

Técnicas de Análise da Regulação da Transcrição Gênica e suas Aplicações na Endocrinologia Molecular

*Chin Jia Lin
Angela Silva Barbosa*

Laboratório de Hormônios e Genética Molecular - LIM/42, Disciplina de Endocrinologia, Departamento de Clínica Médica, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, SP.

*Recebido em 30/06/2002
Aceito em 05/07/2002*

RESUMO

Uma das principais ações dos estímulos hormonais é a modulação da expressão dos genes. Visto que a taxa de transcrição do gene é o maior determinante da sua expressão, os mecanismos moleculares pelos quais a transcrição gênica é regulada têm ganhado interesse crescente e se tornado um dos tópicos principais da Endocrinologia Molecular. Neste artigo, os autores reviram criticamente os aspectos teóricos e as aplicações das técnicas mais utilizadas em estudos da transcrição gênica. As vantagens e os senões dos métodos usados para definição e mapeamento das seqüências regulatórias 5', para teste das interações DNA-proteína, para identificação dos nucleotídeos necessários à interação com fatores de transcrição e para clonagem dos fatores regulatórios *trans* são discutidos. (Arq Bras Endocrinol Metab 2002;46/4:330-340)

Descritores: Regulação da expressão gênica; Transcrição gênica; Fatores de transcrição; Proteínas ligadoras de DNA; Regiões promotoras

ABSTRACT

Techniques for Gene Transcription Regulation Analysis and Their Application in Molecular Endocrinology.

One of the major actions of hormonal stimuli is the modulation of the expression of the genes. Since the rate of gene transcription is the major determinant of gene expression, the molecular mechanisms by which gene transcription are regulated have gained increasing interest and has become a major topic in Molecular Endocrinology. In this paper, the authors critically reviewed both the theoretical aspects and the applications of the techniques mostly used in the studies of gene transcription. The advantages and caveats of methods used for definition and mapping of 5' regulatory sequences, for testing of DNA-protein interactions, in the identification of the nucleotides required for the interaction with transcription factors and in the cloning of transacting regulatory factors are discussed. (Arq Bras Endocrinol Metab 2002;46/4:330-340)

Keywords: Gene expression regulation; Genetic transcription; Transcription factors; DNA-binding proteins; Promoter regions

EM CONDIÇÕES NORMAIS, A INFORMAÇÃO codificada nos genes é transcrita para o RNA mensageiro, que é depois traduzida em proteínas ou subunidades de uma proteína mais complexa. O primeiro passo deste fluxo unidirecional da informação genética - a transcrição dos genes - tem sido alvo de grande interesse da Biologia e da Endocrinologia Molecular. Uma rápida busca nas bases de dados bibliográficos, usando "regulação transcricional" ou "transcrição gênica" como palavras-chave, permite-nos avaliar quanto os mecanismos regulatórios da transcrição gênica interessam à comunidade endocrinológica. Tomando-se como exemplos duas das principais e mais respeitadas revistas de Endocrinologia - *Molecular Endocrinology* e *Endocrinology* - podemos verificar que os trabalhos em transcrição

gênica têm ocupado um espaço cada vez maior (figura 1). Desde a publicação do primeiro número em 1987, o espaço da *Molecular Endocrinology* ocupado pelos trabalhos em transcrição cresceu de 43% para aproximadamente 70% (figura 1). A porcentagem dos trabalhos com enfoque na regulação transcricional publicados na *Endocrinology* cresceu de 12% para mais de 30% entre os anos de 1986 e 2001 (figura 1). Estes dados ilustram bem o grande interesse da Endocrinologia pelos mecanismos de regulação da transcrição gênica, bem como a importância, para os endocrinologistas, do domínio de noções básicas sobre este assunto e do conhecimento das principais técnicas para estudá-lo.

Pode-se afirmar que a Endocrinologia é o ramo da Biologia que se ocupa em estudar todos os aspectos ligados à produção, armazenamento e à ação dos hormônios. Sabendo que uma grande parte das respostas das células/tecidos-alvo decorre de modificações na expressão dos genes (o chamado efeito genômico dos hormônios) fica evidente o porquê do interesse da Endocrinologia nos mecanismos pelos quais a transcrição gênica é regulada. Apesar da expressão dos genes resultar da integração de várias outras etapas regulatórias, variações na transcrição levam, em última análise, a variações na expressão dos genes. Portanto, a compreensão da maquinaria que leva à transcrição ou ao silenciamento dos genes é a chave para entender um dos aspectos mais fundamentais da própria Endocrinologia.

Os genes possuem, além das seqüências que codificam as proteínas (região codificadora), seqüências regulatórias, situadas à montante da região codificadora, que determinam o início e a eficiência da transcrição. Esta modulação é obtida pela interação de fatores reguladores solúveis com as seqüências flanqueadoras 5' dos genes. Os estudos da transcrição gênica se ocupam basicamente da análise desta interação entre os fatores presentes em solução e as seqüências reguladoras, visando

identificar as seqüências e os fatores regulatórios e entender como a interação desses componentes leva à ativação ou ao silenciamento do gene.

O propósito do presente artigo é prover os leitores endocrinologistas com uma breve introdução aos princípios teóricos e às aplicações das principais técnicas de estudo da transcrição gênica. Está fora do escopo deste trabalho uma discussão detalhada sobre os fatores de transcrição e seus mecanismos de ação.

O GENE E OS FATORES DE TRANSCRIÇÃO

Vários conceitos utilizados atualmente para descrever a expressão gênica foram emprestados da Genética. Na década de 60, Jacob e Monod propuseram um modelo no qual um "mensageiro" (que sabemos hoje ser o RNA) estabelece a conexão entre os genes e seus efetores bioquímicos (enzimas ou proteínas estruturais). Implícita no conceito de mensageiro é a necessidade de controle da sua geração. Para isso, postulou-se a existência de um "operador" que serve como um "interruptor" da expressão do gene. O modelo original propõe a existência de repressores que atuam no elemento operador. A expressão do gene poderia ser ativada pelo bloqueio do repressor. Dois outros termos, utilizados inicialmente na Genética para descrever se mutações recessivas ocorrem no mesmo gene (*cis*) ou em genes diferentes (*trans*), são bastante úteis para descrever a relação do gene com o seu "operador" e com os fatores que o regulam. Em contraste com a ligação direta do gene com seu operador, que estão, portanto, em *cis*, os reguladores do operador são fatores solúveis originados de outro compartimento celular e, logo, estão em *trans* em relação ao gene. Estes termos ajudam a correlacionar conceitos genéticos e achados bioquímicos.

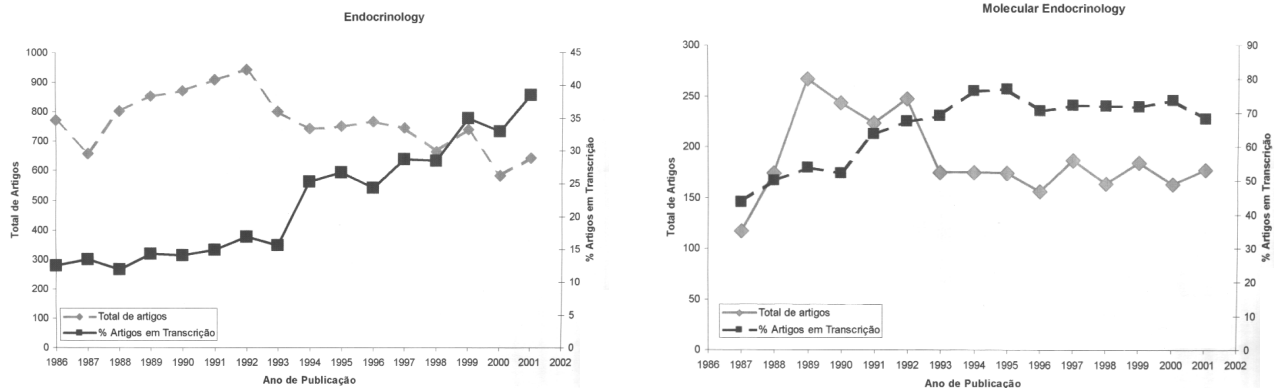


Figura 1. Participação relativa de artigos publicados nos últimos 16 anos, na área de transcrição, nas revistas *Molecular Endocrinology* e *Endocrinology*.

Partindo do modelo de operador podemos definir o gene como uma unidade de transcrição formada pela seqüência que codifica a proteína e as seqüências regulatórias flanqueando os limites 5' e 3' da seqüência codificadora. Os sinais para término da transcrição geralmente ocorrem na extremidade 3' do gene ao passo que a porção localizada à montante da região codificadora tipicamente contém seqüências envolvidas na iniciação da transcrição e outros elementos regulatórios. O termo promotor é muitas vezes utilizado para designar a região regulatória 5' de um gene. Entretanto, esta palavra é utilizada tradicionalmente para descrever as primeiras 100 a 300 bp à montante do início de transcrição e que contém tipicamente o "TATA-box" e seqüências que favorecem a formação de um complexo ativo de transcrição (figura 2). Ao "TATA-box" se liga a TBP (*TATA-binding protein*), um componente-chave do fator de transcrição II D (TF_{II}D, *transcription factor II D*). Em cooperação com outros TF_{II}s e fatores gerais de transcrição, o TF_{II}D forma com a RNA polimerase II o complexo de transcrição (figura 2). Além do promotor, o DNA flanqueador 5' de um gene também possui elementos localizados mais distalmente que se ligam a fatores mais específicos de transcrição e que podem atuar independentemente da orientação. Estes elementos são denominados *enhancers* e podem se localizar a uma grande distância do gene por eles regulados ou estarem contidos num intron. Sabe-se hoje que a distinção entre o promotor e o *enhancer* é menos nítida do que se supunha anteriormente.

Se o promotor e os *enhancers* são componentes que atuam em *cis* com o gene, os fatores de transcrição agem em *trans* para regular a sua expressão. Os fatores de transcrição podem ser gerais quando participam da formação do complexo de transcrição, ou específicos,

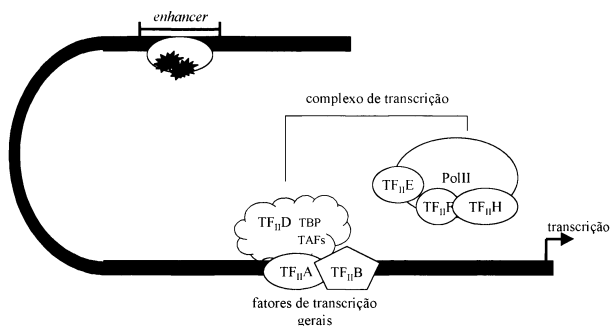


Figura 2. Modelo geral do complexo de transcrição. Interações entre a RNA Polimerase II e os fatores gerais de transcrição são cruciais para que ocorra transcrição em níveis basais. Elementos mais distais, dentre os quais citam-se os *enhancers*, ligam-se a fatores mais específicos e aumentam a taxa de transcrição de genes específicos.

quando reconhecem seqüências específicas de DNA presentes nos *enhancers* e modulam as atividades destes elementos *cis*. Podem ser também classificados como fatores de transcrição os coativadores e co-repressores que não requerem interação direta com DNA e são recrutados pelos dois primeiros para ativar ou reprimir a expressão do gene.

PRINCIPAIS TÉCNICAS UTILIZADAS NOS ESTUDOS DA REGULAÇÃO TRANSCRICIONAL

Quando falamos na análise da maquinaria transcritucional de um gene, estamos nos referindo à delimitação da região regulatória e à caracterização ou identificação dos fatores transcripcionais que nela atuam. Em outras palavras, queremos saber que porções da seqüência situada à montante da região codificadora são relevantes para regular a expressão do gene estudado e quais são as características dos fatores que a elas se ligam (peso molecular aproximado, possível seqüência de reconhecimento, parâmetros cinéticos da sua interação com DNA, formação ou não de oligômeros). O objetivo final é obter a identificação e a clonagem destes fatores. Porções da região promotora, importantes para a transcrição do gene em estudo, podem ser identificadas com o uso de sistemas de promotor/repórter, nos quais fragmentos da seqüência flanqueadora 5' são ligados a um gene heterólogo. Posteriormente o refinamento da localização dos elementos *cis* é feito com técnicas de *footprinting*. As interações específicas DNA-proteína são estudadas por experimentos de retardo da mobilidade na eletroforese. Uma breve discussão de cada uma destas etapas é apresentada nas seções seguintes.

Sistemas de Genes "Repórter"

Um método amplamente empregado para se estudar a expressão de genes eucarióticos é o sistema de gene "repórter". Essa abordagem permite avaliar o efeito que uma seqüência regulatória de um determinado gene - artificialmente inserido à montante de um outro gene, chamado de "repórter" - exerce sobre o nível de transcrição deste último. Em razão desta característica, este é o método de escolha para se delimitar preliminarmente os elementos regulatórios, avaliar os efeitos de agentes hormonais sobre a transcrição gênica e estudar o impacto das mutações em elementos regulatórios *cis*. Este tipo de sistema, utilizado em diferentes tipos de células ou em diferentes ambientes bioquímicos, pode também revelar onde ou em que circunstâncias estão presentes os fatores adequados, necessários para a transcrição.

Primeiramente, constrói-se um plasmídeo quimérico, contendo a seqüência regulatória sob estudo ligada a um gene "repórter". O plasmídeo recombinante é então introduzido em um tipo de célula apropriado e, após um período de 24-72h, o produto do gene "repórter" expresso é quantificado ou avaliado. A quantidade de proteína "repórter" sintetizada sob várias condições ou tratamentos a que as células foram submetidas deve refletir a habilidade da seqüência regulatória inserida em promover a transcrição. Algumas precauções devem ser adotadas ao se utilizar um sistema de expressão transitório. Idealmente, a proteína "repórter" não deve alterar a fisiologia da célula transfectada, nem deve ser sintetizada endogenamente pelo hospedeiro. Se isto ocorrer, é desejável que apresente diferenças em relação à versão "repórter" a fim de que ambas possam ser diferenciadas com segurança. Além disso, é necessário que o "repórter" seja facilmente detectável mesmo em concentrações muito baixas e que, de fato, reflita a expressão do gene *in vivo*, em termos de cinética e magnitude de resposta. Ainda é importante lembrar que esses ensaios permitem a quantificação ou a detecção da atividade ou dos níveis da proteína "repórter", e não dos níveis de RNA. Apesar de ambos estarem freqüentemente correlacionados, um tratamento das células que eventualmente afete a tradução alterará a fidelidade dos valores obtidos (1,2).

Vários sistemas de gene "repórter" são disponíveis atualmente. Cada um apresenta características particulares e algumas limitações. Ao se realizar um experimento, a escolha do sistema mais adequado dependerá do tipo de questão que se pretende abordar e das peculiaridades fisiológicas do modelo sob estudo. Portanto, dependendo da investigação, deve-se optar por uma proteína "repórter" que seja expressa intracelularmente ou, inversamente, que seja secretada e, então, quantificada no meio de cultura. Os ensaios de gene "repórter" podem ser realizados tanto *in vitro* como *in vivo*. Nos ensaios *in vitro*, a proteína "repórter" é quantificada a partir de lisados de células ou tecidos, ou usando-se o meio de cultura das células transfectadas, no caso de proteínas secretadas. A quantificação da proteína, realizada por métodos enzimáticos ou imunológicos, fornece uma estimativa indireta da atividade transcricional do vetor "repórter" que sintetiza a proteína. Nos ensaios *in vivo*, a proteína "repórter" é detectada em células ou tecidos vivos, ou em material fixado para estudos histoquímicos. Diversamente dos ensaios *in vitro*, nesse tipo de abordagem colhem-se, principalmente, informações de caráter qualitativo, podendo-se definir os sítios de produção da proteína "repórter". Dados como a dis-

tribuição tissular de determinados fatores de transcrição ou ainda a especificidade celular de certos promotores, podem ser obtidos com ensaios *in vivo*.

Os genes "repórter" mais amplamente utilizados no estudo da expressão gênica de mamíferos incluem os que codificam enzimas como cloranfenicol acetil-transferase (CAT), luciferase de vaga-lume (*luc*), b-galactosidase bacteriana (*b-gal*), fosfatase alcalina e proteína fluorescente verde (GFP, do inglês *green fluorescent protein*). O gene que codifica o hormônio de crescimento humano (hGH) também é usado. Segue uma breve descrição com ênfase nas aplicações de cada um desses sistemas.

Cloranfenicol Acetil Transferase (CAT)

CAT catalisa a transferência do grupo acetil da acetil-coenzima A, para o substrato, o cloranfenicol (3). O ensaio CAT mais utilizado consiste em incubar o lisado de células, previamente transfectadas, com cloranfenicol marcado com ^{14}C . As formas acetilada e não-acetilada desse composto são separadas por cromatografia de camada fina em placas de vidro siliconizadas (4). A exposição dessas placas a filmes de raios-X fornece uma estimativa qualitativa da atividade de CAT. Uma quantificação dessa atividade enzimática pode ser obtida removendo-se as manchas da placa correspondentes às formas mono e diacetiladas do cloranfenicol e realizando uma contagem em um câmara de cintilação líquida. Existem também métodos de detecção não-isotópicos para quantificar a atividade de CAT.

A principal vantagem do sistema que emprega CAT como "repórter" decorre do fato de se tratar de uma enzima procariótica que apresenta, portanto, atividade endógena mínima em células de mamíferos. Como resultado, obtém-se uma alta taxa de sinal em relação ao *background*. Além disso, CAT apresenta uma meia-vida de cerca de 50h em células de mamíferos e essa estabilidade pode ser satisfatória para experimentos de transfecção transitória. Por outro lado, os ensaios são trabalhosos e longos, o substrato radioativo necessário para a maioria dos experimentos é caro e a sensibilidade do método é inferior àquela exibida por outros sistemas "repórter" desenvolvidos mais recentemente. Esse método se aplica basicamente a ensaios *in vitro*.

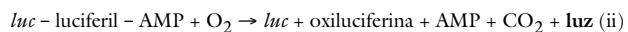
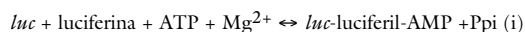
Luciferase de vaga-lume (do inglês, *firefly luciferase - luc*)

O uso da luciferase de vaga-lume como uma enzima "repórter" vem se tornando cada vez mais rotineiro. A enzima luciferase, codificada pelo gene *luc*, clonado do vaga-lume (*Photinus pyralis*) [revisão em (5)], catalisa uma reação de bioluminescência que requer a luciferi-

na como substrato, ATP, Mg²⁺ e O₂. Na presença desses reagentes e mediante a ação da luciferase presente no lisado de células, ocorrerá uma reação de oxidação da luciferina (i e ii) com emissão de um *flash* de luz que decai rapidamente. O sinal luminoso é detectado por um luminômetro. A emissão total de luz é proporcional à atividade da luciferase na amostra que, por sua vez, reflete a taxa de transcrição do gene "repórter" sob ação do promotor em estudo.

O sistema *luc* tem algumas vantagens quando comparado ao sistema CAT usualmente empregado: apresenta uma sensibilidade 10 a 100 vezes superior (6), o processamento das amostras é bem mais rápido, não requer o uso de isótopos, é relativamente mais barato e, como o sistema CAT, apresenta atividade endógena desprezível em células de mamíferos. Por ser sensível à degradação por proteases, a meia-vida da enzima luciferase é de apenas cerca de 3h em células de mamíferos (6,7). Essa característica favorece o emprego desse sistema de gene "repórter" quando se pretende estudar sistemas induzíveis, em que os aumentos observados em relação aos níveis de expressão basal necessitam ser maximizados. Um dos problemas encontrados com o sistema *luc* é a sensibilidade do ensaio a níveis de ATP. Tratamentos que afetam direta ou indiretamente os níveis de ATP intracelular, como forskolina ou prostaglandina E₁, podem causar um aumento da atividade do gene "repórter" *luc* sem, contudo, refletir a ação do promotor em estudo.

Os ensaios de luciferase se aplicam a estudos tanto *in vitro* quanto *in vivo*. A atividade da enzima pode ser detectada em células vivas, *in situ*, quando se usam substratos de luciferase solúveis (ésteres de luciferina) capazes de atravessar a membrana plasmática.



β-Galactosidase (β-gal)

A enzima β-galactosidase, codificada pelo gene *lacZ* da *Escherichia coli* (8), é um dos sistemas "repórter" mais versáteis. Sua atividade pode ser mensurada por métodos colorimétricos ou quimioluminescentes. A clivagem do substrato (várias formas de galactosídeos) pela enzima produz uma solução de coloração amarela, cuja absorvência a 420nm é dada por um espectrofotômetro. No ensaio quimioluminescente, mediante substratos específicos, a atividade da enzima é refletida pela quantidade de emissão luminosa re-

gistrada por um luminômetro. Esse ensaio chega a ser 50.000 vezes mais sensível que o colorimétrico.

A expressão da β-galactosidase é usada como um sistema "repórter" para caracterizar seqüências regulatórias, mas é também freqüentemente adotada como um controle interno para normalizar a variabilidade de outros ensaios "repórter", notadamente CAT e luciferase (9). Nesse caso, a atividade do gene é controlada por um promotor de ação constitutiva, em geral, o promotor do vírus do sarcoma de Roux (RSV, *Roux Sarcoma Virus*).

Além dessas aplicações *in vitro*, níveis da proteína "repórter" β-galactosidase podem ser detectados em células vivas, tecidos seccionados e embriões intactos, na presença do substrato Xgal. Essa reação produz uma coloração azul contrastante com o *background*. O emprego desse sistema em embriões tem permitido a caracterização da expressão de genes específicos em determinados tecidos durante o início do desenvolvimento.

A utilização desse sistema de gene "repórter" deve ser evitada em tipos celulares que têm atividade endógena de β-galactosidase.

Fosfatase alcalina secretada (SEAP, do inglês Secreted Alkaline Phosphatase)

O gene SEAP codifica uma forma de fosfatase alcalina placentária humana, caracterizada pela ausência do domínio de ancoragem à membrana (10,11). Isso confere à proteína a propriedade de ser secretada pelas células e a vantagem de ser quantificada a partir de alíquotas do meio de cultura de células previamente transfectadas. O emprego da proteína SEAP como um sistema "repórter" oferece inúmeras vantagens: múltiplos tratamentos podem ser aplicados às células, uma vez que elas permanecem intactas após as medidas da atividade do gene "repórter", pode-se acompanhar a cinética de expressão gênica a partir de uma mesma cultura e não é necessário preparar lisados de células. Porém, o ensaio não deve ser realizado em tipos de células que expressam baixos níveis de fosfatase alcalina placentária (pulmão, testículo, epitélio do colo uterino). Ainda, espera-se que o desenho experimental não altere a capacidade secretora das células-alvo.

Métodos de detecção colorimétricos, bioluminescentes e quimioluminescentes foram desenvolvidos para esse ensaio.

Hormônio de Crescimento Humano (hGH)

Por apresentar um padrão de expressão restrito às células somatotróficas da adeno-hipófise, o hormônio de crescimento humano é potencialmente interessante como um sistema "repórter" para a maioria dos tipos

de células de mamíferos. Dentre as vantagens conferidas pelo sistema "repórter" hGH destacam-se: detecção direta dos níveis protéicos no meio de cultura e simplicidade e rapidez do ensaio. A quantificação da proteína é feita por radioimunoensaio com a utilização de anticorpos marcados com ^{125}I . O método é menos sensível que o SEAP e é, geralmente, utilizado como controle interno para normalizar outros "repórteres", como SEAP.

Proteína Fluorescente Verde (GFP, do inglês *Green Fluorescent Protein*)

A clonagem do gene GFP de um tipo de água-viva (*Aequorea victoria*) (12,13) possibilitou a expressão da proteína em sistemas heterólogos. A proteína GFP, quando expressa em células procarióticas ou eucarióticas, produz uma fluorescência verde após a excitação das células por luz azul ou UV. Essa propriedade intrínseca da proteína permite que esse sistema "repórter" seja bastante apreciado para avaliar a expressão de genes *in vivo*, principalmente em estudos de desenvolvimento. Como exemplo, citamos a expressão de GFP, sob o controle de um promotor neuronal específico no nemátodo *C. elegans*: a fluorescência decorrente da expressão de GFP permitiu que pesquisadores monitorassem, em tempo real, a formação dos processos neuronais ao longo do desenvolvimento do verme.

Métodos Para o Estudo da Interação DNA-Proteína

Técnicas de Footprinting

Além do emprego dos sistemas de gene "repórter", a localização das seqüências essenciais para a regulação da transcrição pode também ser mapeada pela utilização das técnicas de *footprinting*. Esta técnica baseia-se no fato da ligação de uma proteína a uma fita de DNA proteger este último contra ação de clivagem de um agente químico ou de uma enzima. Assim, quando uma fita de DNA marcada em uma das extremidades sofre a ação degradante de uma substância química ou enzima de maneira controlada (as condições da reação são estabelecidas para que ocorra preferencialmente apenas um ataque por molécula de DNA), uma série de fragmentos de diversos comprimentos é gerada (figura 3a). Todos estes fragmentos terão uma extremidade comum e o seu número e comprimento dependerá da especificidade de seqüência do agente utilizado na reação. Por outro lado, se a ligação de uma proteína a esta fita de DNA produzir um ou mais pontos de proteção contra o ataque, fragmentos de uma determinada faixa de comprimento não serão produzidos quando a reação de clivagem for precedida por incubação da fita de DNA com esta proteína nuclear. Quando produtos de clivagem gerados na presença e na ausência de proteína ligadora de DNA

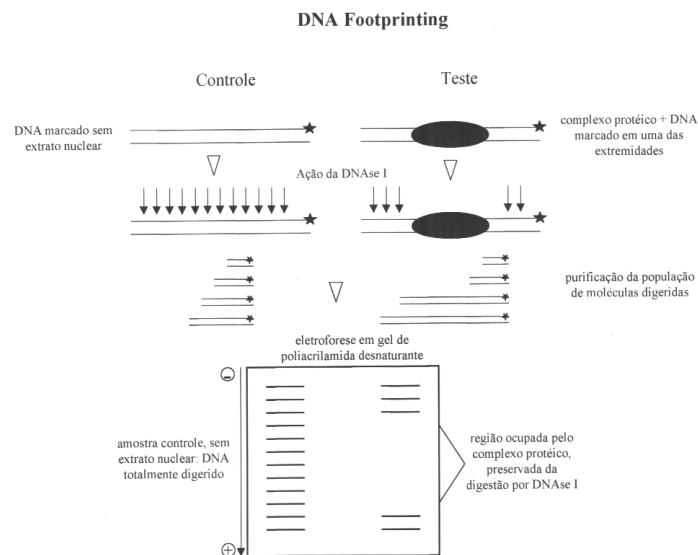


Figura 3a. A posição precisa de interação de uma determinada proteína com um fragmento de DNA marcado radioativamente é revelada por DNA *footprinting*. A enzima DNase é adicionada à mistura em quantidade suficiente para provocar, em média, apenas uma clivagem por molécula de DNA. Na amostra controle não há proteína nuclear. Na amostra teste, a proteína se liga a seu sítio impedindo a ação da DNase nas bases de DNA por ela ocupadas. Após a digestão, as amostras são submetidas à eletroforese em gel de poliacrilamida não-desnaturante e o resultado é visualizado por autoradiografia. Na ausência de proteína (controle) a endonuclease cliva o DNA ao longo de toda sua extensão e um padrão contínuo de bandas é visualizado no gel. Na presença da proteína ligada à sonda, um espaço denominado *footprint* é observado no gel, revelando as bases de DNA que interagem com a proteína.

são resolvidos lado a lado numa eletroforese em condições desnaturantes, observar-se-ão falhas na linha correspondente ao DNA incubado com a proteína. Estas falhas são os *footprints* (pegadas, em inglês) da proteína em questão (figura 3a).

Quando comparadas com o sistema de gene "repórter", as técnicas de *footprinting* permitem um detalhamento mais fino da localização dos elementos regulatórios *cis*. Por outro lado, os sistemas de gene "repórter" apresentam a vantagem de permitir a varredura de grandes porções de seqüência flanqueadora 5' enquanto o *footprinting* é muito mais limitado. Dependendo do agente utilizado o comprimento máximo do fragmento de DNA estudado se situa ao redor de 100 a 200 bp.

Dependendo da resolução desejada, vários agentes podem ser utilizados para *footprinting*. Consideraremos aqui o uso da DNase I, do radical hidroxila e de íons permanganato.

Footprinting com DNase I

A DNase I é uma enzima que se liga à curvatura menor da dupla hélice do DNA e corta a ligação fosfodiéster de ambas as fitas independentemente (14). Ela é uma molécula relativamente grande, o que a impede de atacar a parte do DNA que está muito próxima dos limites da ligação DNA-proteína. Esta propriedade faz com que, na maioria das vezes, o *footprint* de um fator nuclear com a DNase I seja maior do que a área real que esta proteína protege. A vantagem do *footprinting* com DNase I é a relativa facilidade de manipulação dos reagentes utilizados e o não envolvimento do uso de compostos instáveis. Além disso, a maioria das ligações DNA-proteína apresenta *footprints* evidentes com o uso da DNase I. A sua grande desvantagem é a incapacidade de fornecer detalhes mais finos dessa interação, principalmente quando o *footprint* resulta da ligação de um complexo de várias proteínas (15-17).

Footprinting com Radical Hidroxila

O radical hidroxila é muito menor do que a molécula de DNase I. Ele se difunde prontamente e é capaz de acessar bases da molécula de DNA inatingíveis pela DNase I (18). O radical hidroxila pode ser gerado pela reação de peróxido de hidrogênio com [Fe(II)(EDTA)]²⁻ e tem a capacidade de atacar DNA de fita simples ou dupla de maneira independente da seqüência, pela ação na desoxirribose (19,20). Este método apresenta alta resolução, pois o produto de clivagem com o radical hidroxila contém fitas que representam cada uma das bases da seqüência estudada. Ele fornece não apenas detalhes da ligação de DNA-proteína mas também os efeitos decorrentes de modificações na

seqüência de DNA ou da adição de um outro fator ao complexo estudado. No entanto, o grau da proteção dada pelas proteínas contra a atividade do radical hidroxila é muito menor do que a proteção oferecida contra moléculas grandes como DNase I. Por esta razão os *footprints* com radical hidroxila podem ser muito difíceis de serem lidos e interpretados, e muitas vezes requerem o uso de análise densitométrica com computador. Adicionalmente, o radical hidroxila é mais instável do que DNase I, o que aumenta consideravelmente as dificuldades técnicas deste método.

Footprinting com Íon Permanganato

O íon permanganato reconhece especificamente DNA de fita simples e pode ser utilizado para localizar regiões de DNA cuja dupla hélice foi desfeita, identificando, assim, áreas de formação de complexos abertos na região promotora (para possibilitar a transcrição). O permanganato é um agente oxidante virtualmente capaz de reagir com todas as bases (sendo timina a mais reativa) mas é incapaz de quebrar a cadeia de DNA. As modificações são detectadas porque as DNA polimerases são incapazes de continuar a síntese da fita complementar de DNA em presença de uma base modificada pelo permanganato. Assim uma reação de extensão utilizando um iniciador (do inglês *primer*) marcado seguida de eletroforese em condições desnaturantes é o método de escolha para detectar os *footprints* gerados pelo permanganato (21,22). O uso de reação de extensão permite estudar DNAs não marcados ou circulares (plasmídeos, cromossomos) e, desta maneira, avaliar a interação DNA proteína numa configuração mais próxima daquela que é encontrada na célula viva.

ENSAIO DE RETARDO DA MOBILIDADE ELETROFORÉTICA (EMSA)

O método de retardo em gel é uma das formas mais rápidas de se analisar a interação entre DNA e proteína(s) *in vitro*. Simples do ponto de vista conceitual e técnico, a estratégia se fundamenta no fato de que a mobilidade eletroforética de um ácido nucléico se altera quando uma proteína a ele se liga. Essa técnica se popularizou a partir das publicações de Garner e Revzin (23) e Fried e Crothers (24). A proteína de interesse, ou uma preparação de proteínas obtida de um determinado tipo celular sob estudo, é incubada com segmentos de DNA de fita dupla marcados radioativamente, as sondas, para permitir a formação de um complexo. A mistura é, então, submetida à eletroforese sob condições não desnaturantes. O complexo DNA/proteína formado apresenta, em geral, uma

mobilidade eletroforética mais lenta quando comparada àquela exibida pela sonda livre e as bandas correspondentes ao DNA ligado à proteína e ao DNA livre são visualizadas por autorradiografia (figura 3b).

As interações DNA/proteína detectadas por EMSA não são necessariamente específicas e, portanto, realizam-se ensaios de competição para se discriminar um evento de ligação inespecífico de um específico. Nesses experimentos, utilizam-se, geralmente, dois tipos de sondas não radioativas, uma delas contendo uma seqüência qualquer, diferente do suposto sítio de reconhecimento da proteína (competidor inespecífico), e a outra representando a própria sonda original não marcada (competidor específico). Se a interação inicial entre a sonda marcada e a proteína (ou extrato nuclear) for específica, o competidor específico, adicionado à mistura em quantidades crescentes e superiores à sonda marcada, ligar-se-á preferencialmente à proteína, e o complexo, representado pela banda de maior peso molecular no gel, deixa de ser visualizado (figura 3b). Por outro lado, a adição de um competidor inespecífico não altera a aparência do complexo original, mesmo em concentrações muito elevadas (figura 3b).

A identidade da proteína que forma o complexo pode ser revelada com o uso de anticorpos específicos dirigidos contra fatores protéicos suspeitos. O anticorpo se liga à proteína, formando um “super-complexo” de mobilidade eletroforética ainda mais reduzida (figura 3b) ou pode inibir a formação do complexo DNA-proteína. Em ambos os casos, a identidade do fator suspeito de formar o complexo é confirmada.

Em suma, o ensaio de retardo da mobilidade eletroforética é rotineiramente usado para verificar se uma proteína se liga a um determinado sítio de uma molécula de DNA, ou ainda se um dado extrato nuclear contém uma proteína com capacidade de interagir com uma sonda particular. Com algumas modificações, o método pode ainda servir para medir parâmetros cinéticos e avaliar a especificidade da interação DNA/proteína (25-27). Mutações e deleções podem também ser introduzidas nas sondas para verificar qual(is) a(s) base(s) importantes para estabelecer contato de DNA-proteína. Por estas razões, esta técnica tem sido uma das mais aplicadas no estudo da interação entre os elementos *cis* e os fatores *trans*.

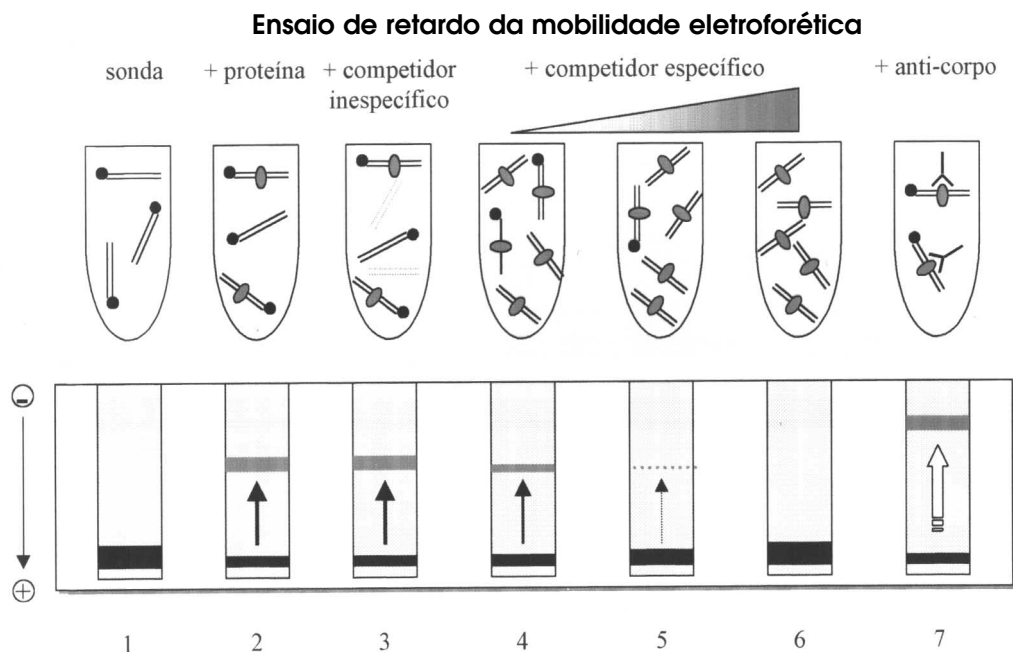


Figura 3b. Um fragmento de DNA marcado radioativamente, contendo o suposto sítio de interação com uma determinada proteína, é incubado com extrato nuclear obtido do tipo celular em estudo. Após um breve período de incubação, a mistura é submetida à eletroforese em gel de poliacrilamida não-desnaturante e o resultado é visualizado por autorradiografia. 1) amostra controle, contendo apenas a sonda marcada, revelada por uma única banda na extremidade do gel; 2) complexo DNA-proteína, representado pela banda de maior peso molecular; 3) adição de um competidor inespecífico não altera a aparência do complexo original; 4, 5, 6) adição de competidor específico não-marcado em concentrações crescentes e superiores à sonda marcada causam o desaparecimento gradual da banda equivalente ao complexo DNA-proteína; 7) o anticorpo adicionado reage com a proteína do complexo, reduzindo ainda mais sua mobilidade eletroforética.

ENSAIOS DE INTERFERÊNCIA

As bases de uma determinada seqüência de DNA podem ser modificadas quando esta é incubada com alguns agentes químicos. Se a modificação ocorrer em bases essenciais à interação do DNA com um fator nuclear esta ligação deixará de ocorrer. Este fenômeno é conhecido como “interferência por modificação”. Os ensaios de interferência aproveitam esta propriedade do DNA modificado para identificar as bases essenciais no contato de DNA com a proteína. Num experimento típico, uma sonda marcada é tratada com um agente químico. Esta reação é padronizada em condições tais que, em média, cada molécula receba apenas uma modificação. Assim, uma série de sondas marcadas que carregam uma modificação em uma de suas bases é gerada. Em seguida estas sondas parcialmente modificadas são incubadas com proteína purificada ou extrato nuclear e submetidas à eletroforese em condições muito semelhantes às do EMSA. As sondas cujas modificações nos nucleotídeos não interferiram na sua ligação com proteína formarão complexos DNA-proteína que migrarão como uma banda retardada. As sondas que sofreram alterações em bases críticas para a ligação protéica não formarão complexos e migrarão como sondas livres. As frações livre e retardada poderão ser recortadas, as sondas eluídas e submetidas a tratamento com um segundo agente químico para clivar especificamente apenas as bases modificadas. Quando resolvidas em paralelo num gel desnaturante, a presença de algumas bandas que ocorrem apenas na fração livre e não na ligada revelará as bases essenciais para o contato do DNA com proteína (28). Os agentes químicos que podem ser utilizados para ensaios de interferência são dimetil-sulfato (DMS), N-etil-N-nitrosouréia e dietil-pirocarbonato (DEPC). Desses, o DMS, que metila as purinas, é o agente mais utilizado (29).

Caracterização e Clonagem de Fatores de Transcrição

A etapa que se segue após o mapeamento dos sítios regulatórios relevantes e a identificação das bases críticas para o contato dos elementos *cis* com os fatores regulatórios é a caracterização e, se possível, a identificação ou clonagem do fator ligante dos sítios estudados. Não há uma diferenciação nítida entre esta etapa e as anteriores e essa divisão foi adotada neste artigo por motivos didáticos. Na verdade, muito da caracterização dos fatores *trans* é feita nas fases descritas em outras seções deste trabalho. Por exemplo, o *footprinting* com radical hidroxila permite identificar, com resolução de um único nucleotídeo, as bases protegidas pelo contato do DNA

com proteína. Combinando este dado com aqueles obtidos pelos ensaios de interferência, isoladamente ou combinado com EMSA, podemos mapear a seqüência de reconhecimento do fator que se liga aos elementos *trans* analisados. Esta seqüência pode ser utilizada para consultar os bancos de dados para uma eventual identificação do fator. Finalmente o EMSA realizado com adição de anticorpos gerados contra fatores de transcrição conhecidos permite que excluamos ou confirmemos a participação dos fatores testados na formação do complexo DNA-proteína. Portanto, informações acumuladas em outras fases de estudo da maquinaria regulatória de um gene são sempre úteis para a identificação do fator de transcrição. Nesta seção discutiremos o uso de hibridação Southwestern (do inglês *Southwestern blotting*) e da coluna de afinidade à DNA na caracterização e purificação de fatores de transcrição.

HIBRIDAÇÃO SOUTHWESTERN

A hibridação *Southwestern* é um procedimento desenvolvido para caracterização de proteínas que têm a propriedade de se ligar ao DNA, especialmente aquelas não descritas, mas cujas seqüências de reconhecimento são bem conhecidas (30,31). Neste método, a mistura de proteína (por exemplo, extrato nuclear) é separada por eletroforese em gel nativo ou desnaturante e depois transferida a uma membrana. Esta membrana é então hibridada com a sonda de alta atividade específica representada pela seqüência de reconhecimento do fator que se pretende caracterizar. A detecção da proteína de interesse é feita por autorradiografia. Por esse método obtêm-se informações tais como o peso molecular aproximado do fator a ser caracterizado ou a presença de isoformas deste mesmo fator. Entretanto, é um procedimento que apresenta um alto grau de dificuldade técnica. O mesmo princípio de hibridação DNA-proteína em membrana do *Southwestern* pode ser aplicado para clonagem de fator de transcrição (32-34). Neste caso, uma biblioteca de expressão em bacteriófago é preparada a partir do cDNA da célula (ou tecido) que expressa o fator de interesse. Esta biblioteca é então utilizada para infectar *Escherichia coli*. Dentro dos hospedeiros os bacteriófagos começarão a se replicar e lisarão as bactérias hospedeiras. As proteínas codificadas pelos cDNAs da biblioteca serão incorporadas como parte da cápside do bacteriófago. Um filtro, que é a réplica da coleção de clones que se formam na superfície da placa de cultura, pode ser preparado e posteriormente hibridado como a membrana do Southwestern e as colônias positivas selecionadas para estudos adicionais.

PURIFICAÇÃO DOS FATORES DE TRANSCRIÇÃO POR CROMATOGRAFIA DE AFINIDADE POR DNA

A capacidade de um fator de transcrição de reconhecer, com alta afinidade e de maneira específica, o elemento *cis* em que ele atua pode ser também utilizada para a sua purificação. Nesta abordagem, extratos protéicos parcialmente purificados são submetidos à separação em colunas de cromatografia preenchidas com resina embebida de DNA cuja seqüência é aquela reconhecida pelo fator de interesse. A fração do extrato que contém o fator de interesse se ligará ao DNA presente na resina e será retida, ao passo que as outras proteínas não se ligarão ao DNA da coluna e sairão rapidamente durante a lavagem desta (35,36). As proteínas assim purificadas poderão ser depois submetidas a análises adicionais para a sua identificação. A grande limitação deste método é a baixa abundância dos fatores de transcrição. Por esta razão, um grande número de células precisa ser utilizado para a preparação do extrato que será aplicado à cromatografia.

AGRADECIMENTOS

Este trabalho foi realizado com apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP, Processos FAPESP: 00/11362-0 (concedido a C.J.L.) e 01/12571-5 (concedido a A.B.S.).

LEITURAS RECOMENDADAS

Kain SR, Ganguly S. Overview of genetic reporter systems. In: Ausubel FM, ed. **Short Protocols in Molecular Biology**. 3rd Edition. New York: Wiley & Sons, 1997. p 9.18 – 9.28.

Johnson W, Jameson JL. Transcription control of genes. In: Jameson JL, ed. **Principles of Molecular Medicine**. 1st Edition. Humana Press, 1998.

REFERÊNCIAS

1. Alam J, Cook JL. Reporter genes: application to the study of mammalian gene transcription. **Anal Biochem** 1990;188(2):245-54.
2. Bronstein I, et al. Chemiluminescent and bioluminescent reporter gene assays. **Anal Biochem** 1994;219(2):169-81.
3. Shaw WV. Chloramphenicol acetyltransferase from chloramphenicol-resistant bacteria. **Methods Enzymol** 1975;43:737-55.
4. Gorman CM, Moffat LF, Howard BH. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. **Mol Cell Biol** 1982;2(9):1044-51.

5. Wood KV. Marker proteins for gene expression. **Curr Opin Biotechnol** 1995;6(1):50-8.
6. Pazzagli M et al. Use of bacterial and firefly luciferases as reporter genes in DEAE-dextran-mediated transfection of mammalian cells. **Anal Biochem** 1992;204(2):315-23.
7. Thompson JF, Hayes LS, Lloyd DB. Modulation of firefly luciferase stability and impact on studies of gene regulation. **Gene** 1991;103(2):171-7.
8. Hall CV et al. Expression and regulation of Escherichia coli lacZ gene fusions in mammalian cells. **J Mol Appl Genet** 1983;2(1):101-9.
9. Hollon T, Yoshimura FK. Variation in enzymatic transient gene expression assays. **Anal Biochem** 1989;182(2):411-8.
10. Berger J et al. Secreted placental alkaline phosphatase: a powerful new quantitative indicator of gene expression in eukaryotic cells. **Gene** 1988;66(1):1-10.
11. Kain SR. Use of secreted alkaline phosphatase as a reporter of gene expression in mammalian cells. **Methods Mol Biol** 1997;63:49-60.
12. Chalfie M et al. Green fluorescent protein as a marker for gene expression. **Science** 1994;263(5148):802-5.
13. Prasher DC et al. Primary structure of the *Aequorea victoria* green-fluorescent protein. **Gene** 1992;111(2):229-33.
14. Galas DJ, Schmitz A. DNase footprinting: a simple method for the detection of protein-DNA binding specificity. **Nucleic Acids Res** 1978;5(9):3157-70.
15. Brenowitz M et al. "Footprint" titrations yield valid thermodynamic isotherms. **Proc Natl Acad Sci USA** 1986;83(22):8462-6.
16. Brenowitz M et al. Quantitative DNase footprint titration: a method for studying protein-DNA interactions. **Methods Enzymol** 1986;130:132-81.
17. Dabrowiak JC, Ward B, Goodisman J. Quantitative footprinting analysis using a DNA-cleaving metalloporphyrin complex. **Biochemistry** 1989;28(8):3314-22.
18. Tullius TD, Dombroski BA. Hydroxyl radical "footprinting": high-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. **Proc Natl Acad Sci USA** 1986;83(15):5469-73.
19. Tullius TD et al. Hydroxyl radical footprinting: a high-resolution method for mapping protein-DNA contacts. **Methods Enzymol** 1987;155:537-58.
20. Tullius TD. DNA footprinting with hydroxyl radical. **Nature** 1988;332(6165):663-4.
21. Sasse-Dwight S, Gralla JD. Footprinting protein-DNA complexes *in vivo*. **Methods Enzymol** 1991;208:146-68.
22. Sasse-Dwight S, Gralla JD. KMnO₄ as a probe for lac promoter DNA melting and mechanism *in vivo*. **J Biol Chem** 1989;264(14):8074-81.
23. Garner MM, Revzin A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. **Nucleic Acids Res** 1981;9(13):3047-60.
24. Fried M, Crothers DM. Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. **Nucleic Acids Res** 1981;9(23):6505-25.

25. Fried MG. Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. **Electrophoresis** **1989**;10(5-6):366-76.
26. Revzin A. Gel electrophoresis assays for DNA-protein interactions. **Biotechniques** **1989**;7(4):346-55.
27. Carey J. Gel retardation. **Methods Enzymol** **1991**;208:103-17.
28. Wissmann A, Hillen W. DNA contacts probed by modification protection and interference studies. **Methods Enzymol** **1991**;208:365-79.
29. Siebenlist U, Gilbert W. Contacts between Escherichia coli RNA polymerase and an early promoter of phage T7. **Proc Natl Acad Sci USA** **1980**;77(1):122-6.
30. Knepel W, Jepeal L, Habener JF. A pancreatic islet cell-specific enhancer-like element in the glucagon gene contains two domains binding distinct cellular proteins. **J Biol Chem** **1990**;265(15):8725-35.
31. Tully DB, Cidlowski JA. Protein-blotting procedures to evaluate interactions of steroid receptors with DNA. **Methods Enzymol** **1993**;218:535-51.
32. Singh H et al. Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. **Cell** **1988**;52(3):415-23.
33. Vinson CR, et al. *In situ* detection of sequence-specific DNA binding activity specified by a recombinant bacteriophage. **Genes Dev** **1988**;2(7):801-6.
34. Singh H. Specific recognition site probes for isolating genes encoding DNA-binding proteins. **Methods Enzymol** **1993**;218:551-67.
35. Kadonaga JT, Tjian R. Affinity purification of sequence-specific DNA binding proteins. **Proc Natl Acad Sci USA** **1986**;83(16):5889-93.
36. Kadonaga JT. Purification of sequence-specific binding proteins by DNA affinity chromatography. **Methods Enzymol** **1991**;208:10-23.

Endereço para correspondência:

Chin Jia Lin
Laboratório de Hormônios e Genética Molecular
Av. Dr. Enéas de Carvalho Aguiar, 155
Prédio dos Ambulatórios (PAMB), 2º Andar, Bloco 6
05403-900 São Paulo, SP