	Característica	Ativação por
Proteína G-unida	Ativa proteínas G heterotrímicas (proteínas reguladoras de ligação a GTP). Estas últimas consistem de três subunidades, $\alpha$ , $\beta$ e $\gamma$ . Sob ativação, a subunidade $\alpha$ transloca para ativar uma proteína-alvo (ver Figura 8.11A para exemplo)	Várias moléculas, incluindo peptídeos, hormônios, p. ex., epinefrina, etc.
Serina-treonino quinase	Ativa proteínas intracelulares pela fosforilação de resíduos de serina ou treonina de proteínas-alvo	Hormônios, fatores de crescimento
Tirosino-quinase	Ativa proteínas intracelulares pela fosforilação de tirosinas das proteínas-alvo, p. ex., receptor de PDGF, receptor de insulina, etc (Figura 8.11B)	Hormônios, fatores de crescimento
Tirosino-quinase associada	Receptores não fosforilam proteínas-alvo diretamente, eles, dependem de uma tirosino-quinase associada. Exemplos incluem receptores de citocinas com atividade JAK (do inglês <i>janus protein kinase</i> ) envolvida na sinalização JAK-STAT	Hormônios, fatores de crescimento
Íon ligado a canal	Envolvido na sinalização sináptica rápida	Neurotransmissores

As duas seções seguintes fornecem exemplos para ilustrar esses dois mecanismos (ver também Karin e Hunter, 1995).

#### Sinalização hormonal por meio da rota do AMP cíclico

O AMP cíclico (cAMP) é um importante mensageiro secundário (Tabela 8.6), que atua em resposta a uma variedade de hormônios e outras moléculas sinalizadoras. Ele é sintetizado a partir de ATP, por uma enzima ligada à membrana, adenilato-ciclase. Os hormônios que ativam a adenilato-ciclase ligam-se a um receptor de superfície celular que é da classe dos receptores associados à proteína G. A ligação do hormônio ao receptor promove a interação do receptor com a proteína G, a qual consiste de três subunidades,  $\alpha$ ,  $\beta$  e  $\gamma$ . Após essa interação, a subunidade  $\alpha$  da proteína G é ativada, causando a sua dissociação e estimulando a adenilato-ciclase.

O aumento do cAMP intracelular, produzido pela adenilato-ciclase ativada, pode então ativar a transcrição de seqüências-alvo específicas que contêm um elemento de resposta a cAMP ou CRE (do inglês *cAMP response element*). Essa função do cAMP é mediada pela enzima proteíno-quinase A. O AMP cíclico liga-se à proteíno-quinase e ativa-a por permitir a liberação de duas subunidades cataliticamente ativas, que entram no núcleo e fosforilam um fator de transcrição específico, CREB (do inglês *CRE-binding protein*, proteína que se liga ao CRE). O CREB ativado ativa, então, a transcrição de genes com o elemento de resposta a cAMP (Figura 8.11A).

#### Ativação de NF- $\kappa$ B pela sinalização da proteíno-quinase C

O NF- $\kappa$ B é um fator de transcrição que está envolvido em uma variedade de aspectos da resposta imune. Em seu estado inativado, o NF- $\kappa$ B é retido no citoplasma, onde ele está complexado com uma subunidade inibitória, I $\kappa$ B. No entanto, esta última pode ser marcada para degradação, após a fosforilação pela proteíno-quinase C. A destruição conseqüente de I $\kappa$ B permite NF- $\kappa$ B translocar para o núcleo e ativar seus vários genes-alvo. A proteíno-quinase C é ativada por diacilglicerol. Esse último é produzido depois que a ligação de vários fatores de crescimento e hormônios a receptores de superfície celular específicos disparam a ativação da atividade da fosfolipase C ligada ao receptor. A enzima ativada converte PIP<sub>2</sub> (fosfatidilinositol 4,5-bisfosfato) a IP<sub>3</sub> (inositol 1,4,5-trifosfato) e diacilglicerol (Figura 8.11B).

#### 8.2.5 O controle da tradução da expressão gênica pode envolver o reconhecimento específico de seqüências reguladoras por proteínas de ligação a RNA, em seqüências não-traduzidas do mRNA

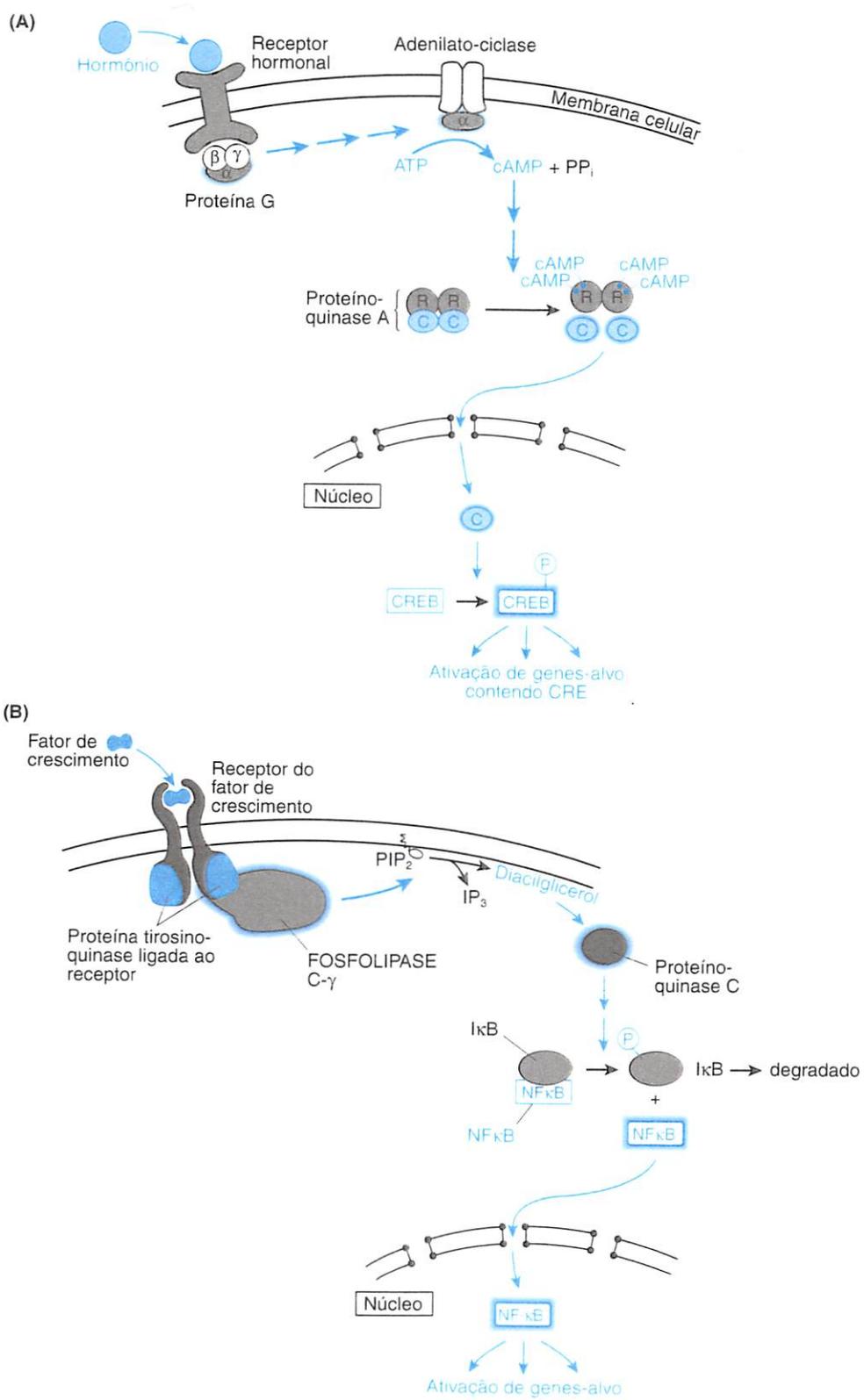
Diferentes formas de controle da tradução da expressão gênica são evidentes, e um número crescente de espécies de mRNA de eucariotos e mamíferos mostrou conter seqüências reguladoras nas suas seqüências não-traduzidas (mais freqüentemente na extremidade 3'; ver Wickens *et al.*, 1997; Day e Tuite, 1998). Várias proteínas de ligação a RNA, eucarióticas e de mamíferos, foram identificadas e mostraram ligar-se a seqüências reguladoras específicas, presentes em seqüências não-traduzidas, fornecendo, dessa forma, as bases para o controle da expressão gênica pelo controle da tradução (Siomi e Dreyfuss, 1997). Diferentes domínios de ligação ao RNA foram identificados e eles incluem elementos que foram previamente associados com propriedades de ligação ao DNA de fatores de transcrição, tais como os dedos de zinco e os homeodomínios (ver Seção 8.2.3 e Siomi e Dreyfuss, 1997).

#### Localização intracelular do RNA

A interação entre elementos reguladores *cis*-atuantes no RNA e proteínas de ligação a RNA, *trans*-atuantes, pode ser considerada para alterar a estrutura do RNA de várias maneiras: facilitando ou atrapalhando interações com outros fatores *trans*-atuantes; alterando estruturas de ordem superior do RNA; criando o encontro de seqüências do RNA inicialmente distantes, ou fornecendo a localização ou os sinais de marcação para o transporte de moléculas de RNA para locais intracelulares específicos. No último caso, sabe-se que numerosos mRNAs eucarióticos e de mamíferos são transportados como partículas ribonucleoprotéicas (RNP) para localizações específicas nas células; foi demonstrado que o transporte sobre microtúbulos e filamentos de actina em alguns casos necessita de motores moleculares específicos (Hazelrigg, 1998). Por exemplo, nos neurônios maduros o mRNA tau está localizado nas porções proximais dos axônios, em vez de nos dendritos, onde estão localizadas muitas moléculas de mRNA; o mRNA da proteína básica mielina é transportado com o auxílio da quinesina para o processamento dos oligodendrócitos.

**Tabela 8.6** Exemplos de mensageiros secundários na sinalização celular

Mensageiro secundário	Características	Exemplos
AMP cíclico (cAMP)	Produzido a partir de ATP, pela adenilato-ciclase. Efeitos são normalmente mediados por proteíno-quinase A	Ativação do fator de transcrição CREB (Figura 8.11A)
GMP cíclico (cGMP)	Produzido a partir de GTP, pela guanilato-ciclase. O papel melhor caracterizado é na recepção visual nos olhos de vertebrados	
Fosfolipídeos/Ca <sup>2+</sup>	Ativados a jusante dos receptores de proteína G-unidas e proteínas tirosino-quinases. Hidrólise de fosfatidilinositol 4,5-bis-fosfato (PIP <sub>2</sub> ) origina diacilglicerol e inositol 1,4,5-trifosfato (IP <sub>3</sub> ), que ativa a proteíno-quinase C e mobiliza Ca de estoques intracelulares	Ver Figura 8.11B



Uma razão para tais mecanismos de localização intracelular do mRNA é que eles fornecem uma maneira mais eficiente para localizar produtos protéicos do que a marcação de proteínas: como é detalhado abaixo, um único mRNA pode originar várias moléculas protéicas, presumindo-se que ele possa se ligar com ribossomos. Diferentes etapas seqüenciais foram consideradas: repressão no início da tradução, transporte dentro da célula, localização (para o destino subcelular específico) e, depois, a tradução dependente da localização. Recentemente, seqüências reguladoras-chave, as quais são necessárias para várias etapas nesse processo, foram identificadas nas seqüências não-traduzidas, predominantemente a UTR 3' de muitas espécies de mRNA (Hazelrigg, 1998). Por exemplo, dois elementos dentro da UTR 3' do mRNA da proteína básica mielina são necessários para etapas distintas de seu transporte, para o processamento dos oligodendrócitos: uma seqüência transportadora de RNA de 21 nucleotídeos e uma região de localização mais longa (Ainger *et al.*, 1997).

#### **O controle da tradução da expressão gênica em resposta a um estímulo externo**

O controle da tradução da expressão gênica pode permitir uma resposta mais rápida a um estímulo ambiental alterado que a alternativa ativação da transcrição. O metabolismo do ferro fornece dois exemplos úteis. O aumento nos níveis de ferro estimula a síntese da proteína de ligação a ferro, a ferritina, sem nenhum aumento correspondente na quantidade de mRNA para ferritina. Inversamente, a diminuição nos níveis de ferro estimula a produção do receptor da transferrina (TfR) sem qualquer efeito na produção do mRNA para o receptor da transferrina. As UTRs 5' dos mRNAs de ambas as cadeias, pesada e leve, da ferritina contêm um único elemento de resposta ao ferro (IRE, do inglês *iron-response element*), uma seqüência reguladora *cis*-atuante específica que forma uma estrutura de grampo. Várias dessas seqüências são também encontradas na UTR 3' do mRNA do receptor da transferrina (ver Klausner *et al.*, 1993). A regulação é exercida pela ligação dos IREs por uma proteína de ligação ao IRE, que é ativada por baixos níveis de ferro (Figura 8.12).

#### **Controle da tradução da expressão gênica durante o desenvolvimento inicial**

A expressão gênica, durante a maturação dos oócitos e nos estágios embrionários iniciais, é regulada em nível de tradução, não

de transcrição. Após a fertilização de um oócito humano, nenhum mRNA é inicialmente produzido, até o estágio celular de 4 a 8 células, quando a transcrição zigótica é ativada, ou seja, a transcrição dos genes presentes no zigoto. Antes desse período, as funções celulares são especificadas pelo mRNA materno, que foi previamente sintetizado durante a oogênese. Enquanto ainda não está esclarecida qual a extensão da regulação da expressão de genes humanos durante esse período, correspondente a dos organismos modelos nesse estágio, extrapolações a partir desses últimos poderiam sugerir que uma variedade de mRNAs é armazenada nos oócitos em uma forma inativa, caracterizada por apresentar caudas de poli(A) curtas. Tais mRNAs foram previamente submetidos à desadenilação, e as caudas de poli(A) curtas resultantes significam que eles não podem ser traduzidos. Subseqüentemente, na fertilização ou, mais tarde, no desenvolvimento, as espécies de mRNAs inativas armazenadas podem ser ativadas pela poliadenilação citoplasmática, que restaura o tamanho normal das caudas de poli(A). A poliadenilação citoplasmática parece utilizar o mesmo tipo de atividade de poli(A)-polimerase que a poliadenilação-padrão de mRNAs recentemente formados (que ocorre no núcleo). No entanto, além do sinal AAUAAA, o mRNA necessita possuir um *elemento de poliadenilação citoplasmático* a montante, rico em uracilas (Wahle e Kuhn, 1997). Outro mecanismo que é utilizado para regular a tradução de alguns mRNAs durante o desenvolvimento é a repressão da tradução (ocultamento), pela qual proteínas de ligação a RNA podem reconhecer e ligar-se a seqüências específicas nas UTRs 3' dos mRNAs, reprimindo, desse modo, a tradução (ver Wickens *et al.*, 1997; Stebbins-Boaz e Richter, 1997).

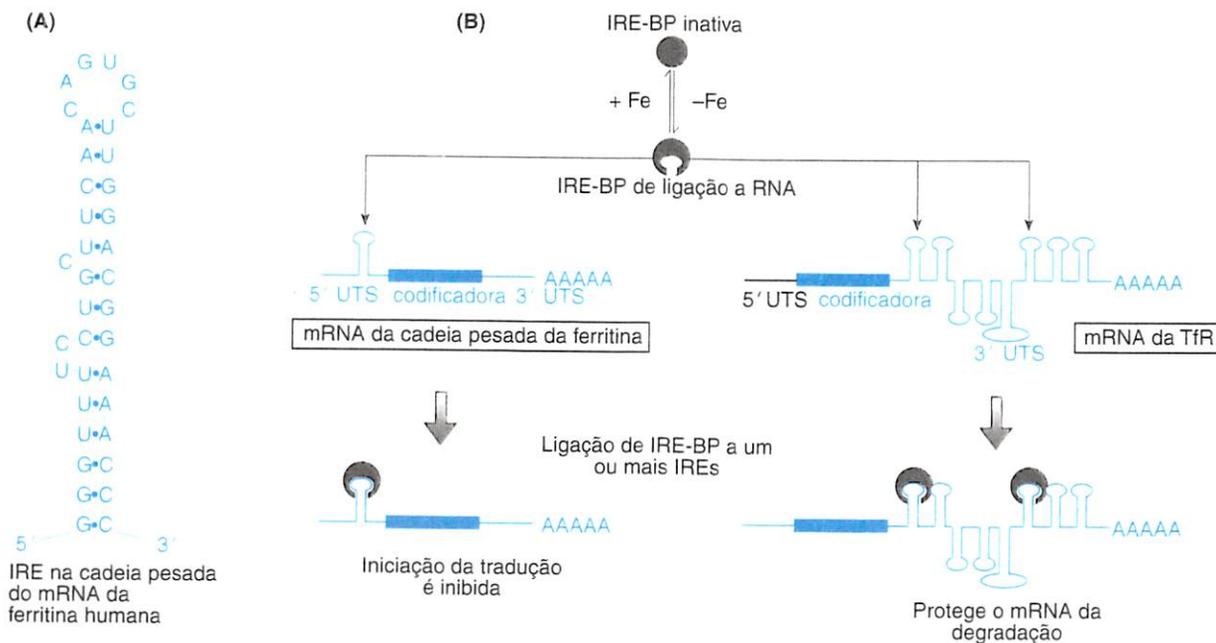
3

### **8.3 Transcrição alternativa e processamento de genes individuais**

Além do controle que é exercido na seleção de genes específicos (ou de seus transcritos) para ativação ou repressão, os mecanismos de controle também podem selecionar entre transcritos alternativos específicos de um único gene. A utilização diferencial do promotor ou os eventos diferenciais de processamento do RNA podem resultar em um grande número de isoformas diferentes, e esses e outros mecanismos têm desafiado a definição clássica de um gene (Quadro 8.3).

**Figura 8.11** Genes-alvo selecionados podem ser ativamente expressos em resposta a estímulo extracelular por transdução de sinal, a partir de receptores de superfície celular ativados.

(A) Um exemplo de ativação de uma proteíno-quinase e translocação para o núcleo: a sinalização hormonal pela rota de transdução de sinal do AMP cíclico-proteíno-quinase A. A ligação de hormônio a um receptor de superfície celular promove a interação do receptor com uma proteína G. A subunidade  $\alpha$  ativada da proteína G dissocia-se do receptor e estimula a adenilato-ciclase ligada à membrana a sintetizar cAMP. Esta última fosforila as subunidades reguladoras da proteíno-quinase A, permitindo a liberação das subunidades catalíticas, que migram para o núcleo e ativam o fator de transcrição CREB (proteína de ligação a CRE) por fosforilação. O CREB ativado liga-se ao elemento de resposta a cAMP nos promotores de genes-alvo. (B) Um exemplo de ativação de um fator de transcrição citoplasmático (NF- $\kappa$ B) e translocação para o núcleo. Em resposta a um fator de crescimento externo, uma tirosino-quinase ligada ao receptor do fator de crescimento fosforila a fosfolipase C que, então, converte fosfatidilinositol 4,5-bisfosfato (PIP<sub>2</sub>) a diacilglicerol e inositol 1,4,5-trifosfato (IP<sub>3</sub>). O diacilglicerol pode ativar a proteíno-quinase C, que fosforila o fator inibitório I $\kappa$ B e, assim, marca-o para a sua degradação. Como resultado, o fator de transcrição NF- $\kappa$ B é liberado e migra para o núcleo para ativar seus genes-alvo.



**Figura 8.12** A proteína de ligação a IRE regula a produção da cadeia pesada de ferritina e do receptor da transferrina por meio da sua ligação aos elementos de resposta a ferro (IREs) nas regiões 5'- ou 3'-não-traduzidas.

(A) Estrutura do IRE na UTR-5' da cadeia pesada da ferritina. (B) A ligação da proteína de ligação a IRE (IRE-BP) aos mRNAs da ferritina e da transferrina (TfR) possui efeitos contrastantes na síntese proteica. Por simplificação, somente alguns dos elementos IRE no mRNA da TfR mostram estar ligados pela proteína de ligação a IRE.

### 8.3.1 A transcrição de um único gene humano pode ser iniciada a partir de uma variedade de promotores alternativos e pode resultar em uma variedade de isoformas tecido-específicas

Vários genes humanos e de mamíferos possuem dois ou mais promotores alternativos, que podem resultar em diferentes isoformas, com diferentes propriedades (ver Ayoubi e van de Vem, 1996). As isoformas podem proporcionar:

- especificidade de tecido (uma ocorrência freqüente; ver o exemplo do gene da distrofia humana abaixo);
- especificidade do estágio de desenvolvimento (por exemplo, o gene do fator II de crescimento, semelhante à insulina);
- localização subcelular diferencial (por exemplo, isoformas solúveis e ligadas à membrana);
- capacidade funcional diferencial (como no caso do receptor da progesterona);
- regulação gênica sexo-específica (ver o caso do gene da metil-transferase *Dnmt1* na Seção 8.4.2 e Figura 8.20).

Um dos exemplos mais celebrados da utilização diferencial de promotores em humanos envolve o gigantesco gene da distrofi-

na (*DMD*), que compreende um total de mais de 79 éxons, distribuídos sobre cerca de 2,4 Mb de DNA, na banda cromossômica Xp21. Ao menos oito promotores alternativos diferentes podem ser utilizados (Cox e Kunkel, 1997). Quatro desses promotores alternativos estão localizados próximo ao sítio de iniciação convencional e compreendem um promotor córtex cerebral-específico, um promotor músculo-específico localizado a 100 kb a jusante, um promotor que é utilizado nas células de Purkinje do cerebelo e localizado 100 kb mais distante a jusante, e um promotor linfócito-específico (Figura 8.13). A utilização desses promotores resulta em isoformas grandes, com um peso molecular de 427 kDa (referidas como Dp427, onde Dp = proteína *D*istrofi-na e freqüentemente recebem um sufixo para indicar a especificidade de tecido, por exemplo, Dp427-M para indicar a isoforma músculo-específica). As quatro isoformas diferem em suas seqüências N-terminais mais extremas, como resultado de utilizarem quatro alternativas diferentes para o éxon 1.

Além dos quatro promotores alternativos, que codificam as isoformas grandes convencionais, ao menos quatro outros promotores alternativos internos podem ser utilizados. A transcrição a partir desses promotores utiliza somente um subgrupo de éxons a jusante, resultando em isoformas significativamente menores: uma isoforma Dp260 produzida nas células da retina;

## Quadro 8.3

## A visão clássica de um gene não é mais válida

Classicamente, um gene era visto como uma entidade que codifica um único RNA ou produto polipeptídico. Ao contrário, o conceito de produtos diferentes sendo codificados por uma única *unidade de transcrição* (um segmento de DNA que é continuamente transcrito em RNA) é há bastante tempo conhecido nos genomas simples, tais como genomas bacterianos, onde os chamados mRNAs policistrônicos são comuns. Eles aparecem por meio da transcrição contínua por vários genes adjacentes (o termo *cistron* é uma palavra essencialmente fora de uso para gene). Tais transcritos multigênicos são, então, processados para gerar dois ou mais polipeptídeos diferentes. Em tais casos, o termo *unidade de transcrição* é claramente não-equivalente funcionalmente ao termo *gene*, porque cada unidade de transcrição corresponde a vários genes. Em genomas complexos, tais como o genoma humano, porém, a grande maioria de genes é transcrita individualmente e, nesses casos, os termos *gene* e *unidade de transcrição* são essencialmente equivalentes. Os genes das imunoglobulinas e receptores de células T, no entanto, fornecem uma complexidade adicional: os genes individuais são muito grandes no DNA da linhagem germinativa (Figura 7.7 e Figura 8.26), mas as unidades de transcrição correspondentes são muito menores e de tamanhos variáveis devido aos rearranjos de DNA célula-específicos em linfócitos B e T individuais (Figura 8.27 e 8.28).

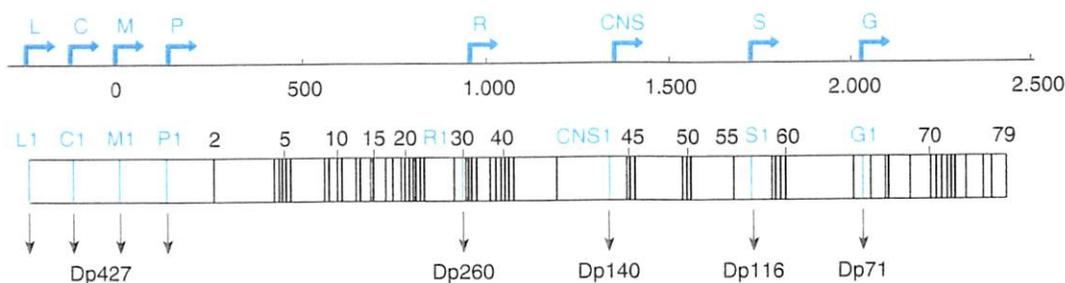
Além disso, devido aos vários outros mecanismos de expressão listados abaixo, genes humanos e unidades de transcrição frequentemente codificam uma variedade de produtos gênicos

Mecanismo	Freqüência e exemplos
Unidades de transcrição	Rara. Exemplos incluem genes de rRNA de 18S, multigênicos 28S e 5,8S (Figura 8.1) e genes mitocondriais ver Seção 7.1.1 e Figura 7.2)
Utilização de promotores	Bastante comum. Ver adiante e Figura 8.13 alternativos
Encadeamento alternativo	Muito freqüente. Ver Seção 8.3.2 e Figura 8.14
Poliadenilação alternativa	Bastante comum. Ver Figura 8.15
Edição de RNA	Extremamente rara. Ver Seção 8.3.3 e Figura 8.16
Clivagem posterior à tradução	Rara. Pode gerar polipeptídeos funcionalmente relacionados, como no caso de vários hormônios humanos, p. ex., insulina humana. Ver Figura 1.23.

uma isoforma Dp140 produzida por muitas células do cérebro e rim; uma isoforma Dp116 produzida nas células de Schwann e uma isoforma pequena Dp71 produzida em muitos tipos celulares (Figura 8.13). Note que a utilização alternativa de promotores implica a utilização alternativa de éxons, mas eventos de encadeamento alternativos, que são independentes da utilização diferencial de promotores, também são muito comuns (Seção 8.3.2). No caso da distrofina, por exemplo, a complexidade adicional das isoformas é introduzida por encadeamento alternativo, especialmente na região C-terminal.

### 8.3.2 Genes humanos freqüentemente codificam mais do que um produto, como resultado de encadeamento alternativo e eventos de poliadenilação alternativa

Além da utilização diferencial de promotores, que implica na utilização alternativa de éxons, uma variedade de eventos de processamento de RNA alternativos pode também resultar em isoformas alternativas. Os mecanismos primários são eventos alternativos de encadeamento (isto é, diferentes daqueles indu-



**Figura 8.13** Ao menos oito promotores podem ser utilizados para gerar expressão específica do tipo celular no gene da distrofina.

As posições dos oito promotores alternativos estão ilustradas na parte superior: L, linfócito; C, cortical; M, músculo; P, Purkinje; R, retinal; CNS, sistema nervoso central; S, célula de Schwann; G, geral. As posições aproximadas dos éxons estão ilustradas abaixo. Note que cada promotor utiliza seu próprio primeiro éxon (em azul: L1, C1, M1, P1, R1, CNS1, S1 e G1) juntamente com os éxons a jusante (em preto). Todos os 78 éxons são utilizados no caso das distrofinas C-, M- e P- completas para gerar um produto que, em cada caso, tem cerca de 427 kDa (Dp427). Os outros promotores estão localizados imediatamente a montante dos éxons, indicados como segue: R, éxon 30; CNS, éxon 45; S, éxon 56; G, éxon 63. Note, também, que a iniciação da transcrição para a isoforma Dp140 não ocorre até o éxon 51, embora seja conhecido que o promotor está no intron 44. Além da diversidade de isoformas geradas pela utilização diferencial de promotores, sabe-se que o encadeamento alternativo ocorre, principalmente na extremidade 3'.

### Quadro 8.4

#### O encadeamento alternativo pode alterar as propriedades funcionais de uma proteína

A seguinte lista, longe de estar completa, é somente uma forma de ilustrar algumas maneiras pelas quais as propriedades biológicas de uma proteína podem ser alteradas como resultado de encadeamento alternativo.

- **Isoformas tecido-específicas.** Exemplos incluem isoformas codificadas pelo gene da tropomiosina. Ver também o exemplo do gene da calcitonina, que está listado abaixo (Seção 8.3.2 e Figura 8.15).
- **Isoformas solúveis e ligadas à membrana.** A localização protéica pode ser regulada pela geração de formas solúveis de numerosos receptores de membrana, p. ex., HLA de classe I e II, IgM, CD8, receptor de hormônio de crescimento, IL-4, IL-5, IL-7, IL, eritropoietina, G-CSF, G-MCSF, LIF (do inglês *leukaemia inhibitory factor*, fator inibidor da leucemia) e o receptor FAS-sinalizador de apoptose.
- **Localização intracelular alternativa.** Um exemplo útil é fornecido pelo gene do tumor de Wilm's *WT1*, que especifica uma proteína com quatro dedos de zinco em seu terminal C e possui muitas isoformas. O encadeamento diferencial pode levar à inclusão ou à exclusão de uma seqüência de 17 aminoácidos no domínio transregulador, e também à inclusão ou à exclusão de uma seqüência de três aminoácidos, lisina-treonina-serina, localizada entre o terceiro e o quarto dedos de zinco (gerando isoformas que são denominadas +KTS e -KTS a partir do código de uma letra para aminoácidos – ver Figura 8.14). Essas isoformas são bastante conservadas na evolução e diferem em vários aspectos, incluindo seus padrões de distribuição subnuclear: as isoformas +KTS estão especificamente localizadas nos sítios de encadeossomos no núcleo, enquanto as isoformas -KTS estão geralmente distribuídas no nucleoplasma (Larsson *et al.*, 1995).
- **Função alterada.** As isoformas +KTS e -KTS do produto gênico *WT1* também diferem em sua capacidade de ligar-se a seqüências de DNA específicas nos genes-alvo. Sabe-se que as primeiras possuem um papel na ligação de fatores de encadeamento; as últimas podem possuir um papel mais geral na ligação de domínios que abrigam fatores de transcrição gerais. Outros exemplos incluem (i) isoformas de fatores de transcrição, que são ativadores ou repressores da transcrição, dependendo da natureza dos domínios que são incluídos ou excluídos do produto protéico – ver Lopez, 1998; (ii) isoformas promotoras de apoptose ou indutoras de apoptose de vários genes, tal como o gene *lch-1* (caspase 2) (ver Jiang *et al.*, 1998); e (iii) isoformas alternativas do gene da calcitonina (Figura 8.15).

zidos pela utilização diferencial de promotores) e eventos alternativos de poliadenilação. Em muitos casos, uma combinação desses mecanismos pode ser originada durante o processamento de um único gene. Juntamente com a possibilidade adicional de utilização diferencial de promotores, esses mecanismos podem resultar em um número muito grande de isoformas para um único gene.

#### Encadeamento (splicing) alternativo

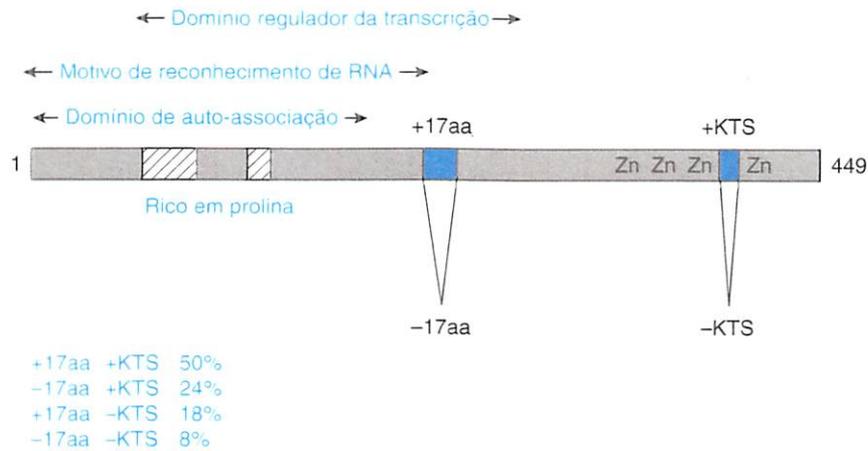
Uma grande porcentagem dos genes humanos sofre **encadeamento alternativo**, pelo qual diferentes combinações de éxons são incluídas em transcritos a partir do mesmo gene, durante o processamento do RNA. Para muitos genes, numerosas isoformas podem ser geradas em nível de RNA, mas freqüentemente a significância funcional é pobremente compreendida. Em alguns genes, o encadeamento alternativo resulta em uma diversidade bastante considerável nas regiões não-traduzidas. Por exemplo, o encadeamento alternativo no fígado resulta em ao menos oito diferentes seqüências UTR 5' para o mRNA do receptor do hormônio de crescimento humano (Pekhlitsky *et al.*, 1992), mas a significância funcional, se existe, não é entendida.

O encadeamento alternativo de éxons codificadores de seqüências também é comum e algumas das isoformas resultantes foram mostradas serem tecido-específicas, de tal forma que éxons individuais, presentes em uma isoforma mas não em outras, podem ser denominados “músculo-específicos”, “cérebro-específicos”, etc. As diferentes isoformas podem fornecer uma variedade de possibilidades de propriedades funcionais alteradas, mas o conhecimento detalhado da significância funcional das diferentes isoformas ainda é comparativamente escasso (Quadro 8.4).

O sistema modelo melhor entendido para a compreensão da regulação do encadeamento é a rota de determinação sexual em *Drosophila*, a qual controla a dosagem gênica. O encadeamento alternativo é utilizado em cada ramificação dessa rota, para controlar a expressão de reguladores da transcrição ou de proteínas associadas à cromatina que influenciam a transcrição e ambos os controles, positivo e negativo, de encadeamento, são evidentes (Lopez, 1998). Em células de mamíferos, os candidatos a reguladores de encadeamento são a família SR de proteínas de ligação a RNA (que possuem um domínio C-terminal distinto rico nos dipeptídeos serina[S]-arginina[R]) e algumas proteínas HnRNP (do inglês *heterogeneous nuclear ribonucleoprotein particle*, partícula ribonucleoprotéica heterogênea nuclear). Essas proteínas são conhecidas por promoverem várias etapas da montagem do encadeossomo (*spliceossomo*) e por se ligarem a **seqüências reforçadoras de encadeamento**, seqüências reguladoras que podem aumentar o reconhecimento do sítio de encadeamento (Lopez, 1998).

#### Poliadenilação alternativa

A utilização de sinais de poliadenilação alternativos também é bastante comum no mRNA humano; diferentes tipos de poliadenilação alternativos foram identificados (ver Edwards-Gilbert *et al.*, 1997). Em muitos genes, dois ou mais sinais de poliadenilação foram encontrados na UTR 3', e os transcritos alternativamente poliadenilados podem apresentar especificidade de tecido; em outros casos, sinais de poliadenilação alternativos podem ser produzidos após o encadeamento alternativo. Como um exemplo deste último, uma combinação de encadeamento alternativo e poliadenilação alternativa do gene da calcitonina (*CALC*) resulta na expressão tecido-específica de duas isoformas. A calcitonina, um hormônio homeostático  $Ca^{2+}$ -circulante, é produzida na tireóide; o peptídeo relacionado ao gene da calcitonina (CGRP), que pode ter ambas as atividades, neuromodulatória e trófica, é sintetizado no hipotálamo (Figura 8.15).



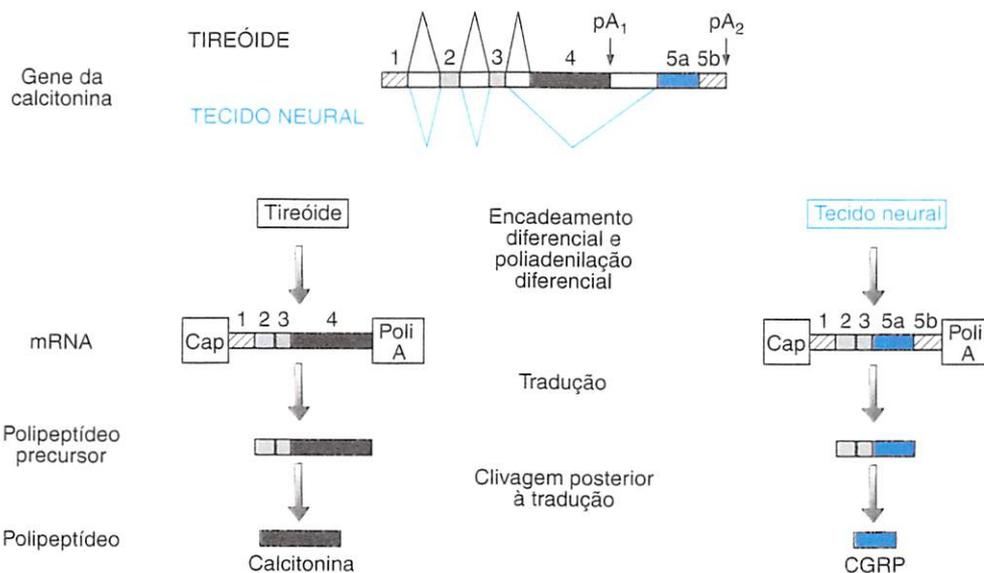
**Figura 8.14** Encadeamento diferencial no gene do tumor de Wilm's *WT1*.

A proteína WT1 contém um domínio regulador da transcrição e um possível motivo de reconhecimento de RNA, no seu terminal N. Quatro dedos de zinco (Zn) são encontrados em seu terminal C, cada um codificado por um éxon distinto. O encadeamento alternativo resulta na inclusão ou na exclusão de um éxon, que codifica uma seqüência de 17 aminoácidos, localizada centralmente, e também uma seqüência que especifica três aminoácidos (K, lisina; T, treonina; S, serina) localizada entre o terceiro e o quarto dedos de zinco. Todas as quatro formas possíveis de mRNA encadeadas são expressas em tecidos normais, que expressam WT1 nas razões aproximadas mostradas.

**8.3.3 A edição de RNA é uma forma rara de processamento posterior à tradução, pela qual mudanças base-específicas são enzimaticamente introduzidas em nível de RNA**

A edição de RNA é uma forma de processamento posterior à tradução que pode envolver a inserção ou a deleção de nucleo-

tídeos, mediada por enzimas, ou a substituição de nucleotídeos individuais em nível de RNA. A edição de RNA por inserção ou deleção parece ser uma propriedade peculiar da expressão gênica nas mitocôndrias de cinetoplastos de protozoários e bolores. A edição de RNA por substituição é freqüentemente empregada em alguns sistemas, tais como mitocôndrias e cloroplastos de plantas vasculares onde mRNAs individuais sofrem múltiplos



**Figura 8.15** Processamento diferencial do RNA resulta em produtos tecido-específicos do gene da calcitonina.

pA e pA<sub>2</sub> representam sinais de poliadenilação alternativos, que são empregados nos tecidos da tireóide e neurais, respectivamente, após o encadeamento alternativo. O éxon 1 e a porção 3' do éxon 5 (5b) codificam seqüências não-traduzidas 5' e 3', respectivamente. A calcitonina é codificada pelas seqüências do éxon 4 na tireóide, enquanto que o peptídeo relacionado ao gene da calcitonina (CGRP), que é sintetizado em tecidos neurais, é codificado pela porção 5' do éxon 5 (5a) como resultado de encadeamento alternativo.

eventos de edição de  $C \rightarrow U$  ou  $U \rightarrow C$ , e foi observada em uns poucos genes de mamíferos (Ashkenas, 1997). Pelo menos quatro classes de edição de RNA ocorrem nas células humanas:

- **Edição  $C \Rightarrow U$ .** A edição no mRNA da lipoproteína *APOB* humana foi bem-estudada. No fígado o gene *APOB* codifica um transcrito de mRNA de 14,1 kb e um produto de 4.536 aminoácidos, apoB100. No entanto, no intestino o mesmo gene codifica um mRNA de 7 kb, que contém um códon de terminação prematuro que não está presente no gene, e codifica um produto, apoB48, que é idêntico em sequência aos primeiros 2.152 aminoácidos de apoB100. Uma desaminase citosina-específica, *Apobec1*, converte uma única citosina, no nucleotídeo 6.666, no mRNA da *APOB* intestinal para uracila, gerando, assim, um códon de terminação (Figura 8.16).
- **Edição  $A \Rightarrow I$ .** Genes codificadores de alguns canais iônicos de penetração de ligantes, incluindo receptores de glutamato e proteínas relacionadas, estão sujeitos a esse tipo de edição de mRNA. Uma adenina é desaminada para originar *iosina (I)*, uma base que não está normalmente presente no mRNA (o grupamento amínico do carbono 6 da adenina é substituído por um grupamento carbonílico  $C=O$ ). A iosina parecia preferencialmente com citosina e também interage com ribossomos, durante a tradução, como se ela fosse uma G. No caso do gene do receptor B do glutamato, por exemplo, a edição do mRNA substitui um códon CAG (glutamina) por um CIG, que é traduzido como se fosse o códon CGG (arginina). Esse tipo de edição, que causa a modificação de uma  $Gln \Rightarrow Arg$ , é freqüentemente referido como edição Q/R, de acordo com o código de uma letra, para os dois aminoácidos envolvidos.
- **Outras classes de edição de RNA.** Duas outras formas documentadas de edição de mRNA em humanos são a edição U

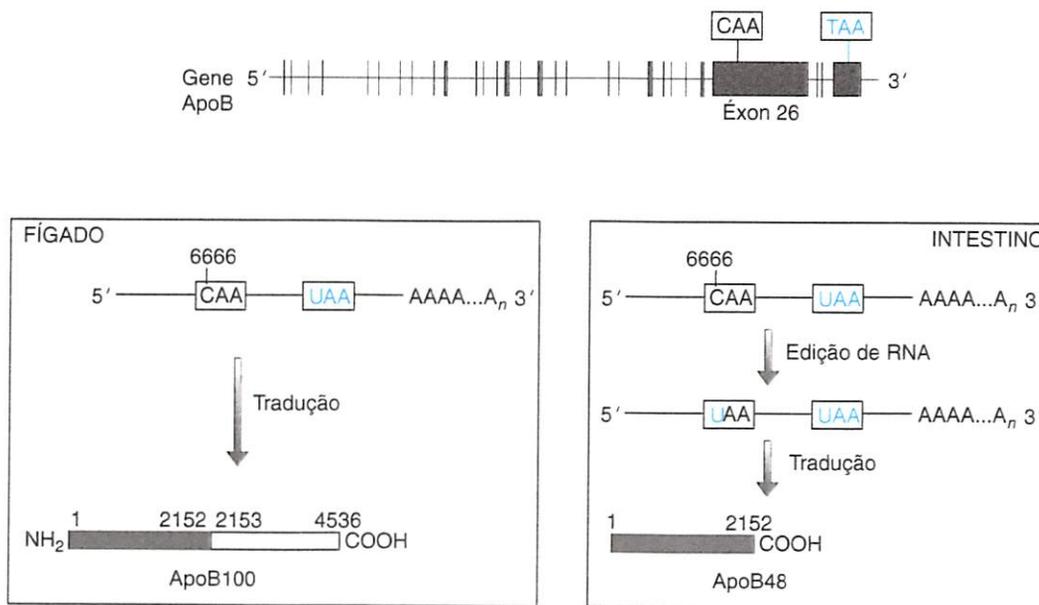
$\Rightarrow C$  no mRNA do gene do tumor de Wilms, *WT1*, e a edição  $U \Rightarrow A$  no mRNA da  $\alpha$ -galactosidase (Ashkenas, 1997).

## 8.4 Assimetria como uma maneira de estabelecimento diferencial da expressão gênica e metilação do DNA como uma maneira de perpetuação da expressão diferencial

O conceito de especificidade de tecido na expressão gênica humana está estabelecido há muito tempo. O que está bem menos claro é como tais padrões foram inicialmente estabelecidos. Uma vez que o conteúdo de DNA de todas as células nucleadas em um organismo é virtualmente idêntico, mecanismos genéticos não podem explicar como a expressão gênica diferencial desenvolveu-se primeiramente nas células. Para explicar isso, C. H. Waddington evocou mecanismos *epigenéticos* de controle gênico durante o desenvolvimento. Atualmente, uma variedade de mecanismos epigenéticos foram identificados, incluindo aqueles que podem perpetuar determinadas etapas da expressão gênica em linhagens celulares somáticas.

### 8.4.1 A expressão gênica seletiva em células de embriões de mamíferos muito provavelmente desenvolveu-se em resposta a eventos sinalizadores célula-célula, de curto alcance

A fim de explicar modelos subsequentes de expressão tecido-, célula- e estágio de desenvolvimento-específicos, algum mecanismo se faz necessário para estabelecer uma assimetria ou um eixo no óvulo fertilizado, ou bem no início do desenvolvimento. Na *Drosophila*, o óvulo é inerentemente assimétrico, devido à



**Figura 8.16** A expressão do gene da apolipoproteína B humana no intestino envolve edição de RNA tecido-específica.

Note que o códon 2153, especificado pela trinca CAA, nas posições nucleotídicas 6666-6668, especifica glutamina no produto ApoB100 sintetizado no fígado. No intestino, porém, o códon CAA é convertido, pela edição de RNA, no códon de terminação UAA, resultando em um produto mais curto, ApoB48.

transferência dos produtos gênicos de células “mantenedoras” (*nurse*), assimetricamente localizadas. O embrião desenvolve-se, inicialmente, como um sincício multinucleado (efetivamente como uma grande célula) e a regionalização depende da resposta de núcleos individuais a gradientes de longo alcance, de moléculas reguladoras. Em mamíferos, no entanto, a célula-ovo é relativamente pequena e o período inicial do desenvolvimento embrionário cria um agregado de células individuais aparentemente simétrico. Contudo, o desenvolvimento torna-se assimétrico.

A geração de assimetria nas células de mamíferos poderia derivar-se a partir de sinais iniciais de posição. Alguns aspectos do desenvolvimento inicial são assimetricamente inerentes, incluindo o local de entrada do espermatozoide durante a fertilização, a ligação do embrião à parede uterina, durante a implantação, e a localização de células com relação às suas vizinhas. Como o embrião se desenvolve em uma bola de células – e, mais tarde, à medida que estruturas mais complexas se desenvolvem, as células individuais irão variar em função do número de células vizinhas disponíveis. Eventos de sinalização intercelulares de curto alcance (pela sinalização direta célula-célula ou eventos sinalizadores intercelulares de curto alcance) podem fornecer uma maneira de identificar a posição celular e disparar a expressão gênica diferencial. Por exemplo, se uma molécula sinalizadora intercelular possui um alcance de, digamos, um diâmetro celular, então as células externas da blástula irão receber sinais diferentes daquelas que estão rodeadas por células vizinhas por todos os lados, e sinais posicionais diferentes podem ser traduzidos em expressão gênica diferencial. Como sistemas celulares específicos desenvolvem-se durante, por exemplo, a organogênese (realizada principalmente entre a 4ª e a 9ª semanas embrionárias), um tipo específico de crescimento celular ou de fatores de diferenciação, podem induzir a expressão de fatores de transcrição específicos para o estágio de desenvolvimento e/ou para tecidos.

#### 8.4.2 A metilação do DNA de vertebrados está amplamente confinada aos nucleotídeos CpG, e modelos de metilação do DNA podem ser herdados quando a célula se divide

Uma vez que modelos diferenciais de expressão tenham sido estabelecidos, mecanismos epigenéticos podem garantir que eles sejam estavelmente herdados quando a célula se divide. Imagine-se que a metilação do DNA exerça um papel principal a esse respeito, permitindo a transmissão estável de estados da cromatina que reprimem a expressão gênica de uma célula diplóide para as células-filha. No entanto, a função precisa da metilação do DNA em eucariotos ainda não está perfeitamente entendida e demonstra claramente diferenças entre as espécies. Alguns organismos, por exemplo, não possuem metilação do DNA detectável, como é o caso da *Drosophila*, da *C. elegans* e da levedura *Saccharomyces cerevisiae*. Em outros organismos em que a metilação do DNA ocorre, os modelos e as funções da metilação do DNA podem ser diferentes.

#### Modelos de metilação do DNA em vertebrados

O modelo de metilação do DNA em vertebrados difere daquele de células bacterianas. Nestas últimas, a adenina e a citosina podem ser metiladas, mas, em vertebrados, a metilação do DNA está restrita aos resíduos de citosina. Somente cerca de 3% das citosinas do DNA humano são metiladas, mas a maioria das

que o são, é encontrada no dinucleotídeo CpG (isto é, as citosinas metiladas são quase sempre aquelas cujo átomo do carbono 3' está ligado, por uma ligação fosfodiéster, ao átomo do carbono 5' de uma guanina). Além disso, uma porcentagem muito menor de citosinas metiladas ocorre nas seqüências CpNpG.

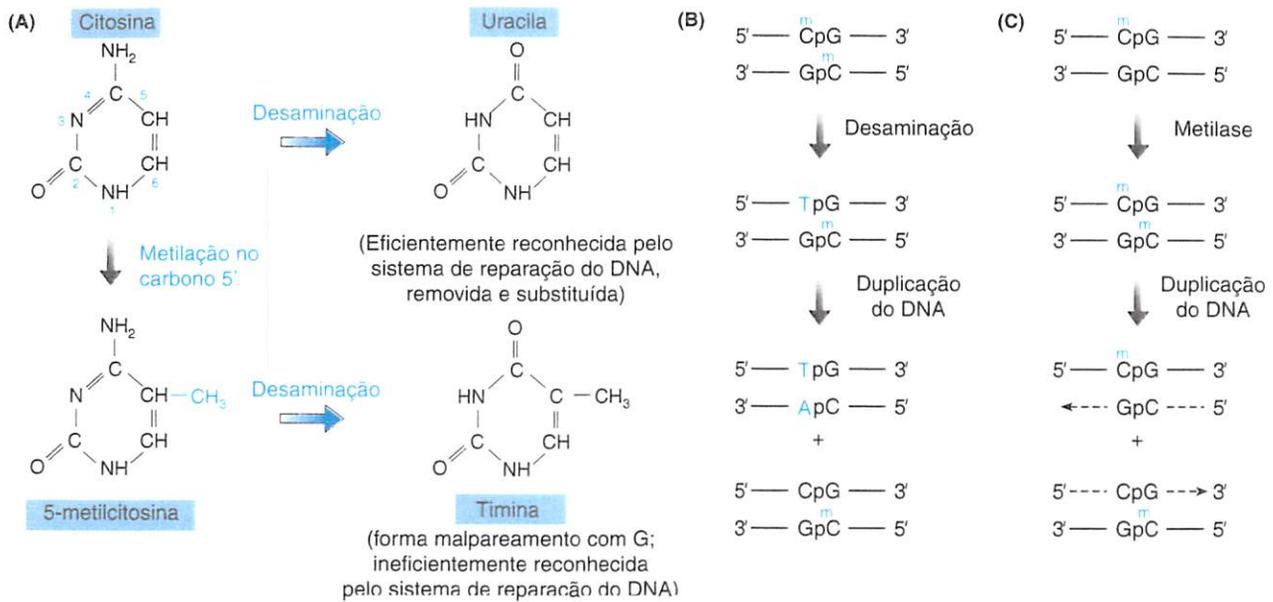
Os resíduos de citosina que ocorrem nos dinucleotídeos CpG no DNA de vertebrados são alvos de metilação por uma metil-transferase citosina-específica. A metilação ocorre no átomo do carbono 5' da citosina, para originar 5-metilcitosina, a qual é quimicamente instável e pode desaminar espontaneamente, originando a timina (Figura 8.17A). Em longos períodos de tempo evolucionário, o número de CpGs no DNA de vertebrados foi gradativamente sendo diminuído, embora sejam conhecidas regiões com freqüência normais (esperadas) de CpG, que, freqüentemente, indicam seqüências ativas em transcrição (ilhas CpG, ver Quadro 8.5).

#### Manutenção e metilação de novo durante o desenvolvimento

Ao contrário das metilases bacterianas, as metil-transferases das citosinas de vertebrados apresentam uma forte preferência pelo reconhecimento de um DNA-alvo hemimetilado (isto é, um que está metilado em apenas uma das fitas). A seqüência CpG apresenta simetria bilateral e, assim, após a replicação do DNA, a nova fita de DNA sintetizada irá receber o mesmo padrão de metilação CpG que o DNA parental (Figura 8.17B). Como resultado, o padrão de metilação CpG pode ser estavelmente transmitido para as células-filhas. A perpetuação de um padrão de metilação pré-existente é, algumas vezes conhecido como manutenção de metilação e é realizado em células de mamíferos, pelo produto do gene *Dnmt1*.

O padrão de distribuição de 5-metilcitosina no genoma de células somáticas diferenciadas varia de acordo com o tipo celular, mas a manutenção da metilação garante que os padrões de metilação de linhagens somáticas individuais sejam um tanto estáveis. Durante a gametogênese e o desenvolvimento embrionário, no entanto, existem mudanças dramáticas na metilação (Razin e Kafri, 1994). Os genomas das células germinativas primordiais do embrião absolutamente não são metilados. Após a diferenciação das gônadas e à medida que as células germinativas começam a se desenvolver, nova metilação *de novo* ocorre, causando uma metilação substancial do DNA de espermatozoides e de óvulos de mamíferos (Figura 8.19). O genoma do espermatozoide é mais fortemente metilado do que o genoma dos óvulos, e diferenças sexo-específicas nos padrões de metilação são encontradas principalmente nos loci pré-programados (ver Mertineit *et al.*, 1998 para referências). A metil-transferase codificada pelo gene *Dnmt1*, além de ser a DNA-metil-transferase de manutenção predominante nas células de mamíferos, também pode ser a principal metil-transferase *de novo*. Esse gene é altamente expresso em células germinativas masculinas, em oócitos maduros e nos estágios embrionários iniciais. A expressão do gene *Dnmt1* está sujeita à regulação sexo-específica, com promotores oócito- e espermatoócito-específicos, introduzindo éxons oócito- e espermatoócito-específicos, que originam diferentes produtos gênicos (Figura 8.20; Mertineit *et al.*, 1998).

O genoma do oócito fertilizado é um agregado dos genomas do espermatozoide e do óvulo, de tal forma que ele e o embrião no estágio bem inicial são substancialmente metilados, com diferenças de metilação nos alelos paternos e maternos de muitos genes. Mais tarde, nos estágios de mórula e blástula inicial, na pré-implantação do embrião, ocorre uma ampla desme-



**Figura 8.17** O dinucleotídeo CpG é pouco freqüente no DNA de vertebrados, porque é sujeito à metilação, e a 5-metilcitosina desaminada está sujeita a uma reparação ineficiente de DNA.

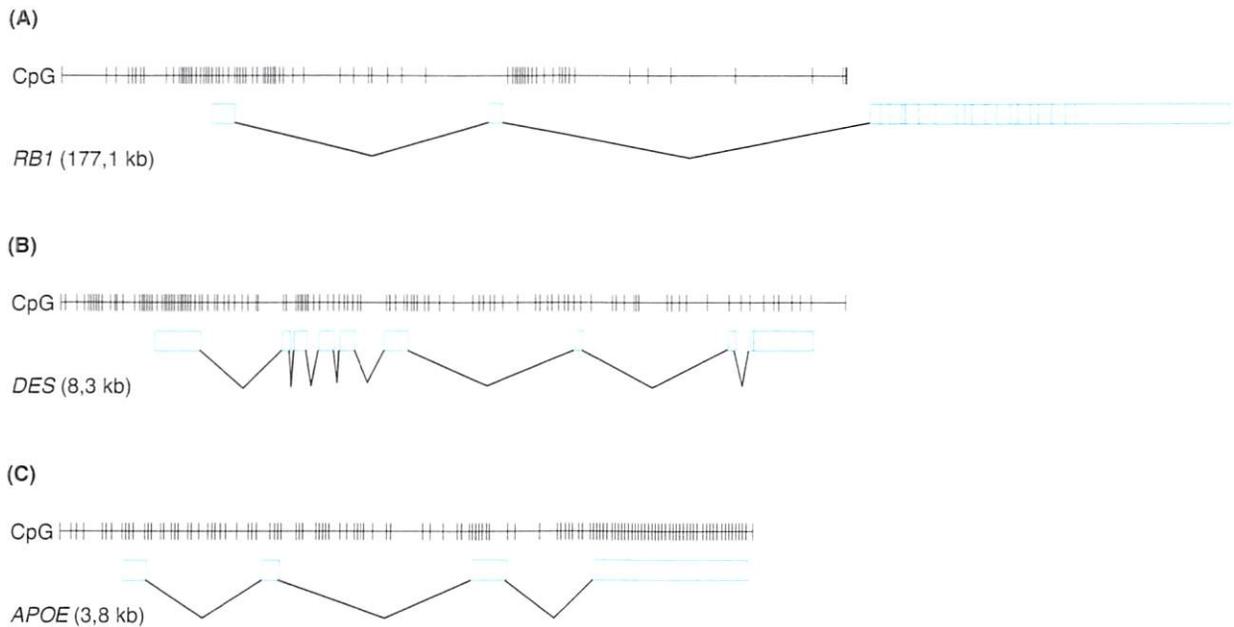
(A) A citosina que ocorre na seqüência 5'-CpG-3' é um alvo para metilação no átomo de carbono 5'. Os produtos desaminados de citosina e seus derivados metilados de 5-metilcitosina são reconhecidos de maneira diferente pelas enzimas que reparam o DNA. (B) O padrão de metilação de CpGs é perpetuado pela necessidade de uma metilase específica em reconhecer uma seqüência-alvo **hemimetilada**. A seqüência CpG possui simetria bilateral. Após a metilação de um alvo hemimetilado (i. e., metilado em apenas uma das fitas), as duas fitas metiladas são separadas na duplicação do DNA e atuam como moldes para a síntese de duas fitas-filhas não-metiladas. Os duplícies filhos resultantes irão, agora, fornecer novos alvos hemimetilados para a continuidade do mesmo padrão de metilação. (C) A desaminação de 5-metilcitosina na seqüência CpG resulta na conversão do dinucleotídeo CpG para os dinucleotídeos TpG e CpA.

## Quadro 8.5

### Ilhas CpG

No DNA de vertebrados, a seqüência CpG é um sinal para metilação por uma DNA-metil-transferase citosina-específica, que adiciona um grupamento metil no carbono 5' da citosina. A 5'-metilcitosina resultante é quimicamente instável e está sujeita à desaminação, resultando em timina (Figura 8.17). Durante longos períodos de tempo evolucionário, o número de dinucleotídeos CpG no DNA de vertebrados tem diminuído gradativamente, devido à lenta, porém constante, conversão de Cpg em TpG (e em CpA na fita oposta). Embora a freqüência geral de CpG no genoma de vertebrados seja baixa (cerca de 20% da freqüência esperada), existem pequenas seqüências de DNA que são caracterizadas por possuírem a freqüência normal de CpG esperada. Tais ilhas de densidade de CpG normal (ilhas CpG) são comparativamente ricas em GC (tipicamente acima de 50% GC) e estendem-se por mais de centenas de nucleotídeos. No caso de genes que apresentam expressão difundida, as ilhas CpG associadas são quase sempre encontradas nas extremidades 5' dos genes, normalmente na região promotora, freqüentemente estendendo-se para o interior do primeiro éxon. No entanto, para genes que mostram padrões de expressão restritos, as ilhas CpG associadas são, comumente, encontradas a alguma distância a jusante do sítio de iniciação da transcrição (ver Jones, 1999 e Figura 8.18).

As ilhas CpG podem ser distinguidas pelas suas suscetibilidades à clivagem com determinadas nucleases de restrição. Por exemplo, a nuclease de restrição *HpaII* reconhece e cliva DNA nas seqüências CCGG, a menos que a citosina central esteja metilada. Como resultado, *HpaII* cliva freqüentemente dentro de ilhas CpG, para dar origem a pequenos fragmentos (por essa razão o nome original das ilhas CpG era de *ilhas HTF*, que significava fragmentos muito pequenos de *HpaII* [do inglês *HpaII tiny fragments*]). Utilizando esse tipo de abordagem, foi estimada a ocorrência de um total de 45 mil ilhas CpG no genoma humano, e aproximadamente mais da metade de todos os genes humanos associados está associada com uma ilha CpG (Antequera e Bird, 1993). Genes associados a ilhas CpG incluem todos os genes de manutenção e genes que são amplamente expressos e talvez cerca de 40% dos genes que são expressos de uma maneira tecido-específica (Larsen *et al.*, 1992). Enquanto a grande maioria de CpGs localizados fora das ilhas CpG é metilada, as ilhas CpG não são metiladas ou apresentam quantidades muito baixas de metilação CpG. As ilhas CpG associadas com sítios de iniciação da transcrição de genes cuja expressão é restrita a certos tecidos não são metiladas, nesses tecidos, mas são metiladas nos tecidos onde os genes não são expressos. No entanto, as ilhas CpG localizadas a jusante do sítio de iniciação da transcrição podem ser metiladas em tecidos onde o gene é expresso (ver Jones, 1999 e texto).



**Figura 8.18** Estrutura das ilhas CpG em três genes humanos.

Barras verticais representam as posições do dinucleotídeo CpG nas seqüências de DNA representando: (A) o gene de desmina (*DES*) humano; (B) a extremidade 5' do gene do retinoblastoma (*RB1*) humano; (C) o gene da apolipoproteína E (*APOE*) humano. Os comprimentos totais dos genes estão indicados entre parênteses após o nome do gene. Boxes representam éxons com padrões de encadeamento mostrados por linhas pretas finas conectadas. No caso do gene *RB1*, somente os primeiros ~10 kb de seqüência são mostrados. Note que ilhas CpG estão freqüentemente localizadas na extremidade 5' de um gene (como em *RB1* e *DES*), mas ocasionalmente podem ser encontradas em outras posições (como no caso de *APOE*).

tilação do genoma (Figura 8.19). Mais tarde ainda, no estágio de pré-gastrulação, ocorre uma ampla metilação *de novo*. No entanto, a extensão dessa metilação varia nas diferentes linhagens celulares:

- a linhagem celular somática é fortemente metilada;
- linhagens derivadas do trofoblasto, as quais dão origem à placenta, saco embrionário, etc., são submetiladas;
- células germinativas primordiais iniciais são poupadas; seus DNAs genômicos permanecem amplamente não-metilados, até após a diferenciação das gônadas e quando as células germinativas se desenvolvem, depois do que ocorre uma ampla metilação *de novo*.

#### 8.4.3 *Imagina-se que a metilação do DNA em animais atue como uma maneira de defesa do hospedeiro contra transposons, bem como uma maneira de perpetuação de padrões de repressão da transcrição*

Embora nem todos os eucariotos pareçam estar sujeitos à metilação do DNA, a sua função nas células animais parece ser extremamente importante, e a mutagênese direcionada ao gene da metil-transferase da citosina em camundongos resultou na letalidade embrionária. A função precisa da metilação do DNA nas células animais, no entanto, permanece obscura. Visões atuais

focalizam, em especial, dois aspectos das células animais: o tamanho do genoma (animais possuem genomas comparativamente maiores, com maior número de genes e também maior quantidade de famílias de DNA altamente repetitivo, pertencentes à classe dos transposons) e o modo de desenvolvimento (especialmente a variação em termos de duração de vida e freqüência de renovação celular). Duas visões um tanto contrastantes, em relação à função principal da metilação do DNA nas células animais, foram o objeto de muita controvérsia: o **modelo de defesa do hospedeiro** e o **modelo de regulação gênica**.

#### **Defesa do hospedeiro como uma função principal para a metilação do DNA**

Como a função de restrição-modificação da metilação do DNA em bactérias (ver Quadro 4.1), o modelo de defesa do hospedeiro prevê que a função principal da metilação do DNA, nas células animais, seja a de conferir uma maneira de proteção ao genoma, mas, nesse caso, controlando a dispersão de transposons (Yoder *et al.*, 1997). Cerca de um terço das seqüências de DNA no genoma humano podem ser classificadas como pertencentes a famílias de (retro)transposons e sabe-se que uma pequena fração dessas seqüências, no genoma humano e em outros genomas, está ativamente transpondo-se (Seção 7.4). Sabe-se que as famílias de transposons no genoma humano e outros genomas são fortemente metiladas (imagina-se que cerca de 90% das 5-metilcitosinas estejam localizadas em famílias de retrotransposons) e, dessa forma, a metilação do DNA tem sido vista como

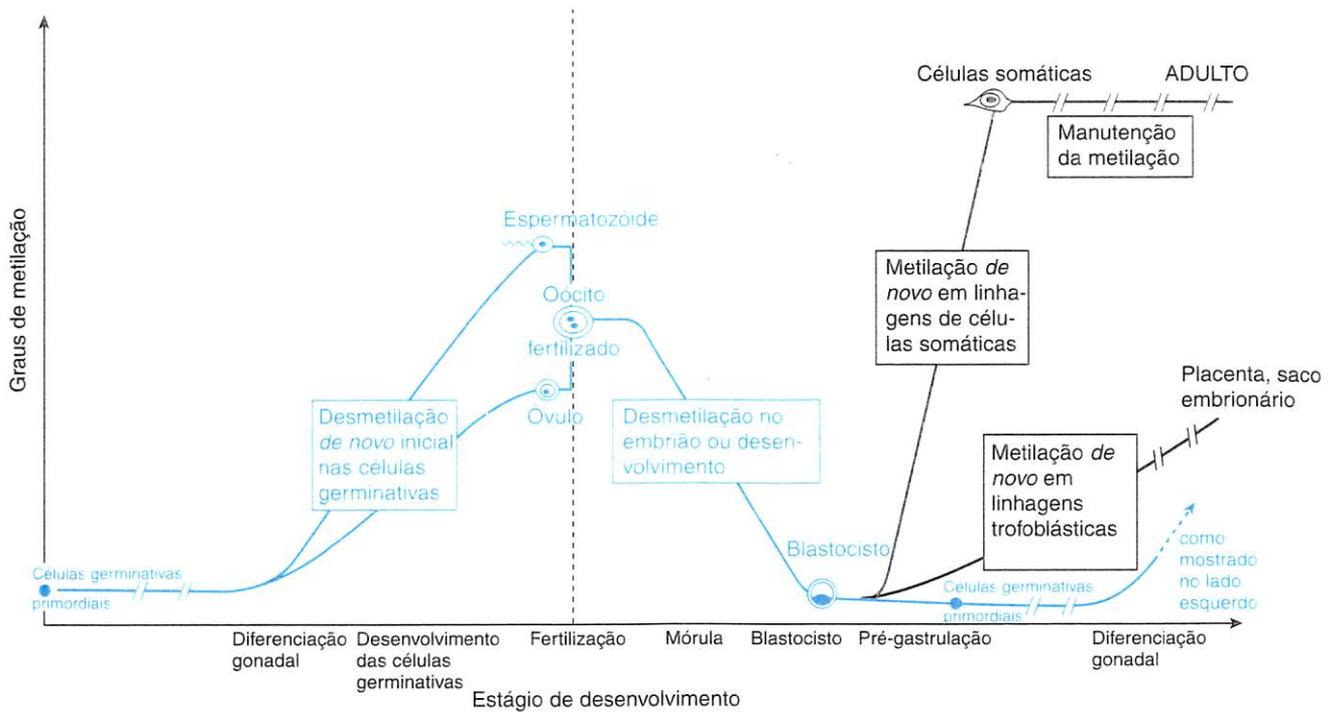


Figura 8.19 Mudanças na metilação do DNA durante o desenvolvimento de mamíferos.

Os estágios do desenvolvimento para gametogênese e desenvolvimento embrionário inicial estão expandidos para clareza; aqueles para o desenvolvimento tardio estão abreviados, como indicado pelos cortes duplos. Note as mudanças muito rápidas na metilação do DNA durante: (i) a *gametogênese* – a metilação *de novo* origina genomas substancialmente metilados no espermatozóide e no óvulo (embora com diferenças no nível total de metilação e no padrão de metilação nesses genomas – ver texto), e em (ii) no *embrião inicial*, quando uma onda de ampla desmetilação genômica ocorre no estágio de pré-implantação (mórula e blástula inicial) e é sucedida, logo após, por uma metilação *de novo*, em ampla escala, começando no estágio de pré-gastrulação. O último é especificamente pronunciado em linhagens somáticas e, em extensão bem menor, em linhagens trofoblásticas, que originam a placenta e o saco embrionário, mas não ocorrem nas células germinativas primordiais (as células do embrião que irão, eventualmente, originar as células do espermatozóide e do óvulo).

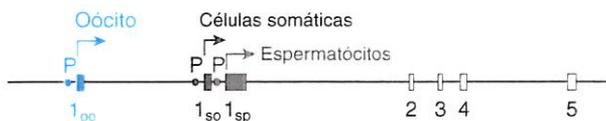


Figura 8.20 Regulação sexo-específica no gene da metil-transferase *Dnmt1*.

O gene *Dnmt1* da metil-transferase parece ser o DNA da metil-transferase de manutenção predominante em células de mamíferos e pode ser o da principal metil-transferase *de novo*, também. Ele é altamente expresso em células germinativas de machos, em oócitos maduros e no embrião inicial. Cinco éxons são utilizados, mas três alternativas para o éxon-1 são empregadas em células diferentes (Mertineit *et al.*, 1998):  $1_{so}$  (éxon 1 em células somáticas) e dois éxons sexo-específicos,  $1_{sp}$  (em espermatócitos) e  $1_{oo}$  (em oócitos). O éxon oócito-específico está associado com a produção de quantidades muito grandes da proteína *Dnmt1* ativa, que é truncada no terminal N e sequestrada no citoplasma, durante os últimos estágios de crescimento. O éxon espermatócito-específico interfere com a tradução e impede a produção de *Dnmt1* durante o estágio de *crossing-over* da meiose de machos.

um mecanismo para reprimir tal transposição, que, descontrolada, poderia ser danosa para as células. No entanto, dados recentemente obtidos a partir de um cordado invertebrado, *Ciona intestinalis*, parecem ser inconsistentes com o modelo de defesa do genoma: múltiplas cópias de um retrotransposon aparentemente ativo e uma grande fração de SINEs altamente repetidas eram predominantemente não-metiladas, enquanto que genes, ao contrário, pareciam ser metilados (Simmen *et al.*, 1999).

### Regulação gênica como a função principal para a metilação do DNA

A metilação do DNA em vertebrados foi vista como um mecanismo para o silenciamento da transcrição e pode constituir uma posição *default*. Sequências de DNA que estão ativamente transcrevendo necessitam estar não-metiladas (ao menos nas regiões promotoras). Enquanto a metilação do DNA em invertebrados pode servir para reprimir transposons e outras famílias de sequências repetidas (ver abaixo), ela pode ter adquirido um papel especial nos vertebrados, como um mecanismo para a regulação da expressão de genes endógenos e redução do ruído

da transcrição (através do silenciamento de uma grande fração de genes cujas atividades não são necessárias em uma célula). Pela redução da expressão gênica desnecessária, a metilação do DNA pode ter permitido o aumento no número de genes e na complexidade que caracteriza os vertebrados (Bird, 1995). O argumento contrário é de que o *status* de metilação das regiões 5' de genes tecido-específicos não pode ser correlacionado com a expressão em diferentes tecidos e que o papel da metilação na expressão gênica é em funções biológicas especializadas resultantes de mecanismos (por exemplo pré-programação, do inglês *imprinting*) que utilizam a expressão gênica alelo-específica (Walsh e Bestor, 1999).

#### Metilação do DNA e expressão gênica

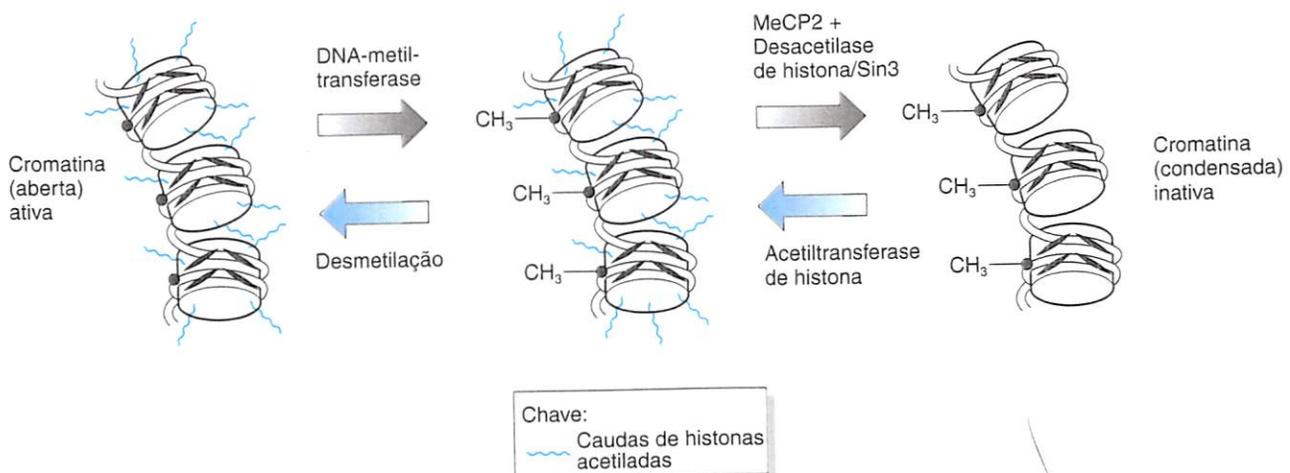
Os DNAs das cromatinas ativa e inativa em transcrição diferem em um número de características, incluindo o grau de compactação e a extensão de sua metilação (Tabela 8.7). Enquanto a metilação das ilhas CpG a jusante de promotores não impede a continuação da transcrição através dessas regiões (Jones, 1999),

não há dúvidas de que regiões promotoras metiladas estão correlacionadas com o silenciamento da transcrição. Além disso, a extensão da acetilação de histonas é um fator importante. As acetiltransferases específicas de histonas adicionam grupamentos acetil aos resíduos de lisina próximos à região N-terminal de histonas. Os terminais N acetilados formam, então, caudas que se projetam a partir do centro do nucleossomo. Imagina-se que, como as histonas acetiladas possuem uma afinidade reduzida pelo DNA, e possivelmente uma pela outra, a cromatina pode ser capaz de adotar uma estrutura mais aberta, que está mais apropriada para a expressão gênica. A desacetilação das histonas, no entanto, promove a repressão da expressão gênica, presumivelmente devido ao fato de a cromatina tornar-se mais condensada.

Recentemente, o processo de metilação do DNA e o de desacetilação de histonas mostraram estar ligados. A repressão de seqüências CpG metiladas, nas regiões promotoras, parece ser mediada por proteínas que se ligam especificamente às seqüências CpG metiladas. Duas dessas proteínas foram identificadas,

**Tabela 8.7** Características associadas com as cromatinas ativa e inativa em transcrição

Característica	Cromatina ativa em transcrição	Cromatina inativa em transcrição
Conformação	Aberta, conformação estendida da cromatina	Conformação altamente condensada; especialmente aparente na heterocromatina (tanto na facultativa como na constitutiva; ver Seção 3.5)
Metilação do DNA	Relativamente não-metilado, especialmente nas regiões promotoras	Metilado, inclusive nas regiões promotoras
Acetilação de histonas	Histonas acetiladas	Histonas desacetiladas



**Figura 8.21** Repressão da transcrição pela desacetilação de histonas pode ser mediada pela metilação do DNA.

Dinucleotídeos CpG são alvos para a metilação e, por sua vez, CpGs metilados são alvos específicos para ligação por proteínas, tal como MeCP2, que atua como um repressor da transcrição e recruta um complexo co-repressor que consiste do fator repressor da transcrição mSin3A e desacetilases de histonas. Estas últimas removem grupamentos acetil das histonas. O processo reverso envolve acetilação seqüencial de histonas e, após, desmetilação do DNA (ver Ng e Bird, 1999).

MeCP1 e MeCP2 (do inglês *methylated CpG binding proteins 1 and 2*, proteínas 1 e 2, que se ligam a seqüências CpG metiladas), esta última mostrou ser essencial para o desenvolvimento embrionário e para funcionar como um repressor da transcrição. A capacidade de MeCP2 reprimir a expressão gênica envolve um complexo de desacetilação de histona (Ng e Bird, 1999). Um modelo possível prevê que um sinal inicial para a repressão da transcrição seja a ligação de MeCP2 a seqüências promotoras CpG metiladas. A proteína MeCP2 ligada é então reconhecida por um complexo que consiste de um repressor da transcrição e uma desacetilase de histona, a qual remove grupamentos acetil dos terminais N das histonas, de forma que a cromatina se torna mais condensada (Figura 8.21).

## 8.5 Controle da expressão gênica de longo alcance e pré-programação (*imprinting*)

### 8.5.1 A estrutura da cromatina pode exercer um controle de longo alcance sobre a expressão gênica

Um assunto predominante na expressão gênica eucariótica, distinguindo-a da expressão gênica bacteriana, é que os genes são *individualmente* transcritos. Promotores e elementos relacionados, a montante, tipicamente controlam a expressão de um único gene, o ponto de iniciação da transcrição localizado na região estando a 1 kb do elemento controlador. Alguns elementos *cis*-atuantes, no entanto, exercem um controle de longo alcance sobre uma região cromossômica muito maior e existem evidências crescentes para a regulação coordenada de agrupamentos de genes. Estudos nos quais os genes foram reposicionados em algum outro local do genoma também têm sugerido que os cromossomos estão organizados em domínios funcionais da expressão gênica (domínios de cromatina). Por exemplo, quando os genes são translocados para novas regiões cromossômicas (tanto como resultado de eventos de quebra cromossômica espontâneos, como por manipulação genética de organismos modelos – Seção 21.3), uma expressão gênica aberrante pode ocorrer freqüentemente, mesmo que o gene inteiro e as seqüências controladoras necessárias nas suas seqüências flanqueadoras imediatas estejam preservados de maneira intacta. Os domínios cromossômicos vizinhos são separados por **elementos de fronteira** (também chamados de **insuladores**), os quais atuam como barreiras para o efeito de reforçadores e de silenciadores remotos (Geyer, 1997).

#### Competição por reforçadores ou silenciadores

Algumas vezes o controle da expressão gênica de longo alcance parece depender da competição por um reforçador, entre genes agrupados. Isso parece ser uma característica da expressão do gene da globina como descrito na Seção 8.5.2.

#### Efeitos de posição induzidos pela heterocromatina

Estudos de rearranjos cromossômicos na *Drosophila* mostraram que a proximidade a centrômeros, telômeros ou blocos de heterocromatina pode suprimir a expressão gênica, presumivelmente pela alteração da estrutura de um grande domínio de cromatina. A distrofia muscular fásio-escápulo-umeral (FSHD), do inglês *fascioscapulohumeral muscular dystrophy*; MIM 158900) é um

possível exemplo de um efeito de posição semelhante, no homem. O gene para essa doença neuromuscular progressiva autossômica dominante localiza-se próximo ao telômero do cromossomo 4q. Quando *Southern blots* do DNA, clivado com *EcoRI*, foram hibridizados com a sonda subtelomérica p13H-11, uma hibridização em um fragmento muito grande, de mais de 30 kb, foi observada. O DNA de pacientes FSHD apresenta, consistentemente, bandas menores, de 14 a 28 kb. Esses pacientes possuem um número reduzido de cópias de uma seqüência repetitiva de 3,2 kb, que é reconhecida pela sonda, e deleções *de novo* de regiões das repetições foram observadas em pacientes FSHD sem história familiar prévia.

Expectativas de que a seqüência de 3,2 kb pudesse conter o gene *FSHD*, no entanto, foram frustradas. Mesmo que ela contenha parte de um homeodomínio (Hewitt *et al.*, 1994), não há evidências de que qualquer parte dela seja transcrita ou expressa. Muito provavelmente o gene *FSHD* esteja localizado proximalmente às repetições de 3,2 kb em tandem, e a deleção o deixe mais próximo do telômero 4q, onde ele é silenciado por um efeito de posição.

#### Outros efeitos de posição

Evidências para efeitos de longo alcance, controlando a expressão gênica sobre grandes domínios cromossômicos, têm surgido a partir de estudos de doenças associadas a quebras cromossômicas em humanos (Kleinjan e van Heyningen, 1998). Exemplos são a aniridia (AN1; MIM 106210), causada por mutações de perda de função do gene *PAX6* situado em 11p13, e a displasia campomélica (CMPD1; MIM 211970), causada por mutações no gene *SOX9* situado em 17q24. Em cada caso, os pacientes afetados são identificados por apresentar quebras cromossômicas claramente causadoras, mas os pontos de quebra podem estar muito distantes (centenas de quilobases) do gene cuja expressão é afetada, e não o interrompem fisicamente. Parece provável que a expressão do gene seja suprimida por um efeito de longo alcance, análogo aos efeitos de posição clássicos descritos acima, refletindo o novo ambiente cromossômico criado pela translocação.

As síndromes de Prader-Willi e Angelman (Seção 16.4.2) juntam efeitos de posição, pré-programação e metilação do DNA. Uma seqüência *cis*-atuante análoga à região controladora do locus da globina foi identificada; ela governa a metilação paterno-específica e a expressão gênica de uma região cromossômica de extensão de megapares de bases na localização 15q11.

#### Inativação do X

A inativação do cromossomo X em mamíferos parece ser iniciada por um único gene, *XIST*, o qual é unicamente expresso no cromossomo X inativado (Seção 8.5.6). Esse efeito não é entendido, mas deve ser mediado por algum tipo de modificação estrutural de longo alcance, na cromatina. Isso ocorre desse modo porque um agente difusível, mediado por *XIST*, não seria capaz de afetar justamente o cromossomo X, no qual o gene *XIST* é expresso.

### 8.5.2 A expressão de genes individuais em agrupamentos de genes pode ser coordenada por uma região controladora de locus comum

Alguns agrupamentos de genes humanos mostram evidências de expressão coordenada de genes individuais no agrupamen-

to. Por exemplo, genes individuais nos agrupamentos da  $\alpha$ -globina,  $\beta$ -globina e os quatro genes *HOX* são ativados em uma seqüência temporal que corresponde exatamente à sua ordem linear no cromossomo. No caso dos genes da globina, existe uma nítida expressão específica do estágio de desenvolvimento: genes diferentes podem ser ativados nos estágios embrionário, fetal ou adulto, para gerar formas levemente diferentes de hemoglobina (troca de hemoglobina; Figura 8.22).

Recentemente, tem-se tornado aparente que a expressão dos genes em cada um dos dois agrupamentos dos genes da globina humana é coordenada por uma região controladora dominante, a região controladora do locus (LCR), do inglês *locus control region*, a qual está localizada a alguma distância a montante do agrupamento de genes (ver Grosveld *et al.*, 1993). Acredita-se que tais LCRs agrupamento-específicos organizam o agrupamento no interior de um domínio de cromatina ativa e atuam como reforçadores da transcrição dos genes da globina. A conformação aberta dos domínios de cromatina ativos em transcrição torna-os mais acessíveis à clivagem pela enzima DNase I. De acordo com essa relação, a LCR da  $\beta$ -globina engloba seqüências curtas nos três principais sítios hipersensíveis à DNase I eritrócito-específicos (HS2, HS3 e HS4) agrupados em uma região de 15 kb, localizada cerca de 50 a 60 kb a montante do gene da  $\beta$ -globina, enquanto a LCR da  $\alpha$ -globina ocorre em um sítio hipersensível à DNase I eritrócito-específico, HS-40, localizado a 60 kb a montante do gene da  $\alpha$ -globina (Figura 8.23). Cada sítio marca a localização do que é efetivamente uma seqüência *reforçadora* (*enhancer*), de cerca de 200 a 300 pb de DNA, que contém curtos elementos de seqüência *cis*-atuantes, incluindo múltiplos elementos de seqüência reconhecidos por fatores de transcrição eritrócito-específicos (Figura 8.6). Sem as respectivas LCRs, a expressão dos genes da globina é insignificante e, no caso da LCR da  $\beta$ -globina, parece que os elementos HS2, HS3 e HS4 interagem uns com os ou-

tros para formar um grande complexo que interage com os genes da globinas, individualmente.

Outros sítios hipersensíveis à DNase I estão localizados nos promotores dos genes da globina, mas mostram especificidade pelo estágio de desenvolvimento. Por exemplo, no fígado fetal, os promotores dos dois genes  $\gamma$  e dos genes  $\beta$  e  $\delta$  são marcados por sítios hipersensíveis à DNase I mas, na medula óssea do adulto, os dois genes  $\gamma$  não são mais ativos em transcrição, e seus promotores não mostram mais sítios hipersensíveis à DNase I. Acredita-se que mudanças específicas do estágio de desenvolvimento, na expressão dos genes da globina, sejam acompanhadas pela competição entre os genes da globina pela interação com a sua respectiva LCR e a ativação estágio-específica de elementos silenciadores gene-específicos. Por exemplo, a transcrição do gene da  $\epsilon$ -globina (*HBE1*) é preferencialmente estimulada pela LCR vizinha, no estágio embrionário. No feto, no entanto, a expressão de  $\epsilon$ -globina é suprimida, acompanhada pela ativação de um silenciador, e a expressão de  $\gamma$ -globina torna-se dominante (Figura 8.23).

Além das LCRs da globina humana, um número de LCRs adicionais diferentes foi identificado (ver Kioussis e Festenstein, 1997). No entanto, o papel da LCR da  $\beta$ -globina humana (o qual contou com a análise de transgênicos humanos) foi desafiado por estudos de direcionamento gênico em camundongos, que mostraram que a LCR da  $\beta$ -globina possui uma função de ajuste, em vez de uma função dominante, e que ela não é necessária para a iniciação da sensibilidade à DNase I e para a expressão dos genes no agrupamento da  $\beta$ -globina no camundongo. Esses achados contraditórios podem, possivelmente, significar que existe uma diferença funcional entre as LCRs humana e murina (ver Grosveld, 1999 para uma revisão).

### 8.5.3 Alguns genes humanos apresentam expressão seletiva de somente um dos dois alelos parentais

Genes ligados ao X nas fêmeas e todos os genes autossômicos são bialélicos, porque ambos os genitores normalmente contribuem com um alelo cada. Nos machos possuidores de um cromossomo X e um cromossomo Y, a grande maioria dos genes ligados ao sexo é monoalélica: a maioria dos muitos genes no X não possui um homólogo funcional no cromossomo Y; e alguns dos poucos genes do cromossomo Y são sabidos serem Y-específicos, por exemplo *SRY*, o principal locus determinante do sexo em machos. Uns poucos genes do cromossomo Y possuem homólogos funcionais no cromossomo X e são, portanto, bialélicos. Em alguns casos de locus X-Y homólogos, ambos os homólogos são normalmente funcionais (Seção 14.3.1 e Figura 14.9).

Estamos acostumados a assumir que os alelos paterno e materno de genes bialélicos são expressos, a menos que uma ou ambas as cópias tenham recebido mutações que afetem a expressão. Nitidamente a expressão pode ser tecido-específica, de forma que em algumas células ambos os alelos parentais são fortemente expressos; em outras, as duas cópias gênicas não são aparentemente expressas. Assim, embora possam existir diferenças celulares tecido-específicas na expressão, a discriminação existente entre a capacidade de os dois alelos parentais serem expressos só é devida a diferenças (mutações) genéticas entre eles. No entanto, conhece-se, em humanos e outros mamíferos, vários genes bialélicos em que a expressão de um alelo parental, paterno ou materno mas não de ambos, é *normalmente* reprimida em algumas células (*exclusão alélica*). Em tais células o gene relevante exibe uma hemizigose funcional: somente uma

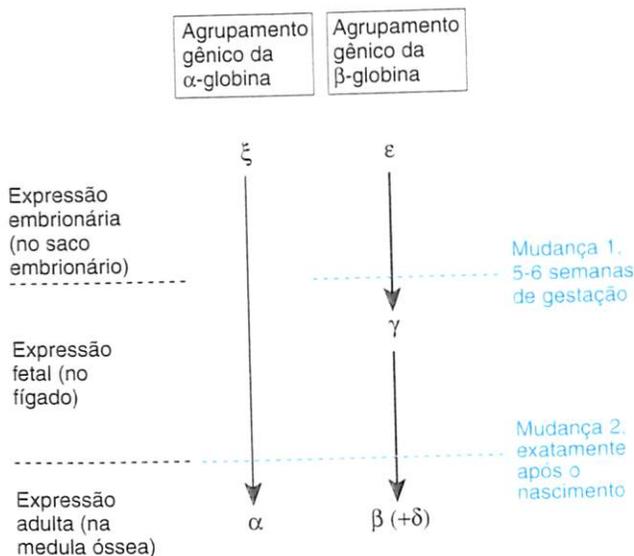
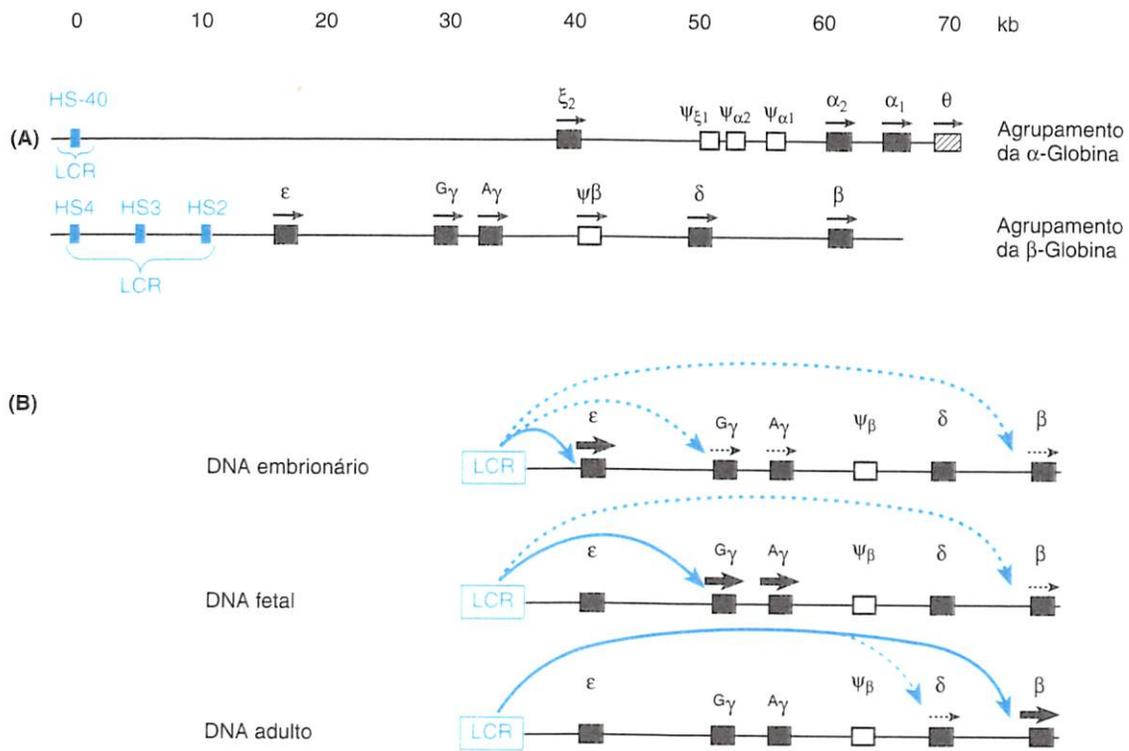


Figura 8.22 Mudanças na hemoglobina humana ocorrem em dois estágios distintos do desenvolvimento.



**Figura 8.23** A expressão gênica nos agrupamentos gênicos da  $\alpha$ - e  $\beta$ -globina é controlada por regiões controladoras de locus comuns.

(A) Organização dos agrupamentos gênicos da  $\alpha$ - e  $\beta$ -globina humanos. As regiões controladoras de locus (LCRs) consistem de um ou mais sítios hipersensíveis à DNase I eritrócito-específicos (HS-40, etc.) localizados a montante do agrupamento. Setas marcam a direção da transcrição dos genes expressos. O status funcional do gene da  $\theta$ -globina é incerto: ele é expresso, mas pode ser um pseudogene expresso (ver Quadro 7.3). (B) Regulação da expressão gênica pela LCR da  $\beta$ -globina. As setas azuis fortes indicam um poderoso efeito reforçador da LCR nos genes indicados, resultando em elevados níveis de expressão; setas pontilhadas azuis indicam os efeitos fracos correspondentes.

### Quadro 8.6

#### Mecanismos que resultam na expressão monoalélica de genes bialélicos em células (de mamíferos) humanas

Mecanismo de expressão monoalélica

Genes relevantes e localização celular

##### **Exclusão alélica de acordo com a origem parental**

Pré-programação genômica

Um pequeno número de genes. As localizações celulares dependem de onde um gene individual é expresso, note, porém, que alguns genes pré-programados apresentam expressão monoalélica em alguns tipos celulares, mas expressão bialélica em outros (Tabela 8.8)

##### **Exclusão alélica, porém independente da origem parental**

Exclusão alélica devido à inativação do cromossomo X

Confinada a certos genes ligados ao X somente em fêmeas; expressão do alelo a partir do cromossomo X não-inativado somente nas células nas quais os genes são expressos

Exclusão alélica seguinte a rearranjos de DNA pré-programados

Expressão do gene da imunoglobulina nos linfócitos B; Expressão do gene do receptor de células T nos linfócitos T

Exclusão alélica por mecanismos desconhecidos

Genes dos receptores olfativos nos neurônios; genes do receptor de células NK; certos genes de interleucinas (IL2, IL4); XIST (em células de embriões iniciais femininos); PAX5 (em células B maduras e progenitoras prematuras)

metade do máximo do produto gênico é normalmente obtida mesmo que as seqüências de ambos os alelos parentais sejam perfeitamente concordantes com a expressão gênica normal ou mesmo idênticas. Em alguns casos, a exclusão alélica pode ser uma propriedade de células ou de tecidos selecionados, enquanto, em outras células do mesmo indivíduo, ambos os alelos podem ser expressos normalmente.

Embora inicialmente considerada uma raridade, a expressão monoalélica de genes bialélicos foi demonstrada para um número crescente de genes, humanos. Uma variedade de mecanismos de expressão pode estar envolvida, e duas amplas classes de mecanismos estão envolvidas (ver Chess, 1998; Ohlsson *et al.*, 1998):

- **Exclusão alélica de acordo com a origem parental (pré-programação).** Em alguns casos, a escolha de qual dos dois exemplares herdados é expressa não é aleatória. Isso significa que, para alguns genes, o alelo cuja expressão é reprimida é sempre o alelo patrilinaramente herdado; em outros ele é sempre o alelo matrilinaramente herdado (Seção 8.5.4).
- **Exclusão alélica independente da origem parental.** Aqui a decisão de qual dos dois alelos é reprimido inicialmente é feita ao acaso, mas, mais tarde, o padrão de exclusão alélica é es-

tavelmente transmitido para as células-filha após a divisão celular. Uma variedade de mecanismos pode estar envolvida (Quadro 8.6). Em alguns casos, uma regulação gênica complexa pode ser necessária. Por exemplo, genes de receptores olfativos são encontrados em grandes agrupamentos ou arranjos nos genes de mamíferos. Em um neurônio olfativo individual, somente um arranjo alélico dos genes do receptor olfativo está ativo (ver Chess, 1998). Uma forma única de controle é a programação de rearranjos de DNA que são necessários para a expressão célula-específica individual dos genes da imunoglobulina nas células B e os genes receptores das células T nos linfócitos T. Devido à complexidade dos últimos mecanismos, eles serão discutidos separadamente na Seção 8.6.

#### 8.5.4 A pré-programação genômica envolve diferenças na expressão de alelos, de acordo com a origem parental

Várias observações em mamíferos sugeriram que os genomas materno e paterno de um indivíduo não são equivalentes (Qua-

### Quadro 8.7

#### A não-equivalência dos genomas materno e paterno

Além das diferenças óbvias entre os cromossomos X/Y, a não-equivalência entre os autossomos e cromossomos X patri e matrilinaramente herdados é sugerida a partir das observações listadas abaixo.

##### Diploidia uniparental induzida experimentalmente em camundongos

O pronúcleo masculino de um oócito de camundongo fertilizado pode ser removido e substituído por um segundo pronúcleo feminino, para gerar um ginogenótipo (algumas vezes chamado um partenogenótipo; todos os 46 cromossomos são de origem materna). Se, ao contrário, o pronúcleo feminino for substituído por um segundo pronúcleo masculino, um androgenótipo é formado. A despeito de possuir cromossomos diplóides normais, tais embriões não conseguem desenvolver-se e morrem antes da metade da gestação. Ginogenótipos mostram deficiências severas em estruturas extra-embriônicas, porém um embrião relativamente normal; em contraste, em androgenótipos o embrião é mais severamente afetado do que as estruturas extra-embriônicas (ver Bestor, 1998 para referências).

##### Diploidia uniparental ocorrendo normalmente em humanos (Seção 2.6.4)

Diploidias uniparentais humanas não são incomuns. Diploidias uniparentais androgenéticas desenvolvem-se como verrugas hidatiformes, que consistem de massas de vilosidades coriônicas hidrofóbicas e outras estruturas placentárias, mas não possuem tecidos embrionários. Diploidias uniparentais ginogenéticas dão origem aos cistos dermóides que se desenvolvem em teratomas ovarianos, consistindo de uma massa de tecidos adultos bem-diferenciados, mas altamente desorganizados, freqüentemente incluindo osso, dente, cartilagem, pele e outros tecidos, mas normalmente não apresentando qualquer estrutura extra-embriônica.

Os indivíduos triplóides representam uma combinação de um genoma diplóide herdado a partir de um progenitor e um genoma haplóide normal do outro. O fenótipo é diferente, dependendo de qual progenitor contribuiu com o genoma diplóide.

##### Dissomia uniparental (Seção 2.6.4)

Algumas diploidias uniparentais possuem um cariótipo 46, XX ou 46, XY normal, mas possuem duas cópias do mesmo cromossomo, herdadas de somente um dos progenitores. Isso pode resultar em fenótipos anormais, que são diferentes conforme a origem parental do cromossomo relevante. Por exemplo, indivíduos 46, XX ou 46, XY que herdaram ambas as cópias do cromossomo 15 de seus pais desenvolvem a síndrome da Angelman; se ambas as cópias do cromossomo 15 são maternalmente herdadas, resulta na síndrome de Prader-Willi (Quadro 16.6).

Mutações subcromossômicas causando fenótipos anormais diferenciais conforme a origem parental

- A deleção de certas regiões cromossômicas produz um fenótipo diferente quando no cromossomo materno ou paterno. O melhor exemplo é a deleção de 15q12, que no cromossomo paterno, produz a síndrome de Prader-Willi e, no cromossomo materno, produz a síndrome de Angelman (Quadro 16.6).
- Certos caracteres humanos são autossômicos dominantes, mas manifestam-se somente quando herdados de um dos progenitores. Em algumas famílias, tumores de massa são herdados como uma característica autossômica dominante, mas expressos apenas em pessoas que herdaram o gene de seu pai. A síndrome de Beckwith-Wiedemann (MIM 130650) é algumas vezes dominante, mas expressa somente por pessoas que herdaram de suas mães. Exemplos de genealogias são mostrados na Figura 3.5.
- A perda de alelos em muitos cânceres (Capítulo 18) envolve, preferencialmente, o alelo paterno.

dro 8.7). Além das diferenças genéticas entre o DNA genômico do espermatozóide e do óvulo, também existem diferenças epigenéticas. A principal diferença está em ambas as causas, a quantidade total de metilação do DNA (o genoma do espermatozóide é mais extensivamente metilado do que o genoma do óvulo) e o padrão de metilação do DNA em classes de seqüências de DNA específicas. Por exemplo, seqüências Line 1 são altamente metiladas nas células do esperma, mas só parcialmente metiladas no óvulo (Razin e Kafri, 1994; Yoder *et al.*, 1997). Em alguns lócus gênicos individuais, também existem diferenças principais entre a extensão da metilação dos alelos paternos e maternos. Por exemplo, o alelo paterno do gene *H19* é fortemente metilado; o alelo materno não é metilado.

Como sugerido pelas observações no Quadro 8.7, diferenças entre os genomas paterno e materno levam a diferenças na expressão entre alelos paternos e maternos. A pré-programação genômica (também chamada de pré-programação gamética ou parental), em mamíferos, descreve a situação onde não existe equivalência na expressão dos alelos em certos lócus gênicos, dependendo da origem parental (Reik e Walter, 1998; Brannan e Bartolomei, 1999; Tilghman, 1999). Em alguns, os tecidos onde o gene é expresso, a expressão ou do alelo patrilinaramente herdado ou do alelo matrilinearmente herdado é consistentemente reprimida, resultando em expressão monoalélica. O mesmo modelo de expressão monoalélica pode ser fielmente transmitido para as células-filha, após a divisão celular. No entanto, como as seqüências nucleotídicas do alelo cuja expressão é reprimida podem ser perfeitamente consistentes com a expressão gênica (e podem mesmo ser idênticas à do alelo expresso), isso é um efeito epigenético, e não genético.

#### Prevalência e evolução da pré-programação

Muitos genes humanos não estão sujeitos à pré-programação, de outra forma não poderíamos ver tantos caracteres mendelia-

nos simples. Pesquisas sistemáticas foram realizadas para identificar regiões cromossômicas pré-programadas em camundongo. Ao contrário de humanos, todos os cromossomos do camundongo são acrocêntricos, e translocações robertsonianas podem permitir que cruzamentos sejam realizados; tais cruzamentos produzem descendentes que possuem ambas as cópias de um determinado cromossomo, derivadas de um único genitor (**dissomia uniparental**, UPD, do inglês *uniparental disomy*, ver Seção 2.6.4). Esses revelaram que a UPD para alguns cromossomos não tem efeito fenotípico; para outros, ela produz fenótipos anormais. Os fenótipos anormais são algumas vezes complementares para origens parentais diferentes, por exemplo o crescimento exagerado é freqüentemente visto na UPD materna e o retardamento no crescimento, na UPD paterna. Para alguns cromossomos a UPD é letal.

Buscas mais profundas em níveis cromossômicos e genéticos mostraram que a pré-programação é uma propriedade de um número limitado de genes individuais ou de pequenas regiões cromossômicas. Atualmente, são conhecidos mais de 30 genes pré-programados em humanos e camundongos (referência eletrônica 1), mas pode-se esperar que a lista cresça. Até o momento, dois principais agrupamentos de genes pré-programados são conhecidos no genoma humano: uma região de 1 Mb em 11p15 (circundando a região de Beckwith-Wiedemann) contém ao menos sete genes pré-programados que podem estar arranjados em dois agrupamentos (Lee *et al.*, 1999); um agrupamento de 2,3 Mb na região 15q11-q13 (circundando o lócus da síndrome de Prader-Willi e Angelman) também contém ao menos sete genes pré-programados (Schweizer *et al.*, 1999; ver Figura 8.24). Os agrupamentos de genes pré-programados contêm exemplos de genes vizinhos com pré-programações parentais diferentes, por exemplo o gene *H19* é expresso somente a partir do cromossomo materno 11, enquanto o gene adjacente *IGF2* é expresso somente a partir do cromossomo paterno.

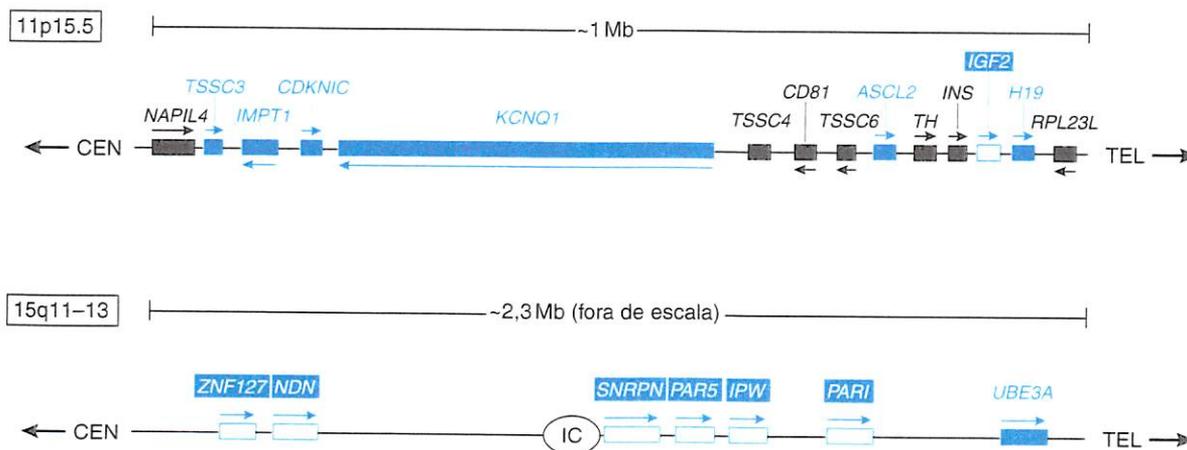


Figura 8.24 Agrupamentos de genes pré-programados em 11p15.5 e 15q11-q13.

Genes desconhecidos quanto à pré-programação são mostrados em preto; genes pré-programados estão em azul. Genes mostrados como boxes azuis cheios, p. ex., *KCNQ1*, *UBE3A*, mostram *repressão preferencial dos alelos maternos* (de forma que, em alguns tecidos, somente o alelo materno é expresso). Genes mostrados em boxes azuis vazios; p. ex., *IGF2*, *ZNF127*, possuem o padrão oposto: *repressão preferencial dos alelos paternos*. Setas indicam a direção da transcrição. IC corresponde a um centro de pré-programação (do inglês *imprinting center*) (ver texto). A região 15q11-q13 foi bem menos estudada, e é provável que outros genes sejam encontrados nela. Note que alguns genes possuem outros nomes na literatura, p. ex., *KCNQ1* (*KvLQT1*), *CDKN1C* (*p57* ou *KIP2*), *CD81* (*TAPA1*). Dados de Lee *et al.* (1999) e Schweizer *et al.* (1999).

A grande maioria dos genes pré-programados conhecida é autossômica. No entanto, o gene *XIST* que possui um papel principal no estabelecimento da inativação do cromossomo X (ver a próxima seção), pode ser considerado um exemplo de um gene pré-programado ligado ao X, uma vez que a expressão do alelo matrilinearmente herdado é preferencialmente reprimida no trofoblasto. Um gene pré-programado ligado ao X que afeta uma função cognitiva também foi sugerido a partir de padrões comportamentais diferenciados na síndrome de Turner. Meninas com a síndrome de Turner não possuem um cromossomo Y, possuem somente um cromossomo X. Se o cromossomo X é herdado da mãe, um comportamento social que tende à desordem é freqüente, mas se ele for herdado do pai, a garota apresenta um comportamento semelhante ao normal para uma garota (Skuse *et al.*, 1997).

A pré-programação ocorre em sementes de plantas, alguns insetos e mamíferos. Nenhum efeito pré-programado importante, como julgado pelo fenótipo, foi observado em organismos-modelo, tais como *Drosophila*, *C. elegans* e *zebrafish* (o peixe ornamental "paulistinha"), embora o potencial para pré-programação possa existir na *Drosophila*. Os mamíferos são incomuns pelo modo como os embriões são totalmente dependentes do fluxo de nutrientes da placenta materna. Como muitos genes pré-programados estão envolvidos na regulação do crescimento fetal, uma explicação enfrenta o conflito entre os genomas parentais: o genoma paterno propaga-se melhor através da criação de um embrião que, agressivamente, remove nutrientes da mãe; o genoma materno suprime isso para proteger a mãe e preservar alguns recursos para gestações futuras. Como visto em casos de diploidia uniparental (Quadro 8.7), genes paternos são preferencialmente expressos no trofoblasto e em membranas extra-embriônicas, enquanto que genes maternos são preferencialmente expressos no embrião.

### 8.5.5 O mecanismo da pré-programação genômica não é claro, mas um componente-chave parece ser a metilação do DNA

Para confirmar a pré-programação de um gene, é necessário identificar um indivíduo que é heterozigoto para uma seqüência variável presente no mRNA maduro. O mRNA de diferentes

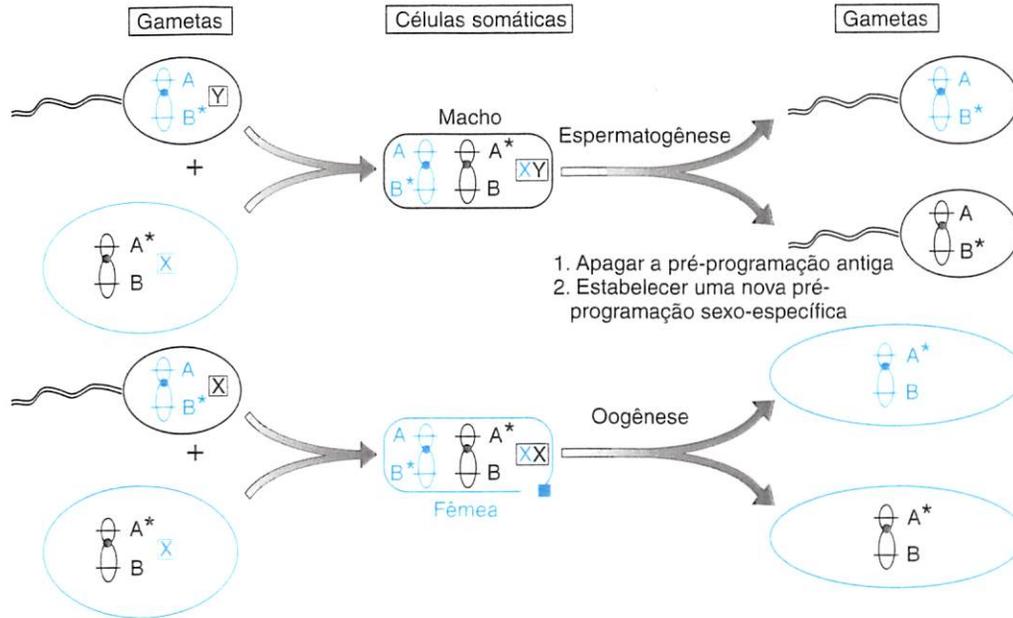
tecidos pode, então, ser testado para a expressão monoalélica ou bialélica, e a origem de cada alelo pode ser determinada pela classificação dos progenitores. Para alguns genes, esse tipo de análise tem demonstrado que a pré-programação está confinada a somente certos tecidos ou a certos estágios do desenvolvimento (Tabela 8.8). Assim, a pré-programação permite um nível extra de controle da expressão gênica, mas não é possível comprimir o seu funcionamento em uma simples história uniforme.

As observações acima sugerem que alguns mecanismos devem ser capazes de distinguir entre alelos patri- e matrilinearmente herdados: à medida que os cromossomos passam através das linhagens germinativas do macho e da fêmea, eles devem adquirir alguma pré-programação para sinalizar uma diferença entre alelos paternos e maternos, no organismo em desenvolvimento. Um componente-chave, ao menos na manutenção do *status* de pré-programação, é a metilação do DNA, alelo-específica (Brannan e Bartolomei, 1999; Tilghman, 1999). A pré-programação de vários genes pré-programados é interrompida em camundongos mutantes que são deficientes no gene da citosina-metiltransferase *Dnmt1*, e todos os genes pré-programados são caracterizados por regiões ricas em CG de metilação diferenciada.

Curiosamente, sabe-se que *Dnmt1* possui éxons sexo-específicos (Seção 8.4.2 e Figura 8.20). Nos oócitos isso resulta em um produto protéico truncado amino-terminal oócito-específico que, de modo concebível, poderia metilar especificamente os alelos maternos de genes, como o receptor do fator de crescimento II semelhante à insulina. O éxon espermatócito-específico de *Dnmt1* interfere com a tradução do mRNA de *Dnmt1*, e é menos claro como padrões de metilação paterno-específicos possam ser adquiridos. Durante o desenvolvimento, a pré-programação poderia ser estavelmente herdada, ao menos para muitos ciclos de duplicação do DNA (porém, ver adiante). Nítidamente, deve também existir um mecanismo para apagar a pré-programação durante a transmissão pela linhagem germinativa, o que é necessário quando, por exemplo, um homem transmite um alelo que ele tenha herdado de sua mãe (Figura 8.25). Novamente, no entanto, alguém poderia sugerir a desmetilação que ocorre no início da embriogênese como uma maneira para obter isso, deixando as células germinativas primordiais essencialmente não-metiladas (Figura 8.19).

**Tabela 8.8** Exemplos de regulação de tecidos e de estágios do desenvolvimento de genes pré-programados em mamíferos

Gene e localização	Alelo reprimido	Diferenças nos padrões de expressão
<i>IGF2</i> (fator de crescimento tipo 2 semelhante à insulina)	Materno	Pré-programado em vários tecidos, mas com expressão bialélica no cérebro, no fígado adulto e nos condrócitos, etc.
<i>PEG1/MEST</i>	Materno	Pré-programado no tecido fetal, mas com expressão bialélica em outros tecidos
<i>UBE3A</i> (ubiquitina proteólise ligase 3)	Paterno	Pré-programado exclusivamente no cérebro, mas com expressão bialélica em outros tecidos
<i>KCNQ1</i> (envolvido no canal de potássio)	Paterno	Pré-programado em vários tecidos, mas com expressão bialélica no coração
<i>WT1</i> (tumor de Wilm)	Paterno	Freqüentemente pré-programado em células da placenta e do cérebro, mas com expressão bialélica no rim



**Figura 8.25** A pré-programação genômica (gamética) necessita apagar a pré-programação na linhagem germinativa.

O diagrama ilustra o destino de um cromossomo transportando dois genes, A e B, que estão sujeitos à pré-programação: A é pré-programado na linhagem germinativa feminina, B é pré-programado na linhagem germinativa masculina, como indicado pelos asteriscos. Como resultado, em células somáticas diplóides, A é pré-programado quando presente em um cromossomo matrilinearmente herdado e B é pré-programado quando presente em um cromossomo patrilinaramente herdado. Um cromossomo individual pode passar por linhagens germinativas masculina e feminina em gerações sucessivas: um homem pode transmitir um cromossomo herdado de sua mãe e uma mulher pode transmitir um cromossomo herdado de seu pai, como indicado pelos gametas no painel esquerdo. Como resultado, deve existir um mecanismo pelo qual a pré-programação antiga seja apagada na linhagem germinativa, antes que uma nova pré-programação sexo-específica seja estabelecida.

O momento da pré-programação tornou-se mais claro recentemente. Na linhagem germinativa feminina, a pré-programação materna, incluindo o padrão materno de metilação, é provavelmente estabelecida durante a maturação do oócito, que é concordante com o achado de que a proteína Dnmt1 não é detectável em oócitos que não estão crescendo, mas é abundantemente produzida em oócitos em crescimento. Na linhagem germinativa masculina, a pré-programação paterna funcional é provavelmente estabelecida antes da meiose, possivelmente no espermatócito primário pós-mitótico (Brannan e Bartolomei, 1999).

Genes pré-programados freqüentemente residem em agrupamentos com genes expressos em cromossomos opostos muitas vezes localizados próximos uns dos outros e freqüentemente contendo genes que parecem codificar um RNA maduro (ver Figura 8.24). Genes adjacentes parecem ser regulados conjuntamente. No caso do agrupamento da síndrome de Prader-Willi/Angelman em 15q11-q13, por exemplo, uma única região adjacente ao gene *SNRPN*, denominada de centro de pré-programação, é a sequência reguladora dominante e parece atuar sobre distâncias comparativamente grandes (ver Figura 8.24, Brannan e Bartolomei, 1999; Tilghman, 1999). No entanto, mecanismos diferentes podem ser encontrados em agrupamentos pré-programados diferentes, ou mesmo dentro de um único agrupamento. Por exemplo, os genes de camundongo *H19*, *Igf2* e *Ins2* são regulados conjuntamente, dividindo dois reforçadores endodérmico-específicos que estão localizados a 3' de *H19*. Porém outros genes no agrupamento não estão sujeitos a esse

controle, sugerindo que múltiplos mecanismos de controle podem ocorrer em um agrupamento gênico pré-programado. Diferentes mecanismos de pré-programação foram considerados em relação à regulação de *H19/Igf2*, incluindo competição por reforçador, mas para explicar uma variedade de achados contraditórios, um centro de pré-programação adjacente a *H19* foi considerado funcionar como um elemento de fronteira (insulador) da cromatina (Tilghman, 1999).

#### 8.5.6 Inativação do cromossomo X em mamíferos envolve a repressão da expressão gênica cis-atuante de alcance muito longo

##### Natureza da inativação do cromossomo X

A inativação do cromossomo X é um processo que ocorre em todos os mamíferos, resultando na inativação seletiva de alelos em um dos dois cromossomos X, nas fêmeas (Migeon, 1994; Lyon, 1999). Ela fornece um mecanismo de compensação de dose que supera diferenças sexuais na razão esperada da dosagem gênica autossômica para a dosagem gênica do cromossomo X (que é 2:1 em machos mas 1:1 em fêmeas). Machos com um único cromossomo X são constitucionalmente hemizigóticos para os genes do cromossomo X, mas as fêmeas tornam-se funcionalmente hemizigóticas pela inativação de um dos alelos cromossômicos X parentais (ver também Seção 2.2.3). Nem todos os genes do cromossomo X estão sujeitos à inativação; os genes que escapam à inativação do X incluem aqueles em que existe um

homólogo funcional no cromossomo Y e alguns em que a compensação de dose não parece ser importante (ver Figura 14.12 para exemplos de genes que escapam da inativação do X).

Em raros indivíduos com um número anormal de cromossomos X (45, X; 47, XXX; 47, XXY, etc.), um único cromossomo X permanece ativo, não importando quantos estejam presentes. Ao contrário, em indivíduos triploídes, um ou dois cromossomos X permanecem ativos e, em tetraploídes, dois cromossomos X permanecem ativos. Assim, deve existir algum tipo de mecanismo de contagem, para garantir que um cromossomo X permaneça ativo para cada dois conjuntos de autossomos.

Em mamíferos, ambos os cromossomos X são ativos no embrião feminino inicial. A inativação do X ocorre em um estágio inicial do desenvolvimento, sendo iniciada no estágio tardio da blástula em camundongos e muito provavelmente também em humanos. Em cada célula que irá originar o feto feminino, um dos dois cromossomos X parentais é aleatoriamente inativado (note que as células do trofoblasto são uma exceção; o cromossomo X paterno é preferencialmente inativado, o que é um exemplo clássico de pré-programação tecido-restrita). Após o cromossomo X paterno ou materno ter sido inativado em uma célula, o mesmo cromossomo normalmente permanece inativo em toda a progênie de células, isto é, o padrão de inativação do cromossomo X é herdado de forma clonal (Figura 2.6). Isso significa que as fêmeas de mamíferos são mosaicos, compreendendo misturas de linhagens celulares nas quais o X paterno é inativado e linhagens celulares onde o X materno é inativado. Além da inativação do cromossomo X em células somáticas femininas, sabe-se que o cromossomo X é transitoriamente inativado durante a gametogênese, em machos e fêmeas.

#### Mecanismos de inativação do cromossomo X

O processo de inativação do cromossomo X é complexo, e mecanismos moleculares distintos estão envolvidos na iniciação e na manutenção da inativação. O centro de inativação do X (*Xic*), que em humanos está localizado em Xq13, controla a iniciação e a propagação da inativação do X. Nesse centro, o gene *XIST* (chamado *Xist* em roedores) codifica um produto de RNA maduro de 15 kb que é unicamente codificado pelo cromossomo X inativo. *XIST/Xist* é portanto um outro exemplo de um gene que está sujeito à expressão monoalélica. Nas células do embrião inicial, a decisão a respeito de qual cromossomo X será inativado é feita aleatoriamente e, assim, a exclusão alélica que *XIST/Xist* apresenta nessas células é independente da origem parental.

*XIST/Xist* é essencial para o funcionamento de *Xic* na iniciação da inativação do cromossomo X, mas não é necessário para a manutenção da inativação do cromossomo X. De alguma forma, uma propagação *cis*-limitada desse produto de RNA atua de forma a cobrir o cromossomo X inativado por distâncias muito longas. Em roedores, a cobertura do RNA *Xist* origina, no cromossomo X inativado, um padrão de bandejamento, sugerindo uma associação preferencial com regiões Giemsa menos ricas em genes. No entanto, o mecanismo que assegura a inativação dos genes no cromossomo X inativo, mas não no X ativo, é desconhecido (ver Duthie *et al.*, 1999 para possíveis modelos).

Embora o gene *Xist* seja essencial para a função de *Xic*, *Xist* não é suficiente sozinho. O elemento controlador de X (*Xce*) afeta a escolha de qual cromossomo X permanece ativo e é diferente de *Xist*, estando localizado a 3' desse. Além disso, a deleção de uma região de 65 kb a 3' de *Xist* produz um efeito que sugere que elementos envolvidos no mecanismo de contagem situam-

se distantes a 3' de *Xist*. Recentemente, um outro gene foi identificado como sendo transcrito a partir da fita oposta àquela que é utilizada para transcrever o gene *Xist*. Uma vez que a unidade de transcrição do novo gene sobrepõe-se completamente ao gene *Xist* e está em orientação reversa, ela foi denominada *Tsix* (Lee *et al.*, 1999). Isso dá origem à idéia de que *Xist* possa ser regulado pelo gene *Tsix* (ver Heard *et al.*, 1999 para possíveis modelos de regulação de *Tsix*).

## 8.6 A organização única e a expressão dos genes Ig e TCR

A organização e a expressão dos genes da imunoglobulina (Ig) e do receptor de células T (TCR, do inglês *T-cell receptor*) é, em muitas maneiras, um tanto diferente da de outros genes. Isso é devido à necessidade de cada indivíduo produzir uma enorme variedade de Igs e TCRs. Um linfócito B ou T individual é *monoespecífico* e produz um único tipo de Ig ou TCR; é a população de diferentes células B ou T, em qualquer indivíduo, que possibilita a síntese de tantos tipos diferentes dessas moléculas. Linfócitos B e T necessitam ser extremamente diversos, porque eles representam as células que fornecem anticorpos de resposta ou respostas mediadas por células a antígenos estranhos: por fornecer um vasto repertório de Igs e TCRs, as possibilidades de serem capazes de reconhecer e ligar-se a muitos tipos diferentes de antígenos estranhos são enormemente aumentadas.

### 8.6.1 Genes de Ig e TCR exibem uma organização única; múltiplos segmentos gênicos podem codificar cada uma das várias regiões diferentes do polipeptídeo

#### Estrutura do polipeptídeo

Uma molécula Ig consiste de quatro cadeias polipeptídicas, duas cadeias pesadas idênticas e duas cadeias leves idênticas (Figura 7.10). As cadeias leves classificam-se em duas categorias: as cadeias leves kappa ( $\kappa$ ) e lambda ( $\lambda$ ), as quais são funcionalmente equivalentes. Os segmentos N-terminais de cada tipo de cadeia são chamados de regiões variáveis (V), as quais são necessárias para a ligação com o antígeno estranho; os segmentos C-terminais remanescentes são regiões constantes (C). No caso das cadeias pesadas, existem alternativas diferentes para a região constante que especificam os tecidos nos quais a Ig será expressa e ditam a classe da imunoglobulina (Tabela 8.9). Semelhantemente, os TCRs, que fornecem respostas imunes mediadas por células a antígenos estranhos, consistem de dois tipos de cadeias. Cada uma de tais cadeias possui regiões variáveis semelhantes à Ig que ligam antígenos estranhos e regiões constantes que ancoram a molécula na superfície da célula (Figura 7.10). Os TCRs que ocorrem com maior frequência possuem uma cadeia  $\beta$  e uma  $\gamma$ ; uma população menor consiste de uma cadeia  $\alpha$  e uma cadeia  $\delta$ .

#### Estrutura do gene

Os genes que codificam os diferentes tipos de cadeias em Igs e TCRs estão localizados em cromossomos diferentes e organizados como agrupamentos de numerosos segmentos gênicos (Tabela 8.10). Cada agrupamento é incomum, pelo fato de que seqüências codificadoras para segmentos específicos da cada cadeia estão freqüentemente presentes em numerosas cópias dife-

Tabela 8.9 Classes e subclasses de Ig

Classes (e subclasses)	Tipo de cadeia pesada	Localização
IgA (IgA1, IgA2)	$\alpha$ ( $\alpha_1, \alpha_2$ )	Ig predominante em secreções soro-mucosas, p. ex., saliva, leite, etc.
IgD	$\delta$	Baixa em soros, mas presente em grandes quantidades na superfície de muitas células B circulantes
IgE	$\epsilon$	Especialmente na superfície de membrana de basófilos e de mastócitos
IgG (IgG1, IgG2, IgG3, IgG4)	$\gamma$ ( $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ )	Principal Ig do soro
IgM	$\mu$	Anticorpo "inicial" predominante

rentes que são repetidas sequencialmente. Por exemplo, embora a região constante da cadeia leve  $\kappa$  humana seja codificada por uma única seqüência  $C\kappa$ , as regiões variáveis são codificadas por uma combinação de um segmento  $V\kappa$  (que codifica a maior parte da região variável) e um curto segmento  $J\kappa$  (segmento de ligação, do inglês joining segment; que codifica uma pequena parte da extremidade C-terminal da região variável), os quais são selecionados a partir de um total de cerca de 76 segmentos  $V\kappa$  e cinco segmentos  $J\kappa$  alternativos. Embora a cadeia leve  $\lambda$  seja semelhantemente codificada por segmentos  $V\lambda$ ,  $J\lambda$  e  $C\lambda$ , o locus Ig da cadeia pesada apresenta algumas diferenças. A região variável é codificada por uma combinação de um segmento gênico  $V_H$ , um segmento gênico  $J_H$  e também um segmento gênico  $D_H$  (codificando um segmento de diversidade), cada um selecionado a partir de muitos segmentos gênicos repetidos. Adicionalmente, existe uma variedade de seqüências  $C_H$  diferentes que especificam a classe da Ig (ver acima). No total, esse agrupamento compreende cerca de 140 segmentos gênicos, dos quais cerca de um terço é incapaz de expressão e abrange cerca de 1.200 kb (Figura 8.26).

Como cada agrupamento gênico de Ig ou de TCR em um linfócito B ou T individual sempre dá origem a, no máximo, um polipeptídeo Ig ou TCR, apesar de incomum, o agrupamento inteiro pode ser funcionalmente considerado como um único tipo de gene. Entretanto, segmentos gênicos individuais não podem ser considerados como os equivalentes funcionais dos *exons clássicos*. Isso é porque os segmentos gênicos individuais nesses agrupamentos são algumas vezes compostos de seqüências codificadoras e não-codificadoras de DNA e podem consis-

tir de vários éxons. Por exemplo, cada uma das seqüências  $C_H$  humanas é composta por três ou quatro éxons clássicos separados por íntrons: após a transcrição em RNA, as seqüências intrônicas são descartadas, e somente as seqüências exônicas são mantidas no mRNA.

### 8.6.2 Rearranjos programados de DNA, nos locos de Ig e TCR, ocorrem durante a maturação dos linfócitos B e T, respectivamente

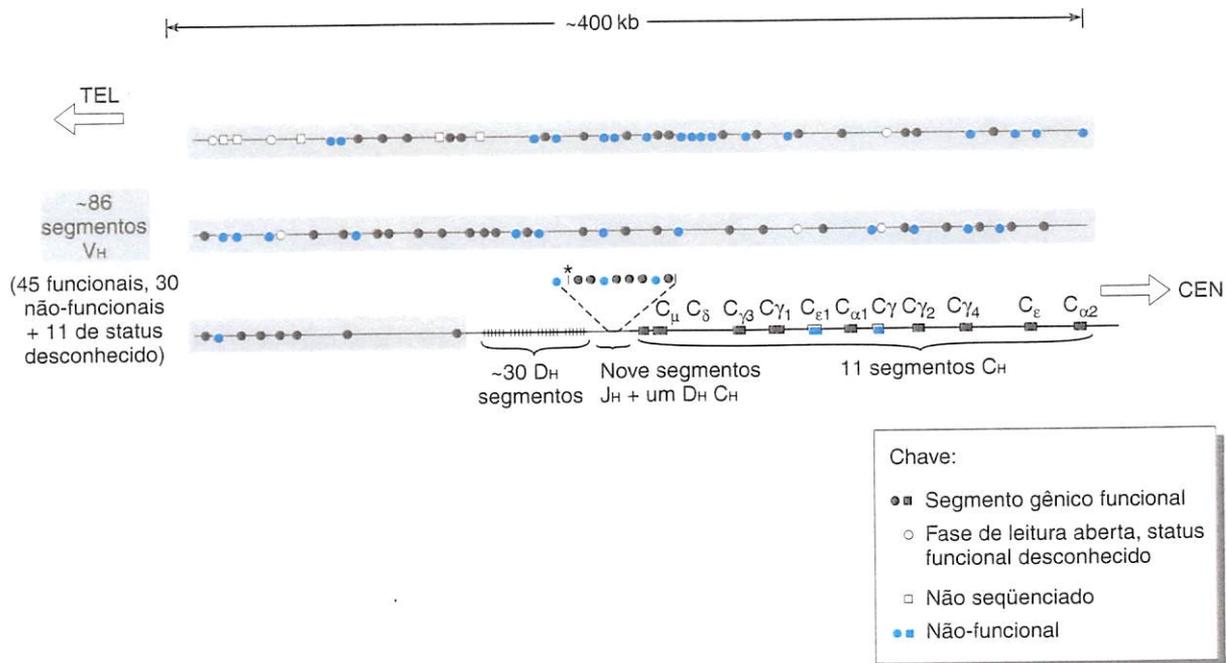
Os rearranjos exclusivos dos segmentos gênicos nos agrupamentos de genes Ig e TCR refletem a maneira muito incomum pela qual recombinações somáticas são necessárias nos linfócitos B e T, antes que os genes Ig e TCR funcionais possam ser montados e, então, expressos (ver adiante). Tais recombinações somáticas resultam na união de combinações diversas de segmentos gênicos diferentes, em linfócitos individuais diferentes. Como consequência, elas podem ser consideradas como eventos tanto *tecido-específicos* (limitados aos linfócitos B e T) como *célula-específicos* que envolvem *encadeamento alternativo de DNA* (como o oposto ao encadeamento alternativo de RNA, que produz combinações diferentes de éxons em nível de RNA – Seção 8.3.2). Como resultado, a organização gênica da linhagem germinativa original é alterada: segmentos gênicos que estavam distantes na linhagem germinativa são encadeados juntos, em nível de DNA. Uma vez que a escolha de qual dos muitos segmentos gênicos repetidos serão recombinados para originar uma unidade V-J ou V-D-J funcional é *célula-específica*, as células B e T individuais produzem Igs e TCRs diferentes. Isso significa que, nesse sentido, cada indivíduo é um mosaico em relação à organização dos genes de Ig e TCR nos linfócitos B e T; mesmo gêmeos idênticos irão divergir geneticamente.

Os rearranjos que irão conduzir à produção de cadeias leves e cadeias pesadas funcionais das Igs são levemente diferentes.

- **Criando uma cadeia leve.** A fim de gerar uma cadeia leve  $\kappa$  de Ig funcional, por exemplo, um evento de recombinação somática junta uma combinação específica de um dos segmentos gênicos  $V\kappa$  e um dos segmentos gênicos  $J\kappa$  (união V-J). Após isso, o encadeamento de uma única seqüência  $C\kappa$  ocorre em nível de RNA (Figura 8.27A).
- **Criando uma cadeia pesada.** Duas recombinações somáticas sucessivas são necessárias, resultando primeiramente na união  $D_H-J_H$  e, depois, na união  $V_H-D_H-J_H$ . Subseqüente-

Tabela 8.10 Locus Ig e TCR funcionais humanos

Locus	Localização	Número de segmentos gênicos			
		V	D	J	C
IGH	14q32.3	86	30	9	11
IGK	2p12	76	0	5	1
IGL	22q11	52	0	7	7
TCRA	14q11-12	60	0	75	1
TCRB	7q32-33	70-100	2	13	2
TCRG	7p15	8	0	5	2
TCRD	14q11-12	6	3	3	1



**Figura 8.26** O locus da cadeia pesada de Ig no cromossomo 14q32 contém cerca de 86 seqüências de regiões variáveis (V), 30 de segmentos de diversidade (D), nove segmentos de união (J, do inglês *joining*) e 11 seqüências de regiões constantes (C).

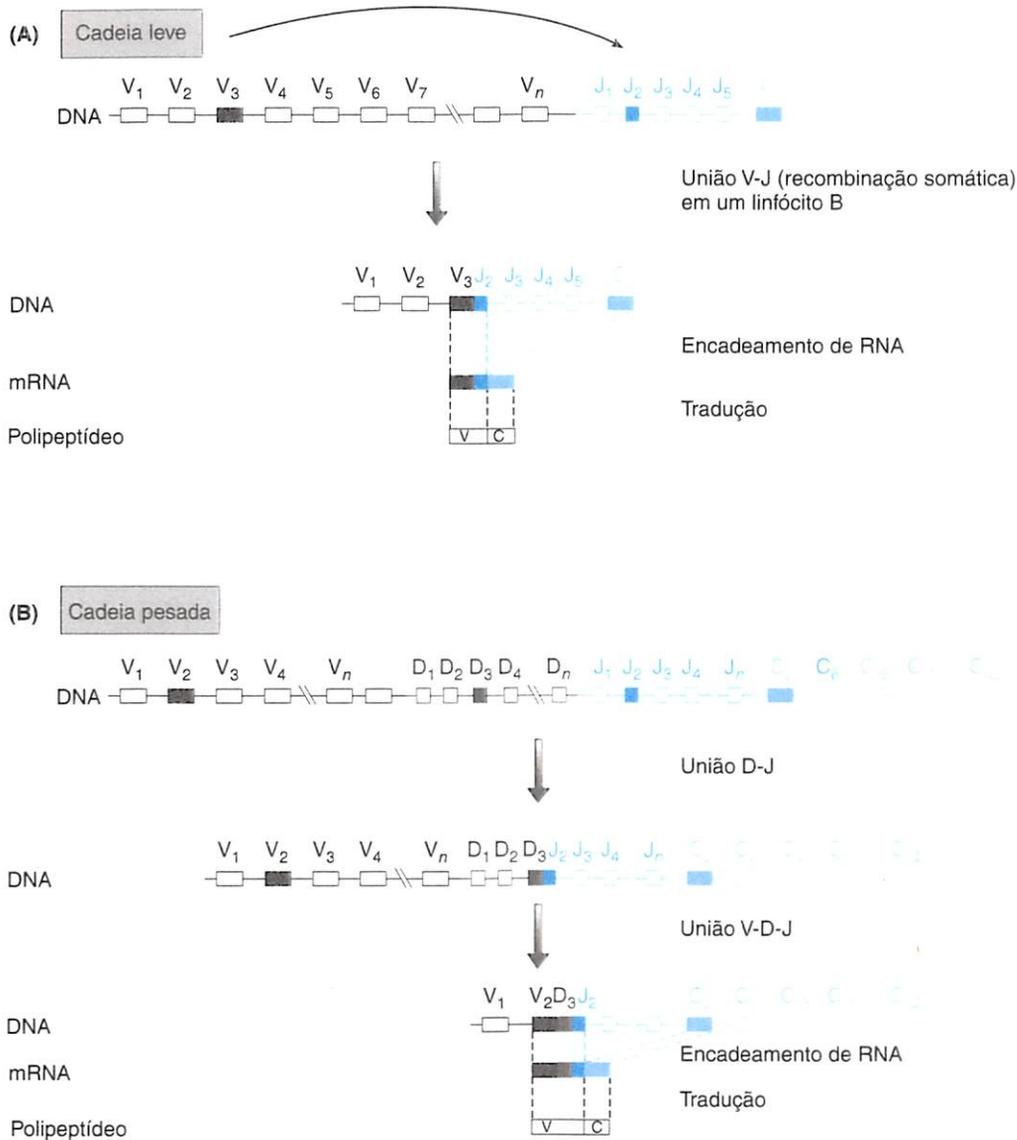
O locus inteiro compreende cerca de 1.200 kb de 14q32 e, por simplificação, é mostrado como três segmentos de 400 kb da extremidade telomérica (parte superior) para a extremidade centromérica (parte inferior). Embora os segmentos  $D_H$  estejam principalmente localizados em poucos agrupamentos, separando os segmentos  $V_H$  e  $J_H$ , ao menos um de tais segmentos está localizado na região do segmento  $J_H$ . Segmentos indicados por círculos abertos possuem as fases de leitura aberta necessárias, mas não foram observados em rearranjos produtivos e, por isso, seus *status* funcionais são desconhecidos. Segmentos, que se sabe serem não-funcionais estão indicados em azul, e representam aproximadamente um terço de todos os segmentos. Note que embora esse seja o único locus da cadeia pesada humana, agrupamentos pequenos de segmentos  $V_H$  e D estão também localizados em 15q11.2 e 16p11.2. Dados adaptados de Cook *et al.* (1994) *Nature Genet.*, 7, pp. 162-168, com permissão de Nature America Inc.

mente, a seqüência codificadora  $V_H$ - $D_H$ - $J_H$  resultante é encadeada, em nível de RNA, com a seqüência  $C_H$  mais próxima, inicialmente  $C_\mu$  (Figura 8.27B).

Porque existem três tipos de locus gênicos Ig funcionais nas células humanas (cadeia pesada, cadeia leve  $\kappa$  e cadeia leve  $\lambda$ ) e porque eles ocorrem em ambos os homólogos materno e paterno, existem seis segmentos cromossômicos nos quais rearranjos de DNA podem resultar na produção de uma cadeia de Ig. No entanto, uma célula B individual é *monoespecífica*: ela produz somente um tipo de molécula de Ig com um único tipo de cadeia pesada e um único tipo de cadeia leve. Isso ocorre por duas razões:

- **Exclusão alélica.** Uma cadeia leve ou uma cadeia pesada pode ser sintetizada a partir do cromossomo materno ou do cromossomo paterno em qualquer célula B, mas não a partir de ambos os homólogos parentais. Como resultado, nas células B existe a expressão monoalélica do locus gênico da cadeia pesada. Esse fenômeno também se aplica aos agrupamentos dos genes de TCR.

- **Exclusão da cadeia leve.** Uma cadeia leve sintetizada em uma única célula B pode ser uma cadeia  $\kappa$ , ou uma cadeia  $\lambda$ , mas nunca ambas. Como resultado dessa necessidade, juntamente com a da exclusão alélica, existe expressão monoalélica em um dos dois agrupamentos de genes de cadeia leve funcionais e *nenhuma expressão* no outro. A decisão de escolher um dos dois alelos da cadeia pesada e um dos quatro possíveis segmentos para criar uma cadeia leve parece ser aleatória. Muito provavelmente, em cada precursor de célula B, são testados rearranjos de DNA eficazes com todos os seis alelos Ig, mas as chances de rearranjos eficazes em mais do que um agrupamento de cadeia leve ou mais do que um de agrupamento de cadeia pesada podem não ser elevadas. Adicionalmente, no entanto, parece existir algum tipo de regulação de retroalimentação negativa: um rearranjo funcional em um alelo de cadeia pesada suprime a ocorrência de rearranjos em qualquer outro alelo, e um rearranjo funcional em qualquer uma das quatro regiões capazes de codificar uma cadeia leve suprime a ocorrência de rearranjos nas outras três.



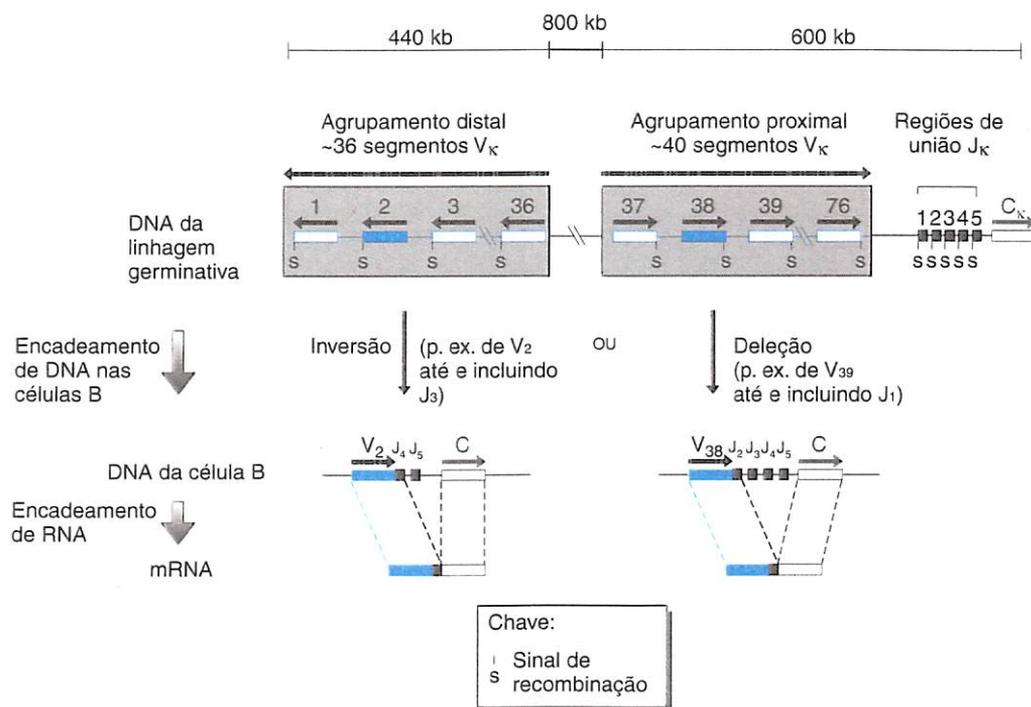
**Figura 8.27** Igs são sintetizadas após a recombinação somática dos segmentos V e J, ou V, D e J, e subsequente encadeamento de RNA para as seqüências C.

**(A)** Síntese da cadeia leve. A recombinação somática (*encadeamento de DNA*) resulta na união de um segmento variável (V) específico a um segmento de união (J) específico; o exemplo mostra uma união  $V_3$ - $J_2$ , a qual é somente uma de muitas possibilidades. A unidade VJ é, então, encadeada com a seqüência da região constante (C) pelo *encadeamento de RNA*. **(B)** Síntese da cadeia pesada. Duas recombinações somáticas seqüenciais produzem a primeira união D-J e depois a unidade VDJ. O encadeamento de RNA subsequente resulta no encadeamento da seqüência VDJ à seqüência  $C_\mu$ . Como as células B maduras, no entanto, recombinações somáticas subsequentes resultam na união da unidade VDJ previamente selecionada a diferentes genes C (*mudança de cadeia pesada*, ver texto e Figura 8.29).

### 8.6.3 Uniões V-J e V-D-J são freqüentemente obtidas por deleções intracromátide e também por inversões com megabases de extensão no caso original

O mecanismo genético que origina as uniões V-J e V-D-J freqüentemente envolve deleções de larga escala, as quais sabe-se ocorrerem por meio de um evento de recombinação intracromátide, similar àqueles utilizados nas uniões V-D-J-C (ver próxima seção). Além disso, a união V-J comumente ocorre como resulta-

do de inversões com megabases de extensão. O lócus gênico da cadeia leve  $\kappa$  humana abrange cerca de 1.840 kb em 2p12 e inclui cerca de 76 segmentos  $V_\kappa$ , a maioria compreendendo pares de segmentos gênicos V duplicados, organizados como dois agrupamentos: um *agrupamento proximal* localizado adjacente aos segmentos  $J_\kappa$  e ao segmento  $C_\kappa$  e um *agrupamento distal*. Isso ocorre como resultado de uma estrutura repetida invertida: segmentos gênicos V no agrupamento  $V_\kappa$  proximal normalmente possuem uma duplicata correspondente em um agrupamento  $V_\kappa$  distal, que está separado do agrupamento proximal por cerca



**Figura 8.28** Inversão ou deleção resultam em encadeamento V-J para produzir Ig funcional de genes de cadeia leve kappa.

O agrupamento gênico da cadeia leve kappa humano contém cerca de 76 segmentos V kappa arranjados em dois grandes agrupamentos em orientações opostas. Segmentos V no agrupamento distal possuem orientação oposta aos segmentos J kappa e à sequência C kappa única. Como resultado, os rearranjos de DNA utilizados para encadear segmentos V kappa distais a segmentos J kappa são inversões de megabases de extensão (Weichhold *et al.*, 1990). Aqueles no agrupamento proximal podem sofrer união V-J por meio de uma recombinação somática, resultando em uma deleção do segmento cromossômico interveniente, muito provavelmente por um evento de recombinação intracromátide, tal como aquele utilizado na mudança de classe (Figura 8.29).

de 800 kb, e em orientação oposta (Figura 8.28). Dependendo de qual agrupamento de segmento V está envolvido, a união V-J ocorre através de duas possíveis rotas:

- Segmentos V no agrupamento distal são unidos aos segmentos J por inversões.
- Segmentos V no agrupamento proximal são unidos aos segmentos J por deleções (Figura 8.28).

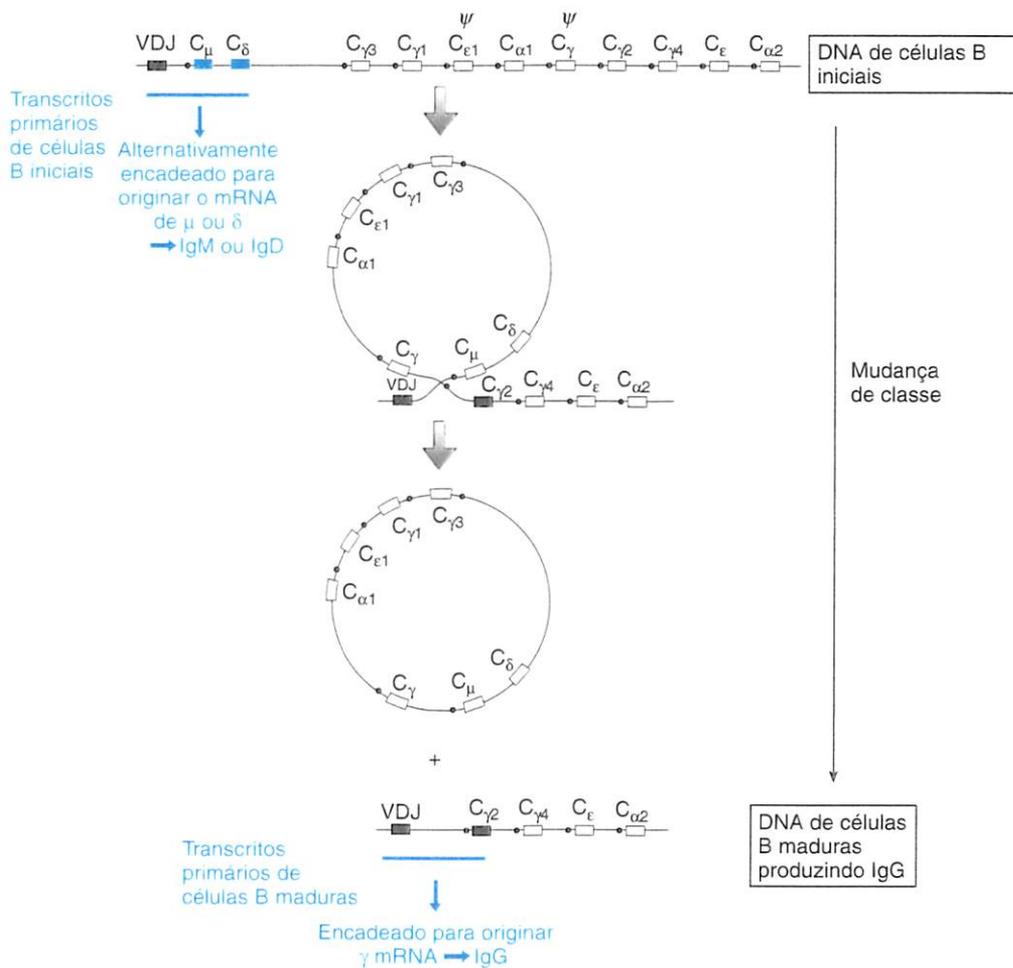
Note que o processo de união é impreciso e, dessa forma, pode também introduzir uma medida de variabilidade na sequência das junções dos segmentos unidos.

#### 8.6.4 A mudança de classe das cadeias pesadas envolve a união diferencial de uma única unidade VDJ a segmentos alternativos de DNA codificadores de regiões constantes

Embora uma célula B produza somente um único tipo de molécula de Ig, a classe da cadeia pesada (Tabela 8.9) pode mudar durante a linhagem celular (mudança de classe ou mudança de isotipo). Tais mudanças envolvem a união diferencial de uma mesma unidade VDJ, que foi unida por duas recombinações so-

máticas sucessivas (Figura 8.27B), a segmentos diferentes codificadores de regiões constantes alternativas. A união inicial de uma sequência VDJ a segmentos de regiões constantes é consumida ao nível do RNA. No entanto, subsequentemente, as mudanças de classe envolvem a união da mesma unidade VDJ, ao nível do DNA, a regiões constantes alternativas, por meio de ainda mais eventos de recombinação somáticos (união V-D-J-C). A mudança de classe envolve a seguinte progressão:

- *Síntese inicial de IgM somente por células B imaturas.* Isso ocorre porque a unidade VDJ é encadeada, ao nível do RNA, à sequência C mu (Figura 8.27B).
- *Síntese tardia de ambas, IgM e IgD, por células B imaturas.* A mudança parcial para criar IgD ocorre porque a unidade VDJ pode ser encadeada, ao nível do RNA, à sequência C delta, como resultado de encadeamento alternativo de RNA (Figura 8.29).
- *Síntese de IgG, IgE ou IgA por células B maduras.* Eventos de mudança de classe envolvem o encadeamento da mesma unidade VDJ às sequências C gamma, C epsilon ou C alpha, respectivamente, ao nível do DNA, como resultado de um evento de recombinação somático (união VDJ-C). O mecanismo envolve deleção da sequência interveniente, pela recombinação intracromátide (Figura 8.29).



**Figura 8.29** A mudança de classe da cadeia pesada de Ig é mediada por recombinação intracromátide.

Note que a união da mesma unidade VDJ à seqüência C<sub>μ</sub> ou C<sub>δ</sub> ocorre em nível de encadeamento de RNA, para gerar cadeias pesadas para IgM e IgD, respectivamente. Ao contrário, mudança de classe para gerar IgA, IgE ou IgG envolve a união da mesma unidade VDJ em nível de DNA a, respectivamente, uma seqüência C<sub>α</sub>, C<sub>ε</sub> ou, como ilustrado na figura, uma C<sub>γ</sub>.

### Leituras adicionais

- Day DA, Tuite MF (1998) Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. *J. Endocrinol.*, 157, 361–371.
- Gray NK, Wickens M (1998) Control of translation initiation in animals. *Annu. Rev. Cell Dev. Biol.*, 14, 399–458.
- Latchman D (1998) *Gene Regulation. A Eukaryotic Perspective*. Stanley Thornes, Cheltenham.
- Roitt I, Brostoff J, Male D (1985) *Immunology*. Gower Medical Publishing, London.
- Russo E, Martienssen RA, Riggs AD (1996) *Epigenetics and Mechanisms of Gene Regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Travers A (1993) *DNA-Protein Interactions*. Chapman & Hall, London.
- van Driel R, Otte AP (1997) *Nuclear Organization, Chromatin Structure and Gene Expression*. Oxford University Press, Oxford.

## Referência eletrônica

1. <http://www.mgu.har.mrc.ac.uk/imprinting>

## Referências

- Ainger K, Avossa D, Diana AS, Barry C, Barbaresi E, Carson JH (1997) Transport and localization elements in myelin basic protein mRNA. *J. Cell Biol.*, 138, 1077–1087.
- Antequera F, Bird A (1993) Number of CpG islands and genes in human and mouse. *Proc. Natl Acad. Sci. USA*, 90, 11995–11999.
- Ashkenas J (1997) Gene regulation by mRNA editing. *Am. J. Hum. Genet.*, 60, 2378–2383.
- Ayoubi TA, Van De Ven WJ (1996) Regulation of gene expression by alternative promoters. *FASEB J.*, 10, 453–460.
- Bestor TH (1998) Cytosine methylation and the unequal developmental potentials of the oocyte and sperm genomes. *Am. J. Hum. Genet.*, 62, 1269–1273.
- Bird A (1995) Gene number, noise reduction and biological complexity. *Trends Genet.*, 11, 94–100.
- Blackwood EM, Kadonaga JT (1998) Going the distance: a current view of enhancer action. *Science*, 281, 60–63.
- Brannan CI, Bartolomei MS (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.*, 9, 164–170.
- Burke TW, Kadonaga JT (1997) The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev.*, 11, 3020–3031.
- Chess A (1998) Expansion of the allelic exclusion principle. *Science*, 279, 2067–2068.
- Cook GP, Tomlinson, IM, Walter G, Riethman H, Carter NP, Buluwela L, Winter G, Rabbitts TH (1994) A map of the immunoglobulin V<sub>H</sub> locus completed by analysis of the telomeric region of chromosome 14q. *Nature Genet.*, 7, 162–168.
- Cox GF, Kunkel LM (1997) Dystrophies and heart disease. *Curr. Opin. Cardiol.*, 12, 329–343.
- Day DA, Tuite MF (1998) Post-transcriptional gene regulatory mechanisms in eukaryotes. *J. Endocrinol.*, 157, 361–371.
- Duthie SM, Nesterova TB, Formstone EJ, Keohane AM, Turner BM, Zakian SM, Brockdorff N (1999) XIST RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material in cis. *Hum. Molec. Genet.*, 8, 195–204.
- Edwards-Gilbert G, Veraldi KL, Milcarek C (1997) Alternative poly (A) site selection in complex transcription units: means to an end? *Nucleic Acids Res.*, 13, 2547–2561.
- Geyer PK (1997) The role of insulator elements in defining domains of gene expression. *Curr. Opin. Genet. Dev.*, 7, 242.
- Grosveld F, Dillon N, Higgs D (1993) The regulation of human globin gene expression. *Baillière's Clin. Haematol.*, 6, 31–55.
- Grosveld F (1999) Activation by locus control regions? *Curr. Opin. Genet. Dev.*, 9, 152–157.
- Hazelrigg T (1998) The destinies and destinations of RNAs. *Cell*, 95, 451–460.
- Heard E, Lovell-Badge R, Avner P (1999) Anti-Xistentialism. *Nature Genet.*, 21, 343–344.
- Hewitt JE, Lyler R, Clark LN et al. (1994) Analysis of the tandem repeat locus D4Z4 associated with fascioscapulo-humeral muscular dystrophy. *Hum. Molec. Genet.*, 3, 1287–1295.
- Jiang Z-H, Zhang W-J, Rao Y, Wu JY (1998) Regulation of lch-1 pre-mRNA alternative splicing and apoptosis by mammalian splicing factors. *Proc. Natl Acad. Sci. USA*, 95, 9155–9160.
- Jones PA (1999) The DNA methylation paradox. *Trends Genet.*, 15, 34–37.
- Karin M, Hunter T (1995) Transcriptional control by protein phosphorylation: signal transmission from the cell surface to the nucleus. *Curr. Biol.*, 5, 747–757.
- Kioussis D, Festenstein R (1997) Locus control regions: overcoming heterochromatin-induced gene inactivation in mammals. *Curr. Opin. Genet. Dev.*, 7, 614–619.
- Klausner RD, Rouault TA, Harford JB (1993) Regulating the fate of mRNA: the control of cellular iron metabolism. *Cell*, 72, 19–28.
- Kleinjan D-J, van Heyningen V (1998) Position effect in human genetic disease. *Hum. Mol. Genet.*, 7, 1611–1618.
- Larsen F, Gundersen G, Lopez R, Prydz H (1992) CpG islands as gene markers in the human genome. *Genomics*, 13, 1095–1107.
- Larsson SH, Charlieru JP, Miyagawak K et al. (1995) Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. *Cell*, 81, 391–401.
- Lee JT, Davidow LS, Warshawsky D (1999) Tsix, a gene antisense to Xist at the X-inactivation centre. *Nature Genet.*, 21, 400–404.
- Lee MP, Brandenburg S, Landes GM, Adams M, Miller G, Feinberg AP (1999) Two novel genes in the center of the 11p15 imprinted domain escape genomic imprinting. *Hum. Mol. Genet.*, 8, 683–690.
- Lopez AJ (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.*, 32, 279–305.
- Lyon MF (1999) X-chromosome inactivation. *Curr. Biol.*, 9, R235–R237.
- Mertineit C, Yoder JA, Taketo T, Laird DW, Trasler JM, Bestor TH (1998) Sex-specific exons control DNA methyltransferase in mammalian germ cells. *Development*, 125, 889–897.
- Migeon BR (1994) X-chromosome inactivation: molecular mechanisms and genetic consequences. *Trends Genet.*, 10, 230–235.
- Ng H-H, Bird A (1999) DNA methylation and chromatin modification. *Curr. Opin. Genet. Dev.*, 9, 158–163.
- Nikolov DB, Burley SK (1997) RNA polymerase II transcription initiation: a structural view. *Proc. Natl Acad. Sci. USA*, 94, 15–22.

- Ogbourne S, Antalis TM** (1998) Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J.*, 331, 1–14.
- Ohlsson R, Tycko B, Sapienza C** (1998) Monoallelic expression: 'there can only be one'. *Trends Genet.*, 14, 435–438.
- Pekhletsy RI, Chernov BK, Rubtsov PM** (1992) Variants of the 59-untranslated sequence of human growth hormone receptor mRNA. *Mol. Cell Endocrinol.*, 90, 103–109.
- Razin A, Kafri T** (1994) DNA methylation from embryo to adult. *Prog. Nucl. Acid Res.*, 48, 53–81.
- Reik W, Walter J** (1998) Imprinting mechanisms in mammals. *Curr. Opin. Genet. Dev.*, 8, 154–164.
- Schoenherr CJ, Paquette AJ, Anderson DJ** (1996) *Proc. Natl Acad. Sci. USA*, 93, 9881–9886.
- Schweizer J, Zynger D, Francke U** (1999) In vivo nuclease hypersensitivity studies reveal multiple sites of parental origin-dependent chromatin conformation in the 150 kb SNRPN transcription unit. *Hum. Molec. Genet.*, 8, 555–566.
- Simmen MW, Leitgeb S, Charlton J, Jones SJM, Harris B, Clark VH, Bird AP** (1999) Nonmethylated transposable elements and methylated genes in a chordate genome. *Science*, 283, 1164–1167.
- Siomi H, Dreyfuss G** (1997) RNA-binding proteins as regulators of gene expression. *Curr. Opin. Genet. Dev.*, 7, 345–353.
- Skuse DH, James RS, Bishop DV et al.** (1997) Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature*, 387, 705–708.
- Smale ST** (1997) Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim. Biophys. Acta*, 1351, 73–88.
- Stebbins-Boaz B, Richter JD** (1997) Translational control during early development. *Crit. Rev. Eukaryotic Gene Expression*, 7, 73–94.
- Steinmetz EJ** (1997) Pre-mRNA processing and the CTD of RNA polymerase II: the tail that wags the dog? *Cell*, 89, 491–494.
- Tilghman SM** (1999) The sins of the fathers and mothers: genomic imprinting in mammalian development. *Cell*, 96, 185–193.
- Wahle E, Kuhn** (1997) The mechanism of 3' cleavage and polyadenylation of 3' eukaryotic pre-mRNA. *Prog. Nucl. Acid Res. Mol. Biol.*, 57, 41–71.
- Walsh CP, Bestor TH** (1999) Cytosine methylation and mammalian development. *Genes Dev.*, 13, 26–34.
- Weichhold GM, Klobeck H-G, Ohnheiser R, Combriato G, Zachau HG** (1990). Megabase inversions in the human genome as physiological events. *Nature*, 347, 90–92.
- Wickens M, Anderson P, Jackson RJ** (1997) Life and death in the cytoplasm: messages from the 3' end. *Curr. Opin. Genet. Dev.*, 7, 220–232.
- Yoder JA, Walsh CP, Bestor TH** (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.*, 13, 335–340.