

# GENÉTICA MOLECULAR BÁSICA

---

DOS GENES AOS GENOMAS

## **Carlos F. M. Menck**

Biólogo. Doutor em Bioquímica pela Universidade de São Paulo (USP).  
Pós-Doutorado pelo Institut des Recherches Scientifiques sur le Cancer (IRSC, França).  
Livre-docência em Genética pela USP. Professor das disciplinas Microbiologia Básica, Biologia Molecular e Biologia Molecular e Celular do Departamento de Microbiologia do Instituto de Ciências Biomédicas da USP (ICB-USP).

## **Marie-Anne Van Sluys**

Bióloga. Especialista em Biologia Celular pela Universidade do Estado do Rio de Janeiro (UERJ). Doutora em Genética e Microbiologia pela Université Paris XI. Professora Titular das disciplinas Diversidade Biológica e Filogenia, Biologia e Evolução em Procaríotos do Departamento de Botânica do Instituto de Biociências da Universidade São Paulo (IB-USP).



- Os autores deste livro e a EDITORA GUANABARA KOOGAN empenharam seus melhores esforços para assegurar que as informações e os procedimentos apresentados no texto estejam em acordo com os padrões aceitos à época da publicação, e todos os dados foram atualizados pelos autores até a data da entrega dos originais à editora. Entretanto, tendo em conta a evolução das ciências da saúde, as mudanças regulamentares governamentais e o constante fluxo de novas informações sobre terapêutica medicamentosa e reações adversas a fármacos, recomendamos enfaticamente que os leitores consultem sempre outras fontes fidedignas, de modo a se certificarem de que as informações contidas neste livro estão corretas e de que não houve alterações nas dosagens recomendadas ou na legislação regulamentadora.
- Os autores e a editora se empenharam para citar adequadamente e dar o devido crédito a todos os detentores de direitos autorais de qualquer material utilizado neste livro, dispondo-se a possíveis acertos posteriores caso, inadvertida e involuntariamente, a identificação de algum deles tenha sido omitida.
- Direitos exclusivos para a língua portuguesa  
Copyright © 2017 by EDITORA GUANABARA KOOGAN LTDA.  
Selo integrante do GEN | Grupo Editorial Nacional  
Travessa do Ouvidor, 11  
Rio de Janeiro - RJ - CEP 20040-040  
Tels.: (21) 3543-0770/(11) 5080-0770 | Fax: (21) 3543-0896  
www.grupogen.com.br | editorial.saude@grupogen.com.br
- Reservados todos os direitos. É proibida a duplicação ou reprodução deste volume, no todo ou em parte, em quaisquer formas ou por quaisquer meios (eletrônico, mecânico, gravação, fotocópia, distribuição pela Internet ou outros), sem permissão, por escrito, da EDITORA GUANABARA KOOGAN LTDA.
- Capa: Bruno Sales
- Editoração eletrônica: Alexandre Uehara
- Ficha catalográfica

---

M514g

Menck, Carlos F. M.  
Genética molecular básica : dos genes aos genomas / Carlos F. M. Menck, Marie-Anne Van Sluys. - 1. ed. - Rio de Janeiro : Guanabara Koogan, 2017.  
528 p. : il. ; 28 cm.

Inclui bibliografia e índice  
ISBN 978-85-277-3167-6

1. Citologia. 2. Biologia molecular. I. Shuys, Marie-Anne van. II. Título.

17-42007

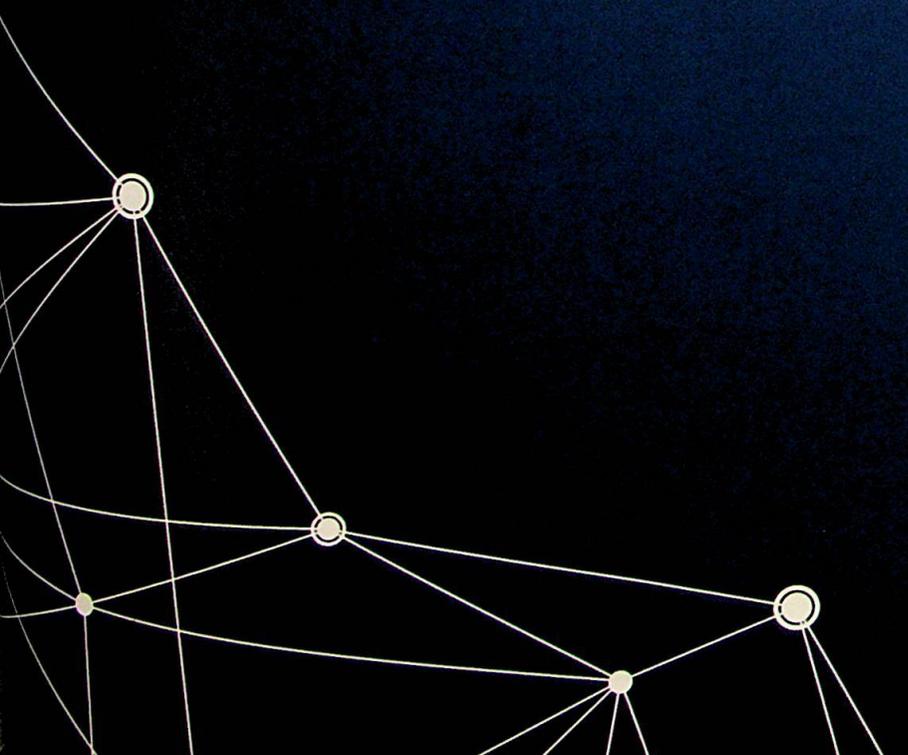
CDD: 574.87  
CDU: 576

## Parte 5

---

# Era da Genômica

INSTITUTO DE BIOCIÊNCIAS - USP  
BIBLIOTECA



# Manipulando o Gene | Técnicas de Biologia Molecular

Rodrigo da Silva Galhardo e Cristina Elisa Alvarez Martinez

Neste capítulo serão descritos os fundamentos das principais técnicas de biologia molecular. Essas metodologias são rotineiras em laboratórios de pesquisa, bem como em investigações clínicas e forenses. Frequentemente, os métodos de manipulação de DNA e RNA *in vitro* envolvem o uso de enzimas purificadas, oriundas de diversos organismos. Primeiramente, serão apresentadas as metodologias básicas de manipulação do DNA, que tornam possível a separação e a clonagem de fragmentos de DNA ou cDNA. Posteriormente, serão explicadas as metodologias que permitem a amplificação e o sequenciamento de DNA em pequena e em larga escala, assim como o estudo da expressão gênica. Algumas aplicações práticas, como a identificação humana por meio da análise de DNA, também serão abordadas.

## Introdução

Até a metade do século 20, a manipulação de genes individuais ou de pequenos trechos regulatórios do DNA parecia uma tarefa intratável para os geneticistas. Tal fato se deve principalmente à natureza da molécula de DNA e da informação nela contida. Apesar de os diferentes genes atuarem como entidades funcionalmente individuais, estes não são isolados em unidades fisicamente discretas nas células. Isso porque eles se encontram como componentes de uma grande molécula de DNA, que contém inúmeros outros genes e regiões não codificantes. No entanto, uma verdadeira revolução nos permitiu compreender melhor a estrutura e a função dos ácidos nucleicos, a partir das décadas de 1970 e 1980, ao ponto de, atualmente, ser comum isolar e estudar um trecho de DNA em particular. Esses avanços deram origem à disciplina à qual hoje nos referimos como biologia molecular, que nada mais é que a incorporação de metodologias de bioquímica e biofísica ao estudo da genética.

Curiosamente, as ferramentas que proporcionaram tal revolução não foram inventadas pelo ser humano, e sim originárias da própria natureza. São enzimas provenientes majoritariamente de microrganismos, envolvidas nos processos fisiológicos básicos ligados ao metabolismo de DNA, tais como DNA polimerases, DNA ligases e nucleases. Essas enzimas podem ser purificadas em grandes quantidades e usadas para manipular DNA em um tubo de ensaio.

Neste capítulo, será abordado o conjunto básico de metodologias mais usadas em estudos de biologia molecular que proporcionaram (e continuam proporcionando) uma verdadeira revolução na biologia moderna. Contudo, é preciso considerar que é impossível descrever em detalhe todas as metodologias atualmente existentes e toda a gama de aplicações possíveis para elas. Pretende-se aqui lançar uma base sólida para o entendimento das ferramentas básicas de biologia molecular e algumas de suas aplicações mais comumente

empregadas. Portanto, parte-se das técnicas mais fundamentais para separação, clonagem, detecção e amplificação de ácidos nucleicos. Em outros capítulos, serão apresentados alguns exemplos de experimentos nos quais essas técnicas fundamentais são empregadas para desvendar a função de genes em sistemas biológicos.

## Enzimas de restrição | Ferramentas básicas da engenharia genética

O primeiro grande passo em direção à biologia molecular moderna foi, sem dúvidas, a descoberta das enzimas de restrição. A importância dessas enzimas como ferramentas de biologia molecular é tão grande que seus principais descobridores – Daniel Nathans, Werner Arber e Hamilton Smith – foram agraciados com o Prêmio Nobel de Medicina e Fisiologia em 1978.

Tais enzimas são endonucleases, ou seja, capazes de promover a clivagem de uma ligação fosfodiéster do DNA; elas estão presentes em bactérias e fazem parte do sistema chamado “restrição-modificação”. A função biológica desse sistema é proteger as bactérias contra o ataque de bacteriófagos, pois as endonucleases de restrição clivam o DNA viral quando este entra na célula, impedindo que o bacteriófago inicie um ciclo infeccioso. A lógica desse sistema consiste em dois fatores. Em primeiro lugar, essas enzimas são altamente específicas: clivam o DNA apenas quando ele apresenta uma determinada sequência de nucleotídeos em particular, comumente chamada de *sítio de restrição*. Em segundo lugar, o reconhecimento dessa sequência é afetado pela metilação do DNA. Assim, a mesma bactéria que produz uma determinada enzima de restrição que reconhece a sequência “X” no DNA viral também protege o seu próprio DNA, por produzir uma enzima chamada metilase, impossibilitando o reconhecimento desse mesmo sítio “X” no seu genoma pela endonuclease. Desse modo, essas enzimas funcionam como um mecanismo de defesa para as

bactérias que as contêm. A nomenclatura dessas enzimas se dá de acordo com o microrganismo do qual cada uma provém. Por exemplo, a enzima *EcoRI* foi isolada de *Escherichia coli*, e a enzima *PstI* é produzida pela bactéria *Providencia stuartii*.

Como mencionado anteriormente, tais enzimas reconhecem sequências específicas no DNA. Existem três tipos de enzimas de restrição (tipos I, II e III). As enzimas de restrição do tipo II cortam o DNA em uma posição definida, dentro do próprio sítio de reconhecimento; até o momento, milhares de enzimas dessa classe já foram identificadas. Já as enzimas dos tipos I e III reconhecem o seu sítio específico, mas cortam a região em seu entorno de maneira aleatória (ou até mesmo a uma grande distância do sítio de reconhecimento) e, por essa razão, apresentam menor utilidade como ferramenta de biologia molecular.

**Existe uma grande variedade de sítios de restrição e tipos de corte**

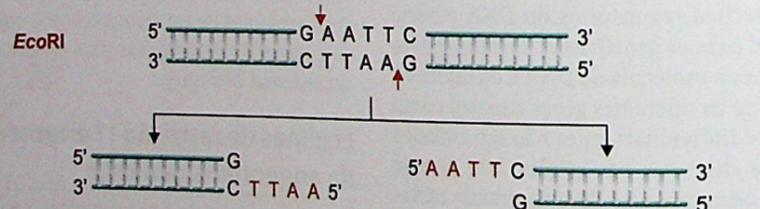
Os sítios de enzimas de restrição apresentam características específicas. A maioria das enzimas reconhece sequências de 6 nucleotídeos, mas muitas outras reconhecem sequências de 4 ou 8 nucleotídeos. Algumas enzimas de restrição e seus sítios de reconhecimento estão ilustrados na Figura 11.1.

Embora haja algumas variações, a maioria das enzimas de restrição funciona de maneira semelhante. Primeiramente, os sítios de restrição costumam ser sequências palindrômicas diretas (p. ex., o sítio da enzima *EcoRI*, mostrado na

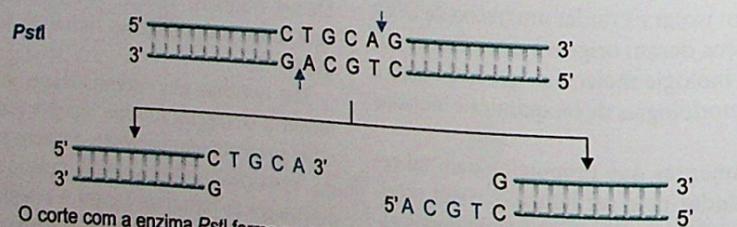
Figura 11.1). A sequência 5'-GAATTC-3' é uma palíndromo, pois a sequência complementar que lida no mesmo sentido (5'-3') é idêntica. Outra característica das enzimas de restrição é a especificidade do corte. A ligação fosfodiéster que é clivada em uma fita é também a região clivada na outra fita da palíndromo. Consideremos as enzimas mostradas na Figura 11.1. A enzima *EcoRI* corta o seu sítio 5'-GAATTC-3' sempre da mesma maneira: entre o primeiro e o segundo nucleotídeo no sentido 5'-3' em ambas as fitas. O DNA resultante desse corte contém extremidades ditas *coesivas*. Tal denominação ocorre justamente pelo fato de que essas pontas podem se reassociar pela formação de pontes de hidrogênio entre as bases nitrogenadas das extremidades protuberantes. No caso de DNA cortado com *EcoRI*, ambas as pontas têm uma extremidade protuberante 5' de 4 nucleotídeos. Outras enzimas geram pontas coesivas diferentes, com protuberância 3' de 2 nucleotídeos, ou protuberâncias 5' de 2 ou 4 nucleotídeos. A enzima *PstI*, mostrada na Figura 11.1, deixa extremidades coesivas com protuberância 3'. Existem outras enzimas que cortam exatamente no centro do palíndromo, tal como *SmaI* (Figura 11.1). Neste caso, as pontas de DNA geradas não apresentam protuberâncias, e são chamadas de pontas *cegas*.

**A frequência de sítios de restrição é variável**

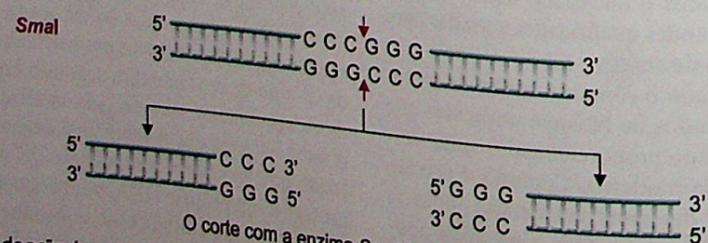
Imagine uma sequência de DNA aleatória composta de proporções iguais de A, C, G e T. A probabilidade de um sítio qualquer de 6 nucleotídeos ocorrer nessa sequência é



O corte com a enzima *EcoRI* forma pontas coesivas com extremidades protuberantes 5'



O corte com a enzima *PstI* forma pontas coesivas com extremidades protuberantes 3'



O corte com a enzima *SmaI* forma pontas cegas

**Figura 11.1** Mecanismo de ação de enzimas de restrição. A figura mostra os sítios de reconhecimento e o corte promovido no DNA pelas enzimas *EcoRI*, *PstI* e *SmaI*.

$\frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4}$ ; ou seja,  $\frac{1}{4.096}$ . Isso quer dizer que há 1 sítio a cada 4 mil pares de base (pb), em média. Seguindo o mesmo raciocínio, enzimas que reconhecem sequências de apenas 4 nucleotídeos irão cortar, em média, 1 sítio a cada 256 pb, e enzimas que reconhecem sítios de 8 nucleotídeos irão cortar uma vez a cada 65.000 pb. No entanto, a sequência do DNA nos cromossomos dos seres vivos obviamente não é aleatória, e nem sempre moléculas de DNA apresentam igual proporção dos 4 nucleotídeos. Assim, até mesmo entre as enzimas com sítios de reconhecimento de mesmo tamanho existe uma grande variedade na frequência desses sítios em diferentes genomas. Logo, há flexibilidade no uso dessas ferramentas e, dependendo da aplicação, pode-se desejar usar uma enzima que corte com maior ou menor frequência o DNA em estudo.

**Enzimas de restrição e os primeiros mapas genéticos moleculares**

As enzimas de restrição são ferramentas básicas no processo mais fundamental da biologia molecular: a clonagem (discutida mais adiante neste capítulo). Contudo, cabe ressaltar que essas enzimas proporcionaram também outro avanço na genética: elas tornaram possível a construção dos primeiros *mapas físicos* de moléculas de DNA. Até a década de 1970, os mapas de cromossomos de organismos bem estudados eram *mapas genéticos*. No Capítulo 19, é mostrado em detalhes o modo como mapas genéticos humanos são obtidos. Basicamente, os mapas genéticos têm como base a posição relativa de genes que estão ligados (*i. e.*, no mesmo cromossomo), e dependem de dois fatores para sua construção: genes que produzem um fenótipo detectável quando alterados e análise de cruzamentos (ou análise de história familiar, em caso de humanos). Já os mapas físicos baseiam-se em características da molécula de DNA em si, ou seja, na sua composição de nucleotídeos.

**Ácidos nucleicos de tamanhos diferentes podem ser separados por eletroforese**

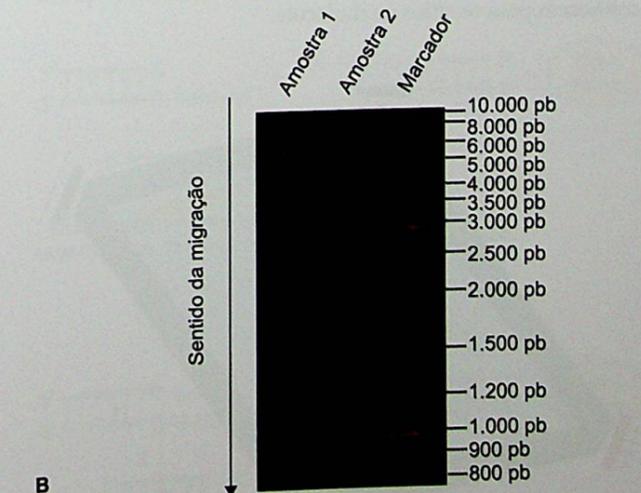
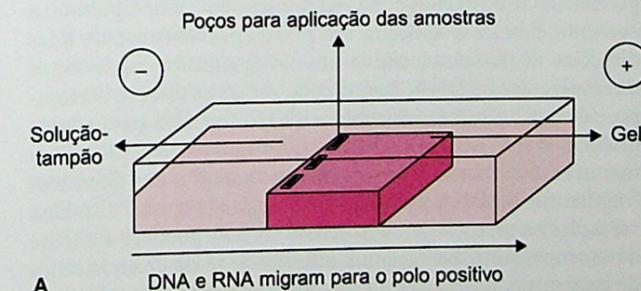
Antes de continuar a discussão sobre as ferramentas utilizadas em biologia molecular, vale apresentar um método fundamental para todas as demais técnicas: a eletroforese em gel. Essa técnica é utilizada para a análise de amostras de DNA de diversas origens, tais como plasmídeos e vírus, e para separar fragmentos de DNA obtidos após digestão com enzimas de restrição ou em vários outros procedimentos de biologia molecular que serão discutidos nas próximas seções. A eletroforese baseia-se na separação de moléculas de acordo com sua carga total. Assim, uma amostra qualquer a ser analisada é submetida a um campo elétrico, e as moléculas com carga total positiva ou negativa migram para o ânodo (polo negativo) ou cátodo (polo positivo), respectivamente. A eletroforese para a análise de amostras de DNA é realizada em fase sólida, utilizando-se uma matriz porosa chamada de *gel*. Uma vez que o DNA tem carga total negativa (pela presença do grupo fosfato na cadeia fosfodiéster externa), sua migração durante a eletroforese em gel depende apenas de conformação (linear ou circular, relaxada ou superenovelada) e de seu tamanho. As amostras de DNA a serem analisadas são colocadas em poços formados no gel, e a corrente elétrica é aplicada na presença de uma solução tampão, o que torna possível que o DNA migre através dos poros do gel na direção do cátodo (Figura 11.2). Quando o DNA está na conformação linear, sua migração

passa a depender apenas de seu tamanho, possibilitando separar e identificar o tamanho de diferentes fragmentos de DNA em uma amostra. Para isso, também é colocada no gel uma amostra de DNA com fragmentos de tamanho conhecido, chamada de *marcador de peso molecular* (Figura 11.2). Após migração em gel, o DNA pode ser visualizado pelo uso de um corante fluorescente capaz de ligar-se ao DNA, como o *bromoeto de etídeo*, uma molécula que se intercala entre as bases nitrogenadas da fita de DNA, gerando fluorescência facilmente visualizada sob luz ultravioleta (Figura 11.2).

**A separação de ácidos nucleicos por eletroforese pode ser feita em géis de agarose ou de poliacrilamida**

De modo geral, a eletroforese torna possível separar fragmentos de DNA na faixa de 10 pb até 50 Kpb. No entanto, a matriz a ser utilizada varia de acordo com o grau de separação e a faixa de tamanho que se deseja analisar, podendo ser *agarose* ou *poliacrilamida*.

A agarose é um polissacarídeo que forma poros de 100 a 300 nm de diâmetro e é bastante utilizada para analisar amostras de DNA até 10 Kpb, sendo a mais comumente usada no dia a dia de um laboratório de biologia molecular. Além disso, é possível variar a concentração de agarose no gel de acordo com a faixa de tamanho do DNA que será analisado: géis contendo maiores concentrações de agarose formam poros



**Figura 11.2** Separação de DNA por eletroforese em gel de agarose. **A.** Esquema de aparato utilizado para a corrida eletroforética horizontal de DNA e RNA. **B.** Visualização sob luz ultravioleta de fragmentos de DNA depois da separação em gel e coloração com brometo de etídeo. Pela comparação com o marcador de peso molecular, é possível determinar os tamanhos dos fragmentos obtidos após digestão com enzimas de restrição.

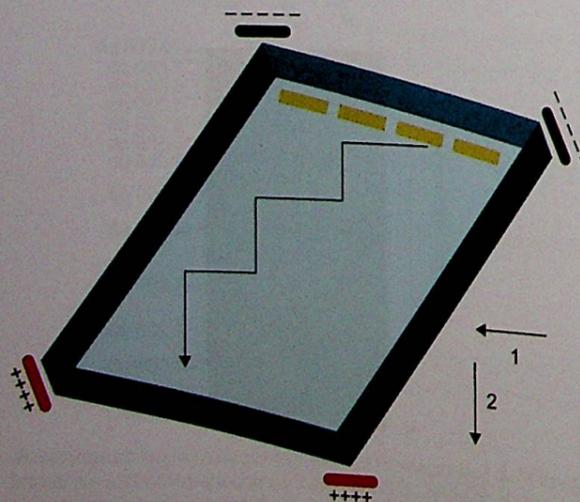
INSTITUTO DE BIÊNCIAS - USP

menores, sendo usados para a análise de fragmentos pequenos; géis com menores concentrações de agarose formam poros maiores, que facilitam a migração de fragmentos grandes.

No entanto, quando a amostra de DNA a ser analisada é composta por fragmentos pequenos, menores que 100 pb, ou quando se deseja obter um grau de separação maior de fragmentos de até 1.500 pb, utiliza-se a *poliacrilamida*. Nesse caso, o gel é composto por uma mistura de dois monômeros: acrilamida e bisacrilamida, e a proporção de cada uma delas, juntamente com a concentração final dos monômeros, determina o tamanho dos poros e, conseqüentemente, o grau de separação que será obtido. Esse tipo de gel torna possível identificar diferenças de apenas um nucleotídeo entre diferentes amostras, o que é bastante útil, por exemplo, para o sequenciamento de DNA.

### Grandes trechos de DNA podem ser separados por eletroforese em campo pulsado

Fragmentos de DNA na faixa de dezenas de Kpb não são capazes de penetrar pelos poros formados pela agarose na eletroforese convencional e, por isso, migram a partir de uma das extremidades da molécula, independentemente do seu tamanho. Desse modo, não é possível obter um grau de separação satisfatório das moléculas nessa faixa de tamanho por meio da eletroforese convencional. Assim, a *eletroforese em campo pulsado* foi desenvolvida para possibilitar a análise precisa de grandes fragmentos de DNA. Em vez de se utilizar de um campo elétrico unidirecional, na eletroforese em campo pulsado a corrente elétrica é aplicada em pulsos provenientes de duas direções, de maneira a causar um movimento em zigue-zague das moléculas de DNA. Assim, cada vez que a direção do campo elétrico é alterada, as moléculas são forçadas para a nova direção, o que ocorre de modo mais rápido com moléculas menores, facilitando a separação por tamanho dos diferentes fragmentos de DNA presentes na amostra (Figura 11.3). Essa variação na técnica de eletroforese tornou possível a análise de cromossomos bacterianos inteiros e até de cromossomos de eucariotos superiores, desde que previamente cortados em fragmentos menores, pelo uso de enzimas de restrição que reconhecem poucos sítios na molécula.



**Figura 11.3** Esquema de funcionamento da eletroforese em campo pulsado. O campo elétrico é continuamente alternado entre as orientações 1 e 2, causando a reorientação do DNA e o seu movimento em zigue-zague, conforme indicado pela seta.

A eletroforese em gel também é comumente utilizada para a análise de amostras de RNA. A molécula de RNA forma inúmeras estruturas secundárias que interferem na análise de tamanho por eletroforese em gel. Assim, o RNA deve ser tratado com um agente desnaturante, geralmente o *formaldeído*, antes de ser aplicado no gel.

### Fragmentos de DNA separados por eletroforese podem ser recuperados de géis

É interessante notar que a eletroforese é uma técnica analítica que possibilita separar vários fragmentos de DNA de uma mistura complexa e determinar seus tamanhos; no entanto, também pode ser usada para obter purificações de um DNA de interesse. A purificação de fragmentos de DNA a partir de amostras separadas por eletroforese é uma tarefa rotineira em laboratórios de biologia molecular. Imagine que você deseja obter uma preparação de apenas um único fragmento obtido por digestão de uma molécula de DNA com enzimas de restrição, sendo que essa digestão gerou quatro outros fragmentos de tamanhos diferentes. Após a eletroforese, o fragmento de interesse pode ser recuperado do gel, pois é possível recortar a banda do gel e purificá-la por meio de metodologias bioquímicas simples. Esse procedimento é fundamental em vários experimentos de biologia molecular, isto que, além de separar um fragmento específico dos demais, também purifica o DNA, livrando-o da presença de enzimas e sais usados em experimentos anteriores.

### Tecnologia do DNA recombinante e clonagem de genes

Para estudos de biologia molecular, é desejável que se possa obter quantidades ilimitadas do fragmento de DNA de interesse, de modo que não seja necessário obter nova preparação de DNA genômico, digestão e purificação da região em estudo. Isso se tornou possível com o advento da *tecnologia do DNA recombinante*, mais popularmente conhecida como *engenharia genética*, que permitiu que fragmentos de DNA fossem clonados – ou seja, reproduzidos em milhões de cópias idênticas. Um desdobramento da clonagem é a construção de *bibliotecas genômicas*, nas quais não se clona apenas um fragmento de interesse e, na verdade, todo o DNA de uma célula é fracionado em pequenos fragmentos e vários clones diferentes são produzidos.

Agora que já se conhece o funcionamento das enzimas de restrição, é possível imaginar que um gene (ou qualquer outra parte de uma sequência de DNA que seja de interesse para estudo) pode ser separado do restante da molécula de DNA por meio da clivagem com uma ou mais enzimas de restrição. No jargão da biologia molecular, essa clivagem é chamada de *digestão* com enzimas de restrição; tomemos como exemplo uma molécula de DNA relativamente simples, tal como o DNA do fago lambda, que tem 48,5 Kpb. É fácil imaginar que a digestão desse DNA e a posterior separação dos fragmentos resultantes por eletroforese isolem um ou mais genes de interesse do restante do DNA. Para genomas maiores, essa tarefa é mais difícil, pois muitos fragmentos de DNA do mesmo tamanho são gerados pela digestão com cada enzima de restrição. Nesse caso, a construção de bibliotecas genômicas ou a reação em cadeia da polimerase (PCR) são frequentemente empregadas, como será visto adiante. Contudo, é necessário introduzir duas ferramentas fundamentais no processo de clonagem: as DNA ligases e os vetores genéticos.

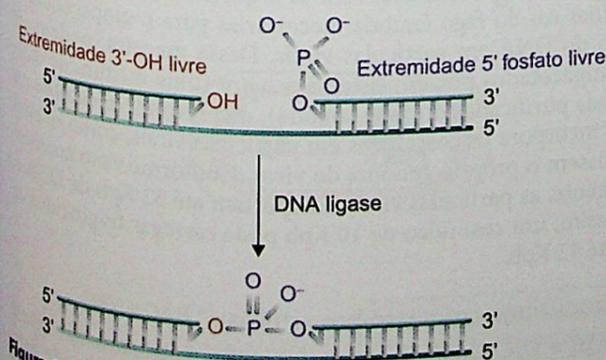
### Uso de DNA ligases em biologia molecular | Novas combinações de sequências de DNA

Assim como as enzimas de restrição, as DNA ligases são enzimas que desempenham importante função na natureza, por realizarem uma função crucial nos processos de replicação e reparo de DNA. A reação que essas enzimas catalisam pode ser colocada em termos simples: a junção de uma extremidade 5'-P livre com outra extremidade 3'-OH livre (Figura 11.4).

Imagine dois DNA de origens diferentes, como ilustrados na Figura 11.5. Ambos foram cortados com a enzima *EcoRI* e, portanto, apresentam as extremidades coesivas, com protuberâncias 3' características. Como a enzima *EcoRI* deixa sempre as mesmas pontas coesivas complementares no DNA, estas podem se associar transientemente pelo emparelhamento de bases entre as protuberâncias em fita simples. Misturando-se os dois fragmentos de DNA de origens diferentes em um tubo de ensaio e adicionando a enzima DNA ligase, é possível produzir uma molécula de DNA *quimérica*, ou *recombinante* (Figura 11.5). Esse tipo de experimento é base fundamental da *tecnologia do DNA recombinante* e possibilita a construção de novas combinações de sequências de DNA. O emparelhamento de poucas bases proporcionado pelas pontas coesivas deixadas por enzimas de restrição aumenta bastante a eficiência das reações com a DNA ligase *in vitro*, por possibilitar uma associação transiente, mas específica, entre as moléculas. A DNA ligase também é capaz de ligar duas pontas cegas de DNA, tais como aquelas produzidas pela ação da enzima *SmaI* (Figura 11.5). Contudo, a eficiência desse processo é menor que a obtida na ligação de duas pontas coesivas compatíveis.

### Vetores permitem a clonagem de genes

Conforme descrito anteriormente, é desejável que um fragmento de DNA em estudo possa ser clonado; para isso, utilizam-se os chamados vetores. De maneira geral, é possível descrever os vetores mais usados para clonagens como moléculas de DNA capazes de se replicar de maneira independente do restante do genoma, dentro de organismos-modelo de fácil manuseio em laboratório (tais como bactérias e leveduras). Por meio do uso da tecnologia do DNA recombinante, fragmentos de DNA de interesse podem ser inseridos em vetores, em uma reação catalisada por uma DNA ligase. Os vetores modificados pela presença do inserto de interesse podem ser reintroduzidos na célula hospedeira, na qual se mantêm de maneira estável, multiplicando-se ao longo das gerações deste organismo.



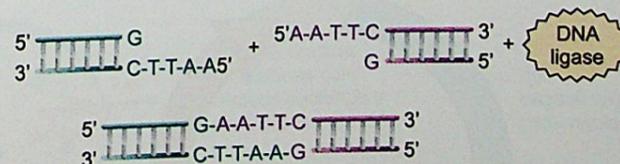
**Figura 11.4** Mecanismo simplificado de ação da DNA ligase, promovendo a formação de uma ligação fosfodiéster entre extremidades 3'-OH e 5'P livres.

### Os plasmídeos bacterianos são os principais vetores para clonagem

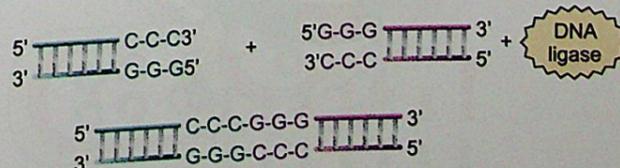
Os vetores de escolha para a maioria dos experimentos de clonagem de trechos pequenos de DNA (de até 10 Kpb) são os plasmídeos da bactéria *E. coli*. Fragmentos de DNA maiores são geralmente instáveis em plasmídeos, podendo ser clonados em outros vetores.

Plasmídeos são moléculas de DNA de replicação autônoma que ocorrem naturalmente em bactérias, apresentando, em sua grande maioria, conformação circular. Os vetores mais utilizados em biologia molecular derivam desses plasmídeos naturais, mas foram modificados extensamente pelos pesquisadores para apresentarem uma série de características desejáveis. As moléculas de DNA plasmidial utilizadas em biologia molecular são relativamente pequenas, e podem ser separadas do DNA cromossômico e isoladas das células por meio de técnicas bioquímicas relativamente simples. Algumas das características comumente encontradas nesses vetores estão listadas a seguir e representadas pelo exemplo da Figura 11.6:

- Presença de um ou vários sítios de restrição: muitos dos vetores mais modernos têm uma região inserida artificialmente que contém um grande número de sítios únicos (ou seja, que só ocorrem uma vez na molécula do plasmídeo), chamada de *sítio múltiplo de clonagem*
- Presença de um *marcador selecionável*, que possibilita uma seleção positiva (ou seja, um crescimento seletivo) das células que apresentam o plasmídeo. Os marcadores selecionáveis mais utilizados são genes de resistência a antibióticos. Assim, apenas bactérias portadoras do plasmídeo podem crescer em meio de cultura contendo o antibiótico em questão
- *Genes repórteres* que possibilitam a distinção entre bactérias que carregam plasmídeos recombinantes (isto é, com inserto) daquelas que têm o plasmídeo “vazio”. Nem todos os plasmídeos usados em clonagem apresentam essa característica, mas ela está presente nos mais populares. A maioria dos vetores mais usados tem parte do gene que codifica a enzima betagalactosidase, *lacZ*. A atividade desta enzima



DNA recombinante produzido pela ligação de duas extremidades coesivas produzidas pela enzima *EcoRI*



DNA recombinante produzido pela ligação de duas extremidades cegas produzidas pela enzima *SmaI*

**Figura 11.5** Produção de DNA recombinante. DNA de origens diferentes (representados por cores diferentes) que foram clivados com enzimas de restrição são covalentemente associados em um novo arranjo pela atividade da enzima DNA ligase.

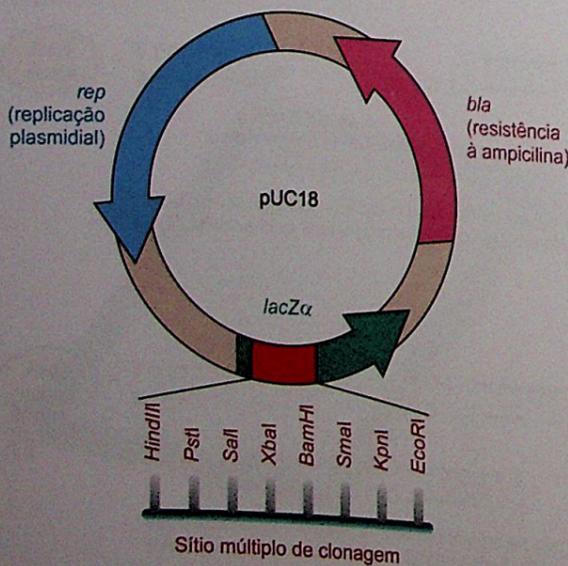
pode ser facilmente detectada em colônias de bactérias, se o meio de cultura for adicionado de um substrato cromogênico. Quando a enzima está presente, este substrato é metabolizado, dando cor azul às colônias. O sítio múltiplo de clonagem desses vetores se localiza precisamente dentro do gene *lacZ* (Figura 11.6). Quando um inserto é introduzido nesse vetor, o gene é interrompido, não produzindo mais a enzima funcional. Como consequência, colônias de bactérias que têm o plasmídeo vazio apresentam cor azul, enquanto aquelas com plasmídeos com inserto exibem cor branca.

**Introduzindo genes em vetores plasmidiais**

A Figura 11.7 mostra os principais passos em um experimento de clonagem usando um vetor plasmidial que reúne as características citadas anteriormente. Um trecho de DNA de interesse digerido com uma enzima de restrição foi isolado e ligado no plasmídeo previamente digerido com a mesma enzima, utilizando-se a enzima DNA ligase. Após a reação de ligação, essas moléculas são introduzidas nas células bacterianas por meio de *transformação genética*. Se as bactérias forem semeadas em meios de cultura contendo o antibiótico apropriado, apenas aquelas contendo plasmídeos irão crescer e formar colônias. Dentre as colônias formadas, é possível identificar aquelas que contêm plasmídeos recombinantes (ou seja, contendo o inserto de DNA desejado) pela coloração diferencial no meio de cultura, que foi acrescido do substrato cromogênico para a enzima betagalactosidase (o composto X-Gal).

O poder dessa clonagem pode ser ilustrado com uma conta simples. Ao crescer uma colônia bacteriana contendo um plasmídeo recombinante em cultura, obtém-se densidades de até 10<sup>9</sup> células/ml. Cada célula bacteriana carrega uma centena de cópias do plasmídeo. Portanto, em apenas 1 ml de cultura bacteriana, é possível obter 10<sup>11</sup> cópias do plasmídeo portando o seu gene de interesse.

É necessário destacar que a gama de plasmídeos existentes para clonagem de genes é muito grande. Muitos desses



**Figura 11.6** Mapa do vetor pUC18, mostrando suas principais características: gene de resistência a antibiótico (*bla*), sequências *rep* para replicação estável em bactérias e sítio múltiplo de clonagem inserido no gene *lacZ*.

plasmídeos apresentam características adicionais relevantes. Alguns são chamados de *vetores ponte*, pois replicam tanto em bactérias quanto em células eucarióticas; outros possibilitam que o gene clonado seja expresso, constitutivamente ou de maneira regulada, porque posicionam o gene inserido sob o controle de promotores conhecidos (para mais detalhes sobre promotores e controle da expressão gênica, ver Capítulo 6). Estes são chamados de *vetores de expressão*. Outra característica importante presente em alguns vetores é a possibilidade de expressar proteínas recombinantes em grandes quantidades e gerando fusões a pequenos peptídeos que permitem a sua purificação. Esses vetores plasmidiais são muito valiosos no estudo funcional de genes.

**Fragmentos grandes de DNA são clonados em outros tipos de vetores**

Além dos plasmídeos, a técnica de clonagem pode ser feita com vários vetores diferentes. Em geral, esses vetores alternativos podem carregar fragmentos de DNA maiores que os suportados pelos plasmídeos.

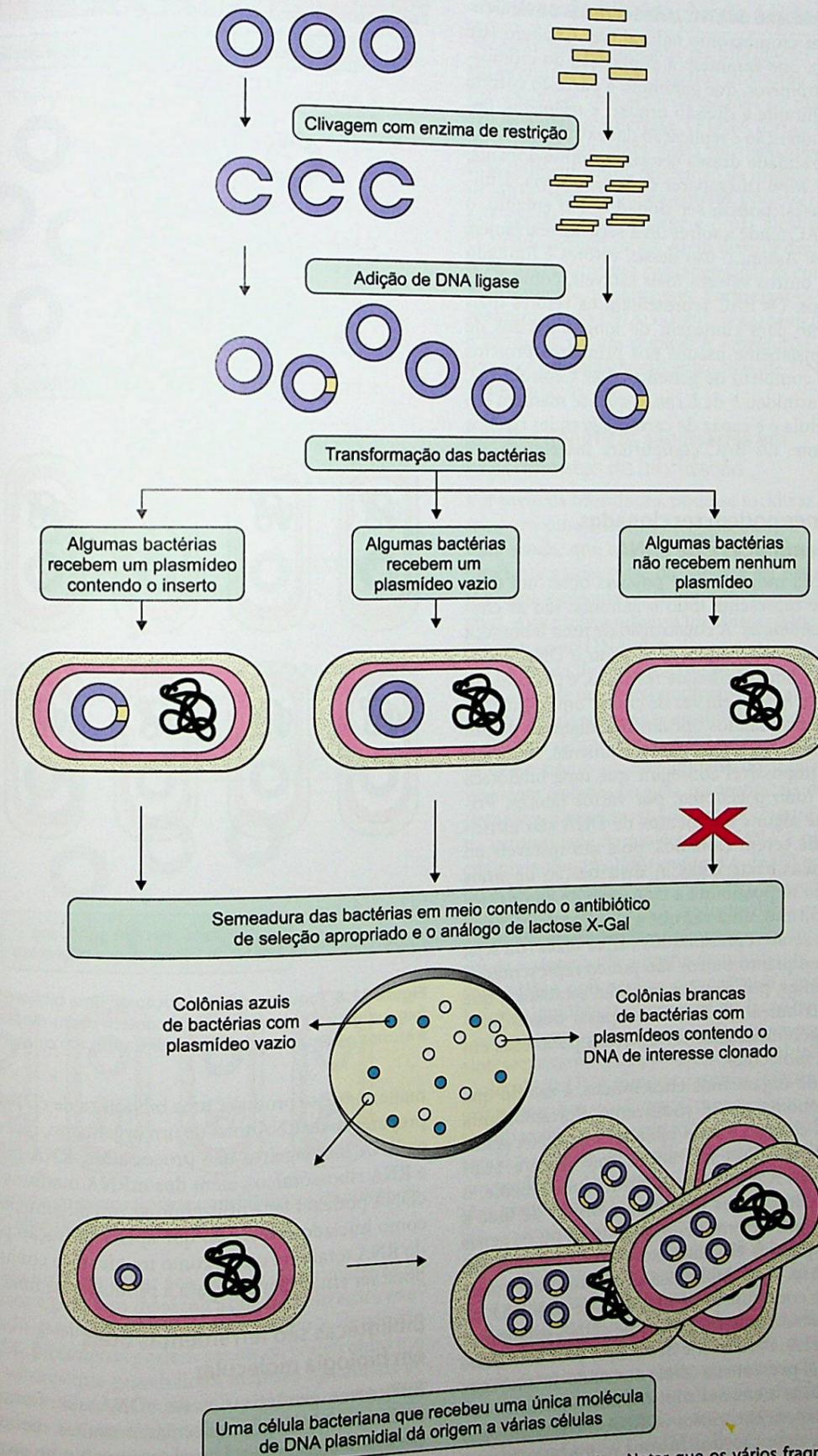
**Bacteriófagos**

Os bacteriófagos proporcionaram uma série de avanços na genética. Por serem modelos de estudo bem conhecidos, não tardaram a ser utilizados como vetores genéticos com o advento da tecnologia do DNA recombinante. O bacteriófago mais usado para este fim é o fago lambda, que tem um genoma linear de 48,5 Kpb, que é empacotado no capsídeo viral. Destes, aproximadamente 15 Kpb codificam genes dispensáveis para o ciclo lítico do fago, e podem ser substituídos por DNA exógeno por meio da metodologia do DNA recombinante. Assim, o fago lambda pode ser usado para clonar fragmentos de tamanho superior aos suportados por plasmídeos, mas o tamanho dos insertos também tem um limite. Os capsídeos virais não incorporam moléculas de DNA menores que 37 Kpb, nem maiores que 52 Kpb. Uma vez que aproximadamente 33 Kpb de DNA endógeno do fago são essenciais para seu ciclo e multiplicação, o tamanho máximo de DNA exógeno comportado por esses vetores fica em torno de 18 Kpb.

**Cosmídeos**

Cosmídeos são vetores artificiais, com características tanto de plasmídeos quanto do bacteriófago lambda. Essas moléculas podem ser mantidas como plasmídeos circulares em bactérias e geralmente apresentam tamanhos menores que 10 Kpb. Contudo, os cosmídeos têm as sequências conhecidas como *sítios cos* do fago lambda, necessárias para o empacotamento do DNA em partículas virais. Dessa maneira, podem ser empacotados por um sistema *in vitro* (uma mistura de enzimas purificadas do fago lambda), que reconhece os *sítios cos* e incorpora os cosmídeos em partículas virais, como se estes fossem o próprio genoma do vírus. Conforme visto anteriormente, as partículas virais comportam até 52 Kpb de DNA e, assim, um cosmídeo de 10 Kpb pode carregar fragmentos de até 42 Kpb.

**Produzindo cromossomos artificiais | BAC e YAC**  
BAC e YAC (do inglês, *bacterial artificial chromosomes* e *yeast artificial chromosomes*) são cromossomos artificiais de bactérias e leveduras, respectivamente. Para a replicação estável como um cromossomo nas células da levedura *Saccharomyces*



**Figura 11.7** Processo de clonagem de um fragmento de DNA em um vetor plasmidial típico. Notar que os vários fragmentos ilustrados representam uma preparação pura de fragmentos idênticos, obtida por digestão de grande quantidade de DNA com uma enzima de restrição e purificação.

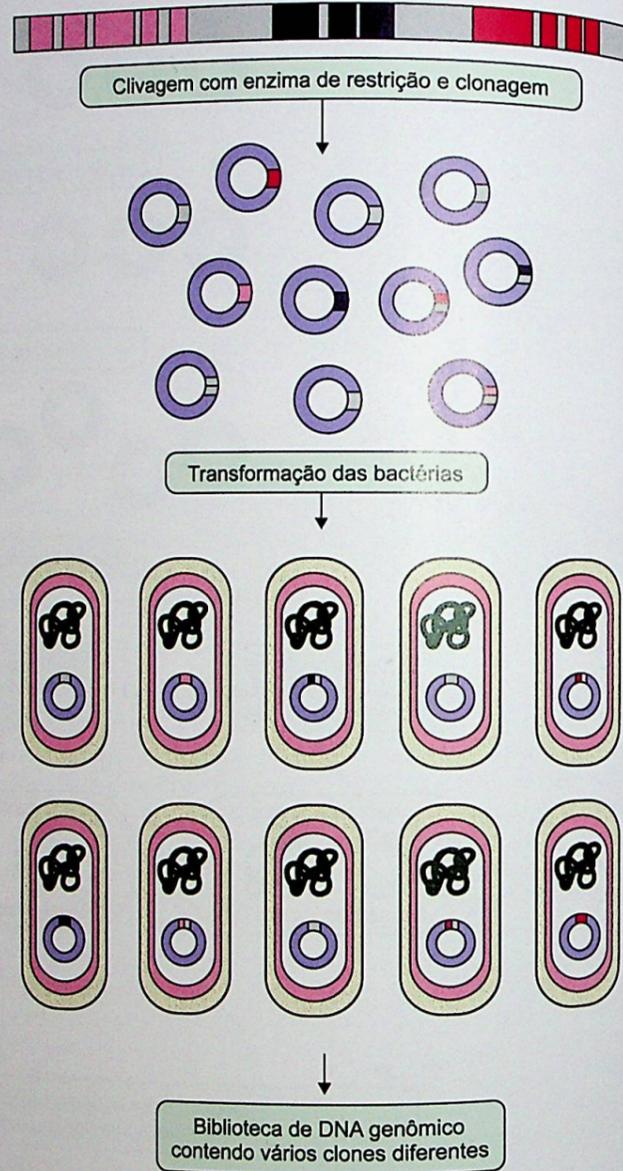
INSTITUTO DE BIOTECNICAS - USP BIBLIOTECA

*cerevisiae*, é necessário que os YAC tenham todos os elementos funcionais de um cromossomo natural (ver Capítulo 10): *origens de replicação*, que garantem a duplicação do cromossomo na fase S; *centrômeros*, que garantem a partição correta dos cromossomos durante a divisão celular; e *telômeros*, necessários para a manutenção e replicação das extremidades do cromossomo. A capacidade desses vetores é impressionante: fragmentos de até 1 Mpb (megapares de base, ou seja, 1 milhão de pares de bases) podem ser clonados. No entanto, o DNA clonado em YAC tende a sofrer uma série de rearranjos, inserções e deleções. Assim, o uso desses vetores é limitado e, quando possível, outros vetores mais estáveis, como BAC devem ser escolhidos. Os BAC representam os vetores mais populares atualmente para clonagem de longos trechos de DNA, e foram extensamente usados nos primeiros projetos de sequenciamento completo de genomas (ver Capítulo 12). São derivados do plasmídeo F de *E. coli*, o qual se mantém em poucas cópias na célula e é capaz de carregar grandes trechos de DNA estávelmente. Os BAC comportam insertos de até 300 Kpb.

### Conjuntos de genes podem ser clonados em bibliotecas genômicas e de cDNA

Além de clonar genes individuais, é possível obter um conjunto de clones que represente todo o genoma: são as chamadas *bibliotecas genômicas*. A construção de uma biblioteca genômica está ilustrada na Figura 11.8. Todo o DNA genômico é digerido com uma enzima de restrição, e clonado em um vetor apropriado. Assim, em vez de clonar um fragmento específico, todos os fragmentos obtidos por digestão de um longo trecho de DNA são clonados. Obviamente, na prática, é virtualmente impossível conseguir que uma biblioteca represente de fato todo o genoma, por vários fatores. Primeiramente, porque alguns fragmentos de DNA são intrinsecamente difíceis de serem clonados, pois são instáveis ou tóxicos para as células bacterianas. A distribuição de sítios de restrição também impossibilita a clonagem de alguns trechos de DNA. Por último, uma vez que a clonagem ocorre ao acaso, alguns genes serão representados várias vezes (ou seja, em vários clones), enquanto outros são pouco representados ou até mesmo ausentes, por simples variação estatística, que obedece a uma distribuição de Poisson. Apesar dessas limitações, bibliotecas genômicas são ferramentas poderosas em estudos de biologia molecular.

Em se tratando de organismos eucarióticos, é sabido que grande parte do genoma não é codificante. Portanto, uma imensa parcela dos clones de uma biblioteca de DNA genômico irá conter apenas DNA não codificante (Figura 11.8) ou genes incompletos. Além disso, muito frequentemente, o objetivo da construção de uma biblioteca é dispor de toda a coleção de genes de um determinado organismo ou os genes que estão sendo expressos em uma condição fisiológica em particular. Nesse caso, uma biblioteca genômica tem pouca utilidade, e pode-se construir uma *biblioteca de cDNA*. O esquema da construção de uma biblioteca de cDNA está representado na Figura 11.9. Para isso, é necessário produzir DNA a partir do RNA total presente na célula ou tecido de interesse, o que é possível graças à enzima *transcriptase reversa*. O Capítulo 6 mostrou que em eucariotos os RNA mensageiros são processados para a remoção dos introns e, posteriormente, são poli-adenilados. A presença da cauda poli-A é, portanto, uma característica dos RNA mensageiros maduros. Uma das

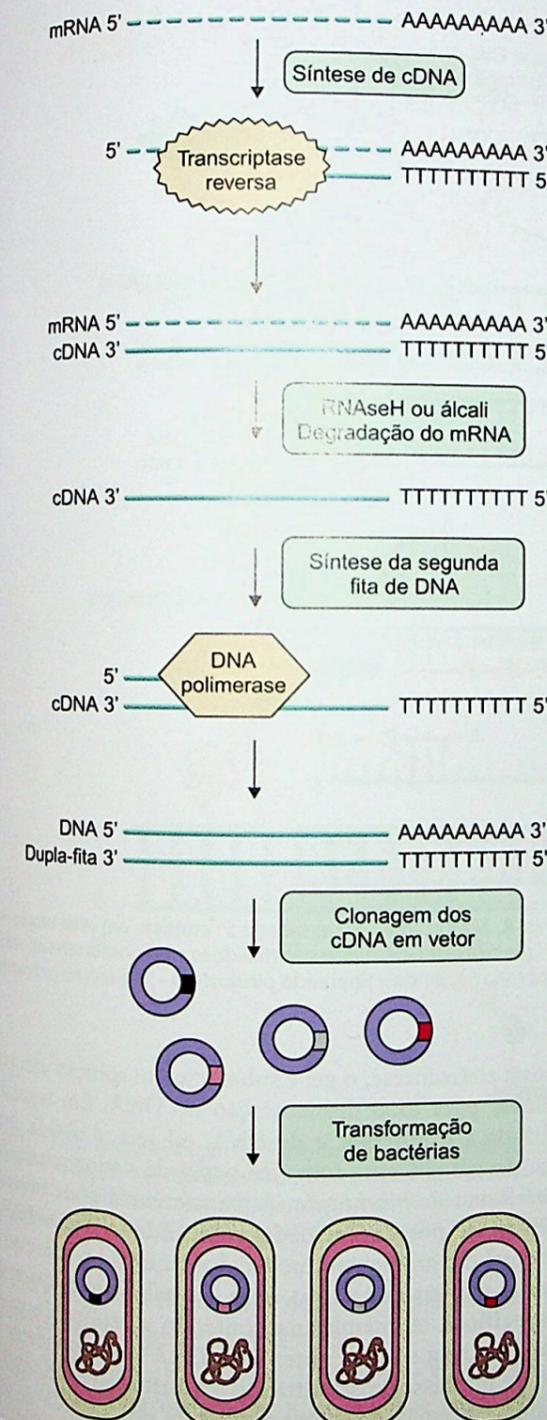


**Figura 11.8** Esquema de construção de uma biblioteca genômica. O fragmento de DNA representado contém exons destacados em cores e introns e demais regiões não codificantes em cinza.

maneiras de se produzir uma biblioteca de cDNA é obter uma preparação de RNA total de um organismo, que contém todos os RNA mensageiros não processados, RNA transportadores e RNA ribossômicos, além dos mRNA maduros. A síntese do cDNA pode ser feita utilizando-se um oligonucleotídeo poli-T como iniciador, de modo que apenas a fração poli-adenilada do RNA total será usada como molde. Este conjunto de cDNA pode ser então clonado para a produção de uma biblioteca.

### Bibliotecas são ferramentas úteis em biologia molecular

Bibliotecas genômicas e de cDNA são ferramentas muito úteis em biologia molecular, e muitas são as suas aplicações. Sua construção é o primeiro passo no sequenciamento de genomas completos utilizando a metodologia de Sanger. Uma biblioteca também pode ser usada para a obtenção de um clone de um gene de interesse. Nesse caso, é necessário



**Figura 11.9** Esquema de construção de uma biblioteca de cDNA utilizando a enzima transcriptase reversa.

identificar esse clone em meio aos demais. Uma maneira de realizar tal busca é pelo uso de técnicas de hibridização, conforme explicado posteriormente neste capítulo e detalhado na Figura 11.14. Bibliotecas podem ser ainda construídas em uma série de vetores que possibilitem análises funcionais de genes, tais como vetores de expressão e vetores-ponte. Tais bibliotecas podem ser usadas em varreduras que busquem identificar genes que desempenham funções específicas na célula. Por último, é possível obter a sequência de vários clones de uma biblioteca de cDNA, de modo a se alcançar um perfil de expressão gênica em condições de interesse.

### Técnicas de hibridação de ácidos nucleicos

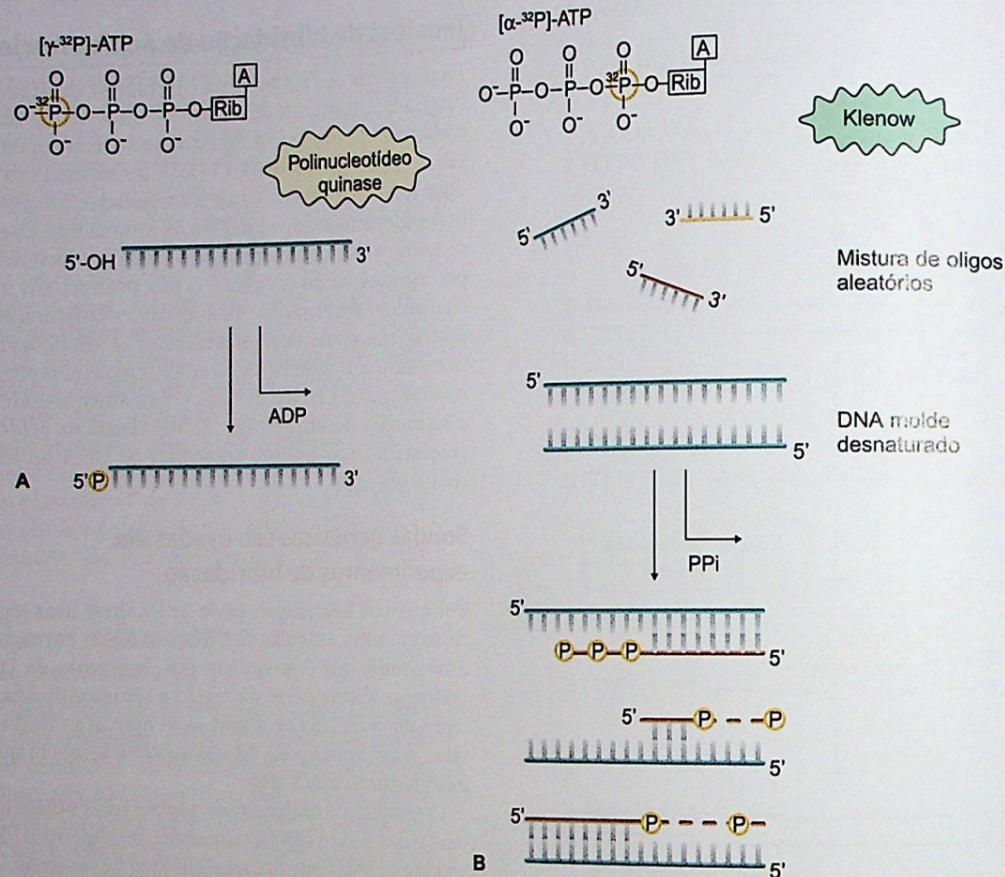
Técnicas que se baseiam na hibridação de ácidos nucleicos surgiram a partir de meados dos anos 1970, viabilizando a realização de uma série de estudos de mapeamento e expressão de genes de interesse. O DNA forma uma fita dupla devido a interações entre bases nitrogenadas *complementares* de fitas independentes, ou dentro da mesma fita. Esse pareamento por bases complementares pode ser facilmente desfeito por agentes como o calor e o pH alcalino, em um processo chamado *desnaturação*. Além disso, a desnaturação do DNA é reversível e, uma vez reestabelecidas as condições ideais, ocorre novamente o anelamento entre regiões complementares e a *renaturação* do DNA; ou seja, o pareamento entre essas bases e formação de um duplex. A *hibridação* de ácidos nucleicos consiste no anelamento entre fitas de DNA ou RNA de diferentes origens.

### Sondas genéticas são usadas em experimentos de hibridação

Por meio da hibridação, pode-se localizar uma região de interesse em uma amostra de DNA ou RNA. Para isso, utiliza-se uma *sonda*, que consiste em um fragmento de DNA correspondente a uma parte da região a ser identificada, e que contém uma molécula ou átomo com marcação facilmente detectável, o que torna possível identificar o local em que ocorreu a hibridação ao DNA alvo.

O método de marcação de sondas mais utilizado tem como base o uso de um átomo radioativo, geralmente  $^{32}\text{P}$  ou  $^{33}\text{P}$ , que produz sondas com alta sensibilidade de detecção. A marcação radioativa de sondas com  $^{32}\text{P}$  pode ser feita pelo acoplamento desse átomo na extremidade do fragmento de DNA previamente isolado, ou pela incorporação do átomo radioativo durante a síntese *in vitro* da sonda (Figura 11.10). A marcação na extremidade da molécula de DNA é realizada pelo uso da enzima *polinucleotídeo quinase* e de  $^{32}\text{P}$  acoplado à molécula do ATP na posição  $\gamma$  ( $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ ). A enzima fosforila o grupo 5'-OH do DNA desfosforilado usando o átomo marcado. Na marcação por incorporação do átomo radioativo, a sonda é sintetizada a partir de um molde, utilizando uma das várias DNA polimerases disponíveis (a mais usada é a porção Klenow da DNA polimerase I de *E. coli*) ou a técnica de PCR. Um dos quatro desoxirribonucleotídeos presentes na reação de síntese contém  $^{32}\text{P}$  na posição  $\alpha$  ( $[\alpha\text{-}^{32}\text{P}]\text{dNTP}$ ), que é retido na molécula de DNA durante a síntese (Figura 11.10). A marcação por síntese é mais eficiente, uma vez que o átomo radioativo é incorporado ao longo de toda a molécula de DNA, geralmente sendo o método de escolha para estudos de hibridação de ácidos nucleicos. No entanto, as sondas marcadas na extremidade têm grande utilidade para outras técnicas de biologia molecular – que ainda serão discutidas neste capítulo. A presença de sondas marcadas radioativamente é facilmente detectada por *autorradiografia*, que consiste na detecção da radiação emitida após exposição em filmes de raios X.

Recentemente, vários métodos não radioativos para a marcação de sondas foram desenvolvidos, e têm se tornado cada vez mais comuns por evitar riscos associados à manipulação de radioatividade. Em um desses métodos, a sonda é marcada por um nucleotídeo acoplado à molécula digoxigeninada por um nucleotídeo acoplado à molécula digoxigeninada (digoxigenina-11-dUTP). Para a detecção da sonda após a hibridação, utiliza-se um anticorpo antidigoxigenina, e o complexo antígeno-anticorpo é detectado por *fluorescência*,



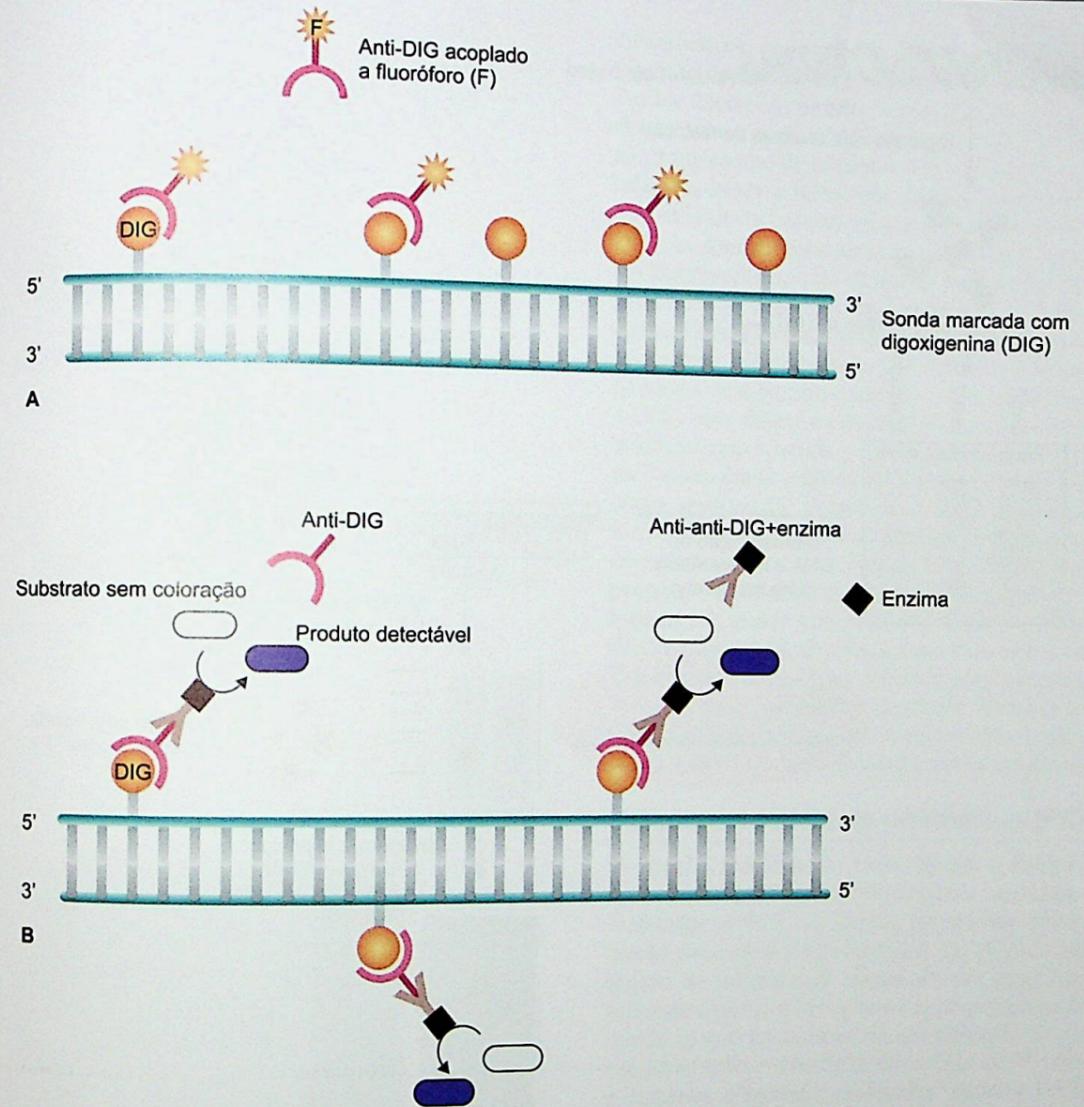
**Figura 11.10** Representação de dois métodos de marcação de sondas radioativas. **A.** Marcação na extremidade 5': consiste em uma reação de fosforilação, e apenas o fosfato da posição  $\gamma$  do ATP é incorporado na sonda. Notar que o oligonucleotídeo sintetizado *in vitro* possui uma extremidade 5'-OH (não fosforilada). **B.** Marcação por síntese: a molécula de AMP é incorporada na sonda – liberando pirofosfato –, e o átomo radioativo deve estar na posição  $\alpha$  do ATP.

quimioluminescência ou métodos colorimétricos (Figura 11.11). A detecção por fluorescência é direta, uma vez que o anticorpo está acoplado a uma molécula que emite luz quando excitada em comprimento de onda específico. Nos métodos indiretos, o anticorpo está acoplado a uma enzima, que catalisa uma reação na presença de substrato específico, formando um produto que emite luz (*método quimioluminescente*) ou um precipitado com coloração visível a olho nu (*método colorimétrico*). Além da digoxigenina, outras moléculas têm sido usadas para marcação de sondas não radioativas. No entanto, em geral, são utilizados princípios similares aos descritos aqui.

**A técnica de Southern blot detecta seqüências de DNA em amostras separadas por eletroforese**

A primeira das técnicas de hibridação usando sondas foi desenvolvida por Edwin Southern em 1975, sendo por isso denominada *Southern blot* (Figura 11.12). Esta técnica tornou possível a detecção e a identificação de um gene de interesse em uma amostra de DNA, ou até mesmo o seu mapeamento em relação a sítios de restrição no cromossomo. Inicialmente, é necessário cortar o DNA pelo tratamento com uma ou mais enzimas de restrição, de modo a obter fragmentos de DNA dentro da faixa de separação de um gel de agarose convencional. Essa mistura de fragmentos é submetida à eletroforese em gel de agarose produzindo, normalmente, um rastro de migração, em função da presença de vários fragmentos de todos os tamanhos possíveis.

Após a eletroforese, o gel é submerso em solução alcalina (pH > 13) para total desnaturação do DNA. Em seguida, é realizada a *transferência do DNA*, ou seja, é obtida uma réplica do gel em uma folha de papel de náilon, chamada de *membrana de hibridação*. A transferência pode ocorrer simplesmente por capilaridade, colocando-se a membrana sobre o gel em ambiente umedecido; ou então o processo é acelerado pelo uso de vácuo, com o auxílio de equipamentos específicos. A membrana contendo a réplica do gel é incubada com a sonda contendo o DNA correspondente à região de interesse, possibilitando a hibridação e a retenção da sonda na região da membrana onde está o DNA de seqüência idêntica. Em geral, utiliza-se um excesso de sonda, o que torna mais provável o anelamento e a hibridação do DNA alvo com a sonda, em vez da renaturação do DNA. Utilizando-se um DNA marcador de peso molecular no gel de agarose, é possível determinar o tamanho do fragmento que contém o gene ou a região de interesse. É importante destacar que a hibridação ocorre até mesmo entre fragmentos que não apresentam 100% de complementariedade, basta apenas que uma parte significativa ou várias partes distintas da sonda tenham complementariedade com o alvo. Assim, é possível encontrar um gene ainda não identificado em um organismo usando como sonda um gene já conhecido de outro organismo, desde que eles apresentem similaridade de seqüência significativa.



**Figura 11.11** Representação de métodos de detecção de sondas marcadas com dUTP-digoxigenina (DIG) por fluorescência. **A.** Detecção direta: após a hibridação, o complexo sonda-alvo é detectado pelo uso de anticorpo antidigoxigenina (anti-DIG). **B.** Detecção indireta (método colorimétrico): utiliza-se também um anticorpo secundário (anti-anti-DIG), que reconhece o anticorpo anti-DIG e está acoplado a uma enzima que catalisa reação detectada por método colorimétrico ou quimioluminescente.

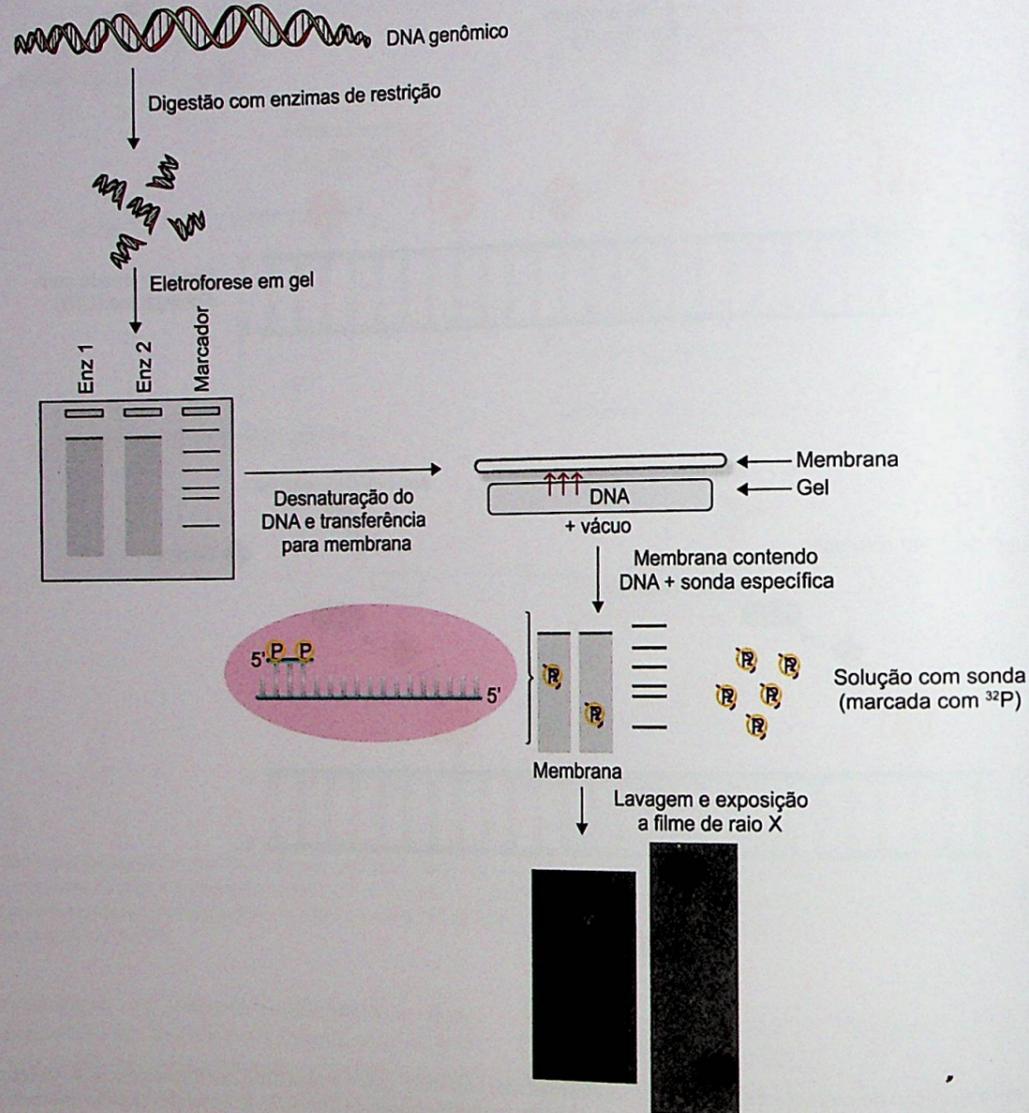
**A técnica de Northern blot detecta a presença de RNA de um gene de interesse**

Pouco tempo após o surgimento da técnica de *Southern blot*, em 1977, foram desenvolvidas modificações que possibilitaram a realização de estudos similares em amostras de mRNA extraído de células, eucarióticas ou procarióticas. A técnica, denominada *Northern blot*, permite *detectar, quantificar* e *estimar o tamanho* dos transcritos de um gene de interesse em uma mistura contendo todas as moléculas de RNA de uma célula, sendo muito usada em estudos de expressão gênica. Ela segue os mesmos princípios do *Southern blot*: a amostra de RNA também é fracionada por tamanho por meio de eletroforese em gel de agarose; no entanto, não há necessidade de tratamento prévio, uma vez que as moléculas de RNA são bem menores e de tamanho conveniente para separação em gel. Além disso, neste caso o gel de agarose deve conter um agente desnaturante, a fim de desfazer estruturas secundárias presentes nas moléculas de RNA. Após a hibridação com sonda específica para o gene de interesse, a quantidade de sonda retida em cada amostra é uma

medida da quantidade de transcritos do gene ali presentes e, portanto, dos níveis de expressão do gene. Para isso, deve-se utilizar um excesso de sonda, o que garante que a sua quantidade não será o fator limitante na hibridação, mas sim a quantidade de RNA complementar presente na amostra.

**Microarranjos de DNA monitoram a expressão de milhares de genes ao mesmo tempo**

A técnica de microarranjos de DNA também se utiliza do princípio de hibridação de ácidos nucleicos, e evoluiu a partir das técnicas de *Southern* e *Northern blot* para tornar possível a análise da expressão gênica global de uma célula. Assim, fragmente da expressão gênica global de uma célula. Para isso, utilizavam-se originalmente fragmentos de DNA obtidos a partir de uma biblioteca de cDNA do organismo de interesse; contudo, com o advento do sequenciamento de genomas em larga escala, passaram a ser utilizados fragmentos pequenos sintetizados *in vitro* com base

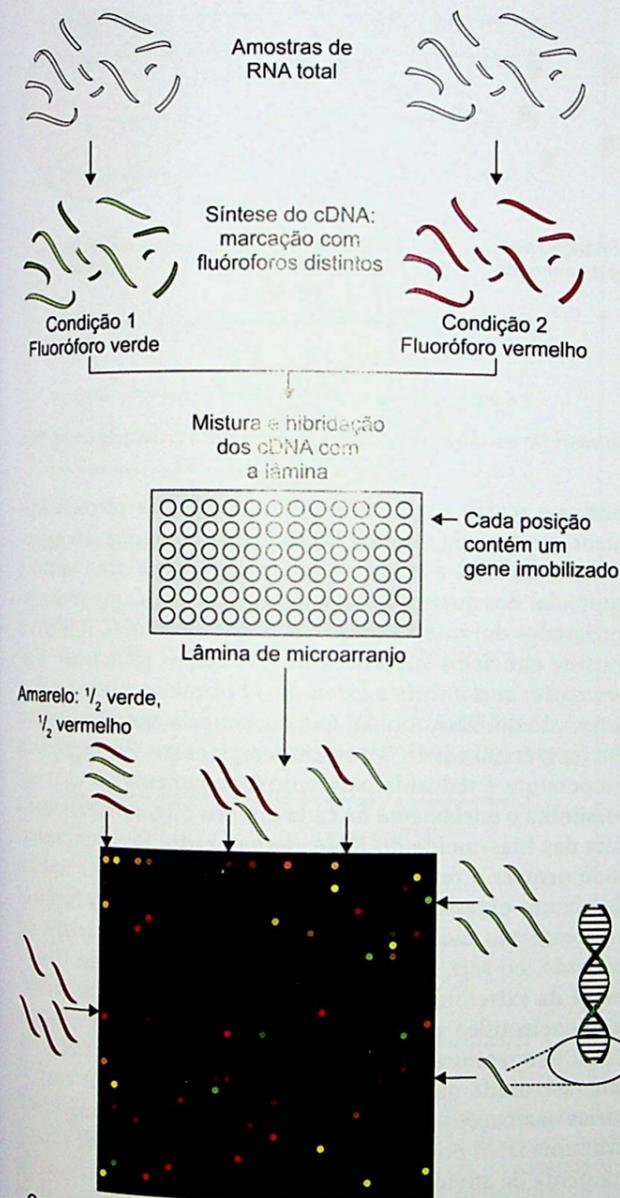


**Figura 11.12** Esquema da técnica de *Southern blot*. A técnica possibilita identificar um gene de interesse em uma mistura de fragmentos obtidos por digestão de um DNA genômico com uma ou mais enzimas de restrição. Quando diferentes enzimas são utilizadas, o gene ou parte dele é encontrado em fragmentos de tamanho distinto, como mostrado na figura.

na sequência conhecida dos genes do organismo. Amostras de cDNA total de células ou tecidos são marcadas com moléculas fluorescentes, os *fluoróforos*, e, assim, todas as moléculas de cDNA presentes na amostra funcionam como sondas, hibridando por complementariedade na posição correspondente na lâmina (Figura 11.13). A intensidade de fluorescência emitida em cada posição da lâmina representa o número total de moléculas de cDNA presentes e, conseqüentemente, o número de transcritos de cada um dos genes. Conforme será possível ver com mais detalhes no Capítulo 13, essa técnica é muito usada para comparar os níveis de expressão de todos os genes representados em uma lâmina entre duas condições diferentes, em um único experimento. Para isso, basta que as amostras marcadas com um fluoróforo diferente (Figura 11.13). Logo, mistura dos dois fluoróforos e representará, portanto, a quantidade relativa de cDNA marcados com os fluoróforos 1 e 2.

**Técnicas de hibridação são úteis para buscar genes em bibliotecas**

Neste ponto, você já sabe o que são e como são construídas bibliotecas de DNA e cDNA. Muitas vezes, desejamos isolar e clonar um gene de interesse de um determinado organismo. Isso pode ser feito se tivermos à disposição uma biblioteca de clones desse organismo e uma sonda correspondente a uma região do gene de interesse, usando-se a técnica de hibridação chamada de *colony blot* (Figura 11.14). Por ela, colônias bacterianas representando vários clones independentes da biblioteca são dispostas em uma placa de petri contendo meio de cultura apropriado. Após o crescimento, elas são transferidas para uma membrana de hibridação, produzindo uma réplica idêntica à placa. Essa transferência é feita por contato direto com a placa. Em seguida, a membrana é submetida a uma série de tratamentos que lisam as células, desnaturam e fixam o DNA. A membrana é então incubada com a sonda para o gene de estudo. O local onde ocorre hibridação



**Figura 11.13** Representação da técnica de microarranjos de DNA para a comparação do padrão global de expressão gênica entre duas condições diferentes. A coloração amarela indica que uma mesma quantidade de moléculas marcadas com fluoróforo vermelho e verde deu origem a híbridos na posição.

entre a sonda e o alvo corresponde a um clone da biblioteca que contém um inserto com parte ou todo o gene de interesse. Uma vez que a membrana é uma réplica perfeita da placa original, é fácil identificar a colônia de bactéria que contém o clone desejado (Figura 11.14).

**Experimentos de hibridação podem ser usados para localizar genes em cromossomos**

A técnica de *Southern blot* depende da extração do DNA total de células ou tecidos e sua digestão com enzimas de restrição, não sendo capaz de fornecer informações sobre a localização de genes no cromossomo, muito importante para estudos

citogenéticos, por exemplo. Em 1969, antes até do desenvolvimento da técnica de *Southern blot*, Joseph Gall e Mary Lou Pardue descreveram uma técnica que usava a hibridação para determinar a posição de uma sequência de DNA de interesse no cromossomo de células *in situ*; ou seja, no seu contexto celular. Na época, a técnica se utilizava de sondas radioativas, o que dificultava bastante a sua utilização. Posteriormente, em 1977, as sondas radioativas foram substituídas por marcadores fluorescentes, cuja detecção tem como base os mesmos princípios já descritos na Figura 11.11. A técnica foi então chamada de *FISH*, do termo em inglês *fluorescence in situ hybridization*. Na hibridação por *FISH*, as células ou tecidos são fixados a uma lâmina e tratadas com agentes que desnaturam o DNA sem alterar a estrutura dos cromossomos. Após a hibridação com a sonda e a lavagem do material, a localização da fluorescência emitida pela sonda é detectada em microscópio apropriado. Usando-se sondas marcadas com fluoróforos diferentes, é possível localizar vários genes de interesse simultaneamente. A técnica de *FISH* foi fundamental para os primeiros estudos de mapeamento de genes em cromossomos humanos e, atualmente, é muito usada no diagnóstico clínico de doenças genéticas associadas a rearranjos cromossômicos. Também tem sido muito utilizada para a localização de transcritos de RNA no interior da célula. Nesse caso, a hibridação ocorre sem uma etapa de desnaturação, o que garante que o DNA celular não será hibridado com a sonda, somente o RNA.

**Reação em cadeia da polimerase (PCR)**

No final da década de 1980, Mullis e Faloona desenvolveram uma técnica que revolucionou e impulsionou a biologia molecular: a PCR (do inglês, *polymerase chain reaction*), ou *reação em cadeia da polimerase*. Tal técnica possibilita a obtenção de milhares de cópias de um gene ou sequência de interesse a partir de poucas ou apenas uma molécula de DNA molde (o que chamamos de *amplificação do DNA*), em uma reação simples realizada *in vitro*. O boxe "Como a PCR surgiu e dominou a biologia molecular" mostra a história da concepção dessa técnica.

A reação de PCR consiste, portanto, na síntese de novas moléculas de DNA, mediada por uma enzima do tipo DNA polimerase. A síntese de DNA pelas DNA polimerases ocorre a partir de uma extremidade 3'-OH de uma pequena molécula iniciadora. Do mesmo modo, a reação de PCR também necessita de iniciadores, pequenos fragmentos de DNA complementares às extremidades da região a ser amplificada, chamados de *oligonucleotídeos*. Os oligonucleotídeos de qualquer sequência desejada podem ser facilmente sintetizados quimicamente, sendo também usados como sondas em experimentos de hibridação ou nas técnicas de sequenciamento de DNA – mostradas na próxima seção deste capítulo. Portanto, para amplificação de um gene de interesse por PCR, é preciso ter informações sobre a sequência da região a ser amplificada ou pelo menos das extremidades dessa região. Com o sequenciamento de genomas, é cada vez mais fácil obter informações sobre a sequência dos genes de interesse, e a técnica de PCR vem cada vez mais substituindo o uso de bibliotecas de DNA para o isolamento de genes.

O uso dos oligonucleotídeos iniciadores confere outra característica importante à PCR, a *especificidade*. O uso de dois iniciadores de tamanho relativamente pequeno (em torno de 20 pares de base cada um) costuma ser suficiente para garantir que a região a ser amplificada seja única, mesmo em genomas

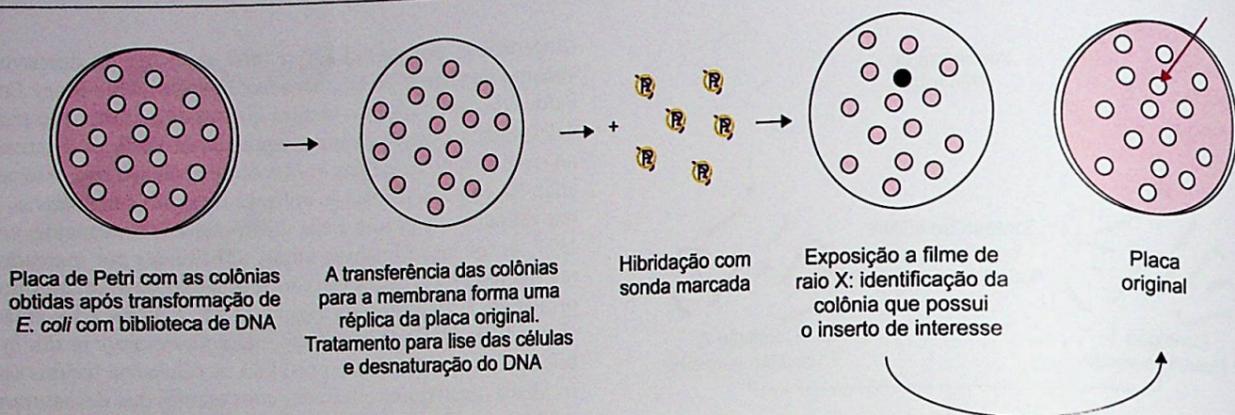


Figura 11.14 Esquema da identificação de uma colônia contendo um inserto de interesse em uma biblioteca de DNA usando a técnica de colony blot.

**Como a PCR surgiu e dominou a biologia molecular**

O primeiro trabalho que descreveu a técnica de PCR foi publicado na revista *Science*, em dezembro de 1985. O estudo original descreve a amplificação *in vitro* de um gene de cópia única de mamíferos, com o objetivo de desenvolver um método diagnóstico rápido e mais sensível para uma doença genética humana. O pesquisador responsável pela descoberta, o químico Kary Mullis, trabalhava na empresa de biotecnologia Cetus Corporation (Califórnia, EUA) e concebeu a técnica enquanto dirigia seu carro, em uma das longas viagens que fazia todos os fins de semana.

No entanto, quase 15 anos antes, o pesquisador indiano Har Gobind Khorana (Prêmio Nobel de Medicina em 1968 por seus estudos sobre o código genético) e seus colaboradores já haviam proposto uma ideia similar para síntese de genes de RNA transportador, sem, contudo, colocá-la em prática. A ideia de Khorana foi esquecida rapidamente, pois, na época, parecia inviável: não havia informações disponíveis sobre as sequências de genes, e a síntese química de oligonucleotídeos para serem usados como iniciadores era um processo extremamente complexo.

Da mesma maneira, a técnica de PCR (como foi proposta no trabalho original de 1985 e descrita em detalhes em um artigo posterior, publicado em 1987) era bastante trabalhosa e jamais teria sido tão impactante se não tivesse sido modificada, 3 anos depois, pelo uso da enzima termoestável Taq DNA polimerase. No início, utilizava-se como polimerase para a reação a enzima Klenow DNA polimerase I purificada de *E. coli*, que perdia a atividade a cada ciclo de desnaturação do DNA. Assim, era necessário adicionar mais enzima nos tubos de reação a cada novo ciclo, o que impedia a automatização do processo. Em 1988, Kary Mullis e seus colegas da Cetus aprimoraram a tecnologia com o uso da enzima purificada de *Thermus aquaticus*, Taq DNA polimerase, a mais usada até o momento. A bactéria *T. aquaticus* foi isolada e descrita por Thomas Brock em 1969, e é encontrada em nascentes termais a temperaturas de até 90°C. A enzima Taq DNA polimerase era bem mais eficiente que a Klenow e facilitou a automatização da técnica de PCR. Além disso, o uso da enzima termoestável possibilitou que os passos de anelamento e extensão fossem realizados a temperaturas mais elevadas, o que reduzia a amplificação de produtos inespecíficos. A partir daí, a técnica de PCR se popularizou definitivamente e, no fim dos anos 1980, tomou conta dos laboratórios de biologia molecular. O pesquisador Kary Mullis ganhou o Prêmio Nobel de Química em 1993 pela invenção da técnica de PCR.

complexos, como o humano. Isso torna comum a tarefa de isolamento de uma região do genoma. Fragmentos amplificados por PCR podem ser facilmente clonados em diversos vetores genéticos, ou usados como sondas em experimentos de hibridação.

**A reação de PCR usa uma DNA polimerase termoestável e amplifica o DNA exponencialmente**

A Figura 11.15 representa um esquema do funcionamento da técnica de PCR. Utiliza-se uma amostra de DNA molde, que

contém a região a ser amplificada, um par de oligonucleotídeos, sendo cada um deles complementar a uma das extremidades do alvo, a enzima DNA polimerase e uma mistura equimolar dos quatro desoxirribonucleotídeos, que serão incorporados durante a síntese das novas moléculas. A técnica consiste em ciclos sucessivos de três etapas principais: *desnaturação*, *anelamento* e *extensão*. O primeiro passo é a *desnaturação* do DNA molde, que ocorre pelo tratamento com alta temperatura (94°C) por um tempo curto. Em seguida, a temperatura é reduzida para aproximadamente 60°C, o que possibilita o *anelamento* de cada um dos oligonucleotídeos a uma das fitas-molde do DNA desnaturado. Eventualmente, pode ocorrer a renaturação do DNA molde, mas a utilização de um excesso de oligonucleotídeos na reação favorece o pareamento molde-oligonucleotídeo. A terceira etapa é a *extensão*, ou seja, a síntese das novas moléculas de DNA a partir da extremidade 3'-OH de cada um dos pares molde-oligonucleotídeo que foram formados. Essa etapa ocorre a 72°C, temperatura ideal para funcionamento da DNA polimerase usada na reação. Como a reação de PCR envolve várias incubações em temperatura de 94°C, é necessário utilizar uma DNA polimerase termoestável, ou seja, que não sofra perda de atividade em altas temperaturas. A enzima Taq DNA polimerase isolada da espécie de bactéria termofílica *Thermus aquaticus* passou a ser a mais usada para amplificação de DNA por PCR, por sua alta eficiência nas condições necessárias para a reação.

Como mostra a Figura 11.15, o primeiro ciclo da PCR dá origem a duas moléculas novas e longas, com a extremidade 5'P definida por cada um dos oligonucleotídeos e término da 3'-OH aleatório, que geralmente vai bem além do limite da região de interesse. No entanto, a cada novo ciclo de PCR, as novas moléculas sintetizadas se tornam moldes para a síntese de mais moléculas, de modo que, a partir do segundo ciclo da PCR, são obtidos apenas fragmentos com extremidades delimitadas pelos dois oligonucleotídeos usados na reação (Figura 11.15). Um total de aproximadamente 30 ciclos é realizado em uma reação de PCR padrão. No início da reação, quando os reagentes estão presentes em excesso, cada ciclo de amplificação produz o dobro de moléculas presentes no ciclo anterior, e a amplificação ocorre em escala exponencial. Em seguida, os reagentes disponíveis se tornam limitantes e a taxa de amplificação começa a reduzir, até que o processo chegue à saturação.

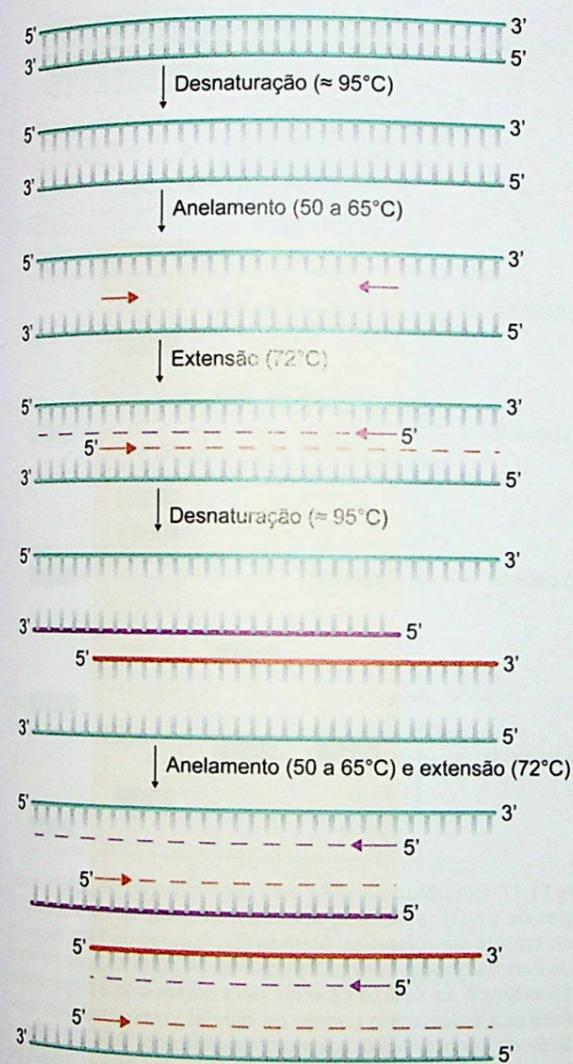


Figura 11.15 Representação dos dois primeiros ciclos de amplificação de DNA pela técnica de PCR. O primeiro ciclo dá origem a dois fragmentos, com extremidade 5' definida pelos oligonucleotídeos e região 3' indefinida. A partir do segundo ciclo, já começam a ser obtidos os fragmentos de tamanho delimitado pelos oligonucleotídeos nas duas extremidades.

**Técnica de PCR**

Agora que você sabe o funcionamento da técnica de PCR, sua simplicidade e eficiência, fica mais fácil compreender como ela rapidamente tomou conta dos laboratórios de biologia molecular, substituindo, em muitos casos, o uso de técnicas mais complexas e trabalhosas, como construção de bibliotecas de DNA e *Southern blot*. Com ela, ficou muito mais fácil isolar um gene de sequência conhecida em diversos organismos. Além disso, como mostra a próxima seção, a técnica possibilitou o desenvolvimento das estratégias de sequenciamento em larga escala. A PCR facilita também a introdução de sítios de enzimas de restrição nas extremidades dos fragmentos amplificados, bastando, para isso, que tais sítios sejam incluídos nas extremidades dos oligonucleotídeos usados para a amplificação. Assim, esses fragmentos podem ser facilmente clonados em vetores de interesse diversos, usando-se os sítios de restrição introduzidos por PCR. A PCR permitiu também a obtenção de genes contendo mutações de interesse, o que

chamamos de *mutagênese sítio-dirigida*. A mutação é inserida na sequência de oligonucleotídeos iniciadores, e os produtos da primeira amplificação passam a conter a mutação, servindo de molde para novos ciclos de amplificação.

Em razão de sua especificidade e da alta sensibilidade da técnica, que é capaz de amplificar amostras que contenham apenas uma molécula do alvo, a PCR é muito usada também no diagnóstico de doenças genéticas e no de doenças infecciosas (pela detecção de sequências de DNA de patógenos em amostras clínicas). Além disso, a PCR veio a contribuir bastante com estudos arqueológicos, possibilitando a análise de DNA obtido a partir de fragmentos de ossos e outros fósseis. A fim de elucidar como essa técnica impactou a identificação humana e a genética forense, ver o boxe "A identificação de indivíduos pela análise de polimorfismos".

**cDNA de interesse também podem ser obtidos por PCR**

A técnica de PCR também foi adaptada para a amplificação de DNA a partir de amostras de RNA, o que chamamos de RT-PCR (do inglês, *reverse transcriptase PCR*). A RT-PCR tem sido muito utilizada para a detecção de transcritos pouco abundantes (de difícil detecção por *Northern blot*), e também possibilita a construção de bibliotecas de cDNA quando apenas pequenas quantidades de mRNA estão disponíveis. No RT-PCR, é necessário adicionar uma etapa de síntese de cDNA a partir do RNA

**A identificação de indivíduos pela análise de polimorfismos**

A técnica de PCR também possibilitou a identificação de indivíduos a partir de amostras de baixa qualidade e/ou contendo quantidades mínimas de DNA, como traços de sêmen, saliva e até um fio de cabelo. Esse tipo de teste é especialmente importante na área forense, para identificação de criminosos a partir de amostras obtidas em cenas de crime.

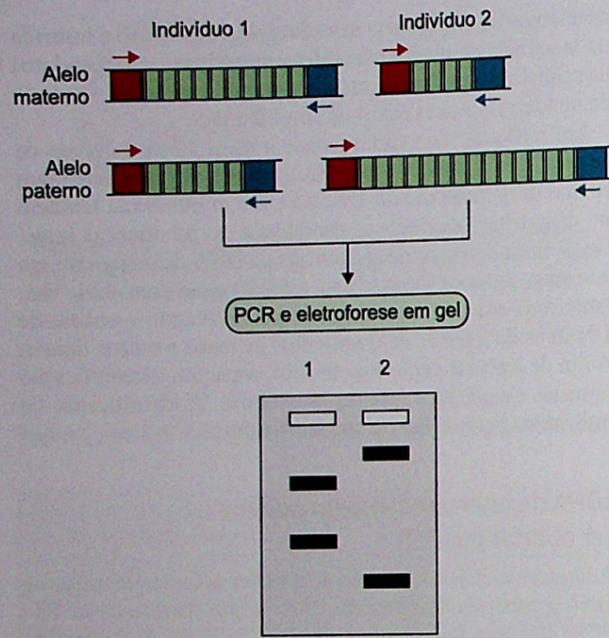
Mesmo antes do uso da PCR, a identificação de indivíduos já se baseava na análise de *polimorfismos* da sequência de DNA. Polimorfismos são variações na sequência de bases em um mesmo local do cromossomo entre os indivíduos de uma população; podem ser variações de uma ou poucas bases e que ocorrem a aproximadamente cada 1.000 pb no cromossomo, ou podem ser variações no tamanho de uma região contendo repetições de uma sequência pequena. Tais regiões são chamadas de *microsatélites*, e têm tamanho bastante variável de pessoa para pessoa. Assim, cada indivíduo tem o seu padrão de polimorfismo.

Muitas vezes, esses polimorfismos modificam o padrão de fragmentos de DNA obtido após digestão com uma ou mais enzimas. Essas variações no padrão de clivagem por enzimas de restrição são chamadas RFLP (do inglês, *restriction fragment length polymorphisms*). Assim, as técnicas tradicionais de identificação de indivíduos consistiam na análise por *Southern blot* do padrão de digestão obtido em amostras de suspeitos, com o padrão gerado pela amostra coletada na cena do crime. Pela análise dos perfis de digestão obtidos com um grupo de sondas diferentes, é possível identificar com alta precisão o indivíduo entre um grupo de suspeitos. No entanto, havia muita dificuldade em obter amostras de DNA de qualidade e quantidade suficiente para a análise por *Southern blot*. Por possibilitar a amplificação do DNA da amostra inicial para posterior análise por *Southern blot*, a PCR eliminou esse problema.

Alternativamente, os polimorfismos podem ser caracterizados diretamente por PCR, detectando-se diferenças no tamanho de fragmentos produzidos pela amplificação de regiões polimórficas. A Figura 11.16 mostra um exemplo de genotipagem de indivíduos usando essa técnica. Para uma identificação sem ambiguidade, vários *loci* polimórficos diferentes devem ser analisados.

Uma técnica similar é utilizada para testes de paternidade, conforme mostrado na Figura 11.17. Para isso, é obtida uma amostra de DNA do filho (geralmente de sangue, mas pode ser feito a partir de saliva ou fio de cabelo) e uma ou mais amostras dos possíveis pais e da mãe. Para cada *locus* analisado, um dos alelos do filho é proveniente da mãe e outro do pai, o que torna possível identificar os padrões paterno e materno.

INSTITUTO DE BIOTECNICAS - USP BIBLIOTECA



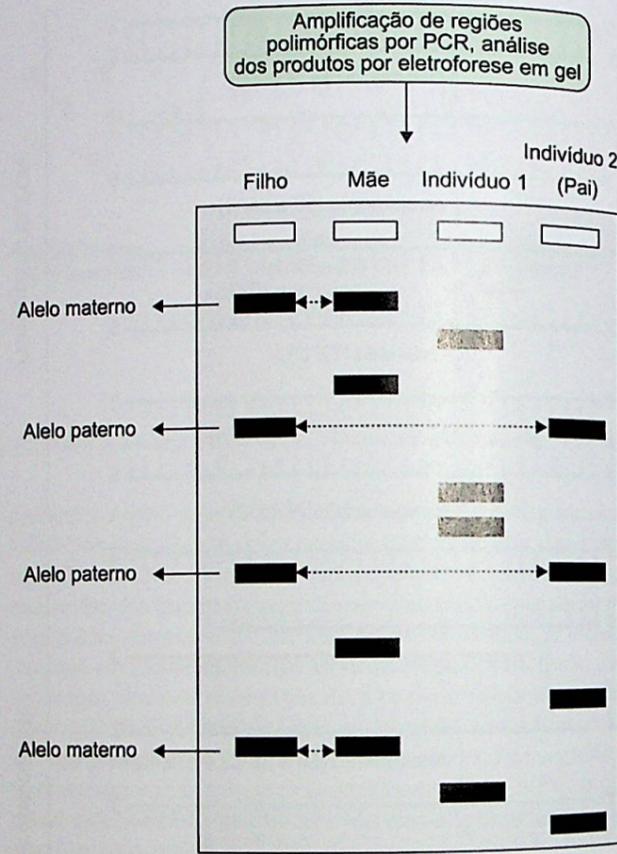
**Figura 11.16** Genotipagem humana utilizando PCR. Regiões polimórficas correspondendo a microssatélites (caixas verdes) são flanqueadas por regiões não variáveis (representadas em rosa e azul). Iniciadores de PCR (setas rosa e azul) podem ser utilizados para amplificar a região de microssatélites. O produto de amplificação de um indivíduo pode conter até duas bandas de tamanhos diferentes, uma correspondente ao alelo materno e outra ao alelo paterno.

inicial, uma vez que a Taq DNA polimerase não é capaz de sintetizar DNA a partir de RNA. A síntese de cDNA é realizada pelo uso de uma enzima do tipo *transcriptase reversa*, uma classe de proteínas provenientes de retrovírus. Estas enzimas são capazes de gerar uma fita simples de DNA complementar ao RNA molde (Figura 11.18). Em seguida, o cDNA produzido é utilizado como molde na reação de PCR padrão. É importante notar que a amostra de RNA não pode conter nenhuma contaminação com DNA celular; caso contrário, o DNA contaminante pode causar amplificação indesejável na reação de PCR. Possíveis contaminações com DNA são eliminadas pelo tratamento com enzimas do tipo desoxirribonucleases (DNase), que reconhecem e degradam exclusivamente moléculas de DNA.

A síntese do cDNA pode ser realizada utilizando-se um oligonucleotídeo específico para um transcrito de interesse, ou usando um oligonucleotídeo inespecífico, capaz de anelar em todas as moléculas de RNA disponíveis na amostra, o que vai gerar uma amostra de cDNA total celular (Figura 11.18). O oligonucleotídeo inespecífico mais utilizado é o *oligo(dT)*, que anela na cauda poliA de mRNA eucarióticos. Após a síntese do cDNA, a reação de PCR é feita com um par de oligonucleotídeos específico para o gene de interesse. Alternativamente ao uso de *oligo(dT)*, pode-se usar os *random hexamer*, que consistem em uma mistura de fragmentos de 6 nucleotídeos distribuídos de maneira aleatória, capazes de anelar em várias regiões distintas ao longo da molécula de RNA.

### Quantificando DNA e RNA | PCR em tempo real

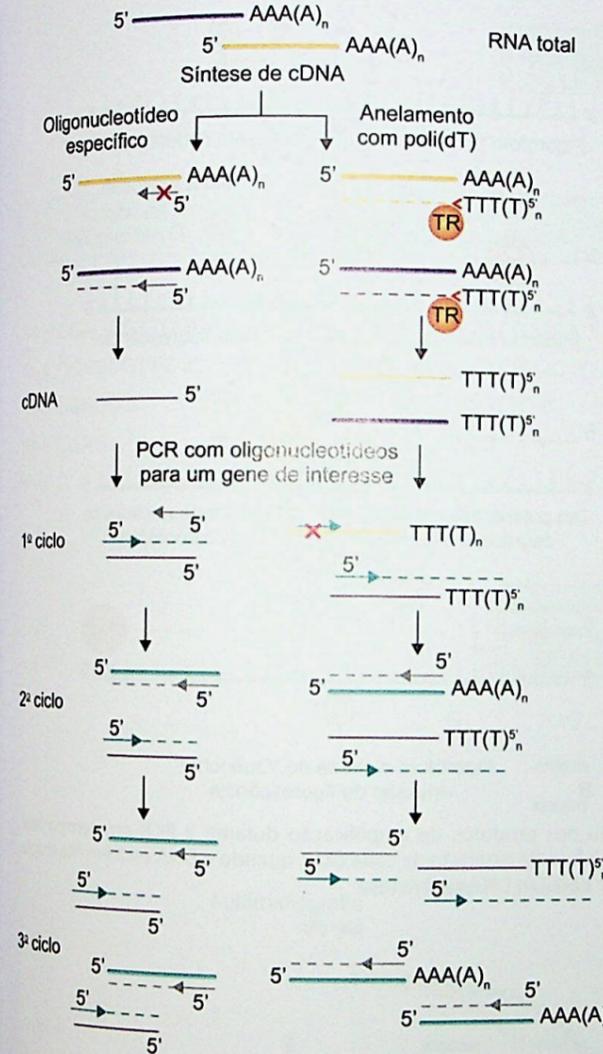
A técnica de PCR também pode ser usada para *quantificar* amostras de DNA e RNA. A quantificação de DNA por PCR é muito usada para a quantificação do número de cópias de genes em estudos de amplificação gênica e no diagnóstico clínico, como,



**Figura 11.17** Exemplo de análise para determinação de paternidade. Produtos de amplificação de regiões de microssatélites são comparados em corrida eletroforética. Metade das bandas do filho deve ser igual às bandas obtidas com o DNA da mãe. A outra metade deve ter correspondência no padrão paterno, para determinar a paternidade. O indivíduo 2 apresenta o padrão de bandas compatível com a paternidade da criança. Neste exemplo simplificado, dois *loci* diferentes foram analisados, mas um número bem maior de *loci* devem ser analisados na prática, para determinar com altos níveis de confiança estatística.

por exemplo, na determinação de carga viral em amostras de pacientes. No caso do RNA, esse tipo de análise é cada vez mais usado em estudos de expressão gênica nas quais se deseja verificar um ou poucos transcritos de interesse. No entanto, a alta sensibilidade da técnica de PCR dificulta bastante a realização de análises quantitativas, pois variações mínimas na eficiência entre amostras são maximizadas pela amplificação em escala exponencial dos produtos, levando a resultados incorretos. Assim, os primeiros métodos de quantificação por PCR com base na análise dos produtos finais da reação eram bastante imprecisos.

Em 1992, Higuchi *et al.* desenvolveram um equipamento de PCR que possibilitava detectar a fluorescência emitida pelo brometo de etídeo conforme essa molécula fluorescente ia sendo incorporada nas novas moléculas de DNA fita-dupla que iam se formando durante a amplificação: surgia o PCR em tempo real. Essa técnica é muito mais precisa e eficaz para quantificação por PCR, já que, em vez de detectar os produtos finais da reação, identifica os produtos de amplificação conforme estão sendo formados a cada novo ciclo. A tecnologia para detecção dos produtos de reação foi bastante aprimorada e, atualmente, utilizam-se marcadores fluorescentes bem mais sensíveis e menos tóxicos que o brometo de etídeo, e que também se ligam aos produtos durante sua síntese.



**Figura 11.18** Esquema da amplificação de um transcrito de interesse por RT-PCR. Após a síntese do cDNA específico (lado esquerdo da figura) ou do cDNA total (lado direito da figura) pela transcriptase reversa (TR), os cDNA servem de molde em uma reação de PCR padrão com oligonucleotídeos para o gene de interesse (indicado pelas setas verdes).

### PCR em tempo real usa marcadores fluorescentes para detectar a quantidade de DNA em cada ciclo de amplificação

Dois tipos de marcadores fluorescentes são os mais utilizados: os que intercalam na dupla-fita de DNA recém-formada após a extensão; e sondas fluorescentes, que hibridam em uma pequena região específica do DNA fita simples durante as fases de anelamento e extensão (Figura 11.19). Nenhum deles interfere na eficiência da reação de amplificação, característica essencial para sua funcionalidade.

A molécula SYBR Green® é o marcador intercalante mais utilizado atualmente e se liga a qualquer DNA fita dupla, o que causa aumento na sua fluorescência. Este tipo de marcação é inespecífico, ou seja, detecta qualquer DNA que esteja sendo amplificado na reação, podendo ser utilizado para qualquer gene de interesse. Uma desvantagem do uso desses marcadores é que eventuais produtos inespecíficos da reação também são detectados, interferindo na análise.

A marcação por sonda fluorescente detecta apenas o produto de interesse, não sofrendo a influência de produtos

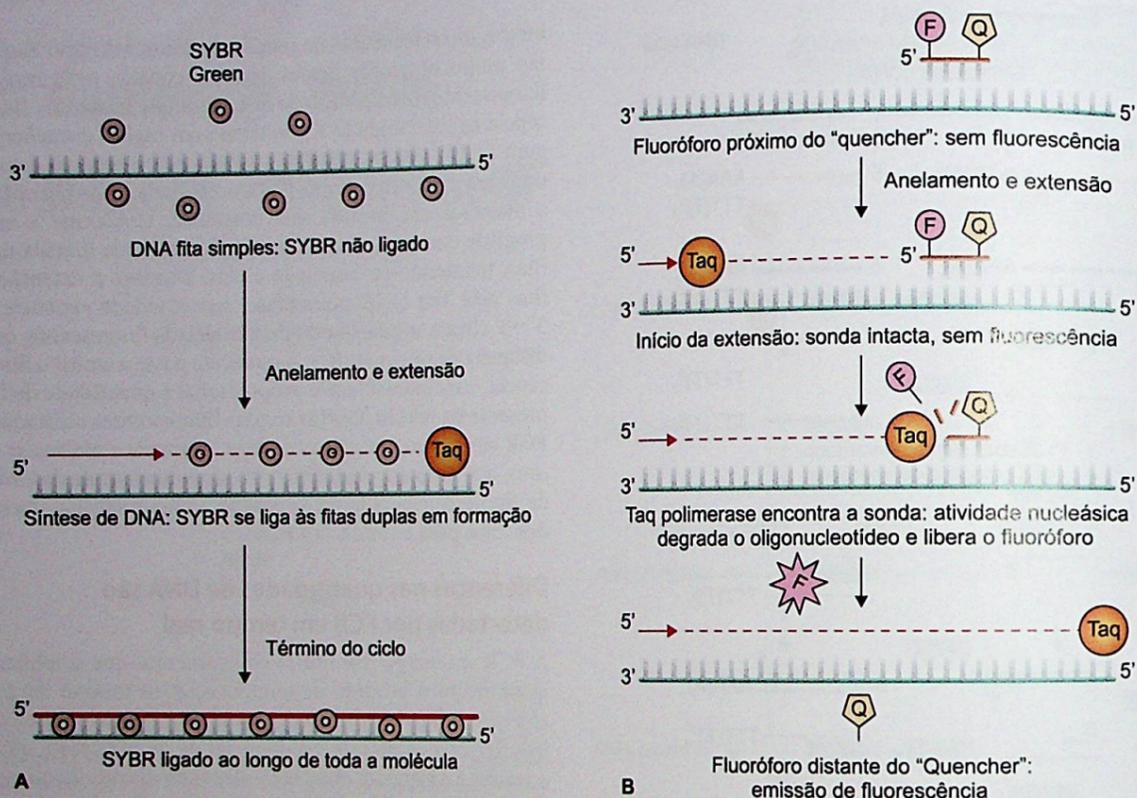
inespecíficos formados na reação. As sondas do tipo TaqMan® são muito utilizadas nesses casos, e contêm uma molécula fluorescente na extremidade 5' e outra que absorve a fluorescência na extremidade 3' (no termo em inglês, *quencher*) (Figura 11.19). Assim, quando a sonda está intacta, o *quencher* está bem próximo da fonte de fluorescência e não permite que a fluorescência emitida seja detectada. Conforme a reação progride e surgem novas fitas simples, a sonda hibrida nessas fitas, mantendo-se associada a elas. Durante a extensão das fitas pela Taq DNA polimerase, sua atividade exonucleásica 5'→3' cliva a sonda, liberando a molécula fluorescente, que se distancia de seu *quencher*. A molécula passa a emitir a fluorescência, cuja intensidade é proporcional à quantidade de DNA presente na reação. Outras sondas fluorescentes utilizadas em PCR em tempo real são as do tipo *Scorpions* e *Molecular Beacons*, que utilizam princípios similares de controle da emissão de fluorescência pela proximidade a uma molécula *quencher* descritos para a sonda TaqMan®.

### Diferenças nas quantidades de DNA são detectadas por PCR em tempo real

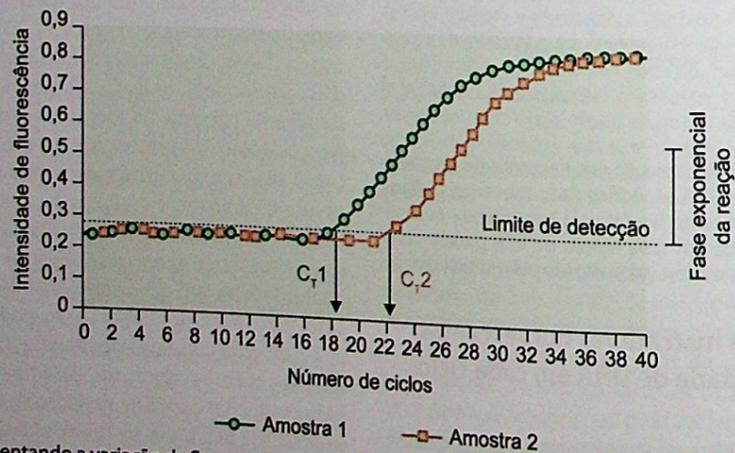
A PCR em tempo real utiliza equipamentos que amplificam e quantificam o produto de amplificação ao mesmo tempo. A alta sensibilidade do método possibilita que os produtos sejam detectados na fase exponencial de amplificação, quando a reação é altamente eficiente e depende apenas da quantidade inicial de molde presente na amostra. Nessa fase, quanto maior a quantidade de molde, menos ciclos de amplificação serão necessários para desenvolver uma quantidade detectável de produto. A quantificação se baseia na análise de um gráfico de amplificação que compara número de ciclos *versus* intensidade de fluorescência (Figura 11.20). O número de ciclos no qual a fluorescência detectada ultrapassa o limite mínimo de detecção é chamado de  $C_T$  (Figura 11.20). Quanto mais DNA molde está presente na reação, menos ciclos serão necessários para detecção dos produtos, e menor será o  $C_T$ . Para a quantificação absoluta do número de cópias de um gene, o valor de  $C_T$  obtido na amostra é comparado a uma curva-padrão de uma amostra de concentração conhecida. No entanto, muitas vezes, desejamos fazer uma quantificação relativa, que compara a quantidade de um gene entre duas condições diferentes. Nesse caso, não é necessária uma curva-padrão, e a comparação entre os  $C_T$  de cada condição demonstra a variação nas quantidades relativas entre as condições. Esse tipo de análise é muito usado para confirmar dados de variação de expressão de genes, como os obtidos em ensaios de microarranjos de DNA de estudos de transcriptoma – conforme mostra o Capítulo 13.

### Sequenciamento de DNA

A habilidade de determinar a sequência de nucleotídeos de moléculas de DNA é parte fundamental da biologia molecular moderna. Frederick Sanger e Walter Gilbert, idealizadores das primeiras técnicas de sequenciamento, foram laureados com o Prêmio Nobel de Química, em 1980, juntamente com Paul Berg, um dos pioneiros na tecnologia do DNA recombinante. Nesta seção, é possível observar que as técnicas de sequenciamento vêm sendo alvo de intensas pesquisas, e estão em constante desenvolvimento. Enquanto as primeiras inventadas podiam determinar a sequência de apenas dezenas ou, no máximo, centenas de bases de cada vez, as mais modernas podem determinar sequências de genomas inteiros em questão de dias.



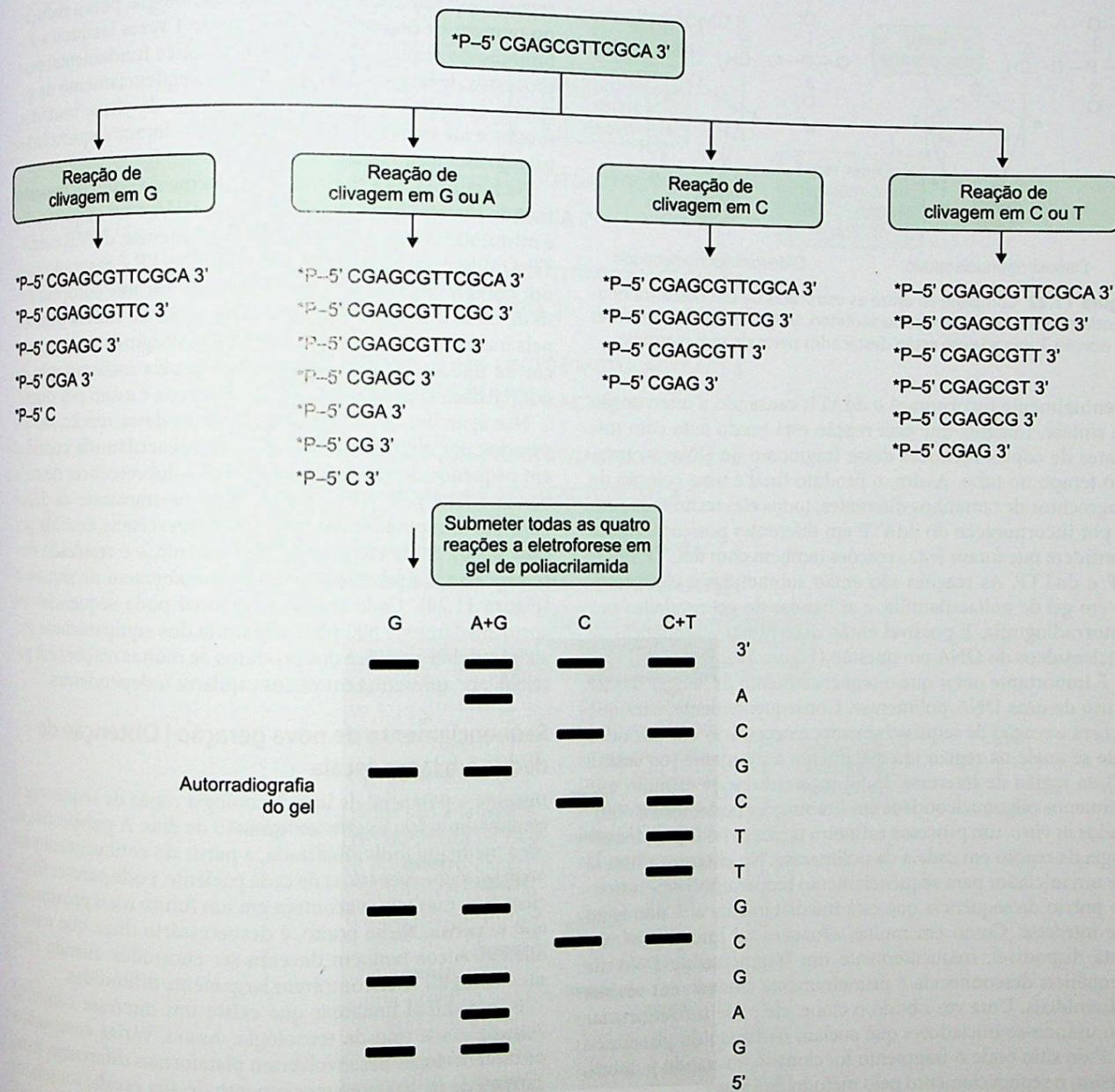
**Figura 11.19** Representação dos dois métodos mais usados para detecção dos produtos de amplificação durante a PCR em tempo real. A. Marcador inespecífico: SYBR Green®. A fluorescência é detectada logo após a fase de extensão de cada ciclo, quando o DNA está em fita dupla. B. Marcador específico: TaqMan®. A fluorescência é detectada após a extensão pela Taq DNA polimerase.



**Sequenciamento químico de Maxam e Gilbert**

Um dos primeiros métodos para sequenciamento foi a técnica desenvolvida por Maxam e Gilbert. Bastante laboriosa, requer o uso de reagentes extremamente tóxicos e, por isso, caiu em desuso, sendo largamente suplantada pelos métodos de terminação de cadeia. Contudo, seu valor histórico é enorme, e segue sendo usada em experimentos específicos. O sequenciamento de Maxam e Gilbert é conhecido como *sequenciamento químico*, e se baseia em quatro tratamentos químicos diferentes, cujo resultado final é a formação de quebras no DNA em localizações específicas. Um dos tratamentos promove uma clivagem a 5' de G, outro a 5' de G

ou A, outro a 5' de T, e outro a 5' de C ou T. Um exemplo de sequenciamento de um pequeno trecho de DNA por essa técnica é mostrado na Figura 11.21. O fragmento de DNA em fita simples a ser sequenciado é marcado na ponta 5' com fosfato radioativo, e então submetido aos quatro tratamentos separadamente, e os produtos de cada reação são analisados por eletroforese desnaturante em géis de poliacrilamida, o que possibilita a separação de fragmentos diferindo de apenas um nucleotídeo. A exposição autorradiográfica desses filmes torna possível a detecção do padrão de bandas e, a partir daí, a sequência do DNA pode ser facilmente deduzida, com a "leitura do gel".



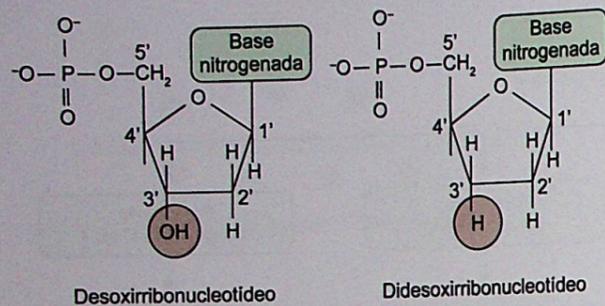
**Figura 11.21** Sequenciamento de um pequeno fragmento de DNA pelo método de Maxam e Gilbert.

**Sequenciamento enzimático de Sanger envolve reações de polimerização do DNA e terminação de cadeia por didesoxirribonucleotídeos**

O método de sequenciamento mais usado até o momento foi desenvolvido por Frederick Sanger. Assim como a tecnologia do DNA recombinante e a PCR, tal técnica também faz uso de enzimas purificadas – neste caso, DNA polimerases. Para compreender melhor como funcionam os métodos de terminação de cadeia, é preciso recordar a estrutura dos nucleotídeos e o modo de ação das DNA polimerases. Didesoxirribonucleotídeos são substratos para DNA polimerase, que requer extremidades 3'-OH livres para a síntese de DNA. A estrutura de um desoxirribonucleotídeo está representada na Figura 11.22. Durante o sequenciamento de DNA de Sanger, são adicionados didesoxirribonucleotídeos (ddNTP) às reações de síntese de DNA *in vitro*. Estes são diferentes

dos desoxirribonucleotídeos por apresentarem uma substituição do grupo 3'-OH por um hidrogênio (Figura 11.22). Logo, esses nucleotídeos podem ser utilizados normalmente pela DNA polimerase durante a síntese de DNA; contudo, uma vez incorporados, impedem que a síntese de DNA continue, por não terem grupo 3'-OH, essencial para a formação de uma ligação fosfodiéster subsequente.

O sequenciamento de Sanger original utiliza quatro reações diferentes de polimerização de DNA *in vitro*. Cada uma dessas reações contém um excesso de desoxirribonucleotídeos normais (dNTP) e um dos quatro didesoxirribonucleotídeos. Tome como exemplo a reação feita com dNTP + ddATP mostrada na Figura 11.23. Eles são misturados ao fragmento de DNA que se deseja sequenciar, juntamente com um iniciador marcado radioativamente na sua ponta 5', e uma DNA polimerase. A DNA polimerase começará a incorporar os nucleotídeos na síntese de uma fita complementar, mas



**Figura 11.22** Comparação entre as estruturas de um desoxirribonucleotídeo e um dideoxirribonucleotídeo. Os grupamentos presentes na porção 3' de cada um estão destacados pelos círculos marrons.

eventualmente incorporará o ddATP, causando a interrupção da síntese. Imagine que essa reação está sendo feita com milhares de cópias idênticas desse fragmento de DNA ao mesmo tempo no tubo. Assim, o produto final é uma coleção de fragmentos de tamanhos diferentes, todos eles terminados em A por incorporação de ddATP em diferentes posições. Agora considere que foram feitas reações também com ddCTP, ddGTP e ddTTP. As reações são então submetidas à eletroforese em gel de poliacrilamida, e as bandas do gel reveladas por autorradiografia. É possível então determinar a sequência de nucleotídeos do DNA em questão (Figura 11.23).

É importante notar que o sequenciamento de Sanger requer o uso de uma DNA polimerase. Conseqüentemente, para que ocorra a reação de sequenciamento, é necessário um iniciador que se anele na região imediatamente a montante (ou seja, a 5') da região de interesse. Já foi mostrado neste capítulo que pequenos oligonucleotídeos em fita simples podem ser sintetizados *in vitro*, um processo rotineiro também no uso da tecnologia da reação em cadeia da polimerase. No entanto, a síntese de um iniciador para sequenciamento requer um conhecimento prévio da sequência que está imediatamente a 5' da região de interesse. Como em muitas situações tal informação não está disponível, frequentemente um fragmento de DNA de sequência desconhecida é primeiramente clonado em vetores plasmidiais. Uma vez obtido o clone, ele pode ser sequenciado usando-se iniciadores que anelam na região do plasmídeo a 5' do sítio onde o fragmento foi clonado, tornando possível, assim, o sequenciamento pelo método de Sanger.

### O método de Sanger foi aprimorado com o uso de dideoxirribonucleotídeos fluorescentes

O método de Sanger foi adaptado para o uso de marcadores fluorescentes. Essa metodologia se tornou muito popular e é até hoje o método mais usado para sequenciamento de DNA. É interessante notar que o salto qualitativo dado com essa nova metodologia foi muito grande, sendo possível somente graças aos avanços em outros campos do conhecimento humano, como a engenharia e a informática, o que possibilitou a construção de máquinas e computadores poderosos nos quais este sequenciamento é feito.

O método fluorescente também se baseia em dideoxirribonucleotídeos, mas, neste caso, uma única reação é feita, em vez das quatro reações necessárias no método original de Sanger. Isso é possível porque cada um dos quatro dideoxirribonucleotídeos é conjugado a uma molécula fluorescente diferente, sem prejudicar seu uso pela DNA polimerase. Pode parecer uma mudança sutil, mas essa simplificação

representou um avanço enorme na tecnologia, pois a redução no número de reações por um fator de 4 vezes facilitou a automação do sequenciamento, característica fundamental para processos de larga escala, tais como o sequenciamento de genomas completos. Centenas de genomas de vírus, bactérias, fungos e até mesmo o genoma humano foram sequenciados usando esta tecnologia.

Na Figura 11.24 é mostrado o esquema do sequenciamento usando fluorescência. Nessa reação, o DNA a ser sequenciado é misturado com o iniciador, DNA polimerase, dNTP, e ddNTP. O diferencial é que cada um dos ddNTP é marcado com um corante fluorescente que emite luz em um comprimento de onda diferente. Assim, a terminação de cadeia causada pela incorporação de ddATP desenvolve fragmentos com uma cor de fluorescência; aquelas causadas pela incorporação de ddCTP dão origem fragmentos de outra cor, e assim por diante.

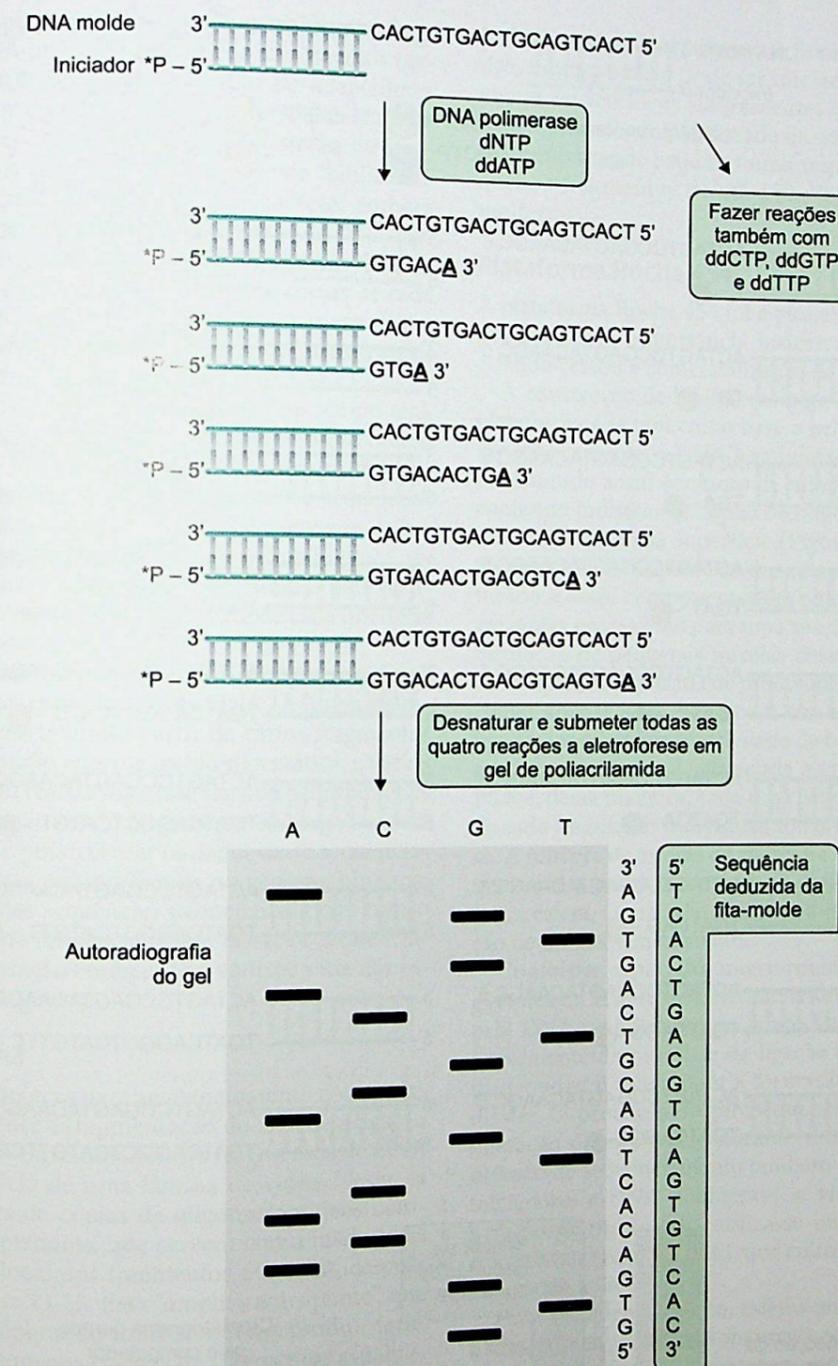
Nos aparelhos modernos, os produtos dessa reação são separados por eletroforese em géis de poliacrilamida contidos em pequenos capilares, e a detecção da fluorescência de cada banda é feita pelo próprio aparelho, que transmite os dados para um computador acoplado. A fluorescência emitida por cada banda é lida em picos de intensidade, e o resultado desse processo é a produção de um  *cromatograma de sequência* (Figura 11.24). Cada reação individual pode sequenciar até aproximadamente 800 pb, e a maioria dos equipamentos atuais possibilita a análise dos produtos de muitas reações em paralelo, por apresentarem vários capilares independentes.

### Sequenciamento de nova geração | Obtenção de dados em larga escala

Imagine o potencial de uma tecnologia capaz de sequenciar o genoma humano inteiro em questão de dias. A perspectiva de uma medicina individualizada, a partir do conhecimento das predisposições genéticas de cada paciente, pode parecer ficção científica, mas talvez aconteça em um futuro mais próximo do que se pensa. Neste ponto, é desnecessário dizer que muitos dilemas éticos também deverão ser encarados quando essas metodologias se encontrarem largamente difundidas.

Não é difícil imaginar que existe um interesse comercial enorme nesse tipo de tecnologia. Assim, várias companhias de biotecnologia desenvolveram plataformas diferentes, todas capazes de realizar sequenciamento de alta escala; ou seja, de vários gigabases de DNA por corrida do aparelho. Neste capítulo serão discutidas apenas algumas plataformas mais populares, embora existam outras em uso e/ou desenvolvimento.

É importante salientar que a complexidade dessas metodologias é enorme, mas serão explicadas aqui porque seu impacto no futuro da biologia e da medicina será, sem dúvida, avastoso. Antes de discutir cada uma delas, pode-se apontar o que elas têm em comum. O sequenciamento de nova geração (*next-generation sequencing*) é também chamado frequentemente de *sequenciamento de segunda geração* ou *sequenciamento maciço em paralelo*. Este último ilustra bem o princípio básico por trás dessas plataformas de próxima geração (ou simplesmente sequenciar várias moléculas diferentes (ou vários fragmentos diferentes provenientes do mesmo genoma) em paralelo. Logo, elas diferem fundamentalmente das metodologias de Sanger e de Maxam e Gilbert, em que uma única sequência de DNA, oriunda de um trecho específico, é obtida em cada reação. Isso é possível graças à integração de ferramentas de nanotecnologia, robótica e informática em aparelhos de alto desempenho.



**Figura 11.23** Sequenciamento de DNA pelo método de Sanger. Notar que, para simplificação, apenas a reação com ddATP foi mostrada, mas reações similares são feitas com ddCTP, ddGTP e ddTTP.

As diferentes plataformas apresentam limitações diferentes e, em geral, obtêm-se apenas a sequência de pequenos fragmentos de DNA, por limitações das técnicas que vão além do escopo deste texto. No entanto, é importante considerar a capacidade assombrosa dessas tecnologias: em seus modelos mais avançados, cada corrida desses aparelhos pode sequenciar dezenas de gigabases de DNA. Para ter uma ideia do que isso representa, vale relembrar os tamanhos dos diferentes genomas. Por exemplo: o genoma da bactéria *Escherichia coli* tem 4,6 Mpb de DNA, e o genoma humano tem aproximadamente 3 Gpb de DNA.

Em primeiro lugar, para compreender o modo como esses métodos de segunda geração mais usados funcionam, serão examinados os seus princípios fundamentais em comum:

1. Construção de uma *biblioteca* que represente o genoma em análise. Ao contrário das bibliotecas genômicas e de cDNA vistas anteriormente, ela não é produzida por clonagem em vetores. Essas bibliotecas são, na verdade, compostas de milhões de fragmentos de DNA diferentes, isolados fisicamente uns dos outros, e cada um deles multiplicado em várias cópias de pequenos agrupamentos (chamados *clusters*) de moléculas idênticas.

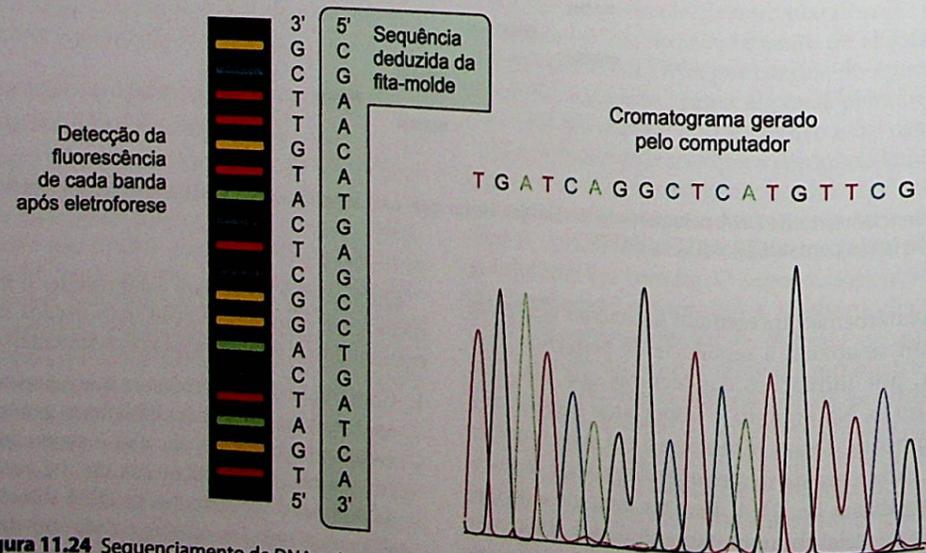
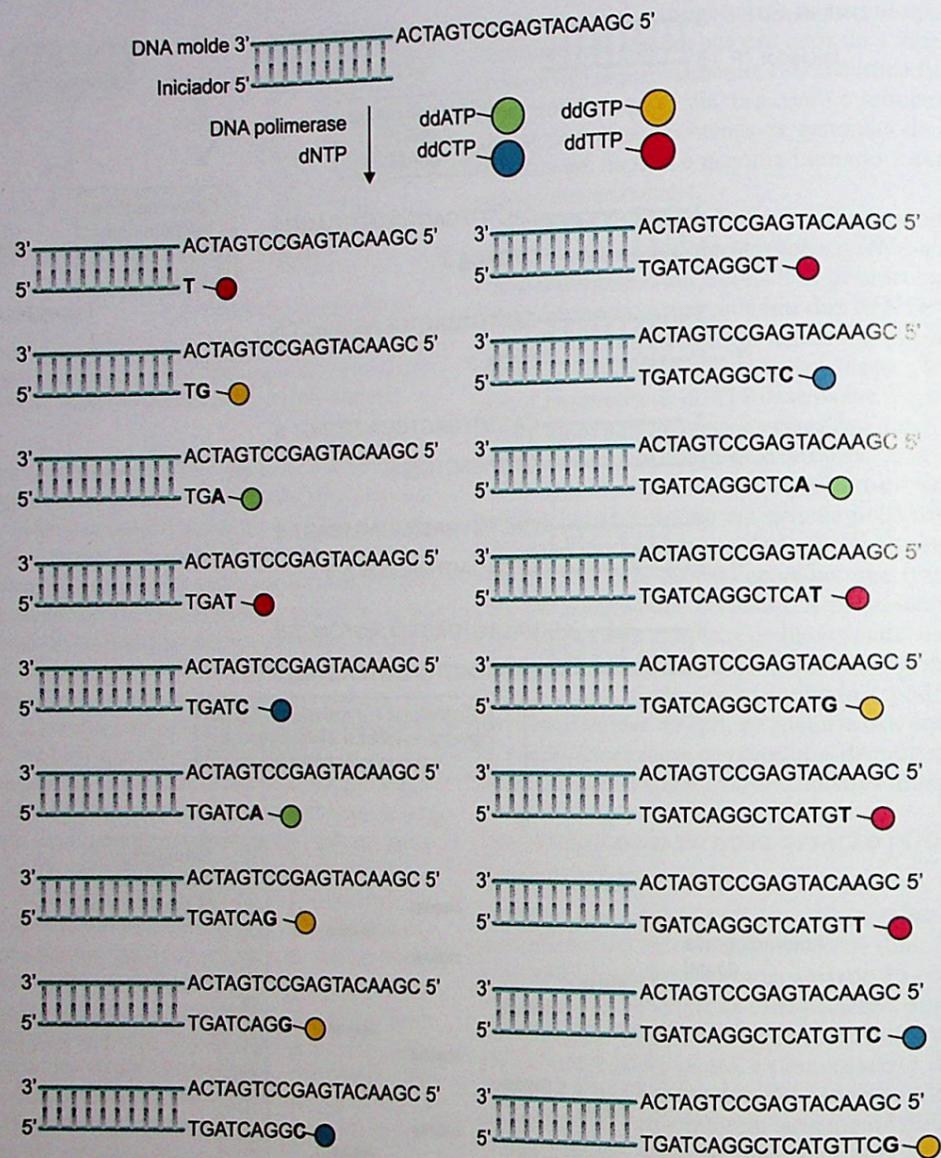


Figura 11.24 Sequenciamento de DNA pelo método de Sanger usando didesoxirribonucleotídeos fluorescentes.

2. Cada um desses pequenos *clusters* é construído por fragmentação do DNA original em pequenos fragmentos (geralmente por força mecânica) e ligação de adaptadores (pequenos trechos de DNA de sequência conhecida) nas duas extremidades dessas moléculas. A coleção de fragmentos ligados aos adaptadores é chamada de "biblioteca" no jargão do sequenciamento de nova geração, embora não represente uma biblioteca de fragmentos de DNA clonados em vetor. As sequências conhecidas desses adaptadores são usadas para a produção de várias cópias de cada molécula por PCR. A produção de várias cópias em *cluster* é fundamental, pois amplifica o sinal obtido pelas reações de sequenciamento subsequentes.
3. Detecção do processo de sequenciamento em tempo real. Ou seja, a cada ciclo de reação nesses aparelhos, se obtém informação sobre uma ou mais bases de cada um dos milhões de fragmentos de DNA que estão sendo sequenciados em paralelo. Essa detecção ocorre pela emissão de luz ou fluorescência por cada um dos *clusters* da biblioteca. As emissões de cada um dos *clusters* são registradas pelo computador a cada ciclo, obtendo a sequência de cada um deles ao final do processo.
4. Em todas as metodologias, o resultado final é a sequência de milhões de pequenos trechos de DNA. A tarefa de remontar um genoma inteiro a partir de vários fragmentos pequenos é um desafio enorme na bioinformática, e vários avanços vêm sendo realizados nesse sentido. Contudo, tais metodologias já então sendo largamente empregadas para várias aplicações. É possível usar os dados desse sequenciamento para detectar polimorfismos no genoma humano, por comparação das sequências pequenas obtidas com o genoma humano de referência. Além disso, os sequenciamentos de nova geração também têm enorme valia nos estudos de expressão gênica.

#### Plataforma Illumina

A construção de bibliotecas para sequenciamento com a plataforma Illumina requer a fragmentação do DNA e a ligação de adaptadores a suas extremidades. Os fragmentos de DNA são ligados à superfície de uma lâmina de vidro. Nesta, já estão ligadas milhões de cópias de oligonucleotídeos complementares aos adaptadores, que servem como iniciadores para a amplificação local dos fragmentos por PCR, conforme ilustrado na Figura 11.25. Essa "amplificação-ponte", que ocorre com DNA se dobrando sobre si mesmo, produz dezenas de milhões de pequenos *clusters* de fragmentos espalhados pela placa.

O processo de sequenciamento usa uma DNA polimerase e nucleotídeos modificados com duas características: primeiro, cada um dos quatro diferentes nucleotídeos é marcado com corantes fluorescentes diferentes. Além disso, esses nucleotídeos têm o terminal 3' bloqueado quimicamente, e tanto a fluorescência quanto o bloqueio 3' podem ser revertidos por tratamentos químicos. O sequenciamento nessa plataforma envolve a adição dos nucleotídeos marcados e bloqueados, que são então incorporados pela DNA polimerase. Como seu terminal 3' está bloqueado, apenas uma incorporação é possível de cada vez para cada *cluster*. A fluorescência emitida por cada *cluster* é lida pelo sistema óptico acoplado e registrada no computador. Na sequência, uma etapa de tratamento químico remove tanto o bloqueio

3' dos nucleotídeos, quanto a fluorescência conjugada. Assim, um novo ciclo pode ser iniciado, adicionando-se novamente nucleotídeos fluorescentes bloqueados na porção 3'. Esta metodologia foi batizada de *sequenciamento por síntese*. A Illumina pode produzir uma sequência de milhões de pequenos fragmentos de até 150 pb em seus aparelhos mais modernos.

#### Plataforma Roche 454

A plataforma Roche 454 foi a pioneira em termos comerciais, portanto, tem importância histórica. Foi utilizada até 2016, quando cessou a comercialização da plataforma.

A construção de bibliotecas para o sequenciamento com a plataforma 454 tem como base o princípio de PCR em emulsão. Basicamente, o DNA fragmentado é ligado a adaptadores e misturado a um conjunto de minúsculas esferas de agarose contendo milhares de cópias de oligonucleotídeos ligados covalentemente a sua superfície (Figura 11.26). Estes são complementares às sequências dos adaptadores. Um óleo é misturado a esses componentes em solução aquosa contendo os reagentes necessários para uma reação de PCR, resultando na formação de pequenas micelas desse óleo. A estequiometria dos reagentes é ajustada de modo que a maioria das micelas irá conter apenas uma esfera com um único fragmento de DNA ligado por complementariedade de bases aos iniciadores. Essa mistura de micelas é submetida a ciclos de amplificação por PCR e, dessa maneira, cada uma produzirá milhares de cópias de cada fragmento individual, todas ligadas à esfera de agarose. A mistura de esferas de agarose é colocada em microplacas que contêm poros de tamanho suficiente para acomodar uma única esfera. O material está pronto para ser submetido à reação de pirosequenciamento.

Conforme explicado anteriormente, os nucleotídeos precursores da síntese de DNA são trifosfatados. Quando usados pela DNA polimerase, um *pirofosfato* é liberado, e a energia proveniente da hidrólise da ligação deste com o restante do nucleotídeo é usada para a formação da nova ligação fosfodiéster. O princípio do pirosequenciamento é a detecção da liberação de pirofosfato durante síntese de DNA. Logo, esse método de sequenciamento também envolve uma reação com iniciadores e DNA polimerase *in vitro*. Para detecção da liberação de pirofosfato, utiliza-se uma sequência de reações enzimáticas (Figura 11.26), que culmina com a emissão de luz pela enzima luciferase.

Para o sequenciamento, esferas ainda menores são adicionadas, contendo todas as enzimas necessárias ao pirosequenciamento ligadas. Soluções com os diferentes nucleotídeos são adicionadas sequencialmente. Imagine uma placa com milhares de microporos, cada um deles contendo um único *cluster* de moléculas de DNA idênticas. Ao adicionar dATP, por exemplo, naqueles fragmentos propriamente anelados com um iniciador e cuja próxima base é T, a DNA polimerase irá incorporar dATP, liberando pirofosfato e culminando com a emissão de luz. Assim, um leitor óptico acoplado a um computador registra todos os microporos nos quais houve emissão de luz. Ao realizar diversas vezes os ciclos com todos os nucleotídeos, é possível determinar sequências de bases dos fragmentos contidos em cada microporo. Assim, as sequências de milhões de diferentes fragmentos pequenos são geradas em paralelo. A plataforma 454 pode sequenciar fragmentos de até 700 pb.

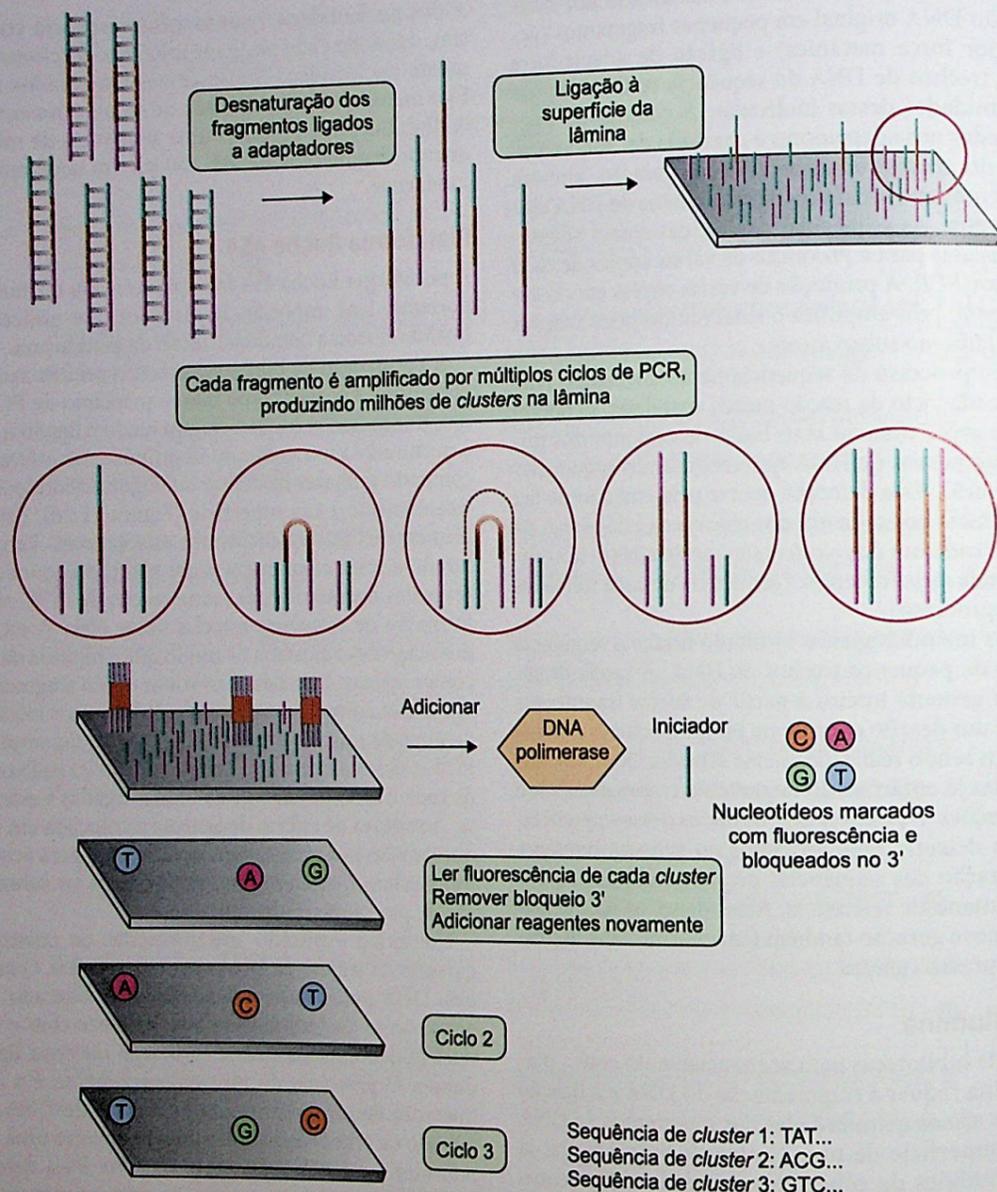


Figura 11.25 Sequenciamento de DNA na plataforma Illumina. Apenas três clusters são mostrados para simplificação.

### Plataforma Ion Torrent (Thermo)

A plataforma Ion Torrent é, em muitos aspectos, similar à plataforma 454 da Roche. Assim como na metodologia 454, PCR em emulsão é usada para a amplificação das bibliotecas, e as esferas com milhares de cópias de cada fragmento são separadas uma das outras em pequenos poços. A metodologia de sequenciamento também envolve a detecção de produtos oriundos da atividade da DNA polimerase. A incorporação de um nucleotídeo à cadeia nascente a partir do dNTP libera não só o pirofosfato, mas também um íon H<sup>+</sup>. A detecção da liberação de H<sup>+</sup> em cada rodada é a base desse sequenciamento. Assim, se um nucleotídeo é adicionado por vez, a incorporação do mesmo em determinado poço é detectada pela liberação de H<sup>+</sup>, uma vez que o chip onde o sequenciamento ocorre funciona como um semicondutor, transformando o sinal detectado em voltagem, registrada pelo aparelho. O processo é repetido diversas vezes, utilizando-se cada um dos

quatro dNTP sequencialmente. Assim como na plataforma 454, o sequenciamento de regiões homopoliméricas (repetições de um nucleotídeo, como a sequência 5'-GGGG-3', por exemplo) apresenta maior taxa de erro.

### Considerações finais e perspectivas

As técnicas básicas de análise e manipulação de ácidos nucleicos descritas neste capítulo permitiram um enorme avanço na compreensão da genética e da biologia dos organismos vivos, sendo o arcabouço para o desenvolvimento de novas metodologias de alta resolução e larga escala que estão surgindo muito rapidamente. Lembre-se, por exemplo, de que a técnica de PCR foi usada pela primeira vez em 1985 e, em apenas 30 anos, passou a ser aplicada em metodologias como o sequenciamento rápido de gigabases de DNA e a quantificação de ácidos nucleicos em alta sensibilidade e tempo real em praticamente qualquer amostra biológica de interesse. Hoje

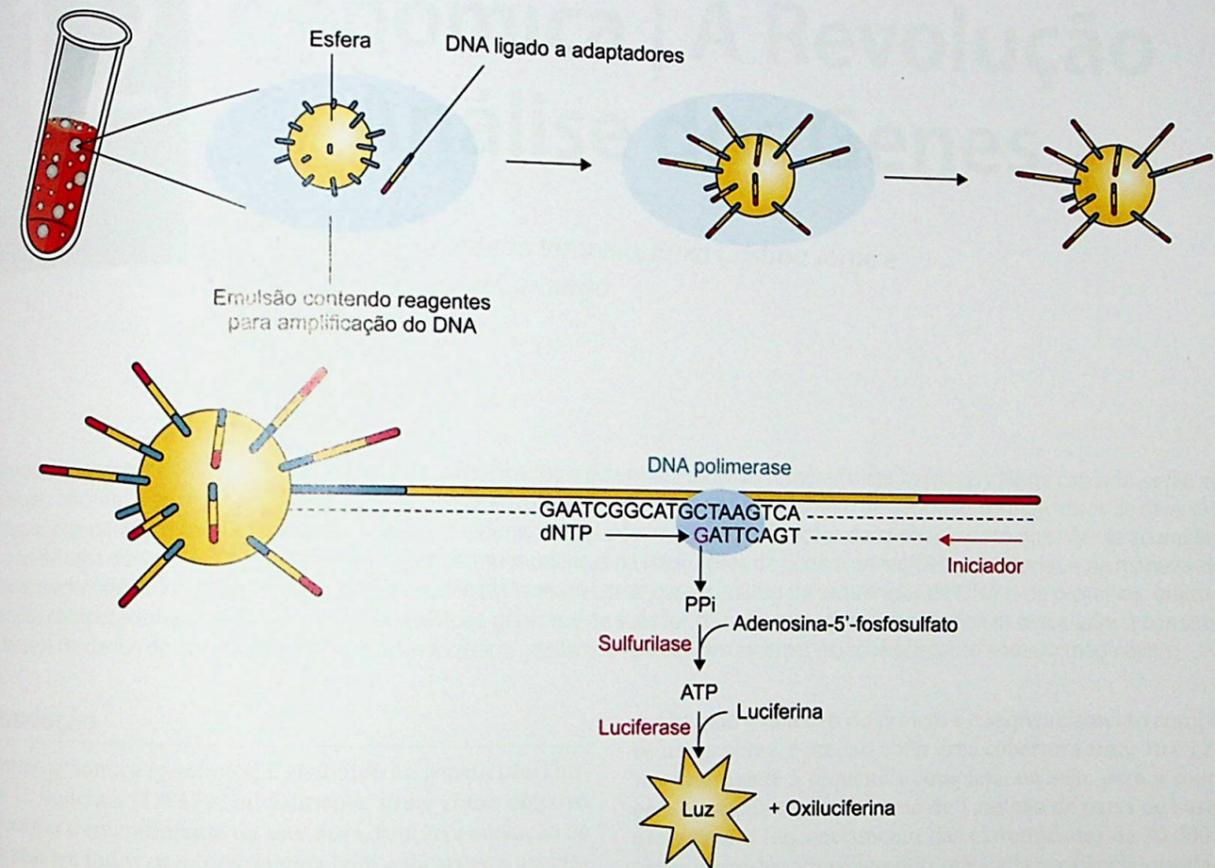


Figura 11.26 Sequenciamento de DNA na plataforma 454. Na parte superior, é mostrado um esquema de produção das esferas contendo um cluster de fragmentos idênticos. Na parte inferior, a detecção de cada ciclo de pirosequenciamento é detalhada.

é possível, portanto, determinar em poucos dias a sequência completa de um genoma humano, abrindo novas perspectivas para a medicina baseada nas características genéticas de cada indivíduo. Da mesma forma, é possível determinar todo o conjunto de microrganismos presente em um ambiente (microbioma) a partir do sequenciamento de todo o DNA ali encontrado.

Novas áreas de estudo também se desenvolveram em função dessas tecnologias nos últimos anos, como a Biologia Sintética, que busca sintetizar genes e até criar genomas para introduzir características de interesse. Agora é possível, por exemplo, sintetizar genes inteiros ligados a vetores de interesse, assim como obter versões com mutações alterando individualmente cada um dos aminoácidos da proteína por ele codificada.

Independentemente das novas metodologias que surgirão, as técnicas descritas neste capítulo são as mais usadas para análise e manipulação de ácidos nucleicos. A compreensão dos princípios básicos dessas técnicas é fundamental para estudos em diversas áreas da Biologia moderna.

### Bibliografia

- Brown TA. Gene cloning and DNA analysis: an introduction. 6. ed. Oxford: Blackwell Scientific Publishers; 2010.
- Hutchison CA 3rd. DNA sequencing: bench to bedside and beyond. Nucleic Acids Res. 2007; 35(18):6227-37.
- Maxam AM, Gilbert W. A new method for sequencing DNA. Proceedings of the National Academy of Sciences of the United States of America. 1977;74(2):560-4.
- Mullis KB. The unusual origin of the polymerase chain reaction. Sci Am. 1990;262(4):56-61,64-5.
- Mullis KB, Faloona FA. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Meth Enzymol. 1987;155:335-50.
- Roberts RJ. How restriction enzymes became the workhorses of molecular biology. Proc Natl Acad Sci U S A. 2005;102(17):5905-8.
- Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science. 1985;230(4732):1350-4.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America. 1977;74(12):5463-7.