

BC/29376

IB/80232



UNICAMP

T/UNICAMP

Si38
g

UNIVERSIDADE ESTADUAL DE CAMPINAS - UNICAMP
INSTITUTO DE BIOLOGIA
DEPARTAMENTO DE GENÉTICA E EVOLUÇÃO



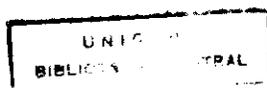
**γ -coixina: isolamento do clone
genômico e caracterização da região
promotora**

Este exemplar corresponde à redação final
da tese de Felipe Rodrigues da Silva
e aprovada pela Comissão Julgadora.
5/11/76 *Adilson Leite*

Felipe Rodrigues da Silva
orientador: Dr. Adilson Leite

Tese apresentada ao Instituto de Biologia da
Universidade Estadual de Campinas
(UNICAMP) para obtenção do grau de
Mestre em Ciências, área de concentração
Genética de Plantas.

1996



UNIDADE T.P.
N.º CHAMADA:
T/UNICAMP
Δ1389
V. 8
T. 29376
P. 668196
C D
PRECO. 891,00
DATA 12-21-96
N.º CPD

C.M.00095404-5

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA CENTRAL DA UNICAMP

Si38g

Silva, Felipe Rodrigues da

γ -coixina: isolamento do clone genômico e
caracterização da região promotora / Felipe Rodrigues
da Silva. -- Campinas, SP : [s.n.], 1996.

Orientador: Adilson Leite.

Tese (mestrado) - Universidade Estadual de
Campinas, Instituto de Biologia.

1. Clonagem Molecular. 2. Plantas - Genética
Molecular. 3. Regulamento Genético. 4. Coix.
5. Fatores de Transcrição 6. Proteínas de Plantas.
7. Cereais 8. Gramínea - Semente. I. Leite, Adilson.
II. Universidade Estadual de Campinas. Instituto de
Biologia. III. Título.

Campinas, 05 de Novembro de 1996

Banca Examinadora:

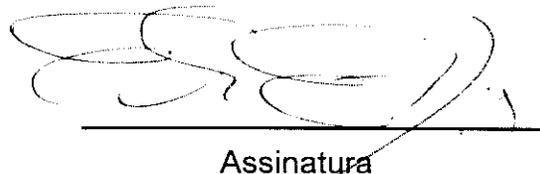
Titulares

Prof. Dr. Francisco J. F. Lara



Assinatura

Prof. Dr. Gonçalo A. Guimarães Pereira



Assinatura

Prof. Dr. Adilson Leite



Assinatura

Suplente:

Prof. Dr. Éberson Sanches Calvo



Assinatura

APROVADA

*“Ah, recomeçar!
Recomeçar como canções e epidemias.*

*“Ah! Recomeçar como as colheitas,
Como a lua e a covardia.*

*“Ah! Recomeçar como a paixão
E o fogo, e o fogo...”*

João Bosco, Caça à Raposa.

*“Coragem, ânimo, disse para mim, não pensemos mais na
Sapiência: pede à Ciência que te ajude.”*

Casabon, personagem de “O Pêndulo de Foucault”, de Umberto Eco

Aos meus pais,
pelo exemplo de coragem,
pelo meu nome,
por todo o carinho e incentivo, sempre,
e, hoje, por terem me mostrado uma série de vezes
como escrever uma tese.

À Tita,

pela amizade, imediata e sempre sincera,
pela alegria e disposição diante de qualquer situação,
por todas as ajudas (incontáveis) ao longo destes oito anos,
pelas farras, pelas festas, pelos telefonemas.
Pelo amor...

“Ah!, ilusão, o amor não é risco na areia, é desenho de giz.”

*“Quando não tinha nada, eu quiz.
Quando tudo era ausência, esperei...”*

*“Se você vier, pro que der e vier,
Comigo...”*

Sem você, nem os versos destas, ou de qualquer
outra canção, iriam valer.

Agradecimentos

Ao meu orientador Adilson Leite e ao amigo Adilson. Ambos fundamentais na minha formação científica. Obrigado pelo bom humor, pela paciência, pela sapiência.

Ao Prof. Dr. Paulo Arruda, por me proporcionar espaço em seu laboratório e pelas discussões acerca do meu trabalho e dos rumos do mundo.

A todos os colegas do laboratório, que auxiliaram sempre, de uma forma ou outra, o desenvolvimento do meu trabalho. Mais especialmente, ao Germano, pela montagem do banco genômico, ao Márcio, pelo auxílio nos experimentos de expressão transitória e à Silvinha pelas inúmeras vezes em que fez minhas suas mãos na bancada. A ela, e ao Dedé, o obrigado por sermos da família FAS. Ao Raul e ao Rodrigo, a confiança de que nossa FARRA entrará para a história deste laboratório. Ao Mano, por um monte de ótimos motivos, mas especialmente por ter sido o primeiro a me convidar para trabalhar no CBMEG.

De novo, a todos os colegas, pelo convívio agradável.

Aos Prof. Dr. Éberson Sanches Calvo, Nilce Meireles e Laura Ottoboni, pela leitura preeliminar deste trabalho e conselhos que o aprimoraram.

Ao seu Chico, pela boa disposição no cuidado do nosso material de campo, ao Maurício, pelo zelo com nossa vidraria e a Tânia, Sandrinha, Claudia e Fabiana pela organização de nossa burocracia.

À CAPES pelo apoio financeiro.

Índice Geral

RESUMO	10
ABSTRACT	12
ABREVIACÕES	14
PREÂMBULO:	15
INTRODUÇÃO	16
A utilidade do estudo de genes de reserva de cereais	16
OBJETIVOS	21
REVISÃO BIBLIOGRÁFICA	23
Cereais:	23
Coix:	24
A semente	27
Proteínas de reserva	31
Mecanismos de regulação	47
MATERIAIS E MÉTODOS	68
Material vegetal	68
Linhagens de bactérias	68
Purificação de DNA de bacteriófago λ	70
Isolamento de DNA nuclear de planta	72
Isolamento em larga escala de DNA de plasmídeo	73
Purificação de fragmentos de DNA a partir de géis de agarose	75
Extração de proteínas nucleares de sementes de Coix	76
Construção da biblioteca genômica	77
Southern Blot	79
Seqüenciamento de DNA	79
Deleção unidirecional com a enzima Exonuclease III	80
Bombardeamento com micropartículas	82
Experimentos de retardamento em gel	89
Métodos computacionais	90
RESULTADOS	92
Seleção do clone genômico.....	92
Seqüenciamento	92
Comparação com seqüências já descritas	98
Experimentos de expressão transitória.....	105
Experimentos de retardamento em gel.....	107
DISCUSSÃO	116
Análise da seqüência do promotor	116
Experimentos de expressão transitória.....	125
Experimentos de retardamento em gel	128
CONCLUSÕES	132
BIBLIOGRAFIA	133

Índice de Figuras e Tabelas

FIGURA 1: ESQUEMA DE UM RAMO DE <i>COIX LACRYMA-JOBI</i>	26
FIGURA 2: CORTE ESQUEMÁTICO DA SEMENTE DE MILHO.	28
FIGURA 3: DESENVOLVIMENTO DOS CORPÚSCULOS PROTÉICOS EM ENDOSPERMA DE MILHO. .	44
FIGURA 4: ESQUEMA DO PLASMÍDEO PDMC202	86
FIGURA 5: ANÁLISE DE RESTRIÇÃO DO CLONE GENÔMICO λ 23.1.....	93
FIGURA 6: ANÁLISE DE RESTRIÇÃO DO CLONE PE/H γ CX23.1.	95
FIGURA 7: MAPA DE RESTRIÇÃO DO CLONE λ 23.1 E ESTRATÉGIA DE SEQÜENCIAMENTO DO CLONE PE/H γ CX23.1.	97
FIGURA 8: SEQÜÊNCIA COMPLETA DO GENE DA γ -COIXINA E TRADUÇÃO DA REGIÃO CODIFICADORA.	99
FIGURA 9: ANÁLISE DA REGIÃO PROMOTORA DOS GENES DE γ -PROLAMINAS.	102
FIGURA 10: ALINHAMENTO DOS PROMOTORES DE γ -PROLAMINA.	106
FIGURA 11: FRAGMENTOS DO PROMOTOR DE γ -COIXINA EMPREGADOS NOS EXPERIMENTOS DE EXPRESSÃO TRANSITÓRIA E RETARDAMENTO EM GEL.	108
FIGURA 12: RESULTADOS DOS EXPERIMENTOS DE EXPRESSÃO TRANSITÓRIA.....	110
FIGURA 13: EXPERIMENTO DE RETARDAMENTO EM GEL DO FRAGMENTO 226.	112
FIGURA 14: EXPERIMENTO DE RETARDAMENTO EM GEL DOS FRAGMENTOS 226 E 75 SIMULTANEAMENTE.	113
FIGURA 15: EXPERIMENTO DE RETARDAMENTO EM GEL DO FRAGMENTO 319.	114
FIGURA 16: EXPERIMENTO DE RETARDAMENTO EM GEL DOS FRAGMENTOS 177 E 142.	115
TABELA I: OLIGONUCLEOTÍDEOS UTILIZADOS NA CONSTRUÇÃO DOS INSERTOS EMPREGADOS NOS EXPERIMENTOS DE EXPRESSÃO TRANSITÓRIA.	83
TABELA II: OLIGONUCLEOTÍDEOS UTILIZADOS NA CONSTRUÇÃO DOS INSERTOS EMPREGADOS NOS EXPERIMENTOS DE EXPRESSÃO TRANSITÓRIA.	84
TABELA III: FRAGMENTOS PROVENIENTES DO CLONE PE/H γ CX23.1 CLONADOS EM pBLUESCRIPT KS+.	94
TABELA IV: FRAGMENTOS UTILIZADOS NO SEQÜENCIAMENTO.....	96
TABELA V: REGIÕES DE ALTA SIMILARIDADE COM SEQÜÊNCIAS ALVO DE FATORES DE TRANSCRIÇÃO CONHECIDOS.	103

Resumo

As prolaminas, proteínas de reserva das sementes de cereais, são codificadas por uma série de genes regulados coordenativamente de forma específica ao nível de tecido e estágio de desenvolvimento. Estas proteínas são agrupadas em classes em função das similaridades na estrutura primária e conseqüente solubilidade. A expressão dos genes que as codificam é regulada principalmente ao nível da transcrição.

As prolaminas de *Coix* são chamadas coixinas e representam mais de 70% do conteúdo protéico do endosperma. De acordo com sua solubilidade diferencial, as coixinas são classificadas em α - e γ -coixinas. Os dados moleculares obtidos até o momento indicam um alto grau de similaridade entre os genes que codificam as prolaminas de milho, *Coix* e sorgo.

A expressão coordenada das prolaminas se dá através de mecanismos comuns a todas as classes. Entretanto, estudos recentes sobre a mutação *opaco-2* mostram claramente que as α - e β -prolaminas são reguladas também por um mecanismo exclusivo para estas classes. Tal fato sugere o envolvimento vários mecanismos regulando a transcrição dos genes de prolaminas.

Devido principalmente à mutação *opaco-2*, os mecanismos envolvidos na regulação gênica das α -prolaminas são bem estudados. Pouco é conhecido sobre estes mecanismos nas γ -prolaminas. Entretanto, enquanto as γ -prolaminas são codificadas por, no máximo, dois genes, as α -prolaminas são codificadas por uma família multigênica. Tal fato torna as γ -prolaminas modelos interessantes no melhoramento molecular de cereais.

Nesta tese o gene da γ -coixina (λ GCX23.1) foi isolado de uma biblioteca genômica construída no fagemídeo λ -dash, utilizando-se um clone de cDNA de γ -coixina como sonda. A seqüência da região promotora do gene foi alinhada com as seqüências dos promotores das γ -prolaminas de milho e sorgo, revelando uma série de elementos conservados. Alguns destes elementos são

semelhantes a elementos *cis* já descritos, como o *prolamin box* e os sítios de ligação dos fatores GCN4 de leveduras e C/EBP de animais.

Experimentos de expressão transitória comprovaram que o promotor do gene da γ -coixina é capaz de dirigir a expressão do gene da β -glicoronidase em endosperma imaturo de milho. A análise de deleção do promotor sugere que diversos elementos podem estar envolvidos na regulação de sua expressão. Experimentos de retardamento em gel provaram que três segmentos distintos do promotor da γ -coixina ligam-se a fatores nucleares de maneira específica.

Abstract

Prolamins, the major cereal storage proteins, are encoded by genes coordinately regulated in a tissue- and developmental stage-specific manner during seed development. There are several prolamin genes encoding proteins of discrete molecular sizes, which are divided in classes according to similarities in their primary structures and solubility. The expression of these genes is regulated primarily at the transcriptional level.

The *Coix* prolamins, known as coixins, represent more than 70% of the total endosperm protein. Coixins can be separated in two fractions based on their differential solubility: α - and γ -coixins. The overall molecular data shows a high degree of similarity among the genes encoding the prolamins of maize, *Coix* and Sorghum.

The coordinated expression of the prolamin genes is accomplished by a common regulatory mechanism. Still, recent studies on the *opaque-2* mutation clearly demonstrated a distinct transcriptional regulation at the class level for the α - and β -prolamins. Therefore, multiple complex regulatory mechanisms are conceivably involved in the regulation of prolamin genes.

Although the mechanisms of gene regulation for the α -prolamins are well studied, mostly because of the *opaque-2* mutant of maize, little is known about the mechanisms controlling the γ -prolamin genes. Also, while the α -prolamins are encoded by multigene families, the γ -prolamins are encoded by one or two genes, which heightens the importance of the γ -prolamins in molecular improving of cereals.

In this work the γ -coixin gene (λ GCX23.1) was isolated from a *Coix* genomic λ -dash library. The library was screened using a γ -coixin cDNA clone as probe.

In order to elucidate the transcriptional regulation of the γ -prolamin genes, the γ -coixin promoter sequence was aligned with the promoter

sequences of the maize and sorghum γ -prolamins genes. Several conserved motifs were found in the upstream region of these genes. Some of these motifs resembles *cis*-acting elements, such as the cereal prolamin box and the binding sites for yeast GCN4 and animal C/EBP transcriptional factors.

Transient expression assays demonstrated the ability of the γ -coixin promoter to drive the expression of the β -glucuronidase reporter gene in immature maize endosperm. Deletion analysis of the promoter sequence suggested that several elements might be involved in its regulation. At least three distinct segments of the γ -coixin promoter bind specifically to nuclear factors, as demonstrated by means of gel retardation assay.

Abreviações

2-ME	β -mercapto-etanol
ATP	adenosina 5'-trifosfato
bHLH	<i>basic Helix-Turn-Helix</i>
CaMV	vírus do mosaico da couve-flor
cpm	cintilações por minuto
CTD	domínio carboxi-terminal
DO₆₀₀	absorbância em 600 nm
DTT	ditiotreitól
EDTA	ácido etilenodiaminotetracético
GUS	β -glicoronidase
HD	homeodomínio
IPTG	isopropil tio- β -D-galactosídeo
kb	kilobases
kDa	kilodaltons
NP-40	Nonideti P-40
o2	gene opacó2
O2	proteína Opaco2
pb	pares de base
PEG	polietilenoglicol
pfu	unidade formadora de placa de lise
PMSF	fluoreto de fenilmetanosulfonil
SDS	dodecil sulfato de sódio
SDS-PAGE	gel de poliacrilamida contendo SDS
TBP	<i>TATA binding protein</i>
Tris	tris (hidroximetil) aminometano
X-gluc	ácido 5-bromo, 4-cloro, 3-indol, β -D-glicurônico

Preâmbulo:

Recentemente, Watson (1993), um dos idealizadores do modelo de dupla hélice para o DNA, dividiu a história completa da genética, do início até onde ele vislumbra o futuro, em sete fases. A primeira foi chamada de Mendeliana, e compreende o período no qual se definiu que o gene constitui a unidade fundamental de hereditariedade. A segunda fase abrange o período em que se estabeleceu a identidade molecular do gene, ou seja, que o DNA é o gene. Na terceira, determinou-se a base molecular da expressão gênica, com o cunho do *Dogma Central* da genética: DNA → RNA → proteína.

A quarta fase compreende o problema de como a expressão gênica é regulada. Na quinta, tem início a utilização das técnicas de DNA recombinante no isolamento e manipulação de genes. Na sexta, o foco de entendimento deixa de ser o gene, isoladamente, e parte-se para a obtenção de toda a informação genética de um determinado organismo. Finalmente, antevê Watson, na sétima fase todo o desenvolvimento de um determinado organismo poderá ser previsto no instante de sua concepção.

É bastante interessante notar que qualquer uma das fases descritas depende daquela que a antecede. Entretanto, ao passo que as três primeiras fases estão concluídas, as demais ainda estão sendo construídas e, se muitas respostas já estão disponíveis, bastante falta ainda para que se possa dominar os problemas abraçados pelas quatro últimas fases.

Sem um entendimento profundo de como se dá a regulação da expressão gênica nos mais diversos organismos, a manipulação de genes far-se-á de maneira absolutamente empírica, a seqüência completa do DNA de um determinado organismo ainda estará muito aquém de fornecer toda a informação genética a seu respeito e a sétima fase será apenas um sonho irrealizável.

Introdução

Os fenômenos celulares responsáveis pela manutenção da vida são regulados pela atividade de proteínas. Calcula-se que existam no genoma humano cerca de 150.000 genes, responsáveis pela codificação de um número semelhante de proteínas. A importância da regulação do nível de expressão de uma determinada proteína fica evidenciada pelo fato de asma, câncer, doenças cardíacas e imunes e infecções virais serem o resultado da super ou subprodução de uma ou algumas poucas proteínas (Tjian, 1995).

Desde muito cedo o melhoramento de cereais utiliza mutantes que apresentam efeitos pleiotrópicos indesejáveis, como é o caso do fenótipo *opaco2* em milho. Entretanto, a maior parte da pesquisa com estes mutantes esteve voltada para a produtividade e qualidade nutricional das sementes e não para questões envolvendo a estrutura física do endosperma, seu desenvolvimento e a base genética deste desenvolvimento (Lopes e Larkins, 1993).

O fenótipo *opaco2*, sabe-se hoje, é causado pela alteração de uma única proteína, um fator de regulação da transcrição, responsável pelo controle do nível de expressão de vários genes específicos de endosperma.

Qualquer abordagem não empírica, tanto para o melhoramento de cereais, quanto para a cura de moléstias que vão do câncer às infecções virais, depende do conhecimento minucioso dos mecanismos por trás das alterações nos níveis de expressão de proteínas.

A utilidade do estudo de genes de reserva de cereais

Entender regulação da expressão gênica de eucariotos

Os estudos da transcrição pela RNA polimerase II têm solidificado o conceito da conservação evolutiva do maquinário de transcrição basal e de

ativação entre eucariotos, desde leveduras ao homem, passando pelas plantas (Maldonado e Reinberg, 1995).

A caracterização de cereais mutantes no desenvolvimento do endosperma começa a mostrar que complexos mecanismos genéticos estão por trás da incrível constância vista em seu desenvolvimento (Goodwin, 1985; Lopes e Larkins, 1993).

Desta forma, a elucidação dos mecanismos envolvidos na regulação da expressão gênica de um modelo eucarioto, como no caso dos genes de proteínas de reserva de cereais, permite estabelecer novos paradigmas ou aprimorar os já existentes.

Entender as particularidades da regulação em plantas

Ainda que os genes de planta se pareçam bastante com genes animais ou de leveduras, existem algumas diferenças sutis entre eles (Kuhlemeier, 1992). Alguns promotores de genes de plantas funcionam em sistemas transcricionais de animais *in vitro*, como mostrado para o promotor do CaMV (vírus do mosaico da couve-flor) dirigindo a expressão de genes em células HeLa (Burke *et al.*, 1990).

Até mesmo em sistemas *in vivo* já é bem documentado o correto funcionamento de promotores de genes de plantas em sistemas heterólogos (i.e., aqueles onde o promotor regulando a expressão do gene não tem origem no organismo no qual se encontra), como no caso de genes de zeína funcionando corretamente em levedura (Langridge *et al.*, 1984).

Muito mais comum, contudo, é que experimentos com sistemas heterólogos simplesmente não funcionem. Os motivos destas discrepâncias não são ainda conhecidos.

Então, mais do que lançar novas luzes nos mecanismos gerais da regulação da expressão gênica em eucariotos, dados obtidos com o estudo dos genes de proteínas de reserva fornecerão subsídio para elucidação das disparidades observadas na regulação dos genes de origem animal e vegetal.

Gerar ferramentas para engenharia genética

As proteínas de reserva de sementes de cereais são, sem exceção, deficientes em determinados aminoácidos essenciais não sendo, desta forma, uma fonte nutricional adequada para seres humanos e animais monogástricos. As coixinas e zeínas, proteínas de reserva de *Coix* e milho respectivamente, apresentam níveis muito baixos de lisina e triptofano (Ottoboni *et al.*, 1990b).

Cereais vêm sendo selecionados, de maneira consciente ou não, desde que começaram a ser cultivados, há cerca de 10.000 anos. No último século, a aplicação de princípios científicos levou a um enorme melhoramento em produtividade, performance agrônômica e qualidade.

O melhoramento clássico, entretanto, é limitado pela variabilidade genética da cultura em questão, ou de espécies próximas. Além disso, o desenvolvimento de cultivares de milho com alta qualidade nutricional é tradicionalmente baseado no emprego de mutações que têm efeitos pleotrópicos indesejáveis.

Três estratégias são atualmente empregadas no melhoramento de cereais. A mais tradicional baseia-se no cruzamento de genótipos contrastantes (eventualmente de espécies distintas próximas) leva a novas combinações de genes e posterior seleção. A segunda estratégia lança mão da biotecnologia, como a cultura de tecidos e a regeneração de plantas, para tornar o processo mais eficiente, ou permitir a incorporação de genes de espécies que não podem ser cruzadas por métodos tradicionais. A terceira é a engenharia genética.

As deficiências nutricionais dos cereais fizeram com que a engenharia genética de proteínas de reserva despertasse grande interesse (Lawrence *et al.*, 1994). Depois de uma grande euforia, mitigada por resultados iniciais pouco satisfatórios, pesquisadores desta área são unânimes quanto a necessidade de se aprofundar nos conhecimentos básicos relacionados a expressão gênica.

A transferência direta de genes e sua conseqüente integração, por exemplo, é hoje um método sob o qual não se tem quase controle. Resultados previsíveis e estáveis no processo de produção de plantas transgênicas necessitam uma melhor compreensão de seus mecanismos (Lazzeri e Shewry, 1993). Um dos grandes problemas com o qual a engenharia genética se depara é conseguir altos níveis de expressão de um gene em um determinado tecido.

Diversos exemplos recentes de promotores tecido-específicos ilustram sua utilidade. Krebbers *et al.* (1993) produziram peptídeos biologicamente ativos em sementes de colza (*Brassica sp.*). Os próprios autores admitem que o uso de um promotor mais forte poderia gerar níveis de expressão até cinco vezes mais altos do que os obtidos. Percebe-se ainda que cereais seriam melhores veículos para produção de tais peptídeos, já que apresentam maior produtividade e facilidade de manejo e processamento (Lazzeri e Shewry, 1993).

Já Stark *et al.* (1992) introduziram em batatas um gene mutante de *E.coli* (o *glgC16*, que codifica a ADP-glicose pirofosforilase, responsável por um passo limitante na via de síntese do amido), sob controle de um promotor específico de raiz. Este gene de *E.coli* é menos sensível à inibição alostérica e menos dependente de ativação que o gene de batata. As batatas provenientes de plantas transgênicas apresentam um aumento de 35% a 60% no conteúdo de amido. Estes dois exemplos ilustram muito claramente a exeqüibilidade da utilização de engenharia genética no melhoramento de plantas. Deixam bem claro também a necessidade de promotores adequados para cada caso particular.

Uma série de trabalhos vem demonstrando a possibilidade de se empregar plantas como biorreatores na produção de matérias-primas para detergentes, plásticos biodegradáveis, lubrificantes industriais, e até mesmo vacinas (Moffat, 1995). Trabalhos como os de Töpfer *et al.* (1995) levam a crer que os primeiros produtos oriundos da manipulação de plantas por engenharia genética produzidos em larga escala possivelmente serão óleos industriais.

Vacinas produzidas em plantas contra doenças que vão de diarreia até AIDS são também cada vez mais prováveis (Moffat, 1995). Haq *et al.* (1995), por exemplo, conseguiram imunizar ativamente camundongos alimentando-os com plantas que expressavam uma subunidade de enterotoxina termolábil de *E.coli* (LT-B). Além de vacinas, plantas surgem como possíveis fontes de anti-soro. Em um trabalho bastante recente, Ma *et al.* (1995) produziram em tabaco anticorpos secretórios contra antígenos de superfície de estreptococos causadores de cárie.

É importante notar, entretanto, que não é sempre que altos níveis de transcrição serão suficientes para se obter a ação esperada de um determinado gene. A expressão gênica envolve mudanças na estrutura da cromatina, ativação gênica e transcrição. Mas envolve também processamento de mRNA, tradução, modificação pós-traducional, dobramento (*foldíng*) correto dos novos polipeptídeos e seu deslocamento para o sítio de ação no grão. A eficiência de cada um destes processos irá influenciar o nível de expressão final do gene. Por isso o processo deve ser analisado não apenas em termos moleculares, mas em um contexto celular. A elucidação da estrutura das proteínas de reserva, por exemplo, surge como pré-requisito fundamental para o desenho de mutantes estruturalmente funcionais (Utsumi, 1992).

Muitos genes de importância econômica potencial para o melhoramento de culturas são expressos de maneira específica nas sementes. As gramíneas armazenam proteínas e amido no endosperma. O estudo dos promotores que controlam a expressão destes genes, além de fornecer valiosas informações dos mecanismos de regulação gênica de eucariotos de um modo geral, e das particularidades de plantas, revelaria elementos que poderiam ser empregados para gerar, nos tecidos de reserva de sementes de cereal, níveis altos de expressão de seqüências codantes interessantes (Blechl *et al.*, 1994).

Objetivos

Os genes que codificam as prolaminas de cereais como milho, sorgo e *Coix* são altamente conservados e provavelmente tiveram uma origem comum, anterior à separação das espécies. O mecanismo de controle de expressão dos genes de prolaminas parece ser conservado nestas plantas, e apresenta componentes classe-específicos, i.e., cada uma das classes de prolaminas apresenta um nível próprio de regulação, além dos tecido- e tempo-específicos comuns. Isto já foi claramente demonstrado nos estudos recentes da proteína *Opaco2*, um ativador transcricional de α - e β -prolaminas (Schmidt *et al.*, 1992; Ueda *et al.*, 1992; Ottoboni *et al.*, 1993; Ueda *et al.*, 1994; Vettore, 1994; Cord-Neto *et al.*, 1995).

Muito pouco se sabe, entretanto, sobre os mecanismos que estariam envolvidos no controle específico das γ -prolaminas (Ueda *et al.*, 1994). A comparação de seqüências regulatórias destes genes pode levar a um melhor delineamento dos elementos *cis* envolvidos neste controle.

Os objetivos deste trabalho são o isolamento e seqüenciamento do gene da γ -prolamina de *Coix* (γ -coixina), a comparação de sua região promotora com seqüências já descritas de milho e sorgo e a caracterização desta região, delimitando porções envolvidas na expressão do gene.

Para tanto, as seguintes etapas foram executadas:

- a) isolamento do clone genômico da γ -coixina a partir de uma biblioteca genômica;
- b) seqüenciamento do clone;
- c) comparação da seqüência obtida com seqüências já descritas;

- d) experimentos de expressão transitória com extensões diversas do promotor dirigindo a síntese de um gene indicador;
- e) análise da interação de fatores nucleares com segmentos delimitados da região promotora através de experimentos de retardamento em gel.

Revisão bibliográfica

Cereais:

Há cerca de 10.000 anos o homem começou a desenvolver a agricultura. O processo de domesticação de plantas praticado desde então, além de fixar o homem primitivo à terra, fornecendo a base de nossa organização social, levou ao surgimento de várias espécies de cereais.

Por sua facilidade de cultivo e alto valor nutricional, aliados ao fato de poderem ser armazenados por longos períodos, as gramíneas selvagens desempenhavam papel fundamental na nutrição dos povos no início do desenvolvimento da agricultura.

Entre os cereais, alguns tiveram importância especial. Tal é o caso do milho, base da alimentação dos incas, maias e astecas, do trigo entre os romanos e egípcios e trigo e arroz entre os asiáticos (Small, 1984)

Atualmente, cereais são as culturas mais importantes mundialmente tanto no tocante à área cultivada quanto produção. Em 1991, por exemplo, mais de 700 milhões de hectares no mundo todo foram cultivados com cereais, o que levou à produção de 1.9 bilhões de toneladas. Milho contribuiu com 25% desta produção (Lazzeri e Shewry, 1993).

Sementes de plantas leguminosas chegam a ter 40% de seu peso seco formado por proteínas. Ainda que com teor menor de proteínas, entre 5% e 20%, grãos de cereais constituem a principal fonte de carboidratos e proteínas na nutrição humana e de rebanhos. Em países subdesenvolvidos, cereais são responsáveis pelo suprimento de cerca de 70% das necessidades protéicas humanas. São também empregados largamente na produção de bebidas alcoólicas (Tosello, 1987; Fincher, 1989; Lazzeri e Shewry, 1993).

Coix:

Origem

Coix lacryma-jobi é uma gramínea pertencente à tribo Andropogoneae, da qual fazem parte também o milho e o sorgo (Targon *et al.*, 1992). É originário do sudeste asiático, onde foi largamente cultivado até que a cultura do milho se tornasse popular (Arora, 1977).

Descrições da planta são encontradas desde os tempos védicos, cerca de 1500 a.C., na Ásia. Em algum momento entre os séculos XIII d.C. e XVII d.C. ela foi introduzida na Europa pelos árabes (Jain e Banerjee, 1974). Seu nome vem do árabe *Damu Ayub*, que significa "lágrimas de Jó" numa clara referência ao formato de suas sementes.

Os primeiros registros escritos da comestibilidade das sementes de *Coix* datam do início do século XVIII (Miller, 1731). Desde meados do nosso século *Coix* pode ser encontrado em todas as regiões tropicais (Shaafhawsen, 1952).

A variedade *Adlay* é uma das mais comuns na espécie *Coix lacryma-jobi*. Sua chegada ao Brasil, onde é popularmente conhecida como "lágrima de Nossa Senhora", se deu por volta de 1938, com sementes trazidas das Filipinas. Desde então esta variedade tem sido submetida a processos de melhoramento visando aumento de produtividade e melhoria na qualidade (Torres e Bergamin, 1951; Shaafhawsen, 1952; Jain e Banerjee, 1974).

O valor nutricional de *Coix* vem sendo objeto de estudos bastante recentes (Hwang *et al.*, 1993), onde se avalia sua utilização como forragem ou mesmo para o consumo humano. Novas variedades têm sido desenvolvidas e lançadas no mercado, como a "Yulmoo1" (Lee *et al.*, 1994), que se caracteriza por ter a casca da semente mais fina e maior produtividade.

Seu valor medicinal é conhecido desde o século XIII, quando era empregada como tônico e diurético (Jain e Banerjee, 1974). Na China, *Coix lacryma-jobi* é tradicionalmente empregado no tratamento de cânceres bronco-

pulmonares, sendo indexado oficialmente na Pharmacopeia Chinesa (Pinkas *et al.*, 1994). Substâncias antitumorais (Numata *et al.*, 1994) e inseticidas (Ary *et al.*, 1989) presentes em suas sementes e substâncias antimicrobianas (Ishiguro *et al.*, 1993) de plântulas estioladas foram recentemente isoladas e descritas.

Descrição

O gênero *Coix* é formado por plantas monóicas que podem atingir até 1,60 m de altura. Algumas são anuais, como *Coix lacryma-jobi* var. *Adlay*, outras perenes. Apresentam polinização anemófila com cerca de 40% de fecundação cruzada.

As sementes são formadas nos ramos apicais da planta, no final do pedúnculo da bainha das folhas. Cada ramo contém de 6 a 9 flores (**Figura 2A**). As espiguetas femininas são tricarpeladas, com apenas um carpelo fértil. O involúcro, uma bainha endurecida modificada semelhante a uma conta, envolve cada espiguetas. A inflorescência masculina emerge através do orifício apical do involúcro, sendo sustentada por uma haste (**Figura 2B**). As espiguetas masculinas são imbricadas em pares ou trios, com uma delas pedunculada. Cada espiguetas é formada por duas floretas estaminadas. A floreta superior pode ser estéril (Shaafhawsen, 1952; Venkateswarlu e Chaganti, 1973; Jain e Banerjee, 1974).

A antese varia imensamente entre os ramos e mesmo entre flores do mesmo ramo. O involúcro da semente praticamente não tem seu formato e tamanho alterados durante o desenvolvimento, mas sua cor sim. Desta forma, a cor é empregada como indicador do desenvolvimento da semente, que dura em torno de 35 dias, arbitrariamente dividido em cinco estágios (Targon *et al.*, 1992).

As proteínas de reserva começam a se acumular no estágio 1 e atingem níveis máximos entre os estágios 4 e 5. Proteínas constituem 20% do peso seco da semente madura de *Coix lacryma-jobi* var. *Adlay* (Ottoboni *et al.*, 1990b). Esta é uma das maiores proporções encontradas em cereais, duas vezes maior do que a observada em sementes de milho.

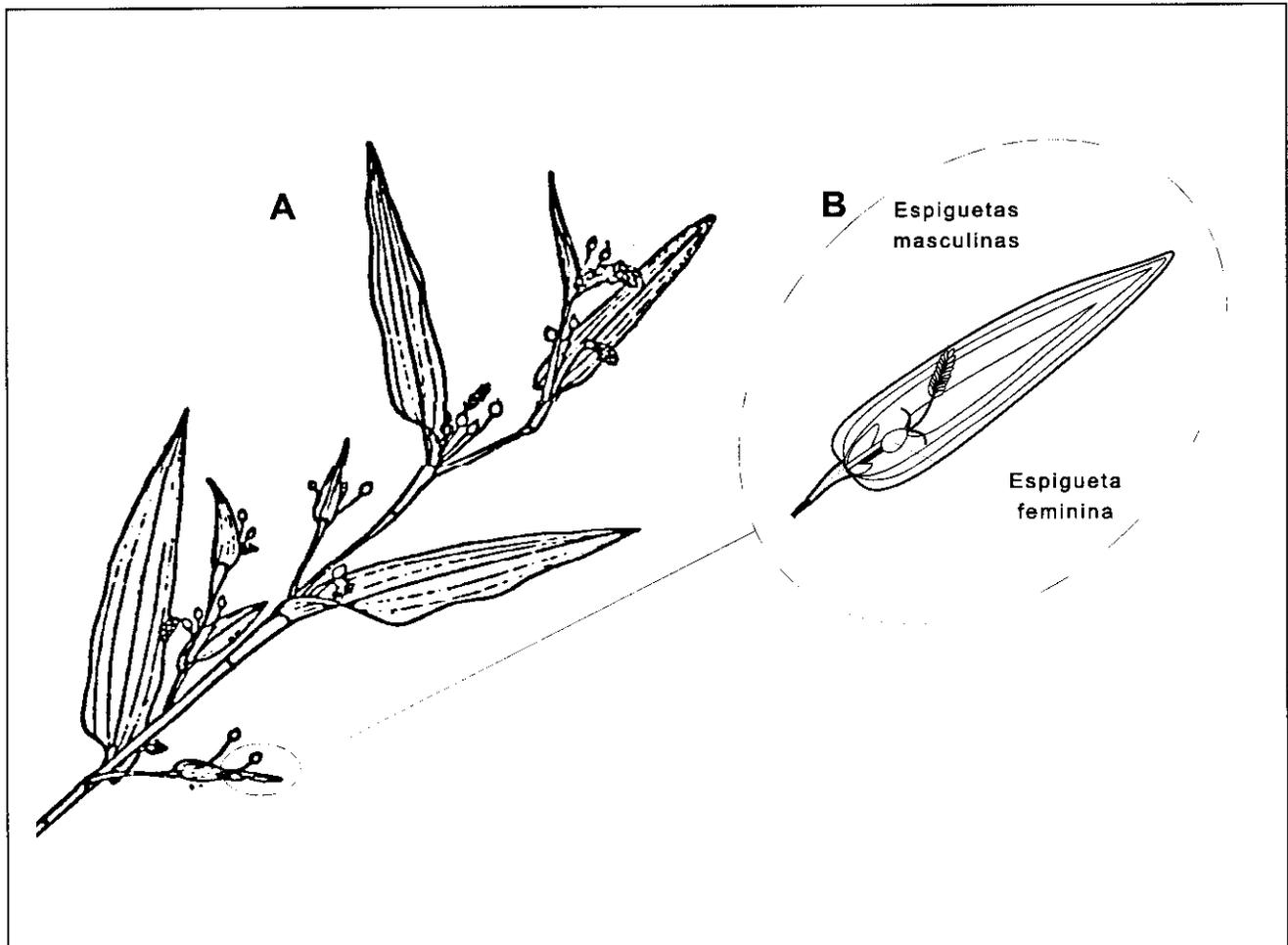


Figura 1: Esquema de um ramo de *Coix lacryma-jobi*.

A) Esquema do ramo de *Coix*, mostrando a posição de formação das sementes, segundo Shaafhawsen (1952). **B)** No detalhe são mostradas as espiguetas masculina e feminina de uma inflorescência.

A semente

Conquanto folhas e raízes possam ser empregadas como silagem e forragem, a parte agronomicamente mais importante dos cereais é a semente, formada pelo endosperma e pelo embrião.

O endosperma se origina da fusão de um dos dois núcleos espermáticos do grão de pólen germinante com os dois núcleos polares da oosfera. É, portanto, um tecido triplóide. A fusão dos núcleos se dá no saco embrionário, que é cercado de tecido materno.

A semente é, então, um tecido geneticamente bastante interessante. O embrião, originado da fecundação da célula-ovo materna pelo outro núcleo espermático do grão de pólen, é um tecido diplóide. O endosperma, que irá fornecer nutrientes para este embrião, como já foi dito no parágrafo anterior, é triplóide (**Figura 1**). Os dois núcleos espermáticos, o responsável pela fecundação da célula-ovo e o responsável pela fecundação da oosfera, são geneticamente idênticos, pois se originam da mitose de uma célula haplóide. O mesmo acontece com os três núcleos maternos envolvidos. Entretanto os dois produtos de fusão, o endosperma e o embrião, desenvolvem-se de maneira completamente diferentes.

Desde o início do século as possíveis origens da dupla fecundação são discutidas (Sargant, 1900). Sua importância pode estar na manutenção de diferenças epigenéticas que regulam o desenvolvimento do endosperma e do embrião (Kermicle e Alleman, 1990). Um tecido triplóide pode permitir quatro níveis efetivos de dosagem gênica por *imprinting* gamético, ao passo que tecidos diplóides só permitem três níveis efetivos. Experimentos com translocações cromossômicas demonstraram a atividade de *imprinting* paternal de genes que regulam o desenvolvimento do endosperma (Lin, 1982). Mutações que levam à formação de endospermas com ploidias variadas mostraram que a manutenção da razão de um genoma paternal para dois maternos é fundamental para o completo desenvolvimento do endosperma de milho (Lin, 1984).

O Endosperma:

O endosperma constitui a maior parte da semente de cereais. Nele estão localizadas as reservas de nutrientes, entre elas as proteínas, mobilizadas para a nutrição da plântula durante a germinação (**Figura 1**).

Este tecido foi estudado sob uma série de enfoques: evolução, papel no desenvolvimento da semente e sua germinação, genética, fisiologia e bioquímica. Ele constitui uma fonte renovável e biodegradável de materiais. Há um grande empenho voltado para melhorar seu emprego como alimento e fonte de produtos secundários, tais como plásticos e óleos (Lopes e Larkins, 1993).

Aleurona

Durante o desenvolvimento da semente, a parte mais externa do endosperma diferencia-se em aleurona, um tecido formado por células morfológicas e funcionalmente distintas (**Figura 1**). Em milho, sorgo e trigo a aleurona é formada por uma única camada de células. Em outros cereais, como cevada e aveia, ela apresenta várias camadas de células. No endosperma, somente as células da aleurona continuam vivas após a maturação do grão (Fincher e Stone, 1986).

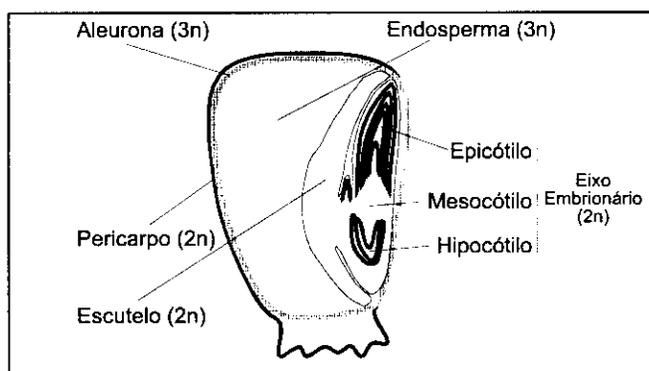


Figura 2: Corte esquemático da semente de milho.

Estão indicadas na figura o nome das estruturas e seu grau de ploidia.

As paredes celulares das células do endosperma maduro são bem menos espessas do que as das células da aleurona. Durante a germinação, a

atividade celular da aleurona é rapidamente voltada, em resposta ao ácido giberélico produzido pelo embrião, para a síntese e secreção de uma bateria de enzimas hidrolíticas. Tais enzimas catalisam a hidrólise das macromoléculas de reserva localizadas nas células do endosperma: amido, proteínas, polissacarídeos da parede celular e DNA (Fincher e Stone, 1986; Rastogi e Oaks, 1986; Fincher, 1989).

Aumento do DNA

Durante o desenvolvimento do endosperma, é comum um aumento considerável do conteúdo de DNA nuclear. Ele se dá através de duplicações cromossômicas endonucleares sem mitose. Este processo é mantido pelo aumento na quantidade e atividade das quinases relacionadas com a fase S do ciclo celular, e pela inativação do fator ligado a fase M, o MPF (M phase-promoting factor) (Grafi e Larkins, 1995).

Em milho, entre 10 e 20 dias após a polinização, o conteúdo de DNA passa de cerca de 3 vezes o genoma haplóide para até mais de 600 vezes o genoma haplóide. Os genes que codificam proteínas e enzimas envolvidas na síntese de carboidratos e proteínas de reserva não se encontram especialmente amplificados, *i.e.*, o genoma todo é copiado em ciclos repetidos de replicação (Lopes e Larkins, 1993). Assim, genes que não são expressos no endosperma também são amplificados.

Este não parece, então, ser um mecanismo para aumentar especificamente a atividade dos genes expressos no endosperma, mas talvez tenha evoluído para cumprir funções adicionais, como o armazenamento de nucleotídeos para futuro uso da plântula em desenvolvimento. A presença de quantidades consideráveis de DNA em células metabolicamente inativas do endosperma de sementes maduras corrobora esta hipótese (McFadden *et al.*, 1988). Trabalhos recentes (Diffley, 1994), ainda incipientes, associando mecanismos de replicação com regulação transcricional, entretanto, não permitem excluir a possibilidade da amplificação estar relacionada à ativação de determinados genes.

O Embrião

Todo o desenvolvimento do endosperma é voltado para a nutrição do embrião no instante da germinação.

O desenvolvimento do embrião de *Coix* é classificado em 9 estágios: pró-embrião, forma de trevo, coleoptilar, monofoliado, bifoliado, trifoliado, tetrafoliado, pentafoliado e hexafoliado, este último sendo considerado o embrião maduro. Os embriões tri, tetra e pentafoliados apresentam 1, 2 e 3 raízes adventícias (seminais), respectivamente. O embrião maduro, assim como o pentafoliado, apresenta 3 raízes seminais (Xi e Ye, 1994).

Na interface entre o embrião e o endosperma encontra-se uma camada única de células especializadas alongadas denominada epitélio do escutelo (**Figura 1**). No que tange à mobilização dos recursos do endosperma, o epitélio é, provavelmente, o componente mais importante da semente. Ele participa tanto da secreção, juntamente com a aleurona, de enzimas digestivas no endosperma, em um primeiro instante, quanto da absorção dos produtos resultantes da degradação (Fincher, 1989). Posteriormente a função de secreção de enzimas parece ser relegada apenas às células da aleurona (McFadden *et al.*, 1988).

A camada de aleurona e o escutelo dos cereais são dois dos poucos tecidos digestivos encontrados no reino vegetal. Como são fundamentalmente autotróficas, as plantas muito raramente desenvolveram tais tecidos. A aleurona e o escutelo são, juntamente com as folhas modificadas de plantas carnívoras, exceções notáveis (Jones e Jacobsen, 1991).

O primeiro depósito dos recursos disponibilizados no endosperma de *Coix* ocorre nas células do escutelo. Nove dias após a antese, no estágio monofoliado, os grânulos de amido se acumulam nas células do escutelo, coleóptilo e mesocótilo. Com a maturação do embrião, os grânulos de amido se depositam em todas as suas células. O aumento na quantidade e tamanho dos grânulos de amido está correlacionado com o surgimento e organização

crescente dos órgãos embrionários, com a quantidade presente nas células do escutelo diminuindo das fases mais tardias até a maturação do embrião.

Dez dias após a antese, no estágio bifoliado, corpúsculos globóides, contendo cristais de proteína e fitina, e corpúsculos lipídicos são encontrados nas células do escutelo. Os corpúsculos lipídicos aumentam em tamanho e número rapidamente ao longo do desenvolvimento do embrião.

Os corpúsculos protéicos logo se transformam em drusas grandes e abundantes, com o surgimento concomitante de corpúsculos protéicos sem cristais. Posteriormente, os corpúsculos protéicos contendo cristais desaparecem, ao passo que aqueles sem cristais seguem crescendo até a maturação do embrião. No estágio trifoliado, 13 dias após a antese, corpúsculos protéicos são formados nas células superiores do coleóptilo. Quanto mais precoce o início do desenvolvimento de um órgão no embrião, maior o número e tamanho dos corpúsculos protéicos contidos neste órgão (Xi e Ye, 1994).

A maturação é um passo essencial no desenvolvimento da semente. Nela ocorre queda da síntese protéica, dessecação e, finalmente, dormência. Nos cereais a maturação ocorre tardiamente. Os níveis de ácido abscísico se elevam, evitando a germinação do embrião (Lopes e Larkins, 1993).

Proteínas de reserva

Em 1746, Beccari isolou pela primeira vez proteínas de origem vegetal. Até então imaginava-se que as proteínas eram características da vida animal apenas (Osborne, 1908).

A fração protéica da semente contém uma enorme variedade de proteínas, presentes individualmente em pequenas quantidades, essenciais para a manutenção do metabolismo celular. Entre as proteínas desta categoria estão incluídas proteínas metabólicas e estruturais, como enzimas e componentes de parede e membrana celulares (Lazzeri e Shewry, 1993). Uma outra categoria é composta por um número pequeno de proteínas, conhecidas como proteínas de reserva, armazenadas em grandes quantidades. Estas

proteínas possuem a função de fornecer nitrogênio, enxofre, carbono e aminoácidos à plântula em germinação, sendo hidrolisadas com rapidez (Higgins, 1984). Hoje sabe-se que as proteínas isoladas por Beccari em 1746 eram proteínas de reserva de trigo (Osborne, 1908; Lazzeri e Shewry, 1993).

Landry e Moureaux (1980) propuseram uma classificação para as proteínas de sementes baseada na divisão proposta por Osborne e Mendel (1914). Estes autores agrupam-nas em quatro classes distintas de acordo com sua solubilidade em extração seqüencial:

- 1) albuminas: solúveis em água;
- 2) globulinas: solúveis em solução salina;
- 3) prolaminas: solúveis em solução alcoólica;
- 4) glutelinas: solúveis em soluções alcalinas e ácidas diluídas, e em soluções salinas ou alcoólicas que contenham um agente redutor.

Certas proteínas de semente parecem ter como única função, a de constituir reservas de nitrogênio, carbono e enxofre durante o desenvolvimento da semente, fornecendo estes nutrientes ao embrião durante a germinação. Essas proteínas com função de armazenamento são encontradas em grandes quantidades, podendo ocupar até 15% do peso seco da semente em cereais, e 25% a 35% em leguminosas. São formadas por uma mistura de polipeptídeos geralmente codificadas por uma grande família gênica. São armazenadas em densas organelas membranosas, os corpúsculos protéicos. Em todos os cereais, exceto aveia e arroz, a principal fração das proteínas de reserva são prolaminas (Leite *et al.*, 1990; Schmidt *et al.*, 1992; Lazzeri e Shewry, 1993; Shewry, 1995).

O termo prolamina foi criado por Osborne (1908). Osborne se referia à grande proporção do aminoácido prolina encontrado por ele neste grupo de proteínas. Hoje o nome é empregado para designar proteínas ricas nos aminoácidos prolina e glutamina e/ou asparagina. Shewry e Tatham (1990) definem as prolaminas como "proteínas de reserva depositadas nos

endosperma de cereais em desenvolvimento, ricas em glutamina e prolina, insolúveis, em seu estado nativo, em água ou soluções salinas diluídas".

As prolaminas recebem nomes específicos que indicam a planta de origem. Em milho, são denominadas zeínas, em *Coix*, coixinas e, em sorgo, kafirinas. A denominação de kafirina deve-se à variedade kafir cultivada no sul da África (Harlan, 1992). As coixinas, prolaminas de *Coix*, correspondem a mais de 70% do total de proteínas encontrado em suas sementes (Ottoboni *et al.*, 1990b).

Uso pela planta

A mobilização das proteínas de reserva de cereais é mediada pela ação de endo- e exopeptidases (Rastogi e Oaks, 1986), com claro predomínio em arroz, sorgo e milho da atividade das endopeptidases (Winspear *et al.*, 1984). Agindo em conjunto, estas peptidases secretadas no endosperma, ou já existentes lá, realizam a hidrólise das proteínas de reserva gerando uma mistura de aminoácidos e pequenos peptídeos. As carboxipeptidases ácidas são as exopeptidases mais abundantes em cevada (Fincher, 1989). Estas carboxipeptidases têm ação muito rápida sobre grandes peptídeos, mas atuam lentamente em di- ou tri-peptídeos. Estes pequenos peptídeos, entretanto, podem ser rápida e ativamente transportados para o escutelo (Rastogi e Oaks, 1986; Payne e Walker-Smith, 1987), onde são transformados em aminoácidos livres para serem transportados à plântula ou utilizados nas reações metabólicas do escutelo (Fincher, 1989).

Muitas destas peptidases poderiam gerar problemas para as células que as sintetizam, dado seu potencial de degradar inespecificamente proteínas celulares. Alguns trabalhos (Doi *et al.*, 1980; Doan e Fincher, 1988) sugerem que estas peptidases sejam sintetizadas em forma de zimógenos.

Na célula

O acúmulo das prolaminas começa logo após a queda do índice mitótico do endosperma, quando suas células iniciam o processo de

amplificação gênica. As proteínas de reserva são sintetizadas por polirribossomos ligados à membrana do retículo endoplasmático e transportadas para seu lúmen, que rapidamente se fissiona para formar os corpúsculos protéicos (Feix e Quayle, 1993).

O acúmulo de proteínas de reserva em corpúsculos protéicos é provavelmente uma adaptação que evita que elas fiquem expostas às enzimas responsáveis pelo *turnover* de proteínas metabólicas. Outra vantagem potencial em separá-las em organelas seria tornar mais fácil seu empacotamento, reunindo-as em um ambiente composto de proteínas com propriedades bioquímicas semelhantes que apresentam capacidade de interagir entre si. Além disso, sua deposição ocorre em um ambiente relativamente desprovido de água, o que facilita a secagem da semente (Lopes e Larkins, 1993).

Propriedades da estrutura primária (seqüência) das proteínas de reserva parecem dirigi-las a organelas específicas ou localizações determinadas no corpúsculo protéico. As γ -prolaminas de trigo, por exemplo, contém informação suficiente para iniciar a formação de corpúsculos protéicos mesmo em sistemas heterólogos (Altschuler *et al.*, 1993).

Especificidade

O acúmulo de coxina segue uma cinética sigmoideal relacionada com o ganho de peso seco do endosperma. A coxina e seus mRNAs começam a se acumular nos estágios iniciais de desenvolvimento do endosperma. Há um claro acúmulo de mRNA polissomal nos estágios de maior taxa de síntese de coxina e, conforme as coxinas são sintetizadas e armazenadas, nos estágios mais tardios, há uma queda evidente no nível de mRNA polissomal, sendo impossível sua detecção nas sementes maduras (Targon *et al.*, 1992). Este padrão é observado também em milho (Feix e Quayle, 1993). Tais fatos, aliados ao fato de as coxinas e as zeínas serem encontradas apenas no endosperma, evidenciam sua especificidade temporal e espacial.

Regulação de transcrição

A quantidade e tipo de mRNAs específicos de zeínas que podem ser isolados nos diferentes estágios de desenvolvimento do endosperma (Feix e Quayle, 1993), aliados à coincidência temporal observada entre a síntese dos RNAs e proteínas de zeínas (Kodrzycki *et al.*, 1989) indicam a ativação transcricional específica de seus genes. Há também uma forte correlação entre o perfil de acúmulo das proteínas de reserva e o nível de RNAs polissomais nas sementes de endosperma de *Coix* (Targon *et al.*, 1992). Estas evidências sugerem que o controle da expressão dos genes das prolaminas se dê basicamente ao nível da transcrição.

Nas gramíneas

Prolaminas do milho

As prolaminas do milho (*Zea mays*), **zeínas**, são subdivididas em seis classes de peso molecular aparente de 10, 14, 16, 19, 22 e 28 kDa, quando analisadas em SDS-PAGE. A composição de aminoácidos das diversas classes de zeínas é similar, no sentido de que todas são ricas em glutamina, leucina, alanina e prolina. As zeínas de 28, 16, 14 e 10 kDa são caracterizadas pelo alto teor de metionina e cisteína (Gianazza *et al.*, 1976).

Esen (1986) estabeleceu as condições para a divisão das prolaminas em função de sua solubilidade em soluções aquosas. Trabalhando com zeínas, prolaminas de milho, ele definiu três classes de solubilidade, posteriormente chamadas de α -, β - e γ -zeínas. Análises estruturais dos polipeptídeos que constituem as β -zeínas levaram à criação de uma quarta classe, as δ -prolaminas (Kirihiro *et al.*, 1988).

Assim, os polipeptídeos de 19 e 22 kDa formam as α -zeínas, os de 14 kDa, as β -zeínas, os de 16 e 28 kDa, as γ -zeínas e os polipeptídeos de 10 kDa as δ -zeínas.

Prolaminas de Coix

A análise das **coixinas**, prolaminas de *Coix*, em SDS-PAGE mostra cinco bandas de pesos moleculares diferentes, denominadas C1 (27 kDa), C2 (25 kDa), C3 (22 kDa), C4 (17 kDa) e C5 (15 kDa) (Ottoboni *et al.*, 1990a) que, por extração diferencial (Esen, 1986), podem ser classificadas em dois grupos principais:

- a) as α -coixinas, solúveis em 40-95% de isopropanol, são constituídas por 4 classes de proteínas de diferentes pesos moleculares aparentes, C1, C2, C4 e C5;
- b) a γ -coixina, solúvel em 0-80% de isopropanol na presença de agentes redutores (2-ME, DTT), é constituída por uma única classe de proteína, C3, apresentando peso molecular aparente de 22 kDa.

A análise seqüencial através de isoeletrofocalização, SDS-PAGE e "Western blot", revelou que C1 é composta, no mínimo, por dois polipeptídeos, C2 por cinco polipeptídeos, C3 e C4 por apenas um polipeptídeo, enquanto que C5 é constituída por 2 polipeptídeos (Leite *et al.*, 1990). A fração que corresponderia, pelos critérios de Esen (1986), às β -coixinas é formada por uma mistura de α e γ -coixinas (Leite *et al.*, 1990). Posteriormente demonstrou-se que uma das α -coixinas (C4) é estruturalmente homóloga às β -zeínas e, a partir de então, esta prolamina passou a ser denominada β -coixina (Leite *et al.*, 1992).

Prolaminas de sorgo

As **kafirinas**, prolaminas de sorgo, assim como as de *Coix*, são divididas por solubilidade diferencial em α -, β - e γ -prolaminas (Shull e Kirleis, 1989; Barros *et al.*, 1991).

Quando separadas por SDS-PAGE, as kafirinas são resolvidas em sete bandas. As α -kafirinas são representadas pelos polipeptídeos de peso molecular aparente igual a 22, 21, 20 e 19 kDa, as β -kafirinas, pelos

polipeptídeos de 16 e 14 kDa (Shull e Kirleis, 1989) e a γ -kafirina por um polipeptídeo com peso molecular aparente de 27 kDa (Barros *et al.*, 1991).

Classificação

Uma série de diferenças pode ser observada entre os genes de prolaminas de cereais. As seqüências de proteínas de classes diferentes, por exemplo, não guardam semelhança entre si. Ainda assim, algumas características estruturais comuns são apresentadas por genes de todas as classes de prolaminas.

Entre as semelhanças destacam-se a ausência de íntrons, a presença de uma seqüência conservada na região promotora, conhecida como *prolamin box* e a organização dos genes em regiões de domínios estruturais que levam à formação de estruturas secundárias próprias para o empacotamento (Feix e Quayle, 1993).

α -prolaminas

São formadas pelos polipeptídeos solúveis em soluções aquosas contendo 40-95% de isopropanol (Esen, 1986). As α -prolaminas constituem a maior fração das proteínas de reserva de milho, compreendendo 80% da zeína total (Shewry e Tatham, 1990), e de *Coix*, onde as α -coixinas representam 85% da coixina total (Targon *et al.*, 1992).

Esta classe de proteínas é codificada por uma família multigênica. Em milho, as classes de 22 e 19 kDa são, cada uma, codificadas por mais de 50 genes localizados em *loci* determinados dos cromossomos 4, 7 e 10. Estima-se que metade destes genes de milho sejam ativos (Leite *et al.*, 1990; Feix e Quayle, 1993). A α -coixina de 25 kDa é codificada por cerca de 36 genes (Ottoboni *et al.*, 1993) e as α -prolaminas de sorgo, por 20 genes (DeRose *et al.*, 1989).

Inicialmente, a homologia existente entre as proteínas dessa classe foi demonstrada através de experimentos de SDS-PAGE e "Western blot",

ocorrendo reações cruzadas com anticorpos produzidos contra coixinas e zeínas e as proteínas de milho, *Coix* e sorgo (Shull e Kirleis, 1989; Leite *et al.*, 1990). Essa homologia foi posteriormente confirmada através de comparações de seqüências de aminoácidos deduzidas de clones genômicos de α -kafirina de 22 kDa (DeRose *et al.*, 1989), α -coixina de 25 kDa (Ottoboni *et al.*, 1993), α -zeínas de 19 kDa (Pedersen *et al.*, 1986) e 22 kDa (Wandelt e Feix, 1989), revelando seqüências muito conservadas.

As α -prolaminas são estruturalmente bastante semelhantes entre si. Esta classe de proteínas é caracterizada por domínios curtos nas extremidades, com 36 ou 37 resíduos na região amino-terminal e 10 resíduos na região carboxi-terminal, no caso das zeínas. Entre estes domínios encontram-se 9 ou 10 repetições de um bloco de 20 aminoácidos, sem seqüência consenso clara. Cada um destes blocos repetidos corresponde a uma α -hélice.

Baseado no alinhamento das repetições presentes nas α -prolaminas de milho, sorgo e *Coix*, e também no padrão de distribuição dos aminoácidos hidrofóbicos e hidrofílicos, Garrat *et al.* (1993) descreveram um modelo estrutural para esta classe de proteínas. Este modelo sugere que cada par de α -hélices esteja arranjado de maneira antiparalela e que cada hélice, da molécula como um todo, esteja arranjada em uma rede hexagonal. A maioria das α -hélices mostram seis faces de resíduos polares e hidrofóbicos alternados, o que permite um denso empacotamento dessas proteínas, característica importante para o armazenamento no interior do corpúsculo protéico.

A α -zeína de 19 kDa é menor que a de 22 kDa devido à perda de um motivo repetido. Dois motivos repetidos ancestrais provavelmente deram origem à parte central das moléculas de α -zeínas através de duas duplicações intragênicas. Um destes motivos ancestrais gerou as repetições pares e o outro as ímpares (Garratt *et al.*, 1993). Estes mesmos autores consideram provável que a α -zeína de 22 kDa tenha originado a de 19 kDa após a divergência entre milho e *Coix*.

A estrutura das α -prolaminas é muito distinta da apresentada pelas demais prolaminas. As β -, γ - e δ -prolaminas, assim como as albuminas 2S das dicotiledôneas mantém, em maior ou menor grau, semelhanças entre si. Desta forma, postula-se que a origem das prolaminas seja polifilética, i.e., as α -prolaminas e as demais (β -, γ - e δ -prolaminas) teriam origens distintas (Shewry e Tatham, 1990; Shewry, 1995).

A α -zeína de 22 kDa tem recebido muita atenção, pois além de constituir o principal componente da fração total de zeínas, esta tem sua síntese reduzida pela mutação *opaco-2* (*o2*) (Kodrzycki *et al.*, 1989). Esse mutante produz um endosperma com alto teor de lisina, porém farináceo e macio, o que prejudica o seu armazenamento e comercialização. O alto teor de lisina é devido à redução da síntese da α -zeína de 22 kDa, o que provoca uma queda de 50-70% no conteúdo de zeína. A diminuição no acúmulo das zeínas é acompanhada por um aumento do conteúdo das frações albumina, globulina e glutelina (Burr e Burr, 1982).

Após a clonagem e caracterização do gene *o2* (Schmidt *et al.*, 1987; Motto *et al.*, 1988), foi verificado que a proteína codificada por este gene pertence à classe de proteínas estruturalmente relacionadas denominadas zíper de leucinas. Dois experimentos independentes demonstraram que a proteína O2 reconhece a seqüência GATGAYRTG no promotor de uma albumina de 32 kDa, denominada B-32 (Lohmer *et al.*, 1991), e a seqüência TCCACGTAGA presente no promotor da α -zeína de 22 kDa (Schmidt *et al.*, 1992). Foi demonstrado também, que a proteína O2 de milho reconhece a seqüência GACATGTC, presente no promotor da α -coixina de 25 kDa (Yunes *et al.*, 1994).

Experimentos realizados em nosso laboratório, incluindo o isolamento de clones de cDNA e genômico de *Coix* apresentando grande similaridade com o gene *o2* de milho (Vettore, 1994), sugerem um mecanismo de regulação conservado entre milho e *Coix* após a divergência das espécies. Da mesma forma, a estrutura conservada das α -prolaminas corrobora esta hipótese.

As α -coixinas correspondem a 85% das coixinas e são formadas por um extenso grupo de proteínas agrupados em quatro classes de tamanho: 27 kDa, 25 kDa, 17 kDa e 15 kDa. Elas se caracterizam por serem ricas nos aminoácidos leucina e alanina (Shewry e Tatham, 1990).

As α -coixinas, assim como uma série de outras α -prolaminas, são sintetizadas como precursores, em média 2 kDa maiores do que as proteínas maduras (Targon *et al.*, 1992). O maior peso molecular se deve à presença do peptídeo sinal que é removido durante o transporte para regiões específicas do corpúsculo protéico (Feix e Quayle, 1993).

β -prolaminas

Pela definição de Esen (1986), as β -prolaminas são formadas pelo grupo de polipeptídeos solúveis em soluções contendo 30-80% de isopropanol. Nesta fração, em milho, encontra-se uma mistura de polipeptídeos de 14 e 16 kDa. O de 16 kDa, devido a suas características estruturais, é classificado como γ -prolamina. Ao contrário do que ocorre com as α -zeínas, a β -zeína é representada somente por uma ou duas cópias desse gene no genoma haplóide do milho (Kiriara *et al.*, 1988). Constitui entre 10 e 15% do total das zeínas (Shewry e Tatham, 1990).

A seqüência de aminoácidos deduzida de cDNAs de β -zeína (Pedersen *et al.*, 1986) não mostra qualquer similaridade com seqüências de α -zeínas. A β -zeína não contém seqüências de aminoácidos repetidas que possam caracterizar um domínio estrutural bem definido. Experimentos de dicroísmo circular e dispersão óptica rotatória, mostraram uma composição deficiente em α -hélices, tendo principalmente, estruturas β -pregueadas (33% a 55%, dependendo da metodologia empregada) (Shewry e Tatham, 1990).

A homologia entre as β -prolaminas foi demonstrada através de experimentos de SDS-PAGE e "Western blot", onde anticorpos contra β -zeína mostraram reação cruzada com as proteínas pertencentes à classe das β -kafirinas (Shull e Kirleis, 1989). A β -zeína encontra-se distribuída na periferia

do corpúsculo protéico, aparentemente desempenhando funções estruturais. Esta idéia é corroborada pelo alto conteúdo de cisteína presente nesta proteína, visto que este aminoácido está envolvido com a formação de pontes dissulfeto (Lending *et al.*, 1988). Além do alto conteúdo de cisteína, esta proteína é também rica em metionina (Shewry e Tatham, 1990).

Muito pouco é conhecido sobre a regulação da expressão das β -prolaminas. Porém, através de experimentos de retardamento em gel, foi identificado um fator protéico, presente somente em endosperma, que se liga ao promotor da β -zeína de 14 kDa. Uma análise posterior através de "footprinting" revelou que o sítio de reconhecimento dessa proteína (chamada Beta-1) está localizado próximo ao elemento TATA. O pico na produção da proteína Beta-1 ocorre em torno de 16 DAP e parece estar ligada à expressão tecido-específica do gene de β -zeína (So e Larkins, 1991). Mais recentemente (Cord-Neto *et al.*, 1995), foi demonstrado que a proteína regulatória O2, responsável pelo controle transcricional das α -prolaminas, está também envolvida na regulação das β -prolaminas.

Pelo menos um dos polipeptídeos da classe de 17 kDa de *Coix* (o codificado pelo clone de cDNA pBCX17.9S), classificado como α -coixina em função de sua solubilidade, é hoje agrupado como β -prolamina devido a similaridades estruturais com as zeínas desta classe (Leite *et al.*, 1992).

γ -prolaminas

As γ -prolaminas de cereais tropicais são ricas em enxofre e solúveis em soluções aquosas contendo 0-80% de isopropanol e agentes redutores (Esen, 1986). Possuem, assim como as β -prolaminas, um elevado teor de cisteína, motivo pelo qual esta classe de proteínas é solúvel somente na presença de agentes redutores. São francamente distintas das α -prolaminas, mas apresentam domínios similares a outras proteínas de reserva de cereais ricas em enxofre, como as β -prolaminas de cereais tropicais e as prolaminas de cereais de regiões temperadas, como as γ -gliadinas (prolaminas de trigo), as

prolaminas HMW de trigo e as albuminas 2S de girassol (Shewry e Tatham, 1990).

Em milho, as γ -prolaminas constituem 5% a 10% da fração total de zeínas e em sorgo, 2% a 5% do total das kafirinas. A γ -kafirina é codificada por um ou dois genes (Freitas *et al.*, 1994). A γ -zeína de 16 kDa é codificada por um único gene (Gallardo *et al.*, 1988), enquanto a γ -zeína de 28 kDa é codificada por apenas um gene em algumas linhagens e em outras, devido a uma duplicação oriunda de recombinação homóloga, por dois genes (Das *et al.*, 1991). A fração correspondente à γ -coixina contém apenas uma proteína de 22 kDa, codificada por um único gene, e corresponde a aproximadamente 15% da fração total de coixinas (Leite *et al.*, 1990).

estrutura

A homologia existente entre a γ -coixina e γ -zeína foi primeiramente demonstrada através de experimentos "Western blot" e "Southern blot" (Leite *et al.*, 1990). Posteriormente, os estudos estenderam-se às proteínas de sorgo, demonstrando-se a homologia entre as γ -prolaminas destas três espécies (Freitas *et al.*, 1994).

Após a clonagem e caracterização de clones genômico e de cDNA de γ -zeína de 28 kDa (Prat *et al.*, 1985; Wang e Esen, 1986; Reina *et al.*, 1990), clone genômico de γ -zeína de 16 kDa (Reina *et al.*, 1990a), clone genômico e de cDNA de γ -kafirina (Barros *et al.*, 1991; Freitas *et al.*, 1994), clone de cDNA (Leite *et al.*, 1991) e genômico (este trabalho) de γ -coixina, verificou-se que essas proteínas possuem na região amino-terminal um domínio formado por repetições do hexapeptídeo Pro-Pro-Pro-Val-His-Leu (Prat *et al.*, 1985), seguido de uma região Pro-Xaa , *i.e.* uma região com resíduos alternados de prolina. A γ -zeína apresenta oito repetições do hexapeptídeo, a γ -kafirina apresenta quatro repetições e três repetições são encontradas na γ -coixina (Freitas *et al.*, 1994). As repetições do hexapeptídeo estão envolvidas com a retenção da γ -zeína no retículo endoplasmático (Geli *et al.*, 1994).

O alinhamento das seqüências das γ -prolaminas de milho, *Coix* e sorgo mostra a conservação de dez resíduos de cisteínas, que devem desempenhar um importante papel na formação de pontes de dissulfeto inter e intra-cadeias (Lending *et al.*, 1988; Freitas *et al.*, 1994; Geli *et al.*, 1994).

Através de dispersão óptica rotatória e dicroísmo circular, foi demonstrado que a γ -zeína de 28 kDa apresenta conteúdo de α -hélice variando entre 19% e 32% e de β -pregueada variando entre 11% e 34%, dependendo do procedimento utilizado (Shewry e Tatham, 1990).

localização celular

Estudos de imunocitocalização têm demonstrado que as γ -zeínas, assim como as β -zeínas, acumulam-se, principalmente, na periferia dos corpúsculos protéicos do endosperma (**Figura 3**), aparentemente desempenhando uma função estrutural, já que apresentam alto teor de cisteína (Lending *et al.*, 1988). O início do desenvolvimento do corpúsculo protéico é marcado pela deposição da β - e γ -zeínas. No decorrer do desenvolvimento do corpúsculo é que se inicia a deposição das α -zeínas no seu interior (Lending *et al.*, 1988; Lending e Larkins, 1989). Os RNAs de γ -zeínas também apresentam níveis mais altos de expressão no início do desenvolvimento do endosperma (Kodrzycki *et al.*, 1989).

Em *Coix* ocorre o oposto: as γ -coixinas são sintetizadas tardiamente no endosperma. Tanto a proteína como seu mRNA aparecem depois dos de α -coixina. As moléculas de α -coixina parecem estar distribuídas por todo o corpúsculo protéico, e não apenas em sua parte mais interna (Targon *et al.*, 1992).

A diferença observada entre milho e *Coix* nos padrões de expressão e localização das prolaminas não era esperada, dadas as semelhanças existentes entre coixinas e zeínas. O motivo desta discrepância ainda está para ser elucidado.

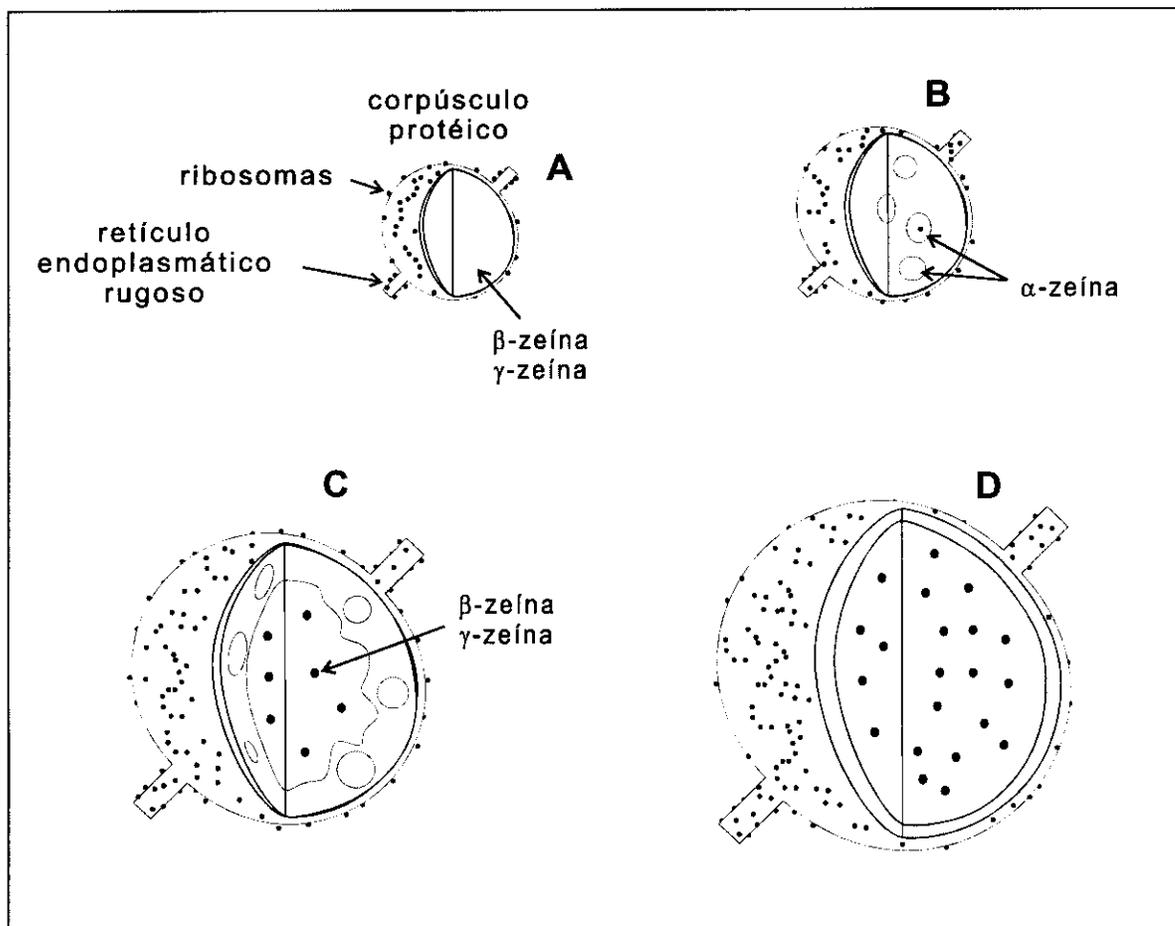


Figura 3: Desenvolvimento dos corpúsculos protéicos em endosperma de milho.

O esquema ilustra a localização das diferentes classes de prolaminas durante o desenvolvimento (A-D) do corpúsculo protéico, segundo o modelo proposto por Lending e Larkins (1989).

regulação

Muito pouco é conhecido sobre a regulação das γ -prolaminas. Ueda *et al.* (1994) mostraram que existe um fator nuclear que se liga a uma seqüência conservada do promotor da γ -zeína de 28 kDa. Mostraram ainda que esta seqüência gera um alto nível de atividade transcricional quando inseridos em diferentes extensões do promotor 35S do CaMV. O fato da seqüência na qual o fator se liga (conhecido como *prolamin box*) ser conservada em todas as classes de prolaminas e em vários outros genes de proteínas de reserva de cereais, entretanto, sugere que ela esteja envolvida com mecanismos regulatórios comuns à expressão coordenada dos genes de prolaminas, e não com a regulação específica das γ -prolaminas.

Demonstrou-se que um determinado fator nuclear liga-se de maneira específica a elementos *cis* localizados entre 900 pb e 1000 pb a montante do ATG inicial nos promotores das γ -zeínas. Este fator é uma proteína de 30 kDa semelhante às HMGs (*high-mobility-group*) e não está presente apenas no endosperma, mas também no embrião e nas folhas, raízes e caule da plântula (Ponte *et al.*, 1994). Nenhuma correlação com atividade é testada neste trabalho.

Este mesmo grupo (Ueda *et al.*, 1992) já havia mostrado que modificações na seqüência do promotor da γ -zeína de 28 kDa faziam com que a proteína O2 o reconhecesse *in vitro* e promovesse a ativação do gene indicador *in vivo*. Novamente, ainda que trabalhando com uma γ -prolamina, o trabalho não traz qualquer contribuição para a compreensão da regulação específica das γ -prolaminas. A γ -zeína é empregada apenas como ferramenta no estudo da regulação de α -zeína de 22 kDa.

Este mesmo grupo mostrou ainda que a eficiência do processamento da extremidade 3' do mRNA da γ -zeína de 28 kDa está correlacionada com o nível de expressão deste gene (Wu *et al.*, 1993). Neste trabalho são identificados os elementos *cis* responsáveis pelo correto processamento da

extremidade 3', mas a especificidade do mecanismo (i.e., se ele é comum a outros genes ou particular ao gene da γ -zeína) não é determinada.

O único trabalho voltado para o entendimento do funcionamento dos promotores de γ -prolaminas (Freitas *et al.*, 1994) mostrou que no promotor de γ -kafirina (a γ -prolamina de sorgo) o elemento responsável pela tecido-especificidade do promotor está localizado na região até 285 pares de base a montante do ATG inicial e que o restante da região promotora parece conter um ou mais elementos responsáveis pela resposta quantitativa deste promotor.

QPM

Há alguns anos, foi verificado que uma determinada variedade de milho *o2*, denominado de QPM (Quality Protein Maize), possui uma modificação no teor de γ -zeína de 28 kDa (Geetha *et al.*, 1991). Essa variedade apresenta uma mutação que se caracteriza pela recuperação do aspecto vítreo do milho normal. Além de apresentar alto teor de lisina, com baixos níveis de α -zeína de 22 kDa, característica do mutante *o2*, esta variedade apresenta 2 a 4 vezes mais γ -zeína de 28 kDa do que o milho normal. Uma análise no conteúdo de mRNA de γ -zeína de 28 kDa, indica que o aumento no acúmulo desta proteína é proporcional ao aumento na taxa de transcrição, produzindo níveis elevados do mRNA correspondente (Wallace *et al.*, 1990; Geetha *et al.*, 1991).

Embora o aumento no acúmulo dessa proteína possa estar relacionado com a dureza e com o aspecto vítreo da semente, isto não explica o alto teor em lisina, pois a γ -zeína é pobre em lisina (Wallace *et al.*, 1990). Habben *et al.* (1995) observaram uma forte correlação entre os conteúdos de lisina e do fator de alongamento 1α (EF- 1α) nos endospermas de milho, sorgo e cevada. O diâmetro dos corpúsculos protéicos nas variedades *o2* (conseqüentemente nas QPM) é de 5 a 10 vezes menor que nas variedades normais. Isto faz com que, mesmo com cerca de metade do conteúdo de zeínas das variedades normais, a superfície total dos corpúsculos protéicos esteja

bastante aumentada (entre duas e cinco vezes maior) nas variedades o2. Isto explicaria o maior nível de uma série de proteínas de citoesqueleto (normalmente associadas com o retículo endoplasmático próximo aos corpúsculos protéicos), entre elas o EF-1 α .

δ -prolaminas

A δ -zeína constitui o menor componente da fração total das zeínas, representando menos que 5% desse conteúdo. Com base nas características de solubilidade da δ -zeína, similares às características das α -zeínas, esta foi inicialmente descrita como α -zeína (Esen, 1986). Entretanto, através do seqüenciamento de um clone de cDNA, foi possível a dedução da estrutura primária, verificando-se que esta proteína deveria constituir uma nova classe, pois não apresentava a homologia esperada com a α -zeína (Kiriwara *et al.*, 1988). Essa classe de proteína é representada por apenas uma ou duas cópias desse gene no genoma haplóide do milho.

Recentemente, foi descrito que os polipeptídeos de zeínas são degradados seqüencialmente da periferia para o interior do corpúsculo protéico durante a germinação da semente, o que pode refletir na localização de cada classe dentro dessas organelas. As γ -zeínas são degradadas antes das outras classes, concordando com sua localização na periferia dos corpúsculos protéicos. As β -, α - e δ -zeínas são degradadas em seguida e de forma mais gradual. Este fato sugere que essas proteínas estejam mais protegidas ao ataque de enzimas proteolíticas (Esen e Stetler, 1992).

Mecanismos de regulação

A regulação da expressão gênica é fundamental para uma infinidade de fenômenos biológicos como a diferenciação e controle celular, desenvolvimento de organismos e resposta a estímulos ambientais. O controle primário da expressão gênica nuclear ocorre ao nível da transcrição e é mediada por proteínas que se ligam a seqüências específicas de DNA

(Latcman, 1990; Kuhlemeier, 1992; Meshi e Iwabuchi, 1995). A divergência entre plantas e animais ocorreu entre 1,5 e 1,1 bilhões de anos. A similitude dos mecanismos de ativação transcricional observada nestes dois grupos leva a crer que a maioria dos processos básicos de transcrição tenha sido conservada por mais de um bilhão de anos (Burke *et al.*, 1990).

Os processos de ativação transcricional são extremamente complexos e se dão através de uma intrincada rede de interações envolvendo uma miríade de componentes, e provavelmente várias vias (Maldonado e Reinberg, 1995). A taxa de transcrição não é determinada apenas pelas propriedades de um único fator de transcrição e seu sítio de ligação, mas pela complexa interação de vários fatores e vários sítios (Kuhlemeier, 1992). Alguns sistemas, como o da regulação da expressão gênica pela luz em plantas (Gilmartin *et al.*, 1990), ilustram muito bem esta complexidade.

O processo inteiro de transcrição pode ser dividido em cinco fases (Maldonado e Reinberg, 1995):

i) formação do complexo pré-iniciação, onde as proteínas relacionadas com a transcrição reconhecem seus sítios;

ii) iniciação, onde os ativadores transcricionais levam à ativação do complexo formado anteriormente;

iii) escape do promotor (*promoter clearance*) que é a saída da RNA polimerase II do promotor para iniciar a transcrição propriamente dita; fase na qual a RNA polimerase II inicial é convertida para uma forma alongadora capaz de deixar o promotor;

iv) alongamento, que é a transcrição propriamente dita, i.e., a síntese do mRNA a partir do molde de DNA. Nesta fase a RNA polimerase II pode encontrar uma série de obstáculos: os físicos, como seqüências de DNA reconhecidas pela própria RNA polimerase II, nucleossomos e outras proteínas ligadas ao DNA, e os funcionais, como níveis limitantes de ribonucleosídeos trifosfatos;

v) terminação.

A regulação da transcrição ocorre, na maioria das vezes, ao nível da iniciação, normalmente envolvendo elementos reunidos no complexo de pré-iniciação (Schleif, 1988).

Não se deve esquecer que existem outros níveis de controle além da transcrição, ainda que a absoluta maioria dos genes descritos até hoje tenha sua expressão predominantemente controlada desta forma. Em alguns genes de plantas, normalmente naqueles sujeitos a controle ambiental ou de desenvolvimento (processos que demandam maior flexibilidade ou velocidade de resposta), não se observa uma estreita correlação entre transcrição e alterações no nível de mRNA nem entre síntese protéica e nível de mRNA. Nestes casos, mecanismos pós-transcricionais podem estar envolvidos na regulação da expressão. Entre estes mecanismos se encontram as modificações no pré-mRNA, transporte do mRNA para o citoplasma, eficiência de transcrição, estabilidade do mRNA e modificações ou *turnover* de proteínas (exemplos podem ser encontrados em Gallie (1993)).

O DNA

A análise de estruturas de cristais de oligonucleotídeos mostrou que a dupla fita de DNA é heterogênea, e que uma parte de suas características estruturais está diretamente relacionada com sua seqüência. A estrutura do DNA é importante em vários processos de reconhecimento molecular, como o que ocorre na ligação de uma proteína com seu sítio no DNA.

As seqüências de DNA envolvidas em transcrição gênica são chamadas de elementos *cis*, e podem ser divididas em duas classes. Uma é representada pelos sítios de ligação do complexo da RNA polimerase, usualmente o *elemento TATA*, também chamado de núcleo do promotor. A outra é formada pelos sítios de ligação dos fatores de transcrição, proteínas que interagem com o complexo da RNA polimerase. Alguns elementos desta segunda classe necessitam de uma posição fixa com relação aos demais

elementos para funcionar corretamente. Outros podem funcionar a distâncias variadas do elemento TATA, e alguns mesmo se sua orientação for invertida, como acontece com os *enhancers* (Kuhlemeier, 1992). Alguns autores (Tjian, 1995) subdividem esta última classe em outras duas. Os elementos que ativam a transcrição são chamados de ativadores e os que a inibem são chamados repressores.

Elemento TATA

A função do elemento TATA na determinação do sítio de início de transcrição vem sendo demonstrada em promotores de plantas há cerca de dez anos (Bruce e Gurley, 1987). Ele dirige, em plantas, fungos e animais, a formação de complexos de iniciação de transcrição estáveis (Buratowski, 1994). O elemento TATA parece ser ainda fundamental na modulação ou estabilização da formação do complexo de iniciação por algumas proteínas transatuantes (ativadores ou repressores). Uma série de trabalhos mostra que modificações na região do elemento TATA modificam a resposta a fatores de transcrição não ligados à transcrição basal (Lin e Green, 1991; Grayson *et al.*, 1995; Zhu *et al.*, 1995). Em alguns genes, a região do elemento TATA está envolvida na determinação da expressão órgão- ou tecido-específica, como é o caso do gene *ghf-1* humano (*growth hormone factor-1*) (McCormick *et al.*, 1991) e do gene da álcool desidrogenase (*adh1*) de milho (Kloeckener-Gruissem *et al.*, 1992).

Foram descritos, até hoje, apenas três promotores em genes de plantas que não contém o elemento TATA (Moffatt *et al.*, 1994). Nos demais, o TATA está presente, ainda que em alguns genes a seqüência descrita para ele seja discordante das descritas para a maioria dos promotores. A maior parte dos elementos TATA descritos foi determinada pela comparação de seqüências, e não experimentalmente.

O único artigo que descreve com detalhes a arquitetura funcional de regiões proximais (*i.e.* aquelas próximas ao início de transcrição) de promotores de plantas mostrou que no gene *pal* (*phenylalanine ammonia-lyase*) de arroz, tanto o elemento TATA como o sinal *cap*, assim como a manutenção da

distância entre estes dois elementos, são necessários para o início correto da transcrição (Zhu *et al.*,1995). Este mesmo trabalho mostra ainda que a composição de bases na seqüência que separa o elemento TATA do sinal *cap* tem pouco efeito na transcrição basal.

Outros elementos cis

Regiões *cis* que controlam a transcrição são tipicamente compostas de vários sítios de ligação de proteínas, o que permite, a partir de variações no tipo e arranjo destes sítios, a criação de complexos nucleoprotéicos únicos, gerando tanto diversidade quanto exatidão na regulação deste processo (Tjian e Maniatis, 1994; Grosschedl, 1995).

Elemento CCAAT

Vários promotores de genes de eucariotos superiores contêm o motivo canônico CCAAT localizado entre 80 e 120 pares de base a montante do sítio de início da transcrição (Albani e Robert, 1995). Análises de mutação demonstraram que esta seqüência exerce efeito positivo em diversos promotores fortes (Johnson e McKnight, 1989) e um motivo similar também funciona como um elemento *cis* em um grande número de genes nucleares de leveduras (Olesen e Guarente, 1990).

Motivos semelhantes ao CCAAT podem ser encontrados em uma série de promotores de genes de planta (Shaw *et al.*,1984; Forde *et al.*,1985; Kreis *et al.*,1986; Matzke *et al.*,1990; Adachi *et al.*,1993; Brignon e Chaubet, 1993; Liaud *et al.*,1993; Higo e Higo, 1996). Alguns autores citam o elemento CCAAT entre os elementos essenciais de promotores de plantas (Pyee e Kolattukudy, 1995), atribuindo a este fator a mesma importância dada ao elemento TATA e empregando sua presença como prova de funcionalidade de genes recém isolados.

Seu envolvimento com a regulação da expressão gênica, contudo, não está claro até o momento (Albani e Robert, 1995), ainda que já tenha sido demonstrada a importância da região que o contém para a expressão máxima

de alguns genes de plantas (Shaw *et al.*,1984; Odell *et al.*,1985). Entre os possíveis papéis deste elemento em plantas, encontra-se a regulação circadiana de alguns genes de *Arabidopsis* (Carré e Kay, 1995) e resposta à luz em genes de várias famílias de plantas, incluindo as gramíneas (Kehoe *et al.*,1994). Exceto nestes trabalhos, entretanto, o elemento CCAAT é descrito apenas pela sua presença.

Por outro lado, alguns trabalhos mostram que sua presença em algumas construções não é suficiente para o correto funcionamento do promotor (Matzke *et al.*,1990). Ainda que o elemento possa ser encontrado em vários promotores de genes de plantas, em outros promotores ele não existe (Walter *et al.*,1994; Chung *et al.*,1995; Guan *et al.*,1996).

Há cerca de dez anos tem sido demonstrado que diferentes fatores reconhecem o elemento *cis* CCAAT (Dorn *et al.*,1987). Destes, o melhor estudado é o C/EBP (*CCAAT/enhancer-binding protein*). Este fator funciona tanto como ativador quanto como repressor da transcrição de vários genes e é capaz de formar dímeros com fatores de classes distintas (Johnson *et al.*,1987; Landschulz *et al.*,1988; Tae *et al.*,1994). Em milho foi demonstrado que as proteínas HMGa/b (*high mobility group*) interagem com certa preferência com os elementos CCAAT e TATA, e o dobramento causado na fita de DNA pela ligação das HMGs parece estar envolvido com estimulação da expressão gênica (Grasser, 1995).

Um fator heteromérico chamado CBF (*CCAAT binding factor*) (também chamado de NF-Y (*Nuclear Factor Y*) ou ainda CP1 (*CCAAT-binding Protein 1*)) (Dorn *et al.*,1987; Chodosh *et al.*,1988)) de eucariotos superiores parece ligar-se de maneira muito específica ao CCAAT. Homólogos de duas de suas três subunidades já foram isolados de milho (Li *et al.*,1992) e nabo (*Brassica napus*) (Albani e Robert, 1995). Nos dois casos, entretanto, as seqüências obtidas mostram que resíduos conservados nas CBFs de animais e leveduras encontram-se substituídas nas CBFs de plantas. Os autores não

foram capazes de determinar se estas substituições refletem diferenças funcionais ou geram diferenças conformacionais nos homólogos de plantas.

Prolamin box

O *prolamin box* é uma seqüência de 7 pb altamente conservada (TGTAAG), geralmente localizada 330 pb a montante do ATG inicial, encontrada na maioria dos genes de prolaminas já isolados (Ueda *et al.*, 1994).

Este elemento parece estar envolvido com a expressão tecido-específica das prolaminas (Schmidt *et al.*, 1992; Hammond-Kosack *et al.*, 1993). Já é sabido, entretanto, que sua presença, apenas, não é suficiente para a manutenção da tecido-especificidade (Ueda *et al.*, 1994) e experimentos com os promotores de α -zeína e γ -kafirina mostraram que a região do *prolamin box* não é necessária para a conservação da tecido-especificidade nestes genes (Matzke *et al.*, 1990; Freitas *et al.*, 1994). Algumas prolaminas, como a δ -zeína, não apresentam seqüências semelhantes ao *prolamin box* em seus promotores (Feix e Quayle, 1993), mas são expressas apenas no endosperma. Assim, o *prolamin box* não deve ser o único elemento relacionado à tecido-especificidade da expressão das prolaminas, ou talvez as evidências que correlacionam o *prolamin box* e tecido-especificidade estejam equivocadas.

O *prolamin box* exerce uma função repressora em alguns promotores (Muller e Knudsen, 1993). Nas glutelinas (prolaminas de trigo), sabe-se que o fator ESBF-I encontra-se precocemente ligado ao *prolamin box*, em uma fase do desenvolvimento da semente que apresenta taxa moderada de transcrição destas proteínas. Posteriormente, quando a taxa de transcrição já é alta, um segundo fator (ESBF-II) encontra-se ligado ao motivo semelhante ao GCN4 localizado próximo ao *prolamin box* (Hammond-Kosack *et al.*, 1993).

Uma das possíveis interpretações para estes achados sobre o *prolamin box* seria que o fator que se liga a ele não teria atividade positiva sobre a transcrição do gene até o instante em que um segundo fator se ligasse ao promotor. Supondo que a ligação do segundo fator dependesse da presença do primeiro no promotor, ou que os dois fatores fossem necessários para a

ativação da transcrição, um promotor que não contivesse um *prolamin box* funcional não poderia ser regulado por este mecanismo.

Proteínas

As proteínas identificam as seqüências nas quais se ligam através do reconhecimento, tanto do padrão de metilação e pontes de hidrogênio específico de cada par de bases, quanto da estrutura geral de uma determinada região da molécula de DNA. Esta estrutura depende também da seqüência nucleotídica (Schleif, 1988).

A cromatina

Histonas são proteínas relativamente pequenas com uma grande proporção de aminoácidos carregados positivamente. São extremamente abundantes, sua massa total sendo bastante próxima da do DNA. Algumas, como H3 e H4 encontram-se entre as proteínas mais conservadas evolutivamente, o que sugere que a absoluta maioria dos seus aminoácidos está implicada com suas funções (Alberts *et al.*, 1989).

O nucleossomo é um cilindro formado por duas cópias de cada uma das histonas H2A, H2B, H3 e H4. Ele constitui a unidade fundamental de empacotamento de DNA. A dupla fita de DNA se encontra espiralada, com 165 pares de base dando duas voltas completas ao redor de cada nucleossomo (Erkine *et al.*, 1995).

Após mais de seis anos de controvérsias, hoje é certo que as histonas são removidas do DNA pelas RNA polimerases durante a transcrição (Kornberg e Lorch, 1995).

Durante muito tempo imaginou-se que estes complexos protéicos cumpriam um papel apenas estrutural na organização da cromatina. Ainda que a maioria dos nucleossomos esteja aleatoriamente posicionada, alguns encontram-se em regiões absolutamente precisas, como os do promotor do gene HSC82 de levedura (Erkine *et al.*, 1995), tornando também mais precisas

as posições ocupadas pelos nucleossomos próximos (Alberts *et al.*, 1989). Hoje sabe-se que o enrolamento de seqüências específicas de DNA ao redor de histonas pode exercer um papel fundamental tanto na ativação como na repressão da transcrição (Wolffe, 1994a).

Esta regulação pode ocorrer de três formas:

- I. Mantendo acessíveis os sítios de ligação de fatores de transcrição. Nucleossomos com posição precisa na fita de DNA fixam a posição da fita, de forma que determinadas seqüências estarão sempre voltadas para as histonas do nucleossomo, ao passo que outras estarão voltadas para a solução.
- II. No caso de sítios bloqueados pela presença de histonas, modificações pós-traducionais ou remoção da(s) histona(s) que bloqueia(m) o sítio, ou ainda a remoção do nucleossomo todo, vão controlar o acesso dos fatores de transcrição.
- III. A presença de nucleossomos pode ainda aproximar elementos regulatórios linearmente distantes permitindo a interação de diversos fatores regulatórios.

Exemplos destes três mecanismos podem ser vistos na regulação do gene da LTR (*long terminal repeat*) do vírus causador de tumor mamário de rato, o MMTV (Yoshinaga *et al.*, 1992). Neste gene, os sítios para o receptor de glicocorticóide são aproximados (forma III de regulação descrita acima) e mantidos expostos (forma I) pelo nucleossomo. A ligação do receptor ao sítio leva à degradação da estrutura do nucleossomo (forma II), permitindo a exposição das demais seqüências regulatórias, com conseqüente início de transcrição.

A alteração da cromatina liberando o acesso a sítios regulatórios parece ser mediada por ativadores globais de transcrição, como os pertencentes à família SNF/SWI de leveduras e seus homólogos em *Drosophila*, rato e no homem (Carlson e Laurent, 1994).

A RNA polimerase II

Eucariotos possuem três enzimas responsáveis pela síntese de moléculas de RNA: as RNA polimerases I, II e III. A RNA polimerase I está envolvida com a transcrição dos grandes genes de rRNA e a RNA polimerase III, com a transcrição dos tRNAs, 5S rRNA e alguns outros snRNAs. A RNA polimerase II é responsável pela síntese de genes que codificam proteínas. As RNA polimerases II de leveduras têm 12 subunidades. Todas já foram clonadas. As diversas RNA polimerases II de plantas isoladas até o momento mantêm uma estrutura semelhante de subunidades, sendo compostas de duas subunidades grandes e oito subunidades pequenas, três delas presentes em duas cópias por molécula e duas em quantidades subestequiométricas (Young, 1991; Kuhlemeier, 1992).

A maior subunidade apresenta peso molecular entre 180 kDa e 220 kDa. A fosforilação do seu domínio carboxi-terminal (CTD) parece ser importante para que o alongamento ocorra. Esta fosforilação evitaria o recrutamento da RNA polimerase II para um complexo de iniciação durante a transcrição (Peterson e Tjian, 1992). O CTD é formado por repetições, com algumas variações, do heptapeptídeo consenso Tyr-Ser-Pro-Thr-Ser-Pro-Ser e o número de repetições parece aumentar conforme a complexidade do eucarioto. Enquanto leveduras apresentam até 27 repetições, o CTD da subunidade maior da RNA polimerase II de plantas tem 34-39 repetições do heptapeptídeo, com substituições na seqüência canônica que parecem ser características de plantas (Dietrich *et al.*, 1990)

Esta enzima multimérica requer uma série de fatores gerais, num total de mais de 20 proteínas (Buratowski, 1994), para iniciar a transcrição sítio específica (Koleske e Young, 1994). Em *Saccharomyces cerevisiae*, a RNA polimerase II parece fazer parte de um complexo gigante contendo cerca de 50 componentes, de massa maior que 3MDa (Halle e Meisterernst, 1996).

Nos últimos 15 anos, na verdade, descobriu-se que o maquinário de transcrição de classe II (i.e., que emprega a RNA polimerase II) é um aparato

surpreendentemente complexo, formado pelos fatores gerais (os ligados à transcrição basal), a RNA polimerase II e os fatores acessórios. Estudos recentes indicam que este maquinário interage ainda com outros complexos protéicos muito grandes, envolvidos na abertura da cromatina e reparo de DNA (Maldonado *et al.*, 1996). Com isso, seu tamanho se aproxima ao de um ribossomo, o que coloca em questão a sua mobilidade e, conseqüentemente, habilidade para encontrar um gene. Alguns autores (Halle e Meisterernst, 1996) sugerem mesmo que o maquinário de transcrição não seja capaz de se deslocar em direção aos genes a serem transcritos, mas, ao contrário, talvez os genes se movam, passando através do maquinário.

Os fatores de transcrição

As proteínas que se ligam às seqüências de DNA ou a outras proteínas, regulando a transcrição gênica, são denominadas fatores *trans*. Algumas destas proteínas são fatores de transcrição gerais, presentes na maioria das células e ativos nas mais diversas condições, talvez até mesmo em todas. Outros fatores são absolutamente especializados (Kuhlemeier, 1992).

Fatores de transcrição são proteínas de natureza modular. Tal fato permite que um número relativamente pequeno de módulos gere o alto nível de especificidade de expressão gênica observado em eucariotos. Um fator de transcrição típico é composto por três módulos: um domínio de ligação ao DNA, um domínio de multimerização e um domínio de ativação. A combinação de módulos em um fator de transcrição o torna único (Tjian e Maniatis, 1994).

Plantas possuem alguns domínios únicos em seus fatores de transcrição. Outros são comuns a fatores de transcrição encontrados nos demais eucariotos. Fatores já conhecidos provenientes de animais e leveduras podem fornecer indícios valiosos sobre os mecanismos de ação e a estrutura dos fatores de plantas, não obstante os fatores destes organismos não serem necessariamente idênticos .

Fatores de transcrição podem responder a estímulos através de três mecanismos distintos. O estímulo pode (1) levar à síntese do fator, (2) gerar modificações no domínio de ligação do fator ao DNA ou ainda (3) modificar o domínio de ativação do fator (Gilmartin *et al.*, 1990).

A fosforilação é um mecanismo comum na geração de modificações nos domínios protéicos. Ela pode regular a atividade de um fator de transcrição modulando sua capacidade de interagir com outras proteínas do complexo ou de se ligar ao DNA, como no caso do TCF (*ternary complex factor*) (Gille *et al.*, 1992), sua capacidade de ativação, como no caso do CREB (*cAMP response element binding protein*) (Gonzales e Montminy, 1989) ou regulando sua localização celular, como no caso de algumas unidades do complexo ISGF3 (*IFN- α stimulated gene factor 3*), as STATs (*signal transducers and activators of transcription*) (Shuai *et al.*, 1993).

Um segundo tipo de modificação covalente pode alterar a atividade de fatores de transcrição. Em algumas bZIPs, como AP-1, c-jun e c-fos, a oxidação de uma cisteína conservada da região básica pode impedir sua dimerização e ligação ao DNA (Wagner e Green, 1994).

Em diversos ativadores a montagem de complexos nucleoprotéicos depende da interação específica de vários componentes. Em alguns, como no caso do promotor da interleucina 2 ou do MHC II (complexo principal de histocompatibilidade classe II), os múltiplos fatores apresentam um comportamento do tipo "tudo ou nada", onde a ausência de um deles faz com que, *in vivo*, nenhum dos outros reconheça seu sítio de ligação (Garrity *et al.*, 1994; Grosschedl, 1995).

Alguns fatores de transcrição, como o LEF-1 em ratos (Giese *et al.*, 1995), YY1, uma proteína *zinc finger* do promotor do gene *c-fos* (Natesan e Gilman, 1993), e HMG I(Y) em humanos funcionam como elementos arquiteturais (Du *et al.*, 1993). Define-se desta forma o fator que causa dobramentos ao se ligar à fita de DNA, posicionando corretamente outros fatores de transcrição e/ou aumentando a afinidade destes por seus sítios. Esta

ação pode se dar por alterações na estrutura do DNA ou por alterações conformacionais nos fatores de transcrição com os quais os elementos arquiteturais mantêm contato. Apesar de sua ação conjunta com outros fatores de transcrição, regulando a expressão gênica, a função arquitetural prescinde de um potencial de ativação transcricional na proteína que a cumpre (Wolffe, 1994b).

Os fatores de transcrição podem ser divididos segundo a forma como agem. Os *fatores ligados à transcrição basal* são aqueles que junto com a RNA polimerase II formam o complexo multiprotéico, ou holoenzima, responsável pela transcrição, nas suas cinco fases. Determinados tipos celulares ou genes possuem *fatores específicos* que podem reconhecer elementos *cis* e modular a atividade da holoenzima (*ativadores* ou *repressores*), normalmente nas fases de ativação e escape do promotor. Por fim, existem ainda fatores protéicos necessários para a comunicação entre determinados ativadores (ou repressores) e a holoenzima, os *fatores intermediários* (Buratowski, 1994; Swaffield *et al.*, 1995; Tjian, 1995).

Fatores ligados à Transcrição basal

Em animais foram identificados seis fatores de transcrição, necessários para a transcrição basal pela RNA polimerase II: TFIIA, TFIIB, TFIID, TFIIE, TFIIF e TFIIH (Buratowski, 1994). Destes, apenas TFIIA e TFIID foram caracterizados em plantas (Kuhlemeier, 1992). O TFIIA de plantas parece ser extremamente parecido com o de animais (Burke *et al.*, 1990).

O TFIID, na verdade, é um complexo protéico formado pela TBP (*TATA binding protein*) e, pelo menos, 8TAFs (*TBP associated factors*) (Tjian, 1995).

A TBP é necessária para a transcrição efetivada por qualquer uma das RNA polimerases (I, II ou III) (Goodrich e Tjian, 1994), sendo por isto chamada de fator de transcrição universal (Struhl, 1994).

De todas as proteínas do complexo basal de transcrição, a TBP é a única que se liga especificamente ao elemento TATA. Ela é uma das proteínas mais conservadas no curso da evolução dos eucariotos. Apresenta mais de 80% de identidade no domínio carboxi-terminal de 180 aminoácidos numa grande variedade de espécies (Struhl, 1994). A conservação das TBPs de diferentes reinos é também funcional. Uma TBP de *Arabidopsis*, a TBP-1, substitui a TBP humana em sistemas de transcrição em células *HeLa* (Mukumoto *et al.*, 1993) e TBPs de milho complementam uma TBP mutante de levedura (Vogel *et al.*, 1993).

A molécula da TBP tem a forma de uma sela, com duas metades bastante simétricas. O que seria a parte de dentro da sela interage com o sulco menor do elemento TATA, causando distorção e desenrolamento no DNA (Kim *et al.*, 1993). A parte de fora da molécula fica acessível para interações com outros fatores de transcrição (Buratowski, 1994). Ainda que tanto a TBP quanto o elemento TATA exibam simetria bilateral, o complexo formado entre os dois tem uma orientação preferencial, o que provavelmente explica a unidirecionalidade da transcrição (Kim *et al.*, 1993).

Mais do que um único papel em todos os complexos do qual faz parte, esta proteína exibe uma dualidade estrutural: ao passo que enquanto parte do TFIID ela se comporta como descrito no parágrafo anterior, nos demais complexos (SL1, TFIIB e SNAP_c) a TBP interage com o DNA de maneira diferente, ou nem mesmo interage diretamente com ele (Struhl, 1994).

A TBP sozinha, entretanto, não intermedeia a regulação por fatores específicos (ativadores/repressores). Este papel requer a presença dos TAFs (Buratowski, 1994).

Os TAFs, por sua vez, não são necessários ou causam qualquer estímulo à transcrição basal (Tjian e Maniatis, 1994). Alguns TAFs servem como pontes de ligação entre os domínios de ativação dos fatores de transcrição e o complexo basal (Goodrich *et al.*, 1993; Gill *et al.*, 1994), se enquadrando, então, na classe dos *fatores intermediários*. TAFs ainda interagem com outros fatores

basais, como o TFIIB (Goodrich *et al.*, 1993). Provavelmente são os TAFs que definem o tipo de promotor ao qual um dado complexo TBP-TAF irá se ligar (Goodrich e Tjian, 1994). Existem pelo menos oito TAFs, e não se descarta a possibilidade da existência de TAFs subestequiométricos ou mesmo TAFs tecido-específicos (Tjian e Maniatis, 1994). Cogita-se, ainda, a possibilidade de TAFs estarem envolvidos em diversos estágios da transcrição: iniciação, escape do promotor, alongamento e terminação.

Além dos TAFs, são descritos em leveduras *fatores intermediários* que interagem não com a TBP, mas com o CTD da RNA polimerase II (Maldonado e Reinberg, 1995; Swaffield *et al.*, 1995).

Em leveduras, antes de se ligar ao DNA, a RNA polimerase II forma uma holoenzima com TFIIB, TFIIF e TFIIH. Esta associação provavelmente é estabilizada pelas SRBs (*suppressors of RNA Pol B* (ou II)) (Koleske e Young, 1994; Maldonado *et al.*, 1996). Este achado diverge do modelo proposto até então, no qual os fatores ligar-se-iam ao promotor de maneira seqüencial: primeiro o TFIID, depois o TFIIB, então a RNA polimerase II e o TFIIF. A este complexo já ligado ao promotor viriam se juntar o TFIIE e o TFIIH (Buratowski, 1994).

O modelo de holoenzima, contrário ao modelo anterior de ligação seqüencial, é corroborado por um outro trabalho (Kim *et al.*, 1994). Neste trabalho, entretanto, a holoenzima identificada é formada por outras subunidades. Dos fatores gerais de transcrição, apenas TFIIF se encontra presente. Estes dois modelos de holoenzima (Buratowski, 1994; Kim *et al.*, 1994), por apresentarem apenas uma fração da RNA polimerase II celular e não conterem todos os fatores gerais de transcrição, não invalidam o modelo de ligação seqüencial. A ordem de ligação dos elementos pode depender de uma série de fatores e os dois achados talvez reflitam dois instantes distintos, isolados, do ciclo celular.

Até o momento, os estudos de transcrição basal empregaram um pequeno conjunto de promotores que contêm elemento TATA. Não se sabe se

outros tipos de promotores requerem este mesmo conjunto de fatores basais (Buratowski, 1994).

Ativadores/Repressores

Uma vez montado o complexo no ativador, os domínios de ativação de alguns fatores devem interagir, direta ou indiretamente, com os componentes do complexo basal de transcrição. Considera-se que o TFIID seja a unidade central de processamento dos sinais regulatórios dos ativadores (Tjian, 1995). A potência final de um ativador de transcrição é determinada por características que vão desde sua afinidade pelo sítio de ligação ao DNA até a força da interação entre seu domínio de ativação e seu alvo (Tjian e Maniatis, 1994).

Alguns ativadores, segundo evidências recentes (Yankulov *et al.*, 1994), parecem estimular também a fase de alongamento pela RNA polimerase II, aumentando a processividade.

Um fator pode ter afinidades diferentes para diversos sítios, pode se ligar ao DNA de maneira cooperativa com outros fatores, ou pode ainda competir com outros fatores pelo sítio de ligação. Esta competição pode se dar porque ambos reconhecem o mesmo sítio. Pode se dar ainda porque seus sítios se sobrepõem, ou estão próximos o suficiente para que os dois fatores não possam estar simultaneamente ligados ao DNA.

A ação de um repressor pode também ser ocasionada por mudanças de conformação do DNA. No caso da proteína YY1, por exemplo, ainda que em alguns contextos ela possa funcionar como ativadora, no gene *c-fos* ela parece organizar a topologia do DNA de uma maneira tal que deixa de existir a interação entre um ativador de transcrição e o complexo basal (Natesan e Gilman, 1993).

Os fatores de transcrição podem ser agrupados em classes em função do emprego de estruturas comuns no reconhecimento do sítio de ligação ou na formação de dímeros. A existência de famílias distintas mostra que há várias soluções para o problema do reconhecimento do DNA por proteínas.

Cada vez se torna mais evidente que os motivos de reconhecimento de DNA, de multimerização e ativação são módulos com plasticidade suficiente para funcionar em mais de uma combinação.

Uma série de fatores de transcrição ainda não pertencem a famílias determinadas. Uma série de outras famílias existem, e imagina-se que muitas serão ainda descobertas. Mesmo com o grande número de fatores de transcrição identificados nos últimos anos, a compreensão dos mecanismos de regulação transcricional em plantas ainda está distante. Os genes alvo da maioria dos fatores ainda não foram delimitados de maneira conclusiva (Meshi e Iwabuchi, 1995).

As classes mais importantes de fatores de transcrição de plantas são as bZIPs (*região básica/zíper de leucinas*), as bHLH (*basic helix-loop-helix*), as MYB, as proteínas que contêm o homeodomínio (HDs), as *MADS box* e as *Zinc Finger* (Kerstetter et al., 1994; Ramachandran et al., 1994; Takatsuji et al., 1994; Bonven et al., 1995; Meshi e Iwabuchi, 1995; Purugganan et al., 1995). Representantes destas classes podem ser encontrados também em fungos e animais. Algumas classes de fatores de transcrição, entretanto, são formadas exclusivamente por proteínas de plantas, como as HD-Zip. As proteínas desta classe distinguem-se por possuírem um zíper de leucinas (como os existentes nas bZIP) adjacente ao domínio HD. Este arranjo torna evidente que os motivos de reconhecimento de DNA, de multimerização e ativação são módulos com plasticidade suficiente para funcionar em mais de uma combinação (Mattsson et al., 1992; Ramachandran et al., 1994; Baima et al., 1995; Kawahara et al., 1995).

A interação entre proteínas envolvidas na regulação da transcrição é muito importante. A possibilidade de formação de heterodímeros é comum nas bZIPs, nas bHLH (Benezra et al., 1990) e nas *MADS box* (Schwarz-Sommer et al., 1992; Goto e Meyerowitz, 1994).

Uma série de interações entre proteínas de classes diferentes já foram descritas. O gene de *Bz1* (UDP-glicose:flavonol 3-O-glicosil transferase, responsável por um dos últimos passos da via de síntese da antocianina) de

milho, por exemplo, é regulado pela MYB C1 e por um membro da família *R/B* (que são bHLH), e esta regulação envolve interações entre as regiões amino terminal das duas moléculas (Goff *et al.*,1990; Roth *et al.*,1991; Meshi e Iwabuchi, 1995). A interação entre proteínas MYB e bHLH na ativação dos genes da via da antocianina é considerado por muitos autores uma das características mais interessantes da pesquisa de fatores de transcrição de plantas (Ramachandran *et al.*,1994). Interações entre *zinc fingers* e bZIPs (Nishio *et al.*,1993) e *zinc fingers* e bHLH (Henderson e Stein, 1994) têm sido também encontradas com relativa frequência.

Em leveduras já foi demonstrado há algum tempo que proteínas MADS e HD interagem na regulação de alguns genes (Herskowitz, 1989). A descrição recente de um gene HD, BEL1, agindo no desenvolvimento floral de *Arabidopsis* (Reiser *et al.*,1995) levanta a possibilidade de membros destas duas famílias interagirem também em plantas (Gasser, 1996).

bZIP (região básica/zíper de leucinas)

Em milho, *opaco2* (*o2*) é o único gene descrito envolvido no controle da expressão de genes de proteínas de reserva. Genes similares foram isolados em *Coix* e sorgo. O produto do gene *o2* de milho é uma proteína bZIP formada por 437 aminoácidos (Schmidt *et al.*,1990), cuja região bZIP encontra-se conservada quando comparadas às proteínas Opaco2 (*O2*) de sorgo e *Coix* (Vettore, 1994).

A proteína Opaco-2 regula a expressão de algumas classes de prolaminas, conforme descrito anteriormente (Schmidt *et al.*,1992), e também de um gene que codifica uma RIP (proteína inativadora de ribossomos) (Bass *et al.*,1992). Sementes que não expressam esta proteína mostram uma redução de 50 a 70% no conteúdo de zeínas (Schmidt *et al.*,1990), a maior parte desta redução devido à queda da expressão dos genes das α -zeínas de 22 kDa.

O domínio de ligação ao DNA das proteínas desta família usualmente apresentam 60 a 80 resíduos formando dois subdomínios distintos.

O primeiro destes domínios, a região básica, responsável pelo reconhecimento da seqüência de ligação ao DNA, é formada por cerca de 30 resíduos. Esta interação ocorre no sulco maior do DNA e o contato entre a proteína e a molécula de DNA determina conformação α -hélice para esta região da proteína. A seqüência alvo de nucleotídeos reconhecida na fita de DNA é primariamente determinada pela combinação dos aminoácidos deste domínio (Kim *et al.*, 1993), mas a região do zíper de leucinas (descrito logo abaixo) e mesmo aminoácidos fora destes dois domínios podem estar envolvidos na determinação da especificidade de ligação (Niu *et al.*, 1994).

O segundo domínio, conhecido como zíper de leucinas, responsável pela dimerização das bZIPs, caracteriza-se por apresentar resíduos repetidos de leucina a cada 7 aminoácidos, numa extensão de 30 a 40 resíduos. A três resíduos das leucinas, no sentido amino terminal, predominam aminoácidos hidrofóbicos (normalmente valina ou isoleucina). O zíper de leucinas quando dimerizado forma uma α -hélice com arranjo *coiled-coil*. Como o *coiled-coil* tem o passo de hélice de 3.5 aminoácidos, cada sétimo resíduo encontra-se no mesmo ambiente estrutural, i.e., na mesma face da molécula. A estrutura desta região determina a dimerização dos zíperes de leucinas através das interações entre os aminoácidos hidrofóbicos e as leucinas (Pabo e Sauer, 1992).

Os heterodímeros de zíperes de leucina detêm papéis importantes na regulação de sua atividade biológica. Alguns fatores de transcrição são ativos apenas quando encontram-se heterodimerizados. Este é o caso da proteína Fos, que é incapaz de formar homodímeros, mas apresenta atividade quando heterodimerizada com a proteína Jun. Outros têm sua atividade diminuída, como é o caso de CREB (cAMP response element binding protein) ao ligar-se a CREM. Heterodímeros podem ainda apresentar especificidade diferente da apresentada pelos homodímeros de seus componentes, reconhecendo sítios distintos ou mesmo levar a novas combinações de domínios de ativação/repressão, mudando, desta forma, a função regulatória das moléculas ligadas a um determinado sítio (Curran e Franza, Jr, 1988; Folkers *et al.*, 1991; Wagner e Green, 1994).

Até 1994, 22 bZIPs já haviam sido identificadas em plantas (Foster *et al.*, 1994). Hoje, 48 proteínas pertencentes a esta classe em plantas podem ser encontradas nos bancos de dados de seqüência. A região básica das bZIPs de plantas apresentam maior similaridade entre si mesmas do que com bZIPs de outros organismos, o que sugere que elas sejam um grupo de bZIPs relacionadas evolutivamente (Niu *et al.*, 1994). Todas as bZIPs de planta que tiveram sua especificidade de ligação determinada até o momento se ligam a seqüências contendo o motivo central ACGT. As seqüências adjacentes ao núcleo ACGT exercem muita influência na afinidade de ligação entre proteína e DNA (Izawa *et al.*, 1993).

Alguns autores dividem as bZIPs em duas subclasses, em função do sítio que reconhecem: uma formada pelas bZIPs similares aos fatores que se ligam ao elemento G (GBFs - *G*-box *b*inding *f*actors) e outra formada pela família TGA (Meshi e Iwabuchi, 1995).

As similares aos GBFs reconhecem o elemento G (G box - CACGTG). A região bZIP destas proteínas se localiza na porção carboxi-terminal da molécula. Na região amino-terminal existe uma região rica em prolina, ou rica em prolina e alanina que, em GBF1 de *Arabidopsis* e em HBP-1a (*histone promoter-binding protein*) de trigo, mostrou possuir potencial ativador (Meshi e Iwabuchi, 1995). São incluídas ainda nesta classe as GBF2 e 3 de *Arabidopsis* (Schindler *et al.*, 1992), GBF4, 9 e 12 de tomate (Meier e Gruissem, 1994) e GBF1 de milho (de Vetten e Ferl, 1995), além da EmBP1 (*Em gene binding protein*) de trigo, TAF-1 (*transcription activator factor*) de tabaco e os CPRF-1 e -3 (*common plant regulatory factor*) de salsa (Weisshaar *et al.*, 1991; Meshi e Iwabuchi, 1995).

As bZIPs agrupadas na subfamília TGA, reconhecem o elemento *cis as-1* (*activation sequence 1*, localizado no promotor do gene 35S do CaMV). As proteínas reunidas nesta subclasse apresentam as seqüências de aminoácidos da região básica extremamente conservadas entre si. Diferente das bZIPs anteriormente descritas, o quarto, e não o terceiro aminoácido após a leucina da

região do zíper de leucinas é que é hidrofóbico. Esta característica faz com que proteínas desta classe formem heterodímeros entre si, mas não com as pertencentes às classes das semelhantes a GBF (Meshi e Iwabuchi, 1995). Na subfamília TGA podem ser incluídas TGA1a e PG13 de tabaco (Fromm *et al.*, 1991), TGA1, 2 e 3, OBF4 (*ocs-element binding factor*), OBF5 e bA19 de *Arabidopsis*, OBF3.1 e OBF3.2 de milho, HBP-1b de trigo e MBF1, MBF2 e MBF3 (*mas-binding factor*) de batata (Feltkamp *et al.*, 1994; Miao *et al.*, 1994).

Algumas bZIPs apresentam características intermediárias (i.e., entre estas duas famílias bem definidas) de especificidade de ligação ao DNA. Entre elas, TGA1b de tabaco, O2 de milho, RITA-1 de arroz e CPRF-2 de salsa. Nestas bZIPs, assim como nas bZIPs GBF, o segundo resíduo hidrofóbico se localiza a três aminoácidos (no sentido N-terminal).

O2 de milho, especificamente, liga-se a sítios alternativos nos promotores de α -prolaminas e do gene *b-32* que, no lugar da seqüência central ACGT, apresenta a seqüência CATG (Yunes *et al.*, 1994).

Várias bZIPs estão presentes simultaneamente em uma mesma célula. O fato de poderem dimerizar em diversas combinações e reconhecerem seqüências similares dificulta bastante esclarecer os fenômenos nos quais as bZIPs identificadas até hoje estão envolvidas.

Materiais e métodos

Material vegetal

As sementes de *Coix lacryma-jobi* L. var. Adlay foram obtidas da coleção de germoplasma do Departamento de Genética da Universidade Estadual de Campinas. As plantas foram mantidas em fecundação aberta na área experimental do Centro de Biologia Molecular e Engenharia Genética da UNICAMP.

Sementes de milho (*Zea mays* L.) F352AG (Agrocere) foram obtidas por autofecundação ou cruzamentos entre os mesmos genótipos na área experimental do Centro de Biologia Molecular e Engenharia Genética da UNICAMP.

Linhagens de bactérias

Em seguida são descritas as linhagens de bactéria (*Escherichia coli*) utilizadas durante a execução dos trabalhos.

DL538

Genótipo relevante: *hsdR mcrA mcrB recD sbcC* (Whittaker *et al.*, 1988)

Esta linhagem foi utilizada como hospedeira do bacteriófago λ DASH durante a seleção de clones recombinantes da biblioteca genômica de *Coix*. A mutação do gene *recD* que elimina a atividade de exonuclease V, bem como a mutação *sbcC*, supressora de mutantes do sistema *recBC* agem diminuindo a perda de fagos recombinantes que carregam fragmentos de DNA genômicos com seqüências invertidas repetidas. As mutações *mcrA* e *mcrB* evitam a restrição de DNAs contendo metil-citosinas nas seqüências 5'-G^mCGC-3' e 5'-AG^mCT-3', o que favorece a construção de bibliotecas genômicas de plantas, cujo DNA é rico em citosinas metiladas (Whittaker *et al.*, 1988).

XL1-Blue

Genótipo relevante: *supE44 endA1 thi-1 lac⁻Δ(lac-proAB) gyrA96 hsdR17 recA1 F' proAB⁺ lacI^f lacZΔM15 Tn10 (tet^r)* (Bullock *et al.*, 1987).

A linhagem *lac⁻*, supressora de mutação âmbar (*supE44*), defectiva em recombinação (*recA1*), deficiente em restrição e proficiente em metilação no sistema Eco K (*hsdR17*), foi utilizada como hospedeira do bacteriófago λZAP durante a seleção de clones recombinantes da biblioteca genômica de *Coix* e nos experimentos de excisão dos fagemídeos dos clones positivos.

Bactérias apresentando o fenótipo F⁺ (contendo um episossomo do tipo F', tal como XL1-Blue) codificam para proteínas de *pilus*, permitindo a infecção por fagos fita simples macho específico, tais como fagos filamentosos do tipo f1. O episossomo F' desta linhagem contém uma deleção na região amino-terminal do gene *lac Z* (*lacZΔM15*), o que permite a α-complementação do segmento do gene presente em determinados plasmídeos como pBluescript. O episossomo F' contém também o gene para superprodução de repressor do operon *lac* (*lac I^f*), bem como os genes marcadores: o que confere resistência a tetraciclina (*Tn10*) e os genes envolvidos no metabolismo de prolina (*proAB*). A linhagem apresenta as marcas cromossômicas para resistência ao ácido nalidíxico (*gyrA*) e requerimento de tiamina para crescimento em meio mínimo. A mutação *endA1* no gene que codifica para endonuclease I, aumenta o rendimento e a qualidade do DNA plasmidial preparado através de técnicas rápidas de minipreparação (Bullock *et al.*, 1987)

DH5α

Genótipo relevante: *F- supE44Δ(lacZYA - argF) U169 (φ 80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*.

Esta linhagem, defectiva em recombinação (*recA1*), foi utilizada na preparação de células competentes para os experimentos de transformação

realizados em todas as subclonagens. A deleção dos genes do operon *lac* ($\Delta lacU169$) e a presença do fragmento *lacZ* $\Delta M15$ no profago $\phi 80$ permite a utilização de α -complementação na seleção de recombinantes. A ausência de superprodução do repressor do operon *lac* (*lac*^r) possibilita a realização do teste de complementação na ausência de IPTG (Hanaham, 1983). Os demais marcadores foram descritos anteriormente para XL1-Blue.

Purificação de DNA de bacteriófago λ

A purificação de DNA do bacteriófago lambda foi realizada conforme a técnica de Yamamoto *et al.* (1970) descrita em Sambrook *et al.* (1989) com algumas adaptações às condições do laboratório.

Inoculou-se 15 ml de meio LB (Sambrook *et al.*, 1989) contendo 10 mM de MgSO₄ com a bactéria *E. coli* (DL-538), incubando-se durante a noite a 37°C, sob agitação vigorosa (300 rpm).

Em um tubo de cultivo, misturou-se 1×10^{10} células ($DO_{600} = 1$ corresponde a 8×10^8 células/ml) com 5×10^7 fagos, incubando-se durante 25 minutos a 37°C sob leve agitação (100 rpm). O conteúdo deste tubo foi inoculado em um Erlenmeyer com 500 ml de meio de cultura LB (Sambrook *et al.*, 1989) contendo 10 mM de MgSO₄, previamente aquecido a 37°C, e mantido sob vigorosa agitação (300 rpm) nesta temperatura. A ocorrência de lise foi verificada através do acompanhamento da DO_{600} do cultivo a partir de 3 horas após o inóculo, com intervalos de 30 minutos, sendo a lise indicada por uma queda brusca da absorbância.

Após essa etapa, adicionou-se 10 ml de clorofórmio ao cultivo, e incubou-se a 37°C durante 10 minutos. O cultivo foi mantido em gelo durante 15 a 20 minutos, sendo em seguida, transferido para tubos de polipropileno de 250 ml e centrifugado a 10.000 rpm em rotor Beckman JA-14 por 10 minutos, a 4°C. O sobrenadante foi transferido para um Erlenmeyer de 2 litros, ao qual adicionou-se RNase e DNase I para uma concentração final de 1 μ g/ml, para eliminação de DNAs e RNAs bacterianos. O lisado foi incubado por 1 hora a

37°C, e adicionou-se NaCl para uma concentração final de 1M (29,2 g para 500 ml). O NaCl auxilia na precipitação de proteínas bacterianas. Após a dissolução do NaCl, o lisado foi mantido em gelo por 1 hora. Em seguida, foi transferido para tubos de polipropileno de 250 ml e centrifugado a 10.000 rpm (rotor Beckman JA-14), durante 10 minutos, a 4°C.

O sobrenadante foi transferido para um Erlenmeyer de 2 litros, e a precipitação das partículas de fagos foi realizada através da adição de polietilenoglicol (PEG-8000) para uma concentração final de 10% (m/v) (50 g para 500 ml de sobrenadante). Essa solução foi mantida em gelo durante a noite.

O lisado foi transferido para tubos de polipropileno de 250 ml e centrifugado a 10.000 rpm (rotor Beckman JA-14), a 4°C, durante 30 minutos. O sobrenadante foi descartado e os sedimentos secos por 10 a 15 minutos à temperatura ambiente. O precipitado foi suspenso em 4 ml de tampão SM. A suspensão foi transferida para tubos Falcon de 15 ml. Adicionou-se 4 ml de clorofórmio, seguida de rápida agitação. O lisado foi centrifugado por 12 minutos a 1.500 rpm (centrifuga de mesa Beckman GPR).

A fase superior foi transferida para um tubo e mantida a 4°C até o momento da ultracentrifugação. Para proceder a ultracentrifugação, adicionou-se 0,5 g de CsCl por mililitro de suspensão de bacteriófago. A suspensão foi depositada no topo de um gradiente descontínuo de CsCl em SM. O gradiente foi montado com as seguintes fases do topo para o fundo do tubo: 1 ml de CsCl 1,45 g/ml (índice de refração 1,3768), 1,5 ml de CsCl 1,50 g/ml (índice de refração 1,3815) e 1,5 ml de CsCl 1,7 g/ml (índice de refração 1,3990). Os tubos foram ultracentrifugados em rotor Beckman SW41 durante 4 horas, a 4°C, a 22.000 rpm. Após a centrifugação, a banda esbranquiçada contendo o fago, depositada entre as fases contendo 1,45 e 1,50 g/ml de CsCl, foi retirada com o auxílio de uma pipeta P1000.

A suspensão de bacteriófagos foi dialisada por duas vezes, durante 1 hora, em 1,5 litro (~1000 vezes o volume de fago recolhido) de solução

contendo 50 mM de Tris-HCl (pH 8,0); 10 mM de NaCl; 10 mM de MgCl₂. Após a diálise, o conteúdo da membrana foi transferido para tubos de microcentrifuga, e adicionou-se EDTA, para uma concentração final de 20 mM, Pronase-E, para uma concentração final de 0,5 mg/ml e SDS, para uma concentração final de 0,5%. Nesta etapa os capsídeos dos bacteriófagos são rompidos e seu conteúdo, o DNA, é liberado. A amostra foi misturada por repetidas inversões e incubada por 1 hora a 37°C.

Após o resfriamento, a amostra foi submetida a extração com fenol equilibrado com Tris-HCl 50 mM (pH 8,0), seguida de uma nova extração com fenol/clorofórmio/álcool isoamílico (25:24:1) e, por último, uma extração com clorofórmio/álcool isoamílico (24:1). O DNA particiona na fase aquosa, e as proteínas na fase orgânica.

Após a última extração, o DNA do bacteriófago foi precipitado com a adição de 0,1 volumes de acetato de sódio 3 M (pH 7,0) e 2 volumes de etanol gelado. A suspensão foi mantida à temperatura ambiente por 30 minutos e, em seguida, centrifugada por 30 minutos, a 4°C, em microcentrifuga. O sedimento obtido foi lavado com etanol 70%. Após a secagem, o sedimento foi solubilizado em 2 ml de TE contendo RNase (40 mg/ml).

Isolamento de DNA nuclear de planta

O DNA de *Coix* foi extraído de plântulas segundo a técnica de Rivin *et al.* (1982), com algumas modificações.

Aproximadamente 1 a 2 g de folhas de plântulas obtidas em condições estéreis e armazenadas a -70°C foram pulverizadas em nitrogênio líquido até a obtenção de um pó fino. A amostra foi homogeneizada em 10 a 20 ml de tampão Tris-HCl 50mM pH8.0; contendo 0,3 M de sacarose e 5 mM de MgCl₂, em um homogeneizador do tipo "polytron" por 30 segundos, na velocidade máxima. O homogenato foi filtrado em 3 camadas de gaze estéril em funil de Buchner com ajuda de bomba de vácuo.

Os núcleos foram sedimentados através de centrifugação a 350 x g por 10 minutos, a 4°C (2.000 rpm em rotor Beckman JA-20). O sobrenadante foi descartado e o sedimento foi suspenso em tampão de lise (Tris-HCl 50mM; contendo 20 mM de EDTA). Adicionou-se sarcosil (n-lauril-sarcosinato de sódio) suficiente para obter-se uma concentração final de 2%, incubou-se por 1 hora a 50°C.

Após a incubação, adicionou-se 0,95 g de CsCl por ml de lisado. As proteínas insolúveis e polissacarídeos foram removidos por centrifugação a 17.000 x g (12.000 rpm em rotor Beckman JA-20), por 15 minutos, a 4°C. Ao sobrenadante coletado adicionou-se brometo de etídio, para uma concentração final de 300 µg/ml. A concentração de CsCl foi ajustada para o índice de refração da solução de 1,3895. As amostras foram transferidas para tubos de ultracentrífuga e submetidas a ultracentrifugação a 50.000 rpm, por 20 horas, a 20°C em rotor VTi-80 (Beckman).

A banda de DNA cromossômico foi removida, cuidadosamente, sob exposição à luz UV e com o auxílio de uma seringa munida de agulha 30x12. O brometo de etídio foi removido por sucessivas extrações com n-butanol equilibrado com água. As extrações foram realizadas até o desaparecimento de cor na fase alcoólica. A solução contendo o DNA foi dialisada por 2 vezes contra 1.000 volumes de TE (Sambrook *et al.*, 1989), por 24 horas cada, obtendo-se um volume final de cerca de: 2 ml. A concentração de DNA foi determinada espectrofotometricamente.

Isolamento em larga escala de DNA de plasmídeo

Para o isolamento de DNA de plasmídeo em larga escala foi utilizada a técnica de lise alcalina descrita em Sambrook *et al.* (Sambrook *et al.*, 1989), com algumas adaptações às condições do laboratório.

A etapa de amplificação descrita no método original foi dispensada para os plasmídeos contendo origem de replicação derivada de Col E1, como no caso de pBluescript KS⁺. A origem de replicação Col E1 apresenta um

controle relaxado de replicação plasmidial, resultando em um alto número de cópias do plasmídeo por célula, e portanto dispensa a amplificação.

A bactéria hospedeira (DH5 α), contendo o plasmídeo, foi inoculada em 25 ml de meio líquido LB/ampicilina (60 μ g/ml) (Sambrook *et al.*, 1989) e incubada a 37°C durante a noite. Esse cultivo foi utilizado para inocular 500 ml de LB/ampicilina líquido, que foi incubado sob vigorosa agitação (300 rpm) durante a noite a 37°C. O cultivo foi centrifugado a 7.000 rpm em rotor Beckman JA-14, por 10 minutos, a 4°C e o sedimento foi ressuspensão em 4 ml de Solução I (Tris-HCl 25 mM pH 8,0, contendo 50 mM de glicose e 10 mM de EDTA).

A suspensão foi transferida para um tubo de centrifuga próprio para rotor Beckman JA-20, adicionou-se 1 ml de solução de lisozima (25 mg/ml em Solução I). A suspensão foi misturada e mantida a temperatura ambiente por 10 minutos. Em seguida, adicionou-se 10 ml de solução II (NaOH 0.2 M; SDS 1%), recentemente preparada, misturando-se por repetidas inversões. A suspensão foi mantida em gelo por 10 minutos. A cada tubo, adicionou-se 7,5 ml de solução gelada de acetato de potássio 3 M pH 4.8, misturando-se por inversão até o desaparecimento das diferentes fases. A suspensão foi mantida em gelo por 10 minutos e centrifugou-se a 12.000 rpm por 20 minutos, a 4°C em rotor Beckman JA-20. O sobrenadante foi filtrado em gaze e transferido para novo tubo de rotor Beckman JA-20. Adicionou-se 0,6 volumes de isopropanol e incubou-se a temperatura ambiente por 10 minutos. Centrifugou-se a 11.000 rpm em rotor Beckman JA-20 por 10 minutos, a temperatura ambiente, e o sedimento obtido foi lavado com etanol 70% a temperatura ambiente.

Após a secagem, o sedimento foi solubilizado em 4 ml de TE. Adicionou-se 4.4 g de CsCl e 0,4 ml de solução de brometo de etídio (10 mg/ml) e a amostra foi centrifugada a 10.000 rpm, durante 5 minutos, a temperatura ambiente em rotor Beckman JA-20 para retirar proteínas precipitadas. A concentração de CsCl foi corrigida para um índice de refração da solução de 1,3860. As amostras foram submetidas a ultracentrifugação em rotor Beckman

VTi-80, a 60.000 rpm, por 20 horas, a 20°C. A retirada da banda correspondente ao DNA plasmidial e a remoção do brometo de etídio foi feita conforme anteriormente descrito para o isolamento de DNA nuclear de plantas. A concentração de DNA foi determinada através de espectrofotometria empregando-se a equação:

$$C = dA \times d \times 50 \mu\text{g de DNA.ml}^{-1}$$

onde:

dA = absorvância a 260 nm

d = fator de diluição

$50 \mu\text{g de DNA.ml}^{-1}$ = 1 unidade de densidade óptica

Purificação de fragmentos de DNA a partir de géis de agarose

Após a digestão com enzimas de restrição apropriadas, o DNA foi submetido à eletroforese em gel de agarose contendo 0,5 $\mu\text{g/ml}$ de brometo de etídio. A concentração do gel de agarose variou de acordo com os tamanhos dos fragmentos produzidos na digestão do DNA. Os insertos foram isolados dos géis utilizando-se papel DEAE (DE-81, Whatman), de acordo com a técnica descrita por Lewis (Chodosh, 1996).

Os fragmentos de DNA foram transferidos eletroforeticamente para papel DE-81 inserido na frente da banda de interesse. A eluição do DNA adsorvido no papel foi realizada incubando-se o papel DE-81 em 800 μl de tampão de eluição (Tris-HCl 10 mM pH 7,9, contendo 10 mM de EDTA e 10 M de NaCl), durante 30 minutos em banho-maria a 68°C. Após este período, o papel DE-81 foi retirado, e a solução restante foi centrifugada por 15 minutos a 4°C. A solução foi transferida para outro tubo de microcentrifuga, tomando-se o cuidado de deixar a solução mais próxima ao fundo do tubo (cerca de 20 μl) no primeiro tubo. O DNA foi precipitado com a adição de 1 ml de etanol gelado e 4 μl de MgCl_2 1M e incubação a -20°C durante a noite. Após centrifugação por 15 minutos a 4°C, os sedimentos foram solubilizados em TE (Sambrook *et al.*, 1989). A concentração de DNA foi determinada espectrofotometricamente

ou, aproximadamente, por análise visual das bandas em gel de agarose com brometo de etídio.

Extração de proteínas nucleares de sementes de Coix

A metodologia empregada na extração de proteínas nucleares de sementes de *Coix* se baseia na descrita por Maier *et al.* (1987) para proteínas nucleares de endosperma de milho.

As sementes de *Coix* foram colhidas conforme o estágio (Targon *et al.*, 1992), imediatamente congeladas em nitrogênio líquido e mantidas a -70°C por até 1 ano.

Todos os procedimentos foram realizados a 4°C. Vinte gramas de sementes no estágio 2 tiveram suas cascas removidas e foram gentilmente maceradas em 90 ml de tampão de homogeneização (Tris-HCl 25 mM, pH8,5; EDTA 25 mM; 0,44 M sacarose; 2,5% ficoll 400 (m/v); 5% dextran T-40 (m/v); 0,5% Triton X-100 (v/v); 2 mM espermina; 0,5 mM espermidina; 10 mM β -mercaptoetanol). O homogenato foi filtrado em *nitex* (poros de 100 μ m) em tubos de vidro e centrifugado por 5 min a 1.000 g.

O sedimento foi suspenso em um total de 10 ml do tampão de solubilização (Tris-HCl 50 mM, pH8,5; MgCl₂ 5 mM; 25% glicerol (v/v); 10 mM β -mercaptoetanol) e centrifugado por 20 min a 20.000 g. O sedimento foi suspenso em 10 ml do tampão de sonicação (Hepes 20 mM, pH 7,5; 25% glicerol (v/v); MgCl₂ 1,5 mM; NaCl 420 mM; EDTA 0,2 mM; PMSF 0,5 mM; DTT 0,5 mM) e tratado com ultra-som para lisar os núcleos.

A suspensão resultante foi centrifugada por 10 min em microcentrífuga para remoção dos grânulos de amido. O sobrenadante foi dialisado durante a noite contra 100 ml do tampão de diálise (Hepes 20 mM, pH 7,5; 20% glicerol (v/v); KCl 100 mM; EDTA 0,2 mM; PMSF 0,5 mM; DTT 0,5 mM), aliqotado em tubos de microcentrífuga e armazenados a -70°C.

Construção da biblioteca genômica

O DNA de *Coix* extraído de plântulas foi digerido parcialmente com a enzima de restrição Mbo I e fracionado através de centrifugação em um gradiente de sacarose por 16 horas a 30.000 rpm (Sambrook *et al.*, 1989).

Frações contendo fragmentos entre 9 kb e 23 kb foram reunidos e clonados no bacteriófago λ DASH (Stratagene) previamente digerido com a enzima de restrição Bam HI. Os bacteriófagos recombinantes foram empacotados *in vitro* com o kit GIGAPACK II - PACKING EXTRACT (Stratagene).

Obtenção de placas de lise

Para seleção de clones, cerca de 7×10^5 placas de lise do bacteriófago foram geradas sobre 23 placas de Petri de 135 mm de diâmetro. Para cada uma das placas, um volume da suspensão de bacteriófagos equivalente a 30.000 pfu foi misturada, em um tubo de ensaio estéril, com 100 μ l de uma suspensão de bactéria XL1-Blue em $MgSO_4$ 10 mM (Sambrook *et al.*, 1989) e incubado a 37°C em um agitador a 100 rpm por 20 minutos. Em seguida, 8 ml de top-agarose previamente aquecida a 60°C foram adicionados ao tubo, acompanhado de leve e breve agitação manual e, rapidamente, todo o conteúdo do tubo foi vertido sobre a placa de Petri contendo meio LB sólido (Sambrook *et al.*, 1989). Após a gelificação da top-agarose as placas foram incubadas a 37°C durante a noite.

Transferência para as membranas

As placas foram mantidas em geladeira por 30 minutos, e em seguida, cobertas com membranas de náilon (Hybond-N, Amersham). As membranas foram marcadas em três posições assimétricas com uma agulha contendo tinta nanquim e foram deixadas sobre as placas durante 1 minuto. As membranas foram então submetidas à incubação por mais 1 minuto em solução de desnaturação (NaOH 0,5 M; NaCl 1,5 M). Após a desnaturação, as

membranas foram incubadas por 5 minutos em solução de neutralização (NaOH 0,25 M; NaCl 1,5 M) sendo em seguida, lavadas em 2X SSC (1X SSC contém 0,15 M de NaCl e 15 mM de citrato de sódio) por 15 minutos. Após a secagem, as membranas foram incubadas por 2 horas em estufa a 80°C.

Sonda

Após dupla purificação em gel de agarose, o fragmento de cDNA de γ -coixina pBCX22.5 (Leite *et al.*, 1991) foi marcado pela incorporação de nucleotídeos radioativos ($[\alpha\text{-P}^{32}]$ dNTPs), através de reação de extensão catalisada pelo fragmento Klenow da DNA-polimerase I, utilizando-se como iniciadores hexanucleotídeos aleatórios (Feinberg e Vogelstein, 1984).

Este clone de cDNA é formado por 840 pb, contendo uma fase de leitura de 600 nucleotídeos precedida por 12 nucleotídeos não traduzidos e 15 nucleotídeos do adaptador de Eco RI (empregado na construção do banco de cDNA do qual este clone foi isolado), seguida de 207 nucleotídeos não traduzidos e do adaptador de Xho I (também utilizado na clonagem).

A sonda foi obtida com atividade específica de 10^8 a 10^9 cpm/ μ g e foi adicionada às soluções de hibridização para uma atividade final de aproximadamente 10^6 cpm/ml. Antes de serem utilizadas nas hibridações as sondas foram desnaturadas em banho-maria fervente por 5 minutos.

Hibridação

As membranas foram pré-hibridadas a 42°C, por 4 horas, em uma solução contendo 5X SSC, 0,05X BLOTTO, 0,02% SDS, 0,1% sarcosil (n-lauril-sarcosinato de sódio) e 50% de formamida deionizada. As hibridações foram realizadas por 12 horas nas mesmas condições da pré-hibridação, utilizando-se como sonda o fragmento de cDNA de γ -coixina pBCX22.5 (Leite *et al.*, 1991).

Os recombinantes positivos foram isolados das placas de Petri (Sambrook *et al.*, 1989) e posteriormente submetidos a mais duas análises sucessivas de hibridação em placas de 85 mm contendo de 1.000 a 3.000 pfu.



Southern Blot

As preparações de *Southern Blot* foram realizadas essencialmente como descrito por Sambrook *et al.* (1989)

Seqüenciamento de DNA

Os seqüenciamentos foram realizados com DNA dupla fita de plasmídeos recombinantes de pBluescript KS⁺. Amostras de 40 µl, contendo 8 µg de DNA plasmidial obtidos através de minipreparação (Sambrook *et al.*, 1989), foram desnaturadas adicionando-se 40 µl de uma solução contendo 0,4 M de NaOH e 0,4 mM de EDTA. As amostras foram incubadas por 5 minutos à temperatura ambiente, e a desnaturação foi interrompida pela adição simultânea de 8 µl de acetato de amônio 2 M pH 4,5 e 176 µl de etanol gelado. As amostras foram mantidas a -20°C por 1 hora e centrifugadas por 20 minutos.

O sedimento foi lavado com etanol 70% e centrifugado novamente por 15 minutos. Após a secagem do sedimento, o mesmo foi dissolvido em 25 µl de água. Alíquotas de 10 µl foram utilizadas nas reações de seqüenciamento segundo a técnica de dideoxinucleotídeos (Sanger *et al.*, 1977). As reações foram realizadas conforme as instruções fornecidas pelo fabricante do "T7 Sequencing Kit" (Pharmacia).

Foram utilizados os iniciadores direto, fornecido pelo kit, e o reverso (Boehringer-Mannheim). Os produtos da reação de seqüenciamento, marcados radioativamente pela incorporação de (³⁵S)dATP_αS (Amersham), foram analisados através de eletroforese em géis desnaturantes de poliácridamida. A eletroforese dos produtos de reação foi realizada carregando-se 2 µl de reação em géis desnaturantes de poliácridamida nas concentrações de 5% e 6% (Sambrook *et al.*, 1989).

Para a confecção dos géis, foram utilizadas placas de 50 cm de altura (LKB), sendo uma delas tratada com Repel-silane (Pharmacia) e a outra com Bind-silane (Pharmacia), de acordo com as recomendações do fabricante.

As amostras foram incubadas por 3 minutos a 80°C antes de serem aplicadas nos géis. Os dois géis foram submetidos a aproximadamente 2.000 volts, sendo que a potência máxima foi estabelecida para 50 watts para o conjunto dos géis, os quais foram montados no sistema de eletroforese Macrophor (LKB). O gel, para leitura da seqüência localizada próxima ao iniciador (gel de 6%) foi submetido à eletroforese até a saída do corante azul de bromofenol. O gel para leitura da seqüência mais afastada do iniciador (gel de 5%) foi submetido à eletroforese por 2 horas adicionais após a saída do corante xileno cianol.

Após a eletroforese, os géis aderidos às placas foram incubados por 30 minutos sob leve agitação em uma solução contendo 10% de ácido acético e 10% de metanol (para retirada de uréia e fixação do DNA), e em seguida, lavados exaustivamente em água corrente. Os géis foram secos a 45°C durante a noite, e submetidos à auto-radiografia através de exposição de filmes sensíveis à radiação (Hyperfilm - Amersham) por 1 a 4 dias.

Deleção unidirecional com a enzima Exonuclease III

A técnica utilizada foi a descrita por Henikoff (Henikoff, 1987) com algumas adaptações às condições do laboratório.

O clone pPH/C (**Tabela IV e Figura 11**) cujo inserto havia sido clonado nos sítios Hind III e Pst I do plasmídeo pBluescript KS⁺, foi digerido com as enzimas Hind III e Apa I. Estas enzimas produzem extremidades 5' e 3' protuberantes, respectivamente. A Exonuclease III age uniformemente sobre extremidades 5' protuberantes ou abruptas, não atacando extremidades 3' protuberantes. Portanto o tratamento do inserto Hind III/Apa I com Exonuclease III permitiu a deleção unidirecional a partir da extremidade Hind III.

Trinta microgramas de DNA foram digeridos com Apa I por 2 horas, a 25°C. O DNA digerido foi então extraído uma vez com fenol/clorofórmio/álcool isoamílico (25:24:1) e uma vez com clorofórmio/álcool isoamílico (24:1). Após a última extração, à fase aquosa foram adicionados 0,1 volumes de cloreto de sódio 2 M e 2 volumes de etanol gelado. A suspensão foi incubada a -70°C, por

30 minutos e, em seguida, centrifugada por 30 minutos, a 4°C. O sobrenadante foi descartado, e o precipitado foi lavado com etanol 70%. O precipitado liofilizado foi suspenso em 57 µl de TE (Sambrook *et al.*, 1989), digerido com Hind III por 2 horas a 37°C e extraído e precipitado como após a digestão com Apa I.

O liofilizado foi solubilizado em 65µl do Tampão de Exonuclease III (Tris-HCl 66 mM pH 8.0, contendo 66 mM de MgCl₂). Uma alíquota de 5 µl foi submetida a eletroforese em gel de agarose para confirmar a completa digestão do DNA. Aos 60µl restantes da amostra, foram adicionadas 300 U de Exonuclease III (GIBCO-BRL). Esta reação foi realizada à temperatura ambiente. Foram retiradas 25 alíquotas, contendo 2,5 µl, a cada 30 segundos. Estas alíquotas foram, imediatamente transferidas para tubos de microcentrífuga que continham 7,5 µl de Mistura de S1 Nuclease (Acetato de Potássio 30 mM; NaCl 250 mM; ZnSO₄ 10 mM; 5% glicerol (v/v); S1 Nuclease 0,2 U/µl). Esta reação foi incubada por 30 minutos, a temperatura ambiente. Adicionou-se então, 1 µl do Tampão de Parada de S1 Nuclease (0,3 M de Trizma Base; 50 mM de EDTA), e incubou-se por 10 minutos a 70°C.

Alíquotas de 2 µl, da reação foram retiradas e submetidas à eletroforese em gel de agarose para conferir a extensão das deleções ocorridas. Ao restante das amostras, adicionou-se 1 µl de Mistura de Klenow (Tris-HCl 10 mM pH 8.0, contendo 200 mM de MgCl₂ e 0,1 U/µl de polimerase Klenow) e 1 µl de Mistura de Nucleotídeos 0,4 mM. As reações foram incubadas por 10 minutos, a 37°C. Após esta incubação, adicionou-se 40 µl da Mistura de Ligase (Tris-HCl 50mM (pH 7,6); MgCl₂ 5 mM; DTT 1 mM; ATP 1 mM; PEG-600 5% (v/v); Ligase 5U/ml), incubando-se, por 2 horas, a 16°C.

Os plasmídeos assim obtidos foram utilizados na transformação de bactérias DH5α. Após a seleção dos clones, foram realizadas minipreparações (Sambrook *et al.*, 1989) para obtenção do DNA utilizado no seqüenciamento.

Bombardeamento com micropartículas

Preparação dos vetores de expressão

O plasmídeo pDMC202 (cedido pelo Dr. Jefferson, R.A. - CAMBIA, Austrália), esquematizado na **Figura 4**, foi empregado como vetor de expressão. Este plasmídeo permite a inserção de seqüências imediatamente a montante da região estrutural do gene da β -glucuronidase, permitindo a avaliação da atividade de tais seqüências na indução da transcrição do gene indicador.

Os oligonucleotídeos mostrados na **Tabela I** foram utilizados em reações de PCR para gerar os fragmentos descritos na **Tabela II** e esquematizados na **Figura 12**. Todas as reações utilizaram o oligonucleotídeo GC4 na geração da extremidade a jusante do fragmento amplificado gerando o sítio para Nco I nesta extremidade em todos os fragmentos.

As reações de amplificação através de PCR foram realizadas utilizando-se a enzima Vent (exo⁻) (BioLabs). Após uma primeira etapa de desnaturação a 94 C por 4 minutos e subsequente adição 0,25 U da polimerase, foram realizados 20 ciclos constituídos por 1 minuto de desnaturação (94°C), 1 minuto de anelamento (50°C) e 1 minuto de extensão (72°C). Após o último ciclo as reações foram mantidas por 10 minutos a 72°C. As reações foram realizadas em Tampão Vent (Tris-HCl 20 mM pH 8,8, KCl 10 mM, (NH₄)₂SO₄ 10 mM, MgSO₄ 10 mM, Triton X-100 0,1% (v/v)), empregando-se 200 ng de DNA molde, 100 pmoles de cada iniciador e 200 μ M de cada um dos desoxinucleotídeos.

Após digestão do produto da reação de PCR com Nco I e a enzima indicada na **Tabela II**, os fragmentos foram inseridos no plasmídeo pDMC202 (aberto com as enzimas apropriadas) e introduzidos em bactérias DH5 α . Todas as construções foram seqüenciadas e não apresentavam erros.

nome	seqüência	anel.	Descrição
iniciado r reverso	GGAAACAGCTATGACCATG	posições 808 a 826 do KS ^{+/}	Anela na região a montante do sítio de policlonagem do plasmídeo KS ^{+/}
DEL	AAG CTCGAG ATTAACCCCTCAC	posições 779 a 799 do KS ^{+/}	Anela na região a montante do sítio de policlonagem do plasmídeo KS ^{+/} . Gera a substituição da G, da C e da A originais pelas T, A e G, respectivamente criando o sítio de Xho I na posição 795
GC1	TATA CTCGAG ATGATCATCAAC	posições -314 a -293 do gene da γ -coixina	Gera a substituição da G, da A e da C originais pelas T, G e A, respectivamente, criando o sítio de Xho I na posição -309
GC2	TAC ACTCGAG TTGCTACATAAAG	posições -162 a -153 do gene da γ -coixina	Gera a substituição da G e da C originais pelas A e G, respectivamente, destacadas na descrição da seqüência, criando o sítio de Xho I na posição -217
GC3	AAAG CTCGAG AGCCGTCGTCAGTAG	posições -162 a -153 do gene da γ -coixina	Gera a substituição da A e das Cs originais pelas C e Gs, respectivamente, criando o sítio de Xho I na posição -157
GC4	CACCT CCAT GGTGTGCGATC	posições -10 a +9 do gene da γ -coixina	Anela na região do início da transcrição do gene da γ -coixina. Gera a substituição da T original pela C, criando o sítio de Nco I na posição +3

Tabela I: Oligonucleotídeos utilizados na construção dos insertos empregados nos experimentos de expressão transitória.

Todas as seqüências são mostradas no sentido 5'-3'. As bases destacadas em negrito estão substituídas nos oligonucleotídeos em relação às seqüências originais do DNA molde das reações de PCR em que foram utilizados. Os dois nucleotídeos grifados em cada seqüência indicam a posição de corte da enzima de restrição indicada.

Preparação de micropartículas

A precipitação do DNA plasmidial, sobre as micropartículas de ouro, foi essencialmente realizada conforme descrito por Sanford *et al.* (1993).

Sessenta miligramas de partículas de ouro de 1 a 3 μ m (Alpha[®] Johnson Matthey) foram pesadas em um tubo de microcentrífuga siliconizado.

Adicionou-se 1 ml de etanol 70%. Após breve agitação, centrifugou-se em microcentrífuga durante 1 minuto. O sobrenadante foi eliminado e adicionou-se 1 ml de água destilada estéril, agitando levemente. O processo de lavagem com água foi repetido mais três vezes.

As partículas foram suspensas em 1 ml de glicerol 50% e conservadas a 4°C por uma semana no máximo.

construção	oligo	DNA molde	extremidade	extensão
rev	iniciador reverso	pP/H C	Hind III	708 pb
3.10	del	del3(10)	Xho I	495 pb
GC1	GC1	pP/H C	Xho I	312 pb
GC2	GC2	pP/H C	Xho I	220 pb
GC3	GC3	pP/H C	Xho I	160 pb
6.10	del	del6(10)	Xho I	161 pb

Tabela II: Oligonucleotídeos utilizados na construção dos insertos empregados nos experimentos de expressão transitória.

A coluna oligo indica apenas o oligonucleotídeo que anela na região a montante e a coluna extremidade indica a enzima empregada na digestão desta região. Em todas as construções o oligonucleotídeo que anela na região a jusante é o GC4 e a enzima empregada na digestão é Nco I. A coluna extensão indica o tamanho do fragmento, em pares de base, após a digestão com as enzimas. As construções 3.10 e 6.10 possuem 46 pb do KS em sua região a montante.

Precipitação de DNA com micropartículas

Todas as soluções empregadas na preparação das micropartículas são estéreis.

Em um tubo de microcentrífuga estéril e siliconizado, sob agitação, contendo 50 µl de micropartículas de ouro, 20 µl de espermidina 0,1 M (base) e 50 µl de CaCl₂ 2,5 M, foram adicionados 5 µl de DNA dos vetores de expressão à concentração de 1 µg/µl. A solução de espermidina, após preparada, foi dividida em alíquotas de aproximadamente 500 µl em tubos de microcentrífuga

e conservada a -70°C por um período máximo de 2 meses. Aliquotas já utilizadas não voltaram a ser congeladas.

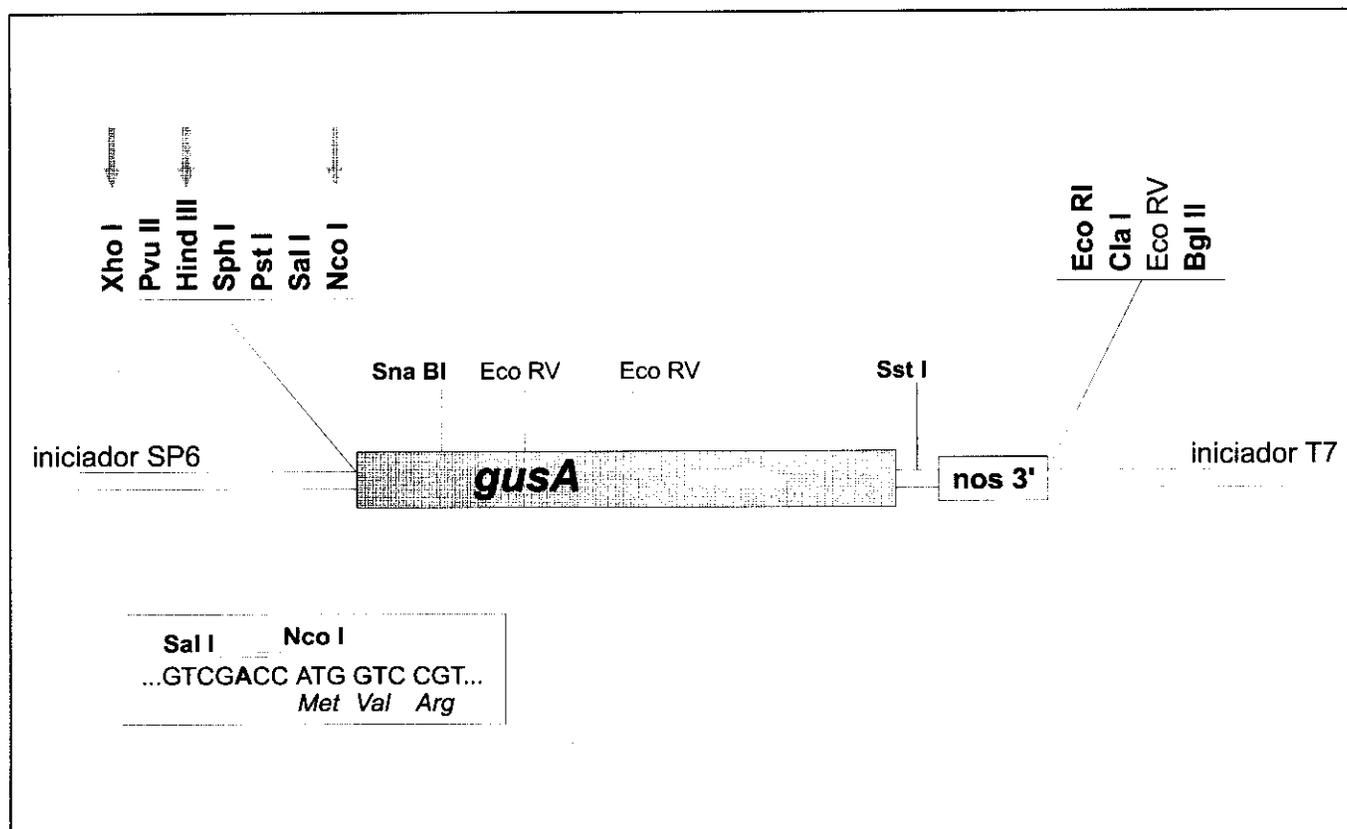


Figura 4: Esquema do plasmídeo pDMC202 empregado na construção dos vetores de expressão.

A região codificadora do gene *gus* está representada pelo quadrado cinza mais escuro. A seta branca mostra a direção em que ocorre a transcrição. Os nomes das enzimas de restrição indicam sua posição de corte. Enzimas que possuem apenas um sítio de restrição no plasmídeo são destacadas em negrito. As empregadas na construção dos vetores de expressão estão indicadas pelas setas cinzas. A adenina na posição -3 relativa ao ATG inicial maximiza a iniciação transcricional em eucariotos. O plasmídeo carrega a resistência ao antibiótico ampicilina como marca seletiva (não mostrada na figura).

A mistura foi agitada durante 5 minutos e centrifugada em microcentrífuga por 15 segundos a 6.000 rpm. O sobrenadante foi cuidadosamente retirado com pipeta de 200 μ l, o sedimento lavado com 140 μ l de etanol 70% e novamente centrifugado por 15 segundos a 6.000 rpm. O processo de lavagem com etanol 70% foi repetido mais duas vezes.

O sedimento foi suspenso em 50 μ l de etanol absoluto e sonicado 3 vezes por três segundos em banho-maria com água a temperatura ambiente. Cinco microlitros das micropartículas foram dispersadas sobre as telas carreadoras, secadas sob vácuo e transferidas para um dessecador.

Bombardeamento de tecidos

Espigas de milho F352AG foram colhidas 16 dias após a polinização e esterilizadas superficialmente por 5 minutos na Solução de Esterilização (hipoclorito de sódio 5% contendo 50 μ l/l de Tween 20). A parte distal das sementes foi eliminada e o pericarpo removido cuidadosamente para expor a camada de aleurona. As sementes foram transferidas para placas de Petri contendo Meio de Cultura MS (Murashige e Skoog, 1962) e posicionadas com a aleurona para cima. As micropartículas foram preparadas com as construções esquematizadas na **Figura 12**.

O bombardeamento dos tecidos foi realizado com um acelerador de micropartículas que utiliza o gás hélio para acelerar as micropartículas (Sanford *et al.*, 1991). A câmara de disparo e componentes internos do acelerador de micropartículas foram esterilizados pulverizando-se álcool 70% e deixando secar em condições estéreis. As telas de aço inox (malha 50) foram esterilizadas através de autoclavagem a 121°C e 1 kgf/cm² durante 20 minutos. Os disparos foram realizados a 1.200 psi com 9 cm de distância entre o tecido alvo e a ponta do canhão. Para cada disparo, vácuo de 710 mmHg foi realizado na câmara de disparo. Gás hélio foi liberado no interior do canhão provocando a quebra do disco de ruptura. Imediatamente após a ruptura, o vácuo foi desfeito permitindo a entrada de ar no interior da câmara de disparo.

Após o bombardeio as placas de Petri foram mantidas na câmara de crescimento durante 48 horas, para subsequente revelação da atividade histoquímica de GUS.

Ensaio histoquímico de GUS

A atividade da β -glicoronidase foi analisada histoquimicamente de acordo com a metodologia descrita por Jefferson *et al.* (1987).

As sementes foram incubados com tampão de revelação X-GLUC (Tampão fosfato de sódio 100 mM pH 7,0; DTT 1mM; Na₂EDTA 10 mM; Triton X-100 0,1%; K₃Fe(CN)₆ 5 mM; K₄Fe(CM)₆.3H₂O 5 mM e X-GLUC 1 mM) a 37°C durante 24 horas. Adicionou-se 20% de metanol ao tampão X-GLUC para inibir a degradação endógena de X-GLUC nos tecidos analisados (Li *et al.*,1993). Após a revelação os tecidos foram conservados em glicerol 50% a 4°C. O número de pontos azuis visualizados nos tecidos decorrentes da atividade de GUS está diretamente correlacionado com as células que receberam uma ou mais micropartículas e que expressaram a atividade da enzima (Morrish e Fromm, 1992).

As contagens dos pontos azuis foram realizadas com auxílio de um estereoscópio. Cada construção foi disparada sobre uma placa de Petri contendo nove sementes preparadas conforme descrito anteriormente e para cada construção 1 repetição foi realizada. A soma dos resultados das nove sementes de cada placa foi considerada um resultado. O desvio padrão (σ) da amostra foi calculado através da fórmula:

$$\sigma = \sqrt{\frac{n\sum x^2 - (\sum x)^2}{n(n-1)}}$$

onde:

n = número de repetições

x = valor de cada repetição

O intervalo de confiança foi calculado, com nível de significância 0,05 (nível de confiança = 95%), através da fórmula:

$$\bar{x} \pm 1,96\left(\frac{\sigma}{\sqrt{n}}\right)$$

onde:

\bar{x} = média das repetições

σ = desvio padrão da amostra

n = número de repetições

Experimentos de retardamento em gel

Preparação das sondas

O clone pP/H C foi digerido com a enzima Taq I, gerando fragmentos de 104 pb , 462 pb e 593 pb. O maior fragmento (593 pb) foi isolado, inserido em KS⁺ e clonado em DH5 α , recebendo o nome de p γ HT (**Figura 11**).

A digestão do inserto do clone p γ HT com a enzima Alu I gera fragmentos de 49 pb, 226 pb e 319 pb. Os dois maiores fragmentos foram isolados, inseridos em KS⁺ e clonados em DH5 α , recebendo os nomes de p γ HT Alu I 226 e p γ HT Alu I 319, respectivamente (**Figura 11**).

Os fragmentos 319 e 226 empregados nos experimentos de retardamento em gel foram isolados através da excisão do inserto dos respectivos clones e purificados em gel de agarose. Os subfragmentos oriundos destes fragmentos foram isolados em gel de agarose a partir da digestão do inserto purificado.

Após a purificação os fragmentos foram marcados através do preenchimento das extremidades 3' com nucleotídeos radioativos ([α -P³²] dNTPs), catalisada pelo fragmento Klenow da DNA-polimerase I e purificados

pelo método *crush & soak* (Sambrook *et al.*,1989). As sondas purificadas apresentavam atividade específica de cerca de 12.000 cpm.ng¹.

Experimentos de retardamento

Um volume contendo aproximadamente 5 µg de proteína de extrato nuclear de sementes de *Coix* no estágio 2 foi adicionado ao tampão de reação (Tris-HCl (pH 7,5) 10 mM, glicerol 10% (v/v), NaCl 50 mM, 0,05 % NP-40 (v/v), 0,5 mM DTT). Nos experimentos que empregavam as sondas 319, 226, 177 e 142, 2 µg de poli-dIdC foram adicionados ao tampão. Naqueles com as sondas menores (75, 53 e 22), apenas 400 ng de poli-dIdC foram adicionados. Nas reações de competição, o competidor, no excesso molar indicado, foi adicionado juntamente com o poli-dIdC ao tampão de reação.

As reações foram pré-incubadas por 5 minutos a 25°C. Cerca de 5 ng de sonda recentemente marcada foram adicionados e a reação incubada por mais 20 minutos. As reações foram corridas em gel não desnaturante 5% poli-acrilamida/bis-acrilamida (75:1) em tampão TBE 0,5X (Sambrook *et al.*,1989), contendo 5% glicerol, a 4°C durante 2 horas a 35 V/cm.

Após a corrida o gel foi transferido para papel-filtro, seco e autoradiografado.

Métodos computacionais

Todos os programas utilizados foram executados em microcomputadores tipo IBM PC (386 SX, 486 DX2 ou Pentium®).

A edição das seqüências, a montagem da seqüência completa do clone genômico de γ -coixina e a tradução da região codante deste gene foi realizada com auxílio dos programas DNASIS v 6.0 (Hitachi) e Gene Runner v3.0 (Hasting Software, Inc.).

A análise de sítios semelhantes aos já descritos para fatores de transcrição foi feito com o auxílio do programa MatInspector (Quandt *et al.*,1995). Além do banco de dados de matrizes que acompanha o programa,

foram construídas novas matrizes com auxílio do programa MatInd (Quandt *et al.*, 1995) para o elemento TATA, sinal *cap* e elemento CCAAT a partir dos dados compilados por Bucher (1990), e para o elemento CATC empregando-se os dados de Forde *et al.* (1985).

O alinhamento das seqüências de DNA dos promotores γ -prolaminas foi realizada com auxílio dos programas Clustal V (Higgins e Sharp, 1989) e ESEE v 1.09g (Cabot e Beckenbach, 1989).

Importante notar que a integração das informações provenientes de cada um dos programas para a confecção final dos resultados foi feita de maneira manual, *i.e.*, os resultados apresentados aqui não são reprodutíveis pela simples repetição dos parâmetros utilizados em um único programa, mas pela interpretação dos resultados obtidos em cada um e subsequente alteração dos dados e parâmetros empregados nos demais.

Resultados

Seleção do clone genômico

Os clones foram isolados através de três seleções sucessivas. Na primeira, 700.000 placas de lise foram transferidas para membranas de náilon (Hybond-N, Amershan) e hibridadas com a sonda pBCX22.5, resultando no isolamento de cinco possíveis clones positivos. Destes, apenas um, o λ 23.1, foi confirmado nas duas hibridações seguintes.

Seqüenciamento

Subclonagem

Devido ao seu tamanho (aproximadamente 22,7 kb), o clone λ 23.1 foi digerido com uma série de enzimas de restrição, como pode ser observado na **Figura 5A**, para obtenção de fragmentos menores (aproximadamente 2 kb) contendo o gene da γ -coixina.

O *Southern blot* do gel (**Figura 5B**) revelou a existência de dois fragmentos de tamanho apropriado para a continuidade dos trabalhos: o fragmento Eco RI/Hind III C, de aproximadamente 3.000 pb e o fragmento Sal I/Hind III B, de cerca de 2.000 pb. Iniciou-se o trabalho de seqüenciamento com o fragmento Eco RI/Hind III C, uma vez que seu maior tamanho aumentava a probabilidade do fragmento conter a região a montante da parte estrutural do gene, normalmente envolvida no controle de transcrição.

O clone pE/H γ CX23.1

O fragmento Eco RI/Hind III C foi isolado em gel e subclonado no plasmídeo pBluescript KS⁺ dando origem ao clone pE/H γ CX23.1 (**Figura 7**). Para obter-se seqüências do interior do clone, o mesmo foi digerido com uma série de enzimas de restrição, mostradas na **Figura 6A**, para as quais o pBluescript KS⁺ possui um único sítio de clivagem, localizado na região de policlonagem.

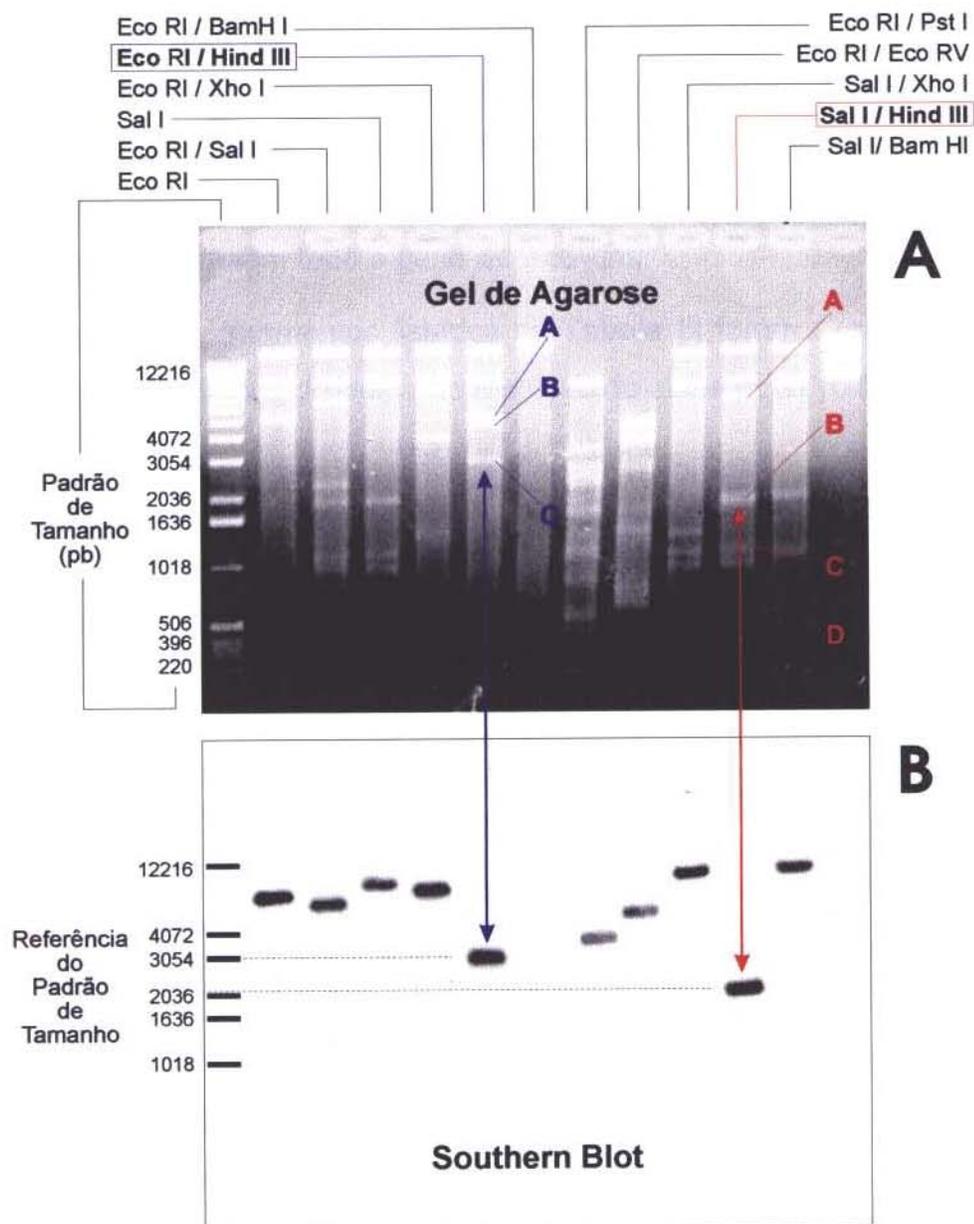


Figura 5: Análise de restrição do clone genômico λ 23.1.

A) O clone genômico λ 23.1 foi digerido com as enzimas de restrição indicadas na figura. As digestões foram submetidas a eletroforese em gel de agarose 0,8%. Estão indicadas as bandas produzidas pelas duplas digestões com as enzimas Eco RI/Hind III e Sal I/Hind III. B) Após a eletroforese, o gel foi fotografado e procedeu-se a transferência alcalina para uma membrana de náilon, com posterior hibridação com a sonda pBCX22.5 (*Southern Blot*). As bandas indicadas na Figura A que hibridaram com a sonda estão identificadas.

O *Southern blot* dos fragmentos gerados (**Figura 6B**) revelou a existência de um sítio de Pst I, já descrito no clone de cDNA, na região codificadora do gene. Este *Southern blot* mostrou ainda que a digestão do clone pE/H γ CX23.1 com a enzima de restrição Sal I ou com as enzimas de restrição Eco RV e Hind III simultaneamente gerava fragmentos menores que 2.000 pb que possivelmente incluíam todo o gene da γ -coixina.

Alguns dos fragmentos listados na **Tabela III** foram seqüenciados e foi possível iniciar o esboço do mapa de restrição do clone pE/H γ CX23.1.

Enzimas	Fragmentos clonados	Nomes dos clones
Pst I/Hind III	B, C e D	pP/H B (ou C ou D)
Eco R V/Hind III	B e C	pEV/H B (ou C)
Sal I	A e B	pSA (ou B)

Tabela III: Fragmentos provenientes do clone pE/H γ CX23.1 clonados em pBluescript KS+.

Os fragmentos obtidos com a digestão do clone pE/H γ CX23.1 com as enzimas indicadas foram designados conforme seu tamanho relativo, o maior sendo o A, o seguinte o B e assim por diante. Compare com a **Figura 6**.

Devido à sua extensão, o clone pP/H C não foi completamente seqüenciado. Não foi possível encontrar, no seqüenciamento parcial do clone pP/H C, sítios de restrição que pudessem ser empregados na obtenção de subclones. Entretanto, o esboço do mapa de restrição do clone pE/H γ CX23.1 indicava que o subclone pP/H C apresentava justamente as seqüências a montante da região estrutural do gene da γ -coixina.

Desta forma procedeu-se a deleção unidirecional com exonuclease III do subclone pP/H C para gerar fragmentos menores que pudessem ser usados na obtenção de sua seqüência completa.

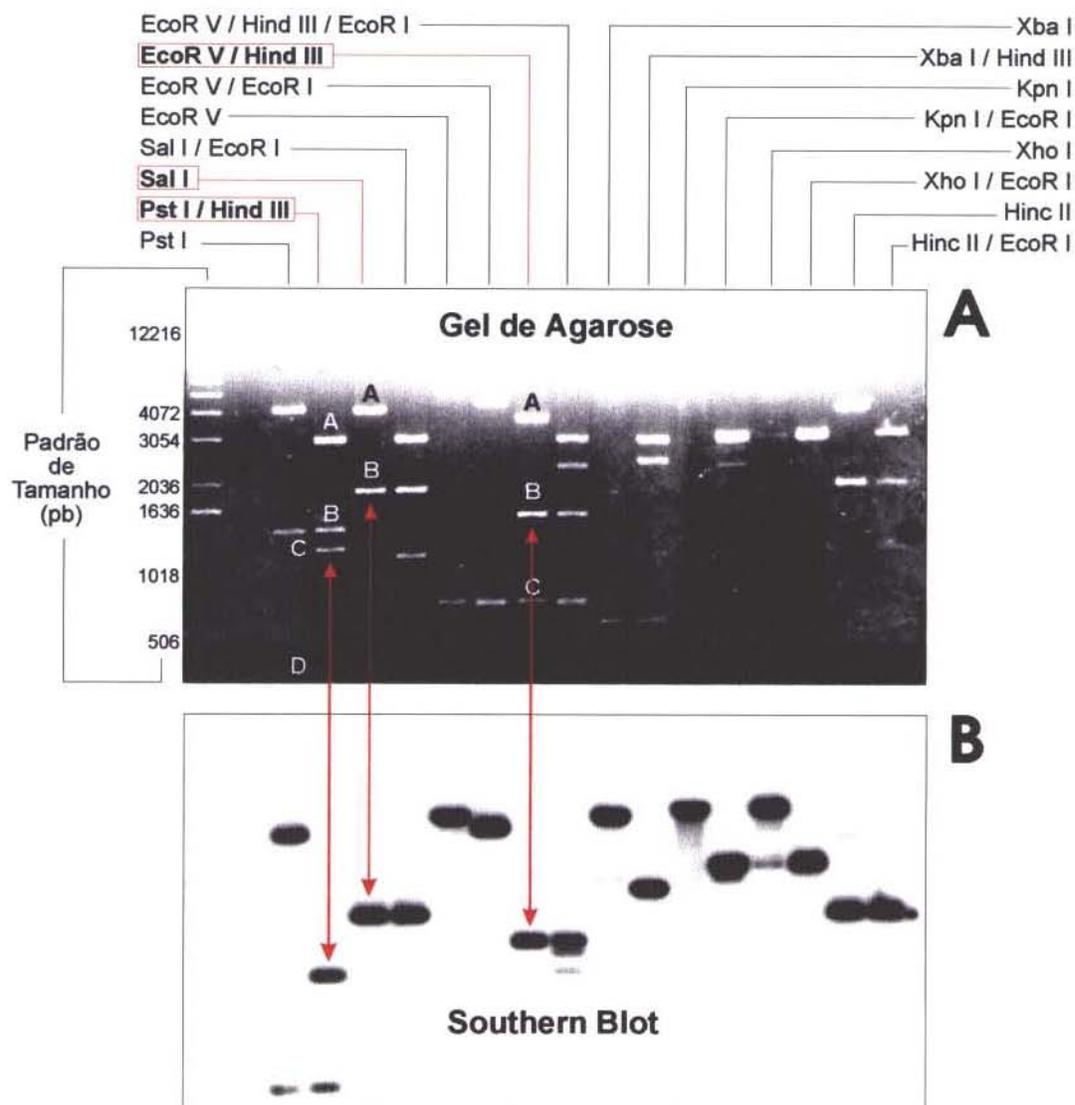


Figura 6: Análise de restrição do clone pE/H γ CX23.1.

A) O clone pE/H γ CX23.1 foi digerido com as enzimas de restrição indicadas na figura. As reações de digestão foram submetidas a eletroforese em gel de agarose 1%. Estão indicadas as bandas produzidas pelas digestões com as enzimas Pst I/Hind III, Sal I e EcoR V/Hind III. **B)** Após a eletroforese, o gel foi fotografado e procedeu-se a transferência alcalina para uma membrana de náilon, com posterior hibridação com a sonda pBCX22.5 (*Southern Blot*). Estão identificadas as bandas indicadas na **Figura A** que hibridaram com a sonda.

A estratégia empregada no seqüenciamento do clone pE/H γ CX23.1, assim como sua posição dentro do clone genômico λ 23.1, está esquematizada na **Figura 7**. Nesta mesma figura, os subfragmentos empregados no seqüenciamento têm definidas suas posições no clone pE/H γ CX23.1. A **Tabela IV** mostra a descrição de todos os clones empregados no seqüenciamento.

Clone	Extremidades	Tamanho (pb)	Origem
pE/H γ CX23.1	EcoRI e HindIII	~3000	sub-fragmento do λ 23.1
pEV/H C	EcoRV	722	sub-fragmento do pE/H γ CX23.1
pP/H C	PstI e HindIII	1161	sub-fragmento do pE/H γ CX23.1
pP/H D	PstI	403	sub-fragmento do pE/H γ CX23.1
pS A	Sal I e EcoRI	~1200	sub-fragmento do pE/H γ CX23.1
pS B	Sal I e HindIII	1835	sub-fragmento do pE/H γ CX23.1
Del3(1)	KpnI e PstI	427	deleção unidirecional do pP/H C
Del3(10)	KpnI e PstI	844	deleção unidirecional do pP/H C
Del3(12)	KpnI e PstI	1032	deleção unidirecional do pP/H C
Del6(10)	KpnI e PstI	569	deleção unidirecional do pP/H C

Tabela IV: Fragmentos utilizados no seqüenciamento.

Todos os fragmentos se encontram clonados em pBluescript KS⁺. A coluna "Extremidades" indica a extremidade dos insertos nestes clones e a coluna "Tamanho" indica o tamanho apenas do inserto.

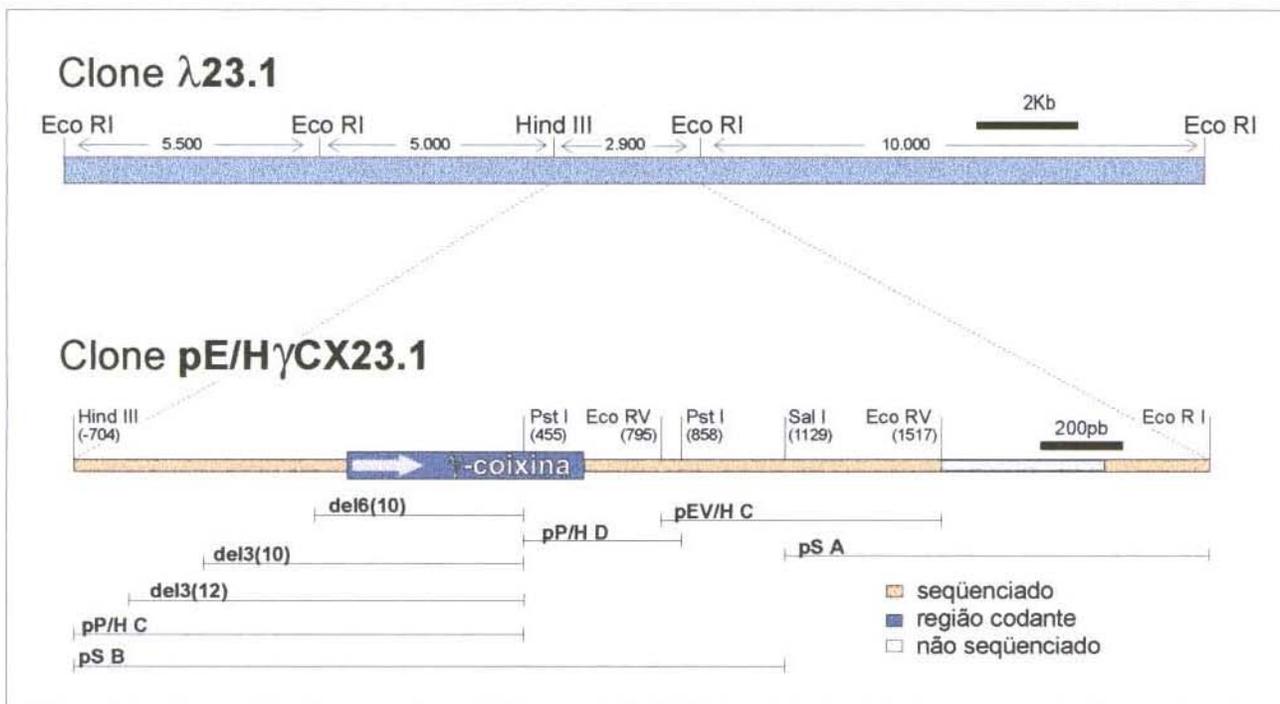


Figura 7: Mapa de restrição do clone λ 23.1 e estratégia de seqüenciamento do clone pE/H γ CX23.1.

Esta figura mostra a posição do fragmento do clone λ 23.1 que originou o inserto do clone pE/H γ CX23.1 (fragmento Eco RI/Hind III C, na **Figura 5**). Em ambos os casos apenas os insertos estão esquematizados, os vetores não. O clone λ 23.1 está clonado em λ dash e o pE/H γ CX23.1 em pBluescript KS⁺. Os números sob os nomes das enzimas de restrição indicam a distância, em pb, relativa ao ATG inicial. O retângulo azul identificado por " γ -coixina" no esquema do clone pE/H γ CX23.1 indica a região estrutural do gene. A seta cinza indica o sentido da transcrição. As linhas sob o esquema do clone pE/H γ CX23.1 indicam as posições dos subfragmentos (**Tabela III**) empregados no seqüenciamento.

Seqüência

Foram seqüenciados 2.225 pb contíguos do gene da γ -coixina no clone pE/H γ CX23.1 contendo uma fase de leitura de 600 pb precedida por 705 pb e seguida por 905 pb (**Figura 8**).

O motivo TGTAAG (conhecido como *endosperm box*, *prolamin box* ou elemento -300), encontrado na maioria genes de prolaminas até hoje descritos (Ueda *et al.*,1994), não está presente no gene da γ -coixina. O posterior alinhamento das seqüências de γ -prolaminas revelou, na região ocupada pelo *prolamin box* nas demais γ -prolaminas, uma seqüência similar ao elemento -300 no promotor da γ -coixina. Na seqüência encontrada em *Coix*, a primeira timidina encontra-se substituída por uma citosina (CGTAAAG).

Na seqüência a jusante da região estrutural do gene de γ -coixina foi encontrado um possível sinal de poliadenilação (AATGAA), localizado a 761 pb do ATG inicial, a 161 pb do códon de terminação da região estrutural do gene e a 23 pb do provável sítio de poliadenilação.

Comparação com seqüências já descritas

Região codificadora

A seqüência obtida para a região codificadora é idêntica à descrita para o cDNA (Leite *et al.*,1991) e caracteriza-se, assim como todas as prolaminas, pela ausência de íntrons.

Região promotora

Foram seqüenciados 705 pares de base da região a montante da região estrutural do gene de γ -coixina. O alinhamento com seqüências já descritas de genes de γ -prolaminas de milho (clone Zc1 (Reina *et al.*,1990a) e Zc2 (Reina *et al.*,1990b)) e sorgo (clone 10S (Freitas *et al.*,1994)) revelou a existência de regiões conservadas em posição e orientação nas quatro seqüências utilizadas nas análises.

-705	AAG CTT GTC TTT GTC CAG AAA TGA GAA ATC ATT GAA GGT GCA CAT GCA	-688	400	CTC AGG CAG GTG GAG CCG CTG CAC DGG CAA CAG GCG ATC TTC GGC GTG	447
-657	TGA TTA AGA AGA CAG GTT TAC ATC ATG GTA TCA ATC AAG ATA CTG AAA	-610	134	Leu Arg Gln Val Gln Pro Leu His Arg Gln Gln Ala Ile Phe Gly Val	149
-609	CAG TAT TCC TTT TAT TTT ATA GGT AAG ATA ATA CAA ATT CTT TTA	-562	448	CTC CTG CAG TCC ATC CAG CAG CAG CCG ATA GGC CAG CCG CTC GCG CTG	495
-561	CTA AAG GAA TCC TTS CRA CTC CCT AAA CGA AAA CTT CTT TAA ATA AAT	-514	150	Val Leu Gln Ser Ile Gln Gln Gln Pro Ile Gly Gln Pro Leu Ala Leu	165
-513	AAA ATA AAC ATT GAT CTG AFA TAT ATG CCA AAT AGC TCA CAC GAC AAC	-466	496	AVG GAG GCG CAA ATA CCG CAG CAA CTG ACG GAG ATG TCC GGT GTG CCG	543
-465	ATT ACA AAC AGG CCC ATA CTA TAC ATC ACA AAG TTT GTT TCA TGA AAA	-418	166	Met Ala Ala Gln Ile Ala Gln Gln Leu Thr Glu Met Cys Gly Val Arg	181
-417	AAC AAA TAA GTA TGC AGG AGG GGA CAA TAA TCC TTG CTT GAC CGS BAA	-370	544	CAG CCA ATG ACA AGT CCG TGC CCT TGC AGC GCT GCT GCC GGC GGT GCC	591
-369	AGT GAA TTT ACA AAG CCA TAT ATC AAC CTA TAT CTA ATT AAT AAG TTC	-322	182	Gln Pro Met Thr Ser Pro Cys Ser Ala Ala Ala Gly Gly Ala	197
-321	GTT ATA TAT AGC CAC GAT CAT CAA CAA CCG TAC CTG TGA AAG CCA	-274	592	CAG TAC TGA ACT GAA GAA ACT ATA TCC TAC TGT AAT AAT AAT GAT GTA	639
-273	ACA AAA TGA GCC ACS CIA AAA TGC AGA ATG AAT CCA TAT GAT GAC GAA	-226	198	Gln Tyr Stop	200
-225	CGT ACA CTC GGC TTG CTA CAT AAA GTG AAT GAT GAG TCA BAA ATA TTT	-178	640	ATA TAG CCG CTG GCG AGC TAA GCT AGT TTA GTC ATT TAG GGC GGC GAT	687
-177	GCC AAG AAA CCG TGA AAG CTA CAC AGC CGT CGT CAG TAG CAC AGS AAC	-130	688	GGG TAA TAA AAG TGT CAT CCA TCA TCA TGG GTG ACA ATA TAT GCG AAC	735
-129	ACA AAA AAC TGT GCT AAT CGA AGC TAT AAA TAA CCC TAG TAT GGC TAT	-82	736	AAA TGA CCT GAA GAT GAA CAA TTG AAA TGA AAA SGA AAA TAT ATT ACA	783
-81	GCA CTT CTC CAT CAC CAC TAC CCA TAT CTT CAG TCT ATT TAC CTT CTC	-34	784	ATT CBA CGA GAT ATC CTC TGG ATC GTA TCA CGT GTC CAC AGG GGT GGA	831
-33	TAT CTA CTC CAG ABA GCA CAG AAG ATC GAC ACC ATG AAG GTG CTG CTC	15	832	TCC ATG CCC CCG GCT GCC CCG GCT GCA GGC GGG GGC GTA GAC CAA AAA	879
16	GTT GCC CTC GCT CTC CTG GCT CTC ACT GGG AGC GCC ACC TCC AGS CAA	63	880	CTT TAG GGA TTC TTT TTT TCA GTT CAA TTT TGC CAA TAA AAA CTA	927
6	Val Ala Leu Ala Leu Leu Thr Ala Ser Ala Thr Ser Thr Gln	21	928	CAT TTA GCC CTA CCI GAT GCG GTC TAA AAI TTT TTT ACA CTA GTT TAG	975
64	ATA GGC GGC ACG TGC TGC CAA CCA CCG CAT CTG CCA CCG CCG CCG	111	976	CCC CTC CTG TAA TGT GTT TCT AGA TCA GGC AAT GTA TTG TAT CAG GGA	1023
22	Ile Gly Gly Thr Cys Gly Cys Gln Pro Pro His Leu Pro Pro Pro	37	1024	CAC GTA TTC ACT TTT TTT TTA TGA ATG TGA CAC GTA ITC ACA TTT GTT	1071
112	GTT CAT CTG CCG CTT CCG GTT CAC CTG CCG CCG GTT CAT CTG CCG	159	1072	GCG GGT AGS CAT ATA TAT ATA TAC CCA GCA APT TCG ATC GAT TAT GAT	1119
38	Val His Leu Pro Pro Val His Leu Pro Pro Val His Leu Pro	55	1120	TAT CAG AAA GTC GAC AAA APT GAG CCA CTG AGC CCA CTC AAA AAA TAA	1167
160	CGC CCA TGC TGC CAC TAC CCT ACT CAA CCG CCC GGG CCT CAG CCT CAG	207	1168	TGG GCT GCG GAA TGT GTT AGS ATA AAC AGA ATT AGA CAA TCA TCT TCA	1215
54	Pro Pro Gln Cys His Tyr Pro Thr Gln Pro Arg Pro Gln Pro Gln	69	1216	ACA AGS ACC TAA ATA AAG TCT CTA CCC TAA TTT GTG TGG CTA GAA AAA	1263
208	CAG CCA CAT CCA TGC CCA TAC CAA CCG CAG CAT CCA AGC CCG TTC CAG	255	1264	AAC ATA TTA TAA GAA CAT CTA CAT CAG GCT CAG CCT ATC ACC AAA ATT	1311
70	Gln Pro His Pro Cys Pro Tyr Gln Pro Gln His Pro Ser Pro Phe Gln	85	1312	TAT GTA CAC CAA AAT TTA TGT ATT GCG CTG ATA AAA AAA ATT ATG TAT	1359
256	TTC CAG CAG CCG GGA ACC TGC GTT GGC CAA GGC ACC CAG ATC CTG GGC	303	1360	TGC GGT AGC CCT ATT AAT TTT GTG AAT ATA AGA AGT TGC TAC AAC GTT	1407
86	Phe Gln Gln Pro Gly Thr Cys Val Gly Gln Gly Thr Thr Ile Leu Gly	101	1408	ACC CAA TAT ATC TCC CRA ATA TCA TAA AAG ATG GGT CCC ACA GGT	1455
304	CAG TGC ATT GAG TTC CTG AGS CAT CAG TGC AGC CCG GCG AGC CCC	351	1456	GTG AAG GGC AAA AGA TGA CCG AAA AGT GAT AGT TGG TGT GGT CTC GTG	1503
102	Gln Cys Ile Gln Phe Leu Arg His Gln Cys Ser Pro Ala Ala Thr Pro	117	1504	CTA TTG TGG CAG ATA TC	1520
352	TAC TGC TCG CCA CAA TGC CAG GCG TTG CAG CAG TGT TGC CAC CAG	399			
118	Tyr Cys Ser Pro Gln Cys Gln Ala Leu Arg Gln Gln Cys His Gln	133			

Figura 8: Sequência completa do gene da γ -coixina e tradução da região codante.

A numeração se refere ao ATG inicial. Os nuceotídeos que fazem parte de possíveis sítios conservados encontram-se destacados em amarelo.

O alinhamento final da região a montante dos genes de γ -prolamina foi precedido pela procura de sítios conhecidos de fatores de transcrição em cada uma das seqüências. A identificação de um sítio conservado em posição, orientação e seqüência nos quatro promotores analisados sugeriria a importância da região contendo tal sítio no controle dos genes de γ -prolamina. Este controle poderia mesmo ocorrer através de mecanismos semelhantes aos descritos para os fatores que originalmente foram empregados na descrição dos sítios.

Análise computacional

Para que a análise fosse bastante completa, construiu-se, com auxílio do programa *MatInd* (Quandt *et al.*, 1995), uma biblioteca de matrizes de sítio de fatores de transcrição a partir dos dados compilados por Bucher (1990) e dos dados propostos por Forde *et al.* (1985). De posse desta biblioteca, utilizou-se o programa *MatInspector* (Quandt *et al.*, 1995) e suas bibliotecas padrão de fatores de transcrição para realizar a inspeção das seqüências.

Empregando-se 0,8 como valor mínimo tanto para a *similaridade de matriz*, quanto para a *similaridade de núcleo*, foram identificadas, nos 705 pb a montante do gene de γ -coixina, 504 fragmentos de seqüências que apresentavam semelhança com seqüências alvo já descritas para fatores de transcrição. Um número de fragmentos próximo a esse foi encontrado também nos 800 pb investigados dos genes de γ -prolamina de milho e sorgo.

Comparação de resultados

A análise do resultado obtido para cada um dos genes de γ -prolamina mostrou que entre os 504 fragmentos encontrados no gene de γ -coixina, 459, provenientes das descrições de 84 matrizes distintas, estavam representados simultaneamente nas seqüências de γ -kafirina e γ -zeínas (**Figura 9B**, barras em azul). Entre estes 459 fragmentos, apenas 32 mantinham conservada posição (*i.e.*, permitiam a sobreposição de suas regiões sem prejudicar o alinhamento

das quatro seqüências) e orientação entre as quatro seqüências (**Figura 9B**, barras em vermelho, **Figura 9A e Tabela V**).

Foi observada a ocorrência, entre estes 32 fragmentos, de regiões de sobreposição, *i.e.*, numa dada região várias matrizes distintas compartilhavam os mesmos nucleotídeos. Neste caso, toda a região delimitada pelas matrizes sobrepostas foi considerada como um bloco único a ser analisado, restando assim 13 regiões distintas conservadas (**Tabela V**). Observando o Grupo 8 (**Tabela V**), por exemplo, dizemos que as matrizes V\$HFH2_01, V\$NF1_Q6, ccaat e I\$HAIRY_01 formam um único grupo porque os sítios descritos por suas matrizes encontram-se sobrepostos no promotor das γ -prolaminas.

Possíveis sítios encontrados

TATA

Das 13 regiões descritas, apenas uma, bastante semelhante ao elemento TATA, obteve resultados de *similaridade de matriz* maiores que 0,95 em todas as seqüências analisadas (**Figura 9A e Tabela V**). Este possível elemento TATA está representado em *Coix* pela seqüência CTATAAATAACCCTA, localizada 105 pb a montante do ATG inicial. Sua posição varia apenas 9 pb entre as seqüências analisadas. A conservação de posição entre os vários clones de γ -prolaminas analisados, o alto valor obtido na comparação com a matriz que descreve o elemento TATA e o fato de haver duas seqüências consenso do sinal *cap* (Bucher, 1990), uma 35 pb e a outra e 24 pb a jusante de sua posição, indicam que o sítio analisado possivelmente representa um elemento TATA funcional.

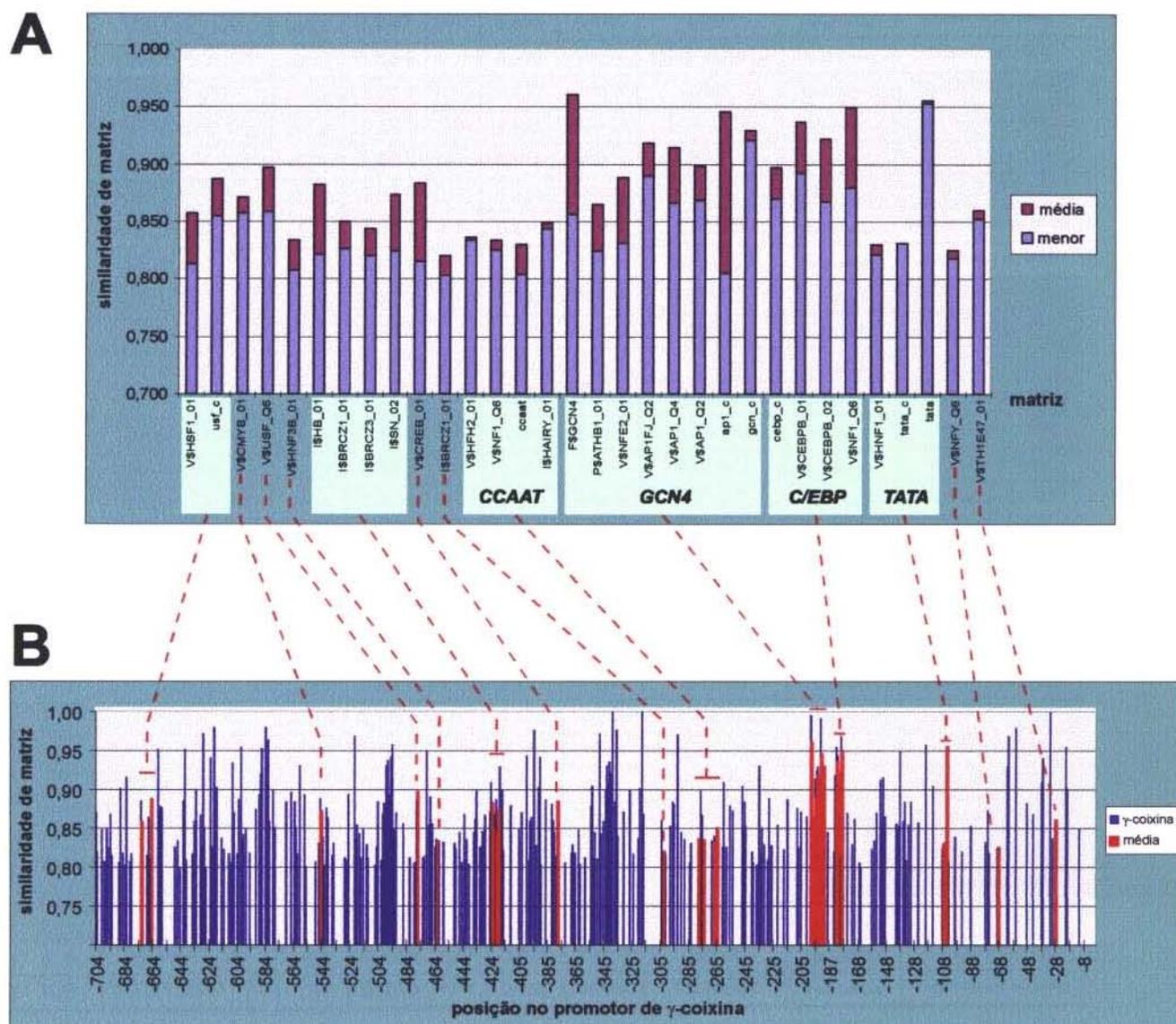


Figura 9: Análise da Região Promotora dos Genes de γ -Prolaminas.

A) Valores de *similaridade de matriz* das 13 regiões que apresentam similaridade com seqüências alvo já descritas para fatores de transcrição e os 32 resultados que a compõem (Tabela V). Os quadros verdes na abscissa agrupam as várias matrizes que formam algumas regiões. A barra vermelha representa a média dos valores encontrados nos promotores de γ -prolaminas para a matriz indicada na abscissa. A barra azul indica o valor mais baixo empregado no cálculo da média. B) Os valores de *similaridade de matriz* de todos os 459 fragmentos encontrados em *Coix* semelhantes a seqüências alvo já descritas para fatores de transcrição estão representados em azul, em função da posição da extremidade 5' do elemento no promotor. Em vermelho, os valores médios encontrados para as 32 matrizes conservadas em posição e orientação nos promotores de γ -prolaminas na posição em que são encontradas no promotor de *Coix*. As linhas pontilhadas relacionam as 32 matrizes conservadas em posição e orientação nos promotores de γ -prolaminas em ambos os gráficos. Note que algumas vezes duas matrizes foram descritas na mesma posição e, neste caso, a de maior valor é a que se observa no gráfico B.

grupo	rel.	matriz	pos.	média		menor		fator		referência
				valor	σ	valor	origem	nome	origem	
1	a ⁻ s ⁻	V\$HSF1_01	-672	0.858	0.031	0.813	zeína 28	HSF1	vertebrados	(Kroeger e Morimoto, 1994)
	a ⁻ s ⁻	usf_c	-665	0.887	0.022	0.855	coixina	USF	vertebrados	(Quandt <i>et al.</i> , 1995)
2	a ⁻	V\$CMYB_01	-545	0.872	0.011	0.858	zeína 28	c-Myb	vertebrados	(Nakagoshi <i>et al.</i> , 1990; Howe e Watson, 1991)
3	a ⁻	V\$USF_Q6	-477	0.897	0.037	0.859	kafirina	USF	vertebrados	(Quandt <i>et al.</i> , 1995)
4	s ⁻	V\$HNF3B_01	-463	0.834	0.018	0.808	kafirina	HNF-3beta	vertebrados	(Quandt <i>et al.</i> , 1995)
5	a ⁻	I\$HB_01	-424	0.883	0.036	0.822	zeína 16	Hb	Drosophila	(Stanojevic <i>et al.</i> , 1989)
	a ⁻	I\$BRCZ1_01	-424	0.851	0.018	0.826	zeína 16	BR-CZ1	Drosophila	(von Kalm <i>et al.</i> , 1994)
	a ⁻	I\$BRCZ3_01	-421	0.844	0.027	0.820	zeína 16	BR-CZ3	Drosophila	(von Kalm <i>et al.</i> , 1994)
	a ⁻	I\$SN_02	-420	0.873	0.037	0.824	coixina	Sn	Drosophila	(Kasai <i>et al.</i> , 1992)
6	a ⁻	V\$CREB_01	-378	0.884	0.054	0.815	coixina	CREB	vertebrados	(Benbrook e Jones, 1994)
7	a ⁻	I\$BRCZ1_01	-303	0.821	0.012	0.803	coixina	BR-CZ1	Drosophila	(von Kalm <i>et al.</i> , 1994)
8	!	V\$HFH2_01	-278	0.836	0.002	0.834	zeína 16	HFH-2	vertebrados	(Overdier <i>et al.</i> , 1994)
	!	V\$NF1_Q6	-274	0.834	0.010	0.825	zeína 28	NF-1	vertebrados	(Quandt <i>et al.</i> , 1995)
	!	ccaat	-267	0.831	0.015	0.804	zeína 16		eucariotos	(Bucher, 1990)
	!	I\$HAIRY_01	-265	0.849	0.004	0.844	zeína 28	Hairy	Drosophila	(van Doren <i>et al.</i> , 1994)
9	!	F\$GCN4	-198	0.960	0.060	0.856	zeína 28	GCN4	leveduras	(Oliphant <i>et al.</i> , 1989)
	!	P\$ATHB1_01	-197	0.865	0.043	0.824	coixina	Athb-1	Arabidopsis	(Sessa <i>et al.</i> , 1993)
	!	V\$NFE2_01	-195	0.889	0.034	0.832	zeína 28	NF-E2	vertebrados	(Andrews <i>et al.</i> , 1993)
	!	V\$AP1FJ_Q2	-194	0.918	0.016	0.890	zeína 28	AP-1, c-Jun/c-Fos	vertebrados	(Quandt <i>et al.</i> , 1995)
	!	V\$AP1_Q4	-194	0.915	0.028	0.866	zeína 28	AP-1	vertebrados	(Quandt <i>et al.</i> , 1995)
	!	V\$AP1_Q2	-194	0.898	0.017	0.869	zeína 28	AP-1	vertebrados	(Quandt <i>et al.</i> , 1995)
	!	ap1_c	-193	0.945	0.081	0.805	zeína 28	AP-1	eucariotos	(Quandt <i>et al.</i> , 1995)
	!	gcn_c	-192	0.930	0.006	0.921	zeína 28	GCN4	eucariotos	(Quandt <i>et al.</i> , 1995)
10	!	cebp_c	-183	0.897	0.018	0.870	kafirina	C/EBPbeta	eucariotos	(Quandt <i>et al.</i> , 1995)
	!	V\$CEBPB_01	-182	0.937	0.026	0.892	kafirina	C/EBPbeta	vertebrados	(Akira <i>et al.</i> , 1990)
	!	V\$CEBPB_02	-182	0.923	0.032	0.867	kafirina	C/EBPbeta	vertebrados	(Johnson e Williams, 1994)
	!	V\$NF1_Q6	-180	0.950	0.040	0.880	kafirina	NF-1	vertebrados	(Quandt <i>et al.</i> , 1995)
11	!	V\$HNF1_01	-108	0.830	0.014	0.822	coixina	HNF-1	vertebrados	(Tronche e Yaniv, 1992)
	!	tata_c	-106	0.832	0.000	0.832	coixina	TBP	retrovirus	(Quandt <i>et al.</i> , 1995)
	!	tata	-105	0.955	0.001	0.953	coixina	TBP	eucariotos	(Bucher, 1990)
12	a ⁻	V\$NFY_Q6	-69	0.825	0.005	0.818	zeína 16	NF-Y	vertebrados	(Quandt <i>et al.</i> , 1995)
13	s ⁻	V\$TH1E47_01	-28	0.860	0.007	0.852	kafirina	Th1/E47	vertebrados	(Hollenberg <i>et al.</i> , 1996)

Tabela V: Regiões de alta similaridade com seqüências alvo de fatores de transcrição conhecidos.

A coluna grupo mostra as 13 regiões de sobreposição encontradas para os 32 fragmentos. Os nomes na coluna Matriz se referem ao nome que as matrizes empregadas recebem no programa MatInspector (Quandt *et al.*, 1995), ou o nome dado para as matrizes descritas em Bucher (1990). A coluna rel. Indica a relevância das seqüências encontradas: o sinal ! indica que as regiões descritas estão conservadas em seqüência, posição e orientação; s⁻ significa que as seqüências encontradas nos quatro promotores são discrepantes; a⁻ significa que o alinhamento da região descrita nos quatro promotores iria prejudicar o alinhamento como um todo; a⁻ s⁻ indica que as seqüências, além de discrepantes, se alinhadas, prejudicariam o restante do alinhamento. A posição (pos.) se refere à distância em pares de base entre a base mais a montante da seqüência descrita pela matriz e o ATG inicial no promotor da γ -coixina. Média mostra o valor da média aritmética do valor encontrado para a matriz em posições equivalentes nos quatro promotores analisados, e σ o desvio padrão para esta média. A coluna menor mostra o menor valor empregado no cálculo da média e de qual promotor ele provém (origem). Fator mostra o nome do(s) fator(es) empregado(s) na determinação das matrizes e em que grupos de organismos tal fator já foi descrito (origem).

CCAAT

Dos três elementos conhecidos como CCAAT (Bucher, 1990) encontrados na região a montante do gene de γ -coixina, apenas um, representado pela seqüência ATGAGCCACGCA, encontra-se conservado em todos os clones analisados (**Figura 9A e Tabela V**). Este possível elemento CCAAT localiza-se 267 pb a montante do ATG inicial, o que, neste gene, significa 188 ou 199 pb a montante dos prováveis inícios de transcrição (sinal *cap* descrito no parágrafo anterior).

GCN4 e C/EBP

Duas outras regiões conservadas nos promotores estudados apresentam alta similaridade com seqüências alvo de fatores de transcrição já descritos. A 192 pb a montante do ATG inicial do gene de γ -coixina, encontra-se um sítio muito semelhante ao descrito para GCN4, um fator de transcrição de leveduras (Hinnebusch, 1984; Oliphant *et al.*, 1989). Em *Coix* o sítio está representado por TGAGTCATAA.

Logo adiante, a 180 pb a montante do ATG inicial, observa-se uma região com seqüência (ATTTGGCAAGAAACCGTG) bastante similar às descritas para o C/EBP, um fator de transcrição específico de fígado de ratos (Johnson *et al.*, 1987).

Demais sítios

Entre as nove regiões restantes, sete ocupam posições cuja justaposição acabaria por impossibilitar o alinhamento de várias outras regiões ou obrigaria a inserção de extensas lacunas em três das quatro seqüências empregadas no alinhamento (**Tabela V**, matrizes identificadas pela letra **a** na coluna **rel.**), enquanto quatro não mantêm similaridade de seqüência entre os diversos genes (**Tabela V**, matrizes identificadas pela letra **s** na coluna **rel.**). Estas regiões de seqüências divergentes são descritas pelo programa *MatInspector* como um mesmo sítio devido à variação na seqüência consenso que pode ser observada em alguns fatores de transcrição.

O alinhamento final dos promotores de γ -prolaminas pode ser observado na **Figura 10**. Nela os sítios conservados que apresentam semelhança com sítios já descritos de fatores de transcrição são identificados pelo nome destes fatores, sem que isso signifique envolvimento de tais fatores com a transcrição destes genes. São indicadas ainda na **Figura 10** a região onde se encontra o elemento -300 (*prolamin box*) e duas regiões bastante conservadas, identificadas por *Palind* e *Desconhecido*, para as quais não foi possível encontrar similaridade com sítios já descritos de fatores de transcrição. A região identificada por *Palind* destaca-se por ser palindrômica e, ainda que distante do sítio de início de transcrição (cerca de 500 pb), estar absolutamente conservada em todas as seqüências. A posição das regiões conservadas encontra-se descrita também na **Figura 10**.

Experimentos de expressão transitória

Para investigar a atividade do promotor de γ -coixina, a superfície do endosperma imaturo de sementes de milho foi bombardeada com construções contendo o gene *gus* sob controle do promotor de γ -coixina.

A opção por sementes de milho, e não de *Coix*, deve-se ao fato de a polinização em milho poder ser controlada - condição necessária para a obtenção de grupos de sementes com estágio determinado de desenvolvimento. Experimentos anteriores demonstram a conservação entre *Coix*, sorgo e milho, dos fatores que controlam a tecido-especificidade das γ -prolaminas. Demonstraram também que a variação nos níveis de resposta à γ -kafirina é mantida nas três espécies (Freitas *et al.*, 1994).

De posse dos resultados do alinhamento do promotor do gene da γ -coixina com promotores de γ -prolamina (**Figura 10**) as seis construções ilustradas na **Figura 11** foram empregadas em experimentos de bombardeamento com micropartículas.

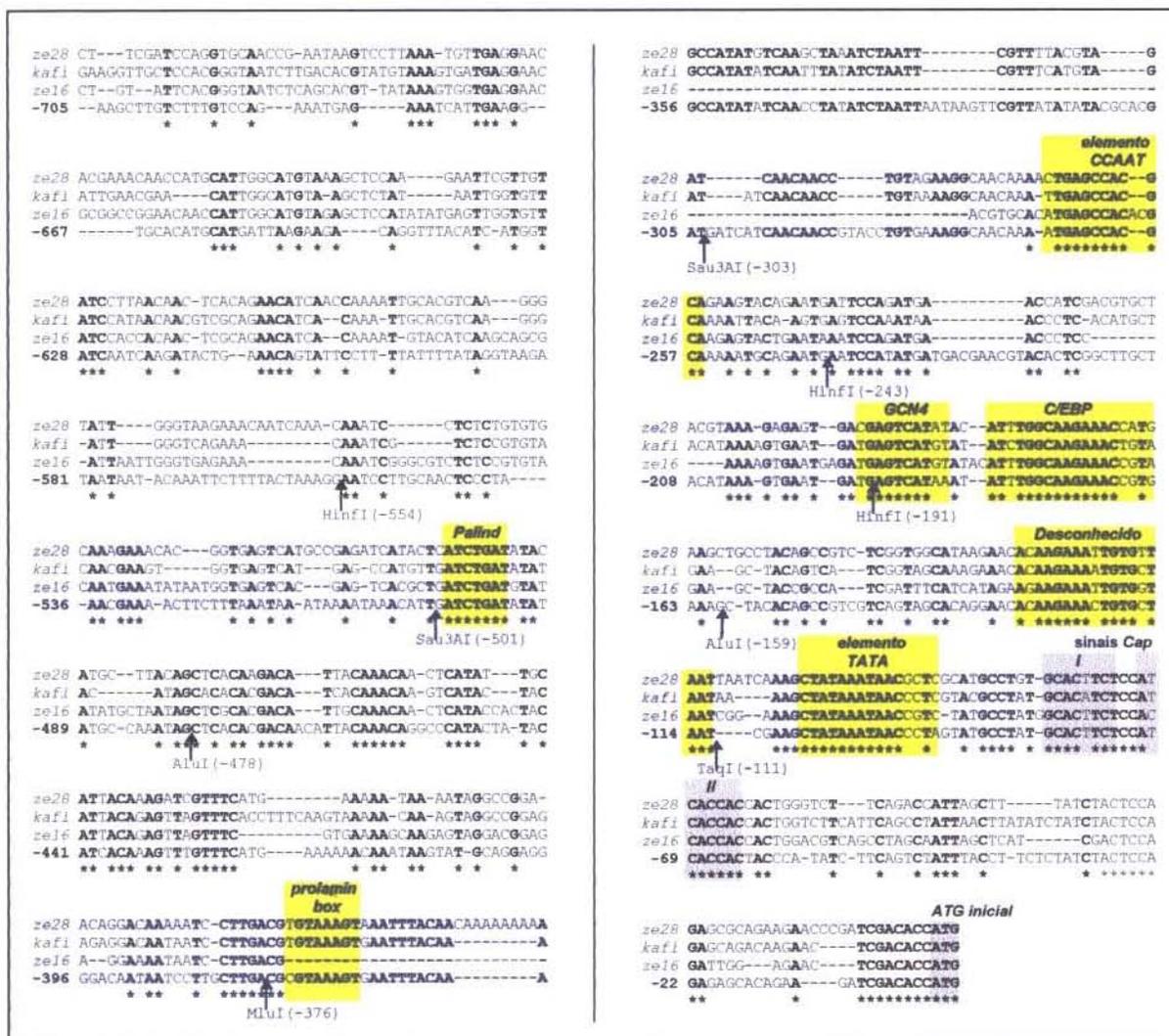


Figura 10: Alinhamento dos promotores de γ -prolamina.

Esta figura mostra o alinhamento final obtido para as regiões a montante da fase de leitura dos genes de γ -prolamina. *ze28* indica a seqüência da zeína de 28 kDa, *kaf1*, a seqüência da γ -kafirina e *ze16* a da zeína de 16 kDa. A seqüência da γ -coixina está representada na última das quatro linhas. Os números à esquerda da seqüência representam a distância, em pb, do primeiro caracter desta linha em relação ao ATG inicial. Nucleotídeos conservados nas quatro seqüências estão representados em negrito e indicados por um asterisco (*). As setas assinalam a posição de corte das enzimas de restrição indicadas. Os nomes sobre as seqüências indicam motivos conservados (destacados em fundo amarelo) descritos em **Resultados**.

Os níveis de expressão obtidos com cada uma das construções encontram-se descritos na **Figura 12**. Com os limites determinados por intervalos de confiança, foi possível distinguir quatro grupos distintos de nível de expressão. Os intervalos de confiança de duas construções, a 3.10 e a gc2, apresentam uma grande amplitude, englobando os três últimos grupos de nível de expressão.

Experimentos de retardamento em gel.

A interação de proteínas nucleares com regiões determinadas da região promotora do gene da γ -coixina foi estudada através de experimentos de retardamento em gel empregando os fragmentos esquematizados na **Figura 11**. Nesta mesma figura encontram-se sumariados os resultados de cada um destes fragmentos.

Região do fragmento 226

226

Conforme se observa na **Figura 13**, o fragmento 226 forma cinco bandas de retardamento ao ser incubado com extrato de proteínas nucleares. A banda E resiste às competições mais fortes, o que sugere que ela seja formada pela interação da sonda com proteínas abundantes no extrato nuclear. As bandas A, B, C e D não resistem à competição específica, mas resistem à inespecífica, indicando especificidade na ligação proteína-DNA.

151

A divisão do fragmento 226 com a enzima Hinf I gera os fragmentos 151 e 75 (**Figura 11**). Não é possível obter um retardamento específico com o fragmento 151.

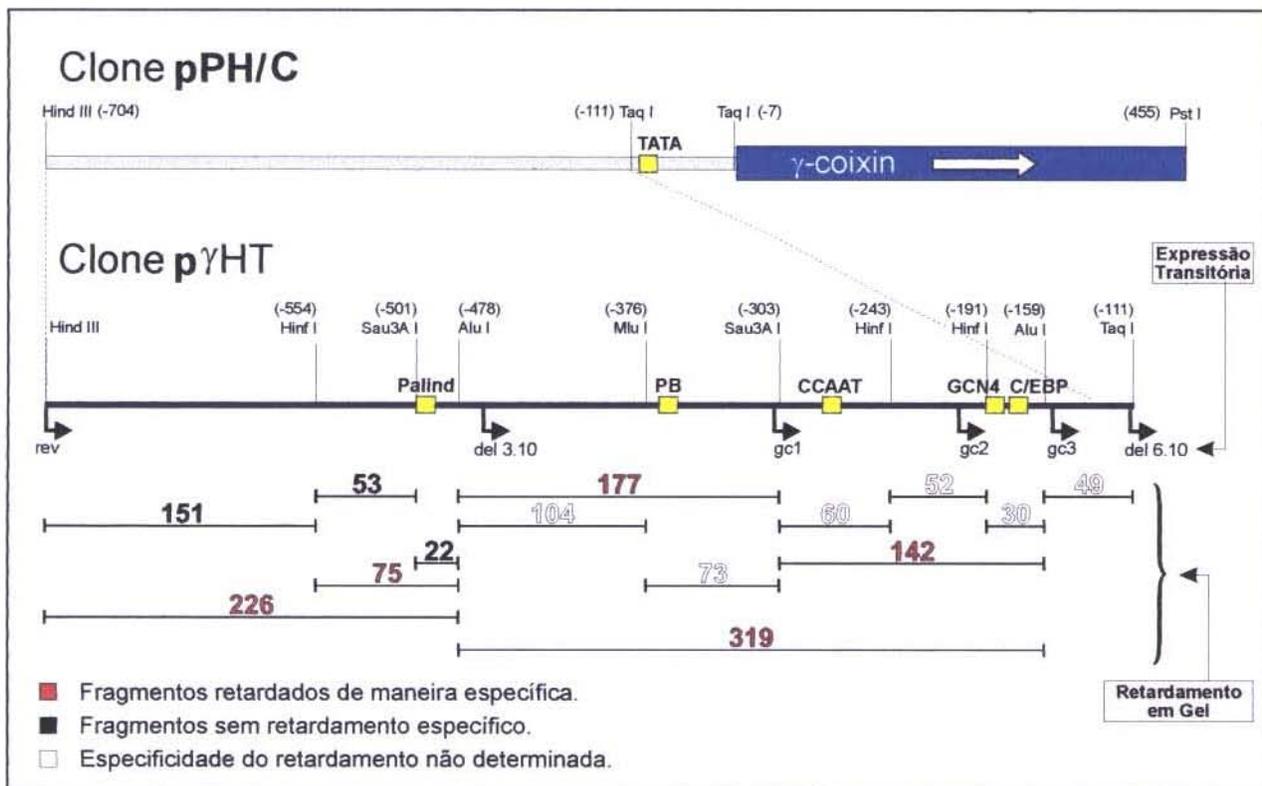


Figura 11: Fragmentos do promotor de γ -coixina empregados nos experimentos de expressão transitória e retardamento em gel.

Os números próximos às enzimas de restrição indicam a posição de corte das enzimas em relação ao ATG inicial do gene de γ -coixina. Os quadrados amarelos identificados nos 2 clones referem-se às regiões conservadas mostradas na **Figura 10**. Os fragmentos empregados nos experimentos de retardamento em gel estão identificados por seu tamanho em pares de base, na região da figura identificada por **Retardamento em Gel**. Os resultados destes experimentos estão indicados pelas cores dos nomes dos fragmentos, conforme legenda na figura. A origem dos fragmentos do clone pPH/C utilizados nos experimentos de expressão transitória está marcada pelas setas identificadas por **Expressão Transitória**.

75

O fragmento 75, como se pode ver na **Figura 14**, forma três bandas de retardamento ao ser incubado com o extrato de proteínas nucleares. A banda C' resiste à competição específica, de onde se infere que pode ser produzida por proteínas abundantes no extrato nuclear. As bandas A' e B', por sua vez, resistem à competição inespecífica, mas não à específica o que indica a especificidade da ligação entre as proteínas que as causam e o fragmento 75. As posições e resistência às competições das bandas A' e C' do fragmento 75 no gel sugerem uma relação com as bandas A e E do fragmento 226. Ainda, a posição da banda B' do fragmento 75 poderia ter relação com uma das bandas na posição ocupada pelas B, C e D do fragmento 226.

53 e 22

A divisão do fragmento 75 com a enzima de restrição Sau3A I gera os fragmentos 53 e 22 (**Figura 11**). Nenhum dos dois apresenta retardamentos resistentes às competições inespecíficas.

Região do fragmento 319**319**

Na **Figura 15** é possível verificar que o retardamento do fragmento 319 apresenta uma definição inferior aos descritos anteriormente. Este fato deve-se, provavelmente, ao seu tamanho, considerado excessivo para este tipo de experimento. Mesmo assim, é possível distinguir quatro regiões de retardamento quando se observa diversos géis com tempos de exposição diferentes. Destas quatro regiões, A, B e C apresentam resistência às competições num padrão que pode ser considerado como resultante de interações específicas entre proteínas nucleares e o fragmento.

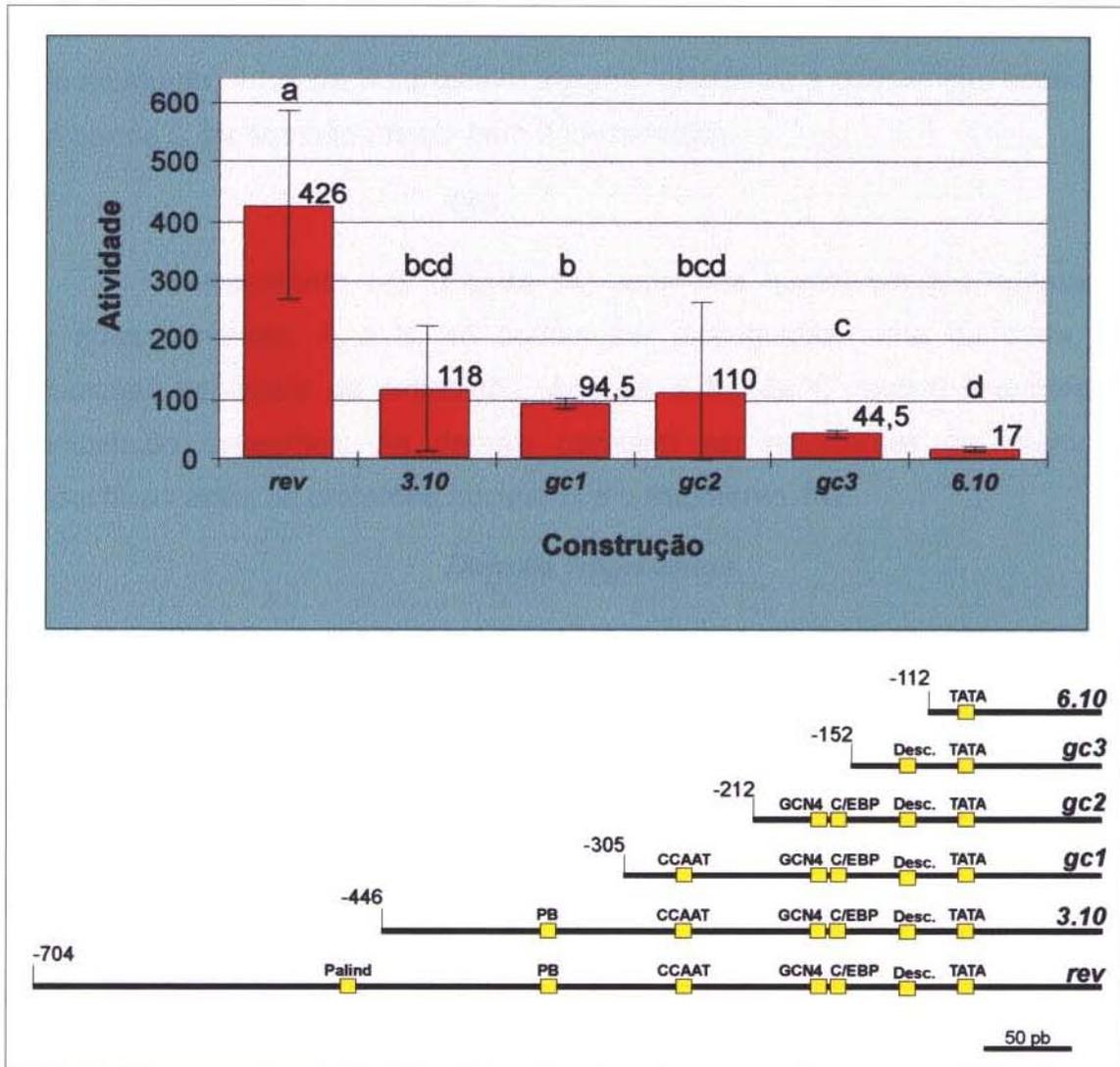


Figura 12: Resultados dos experimentos de expressão transitória.

A altura das barras vermelhas representa a média obtida para cada uma das construções e seu valor está indicado logo acima da barra. As linhas pretas acima e abaixo do topo das barras vermelhas indicam o intervalo de confiança para as médias. As letras sobre as barras indicam a sobreposição dos intervalos de confiança. Construções cujos intervalos de confiança se sobrepõem são identificadas pela mesma letra. A atividade é expressa em número de pontos azuis, conforme explicado em Materiais e Métodos. Os esquemas abaixo do gráfico ilustram a seqüência do promotor de γ -coixina que dirige a expressão do gene *gus* em cada uma das construções. O número à esquerda de cada um dos esquemas indica a posição de sua extremidade 5' relativa ao ATG inicial.

177

A digestão do fragmento 319 com a enzima de restrição Sau3A I gera os fragmentos 177 e 142. O fragmento 177, como pode ser visto na **Figura 15** apresenta 3 bandas de retardamento. A banda A, extremamente fraca e muito pouco definida, só é perceptível com tempos de exposição maiores do que o da foto mostrada. A banda B apresenta alguma resistência à competição específica e a banda C parece não resistir bem à inespecífica.

142

O fragmento 142 (**Figura 16**) apresenta quatro bandas retardadas. As duas primeiras, A e B, só podem ser distinguidas, uma da outra, em condições especiais de exposição. Apenas a banda C resiste levemente à competição específica. As demais parecem ser resultantes de interações específicas entre as proteínas nucleares e o fragmento 142.

Demais fragmentos

Não foi possível determinar a especificidade das interações nos subfragmentos oriundos dos fragmentos 177 e 142 (**Figura 11**). Os resultados obtidos em repetições dos experimentos são inconsistentes.

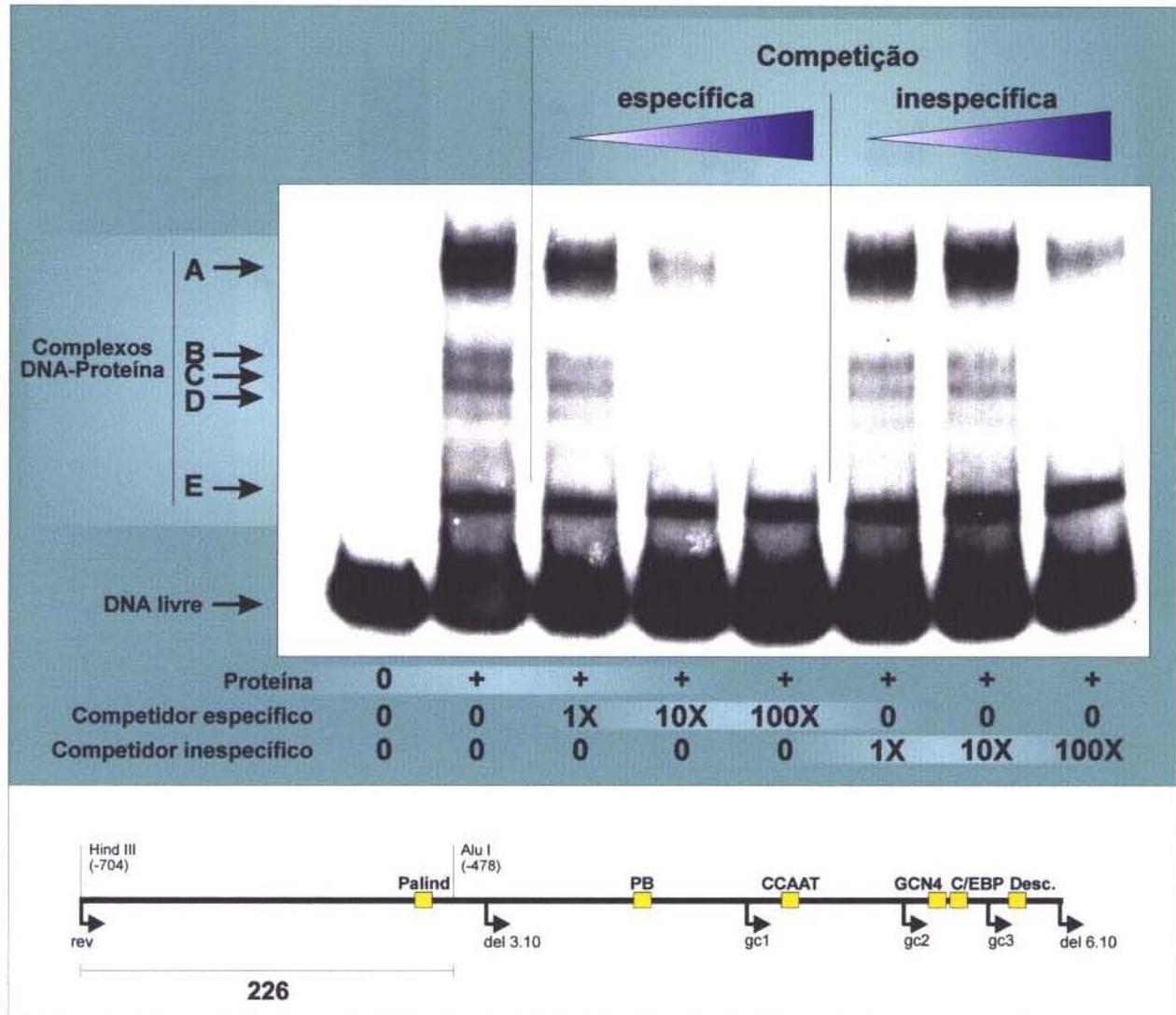


Figura 13: Experimento de Retardamento em Gel do fragmento 226.

O esquema abaixo da foto mostra a localização do fragmento 226 na região a montante do gene de γ -coixina. Os números ao lado dos nomes das enzimas de restrição indicam a posição do corte relativa ao ATG inicial. O sinal de mais (+) indica a presença de proteína na reação de ligação. O zero (0) indica a ausência do reagente. Os números 1X, 10X e 100X descrevem o excesso molar do reagente. Os triângulos azuis identificam grupos de reações com quantidades crescentes do reagente indicado. As letras identificam os complexos DNA-proteína.

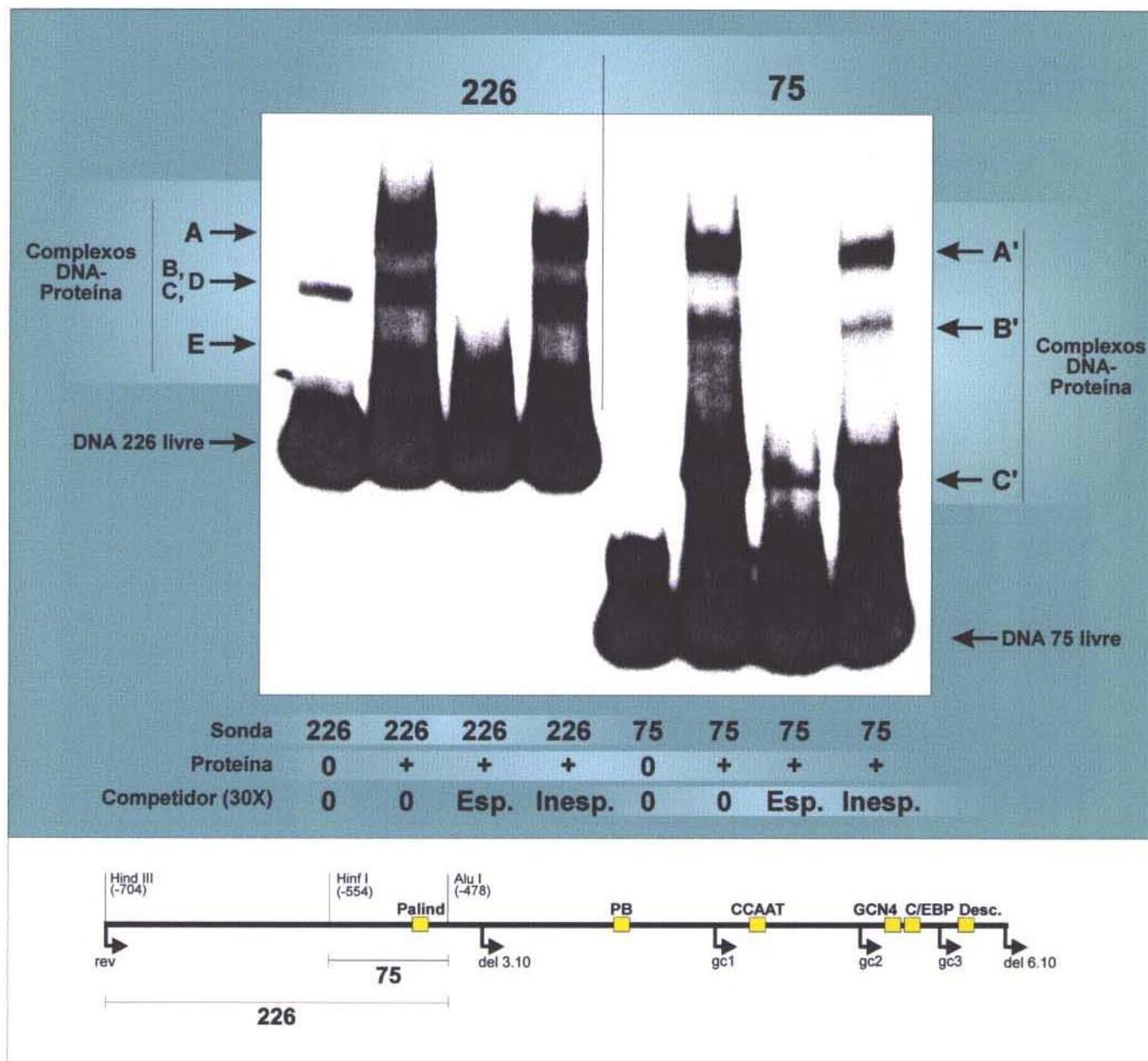


Figura 14: Experimento de Retardamento em Gel dos Fragmentos 226 e 75 Simultaneamente.

O esquema abaixo da foto mostra a localização dos fragmento 226 e 75 na região a montante do gene de γ -coixina. Os números próximos às enzimas de restrição indicam a posição de corte das enzimas em relação ao ATG inicial do gene de γ -coixina. O sinal de mais (+) indica a presença de proteína na reação de ligação. O zero (0) indica a ausência do reagente. *Esp.* indica a presença de competidor específico na reação e *Inesp.* a de inespecífico (em ambos os casos com excesso molar de 30X). As letras identificam os complexos DNA-proteína.

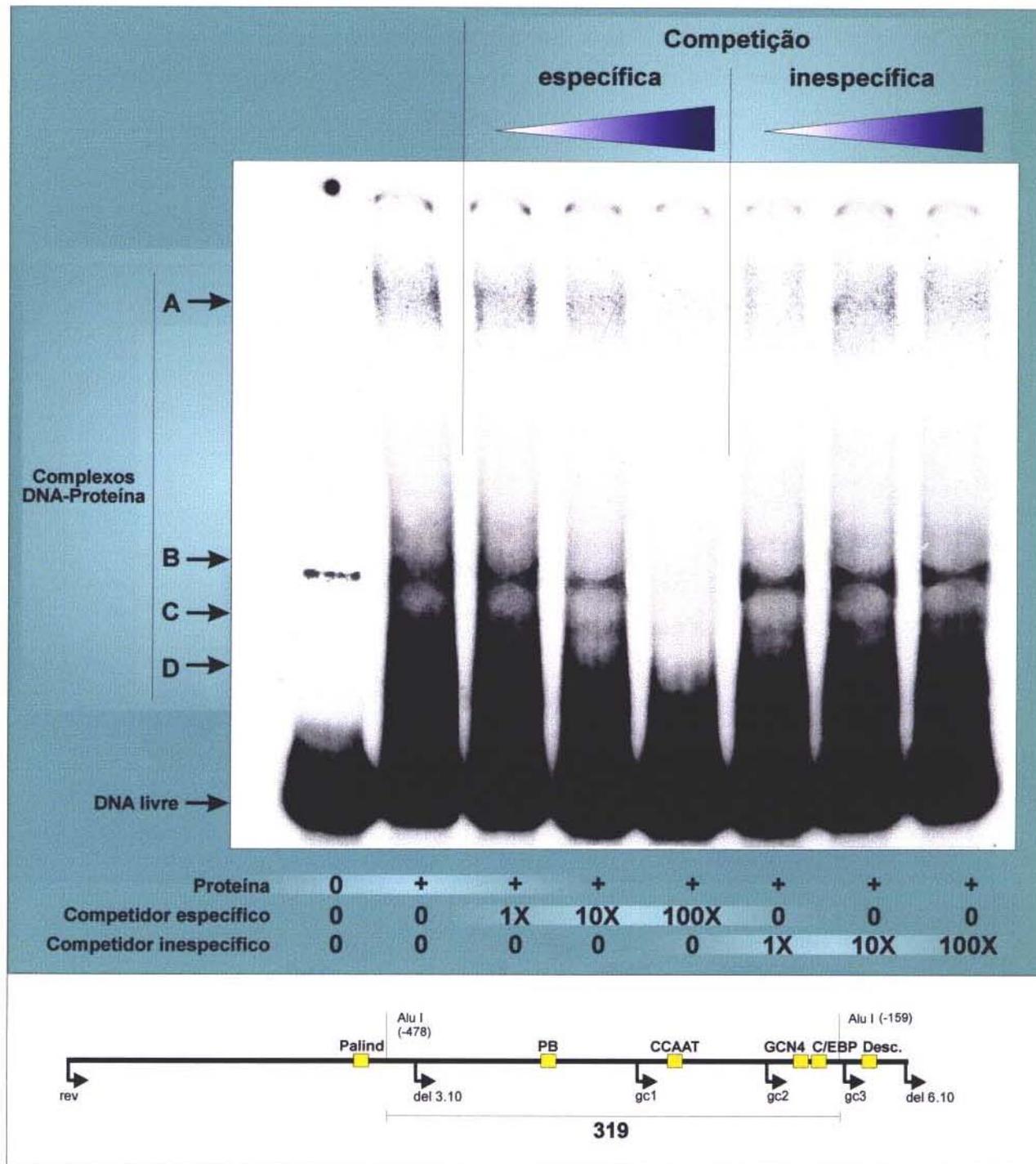


Figura 15: Experimento de Retardamento em Gel do fragmento 319.

O esquema abaixo da foto mostra a localização do fragmento 319 na região a montante do gene de γ -coixina. Os números próximos às enzimas de restrição indicam a posição de corte das enzimas em relação ao ATG inicial do gene de γ -coixina. O sinal de mais (+) indica a presença de proteína na reação de ligação. O zero (0) indica a ausência do reagente. Os números 1X, 10X e 100X descrevem o excesso molar do reagente. Os triângulos azuis identificam grupos de reações com quantidades crescentes do reagente indicado. As letras identificam os complexos DNA-proteína.

Discussão

Análise da seqüência do promotor

Estudos do promotor de um gene de *Drosophila* (Ludwig e Kreitman, 1995) revelaram que, entre duas espécies próximas, o grau de conservação de regiões envolvidas na regulação da expressão deste gene é menor que o normalmente observado para os domínios de ligação de fatores de transcrição (Joly *et al.*, 1993). Devido à presença de sítios múltiplos de ligação para quatro fatores de transcrição, os autores sugerem que a multiplicidade de sítios na região estudada fornece uma redundância funcional que permite flexibilidade na seqüência destes sítios e na estrutura do promotor (Ludwig e Kreitman, 1995).

A **Figura 9A** mostra a presença de um grande número de possíveis sítios envolvidos na regulação da expressão do gene da γ -coixina, que poderiam corresponder às redundâncias observadas no promotor de *Drosophila*. Interessante também perceber na **Figura 9B** (barras em azul) que o número de possíveis sítios é maior na região distal do promotor do que nas regiões próximas ao sítio de início da transcrição.

Uma possível explicação para este fato seria que o surgimento de um sítio para um fator qualquer nas regiões envolvidas na regulação de genes pode gerar efeitos regulatórios indesejáveis. A presença do sítio para um fator de transcrição, ativo em vários tecidos, no promotor de um gene cuja expressão é limitada a um determinado tecido poderia levar à expressão deste gene em tecidos onde seu efeito fosse prejudicial ao organismo, ou eliminar a expressão em tecidos onde o mesmo é necessário. Nesta situação os organismos cujo genoma contivesse este sítio deslocado sofreriam seleção. O surgimento do mesmo sítio em uma região não envolvida na regulação de genes teria menor probabilidade de causar alterações adversas ao organismo, mesmo porque tais regiões costumam ser menos acessíveis aos fatores de transcrição.

A identificação de elementos conservados não significa que estes elementos serão reconhecidos pelas mesmas proteínas em todos os promotores. Não se pode também dizer que um determinado gene será ativado por um fator particular apenas baseado na presença do sítio reconhecido por este fator no promotor. O contexto de promotor é importante, como fica claro no caso do elemento em *cis octamer*, reconhecido por Oct-1 em uma série de genes e por Oct-2 no gene da cadeia κ leve da imunoglobulina onde o reconhecimento por Oct-1 não ocorre (Lewin, 1994).

A despeito destas considerações, a identificação de possíveis sítios regulatórios é importante para o planejamento dos experimentos. A procura de sítios regulatórios potenciais no promotor de γ -coixina foi realizada utilizando-se o programa MatInspector descrito por Quandt *et al.* (1990). Por empregar matrizes ponderadas das seqüências consenso, este programa apresenta maior fidelidade ao universo de informações já disponíveis, quando comparado com métodos baseados simplesmente na coincidência de seqüências. Esta busca foi realizada também nos genes homólogos de milho e sorgo. O cruzamento destes resultados, aliado com o alinhamento das seqüências, restringiu consideravelmente o número de possíveis sítios envolvidos na regulação destes genes, outrossim aumentando a significância dos resultados obtidos.

Elemento TATA

Os elementos TATA dos genes da γ -kafirina e das γ -zeínas que encontram-se alinhados na **Figura 10** com o possível elemento TATA do gene da γ -coixina correspondem aos motivos previamente descritos (Reina *et al.*, 1990; Reina *et al.*, 1990; Freitas *et al.*, 1994).

A conservação das TBPs, e das seqüências por elas reconhecidas, entre animais, leveduras e plantas é muito alta. Isto pode ser reflexo das dificuldades de modificação do modo ancestral de reconhecimento de promotor (Mukumoto *et al.*, 1993), e permite que conclusões obtidas de um modelo possam ser extrapoladas para organismos de grupos distintos.

Partindo da seqüência TATATATA, Mukumoto *et al.* (1993) testaram as 21 mutações pontuais possíveis nas primeiras sete posições e concluíram que, além de seqüência original, os mutantes TATATAAA e TATAAATA são reconhecidos eficientemente pelas TBPs humana e de *Arabidopsis*. Este último elemento (TATAAATA) é idêntico ao encontrado conservado nos promotores de γ -prolaminas (**Figura 10**) e é o que apresenta atividade transcricional mais alta entre os analisados por Mukumoto *et al.* (1993). Este mesmo elemento pode ser ainda encontrado em outros 17 promotores funcionais de plantas (Joshi, 1987).

Bucher (1990), após analisar 502 promotores eucariotos, concluiu que o elemento TATA localiza-se normalmente entre 20 e 36 pb a montante do sinal *cap*. Dois possíveis sinais *cap* foram encontrados no promotor de γ -coixina. Um localizado a 24 pb do elemento TATA e o outro a 35 pb (**Figura 10**).

O segundo *cap* é idêntico ao encontrado no promotor do gene *cab* AB80 de ervilha (*chlorophyll a/b-binding polypeptide*), no qual a funcionalidade do sinal *cap* e do elemento TATA já foi determinada (Cashmore, 1984). Mais interessante, a distância entre o elemento TATA e o sinal *cap* difere em apenas 3 pb nos genes da γ -coixina e *cab* AB80 e as seqüências dos elemento TATA, assim como as distâncias entre o *cap* e o ATG inicial (68 pb), são também idênticas nestes dois genes (Joshi, 1987).

Elemento CCAAT

O valor obtido para similaridade com a matriz do elemento CCAAT nos promotores analisados nos resultados é baixo (**Figura 9A**). Métodos tradicionais de procura de seqüências consenso (código IUPAC, e.g.) poderiam não tê-lo encontrado. A necessidade de um elemento CCAAT nos promotores eucariotos não está clara conforme discutido na revisão bibliográfica. Se as seqüências nos promotores de γ -prolaminas descritas nesta tese são funcionalmente semelhantes ao elemento CCAAT de outros promotores de plantas depende do aprofundamento do trabalho, assim como da determinação de seu papel nos demais promotores onde o sítio foi descrito.

O elemento CATC é uma seqüência semelhante ao elemento CCAAT encontrado no alinhamento dos primeiros genes de prolaminas descritos. Esta seqüência foi descrita em vários genes de cereais, mas não foi encontrada em genes de plantas dicotiledôneas (Forde *et al.*,1985). À época do trabalho, entretanto, o autor dispunha de cerca de 15 seqüências de genes de plantas dicotiledôneas incluindo a região promotora, e um número próximo a este de genes de cereais, a maioria de proteínas de reserva. Hoje, as seqüências de mais de 10.000 genes de plantas (2.600 de cereais) contendo a região promotora encontram-se disponíveis nos bancos de dados.

Desde a descrição do elemento CATC, poucos autores (Kreis *et al.*,1986; Rodriguez-Palenzuela *et al.*,1988; Huang *et al.*,1990; Qu e Huang, 1990; Ottoboni *et al.*,1993; Freitas *et al.*,1994) descreveram seqüências similares em promotores de genes de plantas. Foi mesmo cogitado que o elemento CATC seria, em cereais, o equivalente ao elemento CCAAT. Nenhum destes autores, entretanto, levantou evidências experimentais da função do elemento CATC.

Há cerca de dois anos, Freitas *et al.* (1994) descreveram um possível elemento CATC nos promotores da γ -kafirina e das γ -zeínas. Os resultados apresentados nesta tese revelam que este elemento CATC, que Freitas *et al.* já mostravam não estar conservado no promotor da γ -zeína de 28 kDa, não existe também no promotor da γ -coixina. Os autores não mencionam a existência de um elemento CCAAT nestes promotores.

A análise das regiões promotoras dos genes de γ -prolaminas com o programa MatInspector (Quandt *et al.*,1995), empregando a matriz construída com o auxílio do programa MatInd a partir dos dados empregados originalmente na descrição do elemento CATC (Forde *et al.*,1985), mostrou que esta seqüência não pode ser encontrada conservada em posição nas seqüências analisadas.

Sítios de GCN4 e C/EBP

A síntese de prolaminas é modulada em função da disponibilidade de nitrogênio, tanto na forma de NH_4NO_3 , quanto na forma de aminoácidos. Esta modulação se dá ao nível transcricional. Uma prolamina de cevada apresenta envolvimento de motivos similares ao sítio de GCN4 na resposta ao nitrogênio (Muller e Knudsen, 1993). A proteína GCN4 é uma bZIP de leveduras que modula a transcrição de uma série de genes em função dos níveis de aminoácidos (Hinnebusch, 1984).

Seqüências similares ao sítio de GCN4 já foram encontradas em 31 genes específicos de endosperma e em outros 16 genes de planta. Sua funcionalidade como elemento *cis* já foi demonstrada em, pelo menos, 4 destes genes, onde parece estar sempre relacionado com a resposta aos níveis de nitrogênio (Lohmer *et al.*, 1991; Bäumllein *et al.*, 1992; Muller e Knudsen, 1993).

Nos promotores dos genes das prolaminas de cevada, trigo e centeio e nos das α -prolaminas de 22 kDa de milho motivos similares ao sítio de GCN4 são encontrados oito nucleotídeos a jusante do *prolamin box*. Alguns autores sugerem o nome de *endosperm box* para o conjunto dos dois sítios (cerca de 25 pb) (Muller *et al.*, 1995). Já foi mostrado, em trigo, que fatores transatuantes ainda não isolados ligam-se aos dois motivos: o ESBF-I (*endosperm box factor I*) liga-se ao *prolamin box* e o ESBF-II ao motivo semelhante ao sítio de GCN4. Estes fatores protéicos encontram-se ligados às seqüências do *endosperm box* apenas no endosperma, e não nos demais tecidos de trigo, o que indica a tecido-especificidade de sua atividade (Hammond-Kosack *et al.*, 1993).

Recentemente foi demonstrado que o fator Opaco-2 de milho é capaz de ativar um promotor de glutelina ao ligar-se de maneira específica a um sítio similar ao de GCN4 (Holdsworth *et al.*, 1995). Os autores sugerem que o fator ESBF-II de trigo pode conter um componente relacionado à proteína Opaco-2, e que a diversidade de sítios de ligação descritos para Opaco-2 pode refletir sua interação com diferentes heterodímeros. O fato deste fator ser encontrado ligando-se a um elemento envolvido na resposta aos níveis de nitrogênio, assim

como sua semelhança com o fator GCN4, uma bZIP de leveduras ligada a esta resposta, permitem especular que Opaco-2 e/ou seus heterodímeros seriam os mediadores da resposta à disponibilidade de nitrogênio nas gramíneas. Exceto pela β -zeína, as prolaminas de milho também têm sua proporção na semente modulada pela disponibilidade de nitrogênio (Singletary *et al.*, 1990).

O envolvimento de Opaco-2 com a regulação das γ -prolaminas não foi demonstrado até hoje. A presença de sítios semelhantes aos reconhecidos por bZIPs nos promotores das γ -prolaminas, como C/EBP e GCN4 (**Figura 9**), e a demonstração recente de que a proteína Opaco-2 reconhece um sítio similar ao de GCN4 em trigo (Holdsworth *et al.*, 1995) levantam a possibilidade deste fator, ou uma bZIP semelhante a ele, estar relacionado com a regulação da expressão das γ -prolaminas.

Se a presença de um sítio similar ao reconhecido pelo fator GCN4 de leveduras em um promotor de plantas é relativamente comum, o mesmo não pode ser dito da presença de um sítio similar ao reconhecido pelas C/EBPs.

As C/EBPs constituem uma família de bZIPs envolvidas em uma grande variedade de funções regulatórias da expressão gênica de vertebrados. Entre estas funções encontram-se a ativação de genes constitutivos no fígado, o controle da diferenciação de adipócitos e a regulação de genes em monócitos (Johnson e Williams, 1994).

Neste trabalho, nenhuma evidência direta do envolvimento do sítio semelhante ao de C/EBPs na regulação das γ -prolaminas foi obtida. Um fragmento de 142 pb que contém o sítio similar a C/EBP interage especificamente com proteínas nucleares (**Figura 16**). A deleção de um fragmento de 153 pb contendo este sítio causa uma queda de cerca de 50% no nível de expressão do gene indicador (**Figura 12**).

Entretanto, a extensão dos fragmentos testados nestes experimentos, assim como a presença de outros possíveis sítios de ligação de

fatores de transcrição, não permite delimitar os sítios envolvidos nas interações com proteínas nucleares e/ou aqueles ligados à modulação da expressão.

Até hoje não foi encontrada em plantas uma bZIP semelhante a C/EBPs. As seqüências reconhecidas pelas bZIPs de plantas são extremamente divergentes daquelas às quais as C/EBPs se ligam. As C/EBPs conhecidas podem formar heterodímeros em todas as combinações entre si (Olive *et al.*, 1996). Sua região do zíper de leucinas, contudo, é pouco semelhante a encontrada nas bZIPs de planta já descritas (Pabo e Sauer, 1992), o que torna pouco provável a formação de heterodímeros entre bZIPs já descritas de plantas e proteínas semelhantes a C/EBPs.

Foi demonstrado em nosso laboratório (Yunes *et al.*, em preparação) que dois sítios de ligação de bZIPs próximos podem facilitar a ligação destes fatores. Os fatores que se ligam em cada um dos sítios interagiriam e aumentariam a estabilidade da ligação. Este mecanismo de cooperatividade de ligação é comum entre os elementos que reconhecem sítios similares ao *ocs* (*octopine synthase*) (Singh *et al.*, 1990; Tokuhisa *et al.*, 1990; Ulmasov *et al.*, 1994). A presença de dois sítios conservados de ligação de bZIPs, GCN4 e C/EBP, próximos, como os descritos nos resultados (**Figura 10**) pode indicar que interações como esta ocorram também nos promotores de γ -prolaminas. Se tal hipótese estiver correta, o(s) fator(es) que se liga(m) ao sítio semelhante ao reconhecido pelas C/EBPs de vertebrados seria, pelo menos no tocante a seqüência de reconhecimento de DNA, até agora desconhecido(s) em plantas.

Prolamin box

A diferença observada no nosso trabalho entre a seqüência na região do *prolamin box* das demais gramíneas (TGTAAG) e da de *Coix* (CTGAAAG) (**Figura 10**) poderia explicar a discrepância observada entre *Coix* (Targon *et al.*, 1992) e milho (Lending e Larkins, 1989) nos padrões de expressão e localização das diferentes classes de prolaminas.

Ao contrário do que acontece com as α -prolaminas, não se observa nas γ -prolaminas um sítio similar ao de GCN4 nas proximidades da região do *prolamin box*. Este fato levanta a possibilidade de que o envolvimento do *prolamin box* na regulação das γ -prolaminas de milho, sorgo e *Coix* ocorra por mecanismos distintos aos observados nas α -prolaminas, o que seria coerente com o fato de estas duas classes de prolaminas apresentarem padrões de expressão temporal e espacial distintos.

Seqüências palindrômicas (*Palind*)

Uma série de autores descreve o envolvimento de elementos palindrômicos na regulação mediada por alguns promotores de genes de plantas (Liaud *et al.*, 1993; Nielsen e Gausing, 1993; Shiel *et al.*, 1995; Inada *et al.*, 1996; Weaver *et al.*, 1996). As possíveis proteínas que se ligam a estes elementos não foram ainda caracterizadas. Uma série de classes de fatores de transcrição são capazes de reconhecer seqüências palindrômicas. Entre as classes que contém tais fatores, já estão bem caracterizadas em plantas as bZIPs (Izawa *et al.*, 1993; Meier e Gruissem, 1994; Niu *et al.*, 1994), as bHLH (Consonni *et al.*, 1993; Purugganan e Wessler, 1994) e as *zinc fingers* (Takatsuji *et al.*, 1992; Speulman e Salamini, 1995).

O envolvimento de seqüências palindrômicas não descritas anteriormente na regulação de genes eucariotos acontece com certa freqüência (Antoniewski *et al.*, 1996; Wey e Schafer, 1996). Em alguns casos o isolamento do fator que reconhece o palíndromo envolvido na regulação revela que este fator pertence a uma classe já descrita de fatores de transcrição. Tal é o caso no estudo dos elementos envolvidos na regulação do promotor do gene da cutinase de fungos. Foi demonstrado que um palíndromo era necessário para a indução do gene pelo monômero de cutina (Kamper *et al.*, 1994). Mais tarde, o fator que medeia esta resposta foi clonado e identificado como uma nova proteína do tipo *zinc finger* (Li e Kolattukudy, 1995).

A seqüência palindrômica descrita em nossos resultados (ATCTGAT) encontra-se conservada nos quatro promotores analisados (**Figura 10**). Esta seqüência não é reconhecida por qualquer dos fatores descritos até o momento e nem tampouco foi descrito seu envolvimento nos processos regulatórios de expressão gênica. Entretanto, conforme será discutido, a deleção de um fragmento incluindo este palíndromo provoca uma queda significativa nos níveis de expressão do gene indicador (**Figura 12**). Além disso, fragmentos bastante delimitados que o contém interagem especificamente com fatores nucleares (**Figura 14**).

Sinal de poliadenilação

O processamento da região terminal 3' da molécula de mRNA pode exercer um papel importante na regulação da expressão gênica. Ele estimula o transporte do mRNA do núcleo para o citoplasma (Eckner *et al.*, 1991) e pode influenciar a estabilidade (Birnstiel *et al.*, 1985) e tradutibilidade desta molécula (Jackson e Standart, 1990). Ao contrário do que acontece em mamíferos (Manley e Proudfoot, 1994), os mecanismos envolvidos no processamento da região terminal 3' da molécula de mRNA nuclear de plantas não estão bem definidos.

Sabe-se que para que este processamento ocorra corretamente é necessário um sinal de poliadenilação semelhante ao AATAAA de mamíferos assim como uma seqüência rica em TG a montante do sinal de poliadenilação (Wu *et al.*, 1995).

O possível sinal de poliadenilação (AATGAA) descrito para o gene da γ -coixina (**Figura 8**) é idêntico ao descrito para o gene da γ -zeína de 28 kDa e encontra-se à mesma distância do sítio de poliadenilação (Wu *et al.*, 1993).

Na zeína de 28 kDa, duas outras seqüências são essenciais para o correto processamento da região terminal a jusante da molécula de mRNA (Wu *et al.*, 1993): uma seqüência "rica em TG" (CATGGGTGG) encontrada 38 pb a montante do sinal de poliadenilação e uma seqüência GTAATAATG localizada

13 pb a montante da rica em TG. Seqüências similares podem ser vistas nesta região do gene de γ -coixina. A seqüência CATGGGTGA, semelhante àquela rica em TG, encontra-se 40 pb a montante do sinal de poliadenilação e a seqüência GTAATAAAAG localiza-se 13 pb a montante da rica em TG.

Experimentos de expressão transitória

A grande variação dos resultados obtidos para algumas construções (rev, 3.10 e gc2) parece, à primeira vista, muito discrepante dos resultados mais definidos das construções gc1, gc3 e 6.10 (**Figura 12**). A variação observada para as construções 3.10 e gc2 chega mesmo a obrigar a separação dos resultados em dois grupos: um, o grupo **a**, formado pelos experimentos que empregam o fragmento rev e outro, grupo **bcd**, formado pelos demais experimentos. Esta divisão implica na conclusão de que, para o nível de resolução dos experimentos realizados, a única região na qual se consegue detectar variação na atividade do promotor corresponde àquela que inclui os 258 pb removidos na construção 3.10.

Entretanto, como os experimentos são independentes, a não realização dos testes com as construções 3.10 e gc2 não causaria alteração nos resultados obtidos para as demais construções, e os resultados estariam agrupados em quatro classes distintas, **a**, **b**, **c** e **d**.

Experimentos de deleção como os realizados neste trabalho apresentam o inconveniente de não se conhecer de antemão os elementos que estão sendo eliminados em cada construção. Uma grande extensão do promotor pode ser eliminada e não acarretar mudança significativa nos níveis de expressão do gene indicador sob seu controle. Entretanto, não se pode deste resultado concluir a inexistência de elementos envolvidos na regulação. Poderiam existir nesta região dois elementos, um reprimindo e transcrição e outro eliminando esta repressão, cuja eliminação simultânea mantivesse inalterados os níveis de expressão, por exemplo. Assim, estes experimentos

normalmente não são conclusivos, mas geram indícios importantes sobre os quais outros experimentos podem ser realizados.

rev e 3.10

Os experimentos de expressão transitória demonstraram a capacidade da região a montante da seqüência codificadora do gene de γ -coixina dirigir a expressão de um gene indicador em endosperma de milho. A diferença nos níveis de expressão entre as construções rev e 3.10 (**Figura 12**) deixa clara a necessidade da presença da região a montante da posição -446 do ATG inicial para a expressão máxima do gene indicador. A grande extensão desta deleção (258 pb), entretanto, impede a delimitação do(s) fator(es) *cis* envolvido(s) nesta diferença de resposta. O único elemento conservado nesta região nos promotores estudados corresponde ao elemento *Palind* (**Figura 10**).

gc1

Os experimentos de expressão transitória não detectaram diferenças significativas nos níveis de expressão com a remoção de um fragmento de 141 pb contendo o *prolamin box* (construções 3.10 e gc1). Este resultado diverge de trabalhos anteriores (Matzke *et al.*, 1990; Thompson *et al.*, 1990).

Estes trabalhos, entretanto, são também divergentes entre si. Ao passo que Thompson *et al.* (1990) demonstram que a região do *prolamin box* aumenta a atividade de um gene sob seu controle, e que o *prolamin box* propriamente dito não tem muita influência nesta atividade, Matzke *et al.* (1990) demonstram que a remoção da região contendo o *prolamin box* determina um aumento da atividade do promotor em relação à observada em promotores contendo regiões mais extensas de seqüências a montante dos genes de prolaminas. Nos dois casos, entretanto, os autores testaram a atividade do promotor em sistemas heterólogos: Thompson *et al.* em protoplastos de cenoura e Matzke *et al.* em plantas transgênicas de tabaco, ambas dicotiledôneas.

Em sistemas homólogos (protoplastos de endosperma de milho), regiões contendo o *prolamin box* aumentam a atividade dos promotores (Quayle e Feix, 1992; Ueda *et al.*, 1994). A ligação específica de fatores nucleares a esta região foi também demonstrada nestes trabalhos. Nos dois casos, entretanto, os testes foram realizados através de fusões da região do *prolamin box* com o promotor do CaMV 35S (*i.e.*, não se tratava do promotor no qual o elemento é encontrado naturalmente) e em um sistema que pode não refletir com exatidão o ambiente encontrado no endosperma de milho (protoplastos oriundos de endosperma).

Deve-se destacar, contudo, que a incapacidade de detectar diferenças no nível de expressão, em nossos experimentos de expressão transitória, pode estar relacionada com a grande dispersão dos resultados obtidos nas repetições dos experimentos com o construção 3.10. Os valores médios de expressão obtidos para as construções 3.10 e gc1 são próximos. Contudo, o intervalo de confiança para a média da construção 3.10 é bastante extenso. Assim, ainda que as médias dos valores obtidos com as duas construções sejam semelhantes, o verdadeiro valor da expressão dirigida pelo fragmento 3.10 pode ser até 1,5 vezes maior ou menor que a média obtida nos experimentos com a construção gc1 (**Figura 12**).

Demais deleções

Assim como acontece com a região contendo o *prolamin box*, a deleção de 93 pb incluindo o elemento CCAAT parece não ter efeito significativo nos níveis de expressão do gene indicador (**Figura 12**). Já havia sido demonstrado que a presença deste elemento em algumas construções não é suficiente para o correto funcionamento do promotor (Matzke *et al.*, 1990). O resultado obtido para similaridade com matriz do elemento CCAAT foi baixo, conforme discutido na análise da seqüência do promotor. O conjunto dos resultados poderia indicar que este elemento não é importante na regulação das γ -prolaminas.

A deleção de mais 60 pb contendo os possíveis sítios de ligação de GCN4 e C/EBP (construção gc3, **Figura 12**) reduz o nível de atividade do promotor pela metade. Conforme discutido anteriormente, este é um dos indícios de que fatores envolvidos na regulação da expressão das γ -prolaminas podem estar reconhecendo estes dois sítios.

É importante ressaltar que a afirmação possível a partir dos resultados aqui apresentados é que a eliminação dos 153 pb da região 5' do fragmento gc1, gerando o fragmento gc3, acarreta a redução em cerca de 50% na atividade do promotor. Desta forma, os elementos *cis* envolvidos nesta resposta podem estar localizados em qualquer posição da seqüência eliminada, o que inclui novamente o elemento CCAAT, assim como os possíveis sítios de ligação de GCN4 e C/EBP ou ainda outras seqüências contidas neste fragmento.

A remoção de outros 40 pb, gerando a construção formada pelos 112 pb a jusante do promotor da γ -coixina (construção 6.10, **Figura 12**), causa nova redução, cerca de três vezes, na atividade do gene indicador. Isto sugere o envolvimento de seqüências incluídas nesta região também na regulação dos genes de γ -prolaminas.

Experimentos de retardamento em gel

A interação de proteínas presentes no extrato nuclear de *Coix* com regiões delimitadas do promotor da γ -coixina foi comprovada através de experimentos de retardamento em gel.

A presença de mais de uma banda de retardamento neste tipo de experimento pode ser vista com certa freqüência. Em alguns casos sugere uma série de fatores (proteínas ou complexos protéicos) ligando-se a regiões distintas do fragmento, ou até mesmo a regiões sobrepostas (Carré e Kay, 1995).

A composição das bandas observadas em nossos experimentos não foi determinada. Elas podem ser formadas por ligações sucessivas de

monômeros de um peptídeo, pela interação do fator já ligado ao DNA com outras proteínas, ou ainda pela ligação de vários fatores distintos, reconhecendo um ou mais sítios no fragmento analisado. Não existe, nos experimentos descritos nesta tese, evidências fortalecendo qualquer das possibilidades.

Região do fragmento 226

A região com os resultados mais claros nos experimentos de retardamento em gel é a delimitada pelo fragmento 226 (**Figura 11**). Nesta região, que compreende 226 pb (de -476 a -704 do ATG inicial), um ou mais polipeptídeos ligam-se ao DNA e produzem, pelo menos, cinco bandas de retardamento, das quais quatro parecem ser específicas (**Figura 13**).

As tentativas de delimitação da região de interação com as proteínas neste fragmento, através dos experimentos com os fragmentos 151 e 75 (**Figura 13**) mostraram que o primeiro (151) não é especificamente retardado e o segundo (75) apresenta pelo menos três bandas de retardamento, das quais duas específicas (**Figura 14**). Três hipóteses podem explicar estes resultados.

Na primeira, as três bandas observadas no fragmento 75 seriam geradas pelo(s) mesmo(s) fator(es) que gera(m) as bandas do fragmento 226. Nenhum fator interagiria com o fragmento 151. O diferente número de bandas dever-se-ia às diferentes conformações estruturais assumidas pelos dois fragmentos de DNA quando da ligação com o(s) fator(es) protéico(s) (Lane *et al.*, 1992).

Uma segunda possibilidade seria que um ou mais fatores que se ligam ao fragmento 226 teriam os seus sítios de ligação na região compreendida pelo fragmento 151. A ligação deste(s) fator(es), entretanto, estaria condicionada à ligação de outro(s) fator(es) na região delimitada pelo fragmento 75.

Existiria ainda a terceira alternativa de o sítio responsável por duas das cinco bandas estar localizado na região de corte do fragmento 226 (-554 em relação ao ATG inicial, **Figura 11**). Como nenhum dos dois fragmentos teria

o sítio inteiro, as interações deixariam de existir. Alternativamente, o sítio poderia estar muito próximo da extremidade dos fragmentos gerados (próximo da posição -554, portanto) e, desta forma, assumir uma conformação estrutural diferente da que apresenta quando no interior de uma fita, deixando de ser reconhecido pelo(s) fator(es).

A divisão do fragmento 75 em dois fragmentos menores, de 53 pb e 22 pb, elimina a ligação de fatores nucleares a esta região do DNA. Isto pode ocorrer porque o sítio completo para as interações deixou de existir (o corte na posição -501 do ATG inicial teria dividido o sítio em dois), ou porque, por estar próximo demais de uma das extremidades da dupla fita de DNA, sua conformação já não seja mais reconhecida.

Assim, é possível constatar que o fragmento 75 é o menor a apresentar interações específicas com proteínas nucleares na região que se estende de -478 pb até -704 pb do ATG inicial.

A seqüência eliminada entre a construção rev e a construção 3.10 nos experimentos de expressão transitória compreende todo o fragmento 226 e mais 32 pb representados no fragmento 319 (**Figura 11**). Imagina-se que a diferença do nível de expressão observada entre as duas construções (**Figura 12**) seja devida à interação de fatores de transcrição com a região eliminada.

Nesta região, demonstrou-se que interações específicas ocorrem na porção delimitada pelo fragmento 75 (-478 a -554 pb do ATG inicial) e, provavelmente, próximas da posição -501. Na região do fragmento 75, a única seqüência conservada entre os quatro promotores de γ -prolamina alinhados é a descrita como **Palind**, que está localizada entre -500 e -494 pb do ATG inicial (**Figura 10**).

Já foi demonstrado que seqüências a montante do *prolamin box* são necessárias para expressão superior à basal de um gene indicador sob controle do promotor do gene da C-hordeína (prolamina de cevada) (Muller e Knudsen, 1993). Apesar dos indícios apresentados aqui, o envolvimento do elemento

Palind na formação das bandas de retardamento, assim como sua relação com a diferença nos níveis de expressão entre as construções rev e 3.10 depende ainda de confirmação.

Região do fragmento 319

A presença de, pelo menos, 2 bandas de retardamento específicas, mostra que na região do fragmento 319 também existem sítios interagindo especificamente com proteínas nucleares (**Figura 15**).

Na região compreendida pelo fragmento 319, a porção mais a montante do sítio de início de transcrição, representada pelo fragmento 177, apresenta um comportamento estranho com relação às interações nucleares (**Figura 16**). A única banda que resiste à competição inespecífica, a banda **B**, resiste também à específica, o que sugere que seja oriunda de interações com fatores abundantes no extrato nuclear. As duas outras bandas não resistem bem às competições inespecíficas.

A porção mais a jusante do fragmento 319, a representada pelo fragmento 142, apresenta 3 bandas de retardamento aparentemente oriundas de interações específicas com proteínas nucleares (**Figura 16**). As tentativas de delimitar as regiões envolvidas com estas interações, através de experimentos de retardamento com os subfragmentos 60, 52 e 30 (**Figura 11**) fracassaram.

A diferença entre a seqüência representada por este fragmento e o eliminado entre as construções gc1 e gc3 nos experimentos de expressão transiente é muito pequena (o fragmento 142 é menor 2 pb na região a montante e 7 pb na região a jusante). A deleção desta parte do promotor reduz pela metade a expressão do gene indicador. Novamente, é possível imaginar que o fator, ou fatores, que estão se ligando a esta região (**Figura 16**) sejam os responsáveis pela diferença nos níveis de expressão.

Conclusões

O clone genômico da γ -coixina (λ 23.1) foi isolado e 2.225 pb contíguos foram seqüenciados. A seqüência inclui toda a região estrutural e 705 pb da seqüência a montante da região estrutural, a região promotora.

A comparação com seqüências das γ -prolaminas de milho e sorgo revelou a existência de elementos conservados nos promotores destes genes.

Um fragmento do promotor localizado entre -446 e -704 pb do ATG inicial (**rev**) é necessário para a expressão máxima do gene indicador. Um fragmento contendo apenas 112 pb a montante da região estrutural do gene da γ -coixina (**6(10)**) ainda é capaz de induzir a expressão do gene indicador em endosperma de milho. Seqüências localizadas entre -112 e -446 do ATG inicial parecem ser também importantes na expressão do gene indicador.

Proteínas nucleares de *Coix* ligam-se de maneira específica a seqüências do promotor da γ -coixina. Foi demonstrada a existência de, pelo menos, três locais distintos onde estas interações ocorrem. Um localizado entre -478 e -554 pb, outro entre -303 e -478 pb e um último entre -159 e -303 pb do ATG inicial.

A seqüência de 75 pb localizada entre -478 pb e -554 pb é a única a apresentar interações específicas com proteínas nucleares numa extensa região do promotor (fragmento **226**). A região do promotor representada pelo fragmento 226 se sobrepõe àquela cuja eliminação (construção **3(10)**) causa acentuada redução na atividade do gene indicador. A única seqüência conservada na região delimitada pelos 75 pb que interagem especificamente com proteínas nucleares, entre os promotores de γ -prolaminas analisados, é a representada pelo elemento palindrômico *Palind*. Os resultados das tentativas de se delimitar a seqüência onde ocorrem as interações proteína-DNA no fragmento de 75 pb, ainda que não permitam afirmar o envolvimento do elemento *Palind*, não descartam essa possibilidade.

Bibliografia

- Adachi, T.; Izumi, H.; Yamada, T.; Tanaka, K.; Takeuchi, S.; Nakamura, R. e Matsuda, T. (1993). **Gene structure and expression of rice seed allergenic proteins belonging to the alpha-amylase/trypsin inhibitor family.** *Plant Mol. Biol.* **21**: 239-248.
- Akira, S.; Isshiki, H.; Sugita, T.; Tanabe, O.; Kinoshita, S.T.; Nishio, Y.; Nakajima, T.; Hirano, T. e Kishimoto, T. (1990). **A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family.** *EMBO J.* **9**: 1897-1906.
- Albani, D. e Robert, L.S. (1995). **Cloning and characterization of a Brassica napus gene encoding a homologue of the B subunit of a heteromeric CCAAT-binding factor.** *Gene* **167**: 209-213.
- Alberts, B.; Bray, D.; Lewis, J.; Raff, M.; Keith, R. e Watson, J.D. (1989). The cell nucleus. *in* **Molecular biology of the cell 2nded.** Garland Publishing, Inc. - New York. pp 483-549.
- Altschuler, Y.; Rosenberg, N.; Harel, R. e Galili, G. (1993). **The N- and C-terminal regions regulate the transport of weath γ -gliadin through the endoplasmic reticulum in Xenopus oocytes.** *Plant Cell* **5**: 443-450.
- Andrews, N.C.; Erdjument-Bromage, H.; Davidson, M.B.; Tempst, P. e Orkin, S.H. (1993). **Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein.** *Nature* **362**: 722-728.
- Antoniewski, C.; Mugat, B.; Delbac, F. e Lepesant, J.A. (1996). **Direct repeats bind the EcR/USP receptor and mediate ecdysteroid responses in Drosophila melanogaster.** *Mol. Cell Biol.* **16**: 2977-2986.
- Arora, R.K. (1977). **Job's tears (*Coix lacryma-jobi*) - a minor food and fodder crop from North Eastern India.** *Econ. Bot.* **31**: 358-366.
- Ary, M.B.; Richardson, M. e Shewry, P.R. (1989). **Purification and characterization of an insect alpha-amylase inhibitor/endochitinase from seeds of Job's Tears (*Coix lacryma-jobi*).** *Biochim. Biophys. Acta* **999**: 260-266.
- Baima, S.; Nobili, F.; Sessa, G.; Lucchetti, S.; Ruberti, I. e Morelli, G. (1995). **The expression of the Athb-8 homeobox gene is restricted to provascular cells in *Arabidopsis thaliana*.** *Development* **121**: 4171-4182.
- Barros, E.G.; Takasaki, K.; Kirleis, A.W. e Larkins, B.A. (1991). **Nucleotide sequence of a cDNA clone encoding γ -kafirin protein from *Sorghum bicolor*.** *Plant Physiol.* **97**: 1606-1607.
- Bass, H.W.; Webster, C.; O'Brien, G.R.; Roberts, J.K.M. e Boston, R.S. (1992). **A maize ribosome-inactivating protein is controlled by the transcriptional activator *Opaque-2*.** *Plant Cell* **4**: 225-234.
- Bäumlein, H.; Nagy, I.; Villarroel, R.; Inze, D. e Wobus, U. (1992). **Cis-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene.** *Plant J.* **2**: 233-239.
- Benbrook, D.M. e Jones, N.C. (1994). **Different binding specificities and transactivation of variant CRE's by CREB complexes.** *Nucl. Acids Res.* **22**: 1463-1469.
- Benezra, R.; Davis, R.L.; Lockshon, D.; Turner, D.L. e Weintraub, H. (1990). **The protein Id: a negative regulator of helix-loop-helix DNA binding proteins.** *Cell* **61**: 49-59.

- Birnstiel, M.L.; Busslinger, M. e Strub, K. (1985). **Transcription termination and 3' processing: the end is in site!** *Cell* **41**: 349-359.
- Blechl, A.E.; Lorens, G.F.; Greene, F.C.; Mackey, B.E. e Anderson, O.D. (1994). **A transient assay for promoter activity of wheat seed storage protein genes and other genes expressed in developing endosperm.** *Plant Sci.* **102**: 69-80.
- Bonven, B.J.; Nielsen, A.L.; Norby, P.L.; Pedersen, F.S. e Jorgensen, P. (1995). **E-box variants direct formation of distinct complexes with the basic helix-loop-helix protein ALF1.** *J. Mol. Biol.* **294**: 564-575.
- Brignon, P. e Chaubet, N. (1993). **Constitutive and cell-division-inducible protein-DNA interactions in two maize histone gene promoters.** *Plant J.* **4**: 445-457.
- Bruce, W.B. e Gurley, W.B. (1987). **Functional domains of a T-DNA promoter active in crown gall tumors.** *Mol. Cell Biol.* **7**: 59-67.
- Bucher, P. (1990). **Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences.** *J. Mol. Biol.* **212**: 563-578.
- Bullock, W.O.; Fernandez, J.M. e Short, J.M. (1987). **XL1-Blue: A high efficiency plasmid transforming recA *Escherichia coli* strain with beta-galactosidase selection.** *Biotechniques* **5**: 376-379.
- Buratowski, S. (1994). **The basics of basal transcription by RNA polymerase II.** *Cell* **77**: 1-3.
- Burke, C.; Yu, X.-B.; Marchitelli, L.; Davis, E.A. e Ackermann, S. (1990). **Transcription factor IIA of wheat and human function similarly with plant and animal viral promoters.** *Nucl. Acids Res.* **18**: 3611-3620.
- Burr, B. e Burr, F.A. (1982). **Three mutations in *Zea mays* affecting zein accumulation: A comparison of zein polypeptides, *in vitro* synthesis and processing, mRNA levels and genomic organization.** *J. Cell. Biol.* **94**: 201-206.
- Cabot, E.L. e Beckenbach, A.T. (1989). **Simultaneous editing of multiple nucleic acid and protein sequences with ESEE.** *Comput. Appl. Biosci.* **5**: 233-234.
- Carlson, M. e Laurent, B.C. (1994). **The SBF/SWI family of global transcriptional activators.** *Curr. Opin. Cell Biol.* **6**: 396-402.
- Carré, I.A. e Kay, S.A. (1995). **Multiple DNA-protein complexes at circadian-regulated promoter element.** *Plant Cell* **7**: 2039-2051.
- Cashmore, A.R. (1984). **Structure and expression of a pea nuclear gene encoding a chlorophyll a/b-binding polypeptide.** *Proc. Natl. Acad. Sci. USA* **81**: 2960-2964.
- Chodosh, L.A.; Baldwin, A.S.; Carthew, R.W. e Sharp, P.A. (1988). **Human CCAAT-binding proteins have heterologous subunits.** *Cell* **53**: 11-24.
- Chodosh, L.A. (1996). Mobility shift DNA-binding assay using gel electrophoresis. *in Current Protocols in Molecular Biology* Ausubel, F.M.; Brent, R.; Kingston, R.E.; Moore, D.D.; Seidman, J.G.; Smith, J.A. e Struhl, K. eds. John Wiley & Sons, Inc. - pp 12.2.1-12.2.10.
- Chung, I.K.; Lee, S.Y.; Ito, T.; Tanaka, H.; Nam, H.G. e Takagi, M. (1995). **Lycopersicon peruvianum are highly heterologous but contain short blocks of homologous sequences.** *Plant Cell Physiol.* **36**: 1621-1627.
- Consonni, G.; Geuna, F.; Gavazzi, G. e Tonelli, C. (1993). **Molecular homology among members of the R gene family in maize.** *Plant J.* **3**: 335-346.

- Fincher, G.B. (1989). **Molecular and cellular biology associated with endosperm mobilization in germinating cereal grains.** *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**: 305-346.
- Fincher, G.B. e Stone, B.A. (1986). **Cell walls and their components in cereal grain technology.** *Adv. Cereal Sci. Technol.* **8**: 207-295.
- Folkers, N.S.; Borrelli, E. e Sassone-Corsi, P. (1991). **CREM gene: use of alternative DNA-binding domains generates multiple antagonists of cAMP-induced transcription.** *Cell* **64**: 739-749.
- Forde, B.G.; Heyworth, A.; Pywell, J. e Kreis, M. (1985). **Nucleotide sequence of a B1 hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize.** *Nucl. Acids Res.* **13**: 7327-7339.
- Foster, R.; Izawa, T. e Chua, N. (1994). **Plant bZIP proteins gather at ACGT elements.** *FASEB J.* **8**: 192-200.
- Freitas, F.A.; Yunes, J.A.; Silva, M.J.; Arruda, P. e Leite, A. (1994). **Structural characterization and promoter activity analysis of the γ -kafirin gene from sorghum.** *Mol. Gen. Genet.* **245**: 177-186.
- Fromm, H.; Katagiri, F. e Chua, N. (1991). **The tobacco transcription activator TGA1a binds to a sequence in the 5' upstream region of a gene encoding a TGA1a-related protein.** *Mol. Gen. Genet.* **229**: 181-188.
- Gallardo, D.; Reina, M.; Rigua, J.; Boronat, A. e Palau, J. (1988). **Genomic organization of the 28 kDa glutelin-2 gene from maize.** *Plant Sci.* **54**: 211-218.
- Gallie, D.R. (1993). **Posttranscriptional regulation of gene expression in plants.** *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44**: 77-105.
- Garratt, R.; Oliva, G.; Caracelli, I.; Leite, A. e Arruda, P. (1993). **Studies of the zein-like α -prolamins based on an analysis of amino acid sequences: implications for their evolution and three dimensional structure.** *Proteins* **15**: 88-99.
- Garrity, P.A.; Chen, D.; Rothenberg, E.V. e Wold, B.J. (1994). **Interleukin-2 transcription is regulated *in vivo* at the level of coordinated binding of both constitutive and regulated factors.** *Mol. Cell Biol.* **14**: 2159-2169.
- Gasser, C.S. (1996). **Homeodomains ring a *Bell* in plant development.** *Trends Plant Sci* **1**: 134-136.
- Geetha, K.B.; Lending, C.R.; Lopes, M.A.; Wallace, J.C. e Larkins, B.A. (1991). **opaque-2 modifiers increase γ -zein synthesis and alter its spatial distribution in maize endosperm.** *Plant Cell* **3**: 1207-1219.
- Geli, M.I.; Torrent, M. e Ludevid, D. (1994). **Two structural domains mediate two sequential events in γ -zein targeting: protein endoplasmic reticulum retention and protein body formation.** *Plant Cell* **6**: 1911-1922.
- Gianazza, E.; Righetti, P.G.; Pioli, F.; Galante, E. e Soave, C. (1976). **Size and charge heterogeneity of zein in normal and *opaque-2* maize endosperms.** *Maydica* **21**: 1-17.
- Giese, K.; Kingsley, C.; Kirshner, J.R. e Grosschedl, R. (1995). **Assembly and function of a TCR α enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions.** *Genes Dev.* **9**: 995-1008.

- Gill, G.; Pascal, E.; Tseng, G. e Tjian, R. (1994). **A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation.** *Proc. Natl. Acad. Sci. USA* **91**: 192-196.
- Gille, H.; Sharrocks, A. e Shaw, P. (1992). **Phosphorylation of p62TCF by MAP kinase stimulates ternary complex formation at cFos promoter.** *Nature* **358**: 414-417.
- Gilmartin, P.M.; Sorokin, L.; Memelink, J. e Chua, N. (1990). **Molecular light switches for plant genes.** *Plant Cell* **2(5)**: 369-378.
- Goff, S.A.; Klein, T.M.; Roth, B.A.; Fromm, M.E.; Cone, K.C.; Radicella, J.P. e Chandler, V.L. (1990). **Transactivation of anthocyanin biosynthetic genes following transfer of B regulatory genes into maize tissues.** *EMBO J.* **9**: 2517-2522.
- Gonzales, G.A. e Montminy, M. (1989). **Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133.** *Cell* **59**: 675-680.
- Goodrich, J.A.; Hoey, T.; Thut, C.J.; Admon, A. e Tjian, R. (1993). **Drosophila TAF_{II}40 interacts with both a VP16 activator domain and the basal transcription factor TFIIB.** *Cell* **75**: 519-530.
- Goodrich, J.A. e Tjian, R. (1994). **TBP-TAF complexes: selectivity factors for eukaryotic transcription.** *Curr. Opin. Cell Biol.* **6**: 403-409.
- Goodwin, B.C. (1985). **What are the causes of morphogenesis?** *BioEssays* **3**: 32-36.
- Goto, K. e Meyerowitz, E.M. (1994). **Function and regulation of the Arabidopsis floral homeotic gene PISTILLATA.** *Genes Dev.* **8**: 1548-1560.
- Grafi, G. e Larkins, B.A. (1995). **Endoreduplication in maize endosperm: involvement of M phase-promoting factor inhibition and induction of S phase-related kinases.** *Science* **269**: 1262-1264.
- Grasser, K.D. (1995). **Plant chromosomal high mobility group (HMG) proteins.** *Plant J.* **7**: 185-192.
- Grayson, J.; Williams, R.S.; Yu, Y.T. e Bassel-Duby, R. (1995). **Synergistic interactions between heterologous upstream activation elements and specific TATA sequences in a muscle-specific promoter.** *Mol. Cell Biol.* **15**: 1870-1878.
- Grosschedl, R. (1995). **High-order nucleoprotein complexes in transcription: analogies with site-specific recombination.** *Curr. Opin. Cell Biol.* **7**: 362-370.
- Guan, L.; Polidoros, A.N. e Scandalios, J.G. (1996). **Isolation, characterization and expression of the maize Cat2 catalase gene.** *Plant Mol. Biol.* **30**: 913-924.
- Habben, J.E.; Moro, G.L.; Hunter, B.G.; Hamaker, B.R. e Larkins, B.A. (1995). **Elongation factor 1 alpha concentration is highly correlated with the lysine content of maize endosperm.** *Proc. Natl. Acad. Sci. USA* **92**: 8640-8644.
- Halle, J. e Meisterernst, M. (1996). **Gene expression: increasing evidence for a transcriptome.** *Trends Genet.* **12**: 161-163.
- Hammond-Kosack, M.C.; Holdsworth, M.J. e Bevan, M.W. (1993). **In vivo footprinting of a low molecular weight glutenin gene (LMWG-1D1) in wheat endosperm.** *EMBO J.* **12**: 545-554.
- Hanaham, D. (1983). **Studies on transformation of *Escherichia coli* with plasmids.** *J. Mol. Biol.* **166**: 557

- Haq, T.A.; Mason, H.S.; Clements, J.D. e Arntzen, C.J. (1995). **Oral immunization with a recombinant bacterial antigen produced in transgenic plants.** *Science* **268**: 714-716.
- Harlan, J.R. (1992). Indigenous african agriculture. *in The origins of agriculture - an international perspective* Cown, C.W. e Watson, P.J.eds. Smithsonian Institution Press - Washington. pp 59-70.
- Henderson, E. e Stein, R. (1994). **c-jun inhibits transcriptional activation by the insulin enhancer, and the insulin control element is the target of control.** *Mol. Cell Biol.* **14**: 655-662.
- Henikoff, S. (1987). **Unidirectional digestion with exonuclease III in DNA sequence analysis.** *Method. Enzymol.* **155**: 156-165.
- Herskowitz, I. (1989). **A regulatory hierarchy for cell specialization in yeast.** *Nature* **342**: 749-757.
- Higgins, D.G. e Sharp, P.M. (1989). **Fast and sensitive multiple sequence alignments on a microcomputer.** *Comput. Appl. Biosci.* **5**: 151-152.
- Higgins, T.J.V. (1984). **Synthesis and regulation of major proteins in seeds.** *Plant Physiol* **35**: 191-221.
- Higo, K. e Higo, H. (1996). **Cloning and characterization of the rice CatA catalase gene, a homologue of the maize Cat3 gene.** *Plant Mol. Biol.* **30**: 505-521.
- Hinnebusch, A.G. (1984). **Evidence for translational regulation of the activator of general amino acid control in yeast.** *Proc. Natl. Acad. Sci. USA* **81**: 6442-6446.
- Holdsworth, M.J.; Muñoz-Blanco, J.; Hammond-Kosack, M.C.; Colot, V.; Schuch, W. e Bevan, M.W. (1995). **The maize transcription factor Opaque-2 activates a wheat glutenin promoter in plant and yeast cells.** *Plant Mol. Biol.* **29**: 711-720.
- Hollenberg, S.M.; Sternglanz, R.; Cheng, P.F. e Weintraub, H. (1996). **Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system.** *Mol. Cell Biol.* **15**: 3813-3822.
- Howe, K.M. e Watson, R.J. (1991). **Nucleotide preferences in sequence-specific recognition of DNA by c-myb protein.** *Nucl. Acids Res.* **19**: 3913-3919.
- Huang, N.; Sutliff, T.D.; Litts, J.C. e Rodriguez, R.L. (1990). **Classification and characterization of the rice alpha-amylase multigene family.** *Plant Mol. Biol.* **14**: 655-668.
- Hwang, J.M.; Oshima, M.; Yokota, H.O. e Okajima, T. (1993). **Efficient utilization of adley (*Coix lacrima-jobi* L.) foliage by the green crop fractionation method: 3. Chemical quality and feeding value of pressed cake silage.** *J. Jap. Soc. Grassl. Sci.* **39**: 16-21.
- Inada, H.; Kusumi, K.; Nishimura, M. e Iba, K. (1996). **Specific expression of the chloroplast gene for RNA polymerase (rpoB) at an early stage of leaf development in rice.** *Plant and Cell Physiology* **37**: 229-232.
- Ishiguro, Y.; Okamoto, K.; Sakamoto, H. e Sonoda, Y. (1993). **Antimicrobial substances coixindens A and B in etiolated seedlings of adlay.** *Nippon Nogeikagaku Kaish* **67**: 1405-1410.
- Izawa, T.; Foster, R. e Chua, N. (1993). **Plant bZIP protein DNA binding specificity.** *J. Mol. Biol.* **230**: 1131-1144.

- Jackson, R.J. e Standart, N. (1990). **Do the poly(A) tail and 3' untranslated region control mRNA translation?** *Cell* **62**: 15-24.
- Jain, S.K. e Banerjee, D.K. (1974). **Preliminary observations on the ethnobotany of the genus *Coix*.** *Econ. Bot.* **28**: 38-42.
- Jefferson, R.A.; Kavanagh, T.A. e Bevan, M.W. (1987). **GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants.** *EMBO J.* **6**: 3901-3907.
- Johnson, P.F.; Landschulz, W.H.; Graves, B.J. e McKnight, S.L. (1987). **Identification of a rat liver nuclear protein that binds to the enhancer core element of three animal viruses.** *Genes Dev.* **1**: 133-146.
- Johnson, P.F. e McKnight, S.L. (1989). **Eukaryotic transcriptional regulatory proteins.** *Annu. Rev. Biochem.* **58**: 799-839.
- Johnson, P.F. e Williams, S.C. (1994). CCAAT/enhancer binding (C/EBP) proteins. *in Liver Gene Expression* Tronche, F. e Watson, P.J. eds. R.G. Landes Comp. - Austin. pp 231-258.
- Joly, J.S.; Joly, C.; Schulte-Merker, S.; Boulekbache, H. e Condamine, H. (1993). **The ventral and posterior expression of the zebrafish homeobox gene *eve1* is perturbed in dorsalized and mutant embryos.** *Development* **119**: 1261-1275.
- Jones, R.L. e Jacobsen, J.V. (1991). **Regulation of synthesis and transport of secreted proteins in cereal aleurone.** *Int. Rev. Cytol.* **126**: 49-88.
- Joshi, C.P. (1987). **An inspection of the domain between putative TATA box and translation start site in 79 plant genes.** *Nucl. Acids Res.* **15**: 6643-6653.
- Kamper, J.T.; Kamper, U.; Rogers, L.M. e Kolattukudy, P.E. (1994). **Identification of regulatory elements in the cutinase promoter from *Fusarium solani* f. sp. *pisi* (*Nectria haematococca*).** *J. Biol. Chem.* **269**: 9195-9204.
- Kasai, Y.; Nambu, J.R.; Lieberman, P.M. e Crews, S.T. (1992). **Dorsal-ventral patterning in *Drosophila*: DNA binding of snail protein to the single-minded gene.** *Proc. Natl. Acad. Sci. USA* **89**: 3414-3418.
- Kawahara, R.; Komamine, A. e Fukuda, H. (1995). **Isolation and characterization of homeobox-containing genes of carrot.** *Plant Mol. Biol.* **27(1)**: 155-164.
- Kehoe, D.M.; Degenhardt, J.; Winicov, I. e Tobin, E.M. (1994). **Two 10-bp regions are critical for phytochrome regulation of a *Lemna gibba* *Lhcb* gene promoter.** *Plant Cell* **6**: 1123-1134.
- Kermicle, J.L. e Alleman, M. (1990). **Gametic imprinting in maize in relation to the angiosperm life cycle.** *Development (suppl)*: 9-14.
- Kerstetter, R.; Vollbrecht, E.; Lowe, B.; Veit, B.; Yamaguchi, J. e Hake, S. (1994). **Sequence analysis and expression patterns divide the maize *knotted1*-like homeobox genes into two classes.** *Plant Cell* **6**: 1877-1887.
- Kim, J.; Tzamarias, D.; Ellemberger, T.; Harrison, S.C. e Struhl, K. (1993). **Adaptability at the protein-DNA interface is an important aspect of sequence recognition by bZIP proteins.** *Proc. Natl. Acad. Sci. USA* **90**: 4513-4517.
- Kim, J.L.; Nikolov, D.B. e Burley, S.K. (1993). **Co-crystal structure of TBP recognizing the minor groove as a TATA element.** *Nature* **365**: 520-527.

- Kim, Y.-J.; Bjorklund, S.; Li, Y.; Sayre, M. e Kornberg, R. (1994). **A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II.** *Cell* **77**: 599-608.
- Kirihara, J.A.; Husperger, J.P. e Mahoney, J.W. (1988). **Differential expression of a gene for a methionine-rich storage protein.** *Mol. Gen. Genet.* **211**: 477-484.
- Kloeckener-Gruissem, B.; Vogel, J.M. e Freeling, M. (1992). **The TATA box promoter region of maize Adh1 affects its organ-specific expression.** *EMBO J.* **11**: 157-166.
- Kodrzycki, R.; Boston, R.S. e Larkins, B.A. (1989). **The opaque-2 mutation of maize differentially reduces zein gene transcription.** *Plant Cell* **1**: 105-114.
- Koleske, A.J. e Young, R.A. (1994). **An RNA polymerase II holoenzyme responsive to activators.** *Nature* **368**: 466-469.
- Kornberg, R.D. e Lorch, Y. (1995). **Interplay between chromatin structure and transcription.** *Curr. Opin. Cell Biol.* **7**: 371-375.
- Krebbers, E.; Da Silva, C.A.; Denis, M.; DHondt, K. e Vanderkerckhove, J. (1993). Modification of plant seed storage proteins. *in Plant storage compounds: biosynthesis, interactions and manipulation 1sted.* Shewry, P.R. e Stobart, A.K.eds.Oxford University Press - Oxford. pp 317-324.
- Kreis, M.; Williamson, M.S.; Forde, J.; Schmutz, D.; Clark, J.; Buxton, B.; Pywell, J.; Marris, C.; Henderson, J.; Harris, N.; Shewry, P.R.; Forde, B.G. e Mifflin, B.J. (1986). **Differential gene expression in the developing barley endosperm.** *Phil. Trans. R. Soc. Lond. B* **314**: 355-365.
- Kroeger, P.E. e Morimoto, R.I. (1994). **Selection of new HSF1 and HSF2 DNA-binding sites reveals differences in trimer cooperativity.** *Mol. Cell Biol.* **14**: 7592-7603.
- Kuhlemeier, C. (1992). **Transcriptional and post-transcriptional regulation of gene expression in plants.** *Plant Mol. Biol.* **19**: 1-14.
- Landry, J. e Moureaux, T.H. (1980). **Distribution and amino acid composition of protein groups located in different histological parts of maize grain.** *J. Agric. Food Chem.* **28**: 1186
- Landschulz, W.H.; Johnson, P.F. e McKnight, S.L. (1988). **The leucine zipper hypothetical structure common to a new class of DNA binding protein.** *Science* **240**: 1759-1764.
- Lane, D.; Prentki, P. e Chandler, M. (1992). **Use of gel retardation to analyze protein-nucleic acid interactions.** *Microbiol. Rev.* **56**: 509-528.
- Langridge, P.; Eibel, H.; Brown, J.W.S. e Feix, G. (1984). **Transcription from maize storage protein gene promoters in yeast.** *EMBO J.* **3**: 2467-2471.
- Latcman, D.S. (1990). **Eukaryotic transcription factors.** *Biochem. J.* **270**: 41-49.
- Lawrence, M.C.; Izard, T.; Beuchat, M.; Blagrove, R.J. e Colman, P.M. (1994). **Structure of phaseolin at 2.2 Å resolution. Implications for a common vicilin/legumin structure and the genetic engineering of seed storage proteins.** *J. Mol. Biol.* **238**: 748-776.
- Lazzeri, P.A. e Shewry, P.R. (1993). **Biotechnology of cereals.** *Biotechnology and Genetic Engineering Reviews* **11**: 79-146.
- Lee, J.I.; Kim, S.D.; Park, J.H.; Ahn, B.O. e Lee, S.T. (1994). **A new thin shelled and high-yielding variety of job's-tears "Yulmoo 1".** *RDA J. Agr. Sci. Upland Ind. Crops* **36**: 158-160.

- Leite, A.; Ottoboni, L.M.M.; Targon, M.L.P.N.; Silva, M.J.; Turcinelli, S.R. e Arruda, P. (1990). **Phylogenetic relationship of zeins and coixins as determined by immunological cross-reactivity and Southern blot analysis.** *Plant Mol. Biol.* **14**: 743-751.
- Leite, A.; Freitas, F.A.; Yunes, J.A. e Arruda, P. (1991). **Nucleotide sequence of a cDNA clone encoding γ -coixin from *Coix lacryma-jobi* seeds.** *Plant Physiol.* **97**: 1604-1605.
- Leite, A.; Yunes, J.A.; Turcinelli, S.R. e Arruda, P. (1992). **Cloning and characterization of a cDNA encoding a sulfur-rich coixin.** *Plant Mol. Biol.* **18**: 171-174.
- Lending, C.R.; Kriz, A.L.; Larkins, B.A. e Bracker, C.E. (1988). **Structure of maize protein bodies and immunocytochemical localization of zeins.** *Protoplasma* **143**: 51-62.
- Lending, C.R. e Larkins, B.A. (1989). **Changes in the zein composition of protein bodies during maize endosperm development.** *Plant Cell* **1**: 1011-1023.
- Lewin, B. (1994). Building the transcription complex: promoters, factors and RNA polymerases. *in Genes V 1sted.* Oxford University Press Inc. - New York. pp 847-877.
- Li, D. e Kolattukudy, P.E. (1995). **Cloning and expression of cDNA encoding a protein that binds a palindromic promoter element essential for induction of fungal cutinase by plant cutin.** *J. Biol. Chem.* **270**: 11753-11756.
- Li, L.; Qu, R.D.; Kochko, A.; Fauquet, C. e Beachy, R.N. (1993). **An improved rice transformation system using the biolistic method.** *Plant Cell Rep.* **12**: 250-255.
- Li, X.-Y.; Mantovani, R.; Hooft van Huijsduijnen, R.; Andre, I.; Benoist, C. e Mathis, D. (1992). **Evolutionary variation of the CCAAT-binding transcription factor NF-Y.** *Nucl. Acids Res.* **20**: 1087-1091.
- Liaud, M.F.; Valentin, C.; Brandt, U.; Bouget, F.Y.; Kloareg, B. e Cerff, R. (1993). **The GAPDH gene system of the red alga *Chondrus crispus*: promoter structures, intron/exon organization, genomic complexity and differential expression of genes.** *Plant Mol. Biol.* **23**: 981-994.
- Lin, B.-Y. (1982). **Association of endosperm reduction with parental imprinting in maize.** *Genetics* **100**: 475-486.
- Lin, B.-Y. (1984). **Ploidy barrier to endosperm development in maize.** *Genetics* **107**: 103-115.
- Lin, Y.S. e Green, M.R. (1991). **Mechanism of action of an acidic transcriptional activator in vitro.** *Cell* **64**: 971-981.
- Lohmer, S.; Maddaloni, M.; Motto, M.; Di Fonzo, N.; Hartings, H.; Salamini, F. e Thompson, R.D. (1991). **The maize regulatory locus *opaque-2* encodes a DNA-binding protein which activates the transcription of the *b-32* gene.** *EMBO J.* **10**: 617-624.
- Lopes, M.A. e Larkins, B.A. (1993). **Endosperm origin, development, and function.** *Plant Cell* **5**: 1383-1399.
- Ludwig, M.Z. e Kreitman, M. (1995). **Evolutionary dynamics of the enhancer region of even-skipped in *Drosophila*.** *Mol. Biol. Evol.* **12**: 1002-1011.
- Ma, J.M.; Hiatt, A.; Hein, M.; Vine, N.D.; Wang, F.; Stabila, P.; Dolleweerd, C.; Mostov, K. e Lehner, T. (1995). **Generation and assembly of secretory antibodies in plants.** *Science* **268**: 716-719.

- Maier, U.G.; Brown, J.W.S.; Toloczky, C. e Feix, G. (1987). **Binding of nuclear factor to a consensus sequence in the 5' flanking region of zein genes from maize.** *EMBO J.* **6**: 17-22.
- Maldonado, E.; Shiekhattar, R.; Sheldon, M.; Cho, H.; Drapkin, R.; Rickert, P.; Lees, E.; Anderson, C.W.; Linn, S. e Reinberg, D. (1996). **A human RNA polymerase II complex associated with SRB and DNA-repair proteins.** *Nature* **381**: 86-89.
- Maldonado, E. e Reinberg, D. (1995). **News on initiation and elongation of transcription by RNA polymerase II.** *Curr. Opin. Cell Biol.* **7**: 352-361.
- Manley, J.L. e Proudfoot, N.J. (1994). **RNA 3'-ends: formation and function - meeting review.** *Genes Dev.* **8**: 259-264.
- Mattsson, J.; Söderman, E.; Svenson, M.; Borkird, C. e Engström, P. (1992). **A new homeobox-leucine zipper gene from *Arabidopsis thaliana*.** *Plant Mol. Biol.* **18**: 1019-1022.
- Matzke, A.J.M.; Stoger, E.M.; Scherthner, J.P. e Matzke, M.A. (1990). **Deletion analysis of a zein gene promoter in transgenic tobacco plants.** *Plant Mol. Biol.* **14**: 323-332.
- McCormick, A.; Brady, H.; Fukushima, J. e Karin, M. (1991). **The pituitary-specific regulatory gene *GHF1* contains a minimal cell type-specific promoter centered around its TATA box.** *Genes Dev.* **5**: 1490-1503.
- McFadden, G.I.; Ahluwalia, B.; Clarke, A.E. e Fincher, G.B. (1988). **Expression sites and developmental regulation of genes encoding (1-3, 1-4)- β -glucanases in germinated barley.** *Planta* **173**: 500-508.
- Meier, I. e Grissem, W. (1994). **Novel conserved sequence motifs in plant G-box binding proteins and implications for interactive domains.** *Nucl. Acids Res.* **22**: 470-478.
- Meshi, T. e Iwabuchi, M. (1995). **Plant transcription factors.** *Plant Cell Physiol.* **36(8)**: 1405-1420.
- Miao, Z.; Liu, X. e Lam, E. (1994). **TGA3 is a distinct member of the TGA family of bZIP transcription factors in *Arabidopsis thaliana*.** *Plant Mol. Biol.* **25**: 1-11.
- Miller, P. (1731). **The Gardener's Dictionary**- London.
- Moffat, A.S. (1995). **Exploring transgenic plants as a new vaccine source.** *Science* **268**: 658-660.
- Moffat, B.A.; McWhinnie, E.A.; Agarwal, S.K. e Schaff, D.A. (1994). **The adenine phosphoribosyltransferase-encoding gene of *Arabidopsis thaliana*.** *Gene* **143**: 211-216.
- Morrish, F.M. e Fromm, M.E. (1992). **Cereal transformation methods.** *Curr. Opin. Biotech.* **3**: 141-146.
- Motto, M.; Maddaloni, M.; Ponziani, G.; Brembilla, M.; Marotta, R.; Di Fonzo, N.; Soave, C.; Thompson, R. e Salamini, F. (1988). **Molecular cloning of the *o2-m5* allele of *Zea mays* using transposon marking.** *Mol. Gen. Genet.* **212**: 488-494.
- Mukumoto, F.; Hirose, S.; Imaseki, H. e Yamazaki, K. (1993). **DNA sequence requirement of a TATA element-binding protein from *Arabidopsis* for transcription in vitro.** *Plant Mol. Biol.* **23**: 995-1003.

- Muller, M.; Muth, J.R.; Gallusci, P.; Knudsen, S.; Maddaloni, M.; Motto, M.; Schmitz, D.; Sorensen, M.B.; Salamini, F.; Von Wettstein, D. e Thompson, R.D. (1995). **Regulation of storage protein synthesis in cereal seeds: Developmental and nutritional aspects.** *Journal of Plant Physiology* **145**: 606-613.
- Muller, M. e Knudsen, S. (1993). **The nitrogen response of a barley C-hordein promoter is controlled by positive and negative regulation of the GCN4 and endosperm box.** *Plant J.* **4**: 343-355.
- Murashige, T. e Skoog, F. (1962). **A revised medium for rapid growth and bioassays with tobacco tissue culture.** *Physiol. Plant.* **15**: 473-479.
- Nakagoshi, H.; Nagase, T.; Kanei-Ishii, C.; Ueno, Y. e Ishii, S. (1990). **Binding of the c-myc proto-oncogene product to the simian virus 40 enhancer stimulates transcription.** *J. Biol. Chem.* **265**: 3479-3483.
- Natesan, S. e Gilman, M.Z. (1993). **DNA bending and orientation-dependent function of YY1 in the c-fos promoter.** *Genes Dev.* **7**: 2497-2509.
- Nielsen, P.S. e Gausing, K. (1993). **In vitro binding of nuclear proteins to the barley plastocyanin gene promoter region.** *Eur. J. Biochem.* **217**: 97-104.
- Nishio, Y.; Isshiki, H.; Kishimoto, T. e Akira, S. (1993). **A nuclear factor for interleukin-6 expression (NF-IL6) and the glucocorticoid receptor synergistically activate transcription of the rat α 1-acid glycoprotein gene via direct protein-protein interaction.** *Mol. Cell Biol.* **13**: 1854-1862.
- Niu, X.; Guiltinan, M.J.; Niu, X. e Guiltinan, M.J. (1994). **DNA binding specificity of the wheat bZIP protein EmBP-1.** *Nucl. Acids Res.* **22**: 4969-4978.
- Numata, M.; Yamamoto, A.; Moribayashi, A. e Yamada, H. (1994). **Antitumor components isolated from the Chinese herbal medicine *Coix lacrima-jobi*.** *Planta Medica* **60**: 356-359.
- Odell, J.T.; Nagy, F. e Chua, N. (1985). **Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter.** *Nature* **313**: 810-812.
- Olesen, J.T. e Guarente, L. (1990). **The HAP2 subunit of yeast CCAAT transcriptional activator contains adjacent domains for subunit association and DNA recognition: model for the HAP2/3/4 complex.** *Genes Dev.* **4**: 1714-1729.
- Oliphant, A.R.; Brandl, C.J. e Struhl, K. (1989). **Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein.** *Mol. Cell Biol.* **9**: 2944-2949.
- Olive, M.; Williams, S.C.; Dezan, C.; Johnson, P.F. e Vinson, C. (1996). **Design of a C/EBP-specific, dominant-negative bZIP protein with both inhibitory and gain-of-function properties.** *J. Biol. Chem.* **271**: 2040-2047.
- Osborne, T.B. (1908). **Our present knowledge of plant proteins.** *Science* **28**: 417-427.
- Osborne, T.B. e Mendel, L.B. (1914). **Nutritive properties of proteins of maize kernel.** *J. Biol. Chem.* **18**: 1-6.
- Ottoboni, L.M.M.; Leite, A.; Targon, M.L.P.N.; Silva, M.J. e Arruda, P. (1990a). **Heterogeneity of *Coix*, maize and teosinte prolamins detected by isoelectric focusing.** *Rev. Bras. Genet.* **13**: 313-322.

- Ottoboni, L.M.M.; Leite, A.; Targon, M.L.P.N.; Silva, M.J.; Turcinelli, S.R. e Arruda, P. (1990b). **Characterization of the storage protein in seed of *Coix lacryma-joby* var. Adlay.** *J. Agric. Food Chem.* **38**: 631-635.
- Ottoboni, L.M.M.; Leite, A.; Yunes, J.A.; Targon, M.L.P.N.; Souza, G.A., Filho e Arruda, P. (1993). **Sequence analysis of 22 kDa-like α -coixin genes and their comparison with homologous zein and kafirin genes reveals highly conserved protein structure and regulatory elements.** *Plant Mol. Biol.* **21**: 765-778.
- Overdier, D.G.; Porcella, A. e Costa, R.H. (1994). **The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino acid residues adjacent to the recognition helix.** *Mol. Cell Biol.* **14**: 2755-2766.
- Pabo, C.O. e Sauer, R.T. (1992). **Transcription factors: structural families and principles of DNA recognition.** *Annu. Rev. Biochem.* **61**: 1053-1095.
- Payne, J.W. e Walker-Smith, D.J. (1987). **Isolation and identification of proteins from the peptide-transport carrier in the scutellum of germinating barley (*Hordeum vulgare* L.) embryos.** *Planta* **170**: 263-271.
- Pedersen, K.; Argos, P.; Naravana, S.L.V. e Larkins, B.A. (1986). **Sequence analysis and characterization of a maize gene encoding a high-sulfur zein protein of M_r 15000.** *J. Biol. Chem.* **14**: 6279-6284.
- Peterson, M.G. e Tjian, R. (1992). **Transcription. The tell-tail trigger.** *Nature* **358**: 620-621.
- Pinkas, M.; Trotin, F.; Peng, W. e Torck, M. (1994). **Use, chemistry and pharmacology of ten Chinese medicinal plants.** *Fitoterapia* **65**: 343-354.
- Ponte, I.; Guillén, P.; Debón, R.M.; Reina, M.; Aragay, A.; Espel, E.; Di Fonzo, N. e Palau, J. (1994). **Narrow A/T-rich zones present at the distal 5'-flanking sequences of the zein genes *Zc1* and *Zc2* bind a unique 30 kDa HMG-like protein.** *Plant Mol. Biol.* **26**: 1893-1906.
- Prat, S.; Cortadas, J.; Puigdomenech, P. e Palau, J. (1985). **Nucleic acid (cDNA) and amino acid sequences of the maize endosperm protein glutelin-2.** *Nucl. Acids Res.* **13**: 1493-1504.
- Purugganan, M.D.; Rounsley, S.D.; Schmidt, R.J. e Yanofsky, M.F. (1995). **Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family.** *Genetics* **140**: 345-356.
- Purugganan, M.D. e Wessler, S.R. (1994). **Molecular evolution of the plant R regulatory gene family.** *Genetics* **138**: 849-854.
- Pyee, J. e Kolattukudy, P.E. (1995). **The gene for the major cuticular wax-associated protein and three homologous genes from broccoli (*Brassica oleracea*) and their expression patterns.** *Plant J.* **7**: 49-59.
- Qu, R.D. e Huang, A.H. (1990). **Oleosin KD 18 on the surface of oil bodies in maize. Genomic and cDNA sequences and the deduced protein structure.** *J. Biol. Chem.* **265**: 2238-2243.
- Quandt, K.; Frech, K.; Karas, H.; Wingender, E. e Werner, T. (1995). **MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data.** *Nucl. Acids Res.* **23**: 4878-4884.
- Quayle, T. e Feix, G. (1992). **Functional analysis of the -300 region of maize zein genes.** *Mol. Gen. Genet.* **231**: 369-374.

- Ramachandran, S.; Hiratsuka, K. e Chua, N. (1994). **Transcription factors in plant growth and development.** *Curr. Opin. Genet. Dev.* **4**: 642-646.
- Rastogi, V. e Oaks, A. (1986). **Hydrolysis of storage proteins in barley endosperms. Analysis of soluble products.** *Plant Physiol.* **81**: 901-906.
- Reina, M.; Guillén, P.; Ponte, I.; Boronat, A. e Palau, J. (1990a). **DNA sequence of the gene encoding the Zc1 protein from *Zea mays* W64 A.** *Nucl. Acids Res.* **18**: 6425
- Reina, M.; Ponte, I.; Guillén, P.; Boronat, A. e Palau, J. (1990b). **Sequence analysis of a genomic clone encoding a Zc2 protein from *Zea mays* W64 A.** *Nucl. Acids Res.* **18**: 6426
- Reiser, L.; Modrusan, Z.; Margossian, L.; Samach, A.; Ohad, N.; Haughn, G.W. e Fischer, R.L. (1995). **The BELL1 gene encodes a homeodomain protein involved in pattern formation in the Arabidopsis ovule primordium.** *Cell* **83**: 735-742.
- Rivin, C.J.; Zimmer, E.A. e Walbot, V. (1982). Isolation of DNA and DNA recombinants from maize. *in Maize for Biological Research* Sheridan, W.F. ed. University Press, University of North Dakota - Grand Forks. pp 161-164.
- Rodriguez-Palenzuela, P.; Pintor-Toro, J.A.; Carbonero, P. e Garcia-Olmedo, F. (1988). **Nucleotide sequence and endosperm-specific expression of the structural gene for the toxin alpha-hordothionin in barley (*Hordeum vulgare* L.).** *Gene* **70**: 271-281.
- Roth, B.A.; Goff, S.A.; Klein, T.M. e Fromm, M.E. (1991). **C1- and R-dependent expression of the maize *Bz1* gene requires sequences with homology to mammalian *myb* and *myc* binding sites.** *Plant Cell* **3**: 317-325.
- Sambrook, J.; Fritsch, E.F. e Maniatis, T. (1989). **Molecular Cloning: a laboratory manual** 2nd. Cold Spring Harbor Laboratory - New York.
- Sanford, J.C.; Devit, M.J.; Russell, J.A.; Smith, F.D.; Harpending, P.R.; Roy, M.K. e Johnston, S.A. (1991). **An improved, helium-driven biolistic device.** *J. Meth. Cell Mol. Biol.* **3**: 3-16.
- Sanford, J.C.; Smith, F.D. e Russell, J.A. (1993). **Optimizing the biolistic process for different biological applications.** *Method. Enzymol.* **217**: 483-509.
- Sanger, F.; Nicklen, S. e Coulson, A.R. (1977). **DNA sequencing with chain-termination inhibitors.** *Proc. Natl. Acad. Sci. USA* **74**: 5463-5467.
- Sargent, E. (1900). **Recent works on the results of fertilization in angiosperms.** *Ann. Bot.* **14**: 689-712.
- Schindler, U.; Menkens, A.E.; Beckmann, H.; Ecker, J.R. e Cashmore, A.R. (1992). **Heterodimerization between light-regulated and ubiquitously expressed Arabidopsis GBF bZIP proteins.** *EMBO J.* **22**: 1261-1273.
- Schleif, R. (1988). **DNA binding by proteins.** *Science* **241**: 1182-1187.
- Schmidt, R.J.; Burr, F.A. e Burr, B. (1987). **Transposon tagging and molecular analysis of the maize regulatory locus *opaque-2*.** *Science* **238**: 960-963.
- Schmidt, R.J.; Burr, F.A.; Aukerman, M.J. e Burr, B. (1990). **Maize regulatory gene *opaque-2* encodes a protein with a "leucine-zipper" motif that binds to zein DNA.** *Proc. Natl. Acad. Sci. USA* **87**: 46-50.
- Schmidt, R.J.; Ketudat, M.; Aukerman, M.J. e Hoschek, G. (1992). **Opaque-2 is a transcriptional activator that recognizes a specific target site in 22 kD zein genes.** *Plant Cell* **4**: 689-700.

- Schwarz-Sommer, Z.; Hue, I.; Huijser, P.; Flor, P.J.; Hansen, R.; Tetens, F.; Lonig, W.E.; Saedler, H. e Sommer, H. (1992). **Characterization of the Antirrhinum floral homeotic MADS-box gene *deficiens*: evidence for DNA binding and autoregulation of its persistent expression throughout flower development.** *EMBO J.* **11**: 251-2263.
- Sessa, G.; Morelli, G. e Ruberti, I. (1993). **The Athb-1 and -2 HD-Zip domains homodimerize forming complexes of different DNA binding specificities.** *EMBO J.* **12**: 3507-3517.
- Shaafhawsen, R.V. (1952). **Adlay or Job's tears - a cereal of potentially greater economic importance.** *Econ. Bot.* **6**: 216-227.
- Shaw, C.H.; Carter, G.H. e Watson, M.D. (1984). **A functional map of the nopaline synthase promoter.** *Nucl. Acids Res.* **12**: 7831-7846.
- Shewry, P.R. (1995). **Plant storage proteins.** *Biol. Rev.* **70**: 375-426.
- Shewry, P.R. e Tatham, A.S. (1990). **The prolamin storage proteins of cereal seeds: structure and evolution.** *Biochem. J.* **267**: 1-12.
- Shiel, P.J.; Alrefai, R.H.; Domier, L.L.; Korban, S.S. e Berger, P.H. (1995). **The complete nucleotide sequence of apple mosaic virus RNA-3.** *Archives of Virology* **140**: 1247-1256.
- Shuai, K.; Stark, G.R.; Kerr, I. e Darnell, J.E.J. (1993). **A single phosphotyrosine residue of Stat91 required for gene activation by interferon-gamma.** *Science* **261**: 1744-1746.
- Shull, J.M. e Kirleis, A.W. (1989). **Localization of alfa and beta-kafirins in *Sorghum bicolor*.** *Cereal Foods World* **34**: 784
- Singh, K.; Dennis, E.S.; Ellis, J.G.; Llewellyn, D.J.; Tokuhisa, J.G.; Wahleithner, J.A. e Peacock, W.J. (1990). **OCSBF-1, a maize ocs enhancer binding factor: isolation and expression during development.** *Plant Cell* **2**: 891-903.
- Singletary, G.W.; Doehlert, D.C.; Wilson, C.M.; Muhitch, M.J. e Below, F.E. (1990). **Response of enzymes and storage proteins of maize endosperm to nitrogen suply.** *Plant Physiol.* **94**: 858-864.
- Small, E. (1984). Hybridization in the domesticated-weed-wild-complex. *in Plant Biosystematics* Grant, W.F.ed.Academic Press - London. pp 195-210.
- So, J. e Larkins, B.A. (1991). **Binding of an endosperm-specific nuclear protein to a maize beta-zein gene correlates with zein transcriptional activity.** *Plant Mol. Biol.* **17**: 309-319.
- Speulman, E. e Salamini, F. (1995). **A barley cDNA clone with homology to the DNA-binding domain of the steroid hormone receptors.** *Plant Sci.* **106**: 91-98.
- Stanojevic, D.; Hoey, T. e Levine, M. (1989). **Sequence-specific DNA-binding activities of the gap proteins encoded by hunchback and Krueppel in *Drosophila*.** *Nature* **341**: 331-335.
- Stark, D.M.; Timmerman, K.P.; Barry, G.F.; Preiss, J. e Kishore, G.M. (1992). **Regulation of the amount of starch in plant tissues by ADP-glucose pyrophosphorylase.** *Science* **258**: 287-292.
- Struhl, K. (1994). **Duality of TBP, the universal transcription factor.** *Science* **263**: 1103-1104.
- Swaffield, J.C.; Melcher, K. e Johnston, S.A. (1995). **A highly conserved ATPase protein as a mediator between acidic activation domais and the TATA-binding protein.** *Nature* **374**: 88-91.

- Tae, H.; Luo, X. e Kim, K. (1994). **Roles of CCAAT/enhancer-binding protein and its binding site on repression and derepression of acetyl-CoA carboxylase gene.** *J. Biol. Chem.* **269**: 10475-10484.
- Takatsuji, H.; Mori, M.; Benfey, P.N.; Ren, L. e Chua, N. (1992). **Characterization of a zinc finger DNA-binding protein expressed specifically in Petunia petals and seedlings.** *EMBO J.* **11**: 241-249.
- Takatsuji, H.; Nakamura, N. e Katsumoto, Y. (1994). **A new family of zinc finger proteins in petunia: structure, DNA sequence recognition, and floral organ-specific expression.** *Plant Cell* **6**: 947-958.
- Targon, M.L.P.N.; Ottoboni, L.M.M.; Leite, A.; Ludevid, D.; Puigdomenech, P. e Arruda, P. (1992). **Synthesis and deposition of coixin in seeds of *Coix lacryma-jobi*.** *Plant Sci.* **83**: 169-180.
- Thompson, G.A.; Boston, R.S.; Lyznik, L.A.; Hodges, T.K. e Larkins, B.A. (1990). **Analysis of promoter activity from an alpha-zein gene 5' flanking sequence in transient expression assays.** *Plant Mol. Biol.* **15**: 755-764.
- Tjian, R. (1995). **Molecular machines that control genes.** *Scientific American* **272**: 38-45.
- Tjian, R. e Maniatis, T. (1994). **Transcriptional activation: a complex puzzle with few easy pieces.** *Cell* **77**: 5-8.
- Tokuhisa, J.G.; Singh, K.; Dennis, E.S. e Peacock, W.J. (1990). **A DNA-binding protein factor recognizes two binding domains within the octopine synthase enhancer element.** *Plant Cell* **2**: 215-224.
- Torres, P. e Bergamin, A. (1951). **O cereal Adlay. Estudo do seu valor agrícola e nutricional para aves.** *Anais da Escola Superior "Luiz de Queiroz"* **8**: 669-685.
- Tosello, G.A. (1987). Milhos especiais e seu valor nutritivo. *in* **Melhoramento e produção de milho** Paterniani, E. e Viégas, G.P. eds. Fundação Cargil - Campinas-S.P. pp 375-402.
- Töpfer, R.; Martini, N. e Schell, J. (1995). **Modification of plant lipid synthesis.** *Science* **268**: 681-686.
- Tronche, F. e Yaniv, M. (1992). **HNF1, a homeoprotein member of the hepatic transcription regulatory network.** *BioEssays* **14**: 579-587.
- Ueda, T.; Wawerczak, W.; Ward, K.; Sher, N.; Ketudat, M.; Schmidt, R.J. e Messing, J. (1992). **Mutations of the 22- and 27-kD zein promoters affect transactivation by the Opaque-2 protein.** *Plant Cell* **4**: 701-709.
- Ueda, T.; Wang, Z.; Pham, N. e Messing, J. (1994). **Identification of a transcriptional activator-binding element in the 27-kilodalton zein promoter, the -300 element.** *Mol. Cell Biol.* **14**: 4350-4359.
- Ulmasov, T.; Hagen, G. e Guilfoyle, T. (1994). **The ocs element in the soybean GH2/4 promoter is activated by both active and inactive auxin and salicylic acid analogues.** *Plant Mol. Biol.* **26**: 1055-1064.
- Utsumi, S. (1992). **Plant food protein engineering.** *Adv. Food Nutr. Res.* **36**: 89-208.
- van Doren, M.; Bailey, A.M.; Esnayra, J.; Ede, K. e Posakony, J.W. (1994). **Negative regulation of proneural gene activity: hairy is a direct transcriptional repressor of achaete.** *Genes Dev.* **8**: 2729-2742.

- Venkateswarlu, J. e Chaganti, R.S.K. (1973). **Job's tears (*Coix lacryma-jobi* L.)**. *ICAR Tech. Bull. (Agric.)* **43/44**: 1-54.
- Vettore, A.L. (1994). **Clonagem e caracterização do fator de transativação Opaco2 de *Coix lacryma-jobi***. I.B. - UNICAMP - Tese de Mestrado.
- Vogel, J.M.; Roth, B.; Cigan, M. e Freeling, M. (1993). **Expression of the two maize TATA binding protein genes and function of the encoded TBP proteins by complementation in yeast**. *Plant Cell* **5**: 1627-1638.
- von Kalm, L.; Crossgrove, K.; von Seggern, D.; Guild, G.M. e Beckendorf, S.K. (1994). **The Broad-Complex directly controls a tissue-specific response to the steroid hormone ecdysone at the onset of *Drosophila* metamorphosis**. *EMBO J.* **13**: 3505-3516.
- Wagner, S. e Green, M.R. (1994). **DNA-binding domains: targets for viral and cellular regulators**. *Curr. Opin. Cell Biol.* **6**: 410-414.
- Wallace, J.C.; Lopes, M.A.; Paiva, E. e Larkins, B.A. (1990). **New methods for extraction and quantitation of zeins reveal a high content of γ -zein in modified opaque-2 maize**. *Plant Physiol.* **92**: 191-196.
- Walter, M.H.; Grima-Pettenati, J. e Feuillet, C. (1994). **Characterization of a bean (*Phaseolus vulgaris* L.) malic-enzyme gene**. *Eur. J. Biochem.* **224**: 999-1009.
- Wandelt, C. e Feix, G. (1989). **Sequence of a 21 kDa zein gene from maize containing an in frame stop codon**. *Nucl. Acids Res.* **17**: 2354
- Wang, S. e Esen, A. (1986). **Primary structure of a prolamine-rich zein and its cDNA**. *Plant Physiol.* **81**: 70-74.
- Watson, J.D. (1993). **Looking forward**. *Gene* **135**: 309-315.
- Weaver, L.M.; Yu, F.; Wurtele, E.S. e Nikolau, B.J. (1996). **Characterization of the cDNA and Gene coding for the biotin synthase of *Arabidopsis thaliana***. *Plant Physiol.* **110**: 1021-1028.
- Weisshaar, B.; Armstrong, G.A.; Block, A.; da Costa e Silva, O. e Hahlbrock, K. (1991). **Light-inducible and constitutively expressed DNA-binding proteins recognizing a plant promoter element with functional relevance in light responsiveness**. *EMBO J.* **10**: 1777-1786.
- Wey, E. e Schafer, B.W. (1996). **Identification of novel DNA binding sites recognized by the transcription factor mPOU (POU6F1)**. *Biochem. Biophys. Res. Commun.* **220**: 274-279.
- Whittaker, P.A.; Campbell, A.J.; Southern, E.M. e Murray, N.E. (1988). **Enhanced recovery and restriction mapping of DNA fragments cloned in a new lambda vector**. *Nucl. Acids Res.* **16**: 6725-6736.
- Winspear, M.J.; Preston, K.R.; Rastogi, V. e Oaks, A. (1984). **Comparisons of peptide hydrolase activities in cereals**. *Plant Physiol.* **75**: 480-482.
- Wolffe, A.P. (1994a). **Nucleosome positioning and modification: chromatin structures that potentiate transcription**. *Trends Biochem. Sci.* **19**: 240-244.
- Wolffe, A.P. (1994b). **Architectural transcription factors**. *Science* **264**: 1100-1101.
- Wu, L.; Ueda, T. e Messing, J. (1993). **3'-end processing of the maize 27 kDa zein mRNA**. *Plant J.* **4**: 535-544.

- Wu, L.; Ueda, T. e Messing, J. (1995). **The formation of mRNA 3'-ends in plants.** *Plant J.* **8**: 323-329.
- Xi, X.Y. e Ye, B.X. (1994). **Studies on embryo development and deposition of storage reserves in *Coix lacrima-jobi*.** *Acta Botanica Sinica* **36**: 573-580.
- Yamamoto, K.R.; Alberts, B.M.; Benzinger, R.; Lawhorne, L. e Treiber, G. (1970). **Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification.** *Virology* **40**: 734
- Yankulov, K.; Blau, J.; Purton, T.; Roberts, S. e Bentley, D.L. (1994). **Transcriptional elongation by RNA polymerase II is stimulated by transactivators.** *Cell* **77**: 749-759.
- Yoshinaga, S.K.; Peterson, S.L.; Herskowitz, I. e Yamamoto, K.R. (1992). **Roles of SWI1, SWI2, and SWI3 proteins for transcriptional enhancement by steroid receptors.** *Science* **258**: 1598-1604.
- Young, R.A. (1991). **RNA polymerase II.** *Annu. Rev. Biochem.* **60**: 689-715.
- Yunes, J.A.; Cord-Neto, G.; Silva, M.J.; Leite, A.; Ottoboni, L.M.M. e Arruda, P. (1994). **The transcriptional activator Opaque2 recognizes two different target sequences in the 22-kD-like α -prolamin genes.** *Plant Cell* **6**: 237-249.
- Zhu, Q.; Dabi, T. e Lamb, C. (1995). **TATA box and initiator functions in the accurate transcription of a plant minimal promoter in vitro.** *Plant Cell* **7**: 1681-1689.