

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

◆ **TECNOLOGIA DO DNA RECOMBINANTE** ◆

Alexandra A. C. Nascimento
Enilza Maria Espreafico
Maria Luisa Paçó Larson
Nádia Monesi
Nilce Maria Martinez Rossi
Vanderlei Rodrigues

Revisão 1: Eliana Valéria Patussi
Márcia A. S. Graminha

Revisão 2: Fábio Márcio Squina
Josane de Freitas Sousa
Roberto Ruller
Valeria Valente

•2003•

I. CONCEITOS BÁSICOS

1. INTRODUÇÃO

Até a década de 70, o DNA era o componente celular mais difícil de ser analisado. Sua sequência de nucleotídeos de enorme tamanho e monotonia química era geralmente analisada através de meios indiretos como a sequência de proteínas e análise genética. A partir da década de 70 novas tecnologias foram desenvolvidas permitindo o isolamento e a purificação de genes específicos num processo chamado de clonagem gênica. Na verdade, muitas destas técnicas são provenientes da Microbiologia, Bioquímica, Imunologia e Genética Microbiana e permitiram que a análise do DNA ganhasse um novo enfoque. O DNA tornou-se então, a molécula mais fácil de ser analisada, sendo possível isolar regiões específicas, obtê-las em grande quantidade e determinar a sua sequência numa velocidade de milhares de nucleotídeos por dia.

A Tecnologia do DNA recombinante, como se convencionou denominar este conjunto de técnicas, tem uma ampla aplicação. Ela pode ser usada para estudar mecanismos de replicação e expressão gênica, na determinação da sequência de um gene e conseqüentemente da proteína que ele codifica, ou no desenvolvimento de culturas microbianas capazes de produzir substâncias úteis tais como a insulina humana, hormônio de crescimento, vacinas e enzimas industriais em grandes quantidades. Sua aplicação comercial ou biotecnológica parece ter um potencial inesgotável.

Como conseqüência do desenvolvimento desta tecnologia é atualmente possível realizar investigação de paternidade e o diagnóstico de doenças genéticas e infecciosas através da análise de DNA.

2. CONCEITO DE CLONAGEM MOLECULAR

A origem do termo clonagem vem da Genética Bacteriana que considera uma colônia de bactérias como um clone porque todos os indivíduos são geneticamente idênticos à bactéria inicial.

A técnica central da metodologia do DNA recombinante é a clonagem molecular, a qual consiste no isolamento e propagação de moléculas de DNA idênticas. A clonagem molecular compreende pelo menos dois estágios importantes. Primeiro, o fragmento do

DNA de interesse chamado de *inserto* é ligado a uma outra molécula de DNA chamada de *vetor* para formar o que se chama de *DNA recombinante*. Segundo, a molécula do DNA recombinante é introduzida numa célula hospedeira compatível, num processo chamado de *transformação*. A célula hospedeira que adquiriu a molécula do DNA recombinante é agora chamada de *transformante* ou *célula transformada*.

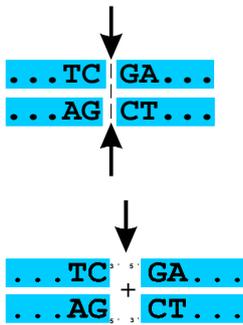
Um único transformante, em condições ideais, sofre muitos ciclos de divisão celular, produzindo uma colônia que contém milhares de cópias do DNA recombinante.

3. ENZIMAS DE RESTRIÇÃO

Em 1953 foi descrito um estranho fenômeno no qual a eficiência da replicação de um bacteriófago (vírus de bactérias) dependia da célula hospedeira na qual ele estava inserido. Algum tempo depois percebeu-se que a inabilidade de certos fagos crescerem em determinadas linhagens bacterianas era devido a presença de nucleases altamente específicas que clivavam o seu DNA. Isto pode ser encarado como um sistema de defesa bacteriano que degrada DNA que lhe é estranho (*restrição*). A bactéria protege seu próprio DNA desta degradação “camuflando-o” através da metilação de algumas bases específicas (*modificação*). Como conseqüência, este sistema é frequentemente descrito como fenômeno da restrição/modificação e existe em um grande número de bactérias.

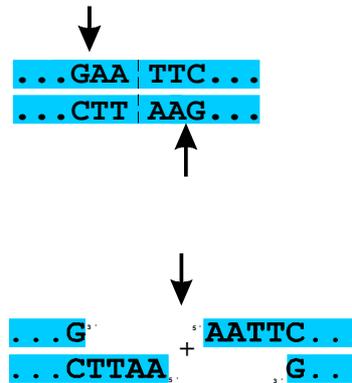
As enzimas de restrição ou endonucleases de restrição são divididas em várias classes, dependendo da estrutura, da atividade e dos sítios de reconhecimento e clivagem. As enzimas do Tipo II, as mais importantes na Tecnologia do DNA Recombinante, são proteínas monoméricas ou diméricas e clivam o DNA no mesmo sítio do seu reconhecimento. O sítio de reconhecimento deste tipo de enzima é normalmente uma seqüência *palindrômica*, isto é, ela tem um eixo de simetria e a seqüência de bases de uma fita é a mesma da fita complementar, quando lida na direção oposta (Figura 1).

a) Clivagem no eixo de simetria



Moléculas com extremidades abrupas

b) Clivagem simetricamente situada ao redor do eixo de simetria



Moléculas com extremidades coesiva

Figura 1. Os dois tipos de clivagem feitos por enzimas de restrição. As setas indicam os sítios de clivagem e as linhas pontilhadas representam o centro de simetria da sequência.

Estas enzimas reconhecem seqüências específicas de 4 a 8 pares de base (pb) na molécula de DNA e fazem dois cortes, um em cada fita. Há 2 tipos distintos de clivagens: a) os dois cortes ocorrem no eixo de simetria da seqüência específica, gerando extremidades abrupas, ou b) os cortes são feitos simetricamente, porém, fora do eixo de simetria, gerando extremidades coesivas. Estes dois tipos de clivagens e suas conseqüências estão mostrados na Figura 1.

Atualmente, mais de 1000 enzimas de restrição já foram identificadas. A nomenclatura desenvolvida foi baseada na abreviação do nome do microrganismo do qual a enzima foi isolada. A primeira letra representa o gênero e as outras duas a espécie, seguido de um algarismo romano (ou outra letra) que indica a ordem da descoberta ou a linhagem da qual ela foi isolada. Por exemplo, a enzima de restrição denominada de *EcoRI* é purificada de uma *Escherichia coli* que carrega um fator de transferência de resistência RI, enquanto que a *Hind III* é isolada da *Haemophilus influenzae*, linhagem d III.

A Tabela 1 mostra a seqüência palindrômica e o local de clivagem de algumas enzimas de restrição. Note que algumas enzimas deixam terminações coesivas enquanto que outras fazem cortes abrupos ou não coesivos.

Microrganismo	Enzima	Sequência Alvo
<i>Escherichia coli</i>	<i>EcoRI</i>	<pre> G A A T T C C T T A A G ^ v </pre>
<i>Bacillus amyloliquefaciens H</i>	<i>BamHI</i>	<pre> G G A T C C C C T A G G ^ v </pre>
<i>Bacillus globigii</i>	<i>BglII</i>	<pre> A G A T C T T C T A G A ^ v </pre>
<i>Haemophilus aegyptius</i>	<i>HaeII</i>	<pre> Pu G C G C Pi P y C G C G B ^ v </pre>
<i>Haemophilus aegyptius</i>	<i>HindIII</i>	<pre> A A G C T T T T C G A A ^ v </pre>
<i>Providencia stuartii</i>	<i>PstI</i>	<pre> C T G C A G G A C G T C ^ v </pre>
<i>Streptococcus albus G</i>	<i>SalI</i>	<pre> G T C G A C C A G C T G ^ v </pre>
<i>Thermus aquaticus</i>	<i>TaqI</i>	<pre> T C G A A G C T ^ v </pre>
<i>Brevibacterium albidium</i>	<i>BalI</i>	<pre> T G G C C A A C C G G T ^ v </pre>
<i>Haemophilus aegyptius</i>	<i>HaeIII</i>	<pre> (A G G C T T C C G A ^ v </pre>
<i>Serratia marcescens</i>	<i>SmaI</i>	<pre> C C C G G G G G G C C C ^ v </pre>

Tabela 1. Algumas endonucleases de restrição: origem e sítios de clivagem. A seta indica o local de clivagem. Pu e Pi referem-se, respectivamente, a qualquer purina e pirimidina.

O interesse por estas enzimas de restrição aumentou em 1973 quando se percebeu que elas poderiam ser usadas para fragmentar o DNA deixando extremidades de fitas simples de DNA que permitiam a ligação dos fragmentos. Isto significava que a recombinação poderia ser efetuada em tubos de ensaio. Além disto, DNA bacteriano poderia recombinar com DNA humano ou de qualquer outra espécie, abrindo a possibilidade de clonar genes humanos ou isolar proteínas de culturas bacterianas.

Uma importante consequência da especificidade destas enzimas de restrição é que o número de clivagens feito por cada uma delas no DNA de qualquer organismo é definido e

permite o isolamento de fragmentos deste DNA. Portanto, cada enzima de restrição gera uma família única de fragmentos quando cliva uma molécula de DNA específica. Enzimas que reconhecem sítios de restrição compostos por 4 pares de bases clivam o DNA em média a cada 256 nucleotídeos ($4^4=256$). Aquelas que reconhecem sítios com 6 e 8 pb clivam o DNA em média a cada 4096 e 65536 pb, respectivamente. No entanto, esta média pode sofrer variações significativas, dependendo principalmente da composição de bases do DNA analisado. Por exemplo, a enzima *NotI* reconhece um sítio de restrição contendo 8pb, incluindo nucleotídeos CpG, que raramente ocorre no DNA de mamíferos.

A família de fragmentos gerados por digestão com enzima de restrição é geralmente detectada pela separação destes fragmentos por eletroforese em gel de agarose. Os fragmentos migram em função de seus pesos moleculares sendo que os menores migram mais rapidamente (Figura 2).

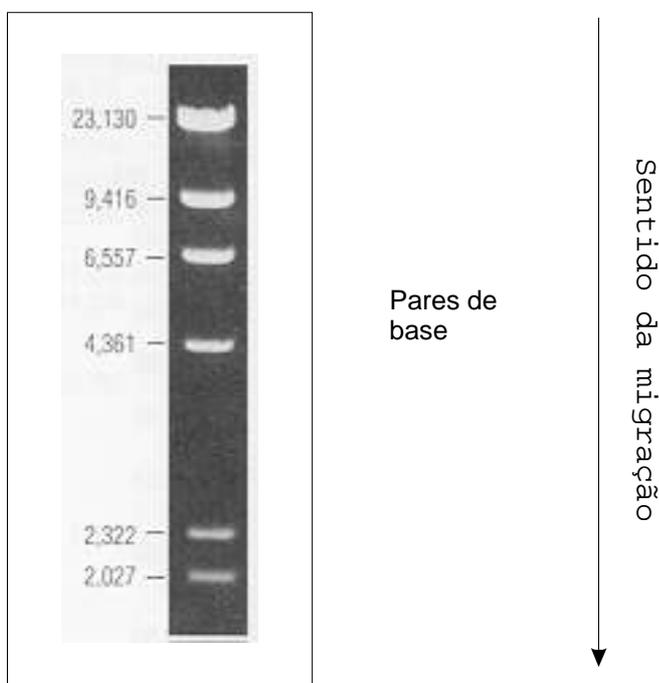


Figura 2. Moléculas de DNA de tamanhos diferentes podem ser separadas por eletroforese. Moléculas menores movem-se mais rapidamente que moléculas maiores, tornando-se portanto separadas em bandas. O DNA é corado com brometo de etídio, uma molécula que fluoresce quando iluminada com luz Ultra Violeta.

4. CONSTRUÇÃO DO DNA RECOMBINANTE

Uma enzima de restrição particular reconhece somente uma seqüência única de bases. DNAs de origens diferentes sob a ação da mesma enzima de restrição produzem fragmentos com o mesmo conjunto de extremidades fitas simples. Portanto, fragmentos de

dois diferentes organismos (por exemplo, bactéria e homem) podem ser ligados por renaturação das regiões de fita simples. Além disto, se a ligação for "selada" com a enzima DNA ligase, depois do pareamento de bases, os fragmentos serão ligados permanentemente.

A técnica de DNA recombinante tem um interesse especial se uma das fontes de DNA clivado for um plasmídeo.

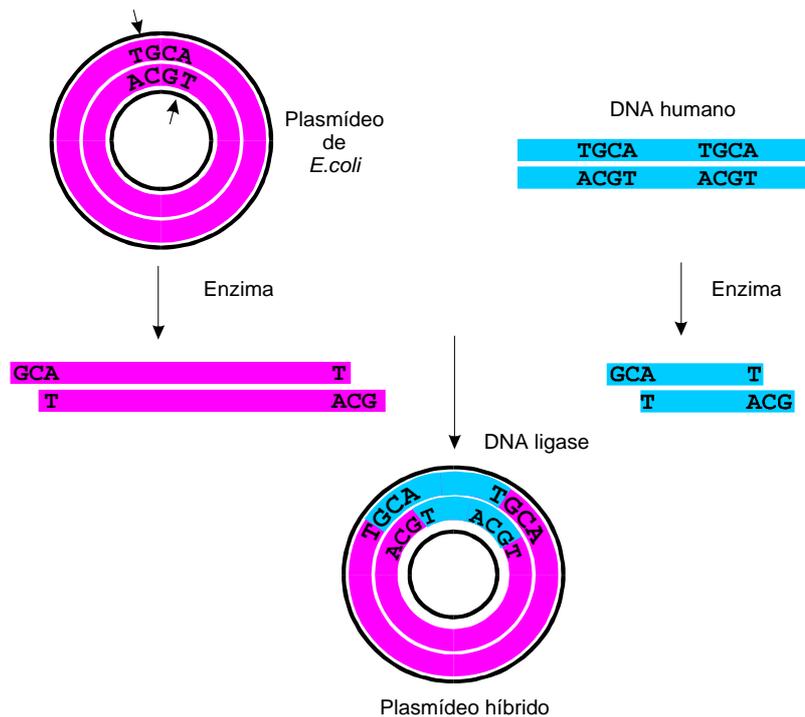


Figura 3. Construção de uma molécula de DNA híbrida a partir de fragmentos de diferentes organismos obtidos com o uso de enzima de restrição.

A figura 3 mostra uma molécula de DNA de plasmídeo que tem somente um sítio de clivagem para uma determinada enzima de restrição. A mesma enzima é usada para clivar DNA humano. Se os fragmentos de DNA humano são misturados com o DNA plasmidial linearizado, permitindo a ligação entre eles, uma molécula de DNA plasmidial contendo DNA humano pode ser gerada. Este plasmídeo híbrido pode ser inserido numa bactéria através de transformação e então o inserto será replicado como parte do plasmídeo. Geralmente, antibióticos são acrescentados ao meio da cultura para selecionar somente as linhagens que portam os plasmídeos (o plasmídeo usado para esta finalidade porta resistência a pelo menos um antibiótico).

5. DNA LIGASE

Conforme mencionado anteriormente, esta enzima promove a ligação dos fragmentos de DNA em vetores previamente clivados por endonucleases de restrição. A DNA ligase requer um grupo OH livre na extremidade 3' de uma das cadeias de DNA e um grupo fosfato na extremidade 5' da outra cadeia (Figura 4). A *E.coli* e o fago T4 codificam uma DNA ligase capaz de selar fragmentos de DNA com dupla fita. DNA ligase isolada de *E.coli* e de outras bactérias requer NAD^+ , enquanto que a isolada do bacteriófago T4 requer ATP como cofator.

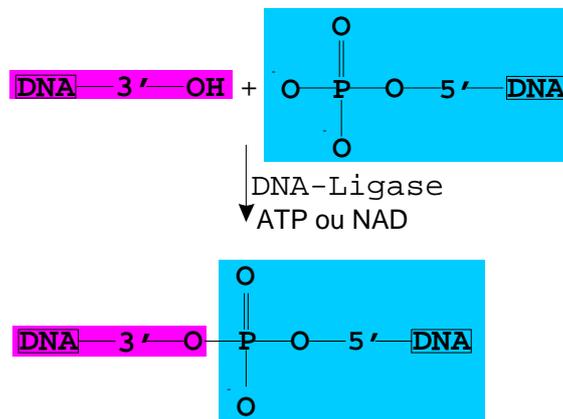


Figura 4. DNA ligase cataliza a junção de duas fitas de DNA que são partes da molécula da dupla-hélice.

5.1. Tipos de fragmentos de DNA que são ligados pela DNA ligase

a) Fragmentos com extremidades coesivas

As extremidades coesivas produzidas por várias enzimas de restrição permitem que dois fragmentos de DNA sejam mais facilmente ligados. Isto porque ocorre inicialmente o pareamento das fitas simples das extremidades coesivas das duas diferentes moléculas, através da formação de pontes de hidrogênio pela complementariedade das bases. Finalmente, a ligação covalente (fosfodiéster) dos fragmentos é realizada pela DNA ligase (ver figura 4).

b) Fragmentos com extremidade não coesivas

DNAs portando extremidades não coesivas são ligados com muito menos eficiência que aqueles que tem extremidades coesivas. Uma concentração muito maior de DNA ligase

e mesmo dos DNAs envolvidos é necessária para que as moléculas com extremidades não coesivas sofram reação de ligação.

No entanto, a ligação deste tipo de fragmento é facilitada pela transformação prévia das extremidades não coesivas em coesivas. Este procedimento pode ser feito através de dois processos:

a) Adição de polidesoxiadenina na extremidade 3' de um fragmento de DNA e polidesoxitimina na extremidade 3' de um outro fragmento de DNA, através da enzima desoxinucleotidil-transferase-terminal ou simplesmente transferase. Esta enzima, uma DNA polimerase incomum, adiciona nucleotídeos à extremidade 3'-OH de fragmentos fita simples proeminentes de uma cadeia de DNA. Para gerar este tipo de extremidade proeminente a molécula é tratada previamente com uma exonuclease 5'-específica para remover alguns nucleotídeos terminais. Após a mistura dos dois tipos de fragmentos complementares e anelamento dos mesmos, as moléculas são unidas preenchendo-se os espaços existentes com o auxílio da DNA pol I e selando-os com DNA ligase (Figura 5)

b) Adição de adaptadores às extremidades não coesivas. Os adaptadores são oligonucleotídeos sintéticos que contém sítios de clivagem para uma ou mais enzimas de restrição. Eles são unidos ao DNA com o auxílio da DNA ligase (Figura 6).

Alternativamente, no lugar de modificar a extremidade não coesiva, pode-se optar pela utilização de T4 ligase em grande concentração, o que permitirá a ligação entre moléculas de DNA sem proeminências.

6. TRANSFORMAÇÃO BACTERIANA

6.1. Conceito de transformação induzida

O processo de transformação constitui um evento de alta importância na técnica de manipulação gênica. A transformação natural descrita por Griffith, em 1928, e por Avery e colaboradores, em 1944, é um evento raro. No entanto, em 1970, Mandel e Higa encontraram que a *E.coli* tornou-se marcadamente competente para transformação com DNA exógeno, quando a bactéria foi suspensa em cloreto de cálcio gelado e submetida a um curto choque térmico à 42°C. Estes mesmos autores também verificaram que as bactérias crescidas até a fase log eram mais competentes do que aquelas isoladas de outros estágios do crescimento.

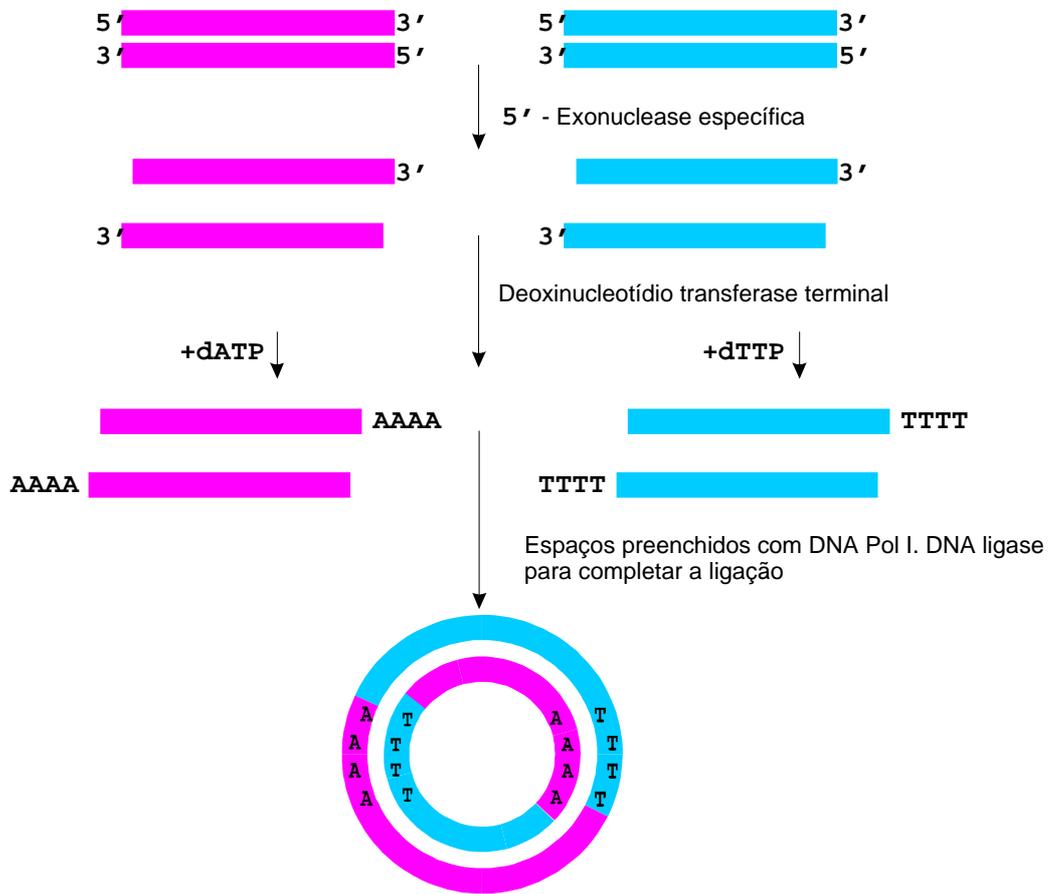


Figura 5. Fragmentos de DNA com extremidades não coesivas podem ser transformadas em coesivas pela adição de poli (A) e poli (T).

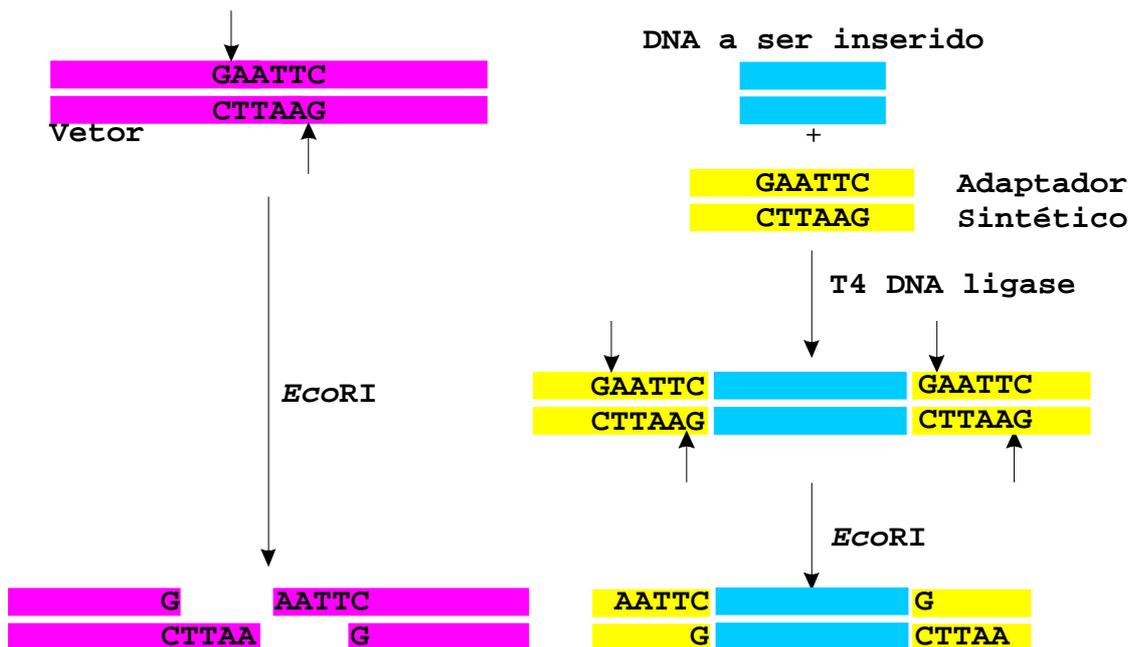


Figura 6. Fragmentos de DNA com extremidades não coesivas podem ser transformados em coesivas pela adição de adaptadores e posterior tratamento com a enzima de restrição que reconhece o adaptador.

O procedimento do cloreto de cálcio que é usado até hoje produz uma eficiência de transformação da ordem de 10^5 a 10^7 transformantes por micrograma de DNA (a eficiência de transformação é geralmente expressa como o número de células transformadas, obtido a partir de um micrograma de DNA plasmidial intacto).

O tamanho e a conformação da molécula do DNA, afetam o processo de transformação. Plasmídeos pequenos são mais facilmente incorporados pela célula bacteriana competente; DNA linear é pobremente incorporado, talvez pelo fato de sofrer degradação pelas exonucleases presentes no espaço periplasmático.

6.2. Mecanismos de captação do DNA

O mecanismo de captação da molécula do DNA pela bactéria competente ainda é desconhecido. Uma hipótese é que as moléculas do DNA passam através de canais situados nas chamadas zonas de adesão, que são locais onde a membrana interna e externa da célula bacteriana unem-se formando poros. Estes poros só estão presentes durante o crescimento bacteriano (fase de crescimento exponencial).

Em condições naturais, a captação do DNA torna-se difícil devido a repulsão eletrostática existente entre as cargas negativas da camada de fosfolipídeos da membrana bacteriana e dos grupo fosfato da molécula do DNA.

O papel do cálcio é explicado pela hipótese de que a 0°C a fluidez da membrana celular é cristalizada, estabilizando a distribuição dos fosfatos carregados. Os íons Ca^{+2} formam um complexo com este grupamento, cobrindo as cargas negativas e assim facilitando a atração eletrostática com as moléculas do DNA na zona de adesão. O choque térmico, complementa este processo de captação, provavelmente criando um desbalanço térmico entre o interior e o exterior da célula bacteriana, auxiliando o bombeamento do DNA através da zona de adesão. A figura 7 ilustra este processo.

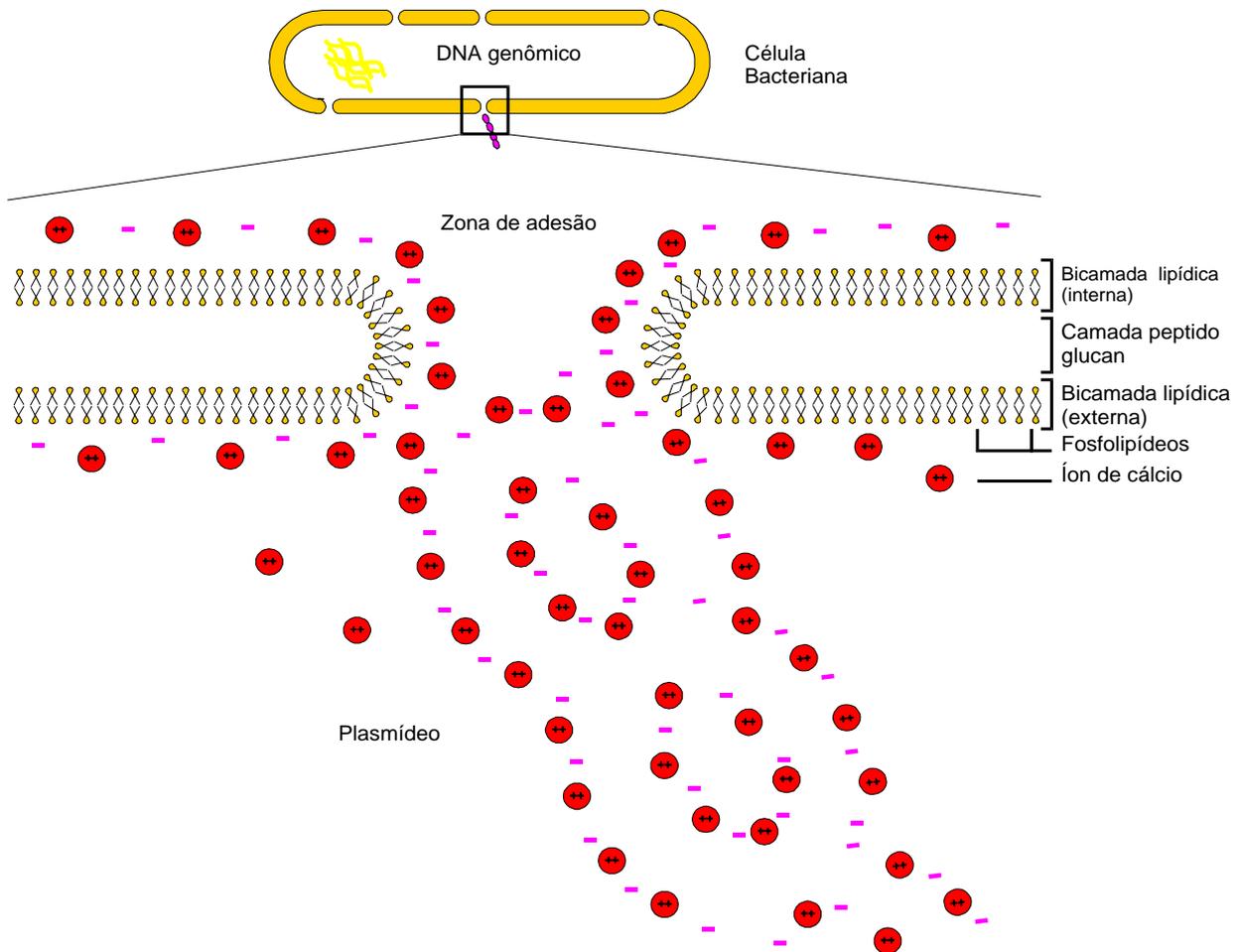


Figura 7. Mecanismo molecular proposto para explicar a transformação de *E. coli* com uma molécula de DNA exógeno.

II. VETORES DE CLONAGEM MOLECULAR

Após o isolamento de uma informação genética, por exemplo um fragmento de DNA obtido pela clivagem com enzimas de restrição, este fragmento deverá ser inserido numa outra molécula de DNA diferente, capaz de amplificar aquela informação genética em centenas de cópias. Este processo de amplificação é obtido através do uso de moléculas de DNA que são os chamados **vetores de clonagem molecular**.

Atualmente, os tipos básicos de vetores usados na metodologia do DNA recombinante apresentam características especiais que os tornam excelentes veículos de clonagem em diferentes situações.

A seguir, vamos apresentar os principais tipos de vetores atualmente em uso na biologia molecular.

1. PLASMÍDEO

Plasmídeos são pequenas moléculas de DNA dupla fita, contendo os elementos necessários para a sua replicação e pelo menos um gene que confere resistência a antibiótico. Estes elementos genéticos extra cromossomais variam de 5 a 400 kilobases e comumente estão presentes em duas ou mais cópias por célula. Os plasmídeos presentes num grande número de cópias são usados como veículos de clonagem desde que capacitem a amplificação do segmento do DNA neles clonado.

Um plasmídeo para ser um bom vetor de clonagem deve conter as seguintes propriedades:

- possuir uma **origem de replicação (O)**, ou seja, uma seqüência de DNA que permita que o vetor seja replicado na célula hospedeira;
- apresentar dois ou mais sítios únicos de clivagem para endonucleases de restrição. O conjunto destes sítios, denominado de **Múltiplos Sítios de Clonagem (MSC)**, é o local onde o inserto é incorporado ao vetor de clonagem;
- possuir um gene que codifica um produto que distingue a célula transformada da célula não transformada. Por exemplo, muitos vetores de clonagem carregam o gene que confere resistência à ampicilina (Amp^R). As células transformadas com tais vetores são capazes de crescer num meio contendo o antibiótico, enquanto que as células não transformadas acabam morrendo.

A figura a seguir ilustra as principais características estruturais de um plasmídeo.

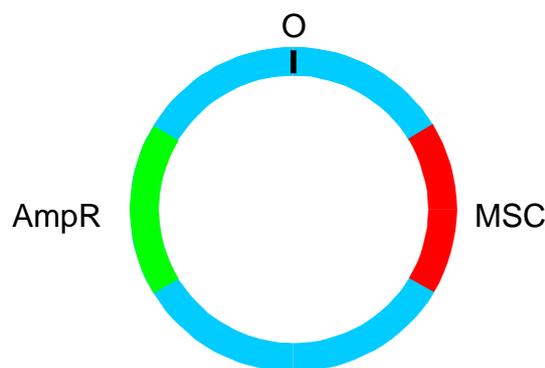


Figura 8. Estrutura molecular de um plasmídeo típico usado em clonagem molecular. Estão representados o gene de resistência (Amp^R), a região de múltiplos sítios de clonagem (MSC) e a origem de replicação (O).

Um dos passos fundamentais no processo de clonagem molecular é o uso de enzima de restrição que produz extremidades compatíveis durante a clivagem do DNA a ser clonado (inserto) e a do DNA receptor (vetor).

Uma vez que o DNA foi ligado ao vetor, esta molécula híbrida deverá ser introduzida numa célula hospedeira geralmente bactérias, por um processo chamado de transformação, para que o vetor possa sofrer replicações e conseqüentemente amplificar o número de cópias do inserto. A bactéria transformada será facilmente reconhecida pela aquisição de um novo fenótipo dado pelo plasmídeo, ou seja, capacidade de crescer em meios contendo antibiótico.

2. FAGOS

Um dos vetores mais utilizados nos processos de clonagem molecular é o denominado **bacteriófago** λ , o qual comporta-se como um vírus da *E.coli*. Antes de analisarmos o uso deste elemento como veículo de clonagem molecular, vamos mostrar resumidamente as suas propriedades biológicas e estruturais.

2.1 Biologia do fago λ

O fago λ é um parasita obrigatório da *E.coli*, o qual necessariamente deve injetar o seu DNA na bactéria hospedeira para a sua multiplicação. Após a infecção da *E.coli* o genoma do fago pode seguir duas vias:

a) No primeiro caso denominado **estado lítico**, o DNA do fago permanece na bactéria como uma molécula independente, havendo a ativação de alguns genes do fago e a concomitante inativação de outros, dentro de um programa estritamente definido. Como resultado o cromossomo do fago é replicado ativamente, ocorre a síntese das proteínas da capa e da cauda e formam-se novas partículas virais. Em aproximadamente 45 minutos após a infecção a célula hospedeira é lisada havendo a liberação de cerca de 100 novos fagos.

b) A outra via que o cromossomo do fago pode seguir é o denominado **estado lisogênico**. Neste caso, o DNA do fago é integrado no cromossomo da bactéria passando a ser chamado profago. No estado lisogênico, todos os genes do profago estão inativos com exceção do gene que produz a proteína repressora. A bactéria hospedeira carregando o profago multiplica-se e este é replicado passivamente e distribuído para as bactérias descendentes.

Em condições naturais a opção entre seguir o estado lítico ou lisogênico depende das condições do meio. Assim, via de regra, em meio rico em nutrientes o estado lítico é preferencial, por exemplo o fago λ na bactéria *E.coli* intestinal. Por outro lado, em meio pobre de nutrientes como é o caso da *E.coli* no solo, o fago prefere o estado lisogênico. Em condições experimentais, o estado a ser seguido depende de um balanço entre os fatores do meio intra e extra celular e de fatores genéticos da bactéria hospedeira e do bacteriófago.

2.2 Genoma do fago λ

O bacteriófago λ é uma partícula viral constituída de aproximadamente 50% de proteínas e 50% de DNA. O DNA do fago λ , na forma como ele é isolado da partícula viral, é uma molécula linear com dupla fita de aproximadamente 48.500 pares de bases. As extremidades da molécula contêm uma fração de DNA fita simples com cerca de 12 nucleotídeos, os quais são complementares na sequência de bases e através delas é que o DNA assume a forma circular quando ele é injetado na célula hospedeira. Estas extremidades são denominadas de **sítios cos**. O genoma do fago λ codifica para aproximadamente 50 proteínas, cujos genes tem um cronograma de expressão bem definido, o que determina a instalação do estado lítico ou lisogênico.

As figuras 9 e 10 ilustram o ciclo biológico do fago λ em uma célula hospedeira e o genoma do fago λ com alguns dos seus principais genes, respectivamente.

2.3 Controle de expressão dos genes do fago λ

Seja qual for a via a ser seguida pelo fago λ , ou seja via lítica ou lisogênica, a expressão dos genes envolvidos nestes circuitos começa pelos produtos dos genes **N** e **Cro**, regulados pelos promotores **pL** e **pR** respectivamente situados à esquerda (**pL**) e à direita (**pR**) do gene **cI**.

Durante o transcurso da **via lítica**, o produto do gene **cro** está diretamente relacionado com a replicação do genoma do fago λ , através da indução da expressão dos genes **OP**. Por outro lado, o produto do gene **N** está diretamente relacionado com a expressão dos genes da região de empacotamento, ou seja genes **A** e **J**, responsáveis pela

síntese das proteínas da cabeça e da cauda do fago λ , como também da expressão dos genes **S** e **R**, diretamente envolvidos com a lise da célula hospedeira.

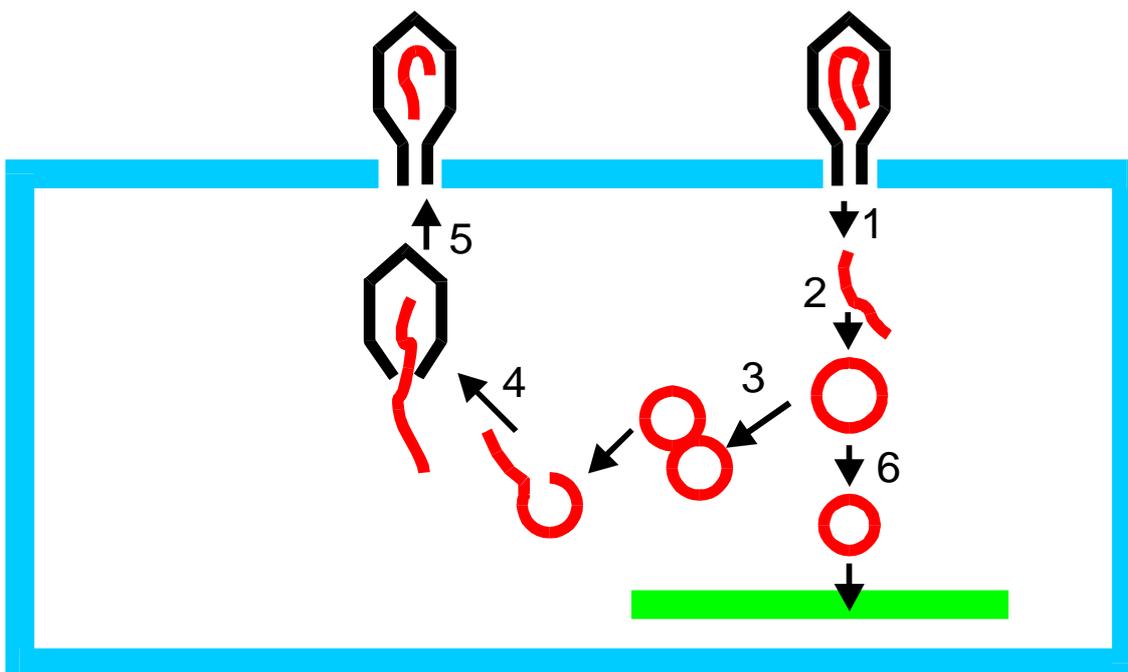


Figura 9. Replicação do fago λ no interior da célula hospedeira. Após adsorção (1) e injeção (2) do genoma do fago λ na bactéria, a via lítica indicada pelos números 3, 4 e 5 leva a formação de novas partículas virais. Alternativamente, a via lisogênica (6) pode ser ativada levando à integração do material genético viral ao genoma da bactéria hospedeira.

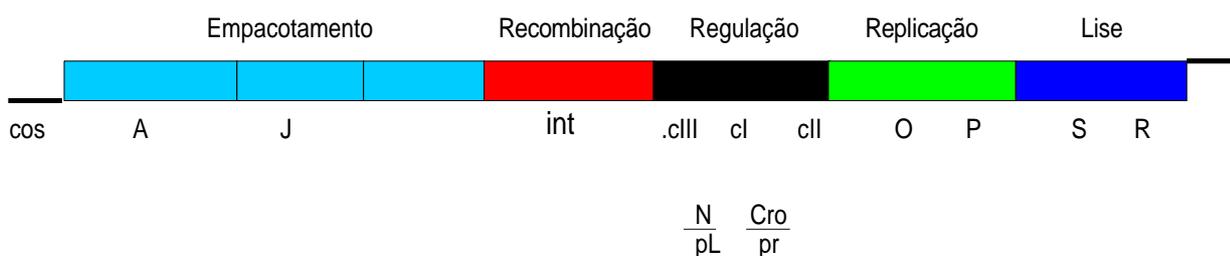


Figura 10. Representação esquemática do genoma do fago λ .

Durante a **via lisogênica**, a transcrição também inicia-se pelos promotores **pL** e **pR** coordenando a expressão dos genes **cII** e **cIII**. O produto destes dois genes coordena a expressão do gene **cI** que produz uma proteína repressora inibindo a expressão dos genes responsáveis pelo empacotamento e a lise. Por outro lado, um outro gene importante é o **int**,

cujo produto relaciona-se com a integração do fago no genoma bacteriano estabelecendo o estado lisogênico.

2.4 O uso do fago λ como vetor de clonagem molecular

Durante o ciclo lítico, os genes envolvidos no processo de lisogênia são dispensáveis e conseqüentemente a região de integração do genoma do fago pode ser totalmente substituída por um outro fragmento de DNA, sem que haja qualquer alteração nos processos envolvidos na via lítica.

Uma das maiores vantagens de usar o fago λ como vetor de clonagem é que o DNA inserido no fago é empacotado *in vitro*. Embora a eficiência de empacotamento seja cerca de 10%, uma vez que os fagos são empacotados teremos 100% de eficiência na infecção da *E.coli* hospedeira. Este processo é altamente eficiente quando comparado com o da transformação bacteriana com plasmídeos. Neste caso, os melhores resultados situam-se ao redor de 10^8 transformantes por μg de DNA o que significa que menos de 1 em 1000 plasmídeos são incorporados na *E.coli* hospedeira.

Atualmente, existem vários derivados do fago λ nos quais um ou mais sítios de restrição flanqueiam a região de recombinação, facilitando a sua substituição por um outro DNA. Estes são os chamados **vetores de substituição**. Um exemplo típico destes vetores é o **EMBL3**, no qual a região de recombinação é flanqueada pelas enzimas de restrição *EcoRI*, *BamHI* e *Sall*. Esse vetor quando clivado pode receber fragmentos de DNA variando de 10,4 a 20 kb, como mostra a figura abaixo.

Outros derivados do fago λ são os chamados **vetores de inserção**. Neste tipo de vetor, os fragmentos de DNA que podem ser inseridos devem ter no máximo até 7kb de tamanho para não alterar os processos de empacotamento do genoma do fago.

Os vetores de inserção derivados do fago λ mais bem utilizados especialmente na clonagem de cDNA, são os vetores $\lambda\text{gt}10$ e o $\lambda\text{gt}11$. Vamos a seguir fazer algumas considerações sobre estes dois tipos de vetores.

No fago $\lambda\text{gt}10$, o inserto geralmente cDNA, é inserido no sítio da *EcoRI* situado no gene repressor **cI**. O fago recombinante terá agora o gene **cI** obstruído e conseqüentemente durante a cultura provocará a lise das colônias de bactérias transformadas. Essas colônias lisadas são visualizadas como círculos com o centro claro (placas de lise), bastante distintos

das colônias não recombinantes, que por possuírem o gene **cI** íntegro não são lisadas e permanecem como colônias turvas.

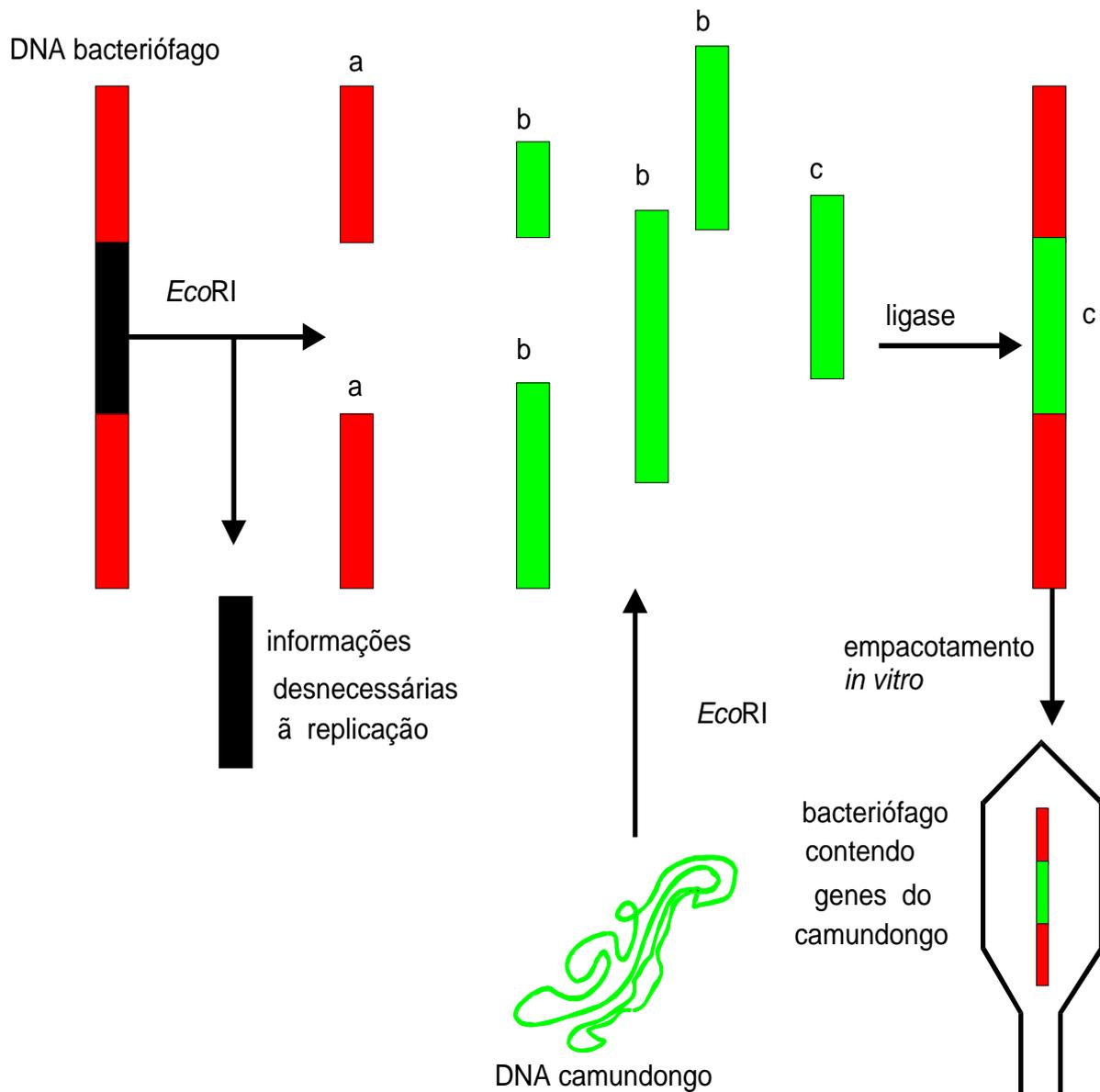


Figura 11. Representação esquemática da clonagem de um fragmento de DNA genômico no fago λ . As regiões indicadas por **a** contêm todas as informações necessárias para replicação em *E. coli* e empacotamento *in vitro*. Os diferentes fragmentos obtidos da digestão do DNA de interesse com enzima apropriada estão indicados pelas letras **b** e **c**, onde somente **c** possui tamanho adequado para o empacotamento.

Por outro lado, o fago λ gt11 contém um sítio de restrição *EcoRI* localizado num gene chamado *LacZ*, a 53 nucleotídeos do códon de término da enzima codificada por esse gene (β -galactosidase). Essa enzima, cuja expressão pode ser induzida por IPTG (isopropil-

β -D-galactosídeo), degrada o substrato X-gal produzindo um precipitado de cor azul. Portanto, colônias de bactérias carregando o DNA do fago com o gene LacZ intacto (sem inserto), na presença do indutor IPTG e do substrato X-gal, ficarão azuis. Enquanto que, aquelas portanto o DNA o fago que incorporou o inserto no sítio de *EcoRI*, não expressam a β -galactosidase e permanecem de cor branca, o que facilita a identificação dos clones recombinantes.

3. COSMÍDEO

A clonagem de fragmentos de DNA no fago λ apresenta uma limitação, pois o fragmento a ser clonado não pode ser maior do que cerca de 15kb (Figura 12). Na maioria das vezes, esta dimensão é suficiente para conter um gene completo, incluindo as sequências flanqueadoras. Entretanto, muitos genes apresentam dimensões da ordem de 35 a 40 kb e nestes casos, a técnica usada para a clonagem deste tipo de fragmento é a chamada **clonagem em cosmídeos**.

Os cosmídeos, são plasmídeos que contêm um fragmento de DNA do fago λ que inclui o sítio **cos**. Estes cosmídeos podem ser usados como veículos de clonagem molecular empregando o sistema de empacotamento *in vitro* que reconstitui a estrutura do fago (cabeça e cauda) e assim é usado para infectar a célula hospedeira.

As enzimas do sistema de empacotamento do fago λ reconhecem os dois sítios **cos** situados de 35 a 49 kb de distância e neste caso, somente fragmentos desta ordem de tamanho serão convenientemente empacotados.

O DNA genômico de interesse é clivado com uma enzima de restrição que produz grandes fragmentos de DNA, os quais serão ligados ao cosmídeo, também clivado com a mesma enzima.

A situação ideal é que cada fragmento do DNA genômico apresente um sítio **cos** nas suas extremidades. Durante o empacotamento, as enzimas do sistema reconhecem os sítios **cos** situados a uma distância dentro de 49Kb, clivam estes sítios e empacotam estas moléculas.

O DNA do cosmídeo é injetado no interior da célula hospedeira, circulariza igual ao DNA do fago e replica como um plasmídeo normal, sem expressão de qualquer função do

fago. As células infectadas serão selecionadas com base na resistência adquirida a um determinado antibiótico. A figura 12 ilustra um típico processo de clonagem em cosmídeo.

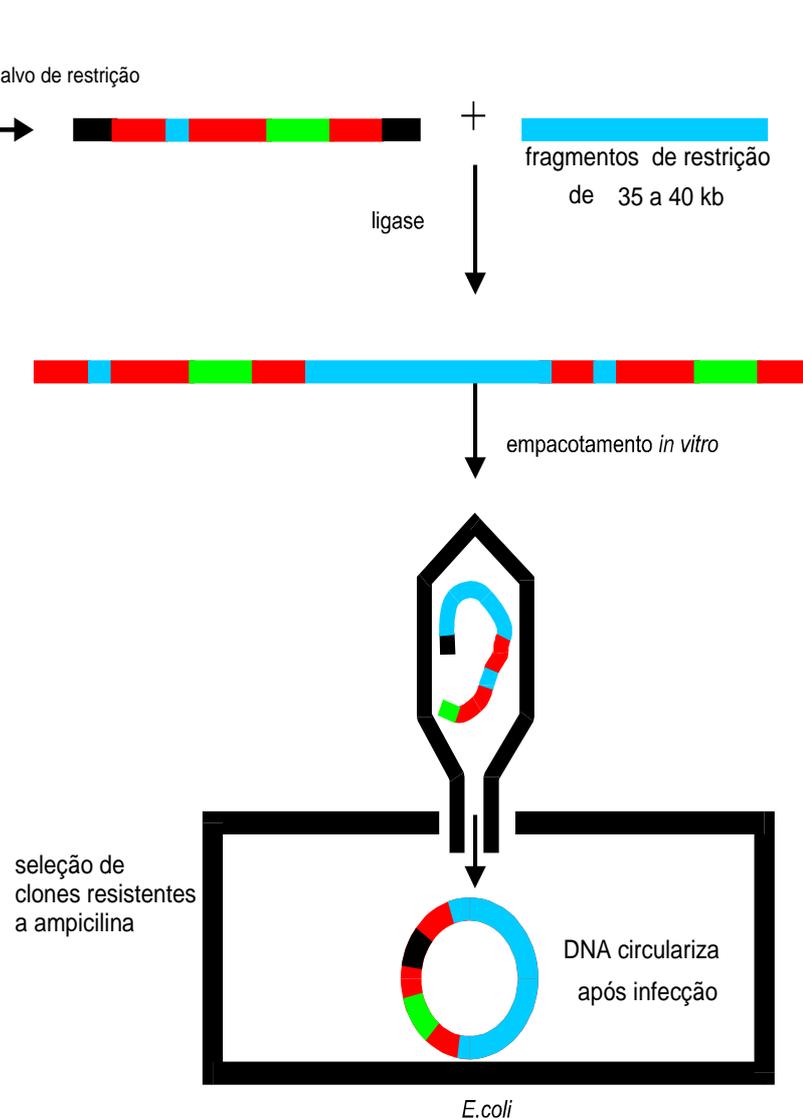


Figura 12. Esquema de clonagem em cosmídeo.

4. VÍRUS

A biologia molecular dos vírus de DNA e RNA foi extensivamente estudada antes do advento da metodologia do DNA recombinante. O **vírus SV40**, isolado de células tumorais de macacos, foi um dos primeiros sistemas virais utilizados para introduzir genes em células de mamíferos. O DNA do SV40 apresenta 5.200 pares de bases e pode ser dividido em duas regiões quanto a expressão gênica viral. Estas regiões são chamadas de região **precoce** e região **tardia**.

A região precoce é expressa através do ciclo lítico, enquanto que a expressão dos genes da região tardia ocorre somente após o início da replicação do DNA viral. Entre estas duas regiões situa-se a origem de replicação do DNA do vírus.

Um produto de expressão da região precoce é o **antígeno T**, o qual é responsável pela transformação maligna de células não permissíveis (células nas quais o vírus não completa a sua replicação), como também pelo início da replicação viral em células permissíveis. Os produtos de expressão codificados pela região tardia correspondem às proteínas que formam o **capsídeo** da partícula viral. A figura 13 ilustra o genoma do vírus SV40.

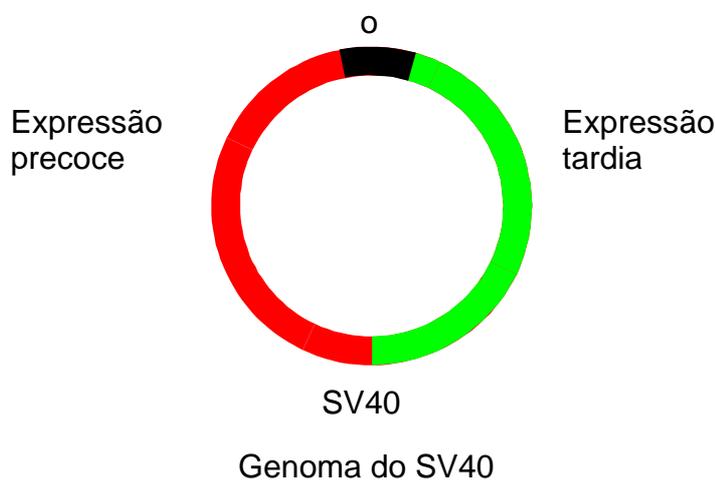


Figura 13. O DNA do vírus SV40

4.1. Clonagem no vírus SV40

A produção de DNA recombinante no vírus SV40, pode ser realizada através da substituição do segmento de expressão **precoce** ou do segmento de expressão **tardia**.

No primeiro caso, o vírus recombinante não produz o **antígeno T**, ao passo que na segunda opção não haverá produção das **proteínas do capsídeo**. Entretanto, estas funções deletadas podem ser suprimidas como veremos a seguir.

4.1.1. Clonagem no SV40 pela substituição da região tardia

Quando a região tardia do SV40 é substituída com outro DNA, perde-se a expressão das proteínas do capsídeo (funções tardias). Para que o sistema de clonagem

funcione adequadamente pode-se realizar a infecção simultânea com um outro vírus SV40 que tem uma deleção nos genes da região precoce, mas a região tardia intacta. Nas células co-infectadas as proteínas da região precoce serão produzidas pelo vírus recombinante e as proteínas do capsídeo pelo vírus deletado. Como resultado, teremos a produção de uma população de partículas virais mistas, ou seja, vírus deletado e vírus recombinante. A figura 14 ilustra este procedimento.

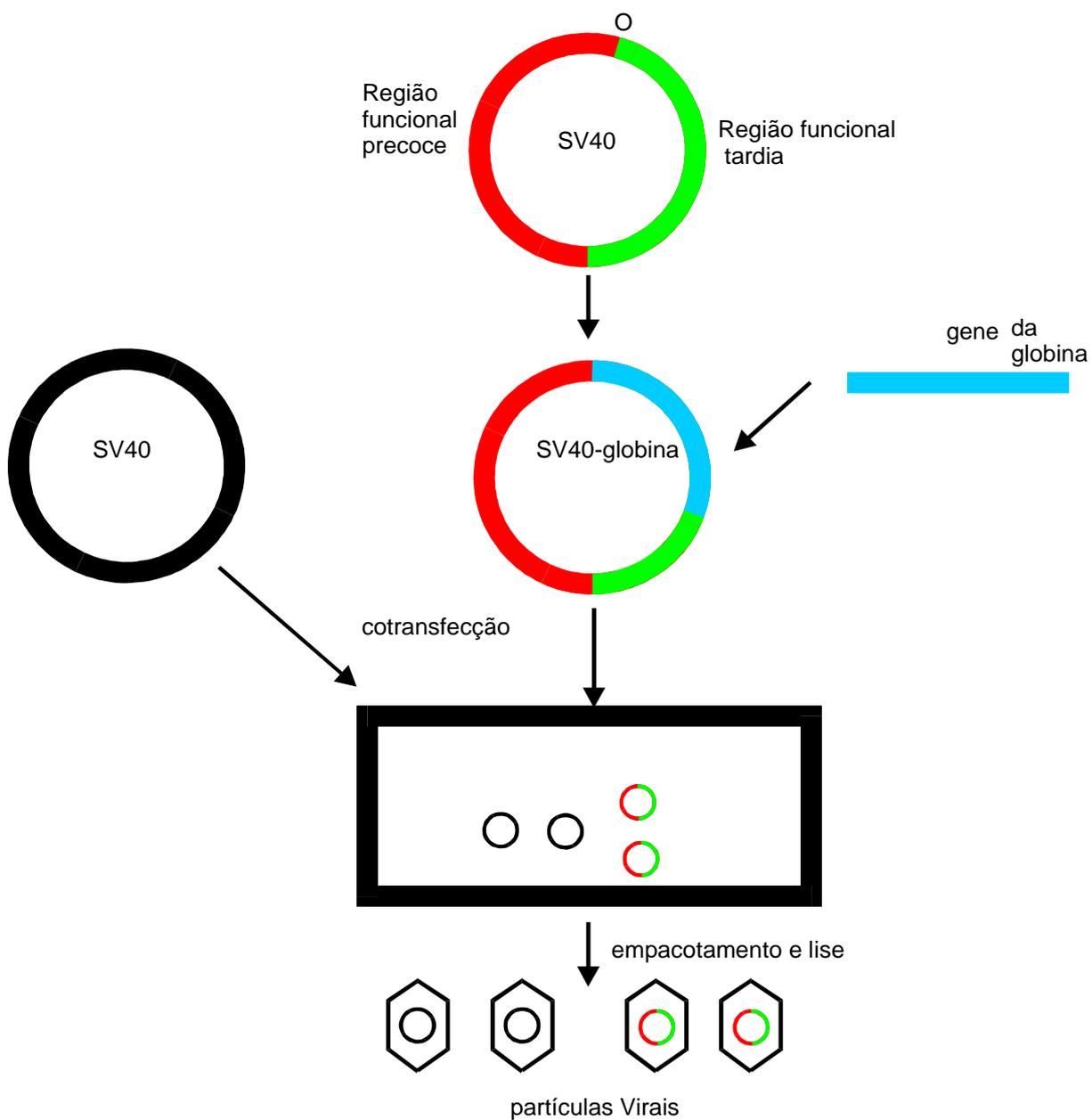


Figura 14. Clonagem no SV40 pela substituição da região tardia

4.1.2. Clonagem no SV40 pela substituição da região precoce

Quando a clonagem no SV40 é feita na região precoce, a produção de partículas virais depende do suprimento, por uma outra fonte, das proteínas codificadas pela região precoce (antígeno T). Para evitar uma infecção mista com um vírus SV40 deletado, usa-se uma linhagem celular, as células COS as quais, apresentam uma porção do genoma do SV40 integrado ao seu próprio cromossomo. Esta linhagem celular produz a proteína viral denominada antígeno T, a qual liga-se na origem de replicação do vírus e induz a sua replicação. A figura 15 ilustra este tipo de clonagem.

Apesar deste sistema viral ter sido usado como vetor de expressão em muitas estratégias de clonagem, existem algumas desvantagens de seu uso. Uma delas é que o gene a ser clonado não pode ser grande, a outra é que as células hospedeira só podem ser células de macacos e finalmente, o genoma da célula hospedeira pode sofrer rearranjos ou deleções.

Atualmente, dois outros sistemas virais mais versáteis estão em uso, são eles o **vírus da vacina** e o **Baculovírus**, os quais podem aceitar genes bem maiores do que aqueles aceitos pelo SV40. Um inconveniente para estes dois sistemas virais é que ambos apresentam um grande genoma e o gene a ser clonado só pode ser inserido por processos de recombinação dentro das células, o qual é bastante lento em relação aos processos tradicionais em uso na metodologia do DNA recombinante.

Finalmente, um sistema viral bastante promissor são os **retrovírus** que são vírus cujo material genético é o RNA e apresentam um ciclo bastante diferente dos mencionados anteriormente que são ciclos líticos. Os retrovírus infectam uma grande variedade de células de mamíferos e o seu RNA é transcrito reversamente para DNA o qual é incorporado ao genoma da célula hospedeira. Neste estado ele produz continuamente novas partículas virais, juntamente com o produto do gene nele clonado. Devido a sua versatilidade, os retrovírus são atualmente os vetores eleitos para uso em terapia gênica.

5. BACTERIÓFAGO M13

O bacteriófago M13 é um fago filamentosos da *E. coli* e contém uma única fita de DNA envolvida por proteínas virais. Durante o seu ciclo, a célula hospedeira é infectada

pela fita (+) a qual servirá como molde para a síntese da outra fita (-). A dupla fita denominada de **forma replicativa**, permanece no interior da célula hospedeira e é amplificada até 200 cópias por célula, quando então a fita negativa não mais se replica e a fita positiva continua a ser sintetizada, adquire as proteínas da cápsula viral e sai da célula hospedeira através de processo não lítico (Figura 16).

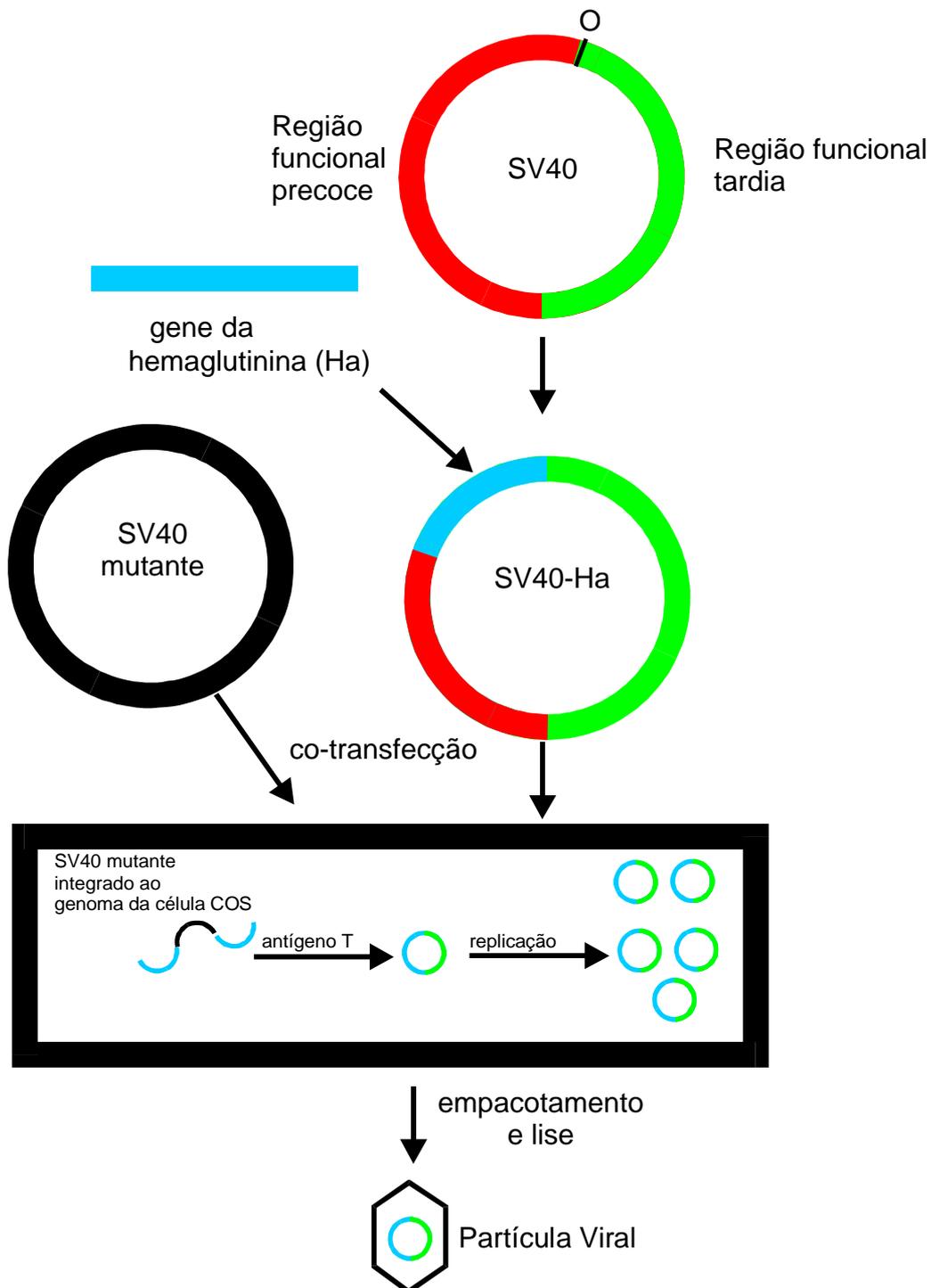


Figura 15. Esquema de clonagem no SV40 pela substituição da região precoce.

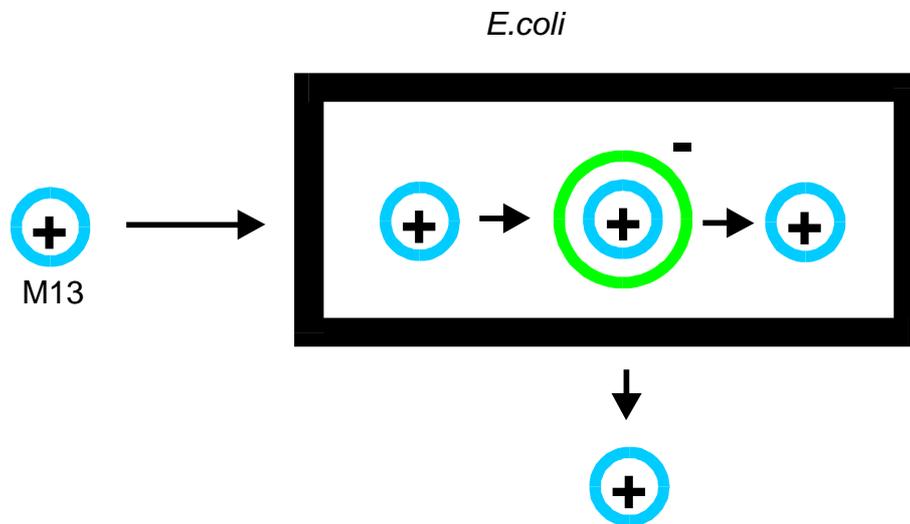


Figura 16. Replicação do bacteriófago M13.

5.1. Clonagem no bacteriófago M13

A grande utilidade deste sistema de clonagem é a sua facilidade na produção de DNA fita simples, facilmente recuperável do sobrenadante de uma cultura líquida da *E. coli* infectada com este vetor. Como veremos futuramente, este DNA fita simples é bastante útil no processo de sequenciamento do DNA pelo método desenvolvido por Sanger e colaboradores (1977).

Para facilitar o seu uso especialmente em estratégias de sequenciamento, foram desenvolvidos vetores da série **M13mp**, os quais contém o MSC inserido no gene que expressa a β -galactosidase.

Mutações foram realizadas de tal modo que a enzima só é ativa quando o vetor está na célula hospedeira, convenientemente preparada. Este processo denomina-se de alfacomplementação. A inclusão do múltiplo sítio de clonagem no gene da β galactosidade visa a identificação dos possíveis recombinantes quando na presença de um substrato cromogênico o X-gal ou Bluogal (5-bromo-4-cloro-indolil- β -D-galactosídeo) e o indutor IPTG. Quando o vetor está intacto (não recombinante), a enzima é funcional e frente ao substrato cromogênico este é clivado produzindo placas de cor azul. Por outro lado, quando um fragmento de DNA é inserido no múltiplo sitio de clonagem, a expressão da enzima é suprimida, portanto as células são incapazes de metabolizar o substrato cromogênico formando placas incolores.

5.2. Estratégia de clonagem em vetores da série M13mp

A utilização de vetores com diferentes orientações do múltiplo sítio de clonagem, tais como M13mp8 e o M13mp9, facilita a inserção de um fragmento de DNA em ambas as orientações. Como veremos adiante, esta estratégia apresenta grande aplicação no programa de sequenciamento de um DNA pelo método de Sanger.

A figura 17 ilustra a inserção de um fragmento de DNA clivado com as enzimas *Pst*I (P) e *Eco*RI (E) nos vetores M13mp8 e M13mp9, ambos clivados com as mesmas enzimas. Esta clonagem orientada permite a inserção do fragmento de DNA na orientação 5'→3' no vetor M13mp8 e na orientação 3'→5' no vetor M13mp9.

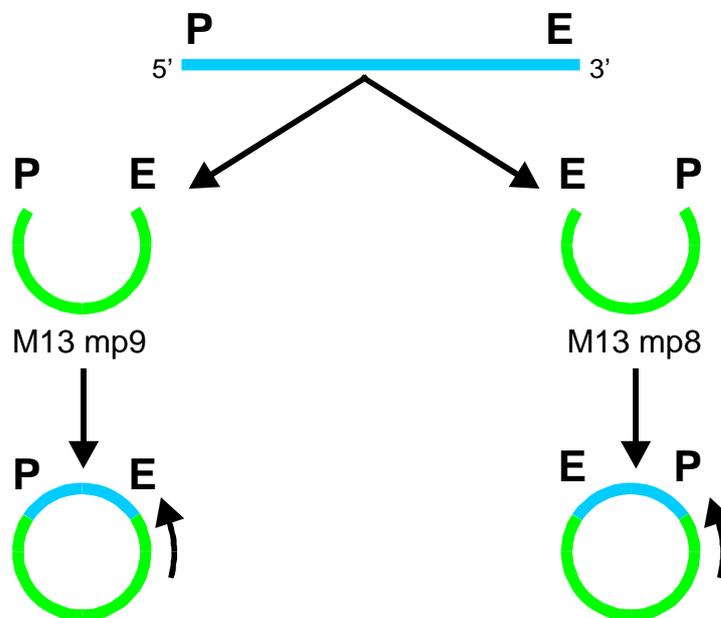


Figura 17. Clonagem de um fragmento de DNA nos vetores M13mp8 e M13mp9 clivados com as enzimas *Eco*RI (E) e *Pst*I (P).

6. FAGOMÍDEOS

Apesar do bacteriófago M13 apresentar grande aplicabilidade especialmente nos processos de seqüenciamento de DNA, foram desenvolvidas outras gerações de fagos, os quais reúnem as vantagens do fago M13 e a do plasmídeo. Os primeiros foram os vetores da série pEMBL e mais tarde os plasmídeos pTZ, pBluescript e pGEM (+/-). Estes vetores são chamados de **fasmídeos** ou **fagomídeos** e apresentam as seguintes características principais:

- fagomídeo apresenta um versátil MSC, permitindo a clonagem de grandes insertos sem maiores dificuldades, além de que, as enzimas situadas neste sítio foram ordenadas de tal modo a permitir uma alternativa interessante durante o processo de sequenciamento.
- utilizando uma combinação estratégica de duas enzimas de restrição, é possível criar terminações 5' e 3' proeminentes, tornando agora uma extremidade (a do inserto) susceptível à ação da Exonuclease III. Esta estratégia, permite a obtenção progressiva e controlada de vários fragmentos deletados do inserto, os quais após a recircularização tornam-se apropriados para o sequenciamento. Este procedimento, facilita o sequenciamento de um grande inserto de DNA, sem a necessidade de múltiplas subclonagens como é usual no sistema M13 ou a síntese de vários oligonucleotídeos internos, usados como iniciadores no processo de sequenciamento.

6.1. Clonagem no fagomídeo

Vamos utilizar como exemplo, o fagomídeo λ ZAP que é particularmente útil para clonagem de cDNA. Este vetor, incorpora a biologia do fago M13 de tal modo que um cDNA clonado neste vetor pode ser automaticamente excisado do fago para um fagomídeo denominado pBluescript.

Basicamente, qualquer DNA clonado no MSC inativa o gene *LacZ*, produzindo uma proteína de fusão quando na orientação correta, podendo assim também ser rastreada com anticorpos.

Quando uma cepa da *E.coli* (F') é infectada com o fago recombinante e superinfectada com o fago auxiliar, esse produz as proteínas do sistema M13. Essas proteínas reconhecem os sinais de iniciação e término da síntese de DNA do fago auxiliar (f1) que estão flanqueando o pBluescript, que por sua vez está incorporado no fago λ ZAP e carrega o inserto. Com a clivagem destas duas seqüências, o DNA é automaticamente circularizado na bactéria para originar a forma recombinante do plasmídeo Bluescript SK(M13).

O inserto clonado no Bluescript é flanqueado por promotores para as RNA polimerases dos fagos T3 e T7. Neste sentido, após o vetor ser convenientemente

linearizado, cópias do RNA relativo ao inserto, podem ser obtidas em ambas as direções através do uso das RNA polimerases T3 e T7.

7. VETORES PARA TRANSFORMAÇÃO DE LEVEDURAS

O grande interesse existente no estudo das leveduras *Saccharomyces cerevisiae* e *Schizosacharomyces pombe* deve-se ao fato de que tratam-se de organismos eucariotos inferiores e como tais possuem organização celular similar à de eucariotos superiores. Além disso, muitas proteínas de leveduras são similares estrutural e funcionalmente às suas homólogas em mamíferos. A utilização de leveduras como um modelo de estudo traz as seguintes vantagens:

- tempo de crescimento reduzido que permite a obtenção de grandes quantidades de leveduras em pouco tempo.
- genoma de pequeno tamanho, cerca de 200 vezes menor que o genoma de mamíferos, o que simplifica análises moleculares e genéticas.
- possibilidade de manter as leveduras como células haplóides ou diplóides, o que permite a obtenção de mutações recessivas em células haplóides, bem como a realização de experimentos de complementação genética. As células de mamíferos são diplóides o que torna impossível a detecção de mutações recessivas.

Existem diferentes marcadores de seleção que são utilizados em leveduras. A maior parte dos genes marcadores utilizados em leveduras são genes envolvidos na biossíntese de aminoácidos e nucleotídeos. Um dos marcadores frequentemente utilizados é o gene de levedura *LEU2* que codifica para uma das enzimas da via de síntese da leucina, a enzima β -isopropilmalato desidrogenase. A linhagem mutante de levedura, *leu2*, não cresce em meio de cultura que não contém leucina. Assim quando o gene *LEU2* é utilizado na transformação da linhagem *leu 2*, as células desta linhagem que adquiriram o gene *LEU2* serão capazes de crescer em meio de cultura deficiente no aminoácido leucina. Esta estratégia é semelhante a resistência à ampicilina adquirida por bactérias que foram transformadas com plasmídeos que contém o gene que promove a resistência à ampicilina.

A tabela a seguir lista alguns dos marcadores de seleção comumente utilizados em levedura.

GENE	ENZIMA	SELEÇÃO
<i>HIS3</i>	Imidazol glicerolfosfato desidratase	Histidina
<i>LEU2</i>	β -Isopropilmalato desidrogenase	Leucina
<i>LYS2</i>	α -Aminoacido redutase	Lisina
<i>TRP1</i>	N-(5'-fosforibosil)- antranilato isomerase	Triptofano
<i>URA3</i>	Orotidina-5'fosfato decarboxilase	Uracil

Tabela 2. Genes marcadores de seleção em levedura. Os meios de cultura utilizados em geral contêm todos os aminoácidos necessários à manutenção da levedura com exceção de um. Assim, por exemplo, células transformadas com o gene *HIS 3* são selecionadas em meio de cultura deficiente em histidina.

Os vetores utilizados na transformação de leveduras são capazes de se propagar tanto em *E. coli* quanto em levedura. A manipulação destes vetores (clonagem, mutagênese, sequenciamento, etc) é realizada em bactéria como para qualquer plasmídeo convencional e então o mesmo é transferido para levedura, organismo onde os ensaios de interesse são realizados. Tais vetores são denominados vetores do tipo ponte (“shuttle vectors”) e contêm entre outros elementos origens de replicação de bactéria e levedura bem como genes marcadores que funcionam para a seleção nos dois organismos.

Diferentes tipos de vetores são utilizados na transformação de leveduras:

Vetores de integração: tais vetores contêm origem de replicação de bactérias e genes marcadores para seleção em bactéria e levedura. Neste caso o plasmídeo não é capaz de replicar de modo autônomo na levedura. O modo pelo qual este tipo de vetor é propagado e é capaz de expressar o marcador de seleção é através da integração em um dos cromossomos de levedura.

Vetores que replicam em levedura: tais vetores também contêm origem de replicação de bactéria e genes marcadores para seleção em bactéria e levedura. Dois tipos de origens de replicação de levedura são empregadas nestes plasmídeos. O primeiro tipo consiste na origem de replicação do plasmídeo de 2 μ m, que é de ocorrência natural em levedura. O segundo consiste de origens de replicação isoladas de DNA cromossomal denominadas ARS (Autonomously Replicating Sequence, sequências de replicação autônoma). Plasmídeos contendo a origem de replicação do plasmídeo de 2 μ m ou sequências tipo ARS replicam em múltiplas cópias em levedura (Figura 18). Uma variante deste tipo de vetor inclui, além da sequência ARS, uma sequência tipo centromérica, denominada CEN, que garante que a replicação destes plasmídeos seja estável, ocorrendo uma única vez por ciclo

celular e que eles sejam segregados de modo preciso durante a mitose e meiose. Vetores contendo sequências tipo CEN são mantidos em cópia única na levedura (Figura 18).

YACs: (Yeast Artificial Chromosome, cromossomo artificial de levedura) são uma variação de um vetor de cópia única, que contêm sequências centroméricas (CEN) e sequências teloméricas, que são sequências das extremidades dos cromossomos. A presença de sequências teloméricas nestes vetores permite que sejam replicados como moléculas lineares, com comportamento semelhante ao do DNA cromossomal. YACs não são utilizados de rotina em experimentos de clonagem, entretanto, têm se mostrado uma ferramenta valiosa na caracterização de grandes segmentos genômicos na medida em que é possível clonar em um YAC fragmentos que vão de 100 a 2000 kb (Figura 19).

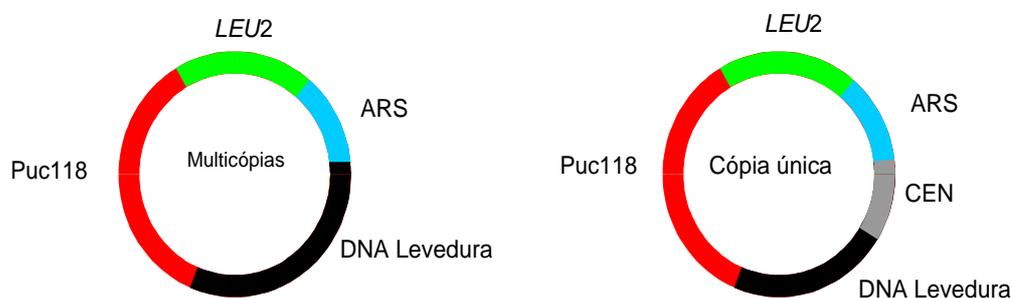


Figura 18. Vetores de levedura. Contêm sequências de plasmídeo (aqui de pUC118) que permitem a propagação e dão resistência a ampicilina em *E.coli*, o gene de seleção em levedura (neste caso *LEU 2*) e sítios únicos de restrição. *Vetores multicópias:* Estes plasmídeos existem na forma circular na levedura. Alguns destes vetores empregam a origem de replicação do plasmídeo de 2 μ m, outros utilizam origens de replicação isoladas de cromossomos que são denominadas ARS. 2 μ m e ARS permitem a replicação do plasmídeo em múltiplas cópias. *Vetores cópia única:* Vetores que contêm origens de replicação tipo ARS podem ser mantidos estávelmente através da adição de sequências centroméricas, CEN. A existência de sequências tipo CEN assegura que o plasmídeo seja mantido em 1 ou 2 cópias na levedura e segregado de modo preciso durante a divisão celular.

GLOSSÁRIO

- **ARS:** (autonomous replication sequence, sequência de replicação autônoma). Sequência que funciona como uma origem de replicação em cromossomos de levedura.
- **BAC:** (bacterial artificial chromosome, cromossomo artificial de bactéria). Vetor utilizado na clonagem de DNA genômico, baseado no fator F de plasmídeo que assegura que o plasmídeo seja mantido em pequeno número de cópias na bactéria.
- **Centrômero:** região do cromossomo que apresenta uma constricção e à qual liga-se o fuso mitótico durante a meiose ou mitose. É necessária para a manutenção da estabilidade e segregação correta dos cromossomos durante a divisão celular.

- **Origem de replicação:** região onde é iniciada a síntese de DNA em um dado fragmento de DNA.
- **Telômero:** extremidade do cromossomo que contém sequências repetitivas de DNA sintetizadas pela enzima telomerase e importantes na manutenção da integridade cromossomal.
- **YAC:** (yeast artificial chromosome, cromossomo artificial de levedura). Sistema de clonagem em levedura que consiste de um vetor linear que tem capacidade de replicação autônoma e que contém um centrômero de levedura e duas sequências teloméricas em suas extremidades.

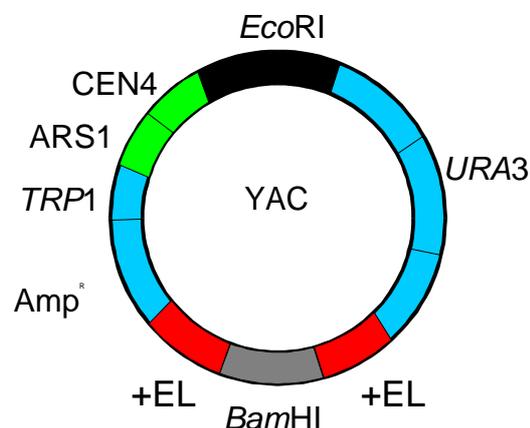


Figura 19. YACs: Contêm um centrômero, dois telômeros, gene(s) marcadores para seleção em levedura e uma sequência tipo ARS. Devido à presença destas sequências os YACs são replicados como pequenos cromossomos lineares na levedura.

III. CONSTRUÇÃO E USO DE BIBLIOTECAS DE DNA RECOMBINANTE

A obtenção de uma molécula de DNA recombinante, através da ligação de um fragmento de DNA de interesse com o DNA de um vetor, é um processo direto e relativamente simples. Entretanto, problemas especiais surgem quando o fragmento de interesse constitui uma fração pequena do DNA total. Este é o caso encontrado comumente quando o objetivo é o isolamento de genes presentes em uma única cópia num genoma complexo, ou quando o objetivo é o isolamento de clones portadores do DNA complementar à RNA mensageiro raro. Deste modo, é claro que quando queremos clonar um determinado gene ou o DNA complementar da mensagem ou mensagens por ele codificado(s) é necessário à obtenção de coleções de clones recombinantes, portadores de

moléculas representantes de todo o genoma, ou de coleções de clones de cDNAs derivados de toda a população de mensageiros da célula ou tecido de interesse. Essas coleções de clones de DNA recombinante são chamadas de bibliotecas: **Biblioteca Genômica**, no caso dos clones terem sido obtidos a partir do DNA genômico ou **Biblioteca de cDNA**, no caso dos clones terem sido construídos a partir de DNA complementar.

O DNA de organismos superiores é bastante complexo: por exemplo, o genoma haplóide de mamífero é composto de aproximadamente 3×10^9 pares de bases (pb). Portanto, se o fragmento de interesse tiver 3000 pb, ele compreenderá somente 1 parte em 10^6 de uma preparação do DNA total. De modo similar, uma espécie de RNAm particularmente rara pode compreender somente 1 parte em 10^5 ou 10^6 da fração de RNA mensageiros de uma célula. Deste modo, para que seja garantida a presença na biblioteca de pelo menos uma versão de todas as seqüências da população alvo, um dos pontos principais na construção de bibliotecas úteis é a obtenção de grandes quantidades de clones. Embora a solução deste problema envolva estratégias específicas no preparo do DNA alvo e na escolha do vetor de clonagem, em linhas gerais a construção de bibliotecas genômicas e de cDNAs segue um procedimento básico bastante semelhante. Nos dois casos, o DNA genômico ou os cDNAs são preparados para a inserção no vetor escolhido. O vetor e o DNA alvo são ligados e introduzidos em *E. coli* através de infecção ou em alguns casos, por transformação direta (Figura 20).

1. BIBLIOTECA DE cDNA

A descoberta da enzima transcriptase reversa, uma DNA polimerase dependente de RNA encontrada predominantemente em vírus de RNA de tumor, tornou possível a síntese *in vitro* de DNA usando-se como molde RNAm. O DNA produto dessa reação é o DNA complementar ou **cópia da mensagem**, abreviadamente chamado de cDNA. A síntese e possível clonagem de cDNAs contribuíram para grandes avanços na análise da estrutura e expressão de genes de eucariotos e propiciaram uma revolução tecnológica nas indústrias farmacêutica e de alimentos. Esses clones são comumente isolados de bibliotecas de cDNA, construídas a partir da população de RNAm isolada da célula ou do tecido de interesse. Portanto, em comparação com bibliotecas do genoma total essas bibliotecas são enriquecidas com seqüências gênicas. Uma vez que o cDNA é cópia da mensagem, ele não contém introns e apresenta significativamente poucas seqüências que não codificam para

proteínas (Figura 21). A ausência de introns permite que cDNAs relativos a células de eucariotos superiores sejam diretamente transcritos e traduzidos em bactérias e em outros microorganismos, permitindo assim a produção em grande escala de polipeptídios importantes tanto na área de pesquisa como na área aplicada. A ausência de introns também facilita a determinação direta da seqüência da mensagem e subsequente dedução da seqüência de aminoácidos da proteína por ela codificada. Isso tem resultado na identificação da seqüência de uma enorme quantidade de proteínas, algumas vezes mesmo antes de terem sido identificadas genética ou bioquimicamente.

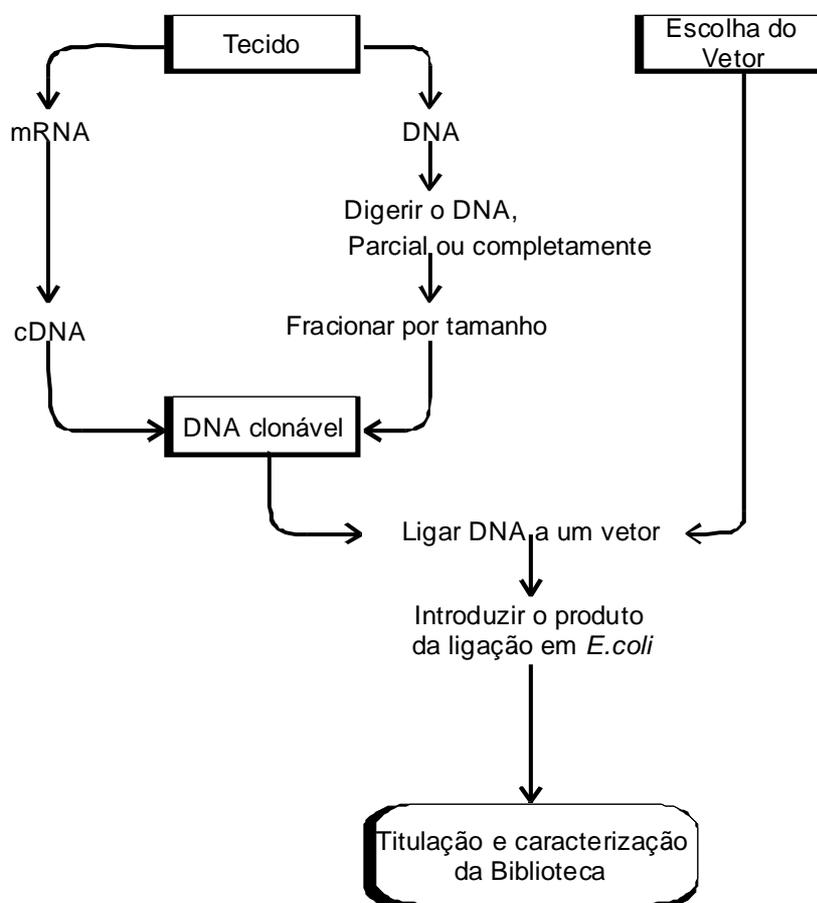


Figura 20. Passos envolvidos na construção de Bibliotecas de cDNA e Bibliotecas genômicas.

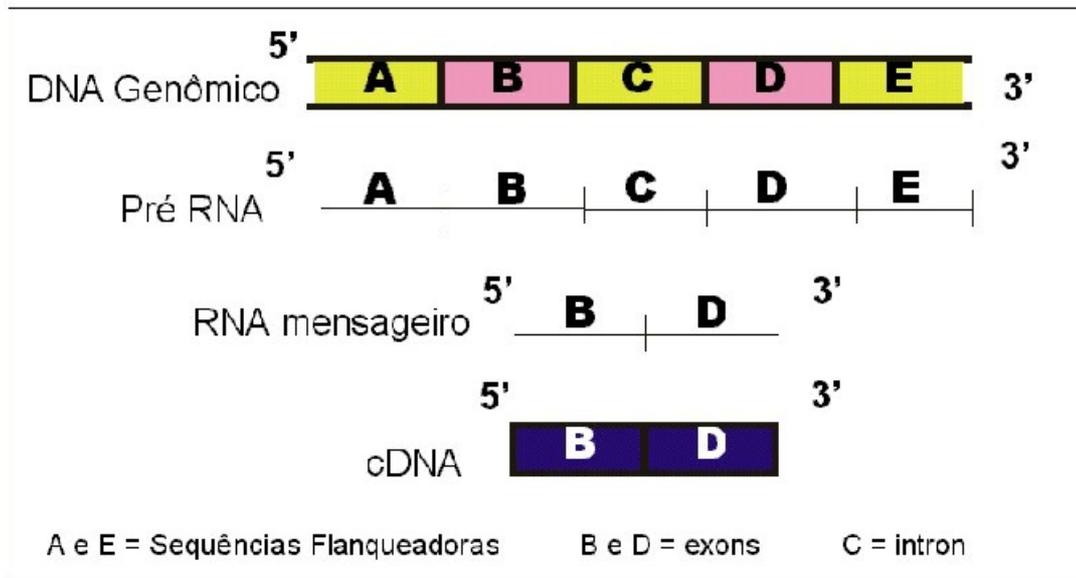


Figura 21. Diagrama representando de forma simplificada as diferenças entre um fragmento de DNA genômico e do cDNA relativo a ele.

De modo resumido, o procedimento para a síntese, clonagem e isolamento de cDNAs específicos envolve 4 etapas:

- 1) Síntese *in vitro* de cDNAs de fita dupla a partir de RNAs mensageiros isolados do tecido de interesse.
- 2) Preparo dos cDNAs para a ligação com o vetor escolhido.
- 3) Introdução dos cDNA recombinantes nas células hospedeiras, onde serão replicados. Isto resultará na amplificação dos recombinantes e dependendo do vetor utilizado, na expressão das proteínas codificadas pelas mensagens que deram origem aos cDNAs.
- 4) Identificação dos clones de interesse através de um processo de triagem dos clones recombinantes.

1.1. Síntese de cDNAs de fita dupla

A construção de uma biblioteca de cDNA inicia-se com o isolamento do RNA total do tecido e subsequente seleção da fração de RNA poli(A)⁺, que compreende a maioria das mensagens presentes na célula. O isolamento de RNA poli(A)⁺ de boa qualidade é

essencial, uma vez que qualquer degradação do RNAm afetará a representatividade das seqüências na biblioteca.

Uma vez obtida a fração de RNA poli(A)+, procede-se à síntese dos cDNAs através de uma série de reações enzimáticas (Figura 22). Nos últimos dez anos foram descritos vários métodos para síntese de cDNA, cada um apresentando vantagens e desvantagens particulares. Como exemplo, apresentamos a seguir um procedimento amplamente usado para esse fim:

a) Síntese da fita de DNA complementar ao RNAm (cDNA) através da enzima transcriptase reversa. Essa enzima é capaz de catalisar *in vitro* a polimerização de DNA a partir RNA e requer um "primer" com a extremidade 3'OH livre para iniciar a síntese. Na maioria das vezes, a molécula utilizada como "primer" é um oligo (dT), que se anela na cauda poli (A)+ do RNA.

b) Após a síntese dessa fita de cDNA, o RNAm é parcialmente removido, pelo uso da RNase A para cortar pedaços do RNA da molécula híbrida RNA-DNA.

c) Utilizando como "primer" os fragmentos de RNA não digeridos pela RNaseA, a DNA polimerase de *E. coli* substitui eficientemente o RNA por DNA, resultando em uma molécula de cDNA de fita dupla.

d) O cDNA fita dupla é tratado com T4 polimerase. Através da atividade 3'-5' exonucleásica dessa enzima remove-se possíveis nucleotídeos extras nas extremidades 3'.

e) O produto final desse conjunto de reações é uma molécula de DNA de fita dupla, cópia exata da mensagem.

1.2. Preparo dos cDNAs para a ligação com o vetor

Uma vez obtido o cDNA, o próximo passo é prepará-lo para a inserção em um vetor de clonagem (Figura. 22). Para isso também existem vários métodos disponíveis, que diferem entre si dependendo do tipo de vetor escolhido: plasmídeo ou fago. O fago λ tem sido o vetor de escolha para a construção de bibliotecas de cDNA. A razão básica desta preferência é a eficiência. Em comparação com plasmídios, o fago λ requer cerca de 16 vezes menos cDNA por μg do vetor para produzir número equivalente de clones recombinantes. Além disso, bibliotecas em fago λ são mais fáceis de serem manipuladas do

que bibliotecas em plasmídios. Entretanto, uma desvantagem dos vetores λ é que não são favoráveis para procedimentos de seqüenciamento do DNA e de mutagênese dirigida. Mais recentemente, outro tipo de vetor foi idealizado para suprir essas deficiências. Essa classe de vetores compreende os fagemídios que, como o próprio nome sugere, são plasmídios contendo porções do genoma de fago e que reúnem vantagens desses dois vetores.

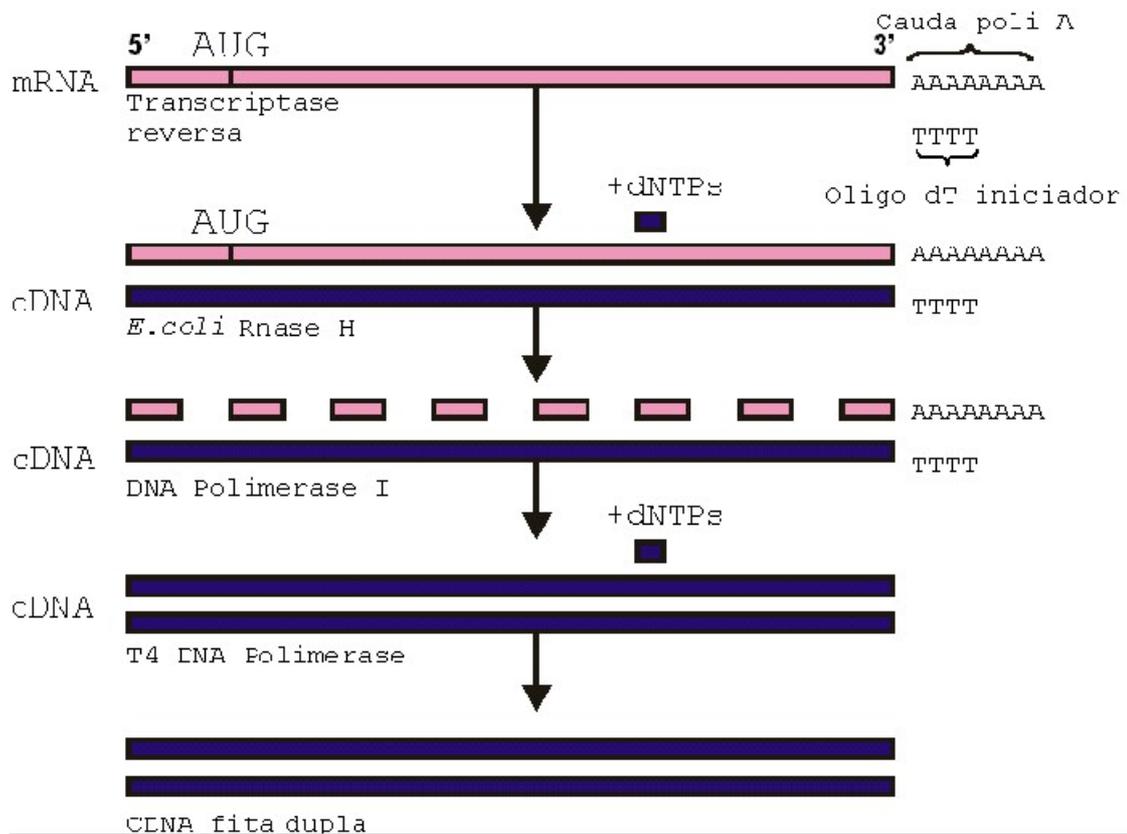


Figura 22. Síntese de cDNA fita dupla.

Para dar uma noção dos passos implicados na clonagem do cDNA relatamos a seguir um procedimento adotado para a clonagem de cDNA em fago λ .

- metilação dos sítios de *EcoRI* através do tratamento do cDNA com *EcoRI*, na presença de S-adenosil metionina. Esse passo é essencial para proteger os sítios *EcoRI* que possam existir no cDNA contra a ação da *EcoRI* em digestão a ser realizada subsequente no processo de clonagem.
- adição de adaptadores, com o sítio *EcoRI*, nas duas extremidades do cDNA. Essa reação resulta em moléculas de cDNA com adaptadores múltiplos ligados nas duas pontas (ver figura 23).

- um único sítio de *EcoRI* é produzido em cada ponta do cDNA pela digestão com *EcoRI*, liberando o excesso de adaptadores ligados na molécula. A metilação prévia do cDNA garante que não haja digestão interna do cDNA.
- antes de ser inserido no vetor, o cDNA é separado do excesso de moléculas de adaptadores. Isso é feito através de filtração em colunas de Sephadex.
- as moléculas de cDNA são ligadas ao vetor, através de reação com T4 ligase.
- o produto da ligação é empacotado com as proteínas formadora do capsídeo do fago.
- os fagos obtidos são utilizados para infectar bactéria de linhagem que favorece o crescimento dos recombinantes.
- após a infecção da bactéria com os cDNA recombinantes temos como produto uma biblioteca de cDNA, que deve conter cópias de toda as seqüências de RNAm da população original. Essa biblioteca pode ser estocada na geladeira, ou submetida a uma triagem em busca do clone de interesse.

2. BIBLIOTECA GENÔMICA

Para se obter informações sobre a estrutura molecular de um gene de eucariotos é necessário o isolamento da porção do DNA genômico que contenha toda a unidade de transcrição, mais as regiões flaqueadoras onde se localizam os elementos controladores da expressão desse gene. O modo mais eficiente para se conseguir isolar essa porção do DNA é através do uso de uma biblioteca genômica, que deve conter clones portando fragmentos de DNA representantes de todo o genoma do organismo em questão. Deste modo, o aspecto principal considerado na construção de uma biblioteca genômica está na obtenção de um número grande de clones portando fragmentos do genoma, obtidos aleatoriamente. O tamanho necessário de uma biblioteca genômica, para que um dado fragmento de interesse esteja entre os fragmentos clonados, é determinado pelo tamanho do fragmento clonado e pelo tamanho do genoma. Deste modo, a estratégia básica na construção de bibliotecas genômicas busca minimizar o número de clones necessários através da clonagem de fragmentos de DNA de maior tamanho possível, e maximizar a eficiência da clonagem,

através da utilização de vetores baseados no fago λ . Basicamente, a construção da biblioteca genômica envolve os seguintes passos:

- isolamento de DNA de alto peso molecular, que é posteriormente quebrado de modo a produzir fragmentos de tamanho compatível com o vetor de clonagem.
- ligação desses fragmentos no vetor e introdução dos recombinantes obtidos nas células hospedeiras.

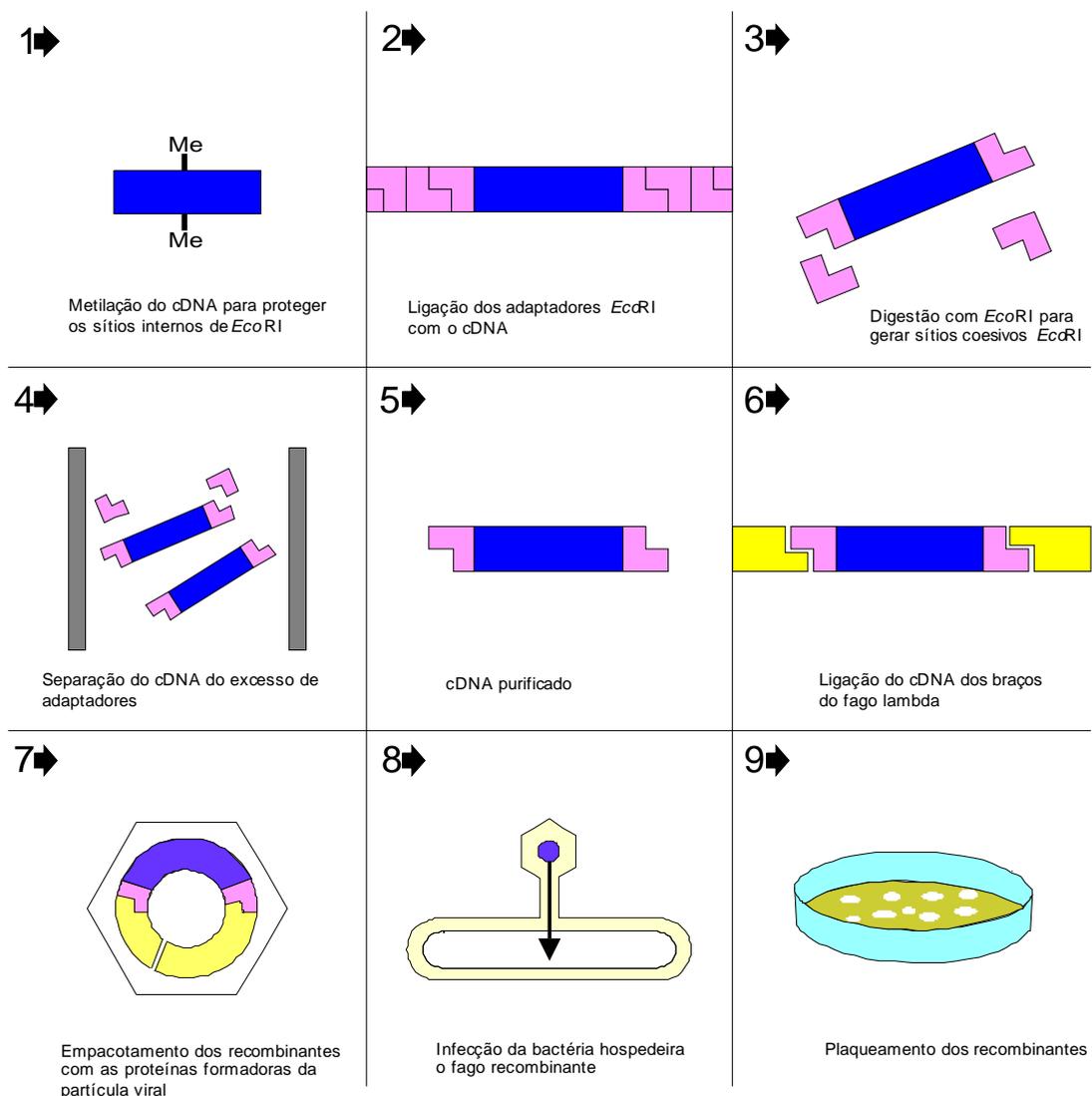


Figura 23. Clonagem de cDNA em fago λ .

Dentre esses passos, o modo de preparo dos fragmentos a serem clonados (insertos) merece ser discutido criticamente dada sua importância na representatividade da biblioteca.

2.1. Preparo do DNA do inserto

A representatividade de uma biblioteca genômica depende grandemente da aleatoriedade com que os fragmentos a serem clonados são gerados, para que não haja exclusão sistemática de nenhuma seqüência. O fracionamento mecânico do DNA é o meio de se obter completa aleatoriedade na fragmentação do DNA. Entretanto, esse método não é comumente adotado devido à trabalhosa manipulação necessária para que os fragmentos assim obtidos possam ser inseridos no vetor. Com bom senso e cuidado é possível conseguir fragmentação suficientemente aleatória através de digestões parciais do DNA com enzimas de restrição. Uma vantagem em se utilizar a digestão parcial do DNA é a produção de fragmentos com pontas coesivas e que, portanto, podem ser diretamente ligados ao vetor. Para garantir que a digestão atinja todo o genoma são utilizadas enzimas de restrição que cortam freqüentemente o DNA e que não apresentam desvios de distribuição de sítios. A enzima *Sau3A I*, que reconhece o sítio de 4 bases GATC, e que gera fragmentos compatíveis com sítios de clonagem de vários fagos e cosmídios, tem provado ser bastante útil na produção de bibliotecas de DNA genômico digerido parcialmente. Após a digestão parcial, os fragmentos gerados precisam ser fracionados antes de serem ligados ao vetor. Sem esse fracionamento, fragmentos pequenos poderão ligar-se entre si produzindo falsos recombinantes. Fragmentos muito grandes que não permitem o crescimento do fago poderão ligar-se aos braços do fago, alterando a relação entre as quantidades de DNA de inserto e vetor.

Um método de fracionamento freqüentemente adotado é o de centrifugação em gradiente de sacarose. Após a digestão parcial, a solução de DNA é aquecida para que possíveis fragmentos agregados se dissociem, e então, é aplicada sobre um gradiente de sacarose de alta salinidade. Após a centrifugação, são coletadas frações desse gradiente. Essas frações são analisadas por eletroforese em um gel de agarose. As frações contendo os fragmentos de DNA de tamanho apropriado são utilizadas para a ligação com o vetor. A figura abaixo ilustra esse procedimento.

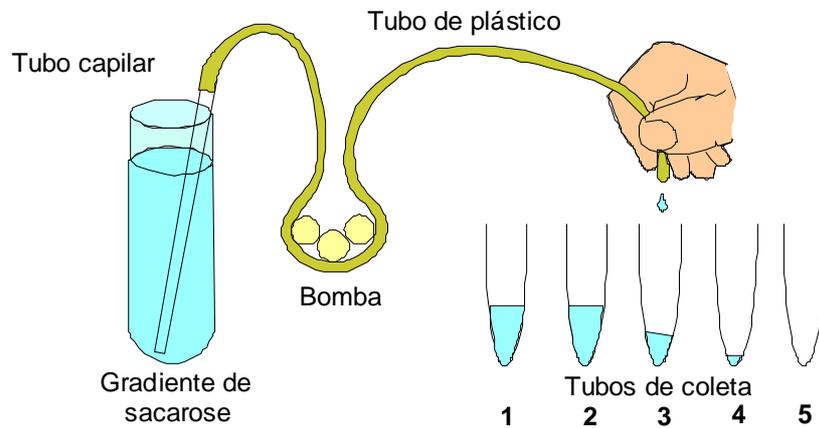


Figura 24. Método para fracionamento do DNA por centrifugação em gradiente de sacarose. O gradiente é aspirado, de baixo para cima, com ajuda de uma bomba peristáltica. A porção mais densa do gradiente contendo o DNA de maior peso molecular, é aspirada primeiro.

Uma vez garantida a aleatoriedade dos fragmentos clonados, é bastante razoável se pensar que a probabilidade de uma dada seqüência de interesse estar presente em uma biblioteca genômica possa ser estimada estatisticamente, com base na distribuição de Poisson. Especificamente, o número de clone independentes (N) que precisam ser triados para uma probabilidade (P) de se isolar uma seqüência específica é dada por:

$$N = \ln(1-P) / \ln[1-(I/G)]$$

onde I é o tamanho médio dos fragmentos clonados, em pares de base, e G é o tamanho do genoma, em pares de base.

Deste modo, precisamos triar cerca de 690.000 clones de uma biblioteca que usa como vetor um fago λ , para termos 99% de chance de isolar qualquer dada seqüência presente em uma única cópia em um genoma típico de mamífero.

$$N = \ln(1-0,99) / \ln[1- (2 \times 10^4 / 3 \times 10^9)] = 690.000$$

2.2. Vetores utilizados na construção biblioteca genômica

Fagos λ e cosmídeos são comumente usados na construção de bibliotecas genômicas. Nos dois casos, fragmentos grandes de DNA gerados por fragmentação aleatória são ligados com o DNA do vetor para formar concatâmeros que podem então ser empacotados em partículas de fago λ . As bibliotecas construídas em vetores λ são armazenadas e propagadas

Como todo fago λ , a extremidade de cada braço apresenta um segmento de fita simples (cos). Esses segmentos são complementares entre si e permitem a recircularização da molécula, e a conseqüente replicação do DNA do fago dentro da célula hospedeira.

Esse tipo de vetor é chamado de vetor de substituição por que no processo de clonagem a parte central do DNA do fago é substituída pelo fragmento de DNA a ser clonado, originando a molécula recombinante.

Mais recentemente outros vetores com capacidade para grandes insertos foram desenvolvidos. Os **cromossomos artificiais de levedura (YACs)** são moléculas lineares de DNA cuja arquitetura mimetiza a estrutura de um cromossomo autêntico (Figura 26). Os YACs recombinantes são criados pela ligação de fragmentos grandes de DNA genômico exógeno (até 2000 kb) com os dois "braços" do vetor YAC e o produto da ligação é introduzido na levedura por transformação. Nos YACs recombinantes, o segmento de DNA genômico exógeno é flanqueado de um lado por um centrômero, uma origem de replicação e uma marca de seleção e pelo outro lado, por uma segunda marca de seleção. As leveduras transformantes com cromossomo artificial são selecionadas como colônias em meio com as drogas de seleção.

Os **cromossomos artificiais de bactérias (BAC)** são moléculas de DNA circulares que carregam uma marca de seleção a antibiótico. Os segmentos de DNA genômico exógeno são clonados no vetor BAC *in vitro* e o produto da reação de ligação é introduzido por eletroporação em cepas de *E. coli*, onde é mantido como um plasmídio de cópia única. Os vetores BACs carregam marcadores que permitem a discriminação dos clones vazios e cheios. A média de tamanho dos clones de BACs é 120 kb, mas alguns clones podem chegar 300kb.

Os vetores de **bacteriófago P1** acomodam fragmentos genômicos de 70 a 100Kb. Neste sistema, as moléculas lineares recombinantes geradas a partir de seqüências do vetor e do genoma são empacotadas *in vitro* em bacteriófago P1. Após a injeção em *E. coli*, as moléculas se tornam circulares pela atividade de uma recombinase, que promove a recombinação de seqüências específicas presentes no vetor. Estes vetores carregam marca de seleção para antibiótico, marca de seleção positiva para seleção dos clones com inserto de DNA genômico exógeno e uma origem de replicação plasmídial P1, que mantém uma ou duas cópias dos plasmídios recombinantes na célula. Uma segunda origem de replicação pode ser ativada para amplificação dos plasmídios antes do isolamento do DNA.

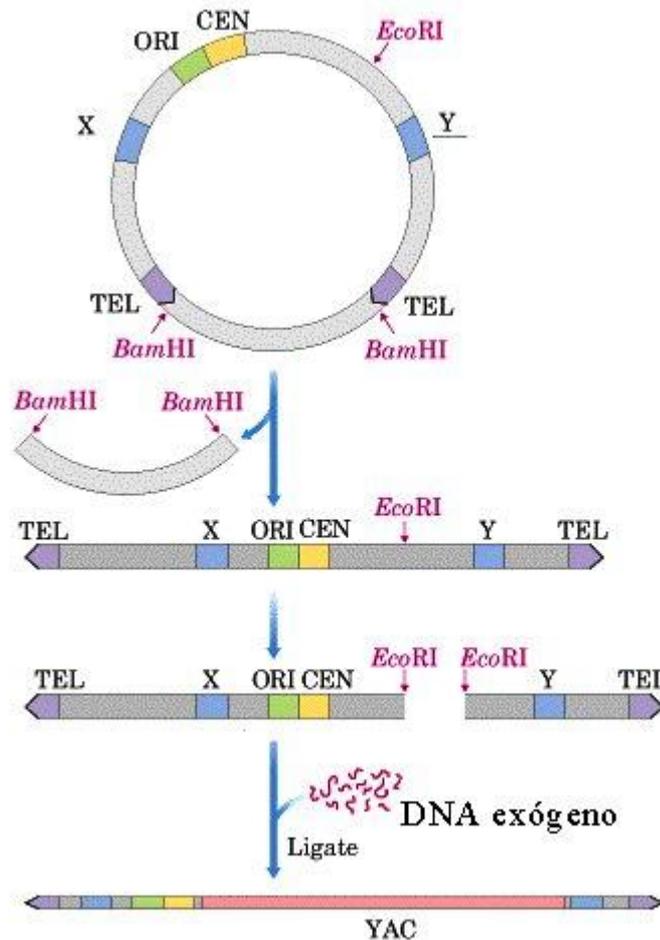


Figura 26. Construindo um cromossomo artificial de levedura (YAC). O vetor YAC carrega uma origem de replicação (ORI), um centrômero (CEN), dois telômeros e marcas de seleção (X e Y). Lehninger, Principles of Biochemistry. Figura 29-16, p. 1139.

Os **cromossomos artificiais P1** combinam as características dos bacteriófagos P1 e dos BACs, incluindo a marca de seleção positiva e as origens de replicação do bacteriófago P1. Os clones circulares recombinantes de PACs gerados durante a reação de ligação in vitro são introduzidos em *E.coli* por eletroporação e mantidos como um plasmídeo de cópia única na célula. Os insertos da biblioteca de DNA do genoma humano em PACs variam de 60kb a 150kb.

IV. DETECÇÃO E ANÁLISE DO CLONE RECOMBINANTE

Um passo crucial na Tecnologia do DNA recombinante é encontrar o clone correto na mistura de clones que foi criada durante os procedimentos da confecção da biblioteca genômica ou de cDNA. Embora seja relativamente fácil selecionar as colônias que abrigam os vetores de clonagem usando os marcadores de resistência a antibióticos que estão presentes nos vetores (Figura 27A), é muito mais difícil selecionar o clone que contenha o gene de interesse. Note que no caso do vetor ser um plasmídeo deverão ser examinadas milhares de colônias de bactérias crescidas em placas de petri contendo agar, para a detecção daquela(s) colônia(s) que produza(m) pequenas quantidades da proteína expressa pelo gene de interesse ou então que possua(m) o gene de interesse sem qualquer expressão protéica que o caracterize (Figura 27B). Será discutida inicialmente a situação na qual a proteína é expressa no hospedeiro. Em seguida será discutida a situação mais comum onde a proteína não é expressa e a análise será feita através da própria presença do DNA de interesse no fago ou na bactéria.

1. QUANDO O GENE DE INTERESSE É EXPRESSO NO HOSPEDEIRO

Se o gene de interesse é capaz de se expressar na célula hospedeira pode-se identificar o clone que carrega este gene pela presença desta proteína entre as colônias recombinantes. Para isto, o hospedeiro não deve produzir esta proteína, a menos que ele carregue o clone que a produza. Se esta proteína for produzida normalmente pela célula hospedeira, será então necessário o uso de uma célula hospedeira defectiva ou mutante para o gene de interesse. Quando este gene for incorporado ele pode ser detectado por complementação daquela função. Se a proteína a ser detectada for específica de mamífero, por exemplo, a célula hospedeira já será naturalmente defectiva. Se a proteína for uma enzima que catalisa uma reação mensurável pode ser possível triar as colônias por sua atividade enzimática.

Se a proteína de interesse não for uma enzima com função detectável, uma outra estratégia será necessária para a identificação do clone correto, como por exemplo, o uso de um anticorpo específico para a proteína de interesse. Anticorpo é uma proteína do soro produzida pelos mamíferos que se combina com alta especificidade com outra proteína, o antígeno. Neste caso, a proteína de interesse é o antígeno. Como o anticorpo se combina especificamente com o antígeno, se o antígeno estiver presente em uma ou mais colônias de

uma placa de petri, a identificação destas colônias poderá ser feita observando-se a reação antígeno-anticorpo.

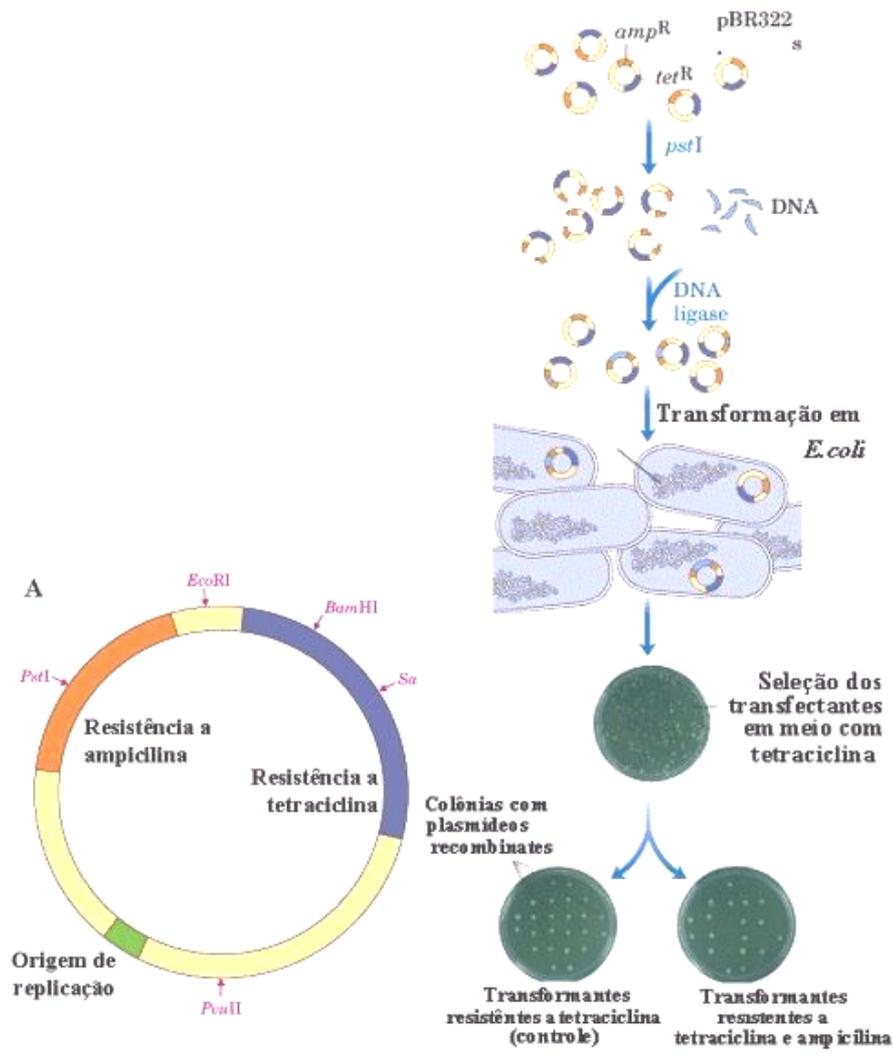


Figura 27. Em A, estrutura do plasmídeo pBR322, um típico vetor de clonagem. Em B, método da inativação insercional para isolar plasmídios contendo DNA exógeno. A inserção do DNA exógeno no plasmídeo pBR322 causa inativação da resistência à tetraciclina, permitindo o isolamento de transformantes contendo o fragmento de DNA clonado (B). Lehninger, Principles of Biochemistry. Figura 29-6, p. 1126.

Para que o antígeno seja produzido na célula hospedeira a biblioteca de cDNA terá de ser construída num **vetor de expressão**. Neste caso os cDNAs são inseridos nestes vetores em regiões que promovam sua expressão em *E.coli*, normalmente em promotores que possam ser regulados. Esta proteína antigênica será expressa como uma **proteína de fusão**, ou seja, será sintetizada subsequente a alguns aminoácidos pertencentes à proteína do vetor bacteriano. Estas proteínas de fusão são geralmente mais estáveis que a

correspondente proteína de eucarioto produzida em bactéria e, portanto podem ser obtidas em grande quantidade. Para visualizar a produção destas proteínas, uma parte dos fagos de cada placa de lise (ou bactérias) recombinantes desenvolvidos numa placa de Petri são transferidos (como um carimbo) para uma membrana de nitrocelulose, tratados para expor suas proteínas e então submetidos a uma solução contendo um anticorpo específico para a proteína desejada. Depois que os anticorpos livres são removidos, um segundo anticorpo é adicionado para localizar a posição do primeiro. O segundo anticorpo é normalmente radioativo e pode ser identificado por autoradiografia utilizando-se um filme de raio X. Se uma colônia radioativa estiver presente, uma mancha no filme de raios-X será observada depois que o filme for revelado (Figura 28). A localização desta mancha no filme corresponde à localização da colônia que produziu aquela proteína, na placa de petri original. Esta colônia será então recuperada da placa original e cultivada.

Uma limitação deste procedimento é que o anticorpo utilizado nesta triagem precisa ser específico para a proteína de interesse. A produção de anticorpos pode ser obtida pela injeção da proteína (antígeno) em um animal. No entanto, esta proteína injetada deve ser pura, caso contrário mais que um anticorpo será formado.

2 O USO DE SONDAS MOLECULARES DE ÁCIDOS NUCLÉICOS PARA IDENTIFICAR O GENE DE INTERESSE

Como pode ser detectado um gene de interesse entre as colônias da biblioteca genômica ou de cDNA, se este gene não é expresso?

Quando o DNA em solução aquosa é aquecido à 100°C, ou exposto à pH alto, as bases complementares que normalmente seguram as duplas hélices juntas são dissociadas em duas fitas simples. Este processo é chamado de *desnaturação*. Estas mesmas fitas simples poderão se reassociar se forem conservadas à 65°C, por longos períodos, num processo chamado de *renaturação* (ou *hibridação*, quando a associação ocorre entre ácidos nucleicos de diferentes origens). Reações de hibridação ocorrem entre quaisquer duas cadeias de ácidos nucleicos de fita simples (DNA:DNA, RNA:DNA ou RNA:RNA) desde que tenham seqüências de nucleotídeos complementares.

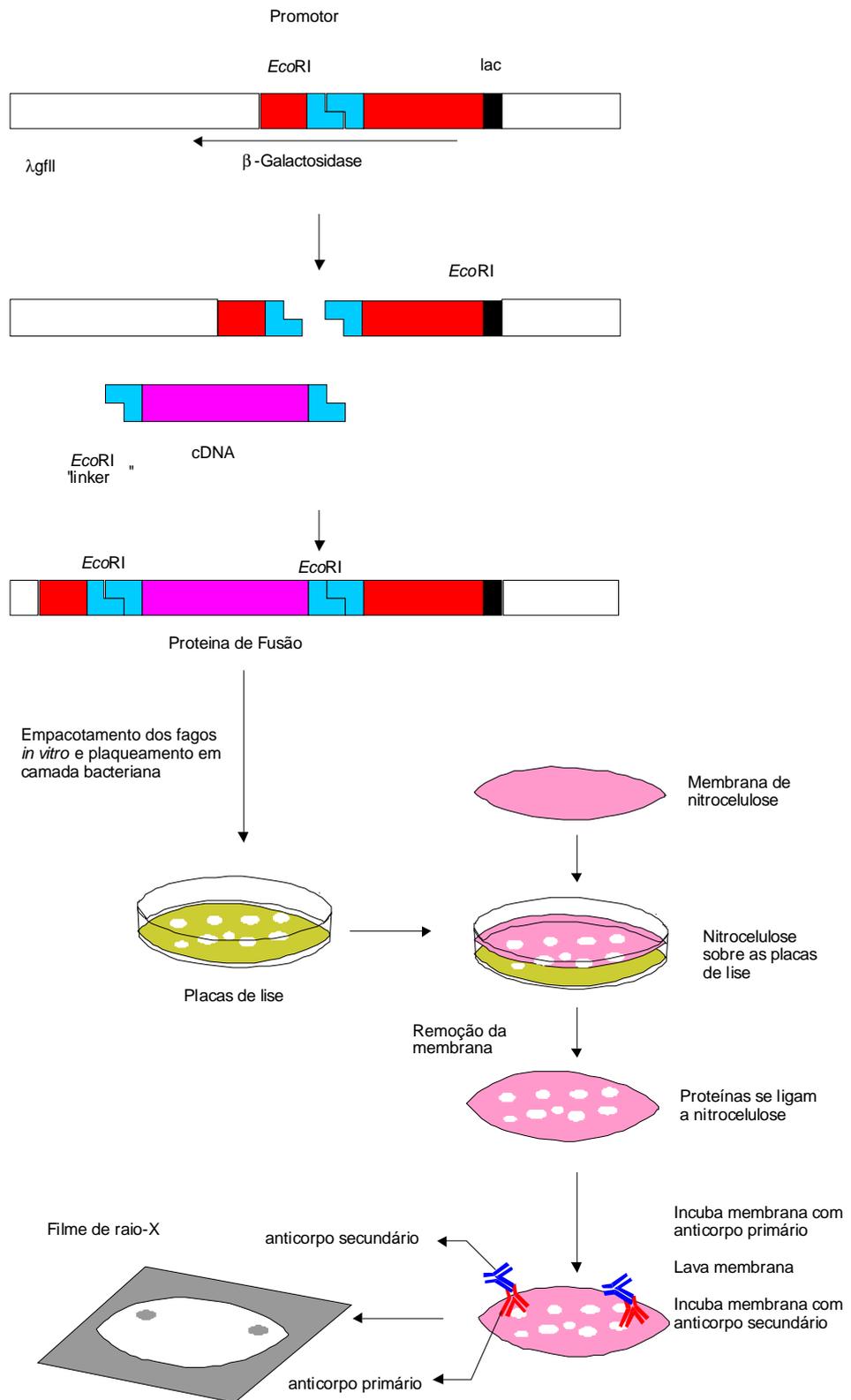


Figura 28. Triagem de bibliotecas de expressão com anticorpos. Se o inserto estiver na orientação correta e na apropriada seqüência aberta de leitura é traduzido em proteína de fusão (antígeno). Os anticorpos identificam as placas de lise específicas destes antígenos.

Este princípio da hibridação molecular é fundamental para caracterizar DNAs recombinantes quando o gene de interesse não é expresso. Sondas de ácidos nucléicos (fragmentos de ácido nucléico marcados radioativamente, geralmente com ^{32}P) são utilizadas para localizar clones, numa biblioteca genômica ou de cDNA, que carreguem uma seqüência de DNA de interesse. Após a confecção da biblioteca genômica (ou de cDNA) os fagos carregando DNA recombinante são espalhados juntamente com uma suspensão de bactérias numa placa de petri contendo meio de cultura sólido. Uma vez visualizada as placas de lise (ou as colônias, no caso do vetor ser um plasmídio) é preparada uma réplica de cada placa de petri com uma membrana de náilon ou de nitrocelulose. Este processo permite a transferência de uma porção de fagos de cada placa de lise (ou de bactérias de cada colônia) para a membrana, de tal maneira que o padrão das placas de lise da placa de petri original seja mantido na membrana. Para identificar qual das placas de lise porta o gene de interesse esta membrana é tratada para expor e desnaturar o DNA das colônias ali presentes e incubados com uma solução de DNA ou RNA fita simples, marcada radioativamente e complementar ao gene de interesse para permitir a hibridação. Dois polinucleotídeos simples fitas somente irão se hibridar se forem complementares. A localização da placa de lise que se hibridou à sonda é determinada após a exposição da membrana a um filme de Raios-X (autoradiografia) e a comparação deste filme com a disposição das colônias na placa de petri original (Figura 29).

Esta sonda molecular radioativa pode ser parte de um gene já clonado para o qual se procura o gene total, pode ser o DNA de um gene vizinho ao gene de interesse, pode ser o RNA de genes que são expressos em tecidos específicos ou em estágios específicos do ciclo celular, ou mesmo o gene de uma outra espécie que codifica para a mesma proteína. Neste último caso, a reação de hibridação pode ser realizada em baixa temperatura (42°C ao invés de 65°C) e em baixa concentração de sal (baixa estringência) para permitir a hibridação mesmo entre DNAs que tenham algumas bases não complementares.

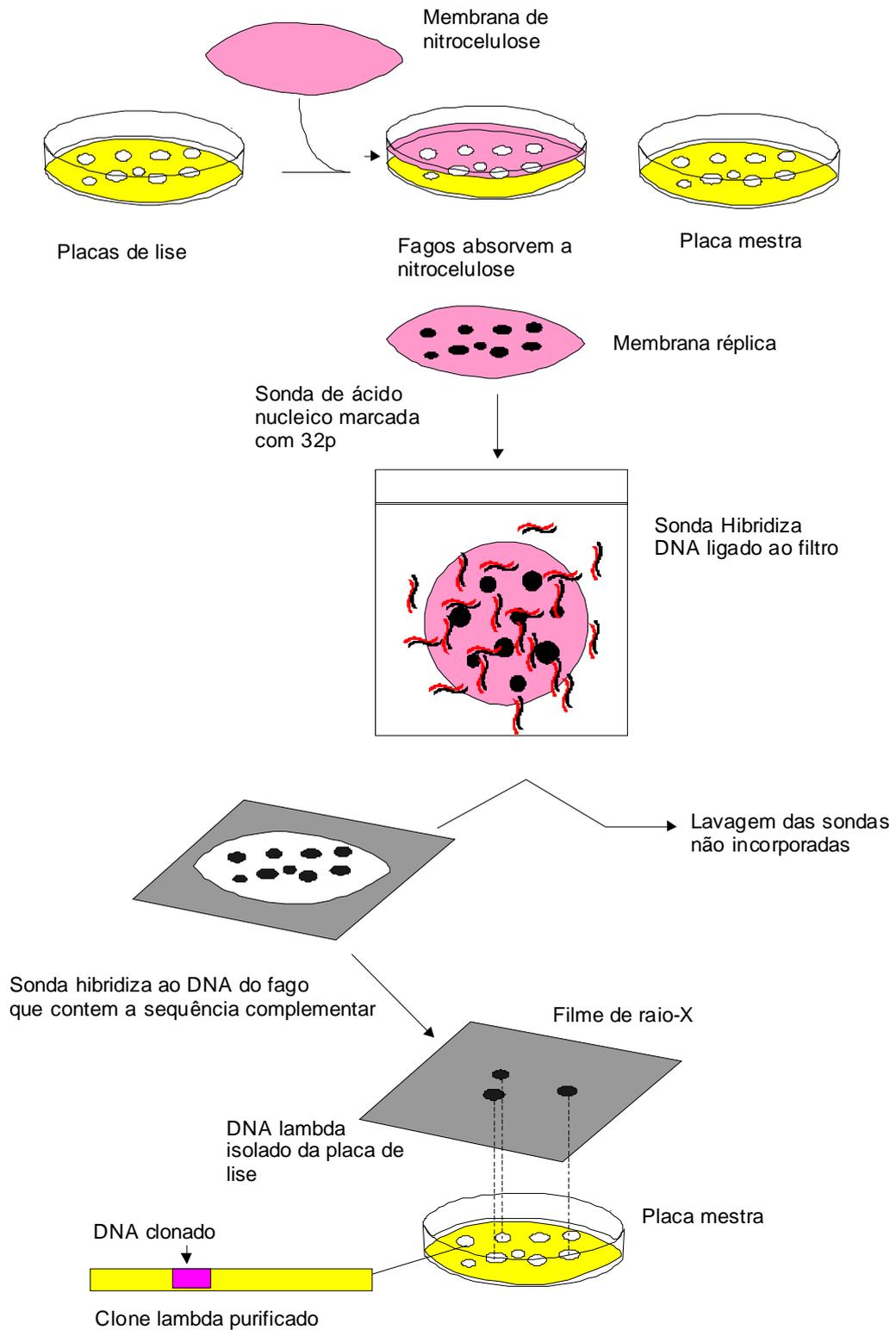


Figura 29. Triagem de uma biblioteca genômica ou de cDNA, construídas no fago λ , com uma sonda molecular para encontrar um clone de interesse. Esta triagem é feita espalhando-se os fagos previamente incubados numa cultura de bactérias, em várias placas de agar. Depois que as placas de lise tornam-se visíveis, são feitas as réplicas em membrana de nitrocelulose, seu DNA é exposto e hibridizado com a sonda molecular. A posição do clone recombinante de interesse é identificada através de autoradiografia. O DNA é isolado dos fagos presentes na placa de lise positiva, identificada na placa de petri original e submetida a análises posteriores.

Genes que respondem a um mesmo mecanismo de regulação podem ser identificados numa biblioteca de cDNA por hibridação diferencial. Isto é, genes expressos em tecidos específicos, em estágios específicos do desenvolvimento ou do ciclo celular e genes regulados por fatores de crescimento são adequados para serem analisados por este tipo de abordagem. Para esta identificação é necessário produzir duas populações celulares (ou extrair mRNA de dois diferentes tecidos) uma na qual eles não se expressam e outra na qual eles se expressam. Suponha por exemplo, que uma biblioteca de cDNA seja preparada usando o mRNA isolado de células tratadas com fatores de crescimento. Esta biblioteca é plaqueada e transferida para dois conjuntos de membranas de nitrocelulose. Um conjunto é ensaiado com DNA marcado radioativamente preparado por transcrição reversa do RNA das células tratadas com os fatores de crescimento. O outro conjunto de filtros é hibridado com cDNA preparado de células não tratadas. Clones positivos identificados por hibridizar mais fortemente com a primeira sonda codificam mRNAs induzíveis pelos fatores de crescimento (Figura 30).

Quando o DNA a ser clonado expressa uma proteína cuja sequência é conhecida pode-se inferir a sequência de um trecho de DNA que lhe deu origem e sintetizar oligonucleotídeos que lhe sejam complementares para serem utilizados como sondas moleculares. Estes nucleotídeos devem ter pelo menos 17 a 20 nucleotídeos de comprimento para ser específico para a sequência procurada. Um problema enfrentado para sintetizar estes oligonucleotídeos é a degenerescência do código genético. Alguns aminoácidos são codificados por quatro ou mesmo seis diferentes códons e, portanto mesmo um pequeno polipeptídeo pode ter sido codificado por várias sequências de DNA. Para contornar este problema as sondas oligonucleotídicas são algumas vezes sintetizadas como uma mistura de todas as possíveis combinações dos códons que são traduzidas naquela sequência protéica (Figura 31). Uma destas sequências é correta e irá hibridar com o clone de interesse, mas a possível hibridação dos outros oligonucleotídeos com sequências sem interesse pode levar a obtenção de clones falsos positivos.

Uma variação deste método é utilizar o menor número possível de oligonucleotídeos como sonda, escolhendo a região da proteína que contém o menor número possível de códons degenerados (por exemplo, metionina é somente codificada pelo códon AUG). Deve-se levar em conta também qual é o códon de uso mais frequente para cada aminoácido, na espécie que está sendo analisada.

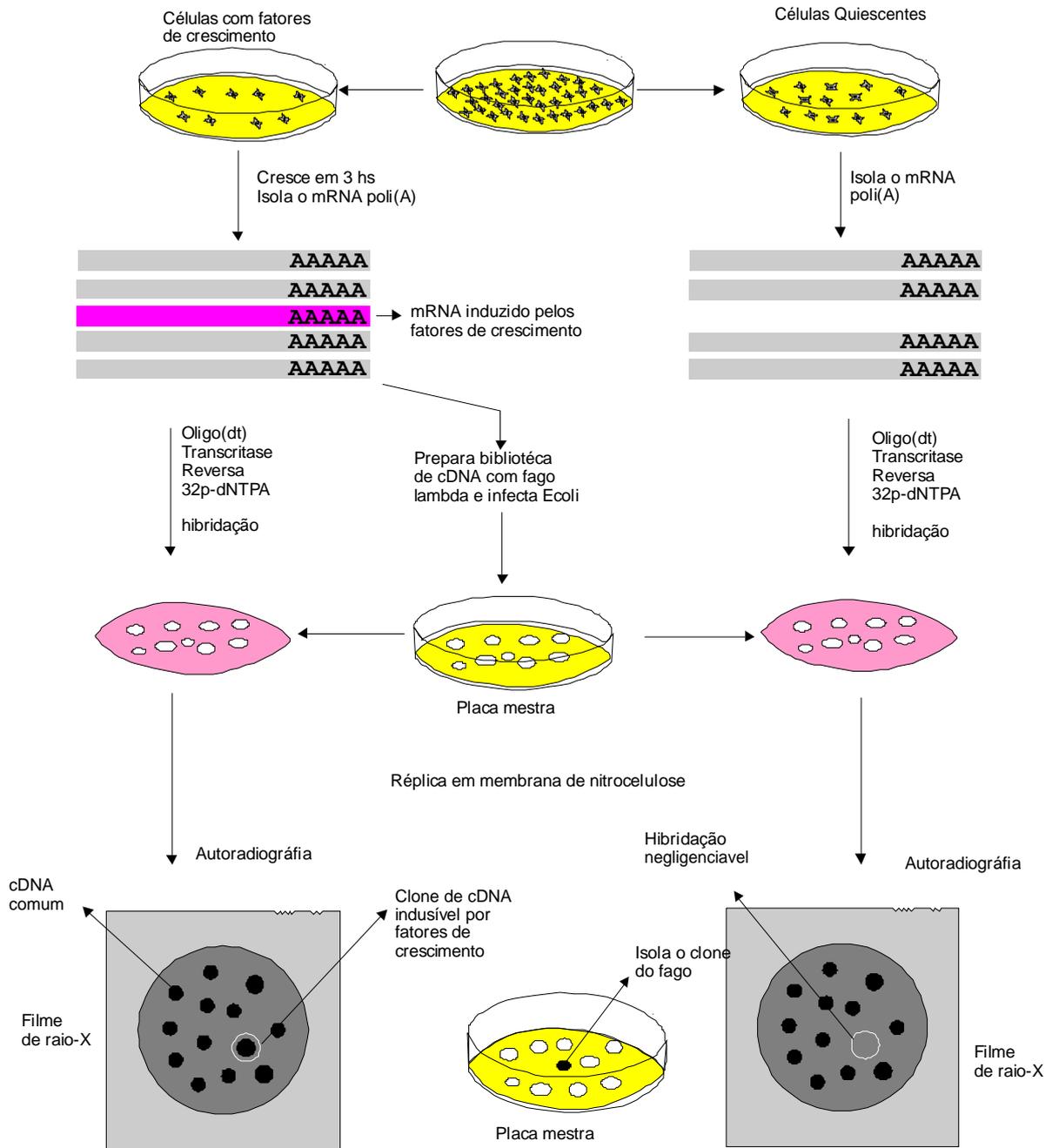
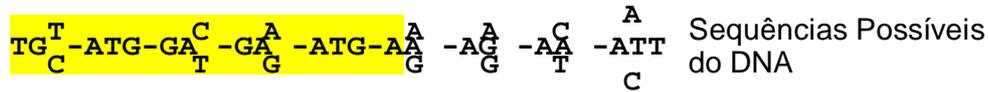


Figura 30. Clonagem de genes regulados por fatores de crescimento através de hibridação diferencial. Esta estratégia é utilizada para clonar uma família de genes que são induzidas quando células quiescentes são estimuladas a crescer pela adição dos fatores de crescimento.

Cys-Met-Asp-Glu-Met-Lys-Arg-Asn-Ile

Sequência Parcial
de Aminoácidos



Parte da sequência com
pouca ambigüidade

A
CGC
G
T

Sondas contendo
todas as possíveis combinações
de condons de parte
da sequência

Sonda tentativa

TGTATGGATGAAATGAA
TGCATGGATGAAATGAA
TGTATGGACGAAATGAA
TGCATGGACGAAATGAA
TGTATGGATGAGATGAA
TGCATGGACGAGATGAA
TGCATGGATGAGATGAA
TGTATGGATGAGATGAA

5' TGCATGGACGAGATGAAGCGCAACATG 3'

Hibridação com cDna

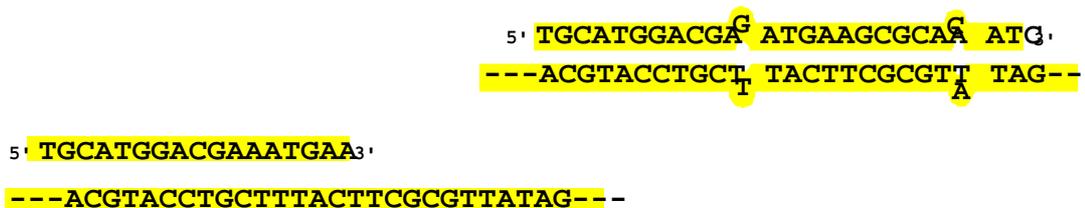


Figura 31. Desenho de sondas oligonucleotídicas baseadas na sequência protéica. Como a maioria dos aminoácidos é especificada por dois ou mais códonos estes oligonucleotídeos são sintetizados usando-se uma mistura dos nucleotídeos das posições ambíguas. A sequência correta estará representada entre os oligos. Uma outra possibilidade é usar uma sonda tentativa cuja escolha é baseada na frequência de uso do códon na espécie em estudo. Neste caso pareamentos incompletos podem ocorrer.

3 ANÁLISE DO DNA E RNA POR ELETROFORESE EM GEL E “BLOTTING”

Várias técnicas foram desenvolvidas baseadas nas propriedades de hibridação dos ácidos nucléicos visando a triagem e isolamento de sequências específicas destas moléculas. A maioria destas técnicas, conforme já mencionado, trabalha com uma réplica do DNA de interesse imobilizado num suporte sólido tal como uma membrana de náilon ou de

nitrocelulose. Um destes métodos, desenvolvido na década de 70 por E. M. Southern, tornou-se conhecido por "*Southern blotting*". Nesta técnica o DNA é digerido com uma ou mais enzimas de restrição e os fragmentos resultantes separados por eletroforese num gel de agarose. Os fragmentos de DNA dupla fita são visualizados por coloração com brometo de etídio, desnaturados 'in situ' com hidróxido de sódio e transferidos para uma membrana por capilaridade, com o auxílio de uma solução de alta concentração salina. Tais condições permitem que o DNA seja retido na membrana no ponto de contacto entre o gel e a membrana, criando uma réplica do gel. O DNA é covalentemente ligado à membrana usando-se calor ou luz ultravioleta. Sondas de ácidos nucleicos (DNA ou RNA marcados radioativamente) podem ser utilizadas para hibridar com o DNA fixado na membrana e a posição na qual a ligação específica ocorrer pode ser detectada por autoradiografia da membrana (Figura 32). O padrão de hibridação pode ser comparado diretamente com a região no gel original que contém a seqüência de DNA de interesse. Southern blotting é uma técnica tão poderosa que permite que as informações obtidas sejam utilizadas para a construção de um mapa de restrição de uma determinada parte do DNA clonado, por exemplo, para identificar a região que contém o gene de interesse. Esta técnica possibilita também que as sondas de ácidos nucleicos sejam utilizadas para diagnósticos de distúrbios genéticos, ensaiando-se o DNA genômico dos indivíduos afetados e de seus familiares. De modo análogo, moléculas de RNA também podem ser identificadas após serem separadas por eletroforese, transferidas para nitrocelulose, sofrerem hibridação com sondas de DNA ou RNA. Esta técnica de analisar RNA, por analogia ao Southern blotting, recebeu o nome de *Northern blotting*. Esta técnica é útil como auxiliar na análise de clones de cDNA porque, o tamanho do mRNA específico pode ser comparado com o tamanho dos cDNAs clonados, revelando a integridade dos clones. Além disto, esta técnica permite indicar em qual tecido ou tipo celular um determinado gene é expresso ou quais são os fatores que regulam a sua expressão.

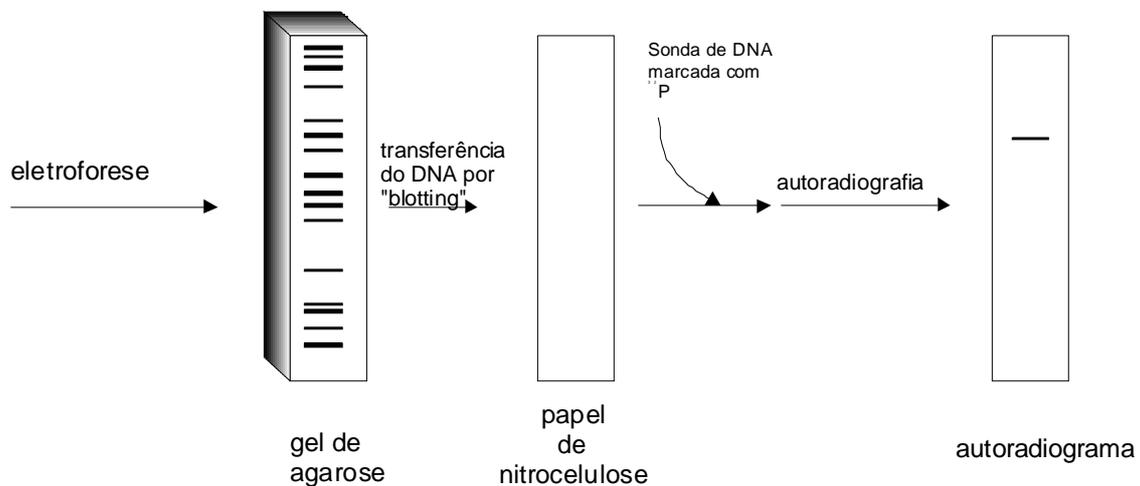


Figura 32. Southern blotting. Um fragmento de DNA específico pode ser identificado separando-se a mistura de fragmentos por eletroforese, transferindo-se para nitrocelulose e hibridizando-se com uma sonda molecular complementar à seqüência e marcada com ^{32}P .

4 COMO OS GENES CLONADOS SÃO UTILIZADOS?

O clone de interesse pode ser explorado sob vários aspectos. Ele pode ser seqüenciado e comparado com outras seqüências já descritas, ser utilizado no estudo da expressão gênica do(s) gene(s) contido(s) no clone, ser alterado especificamente por mutagênese sítio dirigida ou ser usado para gerar um produto de interesse comercial. Para isto quase sempre será necessário transferir o clone, ou parte deste para um vetor mais apropriado para atingir os objetivos desejados. A transferência de parte do DNA clonado para um novo vetor é conhecida como sub clonagem. Algumas destas aplicações serão abordadas nos capítulos seguintes.

5. MAPA DE RESTRIÇÃO

O mapa de restrição do DNA clonado ou de pequenas moléculas de DNAs de fagos, plasmídios ou mitocôndrias pode ser bastante útil na caracterização da molécula. Esta técnica envolve a separação por eletroforese dos fragmentos de DNA obtidos após digestão do DNA total com determinada enzima de restrição. A ordem dos fragmentos na molécula pode ser descoberta com o uso de outras enzimas de restrição, em duplas digestões e/ou por digestões seqüenciais usando 2 enzimas. Assim cada fragmento obtido após digestão com a

enzima x é removido do gel e submetido à digestão com a enzima y e vice-versa (Figura 33).

Informações sobre o mapa de restrição de um fragmento de DNA são utilizadas, por exemplo, para subclonar partes deste fragmento para o seqüenciamento. O mapa de restrição do DNA mitocondrial vem sendo útil na caracterização de linhagens e/ou espécies de vários organismos, auxiliando em estudos taxonômicos e evolutivos.

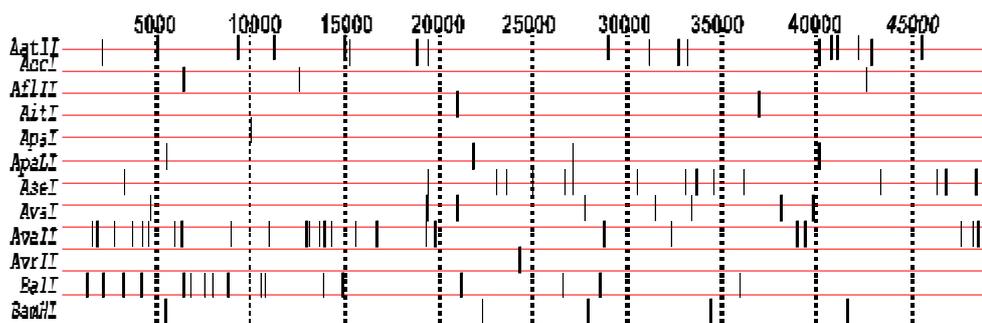


Figura 33. Mapa de restrição do DNA do fago λ para várias endonucleases. As barras verticais indicam os locais de clivagem de cada enzima e os números no alto indicam o tamanho da molécula em unidades de 5000 pares de bases. Esta molécula possui no total 48502 pares de bases.

V. REAÇÃO EM CADEIA DA POLIMERASE (PCR)

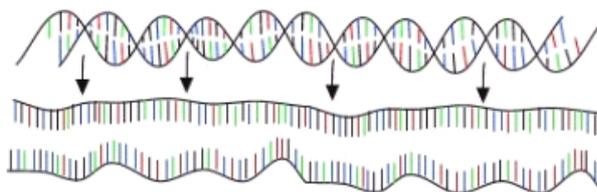
A reação em cadeia da polimerase possibilita a amplificação de uma seqüência rara de DNA a partir de uma mistura complexa, sem a necessidade de clonagem molecular. Esta técnica é amplamente utilizada em pesquisa básica, em medicina forense e no diagnóstico de doenças genéticas e infecciosas.

Inicialmente, é necessária a construção por síntese química de dois oligonucleotídeos de DNA (iniciadores) complementares as extremidades de cada fita de DNA, flanqueando a região de interesse. Estes oligonucleotídeos servem como iniciadores da síntese de DNA *in vitro*, que é catalisada pela DNA polimerase.

Um ciclo de PCR começa com a desnaturação por calor (95°C), que promove a separação da fita dupla de DNA. A reação é resfriada na presença de um excesso dos dois oligonucleotídeos, possibilitando a hibridização dos dois iniciadores com a seqüência complementar presente no DNA alvo. Em seguida, a reação é incubada para atividade da DNA polimerase, produzindo novas fitas de DNAs a partir dos iniciadores e utilizando o quatro desoxirribonucleotídios (dATP, dCTP, dGTP e dTTP) (Figura 34).

30 a 40 de ciclos de 3 etapas:

1) Desnaturação
1 minuto 94 °C



2) Hibridização dos iniciadores
45 segundos 54 °C



3) Extensão
2 minutos 72 °C

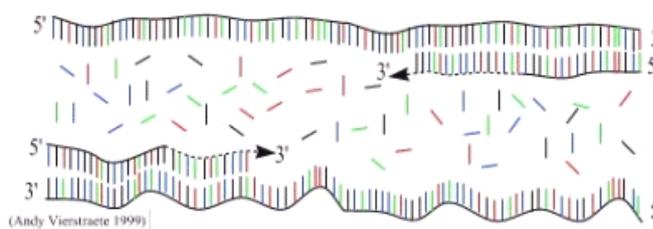


Figura 34. As 3 etapas do ciclo de PCR (<http://allserv.rug.ac.be/~avierstr/index.html>).

A cada novo ciclo da reação inicia-se com o aquecimento para desnaturação da dupla fita de DNA, seguido de resfriamento para hibridação dos iniciadores e síntese de uma nova fita pela DNA polimerase a partir dos iniciadores, sendo que as fitas de DNA recém sintetizadas servem de molde no ciclo seguinte. Portanto, em cada ciclo é sintetizado o dobro do DNA produzido no ciclo anterior (Figura 35). Usualmente, são realizados de 20 a 30 ciclos para amplificação de um segmento de DNA específico dentro de um genoma.

Nas primeiras iniciativas para amplificar fragmentos de DNA utilizava-se a enzima DNA polimerase da *Escherichia coli*, que possui atividade máxima a 37°C. Esta enzima deveria ser adicionada a cada ciclo pois o passo de desnaturação inativa a enzima. Um importante avanço ocorreu com a descoberta da enzima Taq DNA polimerase (Saiki et al, 1988) oriunda da bactéria *Thermus aquaticus*. A Taq DNA polimerase possui atividade ótima a 72°C e permanece razoavelmente estável mesmo a 95°C e com isto, a enzima é adicionada somente no início do processo .

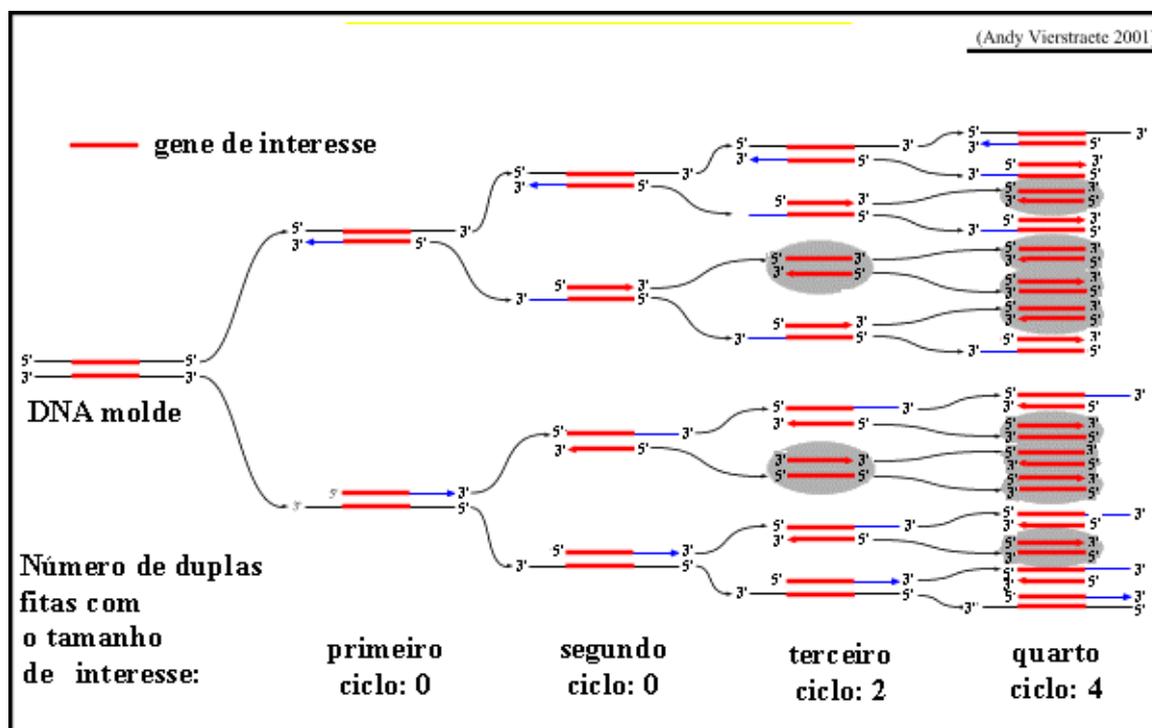


Figura 35. Os primeiros 4 ciclos de um PCR em detalhes. No terceiro ciclo, duas duplas fitas que apresentam o tamanho correto são copiadas (as duas fitas com o mesmo tamanho). No quarto ciclo, 8 duplas fitas que apresentam o tamanho são copiadas ([http:// allserv.rug.ac.be/~avierstr/index.html](http://allserv.rug.ac.be/~avierstr/index.html)).

Além de ser utilizada para a amplificação de um segmento específico de DNA dentro de um genoma, a metodologia de PCR pode ser utilizada para amplificar um segmento ou toda a seqüência de um produto gênico. Isso pode ser feito a partir da população de RNA de uma determinada célula ou tecido. Entretanto, como a Polimerase utilizada na metodologia de PCR é uma DNA Polimerase dependente de DNA, ela não pode usar RNA diretamente como molde para a amplificação. Assim, inicialmente é realizada uma reação de transcrição reversa, onde pela ação da transcriptase reversa o RNA é utilizado como molde para geração de cDNA, que é então utilizado como molde em uma PCR subsequente. Essa abordagem denominada de RT-PCR (Reverse Transcription and Polimerase Chain Reaction) é muito útil para se analisar de maneira rápida a expressão de genes de interesse, pois uma vez que a PCR é realizada a partir de cDNA, o fragmento correspondente ao gene de interesse só será amplificado naquelas células ou tecidos onde o gene é expresso. Esta técnica pode ser utilizada para clonagem de um cDNA de interesse, desde que já se tenha alguma informação de sua seqüência, sem a necessidade de se construir uma biblioteca de cDNA para isolá-lo.

VI. SEQÜENCIAMENTO DE DNA

A partir do final da década de 70 tornou-se possível a determinar a seqüência nucleotídica de fragmentos de DNA. O desenvolvimento da metodologia descrita por Sanger e cols associado a avanços tecnológicos, permite que hoje genomas inteiros sejam seqüenciados.

O princípio do método de terminação da cadeia para seqüenciamento de DNA baseia-se na síntese de DNA *in vitro* realizada na presença didesoxirribonucleotídeos. Um DNA purificado pode ser sintetizado *in vitro* em uma mistura que contém moléculas de fita única de DNA, enzima DNA polimerase, um DNA iniciador que possibilita a polimerase iniciar a síntese de DNA utilizando os desoxirribonucleotídios. Alternativamente, na presença de didesoxirribonucleotídeos (ddATP, ddCTP, ddGTP e ddTTP) na reação, o alongamento da cadeia de DNA é bloqueado pela incorporação do nucleotídeo sem um grupo OH na posição 3'.

No método originalmente desenvolvido por Sanger para seqüenciamento de DNA, uma fita complementar é sintetizada utilizando um mistura de desoxirribonucleotídeos, um deles marcado com P^{32} ou S^{35} . São realizadas 4 reações independentes, contendo uma pequena quantidade de um tipo de didesoxirribonucleotídeo em cada uma das reações. Cada reação produz um conjunto de cópias do DNA inicial que terminam em diferentes pontos da sua seqüência. Os produtos destas quatro reações são separados por eletroforese, utilizando quatro raias paralelas em um gel de poliacrilamida e os fragmentos são detectados através da marcação radioativa. Em cada um das raias, as bandas representam fragmentos que foram terminados em dado nucleotídeo em diferentes posições do fragmento inicial de DNA. Pela análise da ordem das bandas, começando pelo final do gel através das quatro raias, pode-se determinar a seqüência do DNA recém sintetizado (Figura 36).

Este método ainda é amplamente utilizado, sendo que vários avanços tornaram o seqüenciamento de DNA uma técnica simples e rápida. Os principais avanços foram a adaptação da PCR ao método de terminação da cadeia e a utilização de didesoxirribonucleotídeos marcados com diferentes corantes fluorescentes (fluorocromos). A utilização de fluorocromos permite que todas as quatro reações sejam realizadas em um único tubo e o produto da reação de seqüenciamento seja separado em uma única raia do gel.

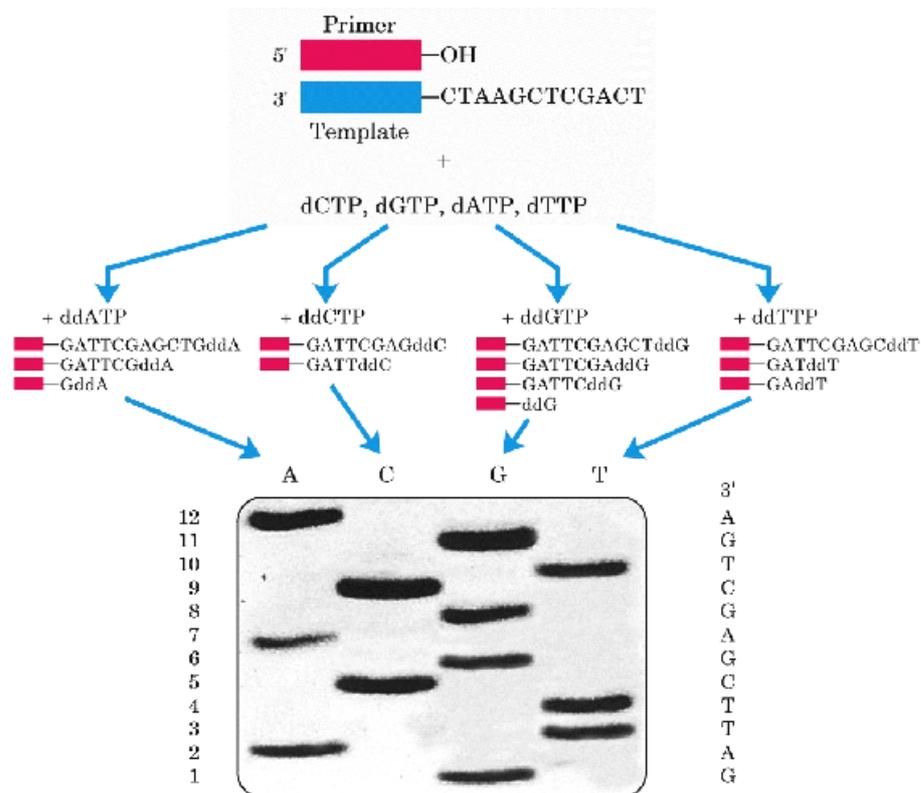


Figura 36. Seqüenciamento de nucleotídeos pelo método do término do crescimento da cadeia. Lehninger, Principles of Biochemistry. Figura 10-36, p. 352.

Atualmente, todo o processo de seqüenciamento de DNA pode ser realizado em seqüenciadores automáticos. Nestes equipamentos o produto da reação de seqüenciamento é separado por eletroforese em gel ou em capilar, os fluorocromos são excitados por um feixe laser, os sinais fluorescentes são amplificadas e detectados por tubos fotomultiplicadores ou nos modelos mais modernos por câmaras CCD. Análises computacionais identificam cada nucleotídeo pelo comprimento de onda de emissão específico de cada fluorocromo. As bases são identificadas de acordo com a forma do pico fluorescente e a distância entre picos sucessivos (Figura 37).

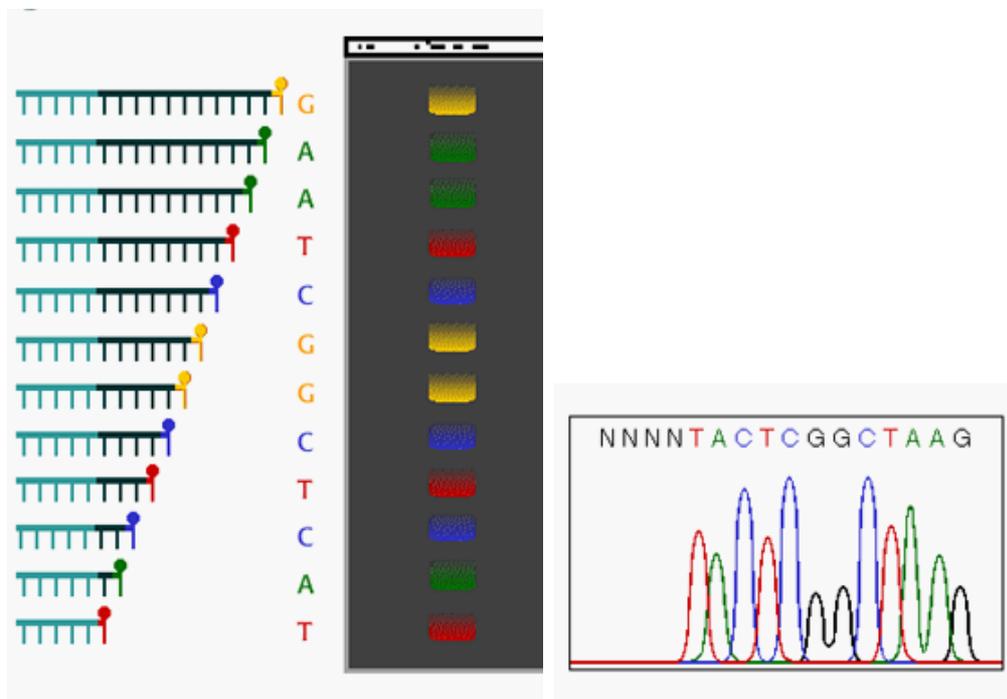


Figura 37. O produto da reação de seqüenciamento é submetido à eletroforese em uma raia de gel. À medida que os fragmentos passam pelo feixe de laser, os fluorocromos são excitados e a luz emitida é detectada por um fotomultiplicador. Esta informação é traduzida na forma de seqüência através de um computador.

VII. EXPRESSÃO DE PROTEÍNAS HETERÓLOGAS

1. INTRODUÇÃO

A descoberta de promotores e repressores específicos do genoma de *E. coli*, de uma variedade de vírus de *E. coli* e de células eucarióticas permitiu a manipulação da expressão de proteínas e a clonagem de genes sob o controle de um dado promotor, cuja expressão pode ser indutível ou contínua. O primeiro, e mais comumente usado, sistema de expressão de proteínas heterólogas em *E. coli* é baseado no operon lac (Figura 38). Neste sistema, o DNA de interesse é clonado em fago (no caso de biblioteca de cDNA de expressão) ou plasmídeo contendo o *lacI* (repressor), *lacP* (promotor) e *lacZ* (gene estrutural transcrito para mRNA da β -galactosidase). A indução da transcrição é obtida pela adição de um análogo de lactose sintético e não degradável (isopropiltio- β -D-galactosídeo, IPTG), o qual se associa ao repressor e inibe-o, deixando o promotor livre para a interação com a RNA polimerase e conseqüente transcrição do gene.

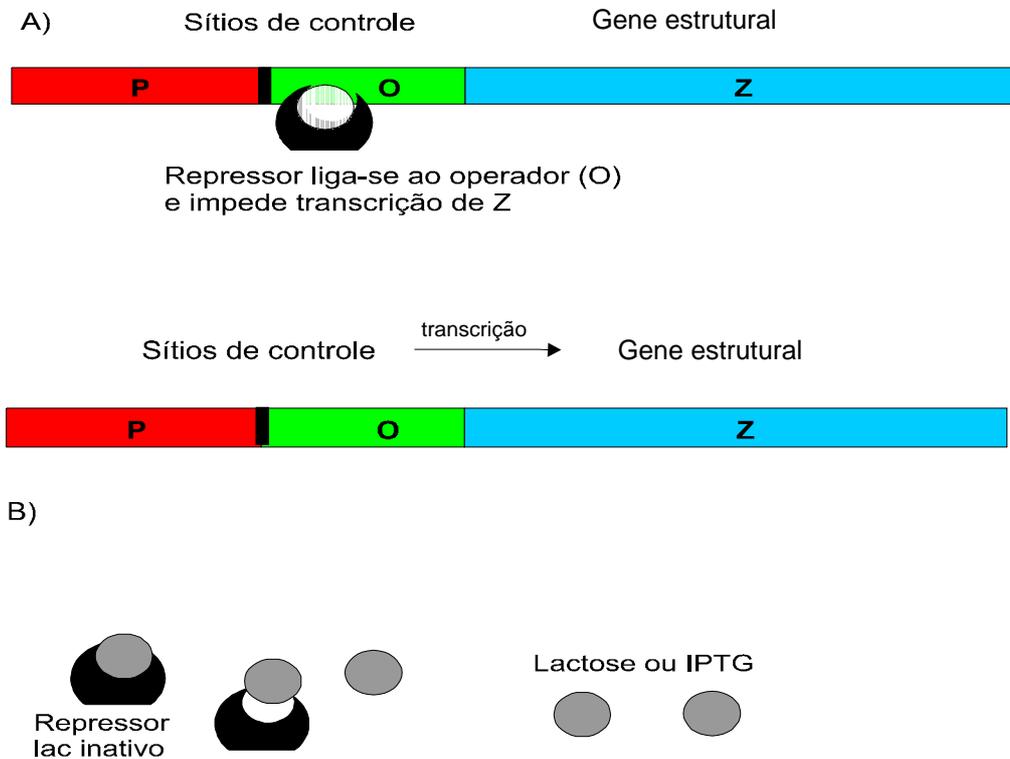


Figura 38. Esquema ilustrando o funcionamento do operon lac. (a) Na ausência de indutor. (b) Na presença de indutor.

O sistema mais comumente utilizado para expressão de proteínas heterólogas é o que utiliza *E. coli* como célula hospedeira. Este sistema é amplamente difundido devido à facilidade e baixo custo de se cultivar *E. coli*, e pela reprodutibilidade e abundância de proteína que produz. Além disso, modificações nos vetores e linhagens de *E. coli* são frequentemente feitas no sentido de aumentar a eficiência e versatilidade do sistema original. Com isto, e com o advento de novos sistemas de expressão em células de eucariotos (levedura, insetos, mamífero), a expressão de proteínas clonadas tornou-se uma abordagem poderosíssima que vem revolucionando os estudos de estrutura, função, purificação e identificação de novas proteínas. Nestes outros sistemas, o recombinante a ser introduzido no hospedeiro alvo pode ser facilmente construído e amplificado em *E. coli*. Por isso os plasmídeos de levedura, mamífero, etc. são construídos por fusão de uma porção de um plasmídeo de *E. coli* (origem de replicação e resistência a um antibiótico) com seqüências específicas para se obter expressão em célula eucariótica, conforme veremos adiante.

2. EXPRESSÃO DE PROTEÍNAS HETERÓLOGAS EM *E. COLI*

O uso de *E. coli* para a obtenção de proteína em quantidade suficiente para o estudo da estrutura e função da mesma ou para aplicações clínicas ou industriais é hoje disseminado e constitui-se em um marco na história do nosso conhecimento de estrutura de proteínas. Este tópico pretende sobretudo descrever os passos básicos envolvidos na construção de recombinantes e expressão, em bactéria, de uma proteína clonada.

2.1 Subclonagem em plasmídeos de expressão

O procedimento de clonagem de um fragmento de DNA para expressão é exatamente igual a qualquer clonagem, no entanto, deve se ter em mente que o propósito será obter a proteína correta. Para tanto, é necessário respeitar o **sinal de tradução de genes procarióticos** (sinal de Shine-Dalgarno), em outras palavras, o DNA deve ser clonado de maneira que sua fase de leitura correta fique em fase com o ATG iniciador (Figura 39). Além disso, um vetor para expressão em *E. coli* deve apresentar as seguintes características:

- origem de replicação
- marcador para seleção: gene que confere resistência a antibióticos como ampicilina, tetraciclina, cloranfenicol, neomicina. Ex. o gene da β -lactamase que confere resistência a ampicilina.
- um promotor para transcrição: como os vetores de expressão são normalmente projetados no sentido de produzir proteína em abundância, o DNA codificante para uma dada proteína deve ser colocado sob o comando de um promotor forte (exemplos: *lacZ*, *tac* (*trp+lacZ*), λ -*pR*, λ -*pL*, *p₁₀* do bacteriófago T7) e regulável, isto é, contendo um repressor para manter os níveis basais de expressão do gene insignificantes até a indução, o que se faz geralmente por adição de IPTG, no caso por ex. do repressor *lacI*, ou choque térmico, no caso do repressor *cIts857*.
- sinal de terminação da transcrição
- sequências para controle da tradução, como por ex., um sítio de ligação ao ribossomo para a iniciação da tradução (Shine-Dalgarno) e um ATG iniciador. Um sinal de terminação da tradução (códon de terminação) também deve estar presente no vetor ou no inserto a ser clonado, ou deve ser adicionado.

- um MSC para facilitar a inserção do gene de interesse na orientação correta.

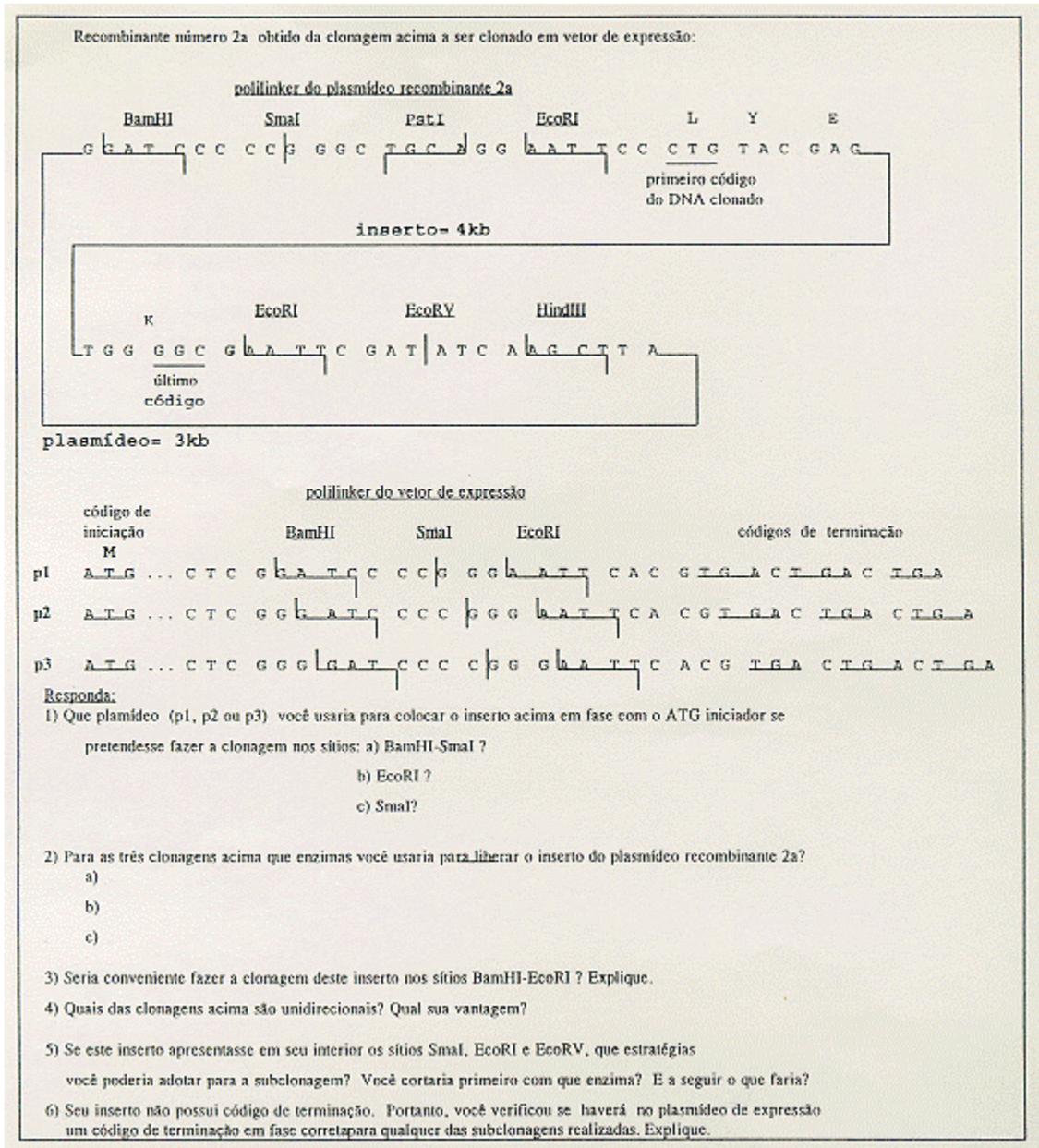


Figura 39. Procedimentos para clonagem de um fragmento de DNA em fase de leitura correta em um plasmídeo de expressão.

Uma vez construído, o vetor de expressão contendo a sequência codificadora da proteína de interesse é introduzido em *E.coli* por transformação.

Ao planejar uma subclonagem para a expressão de proteína, além de se respeitar a fase de leitura correta, deve-se também tentar na medida do possível fazer a clonagem unidirecional do inserto. Na clonagem unidirecional utilizam-se duas enzimas de restrição para digerir o vetor e o DNA inserto. Na bidirecional utiliza-se uma única enzima, de modo que as pontas são iguais, permitindo a inserção do fragmento em ambas as orientações.

Assim, em uma clonagem bidirecional há 50% de probabilidade de se obter os recombinantes na orientação correta (senso) e 50% na orientação invertida (anti-senso).

2.2 Análise dos plasmídeos e seleção dos subclones corretos

A análise de DNA plasmidial para a seleção dos subclones corretos é feita por digestão com enzimas de restrição e confirmada por sequenciamento se julgar necessário. A análise de restrição é um procedimento bastante simples; entretanto deve ser feito com cautela e atenção. Seu objetivo é selecionar o clone recombinante correto quanto ao tamanho do inserto e sua orientação.

Como regra a primeira digestão deve ser feita com a(s) enzima(s) utilizada(s) para a clonagem; isto produzirá dois fragmentos: plasmídeo linear + inserto, e permitirá avaliar se os sítios de clonagem foram recuperados intactos. Se a ligação for realizada entre extremidades cegas, preenchidas pela ação da Klenow polimerase, verifica-se em um catálogo de enzimas de restrição se haverá a formação de um novo sítio, o qual poderá ser clivado para análise. Outra alternativa consiste em clivar em sítios próximos ao da clonagem em ambas as extremidades.

A segunda digestão, no caso da clonagem bidirecional, deve ser planejada para determinar a orientação do inserto. Deve-se escolher um sítio único e não centralizado no inserto e um sítio no MSC que seja ausente no inserto, produzindo de preferência dois fragmentos de tamanhos facilmente distinguíveis no gel. Calcular previamente o tamanho dos fragmentos esperados para ambas as orientações. A digestão completa é fundamental para a correta interpretação dos resultados. A Figura 40 esquematiza um projeto de subclonagem de um gene e a análise de restrição dos subclones obtidos com o objetivo de: a) selecionar os recombinantes, b) selecionar o recombinante correto.

3. PRODUÇÃO DE PROTEÍNAS HETERÓLOGAS EM *E. COLI*

3.1. Produção de proteínas híbridas

De maneira ideal, quando se pensa em expressão heteróloga, espera-se que a proteína de interesse seja estável, não tóxica para a bactéria, solúvel, produzida em grande quantidade e possa ser facilmente purificada. Um procedimento muito utilizado é o de expressar a proteína de interesse em fusão com um “tag” específico que permita a fácil

purificação da mesma através de cromatografia por afinidade em resinas às quais se encontram acopladas ligantes, aos quais o “tag” possa se ligar especificamente. Uma outra vantagem óbvia de se produzir uma proteína híbrida, ou de fusão, é no caso da expressão de polipeptídeos pequenos ou mesmo peptídeos, os quais, sem fusão, seriam instáveis e rapidamente degradados na célula.

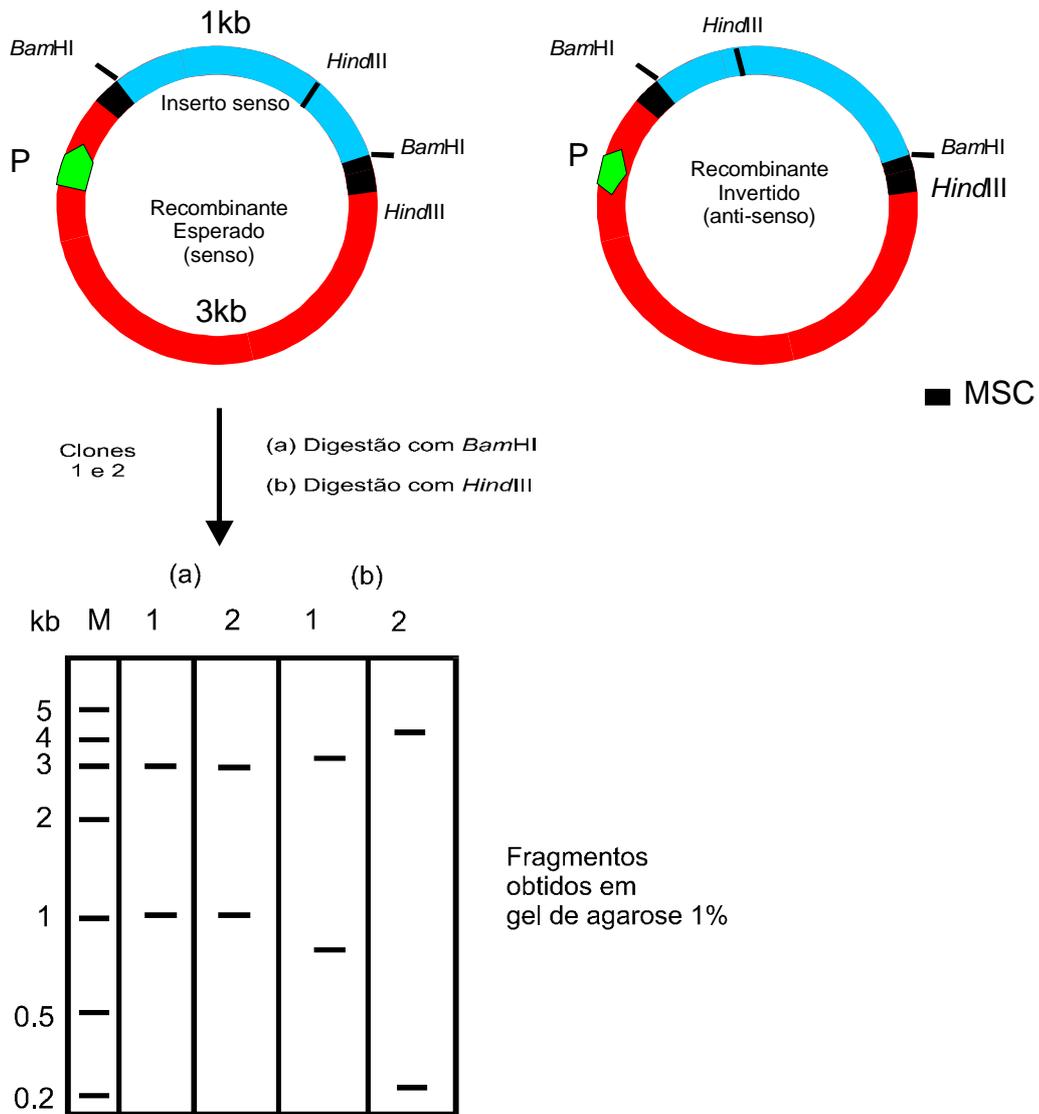


Figura 40. Subclonagem de um fragmento de DNA em plasmídeo de expressão e análise de restrição de dois recombinantes obtidos.

Nota : Os clones 1 e 2 foram digeridos com *Bam*HI e *Hind* III, separadamente.

Resposta: Qual dos dois clones tem o inserto na orientação correta (senso) para a expressão da proteína?

Em geral projeta-se ainda um sítio sensível a uma determinada protease (ex: fator Xa, trombina), inserido imediatamente acima do sítio de clonagem, de maneira que a proteína híbrida possa ser clivada liberando a proteína clonada que pode então ser facilmente purificada utilizando-se a mesma coluna de afinidade. A coluna reterá a proteína de fusão e eliminará no "void" a proteína de interesse. Um exemplo de proteína híbrida obtida por clonagem no vetor pMAL é dado na Figura 41.

O procedimento de proteólise é relativamente trabalhoso e nem sempre eficiente, por isso quando a proteína de fusão não interfere se utiliza a própria proteína híbrida, por exemplo, para experimentos funcionais, produção de anticorpos, etc. Note que neste caso pode ser estratégico ter o mesmo DNA clonado em dois sistemas de fusão diferentes. Isto permitirá que anticorpos produzidos contra uma proteína híbrida sejam purificados por afinidade contra a outra, purificando-se assim apenas anticorpos cujos epitopos localizam-se na proteína clonada.

Dentre os vetores utilizados com sucesso para a produção de proteínas híbridas podemos citar:

- **pGEX**: apresenta a glutathione-S-transferase de *Schistosoma japonicum* (26 kDa) como proteína de fusão, permitindo a purificação da proteína de fusão em coluna de agarose-glutathione
- **pMAL**: apresenta a proteína ligante de maltose (PLM) de bactéria (42 kDa) como fusão, permitindo a purificação da proteína de fusão em coluna com amilose acoplada
- **pQE**: apresenta um "tag" de 6 resíduos de histidina como fusão e permite a purificação das proteínas de fusão em colunas quelantes de Ni²⁺
- **pUR**: cuja fusão é um fragmento da β -galactosidase

A produção de proteínas híbridas é geralmente muito simples, eficiente e de baixo custo financeiro, podendo suprir de imediato necessidades da pesquisa básica, tais como, produção de anticorpos e purificação destes por afinidade, sondas em experimentos variados e para estudos funcionais/estruturais da proteína expressa. Permite, também, a expressão em grande escala, para fins industriais ou clínicos de enzimas, hormônios, anticorpos, etc., quando a atividade da proteína é preservada. Além disto, é possível se obter, por mutagênese, proteína com a atividade de interesse potenciada e livre de efeitos adversos ou atividades indesejadas.

(b)

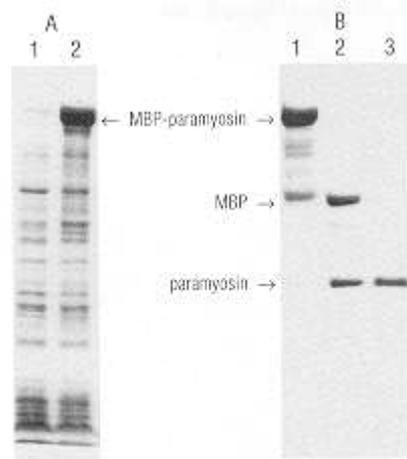
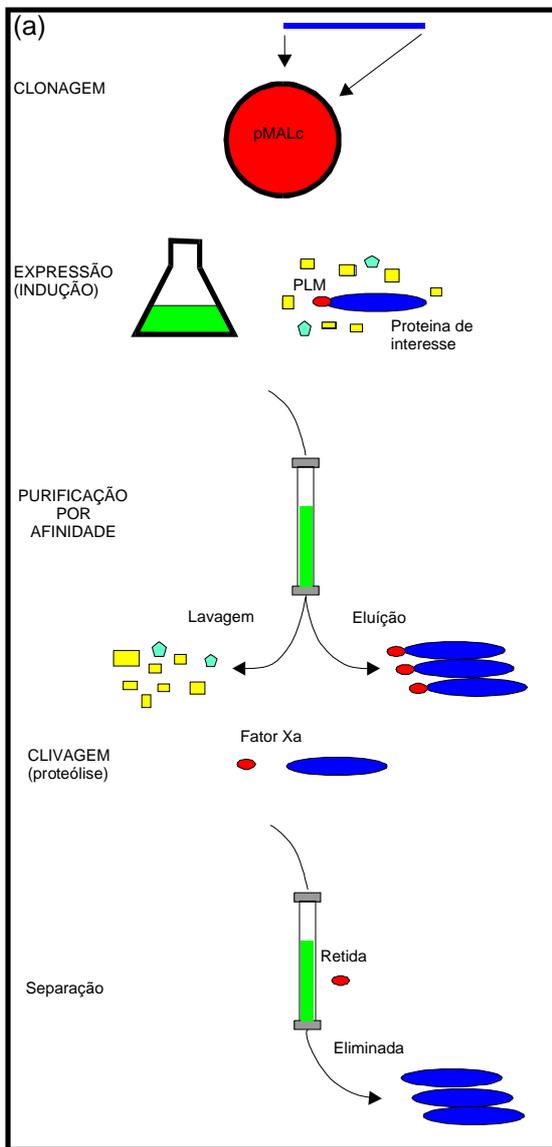


Figura 41. a) Esquema ilustrando os passos para a expressão, purificação e clivagem de uma proteína de fusão. b) Gel de poliacrilamida-SDS de frações da purificação da proteína de fusão PLM-paramiosina. Em A são mostrados os extratos de células não induzidas (1) e de células induzidas (2). Em B, proteína de fusão PLM-paramiosina após eluição da coluna de amilose por adição de maltose (1), após clivagem por fator Xa (2) e fração coletada no "void" após passagem na coluna para a retenção da PLM (3).

É importante estar alerta de que a situação ideal exposta acima nem sempre é atingida. Na realidade, na grande maioria das vezes, proteínas de eucarioto produzidas em bactéria não são solúveis; às vezes podem ser tóxicas para a célula; algumas são expressas em baixos níveis; algumas interferem com a sua fusão inibindo-a de se ligar à resina de afinidade e tornando a purificação menos eficiente; outras formam agregados extremamente insolúveis mesmo na presença de SDS/ β -mercaptoetanol; algumas têm a sua atividade biológica plenamente recuperada, porém outras são inativas.

Assim, dentre os problemas mais comuns que ocorrem com a produção de proteínas heterólogas em *E. coli*, podemos citar:

- Proteínas tóxicas: algumas proteínas são tóxicas para a célula hospedeira. Neste caso, pode-se proceder a secreção. Uma alternativa à produção de proteína citoplasmática, é a

produção de proteínas que são secretadas. Para tanto basta clonar o DNA codificante em fusão com uma seqüência codificante para um peptídeo sinal de procarioto. Este peptídeo é clivado pela peptidase sinal quando a proteína é secretada para o periplasma. Embora este método freqüentemente traga problemas com o rendimento ou a clivagem do peptídeo sinal, há vantagens em alguns casos: no caso de proteínas tóxicas para a bactéria; algumas proteínas degradadas por protease no citoplasma são estáveis no periplasma; algumas que são inativas quando produzidas intracelularmente são ativas quando secretadas; a proteína produzida já tem a sua metionina N-terminal processada. Um exemplo de sucesso é a expressão do hormônio fator de crescimento epidermal humano (hEGF) sob o comando do promotor da fosfatase alcalina (*phoA*) e com a seqüência sinal desta. A indução é obtida sob privação de fosfato do meio.

- Proteínas instáveis: isto pode ser resolvido reduzindo-se a temperatura de crescimento ou mudando-se para uma linhagem de bactéria deficiente em uma ou mais proteases. Mesmo assim é comum se obter algum nível de fragmentação da proteína expressa (ver Fig. 39 b, raia 1).
- Baixos níveis de expressão: pode ocorrer pelas razões acima dentre outras, como, por exemplo, instabilidade do mRNA, término prematuro da mensagem, tradução ineficiente.
- Proteínas insolúveis: proteínas de eucariotos produzidas em bactéria são geralmente precipitadas na forma de corpos de inclusão e requerem procedimentos adicionais de desnaturação para solubilizá-las e de renaturação para mantê-las solúveis e funcionais. Isto nem sempre é um problema, pois a formação de corpos de inclusão pode proteger a proteína contra a degradação por proteases bacterianas e também facilitar a purificação, uma vez que são corpos densos, precipitados à baixa velocidade de centrifugação, enquanto a maior parte das proteínas bacterianas permanecem no sobrenadante. Mas algumas vezes não se consegue renaturação adequada da proteína purificada. Neste caso deve-se tentar atenuar a formação de corpos de inclusão alterando as condições de expressão, por exemplo, crescendo-se a cultura a temperatura mais baixa após a indução ou utilizando um promotor mais fraco.

3.2. Produção de proteínas intactas

A vantagem de se produzir proteína intacta sem fusão é que ela poderá ser utilizada sem a preocupação de que o segmento de fusão esteja interferindo em sua estrutura e atividade. A principal desvantagem é que nem sempre se encontra um método eficaz para a purificação da mesma. Contudo há exemplos demonstrando que para estudos funcionais e de estrutura é preferível investir na produção de proteína intacta.

Vários vetores encontram-se disponíveis no mercado para a expressão de proteínas sem fusão em *E.coli*. Os mais referidos são da série pET (plasmid for expression by T7 RNA polimerase), cuja expressão está sob o controle do promotor de transcrição $\phi 10$ e dos sinais de iniciação de tradução *s10* da proteína do gene 10 (a principal proteína do capsídeo) do bacteriófago T7. A grande vantagem deste vetor é que ele é transcrito pela T7 RNA polimerase, a qual é muito seletiva e ativa, sendo capaz de alongar cadeias de RNA aproximadamente 5 vezes mais rápido que a RNA polimerase de *E. coli*. Alguns vetores da série pET apresentam o promotor T7-lac, colocando a expressão da proteína sob o controle lac e reduzindo portanto o background de expressão da proteína alvo na ausência de IPTG.

4. EXPRESSÃO DE PROTEÍNAS HETERÓLOGAS EM ORGANISMOS EUCARIOTOS

Ao se escolher um sistema de expressão, deve-se sempre considerar a aplicação final da proteína a ser expressa. Embora o sistema de expressão de proteínas heterólogas em *E. coli* seja bastante difundido, no caso de proteínas eucarióticas complexas torna-se, as vezes, necessário lançar-se mão de sistemas de expressão em células eucarióticas para se produzir proteínas na sua forma nativa, pois estes sistemas permitem modificações pós-traducionais essenciais para a função de determinadas proteínas, como, por exemplo, glicosilação. Consideraremos alguns destes sistemas a seguir.

4.1. Expressão em células de mamíferos

Os vetores para expressão em mamíferos contém sequências para facilitar a propagação em bactéria (origem de replicação e um marcador para seleção), bem como sequências específicas para se obter expressão em células eucarióticas.

Assim, um vetor para expressão em células de mamíferos deve apresentar as seguintes características:

- origem de replicação para células eucarióticas: por exemplo a do vírus SV40
- promotor: como, por exemplo, do SV40 ou do citomegalovírus (CMV). Existem também vetores com promotores induzíveis como, por exemplo, o da tetraciclina e o responsivo à ecdisona.
- sinais para o término da transcrição e para a poliadenilação do transcrito
- marcador para seleção: como por exemplo o gene que codifica a aminoglicosídeo fosfotransferase, conferindo resistência à geneticina (um análogo de neomicina).
- sequência para ligação ao ribossomo (sequência de Kozak)

Os sistemas de expressão em células de mamíferos podem ser divididos em dois tipos: a) aqueles envolvendo expressão transitória da proteína em questão (1-4 dias após a introdução do DNA); b) aqueles envolvendo a expressão estável e que requerem, portanto, a integração do gene no DNA cromossômico.

4.2 Expressão em fungos

Fungos são eucariotos unicelulares potencialmente capazes de realizar todas as modificações pós-traducionais observadas em células de mamíferos. A facilidade de cultivo em larga escala e o fato de *S. cerevisiae* ser um microrganismo seguro, legalmente usado na indústria de alimentos e farmacêutica, são características adicionais que tornam a expressão heteróloga de proteínas neste sistema particularmente atrativa e com perspectivas para uso, por exemplo, no desenvolvimento de vacinas orais através da imobilização de antígenos na superfície deste microrganismo. Exemplos da expressão de proteínas heterólogas em fungos encontram-se descritos nos itens 7.1.2 e 7.2.

5. SISTEMA DE EXPRESSÃO EM CÉLULAS DE INSETO UTILIZANDO BACULOVÍRUS COMO VETOR

É um dos sistemas de expressão em células eucarióticas mais poderosos e versáteis. O genoma do Baculovírus é muito grande para permitir a inserção direta de genes heterólogos; por isso, estes são clonados em vetores de transferência, os quais contêm sequências flanqueadoras que são homólogas (5' e 3' ao inserto desejado) ao genoma do Baculovírus. Procede-se, então, à co-transfecção do vetor de transferência contendo o gene

de interesse clonado e do DNA linearizado do vírus *Autographa californica* (AcNPV) em células do inseto *Spodoptera frugiperda* (Sf). Nestas células ocorre recombinação homóloga, com transferência do gene heterólogo do vetor para o DNA do AcNPV. A infecção das células Sf com o DNA do vírus resulta na parada total da expressão das proteínas do hospedeiro, permitindo alta produção do mRNA e da proteína recombinantes.

Uma variedade de vetores de transferência foram construídos para utilização neste sistema. Cada vetor contém :

- uma origem de replicação de *E.coli*
- um marcador de resistência à ampicilina
- o promotor da polihedrina ou da p10 (proteínas do baculovírus): promotores fortes.
- um MSC para permitir inserção do gene de interesse
- sequências flanqueadoras pertencentes ao genoma do vírus para facilitar a recombinação homóloga

Vantagens deste sistema de expressão:

- Modificações pós-traducionais
- Obtenção de proteínas solúveis e funcionalmente ativas
- Altíssimos níveis de expressão
- Capacidade para grandes insertos
- Expressão simultânea de vários genes; vários plasmídeos com promotores múltiplos encontram-se disponíveis comercialmente, permitindo a expressão de duas ou mais proteínas numa única célula infectada.
- Facilidade de purificação; vetores apresentando “tags” de 6xHis e GST permitem a purificação por afinidade das proteínas recombinantes.
- Facilidade de monitoramento da expressão: vetores Biocolors permitem a obtenção da proteína de interesse em fusão com GFP (green), BFP (blue) ou YFP (yellow), permitindo o monitoramento da expressão sem a necessidade da utilização de anticorpos.

6. EXPRESSÃO EM CÉLULAS DE *Drosophila*

Combina algumas das melhores características dos sistemas de expressão em células de mamífero com aqueles em células de insetos utilizando baculovírus. É um sistema mais barato e eficiente que expressão em células de mamíferos e mais rápido que expressão

utilizando baculovírus. Este sistema utiliza vetores plasmidiais para expressão transitória ou estável da proteína de interesse. Encontram-se disponíveis comercialmente vetores para expressão: a) constitutiva (promotor da actina), b) induzível (promotor da metalotioneína, expressão induzida pela adição de cobre ou cádmio) e c) induzível e com secreção da proteína para o meio. Todos os vetores apresentam “tags” de 6xHis no C-terminal, facilitando a purificação por afinidade da proteína de interesse.

7. EXPRESSÃO DE PROTEÍNAS HETERÓLOGAS: APLICAÇÕES E PERSPECTIVAS

7.1- Bibliotecas de expressão e identificação de novas proteínas através de anticorpos ou outros ligantes específicos

7.1.1- Sistema convencional - expressão de biblioteca de cDNA em *E. coli*

Vocês já viram em capítulos anteriores que uma biblioteca de cDNA pode ser construída em vetor de expressão (fago ou plasmídeo). Para essa finalidade, um dos vetores mais utilizados atualmente, pela sua versatilidade, é o λ ZAP e o hospedeiro mais usado é *E. coli*. Quando se clona um dado fragmento de DNA pode-se planejar previamente se deseja-se uma clonagem uni ou bidirecional. Quando se sintetiza uma biblioteca de expressão é usual que se opte pela clonagem unidirecional. Para isso basta digerir o vetor com duas enzimas que geram pontas incompatíveis e preparar o inserto com adaptadores que geram pontas compatíveis com as do vetor, de tal maneira que a orientação seja a correta (fita sense, sentido 5'→3', a partir do promotor) para a expressão de proteína. A biblioteca é amplificada como qualquer outra e a expressão de proteína é induzida pela adição de IPTG; uma réplica da placa de ágar contendo as placas de lise é obtida em um filtro de nitrocelulose colocado sobre a mesma por um certo tempo para que as proteínas sejam aderidas. Este filtro é em seguida tratado como em um Western blot para a detecção de proteína, com anticorpo ou um outro ligante específico para a proteína de interesse. O sistema λ ZAP é interessante porque após o isolamento do clone o fagomídeo pode ser facilmente obtido por excisão *in vivo* por co-infecção com fago filamentosso auxiliar (Figura 42).

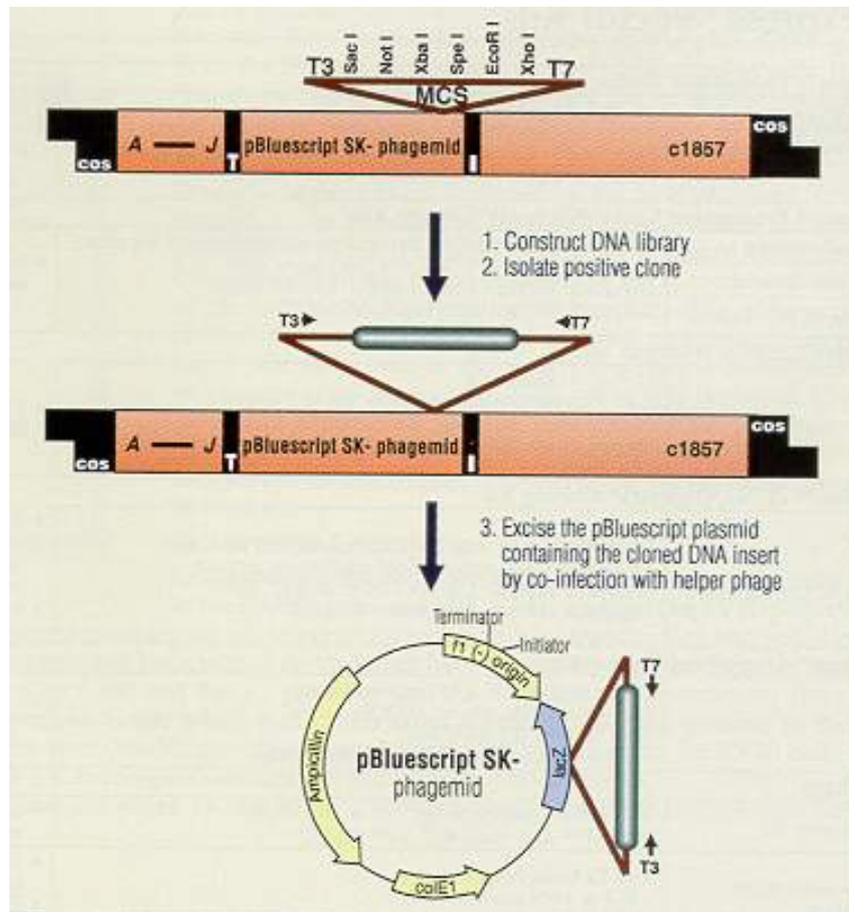


Figura 42. Esquema do fago λ ZAP, isolamento do clone e excisão do fagemídeo pBluescript por coinfeção com fago filamentoso auxiliar.

A proteína clonada pode então ser expressa e suas propriedades estudadas, inclusive pode ser prontamente purificada para a injeção em coelhos a fim de gerar anticorpos. Estes podem ser usados como contra-prova da clonagem e como sondas importantíssimas para diagnóstico, imunocitoquímica, etc. Há também a possibilidade de se construir a biblioteca em um λ ZAP modificado que contém o promotor de citomegalovírus (CMV) e outros elementos necessários para a expressão do gene clonado em células eucarióticas.

7.1.2. Novos sistemas para a detecção de receptores ou proteínas ligantes

A identificação de moléculas que interagem com uma dada proteína, isto é peptídeos ligantes, proteínas ligantes ou receptores, é fundamental para o mapeamento da via de ação de uma dada proteína na célula. O sistema convencional de "screening" (seleção) descrito acima pode ser utilizado. Mas, novos sistemas começam a aparecer com o

objetivo de selecionar ligantes ou anticorpos. Um destes é a clonagem do DNA (biblioteca de cDNA comum ou bibliotecas combinatoriais de peptídeos sintéticos) em fusão com uma proteína que forma a capa do fago lambda (Felici et al., 1991). Isto tem permitido isolar ligantes de uma biblioteca complexa por passagem das partículas de fago sobre determinado ligante específico imobilizado em uma matriz. Uma evolução deste sistema que parece ser promissora é a clonagem da biblioteca de cDNA de interesse em fusão com uma proteína da parede bacteriana, de tal maneira que as proteínas expressas ficarão expostas recobrendo a superfície da bactéria, a qual poderá então ser incubada com um ligante marcado por fluorescência que marcará aquela célula que expressar o receptor ou ligante específico. A célula marcada poderá, então, ser selecionada (isolada) por um aparelho de seleção de células, "Fluorescent Assisted Cell Sorters", comumente referido como FACS. Para a análise de milhões de bactérias, entretanto, será necessário o desenvolvimento de FACS mais sofisticados. A Figura 43 exemplifica esquematicamente uma proteína (anticorpo, neste caso) expressa na superfície de uma bactéria (Little et al., 1993). Este sistema tem sido considerado promissor para a produção de vacinas orais e kits para diagnóstico.

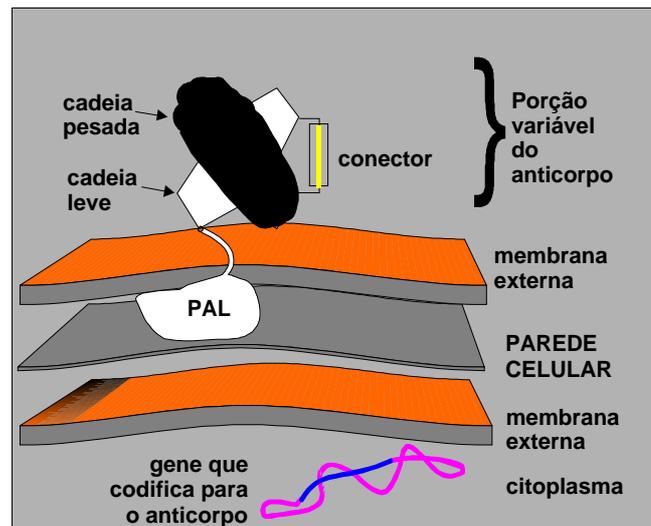


Figura 43. Esquema ilustrando a apresentação na superfície bacteriana de uma proteína heteróloga expressa em fusão com uma proteína da parede bacteriana. Este caso demonstra a expressão dos domínios variáveis de um anticorpo unidos por um peptídeo "conector" em fusão com a lipoproteína associada a peptídeoglicano (PAL). Foi projetado para preservar a conformação nativa da região variável do anticorpo de maneira a manter suas propriedades ligantes intactas.

Além dos exemplos citados acima, o "sistema de dois híbridos" (Fields & Song, 1989), para expressão em levedura, tem sido bastante utilizado para a detecção e

identificação de proteínas capazes de interagir com uma determinada proteína (Figura 44). Neste sistema dois plasmídeos são construídos para posterior co-transfecção e expressão dos híbridos. O primeiro híbrido (isca) é a proteína conhecida (**x**) em fusão com o domínio ligante de DNA do fator de transcrição GAL4; o segundo são genes desconhecidos de uma biblioteca de um determinado tecido ou espécie (proteína **y**) em fusão com o domínio da GAL4 ativador da transcrição. É um processo engenhoso, cuja lógica baseia-se no fato de que a GAL4, como a maioria dos fatores de transcrição, é uma proteína com dois domínios bem definidos (um ligante de DNA e o outro ativador da maquinaria de transcrição), ambos indispensáveis para o efeito final de ativação da transcrição de genes situados sob seu controle (controle do promotor GAL). Quando se expressa estes domínios separadamente, a proteína será inativa, exceto se os dois polipeptídeos se associarem de tal maneira que encontrar-se-ão juntos na região promotora para exercerem seu papel de ativação da transcrição. Quando se expressam estes híbridos em uma linhagem de levedura, cujo determinado gene marcador de seleção está sob o comando de GAL4 e, cultivam-se as células na ausência de um nutriente essencial sintetizado pelo gene marcador, a ativação deste gene torna-se essencial para a sobrevivência das células. Um gene comumente utilizado como marcador de seleção neste sistema é o *HIS3*, cujo produto é uma enzima da via de biossíntese de histidina; neste caso, as células são capazes de crescer em meio com histidina, mas são incapazes na sua ausência, a menos que a transcrição do gene *HIS3* seja ativada pelo fator GAL4 “reconstituído” (ativo). Isto significa que no interior das células que sobreviveram encontram-se dois híbridos que formam um complexo; isto é, a proteína conhecida **x** interage com uma proteína desconhecida **y**, a qual pode ser então identificada por sequenciamento de seu cDNA e caracterizada. Merece destaque o fato de que a interação entre as proteínas **x** e **y**, por este processo, ocorre no interior de uma célula eucariótica viva, sendo, por isso, provavelmente mais confiável que outros métodos utilizados para o estudo de interações entre proteínas.

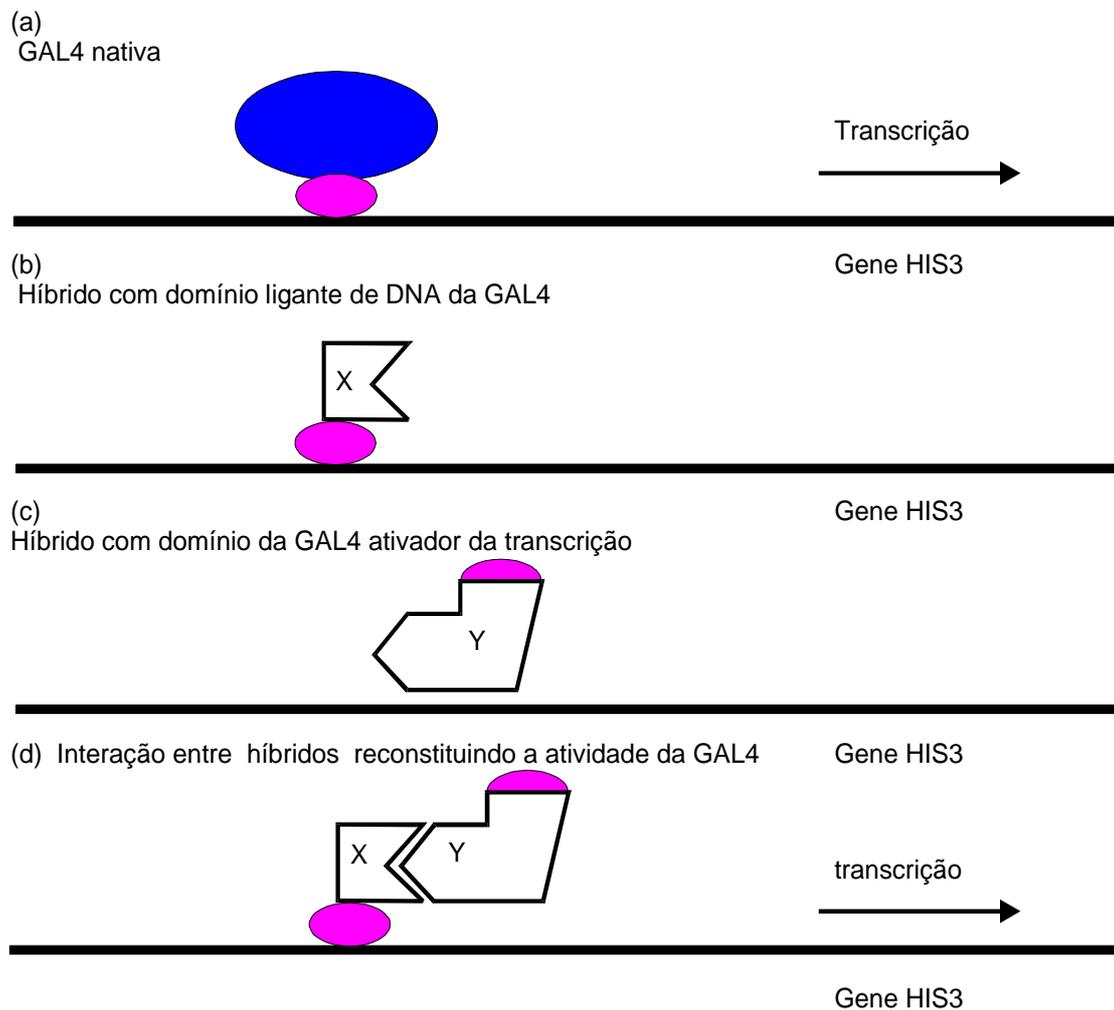


Figura 44. Estratégia para detectar proteínas ligantes utilizando o sistema de dois híbridos.

7.2. A utilização da tecnologia do DNA recombinante no conhecimento de receptores de membrana

Para informações específicas em clonagem de receptores de superfície celular, incluindo abordagens que visam descobrir ou projetar novas drogas efetivas direcionadas a esses receptores, consulte a revisão de Luyten & Leysen (1993). Estes autores enfatizam que a expressão heteróloga de receptores clonados proporciona uma fonte rica, reproduzível, inexaurível e barata de subtipos de receptores humanos. Esses receptores expressos podem ser utilizados no “screening” de drogas e também para o “design” de novas drogas através de um melhor entendimento da relação estrutura-função, visto que a clonagem molecular desses receptores revelou uma abundância de subtipos, bem como modelos estruturais dos mesmos (Figura 45). Agora, uma nova era pode ter lugar, em que as drogas começam a ser projetadas e construídas com base na estrutura de seu alvo no organismo (Bugg et al., 1994).

Exemplos da utilização da tecnologia do DNA recombinante neste caso incluem a expressão heteróloga de receptores acoplados a proteína G nos fungos *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* e *Pichia pastoris*, visando estudos estruturais e funcionais. Células de *S. cerevisiae* expressando o receptor de somatostatina de rato foram usadas com sucesso no “screening” de análogos de somatostatina e bibliotecas combinatoriais para identificar agonistas e antagonistas seletivos para subtipos com potentes efeitos *in vivo* (M. H. Pausch, 1997).

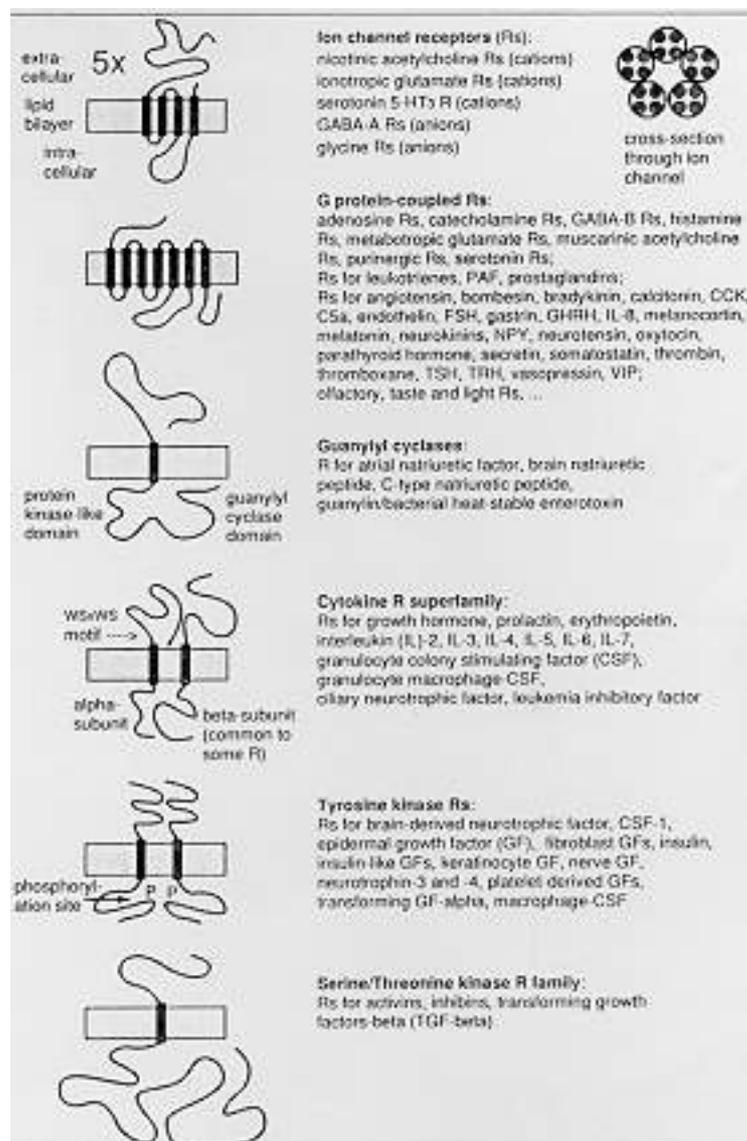


Figura 45. Modelos estruturais de famílias de receptores de superfície celular. Os domínios transmembrana e alguns domínios funcionais são representados. O enrolamento da cadeia proteica é hipotético, pois o único que teve sua estrutura determinada por cristalografia até o momento foi o domínio extracelular do receptor de hormônio de crescimento.

7.3. Expressão de proteínas para intervenção terapêutica

Podemos dizer que há duas maneiras gerais de se fazer terapêutica por expressão de proteínas:

a) a expressão heteróloga de uma proteína humana em bactéria ou célula eucariótica para a produção em grande escala, purificação e utilização desta para injeção em pacientes: o exemplo mais conhecido é provavelmente o da insulina (Bristow, 1993), a qual foi o primeiro produto comercial obtido pela tecnologia do DNA recombinante e aprovado para uso em humanos em 1982 pela U.S.FDA (Food and Drug Administration). A Figura 46 mostra as estratégias empregadas para a produção de insulina para fins terapêuticos; a bovina e a suína sem modificações químicas são menos eficientes. A insulina produzida pela tecnologia do DNA recombinante apresenta a vantagem de eliminar a possibilidade de contaminantes ou agentes infecciosos que podem estar presentes nas proteínas purificadas de animais. Uma das maiores expectativas neste campo é a possibilidade de se obter mutantes com atividade muito superior ao da proteína nativa e sem certos efeitos adversos. Outros exemplos de produtos obtidos pela tecnologia do DNA recombinante e utilizados com finalidade terapêutica são o anticoagulante TPA (ativador de plasminogênio tipo tecidual), o fator VIII de coagulação e a eritropoietina (hormônio que estimula a produção de eritrócitos).

A determinação e caracterização refinada da estrutura de proteínas começa a ganhar maior impacto pela possibilidade de expressão/purificação de proteína em larga escala e mutagenizadas. Além de nos proporcionar a satisfação de se conhecer uma nova estrutura, esse conhecimento nos permite projetar drogas e pensar em outras estratégias terapêuticas, tais como, expressar apenas um domínio ativo da proteína.

Por exemplo, os domínios variáveis de anticorpos podem ser expressos como uma única cadeia polipeptídica ativa. Pode-se pensar também na expressão de receptores solúveis (ou domínios ativos destes) para um dado vírus, no sentido de inibir a sua interação com o receptor da célula; um exemplo seria a expressão do receptor CD4 para o vírus da AIDS. Uma idéia espetacular que começa a emergir a partir do nosso conhecimento da estrutura atômica e função de fatores de transcrição, é a possibilidade da síntese de drogas que inibiriam ou ativariam especificamente um dado fator de transcrição (Figura 47).

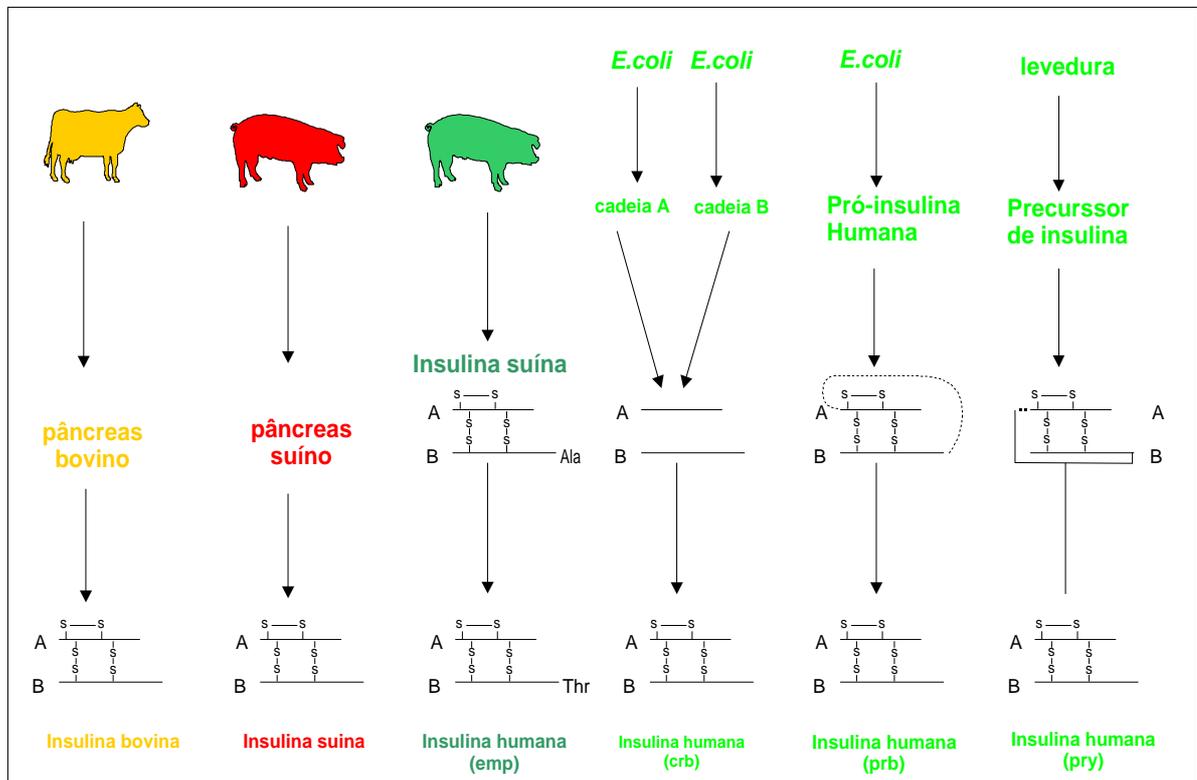


Figura 46. Representação esquemática de estratégias usadas para produzir insulina para fins terapêuticos.

Um outro exemplo da potencial aplicação da expressão de proteínas heterólogas em intervenção terapêutica seria a produção de grandes quantidades de anticorpos em plantas, visando seu uso em imunoterapia. As plantas são capazes de sintetizar e montar virtualmente qualquer tipo de molécula de anticorpo, desde fragmentos até cadeias completas e mesmo anticorpos multiméricos. Ma & Ben (1995) descreveram a expressão e montagem em plantas da molécula de IgA, que é a forma predominante de imunoglobulina em secreções de superfícies mucosas como o trato gastrointestinal. Até então a produção de IgA em sistemas heterólogos tinha sido frustrante devido a complexidade desta molécula. Anticorpos produzidos em plantas podem vir a ser particularmente úteis para imunoterapia tópica. No caso, por exemplo, de cáries, Ma et al. (1990) mostraram que a aplicação tópica de anticorpos monoclonais contra uma proteína da superfície (de adesão) do *Streptococcus mutans* em dentes humanos conferiu proteção a longo prazo contra esse microrganismo em adultos.

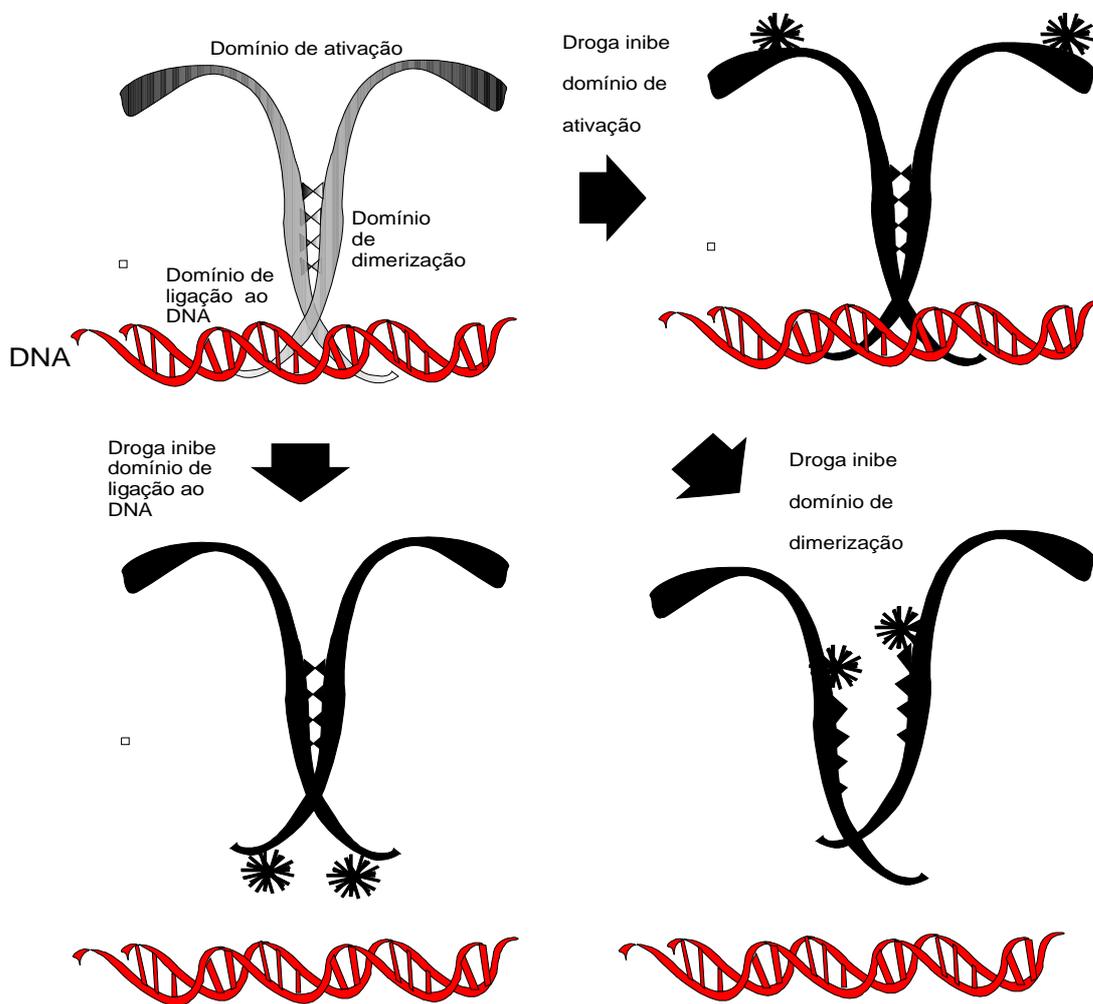


Figura 47. Modelos hipotéticos da ação de drogas que poderiam ser efetivas em terapêutica atuando sobre um fator de transcrição.

b) a terapia gênica por transfecção de células retiradas do próprio paciente com retrovírus, ou outros tipos de vírus, com o intuito de repor um gene mutante ou para o tratamento de doenças multi-fatoriais como câncer e doenças cardíacas. A primeira terapia gênica realizada para substituir um gene mutante único, iniciada em 1990, foi a terapia para a deficiência da adenosina desaminase (ADA), a qual consiste em implantar linfócitos T transfectados com retrovírus expressando a proteína ausente nos portadores desta doença. Esta tem sido considerada o protótipo das terapias gênicas e tem sido bem sucedida (ver o artigo de F. Watson, 1993).

VIII. ANÁLISE DE EXPRESSÃO ATRAVÉS DE MICROARRANJOS DE DNA

Como vimos até agora, a tecnologia do DNA recombinante vem permitindo grandes avanços na identificação e caracterização de genes específicos. Com relação aos estudos de expressão gênica, abordagens amplamente utilizadas, como Northern blot e RT-PCR, são bastante sensíveis e precisas, porém permitem o estudo de um ou poucos genes de cada vez. Os Projetos Genoma, além de terem gerado informação sobre as sequências dos genomas de diversos organismos, impulsionaram um grande desenvolvimento tecnológico que resultou no aparecimento de novas metodologias de análises genéticas. Uma dessas novas metodologias, os microarranjos de cDNA, permite não só a análise de expressão de genes específicos mas de padrões de expressão característicos de certos tipos celulares, estágios de desenvolvimento, condições patológicas e muito mais. Com a utilização dessa nova metodologia é portanto possível analisar a expressão de milhares de genes simultaneamente, determinando-se quais deles são expressos em tipos celulares específicos, em momentos e condições particulares, sempre através da comparação de duas situações diferentes.

Como no caso do Northern blot, a metodologia de microarranjos baseia-se na propriedade de hibridação dos ácidos nucleicos. Mas ao contrário do Northern, onde o RNA total de determinadas células ou tecidos está imobilizado em uma membrana de nylon e o fragmento correspondente ao gene de interesse é marcado para a hibridação; no microarranjo, os milhares de fragmentos de DNA que representam genes específicos estão imobilizados num suporte sólido, que pode ser uma membrana de nylon, ou uma lâmina de vidro e o RNA total é que é marcado antes de hibridar com os fragmentos gênicos imobilizados.

A metodologia que usa a lâmina de vidro como suporte é a que permite a análise de maior número de genes. Na figura 48 podemos ver um esquema representando os principais passos de um experimento com microarranjos de DNA.

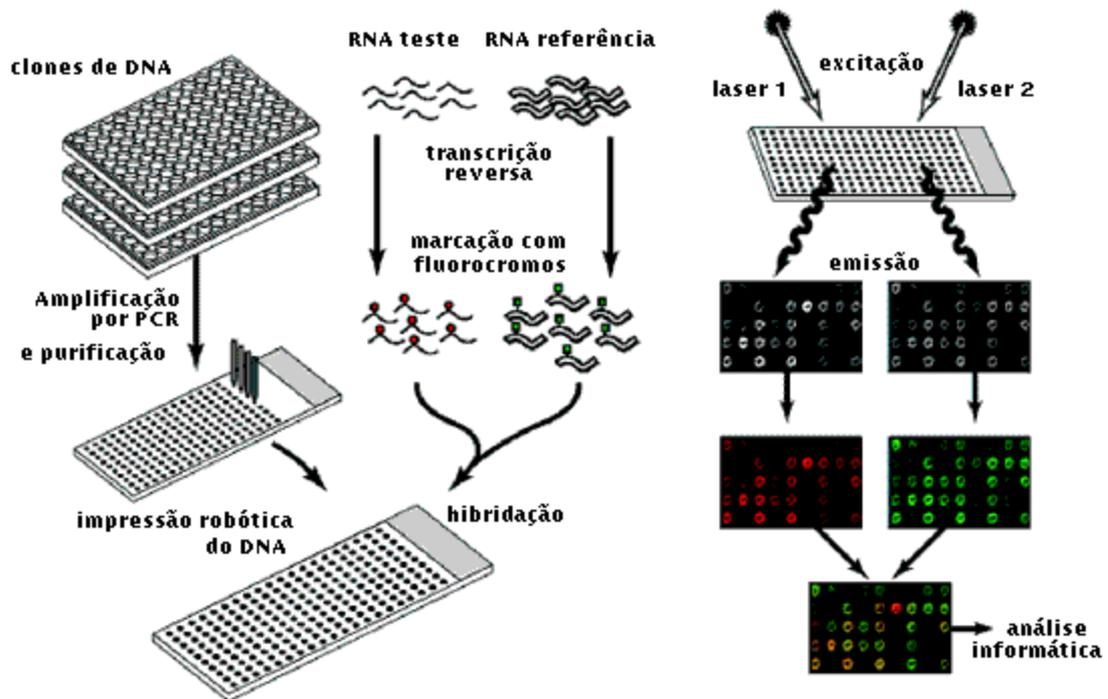


Figura 48. Esquema da preparação e análise de microarranjos de DNA.

Primeiramente, os fragmentos representando genes específicos (derivados, por exemplo, de uma biblioteca de clones de cDNA) são amplificados por PCR e aplicados sobre a lâmina por um robô que automaticamente fixa à lamina os fragmentos de DNA em minúsculos círculos com diâmetro da ordem de 0,1 mm ou menos, formando os *spots*. Cada *spot* na lâmina contém milhares de cópias do fragmento de DNA que representa um determinado gene e sua posição exata fica registrada. Então, o RNA total de células em duas diferentes condições é isolado e marcado com dois fluorocromos (corante fluorescente) distintos. A marcação do RNA geralmente ocorre através da síntese de cDNA utilizando a transcriptase reversa, onde um dos nucleotídeos fornecidos está conjugado com o corante fluorescente. Para o RNA extraído de células na condição 1, a síntese de cDNA é feita fornecendo-se um dos nucleotídeos marcado com um fluorocromo vermelho e para o

RNA extraído de células na condição 2, a transcrição reversa é feita na presença de um nucleotídeo marcado com um fluorocromo verde. Os dois cDNAs marcados são misturados e incubados sobre a lâmina contendo o microarranjo, para que possa ocorrer a hibridação entre o cDNA marcado e os fragmentos de genes na lâmina. Os *spots* correspondentes aqueles genes que forem expressos na condição 1, ficarão marcados de vermelho, pois havia RNA corresponde a ele na amostra da condição 1 que foi convertido em cDNA marcado com o fluorocromo vermelho. Aqueles que forem expressos na condição 2 ficarão marcados de verde. Os genes que forem expressos em quantidades equivalentes nas duas condições ficarão amarelos, dada a sobreposição de marcação com os cDNAs de ambas as situações. Estes resultados são obtidos através da utilização de programas de computador que analisam a imagem do microarranjo hibridizado e quantificam a intensidade do sinal de cada *spot*.

Este tipo de abordagem está sendo amplamente utilizado para o monitoramento de padrões de expressão gênica característicos de diferentes situações, como por exemplo, leveduras crescendo na presença de glicose *versus* leveduras crescendo na presença de álcool ou células de diferentes tipos de câncer em relação a células normais. Isto é possível devido a disponibilização da seqüência genômica dos organismos, no caso da levedura por exemplo, foram analisados microarranjos de DNA contendo cerca de 6000 clones o que representa praticamente todos os genes desse organismo. Nesses experimentos observou-se que quando as leveduras passam de uma condição de fermentação de glicose para crescimento em meio com etanol, seu padrão de atividade gênica se altera visivelmente, sendo que cerca de 900 genes são transcritos mais ativamente e outros 1200 têm sua atividade diminuída.

Os resultados de experimentos com microarranjos de DNA normalmente revelam um grande número de genes que tem sua expressão alterada quando se muda de uma condição para outra. Muitos deles são normalmente genes de função desconhecida e a análise do microarranjo pode revelar um grupo de genes com os quais eles se relacionam através do padrão de expressão, o que pode ajudar bastante na predição da possível função desses genes. Esta análise é feita usando-se uma técnica chamada de *clustering*, através da qual grupos de genes que têm expressão regulada de maneira coordenada podem ser identificados. Os resultados dessa análise são organizados e mostrados em imagens criadas

por programas de análise que permitem a fácil visualização desses grupos e fornecem também a informação sobre o nível de expressão dos mesmos em cada uma das situações. Genes que são ativados ou reprimidos simultaneamente numa dada circunstância podem compartilhar funções, fazendo parte de uma mesma maquinaria multi-protéica ou atuando em processos celulares comuns. Esta é, portanto, uma abordagem de grande utilidade para o atual momento da ciência, a fase pós-genômica ou do genoma funcional.

IX. REFERÊNCIAS BIBLIOGRÁFICAS

- Bristow, A.F.** Recombinant-DNA-derived insulin analogues as potentially useful therapeutic agents. *Trends in Biotechnology* 11: 301-305, 1993.
- Bugg, C.E., Carson, W.M., and Montgomery, J.A.** Drugs by Design. *Scientific American* 269: 60-66, 1994.
- Felici, F., Castagnoli, L., Musacchio, A., Jappelli, R. and Cesareni, G.** *J. Mol. Biol.* 222: 301-310, 1991.
- Fields, S., and Song, O.** Novel genetic system to detect protein-protein interactions. *Nature* 340: 245-246, 1989.
- Little, M., Fuchs, P., Breitling, F., and Dübel, S.** Bacterial surface presentation of proteins and peptides: an alternative to phage technology? *Trends in Biotechnology* 11: 3-5, 1993.
- Luyten, W.H.M.L., and Leysen, J.E.** Receptor cloning and heterologous expression--towards a new tool for drug discovery. *Trends in Biotechnology*, 11:247-254, 1993.
- Ma, J. K-C., and Hein, M. B.** Immunotherapeutic potential of antibodies produced in plants. *Trends in Biotechnology* 13: 522-527, 1995.
- Ma, J. K-C., Hunjan, M., Smith, R., Kelly, C., and Lehner, T.** An investigation into the mechanism of protection by local passive immunization with monoclonal antibodies against *Streptococcus mutans*. *Infection and Immunity* 58: 3407-3414, 1990.
- Monaco, A.P. AND Larin, Z.** YACS, BACs, PACs and MACs: artificial chromosomes as research tools. *Trends in Biotechnology* 12: 280-86, 1994
- Pausch, M. H.** G-protein-coupled receptors in *S. cerevisiae*: high-throughput screening assays for drug discovery. *Trends in Biotechnology* 15: 487-494, 1997.
- Sambrook, J.; E.F.Fritsh and T. Maniatis** - Molecular cloning. A laboratory manual, 2nd ed., Cold Spring Harbor Laboratory Press, 1989.
- Watson, F.** Human gene therapy-progress on all fronts. *Trends in Biotechnology* 11: 114-117, 1993.

ÍNDICE

I. CONCEITOS BÁSICOS	1
1. INTRODUÇÃO	1
2. CONCEITO DE CLONAGEM MOLECULAR	1
3. ENZIMAS DE RESTRIÇÃO	2
4. CONSTRUÇÃO DO DNA RECOMBINANTE	5
5. DNA LIGASE	7
5.1. <i>Tipos de fragmentos de DNA que são ligados pela DNA ligase</i>	7
a) Fragmentos com extremidades coesivas	7
b) Fragmentos com extremidade não coesivas	7
6. TRANSFORMAÇÃO BACTERIANA	8
6.1. <i>Conceito de transformação induzida</i>	8
6.2. <i>Mecanismos de captação do DNA</i>	10
II. VETORES DE CLONAGEM MOLECULAR	11
1. PLASMÍDEO	12
2. FAGOS	13
2.1 <i>Biologia do fago λ</i>	13
2.2 <i>Genoma do fago λ</i>	14
2.3 <i>Controle de expressão dos genes do fago λ</i>	14
2.4 <i>O uso do fago λ como vetor de clonagem molecular</i>	16
3. COSMÍDEO	18
4. VÍRUS	19
4.1. <i>Clonagem no vírus SV40</i>	20
4.1.1. <i>Clonagem no SV40 pela substituição da região tardia</i>	20
4.1.2. <i>Clonagem no SV40 pela substituição da região precoce</i>	22
5. BACTERIÓFAGO M13	22
5.1. <i>Clonagem no bacteriófago M13</i>	24
5.2. <i>Estratégia de clonagem em vetores da série M13mp</i>	25
6. FAGOMÍDEOS	25
6.1. <i>Clonagem no fagomídeo</i>	26
7. VETORES PARA TRANSFORMAÇÃO DE LEVEDURAS	27
III. CONSTRUÇÃO E USO DE BIBLIOTECAS DE DNA RECOMBINANTE	30
1. BIBLIOTECA DE CDNA	31
1.1. <i>Síntese de cDNAs de fita dupla</i>	33
1.2. <i>Preparo dos cDNAs para a ligação com o vetor</i>	34
2. BIBLIOTECA GENÔMICA	36
2.1. <i>Preparo do DNA do inserto</i>	38
2.2. <i>Vetores utilizados na construção biblioteca genômica</i>	39
IV. DETECÇÃO E ANÁLISE DO CLONE RECOMBINANTE	43
1. QUANDO O GENE DE INTERESSE É EXPRESSO NO HOSPEDEIRO	43
2. O USO DE SONDAS MOLECULARES DE ÁCIDOS NUCLÉICOS PARA IDENTIFICAR O GENE DE INTERESSE	45
3. ANÁLISE DO DNA E RNA POR ELETROFORESE EM GEL E “BLOTTING”	51
4. COMO OS GENES CLONADOS SÃO UTILIZADOS?	53

5. MAPA DE RESTRIÇÃO.....	53
V. REAÇÃO EM CADEIA DA POLIMERASE (PCR).....	54
VI. SEQÜENCIAMENTO DE DNA	57
VII. EXPRESSÃO DE PROTEÍNAS HETERÓLOGAS	59
1. INTRODUÇÃO.....	59
2. EXPRESSÃO DE PROTEÍNAS HETERÓLOGAS EM <i>E. COLI</i>	61
2.1 <i>Subclonagem em plasmídeos de expressão</i>	61
2.2 <i>Análise dos plasmídeos e seleção dos subclones corretos</i>	63
3. PRODUÇÃO DE PROTEÍNAS HETERÓLOGAS EM <i>E. COLI</i>	63
3.1. <i>Produção de proteínas híbridas</i>	63
3.2. <i>Produção de proteínas intactas</i>	68
4. EXPRESSÃO DE PROTEÍNAS HETERÓLOGAS EM ORGANISMOS EUCARIOTOS	68
4.1. <i>Expressão em células de mamíferos</i>	68
4.2 <i>Expressão em fungos</i>	69
5. SISTEMA DE EXPRESSÃO EM CÉLULAS DE INSETO UTILIZANDO BACULOVÍRUS COMO VETOR.....	69
6. EXPRESSÃO EM CÉLULAS DE <i>DROSOPHILA</i>	70
7. EXPRESSÃO DE PROTEÍNAS HETERÓLOGAS: APLICAÇÕES E PERSPECTIVAS	71
7.1- <i>Bibliotecas de expressão e identificação de novas proteínas através de anticorpos ou outros ligantes específicos</i>	71
7.1.1- Sistema convencional - expressão de biblioteca de cDNA em <i>E. coli</i>	71
7.1.2. Novos sistemas para a detecção de receptores ou proteínas ligantes	72
7.2. <i>A utilização da tecnologia do DNA recombinante no conhecimento de receptores de membrana</i>	75
7.3. <i>Expressão de proteínas para intervenção terapêutica</i>	77
VIII. ANÁLISE DE EXPRESSÃO ATRAVÉS DE MICROARRANJOS DE DNA	80
IX. REFERÊNCIAS BIBLIOGRÁFICAS	84