BC/31046
IB/80398

UNICAMP

T/Unicamp

Y92i

SECRETÁRIA
DE
PÓS GRADUAÇÃO
I. B.

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

Interação da proteína Opaco2 com o promotor de um gene de α-coixina: especificidade e aspectos termodinâmicos da interação proteína-DNA

Este exemplar corresponde à redação final da tese defendida pelo (a) candidato (a)
[illegible]
e aprovada pela Comissão Julgadora
30/04/97

José Andrés Yunes
Orientador: Dr. Paulo Arruda

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

Campinas - SP
1997

UNIDADE IB
N.º CHAMADA:
T/UNICAMP
Y92i
V. Ex.
TOMBO BC/ 31046
PROC. 281/97
C [ ] D [X]
PREÇO R$ 11,00
DATA 12/07/97
N.º CPD

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE BIOLOGIA - UNICAMP

Yunes, José Andrés

Y92i Interação da proteína Opaco2 com o promotor de um gene de α-coixina: especificidade e aspectos termodinâmicos da interação proteína-DNA/José Andrés Yunes. -- Campinas, SP:[s.n.], 1997. 106f: ilus.

Orientador: Paulo Arruda

Tese (doutorado) - Universidade Estadual de Campinas, Instituto de biologia.

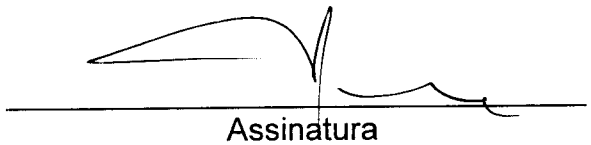
1. Sementes - Proteínas. 2. Plantas - Genética molecular. 3. Fatores de transcrição. 4. Coix. I. Arruda, Paulo. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

LOCAL E DATA: Campinas, 30 de Abril de 1997

BANCA EXAMINADORA:

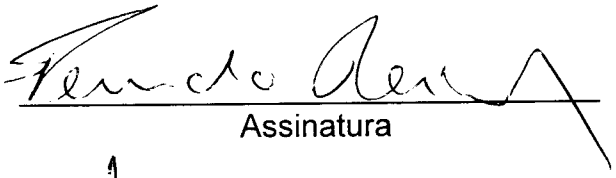
TITULARES:

Prof. Dr. Fernando F. Costa



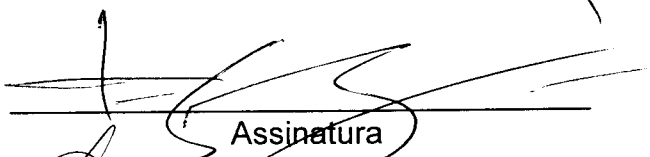
Assinatura

Prof. Dr. Fernando C. Reinach



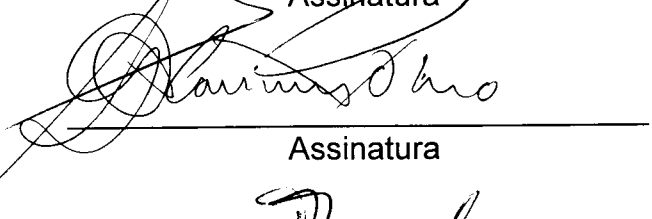
Assinatura

Prof. Dr. Francisco G. Nobrega



Assinatura

Prof. Dr. Glaucius Oliva



Assinatura

Prof. Dr. Paulo Arruda (Orientador)

Assinatura

SUPLENTES:

Dr. Éberson S. Calvo

Assinatura

Prof. Dr. Gonçalo A. G. Pereira

Assinatura

APROVADA

"Que quero dizer-vos? Quero dizer-vos - não vos riais de mim - que não sei donde é que vim até aqui, a esta vida mortal ou, se preferis, a esta morte vital. Não o sei; somente sei que, ao vir a esta vida, receberam-me as carícias de meus pais" Sto. Agostinho

Quando era criança, me lembro, fazendo alguma "proeza" e gritando alto paaapi, mamaaa, olha aqui. A necessidade de carinho, de atenção, mas é também sinal daquilo que somos: *para os outros*. Hoje lhes dou um pouco daquilo que sou, ou melhor, daquilo que vocês são em mim. E é isso mesmo, quem me vê, vê voces. Claro, talvez meio desarrumado!

E também ao Pablo, Santiago, Virginia, Yamile e Cristina, filhos da mesma ninhada. Familia abençoada. Vida boa, bagunçada, divertida, nunca parada. Quantas histórias! Quanto ainda pela frente! Caminhando, esperando, caindo, levantando, brigando, ajudando, sempre em frente, na certeza de estarmos juntos...

...e no caminho da paz guiar nossos passos.

Um dia, no avião, na época em que eu ainda tinha medo do mar, tu me ensinastes a rezar o Pai Nosso. E eu, que não tenho boa memória, lembro disso. Oh semente da boa que tu plantastes e ajudastes a cuidar! Como esquecer de ti. Bem sei do teu silêncio, da tua humildade, da tua pobreza, mas deixo aqui o teu nome, Maria Bonoris, madrinha.

Agradeço:

Ao Prof. Paulo Arruda, pelos mais de 8 anos de orientação. Por ter sempre me incentivado, com liberdade, na condução dos trabalhos. Pelas inúmeras oportunidades de participação em congressos e projetos de pesquisa. Hoje também agradeço muito o fato dele ter sido sempre exigente. A sua amizade e entusiasmo, a sua preocupação com o futuro da ciência e educação brasileiras, guardo como exemplo a ser seguido.

Ao Adilson Leite, pelo ensino e orientação "molecular", pelas oportunidades de trabalho, pelas divertidas discussões filosóficas, teológicas, metafísicas e do além. Pela correção da tese na pré-banca.

À Laura M.M. Ottoboni, por ter partilhado o clone do gene da α-coixina.

Ao meu querido amigo Germano Cord Neto, o "cordeiro", pela companhia e ajuda de tanto tempo. Na beleza da simplicidade a palavra "amigo" diz tudo.

Ao Márcio J. da Silva, messie le Marcion, por um ano de carona, pelos ensaios de expressão transitória, pelas várias e criativas dicas na condução dos experimentos (quem o conhece, sabe o que quero dizer). Por ter me ensinado a relativisar o trabalho.

Ao André L. Vettore, o "Andreinha", pelos experimentos de "gel shift", feitos com precisão e sem demora. Foi muito bom trabalhar junto. E aproveito para agradecer à Berenice, pelos inumeraveis geis de sequenciamento e alegres conversas.

À minha afilhada Nilda D. Montes Villanueva, que me salvou na hora da regressão não-linear. Muito obrigado. E também à Lurdes Contreras Castillo, que me ensinou a programar no SAS.

Ao Gonçalo A. Souza Filho, pela caracterização dos clones de α-coixina.

Ao amigão, não só pelo tamanho, Edson L. Kemper, que me ajudou na construção de alguns vetores de expressão e muitas vezes nos experimentos de

transformação de protoplastos. O Paulo de Luca também me ajudou na preparação de alguns DNAs CsCl. Obrigado Luquinha.

Ao meu querido amigo Pedro Mauricio, pelo enorme trabalho de ajuda a nós todos, no silencio, no capricho.

Ao Sr. Chico, que cuida dos campos de milho.

Ao Felipe R. da Silva, profeta do World Wide English, pela inestimável contribuição na introdução desta tese, pelas tardes juntos, lado a lado na frente da telinha, escutando música e falando à beça. Pela ajuda com os computadores.

Ao Marcelo Menossi, pela ajuda e amizade desde os primeiros anos aqui em Campinas. Pelos trabalhos em conjuntos, uns loucos, outros futuristas. Pela correção da tese na pré-banca.

Ao Prof. Gonçalo A. Guimarães Pereira, pela correção da tese na pré-banca.

Ao Fabio Papes, que eivou esta tese de correções, contribuindo significativamente para que ficasse mais fácil de ler e entender.

Aos colegas do Fundão de Deus: Fernando A. Freitas, Xico (Antônio Francisco Ferreira Neto), Ricardo A. Dante, Luciano A. Glavina, Edson L. Kemper e Ivan Maia, pela alegre e tranquila convivência no laboratório. E em memória de Mário Lopes Junior, um amigo que espero reencontrar.

À querida Silvia R. Turcinelli, pelas tantas ponteiras, soluções,...

Aos queridos amigos que tornam (ou tornavam) tão agradável vir trabalhar, e pelos favores constantes. Eu sei que muitos acham que aqui é o lugar da vala comum, mas vocês sabem que não é bem assim. É que eu escolhi agradecer separadamente o que de algum modo estava relacionado com o trabalho de tese, senão não acabava mais: Celso, Cintia, Adriana, Karen, Susana, Rodrigo, Alba, Éberson, Marymar, Kcal, Tita, Emi, Agostinho, David.

Aos amigos do CBMEG, alunos e professores, que se eu fosse querer citar, com certeza acabaria por esquecer algum.

A ajuda sempre presente do pessoal da secretaria do CBMEG, Sandra, Cláudia, Tânia e Fabiana. E às secretárias do departamento de Genética e Evolução, Teresa, Silvia e Célia, pela atenção e auxílios prestados.

Aos funcionários Sr. Manoel, Joãozinho e Chico.

Ao CNPq, que desde a iniciação científica (~10 anos atrás) vem me apoiando financeiramente.

A este Brasil querido, que me acolheu, me ensinou e me adotou como filho. Como diz o Hino: *...és mãe gentil, pátria amada Brasil.*

A Cesar Felski (capivara) e Zolmar Sinhorin, que me receberam e hospedaram durante o primeiro ano em Campinas, e fizeram todo o possível para que eu pudesse estudar tranquilo. Um forte abraço.

À minha família republicana: Chita Maria (minha filha canina), Jorge (SJ), Dudú Nicola, Kcal (Cláudia, ex-ex-namorada), Maria Elena (Lenin, Zeperina), Guilherme, Gonçalo (Gogo), Ana Paula, Dra. Ana Paula, Eric (holandes), Eliane, Patricia, Nancy, Luzão, Edson, Edson "baiano", Lilian, Tito, Camilo, Bia, Brochi, Dra. Ana Maria, Alessandra, Germano, Luiz ("Zé banherinho"), Sergio ("Filisbin"), Dener (irmão de coração e alegria), e por último, cheio de alegria, com muito carinho, um beijo grande na Fagoni, namorada, amiga, recreio dos meus sentidos, doçura pro meu coração.

Então... ☺ após ter agradecido a vocês, só me falta agradecer Quem me faz respirar a todo instante. Mas não sei como! Foi tanto o que Ele me deu que vocês nem imaginam. Não somente em termos subjetivos, no sentido de coragem, alegria, criatividade, paciência e essas coisas, mas também muito concretamente, através de vocês todos e às vezes de maneira inesperada, misteriosa se preferirem, porém insisto, muito concretamente, no dia a dia do laboratório. Talvés o melhor seja dizer: Gloria ao Pai, ao Filho e ao Espirito Santo, como era no princípio, agora e sempre!

ÍNDICE

RESUMO ix

SUMMARY xii

INTRODUÇÃO 1

1. Proteínas de reserva da semente 1

2. Proteínas de reserva da semente de *Coix* 3

3. As α-coixinas de 25 kDa em *Coix* correspondem às α-zeinas de 22 kDa de milho 4

4. Biossintesse das proteínas de reserva da semente 5

5. Regulação transcricional dos genes de prolaminas 6

6. Transcrição gênica em eucariótos 6

   6.1. Estrutura de cromatina 7

   6.2. Metilação 7

   6.3. Interação com fatores de transcrição 8

      6.3.1. As sequências de DNA envolvidas na transcrição 9

      6.3.2. As proteínas envolvidas no processo de transcrição 10

         6.3.2.1. Fatores gerais de transcrição 11

         6.3.2.2. Fatores reguladores da transcrição 12

      6.3.3. Mecanismos de ativação via fatores reguladores da transcrição 12

7. Genes reguladores da biossíntese das proteínas de reserva 15

8. O locus *Opaco2* codifica uma proteína regulatória da família bZIP 16

9. Existência de um gene tipo *Opaco2* em *Coix* e sorgo 18

10. A proteína Opaco2 é capaz de se ligar a diferentes sequências de DNA 18

APRESENTAÇÃO DOS TRABALHOS E OBJETIVOS 22

Trabalho 1 25

Trabalho 2 38

DISCUSSÃO 89

CONCLUSÕES 95

BIBLIOGRAFIA 97

RESUMO

As proteínas de reserva da semente dos cereais podem ser agrupadas em diferentes classes de polipeptídeos relacionados entre si, cujos genes apresentam características estruturais e regulatórias semelhantes. Estes genes são abundantemente transcritos, num padrão de expressão coordenado espacial e temporalmente. A regulação transcricional parece ser um dos principais mecanismos no controle desta expressão. Vários destes genes já foram clonados e mutações que afetam uma certa classe das proteínas de reserva foram caracterizadas. Isto torna o sistema das proteínas de reserva um modelo muito apropriado para o estudo da regulação da expressão gênica em plantas. Uma das mutações que afeta a síntese de proteínas de reserva em milho, a *opaco2* (*o2*), reduz o acúmulo de várias proteínas, entre elas uma albumina denominada b-32 e as classes das α- e β-zeínas. O gene *Opaco2* (*O2*) foi caracterizado como sendo um *locus* regulatório que codifica um fator da transcrição da família dos "zíper de leucinas" (bZIP), capaz de controlar distintas classes de genes expressos no endosperma, através do reconhecimento de elementos *cis*-regulatórios significativamente distintos.

Neste estudo, demonstrou-se que o gene da α-coixina de 25 kDa, uma α-prolamina de *Coix*, cereal filogeneticamente aparentado ao milho, pode ser transativado pela proteína Opaco2 (O2). Ensaios de ligação *in vitro* e experimentos de "footprint" mostraram que a sequência 5'-AAAATTGACTAGGAGACATGTCATCTCTA-3' do promotor da α-coixina é reconhecida e protegida, tanto pela proteína O2 de milho quanto pela de *Coix*. Interessante notar que esta região *cis* para O2 não apresenta o motivo central ACGT, postulado como necessário para a ligação eficiente das bZIPs de plantas.

Através de ensaios de retardamento em gel, usando uma proteína O2 de *Coix* recombinante e oligonucleotídeos correspondentes às regiões *cis* inteira ou

truncada, demonstrou-se a existência de dois sítios de ligação localizados em posições adjacentes, denominados *O2u* (5'-TTGACTAGGA-3') e *O2d* (5'-GACATGTCAT-3'). Ambos os sítios apresentam similaridade com o sítio de ligação da proteína GCN4, podendo ser simultaneamente reconhecidos pela O2.

Diferentes mutações foram criadas na região *cis* e analisadas através de "footprint" quantitativo. Observou-se assim, que a ligação da O2 aos sítios adjacentes é melhor descrita por um modelo cooperativo. Analisando simultaneamente os dados numéricos do "footprint" quantitativo das sequências *cis* normal e mutantes, calcularam-se constantes termodinâmicas relativas para a ligação da proteína O2 a cada sítio e para a cooperatividade. Esta análise indicou que a afinidade intrínseca de ligação da proteína O2 a cada um dos sítios é semelhante. Algumas mutações afetaram a cooperatividade enquanto outras não, indicando que esta não depende exclusivamente da afinidade de ligação da proteína ao DNA.

Por ensaios de expressão transitória em endosperma imaturo de milho e protoplastos de mesofilo de tabaco, foi demonstrado que a sequência *cis* completa é necessária para mediar a transativação por O2 do gene indicador β-glucuronidase (*gus*) sob controle do promotor da α-coixina de 25 kDa. A análise das diferentes sequências *cis* mutantes nos experimentos de expressão transitória em protoplastos de tabaco mostrou que os sítios *O2u* e *O2d* atuam sinergisticamente. A comparação dos resultados de atividade transcricional com os de energia livre relativa obtidos para cada uma das diferentes sequências *cis* sugere que o sinergismo é decorrente, em parte, da ligação cooperativa ao DNA. Por outro lado, a ligação da proteína O2 ao sítio *O2u* tem maior efeito nos níveis finais de transcrição se comparado com a ligação ao sítio *O2d*. Desta maneira, parte do sinergismo pode estar relacionado ao processo de interação do complexo O2-DNA com o complexo geral de iniciação da transcrição.

Do ponto de vista evolutivo, a ocorrência destes dois sítios adjacentes, de sequência distintas, e a ligação cooperativa da O2 podem ser devidos a uma necessidade de flexibilidade no reconhecimento de sítios no DNA por proteínas regulatórias.

SUMMARY

The storage proteins of cereal seeds are composed of different classes of related proteins encoded by genes with common regulatory and structural features. These genes are abundantly transcribed and show a spatial and temporal coordinated expression pattern. Transcriptional regulation seems to be one of the major mechanisms controlling the expression of the storage protein genes. Because several of these genes have been cloned and some mutations affecting certain protein classes have been identified, the seed storage protein system is an useful model to study the regulation of gene expression in plants. One of the mutations affecting maize storage protein synthesis, the *opaque2 (o2)*, reduces the accumulation of several proteins, including an albumin named b-32 and the α– and β–zeins. The *Opaque2 (O2)* gene has been characterized as a regulatory *locus* that encodes a transcriptional factor. The Opaque2 protein (O2) is a leucine zipper (bZIP) transcriptional factor, which controls the expression of distinct classes of endosperm genes through recognition of significantly different *cis*-acting elements in their promoters.

We show here that the gene encoding for the 25 kDa α-coixin, the α-prolamin of the maize-related grass *Coix*, can also be transactivated by the O2 protein. Binding assays *in vitro* and footprint analysis demonstrated that the 5'-AAAATTGACTAGGAGACATGTCATCTCTA-3' sequence of the α–coixin promoter can be recognized and protected by both the maize and *Coix* O2 proteins. Interestingly, this O2 target region does not contain an ACGT core sequence, which has been postulated to be necessary for efficient binding of plant bZIP proteins.

In order to further characterize this *cis*-acting sequence, we performed gel shift assays using a recombinant *Coix* O2 protein and two oligonucleotide probes containing either the whole or a truncated O2 target sequences. Our results indicate that the α-coixin O2 target sequence comprise two closely adjacent

binding sites, which we named *O2u* (5'-TTGACTAGGA-3') and *O2d* (5'-GACATGTCAT-3'). Both sites are related in sequence to the GCN4 binding site, and can be simultaneously bound by O2.

Quantitative DNase I footprint analysis using a series of DNA templates having a mutation in one or both sites, or disjointing sites, indicated that O2 binding to the α-coixin target sites is best described by a cooperative model. Simultaneous numerical analysis of footprint titration data from the normal and mutant templates were used to determine relative thermodynamic constants for local site binding and cooperativity. This analysis indicated that O2 interacts with similar intrinsic affinity with both sites. Some mutations affected cooperativity, while others did not, indicating that cooperative binding is based on other properties in addition to primary sequence recognition.

Employing transient expression experiments in immature maize endosperm and tobacco mesophyll protoplasts, we demonstrated that the entire *cis*-acting sequence is necessary in mediating the O2 activation of expression of the β–glucuronidase (*gus*) reporter gene placed under the control of the 25 kDa α-coixin promoter. By analyzing the different mutated templates in the transient tobacco protoplast expression assay, we found that the two adjacent sites act synergistically, but binding at the O2u site is more important for enhancer activity. Comparisons of the transcriptional activities with the resolved relative binding free energies for the various templates, suggest that this synergy is mediated both by cooperative DNA binding and by the interaction of the O2-DNA complex with the transcriptional machinery.

It is suggested that the occurrence of cooperative binding to two adjacent sites, different in sequence, may derive from an evolutionary requirement of flexibility in binding site recognition.

INTRODUÇÃO

O crescimento e desenvolvimento dos organismos, a diferenciação dos vários tecidos e orgãos e a resposta a estímulos extracelulares requerem tanto a presença de genes capazes de codificar proteínas estruturais e funcionais, quanto a existência de mecanismos de regulação da expressão destes genes. Tais mecanismos regulatórios asseguram que aquelas proteínas sejam produzidas em quantidade adequada e no momento correto com relação ao ciclo celular e ao processo de desenvolvimento.

A regulação do nível de expressão de certos genes é de tal modo importante para o organismo que, por exemplo, a asma, o câncer, doenças cardíacas e imunes e infecções virais são o resultado da super ou subprodução de uma ou poucas proteínas (Tjian, 1995).

Durante o desenvolvimento da semente de plantas superiores, grande variedade de proteínas é produzida e muitas são suprimidas, proporcionando um excelente sistema para o estudo dos diferentes aspectos da regulação gênica.

1. Proteínas de reserva da semente

A fração protéica da semente contém uma ampla variedade de proteínas pouco abundantes que são essenciais na manutenção do metabolismo celular, e uma outra categoria composta por um número relativamente pequeno de proteínas armazenadas em grandes quantidades, cuja principal função é fornecer nitrogênio e aminoácidos à plântula em germinação (Higgins, 1984).

Além de não apresentarem qualquer outra função conhecida, as proteínas de reserva caracterizam-se por uma composição peculiar de aminoácidos, sendo ricas em asparagina, glutamina e arginina ou prolina, mas deficientes em lisina, treonina e triptofano, nos cereais, e em cisteína, metionina e triptofano, nas leguminosas.

Também é característico o fato de que proteínas de reserva são acumuladas em pequenas organelas esféricas unidas à membrana do retículo endoplasmático (RE), com poucos *micra* de diâmetro, denominadas de corpúsculos protéicos (Pernollet e Mossé, 1983).

Para tanto, as proteínas de reserva são sintetizadas em polissomos unidos à membrana e contém uma pequena sequência aminoterminal, chamada peptídeo sinal, que facilita a translocação do polipeptídeo nascente para o lúmem do RE (Kreil, 1981). No início da tradução os polissomos atracam-se ao RE e o peptídeo sinal atravessa a membrana. A seguir, enquanto a síntese do restante do polipeptídeo continua, o peptídeo sinal é clivado por uma endopeptidase (Muntz *et al.*, 1985).

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

- Albuminas: solúveis em água;
- Globulinas: solúveis em solução salina;
- Prolaminas: solúveis em solução alcoólica; e
- Glutelinas: solúveis em soluções alcalinas e ácidas diluídas, e em soluções salinas ou alcoólicas que contenham um agente redutor.

Nos cereais - com exceção da aveia, onde 80% das proteínas da semente são globulinas, e do arroz, onde 85% são glutelinas - as prolaminas representam aproximadamente 50% das proteínas da semente e as glutelinas ao redor de 40% (Larkins, 1981).

As prolaminas geralmente são monoméricas, têm composição característica de aminoácidos (ricas em prolina e glutamina, e pobres em aminoácidos básicos como lisina) e são comumente agrupadas em classes de acordo com seus pesos moleculares (Pernollet e Mossé, 1983).

Além disso, as prolaminas recebem nomes específicos que indicam a planta de origem. Assim, por exemplo, as prolaminas de milho (*Zea mays*) são

chamadas zeínas, as de trigo (*Triticum aestivum*), gliadinas, as de cevada (*Hordeum vulgare*), hordeínas e as de *Coix* (*Coix lacryma-jobi*), coixinas.

As glutelinas formam grandes agregados através de pontes de dissulfeto e são um grupo bastante heterogêneo, sendo, por isso, particularmente difícil de caracterizar (Larkins, 1981). Certos subgrupos de glutelinas mostraram ser prolaminas polimerizadas (Pernollet e Mossé, 1983).

Devido à sua contribuição para a qualidade nutritiva da semente, estas proteínas e seus genes têm sido motivo de intenso estudo em várias espécies.

Sendo a síntese destas proteínas semente-específica e muito intensa num período delimitado do desenvolvimento da semente, os genes correspondentes oferecem também um interessante modelo para o estudo dos mecanismos de regulação da expressão gênica.

2. Proteínas de reserva da semente de *Coix*

Em *Coix lacryma-jobi*, var. Adlay, cereal pertencente à mesma tribo do milho (Andropogoneae) o endosperma contém aproximadamente 20% de proteínas, entre albuminas, globulinas, prolaminas e proteínas residuais (Ottoboni *et al.*, 1990). Esta é uma das maiores proporções encontradas em cereais, duas vezes maior do que a observada em sementes de milho.

As coixinas representam mais de 70% do conteúdo protéico do endosperma. Através de eletroforese em gel de poliacrilamida contendo dodecil sulfato de sódio (SDS-PAGE), as coixinas podem ser separadas em 5 componentes com pesos moleculares de 27, 25, 22, 17,5 e 15 kilo Daltons (kDa), chamados de C1, C2, C3, C4 e C5, respectivamente (Ottoboni *et al.*, 1990).

Análises sequenciais utilizando focalização isoelétrica (IEF) seguidas de SDS-PAGE revelaram que as classes C1, C2, C3, C4 e C5 incluem 4, 5, 1, 1 e 2 polipeptídios, respectivamente (Ottoboni *et al.*, 1990).

Esen (1986) estabeleceu as condições para a divisão das prolaminas em função da solubilidade em soluções aquosas. Trabalhando com zeínas, definiu três classes de solubilidade, posteriormente denominadas de α-, β- e γ-zeínas.

Pelo procedimento de Esen (Esen, 1986), as coixinas podem ser divididas em duas frações: α- e γ-coixinas. A fração das α-coixinas é solúvel em 40-95% de isopropanol, corresponde a 80% das coixinas totais e inclui as classes C1, C2, C4 e C5. A fração das γ-coixinas, é solúvel em 0-80% de isopropanol na presença de agentes redutores (β-Mercapto etanol, Ditiotreitol) e corresponde unicamente à classe C3 (Ottoboni *et al.*, 1990).

Posteriormente, demonstrou-se que uma das α-coixinas, a C4, era estruturalmente homóloga às β-zeínas; a partir de então esta prolamina passou a ser denominada β-coixina (Leite *et al.*,1992).

3. As α-coixinas de 25 kDa em *Coix*, correspondem às α-zeínas de 22 kDa de milho

As zeínas, quando submetidas a SDS-PAGE, são separadas em 6 componentes, com pesos moleculares de 28, 22, 19, 16, 14 e 10 kDa.

Segundo a classificação de Esen (Esen, 1986), os polipeptídeos de 19 e 22 kDa compõem as α-zeínas, os de 14 kDa, as β-zeínas e os de 16 e 28 kDa, as γ-zeínas. As zeínas de 10 kDa foram inicialmente classificadas como β-zeínas; porém, análises estruturais destes polipeptídeos levaram à inclusão de uma quarta classe, a das δ-zeínas (Kirihara *et al.*,1988).

Assim como em *Coix*, as α-zeínas constituem a maior fração das proteínas de reserva de milho, compreendendo 80% da fração total de zeínas (Shewry e Tatham, 1990).

Não somente em termos de solubilidade, como também através de hibridização cruzada de DNA e por métodos imunológicos, foi demonstrada a existência de uma alta homologia entre as coixinas e as zeínas (Leite *et al.*, 1990).

Pelo sequenciamento de um clone genômico contendo três genes de coixinas de 25 KDa, foi possível evidenciar a alta homologia existente entre estes genes e aqueles das zeínas de 22 KDa. Tal homologia foi confirmada não somente na região codificadora, mas também nas regiões regulatórias do gene (Ottoboni *et al.*, 1993). Em um estudo posterior de modelagem tridimensional comprovou-se que as α-zeínas de 22 kDa e as α-coixinas de 25 kDa são estruturalmente semelhantes entre si (Garratt *et al.*, 1993).

4. Biossíntese das proteínas de reserva da semente

Pressupõe-se que a alta quantidade de proteínas de reserva na semente dos cereais seja consequência de uma alta taxa de síntese protéica durante o seu desenvolvimento, alcançada graças ao grande número de genes, à alta eficiência de tradução e processamento do mRNA e, principalmente, à grande taxa de transcrição dos genes (Pernollet, 1985).

As proteínas de reserva, similarmente ao que ocorre com outras proteínas, são codificadas por famílias multigênicas, originadas possivelmente através de um complexo processo de duplicações gênicas. Em milho, por exemplo, as zeínas 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 sejam ativos (Feix e Quayle, 1993). Em trigo, estima-se a existência de 50 a 100 genes de gliadinas (Pernollet, 1985). Em *Coix*, a α-coixina de 25 kDa é codificada por cerca de 36 genes (Ottoboni *et al.*,1993).

Nos cereais, o acúmulo das proteínas de reserva se dá no endosperma, com a taxa máxima de biossíntese acontecendo, de modo geral, ao redor da 3ª semana após a antese, coincidindo também com a fase de expansão das células do endosperma (Pernollet, 1985). Caracteristicamente, este acúmulo segue uma cinética sigmoidal, altamente correlacionada com o ganho de peso seco da semente (Dure, 1975).

Entretanto, a análise detalhada do acúmulo das proteínas de reserva demonstrou a existência de diferentes taxas de síntese para cada uma das

proteínas durante o desenvolvimento da semente (Pernollet, 1985). Isto pôde ser exemplificado pelo processo de formação dos corpúsculos protéicos no milho, onde inicialmente verifica-se a deposição das β- e γ-zeínas; em seguida, com o amadurecimento do endosperma, dá-se o acúmulo de α-zeínas no interior dos corpúsculos protéicos, culminando com o preenchimento do seu interior e limitando as β- e γ-zeínas a uma distribuição periférica (Lending e Larkins, 1989).

Assim como as zeínas, as coixinas apresentam um padrão típico de acúmulo temporal que varia entre os diferentes componentes (Targon *et al.*, 1992).

5. Regulação transcricional dos genes de prolaminas

O acúmulo das proteínas reflete o equilíbrio estabelecido entre as taxas de síntese e de degradação. As proteínas de reserva da semente são estáveis desde a fase de deposição até a embebição e germinação da semente (Madison *et al.*, 1981); conclui-se, portanto, que é sua taxa de síntese quem principalmente determina a taxa de acúmulo (Higgins, 1984).

Por outro lado, existe uma estreita correlação entre o nível de RNA mensageiro (mRNA) e a taxa de síntese de cada proteína de reserva nos diferentes estágios de desenvolvimento do endosperma (Kodrzycki *et al.*, 1989; Targon *et al.*, 1992; Feix e Quayle, 1993). Conseqüentemente, a biossíntese destas proteínas parece ser principalmente controlada ao nível da transcrição (Higgins, 1984), sendo pouco provável a ocorrência de um controle pós-transcricional (Gallie, 1993).

6. Transcrição gênica em eucariotos

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

- Formação do complexo de pré-iniciação da transcrição, onde os fatores de transcrição reconhecem seus sítios específicos no promotor do gene;

- Iniciação, onde os ativadores da transcrição levam à ativação do complexo formado anteriormente;
- Escape do promotor (*promoter clearance*), fase na qual a RNA-polimerase II adota uma conformação alongada capaz de deixar o promotor para iniciar a transcrição propriamente dita;
- Alongamento, que representa a transcrição propriamente dita, onde ocorre a síntese do RNA a partir do molde de DNA;
- Terminação.

O primeiro requerimento para a transcrição de um gene é que o mesmo se encontre em estado ativado; para tanto, parecem exercer influência fatores como a estrutura da cromatina e o grau de metilação do DNA.

6.1. Estrutura de cromatina

Quanto à estrutura de cromatina, pressupõe-se que genes ativamente transcritos se encontrem menos protegidos pelo complexo nucleossomal, permitindo o acesso da polimerase do RNA. Até certo ponto, tal hipótese é comprovada pela maior sensibilidade dos genes transcritos ativamente à degradação do DNA por deoxirribonucleases (Apel *et al.*, 1986).

Sabe-se também que o enrolamento de seqüências específicas de DNA ao redor das histonas pode exercer um papel fundamental, tanto na ativação como na repressão da transcrição (Wolffe, 1994a).

6.2. Metilação

Vários organismos, incluindo plantas superiores, contém, na fita de DNA, bases que foram modificadas após o processo de replicação, por metilases capazes de reconhecer sequências específicas. Um crescente número de trabalhos tem discutido a importância da metilação do DNA na regulação da replicação, expressão gênica e diferenciação celular (Vanyushin e Kirnos, 1988; Chomet, 1991; Finnegan et al., 1993). Trabalhos que incluem obtenção de plantas transgênicas também têm lidado com tal assunto, uma vez que o padrão

de metilação do DNA da planta pode inibir a transcrição do gene exógeno transferido (Hepburn *et al.*, 1983; Ingelbrecht *et al.*, 1994; Meyer, 1995).

A transcrição dos genes de proteínas de reserva também parece ser controlada pela metilação do DNA. Em milho, observou-se uma extensiva hipometilação de genes de zeínas e glutelinas, especificamente em células do endosperma. Pelo fato de esta hipometilação ocorrer antes da transcrição dos genes de zeínas e glutelinas, sugere-se que o processo de desmetilação é um dos primeiros passos rumo à ativação destes genes (Bianchi e Viotti, 1988).

No mutante *lys3a* de cevada, a transcrição dos genes de B- e C-hordeínas é altamente reduzida. Por outro lado, as D-hordeínas são normalmente expressas. Recentemente, verificou-se que o nível de transcrição destes genes está inversamente relacionado ao grau de metilação de seus promotores (Sørensen *et al.*, 1996).

A metilação do DNA parece inibir a transcrição ao impedir a ligação de proteínas regulatórias a regiões específicas no promotor dos genes (Dynan, 1989; Inamdar *et al.*, 1991; Ehrlich e Ehrlich, 1993). Pouco é conhecido sobre os mecanismos que controlam os processos de metilação ou desmetilação de sequências particulares em determinado tecido e/ou período de desenvolvimento do organismo (Dynan, 1989). Sabe-se, entretanto, que o processo de metilação está relacionado com mudanças na conformação da cromatina (Keshet *et al.*, 1986; Chomet, 1991; Sørensen *et al.*, 1996)

6.3. Interação com fatores de transcrição

A regulação da transcrição ocorre, na maioria das vezes, ao nível da iniciação. Normalmente é um processo onde sequências específicas de DNA (elementos *cis*), através da interação com fatores de função regulatória (fatores *trans*), ou ainda de maneira direta, modulam a formação do complexo de pré-iniciação (Kuhlemeier *et al.*, 1987; Schleif, 1988).

6.3.1. As sequências de DNA envolvidas na transcrição

Os elementos *cis* podem ser divididos em dois grupos. Um é representado pelos sítios de ligação do complexo da RNA polimerase (fatores gerais de transcrição), usualmente o *elemento TATA*, também denominado núcleo do promotor.

A análise de 79 genes de planta mostrou que a sequência consenso para o *elemento TATA*, ou TATA-box, em plantas, é 5′-TCAC**TATATATA**G-3′, enquanto que para genes de proteínas de reserva de cereais (análise de 12 genes) é 5′-TAAC**TATAAATA**G-3′ (Joshi, 1987).

Para o local de início da transcrição em eucariotos foi proposta a sequência consenso PyAPyPy, onde Py representa uma base pirimidínica (Breathnach e Chambon, 1981). No caso de plantas é 5′-CTCATCA-3′, e para os genes de proteínas de reserva de cereais 5′-ATCATCA-3′ (Joshi, 1987).

A outra classe de elementos *cis* é formada pelos sítios de ligação dos fatores reguladores da transcrição, proteínas que interagem com o complexo da RNA polimerase. Alguns elementos desta segunda classe necessitam adotar uma posição fixa em relação aos demais elementos para desempenhar corretamente sua função; outros podem funcionar a distâncias variadas do *elemento TATA*; outros o fazem mesmo se sua orientação estiver invertida, como acontece com os "enhancers" (Kuhlemeier, 1992). Com respeito a esta última classe, alguns autores fazem uma distinção entre os elementos que ativam a transcrição, chamados de intensificadores ("enhancers"), e os que a inibem, chamados silenciadores ("silencers") (Tjian, 1995).

Supondo que sequências envolvidas no controle da expressão de um grupo de genes relacionados quanto à função e expressão foram conservadas durante o processo evolutivo, torna-se possível, através da comparação entre os diversos genes de proteínas de reserva já seqüenciados, identificar os possíveis elementos *cis* envolvidos na determinação do padrão de expressão destes genes (Kreis *et al.*, 1986).

Um dos primeiros elementos *cis* identificado pelo alinhamento das sequências dos promotores de diversos genes de prolaminas foi o *prolamin box* ou *endosperm box.* Este sítio de 7 pares de base (pb) (5'-TGTAAAG-3'), geralmente localizado a 330 pb do ATG inicial da tradução, e cuja função é ainda desconhecida, caracteriza-se por ser altamente conservado na maioria dos genes de prolaminas já isolados (Kreis *et al.*,1986).

Com o advento dos métodos de transformação de plantas (Weising *et al.*, 1988), aliado a técnicas de mutação induzida *in vitro*, tornou-se possível a introdução e posterior análise da expressão de genes com mutações ou deleções nas sequências-consenso de seus promotores, permitindo desta forma a caracterização da função biológica destes possíveis elementos *cis* no processo de expressão gênica (Benfey e Chua, 1989).

Tipicamente, a região promotora dos genes é composta por vários sítios de ligação de proteínas. A partir de variações no tipo e no arranjo destes sítios, ocorre a formação de complexos nucleoprotéicos distintos e únicos, gerando diversidade e exatidão na regulação da transcrição (Tjian e Maniatis, 1994; Grosschedl, 1995).

6.3.2. As proteínas envolvidas no processo de transcrição

Os eucariotos possuem três enzimas responsáveis pela síntese de moléculas de RNA; dentre estas, a RNA-polimerase II é a que transcreve os genes que codificam polipeptídeos.

A polimerase de RNA de eucariotos não é capaz de iniciar a transcrição dos genes sem a presença de vários fatores reguladores de transcrição (fatores *trans*), os quais interagem com as sequências regulatórias do DNA (elementos *cis*), comandando a transcrição gênica (Dynan e Tjian, 1985).

Os fatores de transcrição podem ser divididos em dois tipos segundo a forma através da qual agem: (i) os fatores gerais de transcrição (GTFs, *general transcription factors*), que são os que, juntamente com a RNA-polimerase II, formam o complexo multiprotéico capaz de reconhecer um promotor (na região

do *elemento TATA*) e iniciar corretamente e manter a transcrição em um nível basal; (ii) e os fatores reguladores da transcrição, que se ligam a outros elementos *cis* no DNA e interagem com os GTFs, para assim ativar ou inibir (*ativadores* ou *repressores*) a transcrição (Goodrich *et al.*, 1996).

6.3.2.1. Fatores gerais de transcrição

O maquinário de transcrição que emprega a RNA polimerase II é formado por um complexo arranjo de proteínas. Em animais foram identificados seis fatores 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 TFIID, na verdade, é um complexo protéico formado pela TBP (*TATA binding protein*) e vários fatores TAF (*TBP associated factor*); estes últimos parecem desempenhar um papel central na intermediação entre o complexo basal e os fatores reguladores da transcrição (Tjian e Maniatis, 1994). Ver Figura 1.

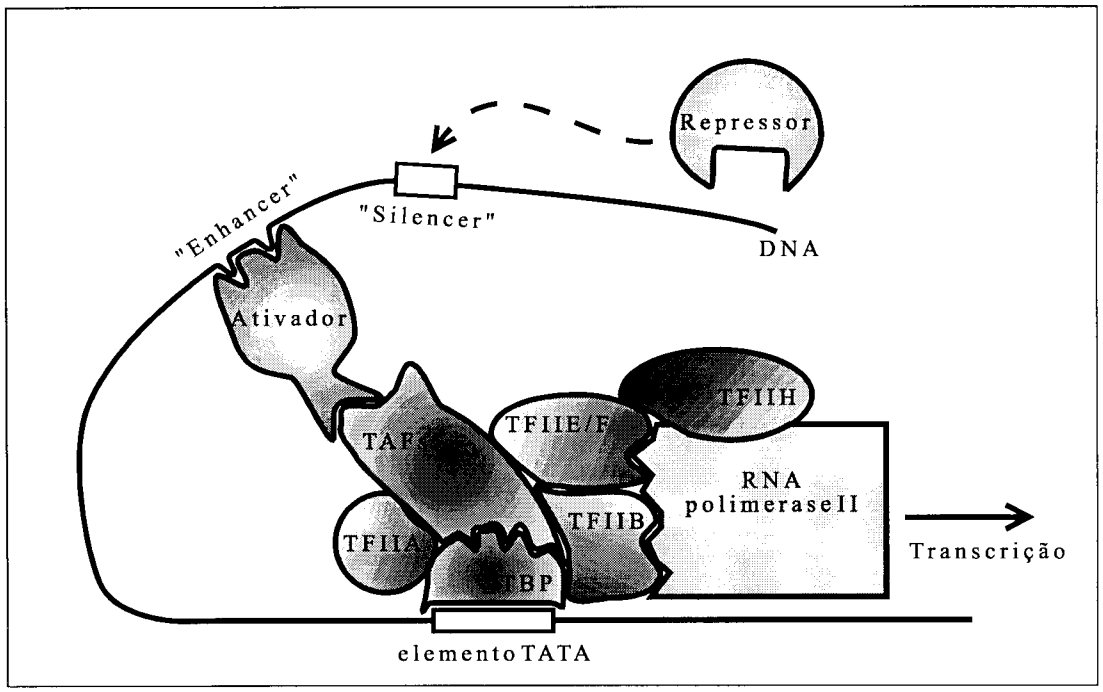


Figura 1. Representação esquemática mostrando os fatores de transcrição e suas interações com as sequências alvo no DNA, com a RNA-Polimerase II e entre si.

6.3.2.2. Fatores reguladores da transcrição

Os fatores de transcrição são geralmente proteínas de estrutura modular. Um fator de transcrição típico é composto por três módulos: (i) um domínio de ligação ao DNA, (ii) um domínio de interação protéica, geralmente de multimerização, e (iii) um domínio de ativação.

Os fatores de transcrição têm sido classificados em função do domínio de ligação ao DNA e de formação de dímeros. Uma série de fatores de transcrição não puderam ser classificados em famílias determinadas, o que leva a supor que outras famílias serão ainda caracterizadas.

Cada vez se torna mais evidente que as regiões de reconhecimento de DNA, de interação protéica e ativação são módulos com plasticidade suficiente para originar várias combinações funcionais. Isto permite que, com um número relativamente pequeno de diferentes módulos, seja possível obter inúmeros fatores com especificidades distintas, gerando a diversidade necessária à regulação do complexo padrão de expressão gênica (Tjian e Maniatis, 1994).

As classes mais importantes de fatores reguladores da transcrição de plantas são as bZIP (*basic-region/leucine-zipper*), as bHLH (*basic helix-loop-helix*), as Myb, as proteínas que contém o homeodomínio (HD), 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 são também encontrados em fungos e animais.

6.3.3. Mecanismos de ativação via fatores reguladores da transcrição

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 em diversos sítios do promotor do gene (Kuhlemeier, 1992).

Uma vez montado o complexo no sítio do DNA, os domínios de ativação de alguns fatores devem interagir, direta ou indiretamente, com os componentes

do complexo basal de transcrição. A potência final de um regulador de transcrição é determinada por características que vão desde sua afinidade pelo sítio de ligação ao DNA até a intensidade da interação entre seu domínio de ativação e seu alvo no complexo basal (Tjian e Maniatis, 1994).

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

Em diversos elementos *cis*, a montagem de complexos nucleoprotéicos ativos depende da interação específica de vários componentes no mesmo promotor. Por exemplo, no caso do promotor da interleucina 2 ou do MHC II (complexo principal de histocompatibilidade de classe II), a ausência de um dos vários fatores faz com que, *in vivo*, nenhum dos outros reconheça seu sítio de ligação (Garrity *et al.*, 1994; Grosschedl, 1995).

A homo- ou heterodimerização dos fatores reguladores da transcrição detêm um papel importante na regulação da atividade biológica. Alguns fatores de transcrição são ativos apenas quando na forma de heterodímero. 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 do CREB (*cAMP response element binding protein*) ao ligar-se ao elemento CREM. Heterodímeros podem ainda apresentar especificidade diferente da apresentada pelos homodímeros de seus componentes. A heterodimerização pode também levar a novas combinações de domínios de ativação/repressão, mudando, desta forma, a função regulatória das proteínas ligadas a um determinado sítio no DNA (Curran e Franza Jr, 1988; Folkers *et al.*, 1991; Wagner e Green, 1994).

Existem ainda proteínas que causam dobramentos na fita de DNA ao se ligarem a ela, posicionando corretamente outros fatores de transcrição e/ou

aumentando a afinidade destes por seus sítios, como ocorre com o fator LEF-1 em ratos (Giese *et al.*, 1995) e o HMG I(Y) em humanos (Du *et al.*, 1993). Por outro lado, as mudanças na conformação do DNA podem ter um efeito negativo sobre outros fatores; é o caso da YY1, uma proteína *zinc finger* de humanos. Ainda que em determinados contextos a YY1 possa funcionar como ativadora, no gene *c-fos* ela parece organizar a topologia do DNA de uma maneira tal que previne a interação entre um ativador específico da transcrição deste gene e o complexo basal (Natesan e Gilman, 1993). É importante notar que tais fatores, os quais podem ser agrupados por sua função arquitetural, não apresentam domínios de ativação da transcrição, sendo incapazes de promover transativação por si próprios (Wolffe, 1994b).

Alguns dos fatores *trans*-regulatórios estão presentes na maioria das células e ativos nas mais diversas condições. Outros fatores são bastante especializados. Estes últimos são produzidos somente no tecido e no momento adequados para a síntese de uma dada proteína, ou então eles podem já estar presentes, porém em uma forma inativa, sendo posteriormente ativados nos tecidos e estágios de desenvolvimento adequados (Maniatis *et al.*, 1987; Kuhlemeier, 1992).

A existência de mecanismos pós-traducionais para a ativação dos fatores *trans* fornece uma alternativa para explicar o modo pelo qual ocorre a regulação da expressão dos genes regulatórios (aqueles envolvidos na síntese destes fatores *trans*). Se a regulação destes genes ocorresse ao nível da transcrição, haveria necessidade da existência de "cascatas regulatórias" envolvendo a transcrição sequencial de genes codificando fatores regulatórios de outros genes. Com o mecanismo de modulação pós-traducional, pelo menos alguns dos genes regulatórios dispensariam o controle ao nível transcricional, sendo expressos constitutivamente. Assim, com o desenvolvimento dos tecidos, indutores da diferenciação celular poderiam atuar na conversão de uma certa quantidade dos fatores regulatórios apropriados para a forma ativa, promovendo a expressão gênica no tecido e estágio específico (Maniatis *et al.*, 1987).

Os mecanismos pós-traducionais de modulação dos fatores *trans* podem ser de três tipos: (i) gerando modificações no domínio de ligação do fator ao DNA; (ii) modificando o domínio de ativação do fator; (iii) ou restringindo a localização intra-celular do fator (Gilmartin *et al.*, 1990).

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

Um segundo tipo de modificação covalente capaz de alterar a atividade de fatores de transcrição foi descrito para algumas bZIPs, como AP-1, c-jun e c-fos. Neste caso, a oxidação de uma cisteína conservada da região básica parece impedir a dimerização necessária para a ligação ao DNA (Wagner e Green, 1994).

7. Genes reguladores da biossíntese das proteínas de reserva

O estudo da taxa de acumulação das proteínas de reserva e de suas quantidades finais em diferentes mutantes fenotipicos da semente permitiram o descobrimento de vários genes que regulam a biossíntese destes polipeptídeos, principalmente em cereais.

Mutações nestes genes reguladores têm drásticos efeitos quantitativos e qualitativos no perfil das prolaminas da semente, e freqüentemente são também acompanhadas por significativas mudanças em algumas das outras frações protéicas. Este último fato geralmente acarreta um vantajoso aumento no conteúdo de lisina do endosperma, às custa porém de um decréscimo no conteúdo protéico total da semente e queda de produtividade.

Em milho, nos mutantes *opaco2* (*o2*) e *endosperma defetivo B30* (*De-B30*) ocorre redução preferencial do nível das zeínas de 22 KDa. Em *opaco7* (*o7*) há redução no nível das zeínas de 19 KDa, e nos mutantes *floury2* (*fl2*), *mucronate* (*Mc*) e *opaco6* (*o6*) ocorre suprissão da síntese de todas as zeínas com a mesma efetividade. Foi constatado que a menor produção de polipeptídeos de zeínas nestes mutantes é resultante de uma menor população de seus mRNAs nos endospermas *o2*, *o7* e *fl2*, ou de um nível menor de mRNA traduzível, em *o6*, *Mc* e *De-B30* (Soave e Salamini, 1984; Motto *et al.*, 1989).

O mecanismo de ação destes *loci* a nível molecular é desconhecido. No entanto, por controlarem a ação de genes situados em diversos pontos do genoma, o processo de regulação deve ocorrer através de fatores difusíveis de ação regulatória, ou seja fatores *trans* (Soave e Salamini, 1984).

8. O locus *opaco2* codifica uma proteína regulatória da família das bZIP

Em plantas homozigotas *opaco2* (*o2*), a síntese da classe das α-zeínas, particularmente da família de 22 KDa, encontra-se grandemente reduzida (Burr e Burr, 1982), em parte devido a um decréscimo nos níveis dos transcritos que codificam os polipeptídeos desta classe (Kodrzycki *et al.*, 1989).

O isolamento e sequenciamento de clones de cDNA do gene *Opaco2* (*O2*) caracterizou a proteína codificada por este *locus* como sendo um fator regulador da transcrição pertencente à classe das bZIP (Hartings *et al.*, 1989; Schmidt *et al.*, 1987; Schmidt *et al.*, 1990).

A atividade dos fatores da classe bZIP depende de duas funções mediadas por diferentes domínios: o domínio básico, responsável pela interação direta com o DNA, e o domínio das repetições de leucinas, responsável pela interação com outras proteínas através de dimerização. De acordo com o modelo "scissors grip", os resíduos de leucina separados por sete aminoácidos ficariam orientados na mesma face de uma α-hélice, criando uma face hidrofóbica. As α-hélices de monômeros separados, apresentando conformação "coiled coil", dimerizam paralelamente, enquanto que as hélices formadas pelos

domínios básicos repelem-se mutuamente, produzindo uma estrutura em forma de "Y" (Figura 2). Os dois braços do "Y", representando os domínos básicos, intercalam-se entre as fitas de DNA, penetrando pelo sulco maior (Vinson *et al.*, 1989).

Estudos posteriores postularam a ocorrência de alterações conformacionais na proteína após a ligação ao DNA, levando à proposição do modelo "induced helical fork" (O´Neil *et al.*, 1990). Neste modelo, o domíno básico parcialmente estruturado interage com o sítio de DNA produzindo uma α-hélice que é então estabilizada através de contatos específicos com o DNA. Este modelo foi posteriormente comprovado através de estudos de cristalografia de raios-X realizados com complexos proteína/DNA (Ellenberger *et al.*, 1992).

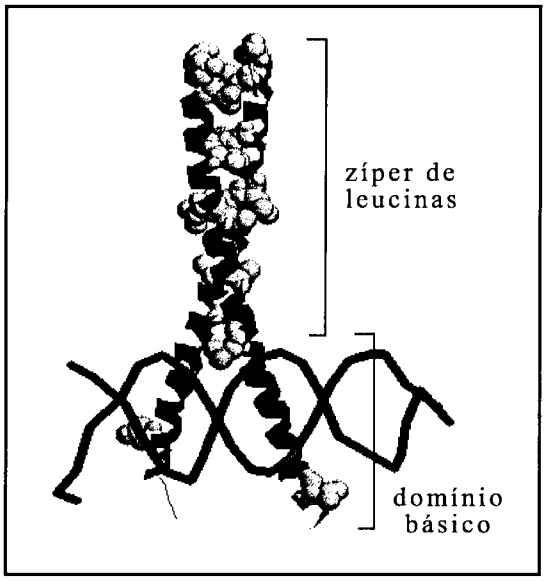


Figura 2. Representação gráfica do complexo GCN4/sítio AP-1, construída a partir das coordenadas depositadas no "Protein Data Bank", código 1YSA (Ellenberger *et al.*, 1992).

Até 1994, 22 bZIPs haviam sido identificadas em plantas (Foster *et al.*, 1994). Hoje, mais de 50 proteínas pertencentes a esta classe podem ser encontradas nos bancos de sequências de genes de planta; porém, até o

momento, a proteína Opaco2 é a única destas bZIPs para a qual uma função regulatória foi especificada *in situ*, ou seja, onde os genes por ela controlados são conhecidos.

9. Existência de um gene tipo *Opaco2* em *Coix* e sorgo

A semelhança nos genes de proteínas de reserva de milho, *Coix* e sorgo levou à suposição da existência de um *locus* regulatório do tipo *opaco2* também nestas espécies. Em experimentos de "Southern blot", utilizando um fragmento do clone *Opaco2* de milho como sonda, foi possível comprovar a presença de sequências similares à do gene *Opaco2* em *Coix* e sorgo (Ottoboni *et al.*, 1993).

Posteriormente, genes similares ao *opaco2* de milho foram isolados e caracterizados em sorgo (Pirovano *et al.*, 1994) e *Coix* (Vettore, 1994), possibilitando um interessante modelo para o estudo da evolução do mecanismo de regulação transcricional.

10. A proteina Opaco2 é capaz de ligar-se a diferentes sequências de DNA

O primeiro gene cujo promotor foi investigado para identificação do sítio de ligação da proteína Opaco2 (O2) foi o gene *b-32*, que codifica uma proteína RIP (ribosome-inactivating protein), expressa unicamente no endosperma de milho e sabidamente sob influência do locus *Opaco2* (Bass *et al.*, 1992). Através de experimento de "footprint" *in vitro*, foram identificados cinco sítios de ligação da proteína O2 no promotor do gene *b-32*, chamados B1 a B5, os quais apresentam a sequência consensso 5´-GATGAPyPuTGPu-3´. Em ensaios de transformação transitória de protoplastos de tabaco, verificou-se um incremento de 82 vezes na expressão de um gene indicador sob controle do promotor *b-32* quando simultaneamente era transferido o gene *O2* sob controle de um promotor constitutivo, corroborando assim a transativação do gene *b-32* pela proteína O2 (Lohmer *et al.*, 1991).

No gene da zeína de 22 kDa, a proteína O2 liga-se à sequência 5´-TCCACGTAGA-3´, que se encontra aproximadamente 10 nucleotídeos

distante do *prolamin box*. Um multímero desta sequência, quando adicionado ao promotor mínimo GAL1, é suficiente para ativar a expressão do gene indicador *lacZ* em levedura, se conjuntamente for expresso o gene *O2* (Schmidt *et al.*, 1992).

Posteriormente, através de ensaios de transativação em protoplastos de endosperma e em protoplastos de células cultivadas *in vitro*, utilizando o gene indicador *gus* sob controle de promotores de zeínas com alterações no sítio TCCACGTAGA, conclui-se que uma única substituição na região central ACGT é suficiente para suprimir a ligação da proteína O2 *in vitro*, bem como a resposta à transativação *in vivo*. Também se verificou que a sequência TCC era essencial para a alta afinidade de ligação da O2 ao promotor e por consequência para a transativação *in vivo* (Ueda *et al.*, 1992).

Tendo em vista a diferença existente entre os sítios de ligação da proteína O2 nos promotores dos genes *b-32* e da zeína de 22 kDa, havia a necessidade de explicar como um fator de transcrição era capaz de se ligar a sítios tão diferentes entre si.

A princípio, generalizando o fato de que a maior parte dos fatores bZIP de plantas apresentam alta afinidade de ligação a sequências contendo o motivo central ACGT, pensava-se que este ACGT era imprescindível para a interação de qualquer bZIP com o DNA, inclusive para a O2. De fato, em ensaios de "gel shift" a proteína O2 mostrou alta afinidade por uma ampla coleção de sequências contendo o motivo central ACGT (Izawa *et al.*, 1993).

Dentre os cinco sítios descritos no promotor do gene *b-32*, somente o B5 apresentava o motivo ACGT, portanto, pensava-se que somente este seria um verdadeiro sítio de ligação para a proteína O2, enquanto que os demais quatro seriam produto de artefatos de técnica (Izawa *et al.*, 1993).

Entretanto, analisando a interação da proteína O2 de milho com o promotor de um gene de α-coixina, encontrou-se um sítio de ligação de O2 sem o motivo ACGT e que apresentava maior similaridade com os sítios encontrados

no gene *b-32* do que com aquele descrito para o gene da α-zeína (Yunes *et al.*, 1994b).

Corroborando o resultado obtido com o gene da α-coixina, de Pater *et al.* (1994), estudando a interação da proteína O2 com o promotor do gene *psl*, que codifica uma lectina da semente de ervilha, mostraram que a proteína O2 reconhece um sítio sem o motivo ACGT e que se enquadra na sequência consenso proposta para o gene *b-32.* Interessante notar também que este sítio é idêntico ao sítio de ligação da proteína GCN4 (Hill *et al.*, 1986), uma bZIP de leveduras que controla vários dos genes do metabolismo de nitrogênio. De fato, a proteína O2 é capaz de substituir a função da GCN4 em levedura (Mauri *et al.*, 1993).

Recentemente, trabalhando com o promotor do gene de uma proteína de reserva de trigo, uma glutenina de baixo peso molecular, Holdsworth *et al.* (1995) comprovaram que a proteína O2 reconhece e é capaz de transativar este promotor *in vivo* ao se ligar à sequência 5′-TGAGTCATAT-3′, que é o mesmo elemento *cis* antes conhecido como GLM (*GCN4 like motif*). Este resultado, além de comprovar a capacidade da proteína O2 em reconhecer sítios do tipo GCN4, sugere também que outros cereais além daqueles pertencentes à tribo do milho (Andropogoneae) provavelmente possuem um *locus* do tipo *Opaco2* que regula a expressão das proteínas de reserva.

Por outro lado, nos promotores dos genes da β-coixina e da β-zeína, foram encontrados sítios de ligação da proteína O2 apresentando o motivo ACGT. Experimentos de "Northern blot" e a análise do perfil de zeínas no endosperma de sementes de milho normal e mutante *opaco2* confirmaram que as β-zeínas têm sua expressão influenciada pelo gene *Opaco2.* Além disso, ensaios de expressão transitória demonstraram que a O2 é capaz de transativar a expressão de um gene indicador controlado pelo promotor da β-coixina ou da β-zeína (Cord Neto *et al.*, 1995).

Uma vez que ao *locus Opaco2* era atribuído unicamente o controle da expressão da classe das α-zeínas de 22 kDa (Schmidt *et al.*, 1992), o fato de as β-zeínas também responderem ao controle deste *locus* constituiu um no primeiro indício de que a proteína O2 seria responsável pelo controle da expressão de vários genes distintos, de tal modo que um certo relaxamento na especificidade de ligação ao DNA seria vantajoso ou até mesmo necessário (Yunes *et al.*, 1994a).

Recentemente, comprovou-se que a O2 é capaz de reconhecer não um mas três sítios no promotor da α-zeína de 22 kDa, denominados Z1, Z2 e Z3. O sítio Z3 apresenta alta afinidade por O2 e corresponde à sequência 5´-TCCACGTAGA-3´ anteriormente descrita por Schmidt *et al.* (1992). Os outros dois sítios apresentam baixa afinidade por O2. O Z1 é similar aos sítios de O2 descritos para os genes *b-32* e da α-coixina. Já o Z2 apresenta apenas uma baixa similaridade com os outros sítios e está localizado adjacente ao sítio Z3. Ensaios de expressão transitória mostraram que Z1 e Z3 são igualmente efetivos na mediação da transativação pela proteína O2. Para o Z2 observou-se que, embora ele não seja capaz de isoladamente conferir transativação por O2, incrementa significativamente os níveis de transativação quando em combinação com qualquer um dos outros dois sítios (Muth *et al.*, 1996).

APRESENTAÇÃO DOS TRABALHOS E OBJETIVOS

O corpo desta tese está constituído por dois trabalhos, um dos quais já publicado e o outro submetido à publicação. Ambos os trabalhos tratam da interação da proteína Opaco2 com o promotor do gene da α-coixina de 25 kDa.

A realização desses dois trabalhos baseou-se nos estudos realizados de caracterização de genes de proteínas de reserva de milho, *Coix* e sorgo, onde constatou-se o envolvimento da proteína Opaco2 na regulação da expressão de diversos genes que atuam durante o desenvolvimento do endosperma. Esses estudos constam nos trabalhos abaixo relacionados.

Leite, A., Freitas, F.A., Yunes, J.A. and Arruda, P. (1991). Nucleotide sequence of a cDNA encoding γ-coixin from *Coix lacryma-jobi* seeds. *Plant Physiol.* 97: 1604-1605.

Leite, A., Yunes, J.A., Turcinelli, S.R. and Arruda, P. (1992). Cloning and characterization of a cDNA encoding a sulfur-rich coixin. *Plant Mol. Biol.* 18: 171-174.

Ottoboni, L.M., Leite, A., Yunes, J.A., Targon, M.L., de Souza Filho, G.A. and 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.

de Freitas, F.A., Yunes, J.A., da Silva, M.J., Arruda, P. and Leite, A. (1994). Structural characterization and promoter activity analysis of the γ-kafirin gene from sorghum. *Mol. Gen. Genet.* 245: 177-186.

Yunes, J.A., Cord Neto, G., Leite, A., Ottoboni, L.M. and Arruda, P. (1994). The role of the Opaque2 transcriptional factor in the regulation of protein

accumulation and amino acid metabolism in maize seeds. *An. Acad. Bras. Cienc.* 66: 227-237; quiz 237-238.

Cord Neto, G., Yunes, J.A., da Silva, M.J., Vettore, A.L., Arruda, P. and Leite, A. (1995). The involvement of Opaque 2 on β-prolamin gene regulation in maize and *Coix* suggests a more general role for this transcriptional activator. *Plant Mol. Biol.* 27: 1015-1029.

Trabalho 1: **The transcriptional activator Opaque2 recognizes two different target sequences in the 22-kD-like α-prolamin genes.**

José A. Yunes, Germano Cord Neto, Márcio J. da Silva, Adilson Leite, Laura M.M. Ottoboni e Paulo Arruda. *Plant Cell* 6 (1994): 237-249.

Em função da alta homologia estrutural e funcional dos genes de α-coixina de 25 kDa e α-zeina de 22 kDa, bem como, dos resultados preliminares de "Southern blot" indicando a existência de sequências similares ao gene *Opaco2* de milho em *Coix* (Ottoboni *et al.*, 1993), iniciou-se este estudo, com os objetivos a seguir:

1. Verificar se o gene da α-coixina de 25 kDa apresenta no seu promotor algum sítio de ligação da proteína O2.
2. Determinar a sequência e local deste(s) possível(is) sítio(s) de ligação da proteína O2.
3. Determinar a função biológica deste(s) possível(is) sítio(s) de ligação da proteína O2, através de ensaios de expressão transitória em protoplastos de tabaco e em sementes de milho.

Trabalho 2: **The Opaque2 transcription factor binds to two adjacent GCN4-like motifs in a cooperative manner.**

José A. Yunes, André L. Vettore, Marcio J. da Silva, Adilson Leite e Paulo Arruda.

Uma vez clonado o gene *Opaco2* de *Coix*, verificou-se que a proteína O2 de *Coix* reconhece no promotor do gene da α-coixina de 25 kDa o mesmo sítio de ligação da proteína O2 de milho (Vettore, 1994). O fato de este sítio apresentar uma sequência de nucleotídeos aproximadamente duas vezes maior que aquela dos sítios de ligação da O2 nos genes *b-32* ou da α-zeína de 22 kDa parecia indicar a existência de dois sítios adjacentes. Isto era ainda reforçado pela presença de duas sequências adjacentes similares ao sítio de ligação da proteína GCN4. Os objetivos principais deste trabalho foram:

1. Verificar se a região de ligação da proteína O2 no promotor da α-coixina de 25 kDa é constituida por dois sítios adjacentes.
2. Delimitar cada um dos sítios.
3. Determinar se a proteína O2 é capaz de se ligar simultaneamente a ambos sítios.
4. Determinar se a ligação da proteína O2 nestes sítios é cooperativa.
5. Calcular as constantes de equilibrio de ligação da proteína O2 para cada um dos sítios e para a cooperatividade.
6. Determinar o efeito da ligação da proteína O2 em cada um dos sítios, bem como o efeito da cooperatividade, no processo de transcrição.

The Plant Cell, Vol. 6, 237–249, February 1994 © 1994 American Society of Plant Physiologists

The Transcriptional Activator Opaque2 Recognizes Two Different Target Sequences in the 22-kD-like α-Prolamin Genes

José A. Yunes,[a] Germano Cord Neto,[a] Marcio J. da Silva,[a] Adilson Leite,[a] Laura M. M. Ottoboni,[a] and Paulo Arruda [a,b,1]

[a] Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, 13081, Campinas, SP, Brazil
[b] Departamento de Genética e Evolução, IB, Universidade Estadual de Campinas, 13081, Campinas, SP, Brazil

The maize Opaque2 (O2) protein is a "leucine zipper" DNA binding factor that interacts with the sequence TCCACGTAGA in the promoters of the 22-kD α-zein genes and activates its transcription. A completely different consensus sequence (GATGAPyPuTGPu) identified in *b-32*, a gene that encodes an abundant albumin that is also under control of the *O2* locus, can also be bound by the O2 protein. We showed that the gene encoding the 22-kD-like α-coixin, the α-prolamin of the maize-related grass *Coix*, can also be transactivated by the O2 protein. A binding assay in vitro and footprint analysis demonstrated that the GACATGTC sequence of the α-coixin promoter can be recognized and protected by the maize O2 protein. Employing transient expression experiments in immature maize endosperm and tobacco mesophyll protoplasts, we demonstrated that the O2 protein can activate expression of the β-glucuronidase reporter gene placed under the control of the 22-kD-like α-coixin promoter. We also demonstrated that a 22-kD-like α-coixin pseudogene promoter is transactivated by the maize O2 protein.

INTRODUCTION

Prolamins, the major storage proteins of cereals (Osborne and Mendel, 1914), have proved to be a useful model system to study the regulation of gene expression in plants. We are currently studying the prolamins from *Coix lacryma-jobi*, which belongs to the grass tribe Andropogoneae. Other members of this tribe include maize, *Tripsacum*, and sorghum. Comparisons between the genes from these related species have provided insights into the structure and evolution of the prolamin genes. The prolamins of *Coix* and maize, coixins and zeins, respectively, can be separated into polypeptides of distinct molecular mass by SDS-PAGE (Ottoboni et al., 1990) and grouped into four classes, α-, β-, γ-, and δ-prolamins (Esen, 1986; Leite et al., 1990). In maize and *Coix*, the α-prolamins account for more than 70% of total prolamins and are composed of polypeptides of 19- and 22-kD in maize (Burr et al., 1982) and 25- and 27-kD in *Coix* (Leite et al., 1990). These size class proteins are encoded in both cereals by multigene families (Hagen and Rubenstein, 1981; Ottoboni et al., 1993), and their expression is under strict tissue-specific and developmental control (Shotwell and Larkins, 1989; Targon et al., 1992).

[1] To whom correspondence should be addressed at Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas–Unicamp, Cidade Universitária "Zeferino Vaz," Cx.P. 6109, 13081-970-Campinas-SP, Brasil.

Recently, we have shown that the 25-kD α-coixin genes share high homology with the 22-kD α-zein genes in respect to both protein structure and to the regulatory sequences present in their promoter regions (Ottoboni et al., 1993). For this reason, we designated the 25-kD α-coixin gene family as 22-kD-like α-coixins (Ottoboni et al., 1993).

Several mutations are known to affect the level of zeins in the maize endosperm (Motto et al., 1989). One of these, *opaque2* (*o2*), exerts its major effect on the level of the 22-kD α-zeins (Soave et al., 1976; Burr and Burr, 1982). A homozygous loss-of-function mutation at *O2* results in an overall reduction of 50 to 70% in zein content. This is a consequence of a lower rate of transcription, particularly from those genes encoding the 22-kD α-zeins (Kodrzycki et al., 1989).

Another endosperm protein, a 32-kD albumin designated b-32, whose expression is temporally and quantitatively coordinated with the deposition of zeins, is greatly affected by the homozygous *o2* mutation. Such a mutation reduces the b-32 mRNA content to 5 to 10% of the wild-type level (Di Fonzo et al., 1986).

The *O2* gene has been cloned by transposon tagging (Schmidt et al., 1987; Motto et al., 1988) and shown to be a basic "leucine zipper" (bZIP) DNA binding protein (Hartings et al., 1989; Schmidt et al., 1990). Subsequent experiments have shown that the O2 protein is capable of transactivating *b-32* gene transcription in vivo by interacting with five putative

binding sites containing the consensus sequence GATGAPyPuTGPu (Lohmer et al., 1991).

A quite different target sequence (TCCACGTAGA) has been shown to be the O2 binding site in the 22-kD α-zein promoter (Schmidt et al., 1992). This sequence was verified by mobility shift assays using the O2 protein from maize endosperm (Schmidt et al., 1992). Furthermore, when a multimer of this O2 target site is placed upstream of a minimal promoter, it mediates transactivation of the reporter gene by O2 in various transient expression systems (Schmidt et al., 1992; Ueda et al., 1992).

We have previously shown the presence of putative O2 boxes in the promoter region of the 22-kD-like α-prolamins and also found sequences homologous to the *O2* gene in *Coix* and sorghum (Ottoboni et al., 1993). In this study, we report the binding to and transactivation of the 22-kD-like α-coixin promoter by the maize O2 protein in both homologous and heterologous expression systems. Surprisingly, we found that the core of the α-coixin O2 target site (GACATGTC) shares more homology with the B1 and B4 O2 binding sites of the *b-32* gene than with that of the 22-kD α-zein gene. Moreover, the α-coixin O2 box does not contain the ACGT core, an essential condition for the high-affinity binding of the O2 protein (Schmidt et al., 1992; Izawa et al., 1993).

RESULTS

Maize O2 Protein Binds to the 22-kD-like α-Coixin Promoter

Sequence analysis of the 22-kD-like α-prolamin genes from maize, *Coix*, and sorghum revealed two homologous O2 target sites in their promoters (Ottoboni et al., 1993). The target sites are designated as O2a and O2b boxes in this study. In the *α-3B* α-coixin promoter, the first box, O2a, is located at position −301 with respect to the ATG, and the second, O2b, at position −281. However, both boxes showed significant discrepancy in relation to the O2a target sequence 5′-TCCACGTAGA-3′ present in the *22Z-4* 22-kD α-zein gene promoter (Schmidt et al., 1992). Interestingly, as can be seen in Figure 1, other 22-kD α-prolamin genes share such a discrepancy, including two recently isolated 22-kD α-zein genes (Thompson et al., 1992).

To determine whether the O2-like boxes of the 22-kD-like α-coixin promoter are recognized by the O2 protein, we conducted a DNA binding assay using the maize O2 protein in a β-galactosidase (β-Gal)::O2 fusion (Schmidt et al., 1990). As shown in Figure 2A (lane 2), a 260-bp Sau3AI restriction fragment that is immediately 3′ of the "prolamin-box" (Brown et al., 1986b) and contains the two related O2 boxes (Figure 2B) was bound by the O2 protein. Digestion of the Sau3AI fragment with EcoRI, which separates the two O2-like boxes (Figure 2B), showed that the resulting 225-bp fragment retains a strong binding capacity (Figure 2A, lane 4) and indicated that the O2

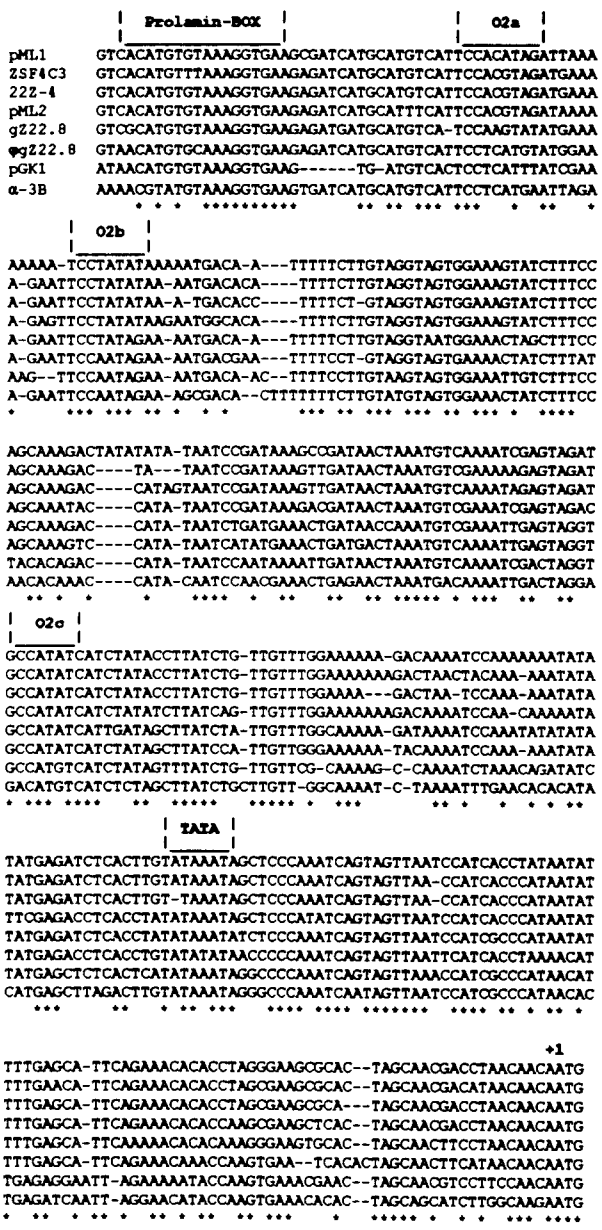


Figure 1. Sequence Alignment of 22-kD α-Prolamin Promoters.

The ∼350-bp promoter fragments extending from the ATG of six 22-kD α-zein clones are aligned. They include *pML1* and *pML2* (Wandelt and Feix, 1989), *ZSF4C3* (Liu and Rubenstein, 1992), *22Z-4* (Schmidt et al., 1992), and *gZ22.8* and *ψgZ22.8* (Thompson et al., 1992) from maize; the 22-kD-like α-kafirin clone *pGK1* (DeRose et al., 1989) from sorghum; and the *Coix* 22-kD-like α-coixin clone *α-3B* (Ottoboni et al., 1993). Dashes indicate deletions introduced to maximize the homology comparison. Identical nucleotides are marked by asterisks. The locations of the putative DNA regulatory sequences "prolamin-box;" O2 target sites O2a, O2b, and O2c; TATA box; and the translation start codon (+1) are indicated above the sequence data.

protein does not bind to the O2a box of the α-coixin promoter. Because in this EcoRI-Sau3AI digestion the O2a box remained in a 35-bp fragment that is too small to be retained by the β-Gal::O2 fusion, the binding was confirmed by digesting the α-coixin promoter fragment with EcoRI and XbaI. As shown in Figure 2A (lane 6), only the 310-bp fragment that contains the O2b (and O2c, see below) box (Figure 2B) was retained. The Sau3AI digestion of the 22-kD α-zein promoter (plasmid Pα-Z4103), which was included in the binding assay as a positive control, showed that the O2 protein binds to a 200-bp fragment that contains the O2a box (Figure 2A, lane 8), as was determined previously (Schmidt et al., 1992).

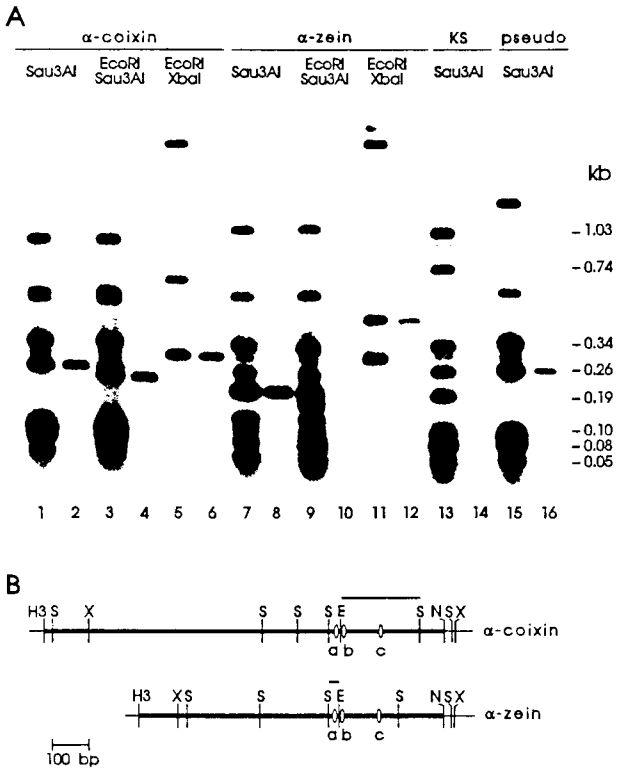


Figure 2. Selective Binding of the O2 Protein to Restriction Fragments from the 22-kD-like α-Coixin and 22-kD α-Zein Promoter Sequences.

(A) Plasmids containing the promoter sequence from α-zein *22Z-4* (Pα-Z4103), α-coixin *α-3B* (Pα-Cx103), α-coixin pseudogene (pseudo) *α-3A* (Pα-PCx103), and pBluescript KS+ (KS) vector were digested with Sau3AI, EcoRI and Sau3AI, or EcoRI and XbaI, end labeled, and incubated with the immunoselected β-Gal::O2 fusion protein. Odd-numbered lanes show the labeled restriction fragments that were incubated with the fusion protein. Even-numbered lanes show the fragments that were specifically retained by the β-Gal::O2 fusion. The lengths of the pBluescript KS+ Sau3AI fragments are given at right in kilobases.

(B) Partial restriction map of the promoter region from the 22-kD-like α-coixin *α-3B* and 22-kD α-zein *22Z-4* genes cloned in Pα-Cx103 and Pα-Z4103, respectively. The putative O2 target sites O2a, O2b, and O2c are represented by ovals. The solid lines above the map indicate the restriction fragment that was bound by the β-Gal::O2 fusion. Abbreviations for restriction sites are as follows: E, EcoRI; H3, HindIII; N, NcoI; S, Sau3AI; X, XbaI.

When the α-zein promoter was digested with EcoRI-Sau3AI, thus separating the O2 boxes present in the 200-bp Sau3AI fragment (Figure 2B), no binding was observed (Figure 2A, lane 10). This confirmed that the O2 protein does not recognize the O2b (and O2c, see below) in the α-zein promoter. When the α-zein promoter was digested with EcoRI-XbaI, only the 455-bp fragment containing the O2a box (Figure 2B) was retained (Figure 2A, lane 12). As shown in Figure 2A (lane 14), no binding was observed to the restriction fragments from the pBluescript KS+ vector included as a negative control.

The clone Pα-PCx103, containing the promoter of the 22-kD-like α-coixin pseudogene *α-3A* (Ottoboni et al., 1993), was also assayed for its retention by the O2 protein. Like the normal gene (*α-3B*), the 260-bp Sau3AI restriction fragment was bound by the O2 protein (Figure 2A, lane 16).

O2 Protein Recognizes the GACATGTC Sequence in the 22-kD α-Coixin Promoter

The site of interaction of β-Gal::O2 on the 260-bp α-coixin promoter fragment was mapped using DNase I footprint, as shown in Figure 3. The assay was performed on both strands by using increasing amounts of the protein A–Sepharose beads containing the immunoprecipitated β-Gal::O2 fusion (Schmidt et al., 1992). Surprisingly, the O2a and O2b boxes were not protected. Instead, a third region that we have designated O2c (Figure 1) was strongly footprinted.

As shown in Figure 3, the footprint observed in each strand produced an overlapping region spanning 23 nucleotides that is located at position −181 to −159 with respect to the ATG initiation codon. According to the assumption that bZIP proteins bind palindromic sequences, the region footprinted in the α-coixin promoter encompasses the perfect palindrome 5′-GACATGTC-3′. As shown in Table 1, the O2c target site of the α-coixin promoter, albeit completely different from that identified in its α-zein relative (Schmidt et al., 1992), has some homology with the O2 target sites B1 and B4 reported for the *b-32* gene (Lohmer et al., 1991).

The nucleotide positions of the α-coixin O2c box were numbered relative to the center of symmetry. Following the nomenclature used previously for the GCN4 binding site (Oliphant et al., 1989), the central two nucleotides A and T of the GACATGTC palindrome were designated as −0 and +0, respectively.

Among the 22-kD-like α-prolamin genes (Figure 1), the α-kafirin clone *pGK1* (DeRose et al., 1989) contains the most homologous O2c box, which differs from that found in the α-coixin promoter by an A-to-C replacement at position −2 (Table 1). Interestingly, the O2c box in the α-coixin pseudogene *α-3A* contains the same modification (Table 1) and is still recognized by the O2 protein (Figure 2A). Because the O2a site in the α-kafirin promoter is quite different from that of the *22Z-4*

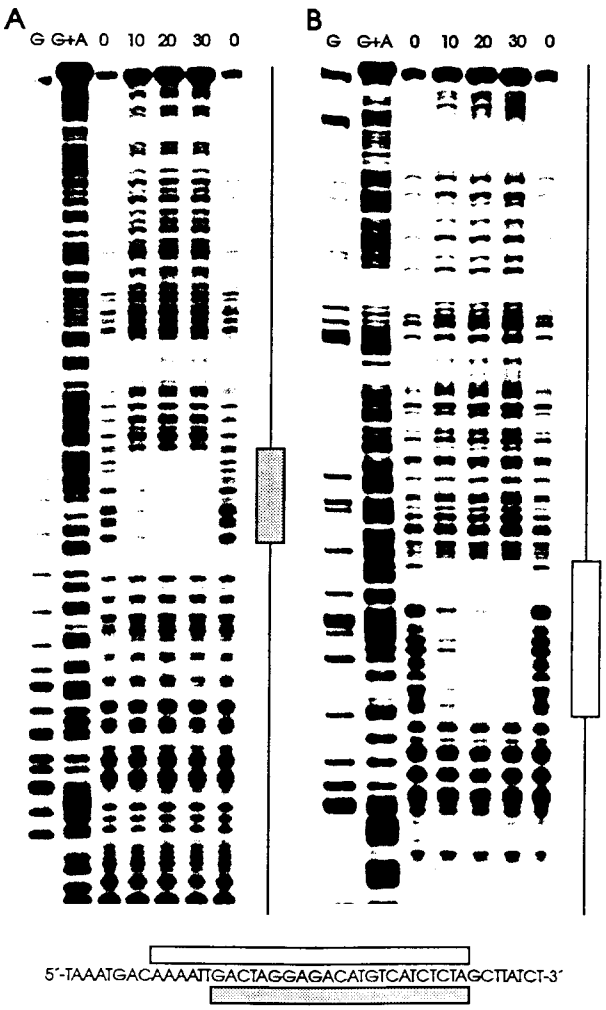


Figure 3. Footprint in Vitro of the O2 Fusion Protein on the 22-kD–like α-Coixin Promoter.

The 260-bp Sau3AI restriction fragment from the *α-3B* α-coixin promoter that was retained by the β-Gal::O2 fusion in the DNA binding assays was used for the footprint analysis. The lanes G and G+A in **(A)** and **(B)** are the Maxam and Gilbert sequencing ladders. The labeled α-coixin promoter fragment was incubated with DNase I without prior incubation with β-Gal::O2 (lanes 0) or with increasing amounts of the fusion protein (lanes 10, 20, and 30).

(A) Footprint analysis performed on the bottom strand.

(B) Footprint along the top strand.

The box beside each panel represents the region of footprint along box O2c. A portion of the sequence spanning box O2c is shown along the bottom of **(A)** and **(B)**. The protected nucleotides are indicated by the shaded and open boxes along the DNA strand given in single-letter code. Note that in **(A)**, the G and G+A sequencing ladders should be read C and C+T, respectively, when referring to the sequence shown along the bottom of the panels.

α-zein clone (Figure 1), we predicted that the O2c box is the preferred binding sequence in the *pGK1* α-kafirin clone.

In the *22Z-4* clone, as in all other 22-kD α-zein genes (Figure 1), the O2c box differs from that of the α-coixin by an A-to-C mutation and a G-to-A mutation at positions −2 and +1, respectively (Table 1). These alterations, as shown in Figure 2A (lane 10), were sufficient to prevent O2 binding to this box under conditions in which strong binding to the α-coixin box was apparent. The other differences in *22Z-4* at positions −4 and −5, although they improved either the similarity with boxes B1 and B4 of *b-32* or the symmetry with the nucleotides at positions +4 and +5, were not sufficient to restore the O2 binding affinity. Therefore, it appeared that within the 23-nucleotide footprinted sequence in the α-coixin promoter, the motif GACATGTC represents the main O2 binding requirement at the nucleotide level. Moreover, although it is not present in the B2, B3, or B5 O2 boxes of the *b-32* gene (Lohmer et al., 1991), the CATG core, as shown in Table 1, seems to be an important consensus motif.

To confirm the footprint result, we performed O2 DNA binding assays on plasmids that contained either the intact 285-bp EcoRI-NcoI α-coixin promoter fragment or an otherwise identical clone from which the sequence CAATAGA from the O2b box had been deleted. These plasmids, designated P285α-C and P285ΔO2b, respectively, are represented in Figure 4B. As shown in Figure 4A, a single restriction fragment of 225 bp in P285α-C (lane 4) and a fragment of 218 bp in P285ΔO2b (lane 6) were retained by the β-Gal::O2 fusion, confirming that the O2 binds to the O2c box and not to the O2b in the α-coixin promoter.

Transcriptional Activation of the 22-kD α-Coixin Promoter Is Mediated by the O2 Protein

Immature Endosperm

To determine whether the binding affinity in vitro of the α-coixin promoter reflected the situation in vivo, we conducted a histochemical spot-counting transient assay in situ. In this assay, immature maize endosperms were bombarded with DNA-

Table 1. Alignment of α-Coixin-like O2 Target Sites and Flanking Sequences

Box	Gene	DNA Sequence	Reference
B1	*b-32*	TGGGAT**GACATGGC**TAATATA	Lohmer et al. (1991)
B4	*b-32*	TATGAT**GACATGAC**-TAT-TA	Lohmer et al. (1991)
O2c	*α-3B*	CTAGGA**GACATGTC**ATCTCTA	This study
O2c	*α-3A*	CTAGGA**GCCATGTC**ATCTCTA	This study
O2c	*pGK1*	CTAGGT**GCCATGTC**ATCTATA	DeRose et al. (1989)
O2c	*22Z-4*	GTAGAT**GCCATATC**ATCTATA	Schmidt et al. (1992)

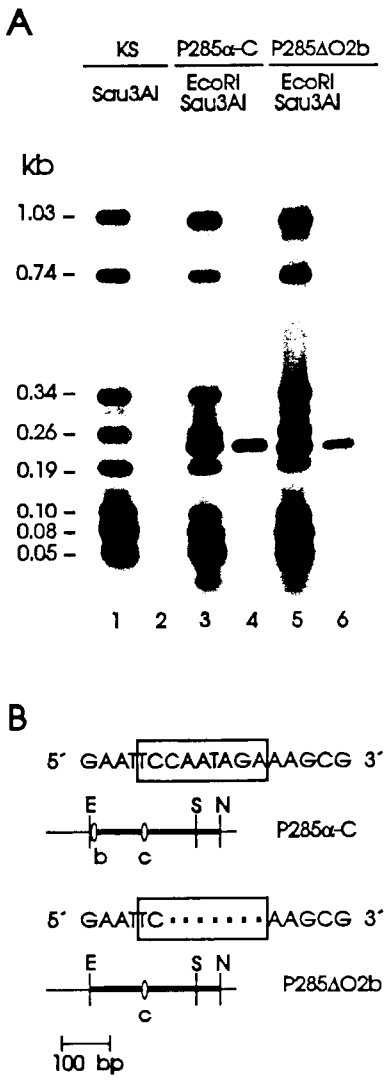


Figure 4. Binding of the O2 Protein to the Wild Type and Mutant 22-kD-like α-Coixin Promoter Fragment.

(A) A plasmid containing the wild 285-bp α-coixin promoter fragment (P285α-C), a mutant version in which the O2b box was deleted (P285ΔO2b), and the pBluescript KS+ (KS) vector was digested with EcoRI and Sau3AI or Sau3AI, end labeled, and incubated with the immunoselected β-Gal::O2 fusion protein. Odd-numbered lanes show the labeled restriction fragments that were incubated with the fusion protein. Even-numbered lanes show the fragments that were specifically retained by the β-Gal::O2 fusion. The lengths of the pBluescript KS+ Sau3AI fragments are given at left in kilobases.

(B) Partial restriction map of the subcloned promoter region from the α-coixin *α-3B* gene. The putative O2 target sites O2b and O2c are represented by ovals. The wild-type sequence spanning the O2b box (boxed sequence) is represented in single-letter code over its corresponding restriction map. The deleted nucleotides in the O2b box of P285ΔO2b (deleted version of boxed sequence shown above) are shown as points. Abbreviations for restriction sites are as follows: E, EcoRI; N, NcoI; S, Sau3AI.

coated microprojectiles. The success of this method in the analysis of promoter–regulatory protein interaction has been demonstrated previously (Oeda et al., 1991; Roth et al., 1991).

Figure 5 shows the different constructs used to analyze the role of the O2a and O2c boxes of α-coixin and α-zein promoters. Such constructs consist of reporter plasmids expressing the β-glucuronidase (*GUS*) gene under the control of the intact promoters (Pα-CxGUS and Pα-Z4GUS in Figure 5) or deleted versions containing only the fragment from the EcoRI site to the ATG codon (P285α-CGUS and P276α-Z4GUS in Figure 5). The deleted versions lacked the O2a box, the prolamin-box, and any other possible upstream *cis*-acting element. As a negative control, we used a plasmid in which the *GUS* gene was placed under the control of the α-coixin promoter fused in the inverted orientation (Pα-Cs[–]GUS in Figure 5). These constructs were delivered into immature maize endosperm alone or together with the effector plasmid (pRT101-O2) expressing the intact O2 protein from maize.

As shown in Figure 6, the intact α-coixin (bar A) and α-zein (bar C) promoters conditioned similar levels of activities. When the sequence upstream of the EcoRI site was deleted from both promoters, the activity was reduced ~50% for the α-zein (Figure 6, bar D), whereas for the α-coixin the activity was not significantly ($P < 0.05$) altered in relation to the intact promoter (Figure 6, bar B). These results agree with the DNA binding assay, indicating that the preferred O2 target site in the α-zein promoter is the O2a box, whereas in the α-coixin it is the O2c box. As expected, no activity was observed from the negative control construct (Figure 6, bar F).

When cotransformed with the effector plasmid pRT101-O2, a reproducible level of enhancement in GUS activity was observed for all the reporter plasmids (Figure 6, bars A+G to D+G), exempting the negative control (Figure 6, bar F+G). In addition, after cotransfection, blue spots were also observed in the pericarp, as shown in Figure 7 (compare Figures 7A, 7C, 7F, and 7H with 7B, 7D, 7G, and 7I, respectively), unequivocally indicating the occurrence of O2-mediated transactivation.

Strikingly, transactivation was also observed for the P276α-Z4GUS construct, suggesting that another sequence of the −276-bp α-zein promoter may also be recognized by the O2 protein in vivo. Because the −276-bp promoter lacks its preferred O2a target site in this construct, this enhancement of expression was expected to be lower than that observed for the intact α-zein *22Z-4* promoter (Pα-Z4GUS), which contains both O2 boxes. However, as can be seen in Figure 6 (bars C+G and D+G), both constructs reached the same level of expression. This could be explained by a condition in which a large amount of O2 protein, which is transiently expressed by pRT101-O2, increased the chance of O2 binding to the putatively less efficient O2 target site of the α-zein promoter.

Because the O2 protein is not able to transactivate the rice *actin1* promoter (data not shown), our results suggested a model in which the O2-mediated enhancement of transcription is a consequence of direct interaction with these prolamin gene promoters and not due to a general increase in transcription.

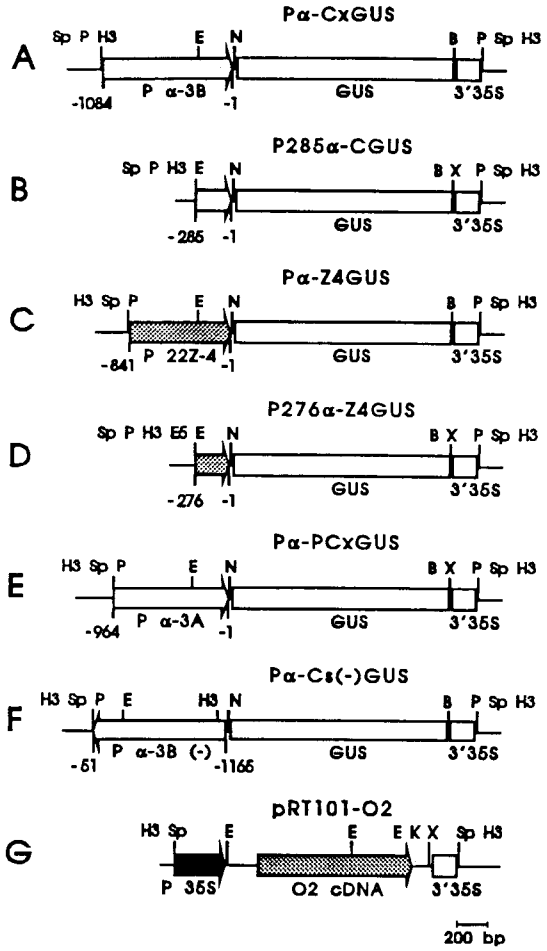


Figure 5. Schematic Representation of Chimeric *GUS* Constructs Used To Analyze the Interaction between the O2 Protein and the 22-kD–like α-Coixin and 22-kD α-Zein Promoters.

Distances are relative to the A residue of the initiator ATG codon. Unique restriction site enzymes are B, BamHI; E, EcoRI; E5, EcoRV; H3, HindIII; K, KpnI; N, NcoI; P, PstI; Sp, SphI; X, XbaI.

(A) Pα-CxGUS, the *GUS* + 35S poly(A) cassette under the control of a 1.08-kb *α-3B* coixin promoter.

(B) P285α-CGUS, the *GUS* + 35S poly(A) cassette under the control of a 285-bp *α-3B* coixin promoter fragment that contains only the putative boxes O2b and O2c.

(C) Pα-Z4GUS, the *GUS* + 35S poly(A) cassette under the control of a 0.84-kb *22Z-4* zein promoter.

(D) P276α-Z4GUS, *GUS* + 35S poly(A) cassette placed under the control of a 276-bp *22Z-4* zein promoter fragment that contains only the putative boxes O2b and O2c.

(E) Pα-PCxGUS, a 0.96-kb fragment of the *α-3A* coixin pseudogene promoter controlling *GUS* + 35S poly(A) reporter gene expression.

(F) Pα-Cs(−)GUS, a 1.1-kb *α-3B* coixin promoter in a reverse orientation driving *GUS* + 35S poly(A) reporter gene expression.

(G) The effector plasmid pRT101-O2 consisting of an *O2* cDNA from clone *O2cDNA1-4* under the control of the CaMV 35S promoter and 35S poly(A) signal. The translated sequence of the *O2* cDNA is represented as a shaded arrowhead.

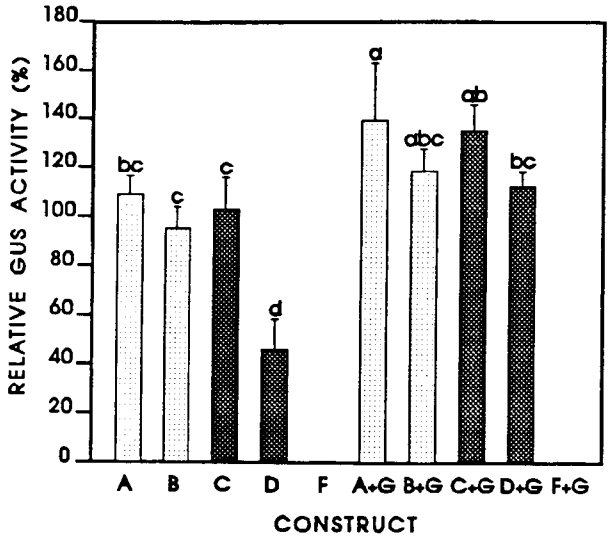


Figure 6. Transient Expression from the 22-kD–like α-Coixin and 22-kD α-Zein Whole and Truncated Promoters Vectors in Immature Maize Endosperm.

GUS activity was normalized according to the expression of the pACT1-F vector (100% activity equals average 673 [± 92] blue spots per plate). The bars correspond to activity of the construct with the same nomenclature as given in Figure 5. Bars A to D and F indicate activity of reporter constructs transfected alone, and bars A+G to F+G indicate activity upon cotransfection with the effector plasmid (Figure 5G) in a 1:2 molar concentration. The values represent the mean GUS activity, and standard deviation is represented by an error bar. Lowercase letters over the bars represent the significant differences ($P < 0.05$) as determined by analysis of variance and SNK multiple range tests (Steel and Torrie, 1980).

Tobacco Mesophyll Protoplasts

To ensure that the pRT101-O2 enhancement of expression from the α-coixin promoter is not a mere consequence of repressor titration by heterodimerization (Pysh et al., 1993), we conducted transient expression assays using the same set of constructs in tobacco mesophyll protoplasts. Our rationale was based on the assumption that tobacco does not contain an endosperm repressor. Moreover, by using tobacco protoplasts we eliminated the interference of both the endogenous O2 protein and other endosperm-specific *trans*-acting factors.

The GUS activity driven by the α-coixin and α-zein promoters was greatly enhanced upon cotransfection with the pRT101-O2 effector plasmid, as shown in Figure 8. The levels of activity from both α-coixin and α-zein intact promoters were statistically similar (Figure 8, bars A+G and C+G). Both truncated constructs were significant less efficiently transactivated by the O2 protein than the whole promoters (Figure 8, bars B+G and D+G), although the upstream deletion in the α-coixin promoter was less detrimental to the transactivation efficiency than the same deletion in the α-zein promoter. In fact, the level of

GUS activity obtained after cotransformation of the truncated α-coixin promoter was approximately twice that observed for the truncated α-zein. No activity was obtained when using the construct with the inverted α-coixin promoter (Figure 8, bars F and F+G).

These results confirmed those observed in the transient expression assay in immature endosperm. The O2a box accounted for most of the O2 transactivation of the α-zein promoter, whereas the O2c box accounted for most of the O2 transactivation of the α-coixin promoter. However, the following two questions can be raised. Why was the truncated α-zein promoter, which does not contain the O2a box, substantially transactivated? Why did the whole α-coixin promoter result in higher GUS activity when compared to the truncated version if the O2c box is already present in the deleted promoter? It appeared that in our transient expression assay, not only the preferred O2 target site but also other sequence motifs are involved in the transactivation of both promoters. Such sequences may represent alternative weakly O2 binding sites.

O2-Mediated Transcription of an 22-kD α-Coixin Pseudogene Promoter

The presence of in-frame stop codons is a common feature of multigene families containing tandemly repeated gene blocks. They are called pseudogenes, and they have been found in several cereal storage protein genes. The presence of such "premature" stop codons, however, does not imply that the gene is inactive. In fact, zein cDNA clones containing in-frame stop codons have been isolated, indicating that such pseudogenes may not be transcriptionally silent (Viotti et al., 1985). Recently, it was shown that a C-hordein pseudogene is transcriptionally active and that its amber codon can be partially suppressed during translation (Entwistle et al., 1991). In contrast, α-glutenin (Halford et al., 1989) and γ-gliadin (Rafalski, 1986) pseudogenes of wheat are reported to be transcriptionally silent.

A genomic clone containing three contiguous 22-kD–like α-coixin genes, two of which (*α-3A* and *α-3C*) harbor a

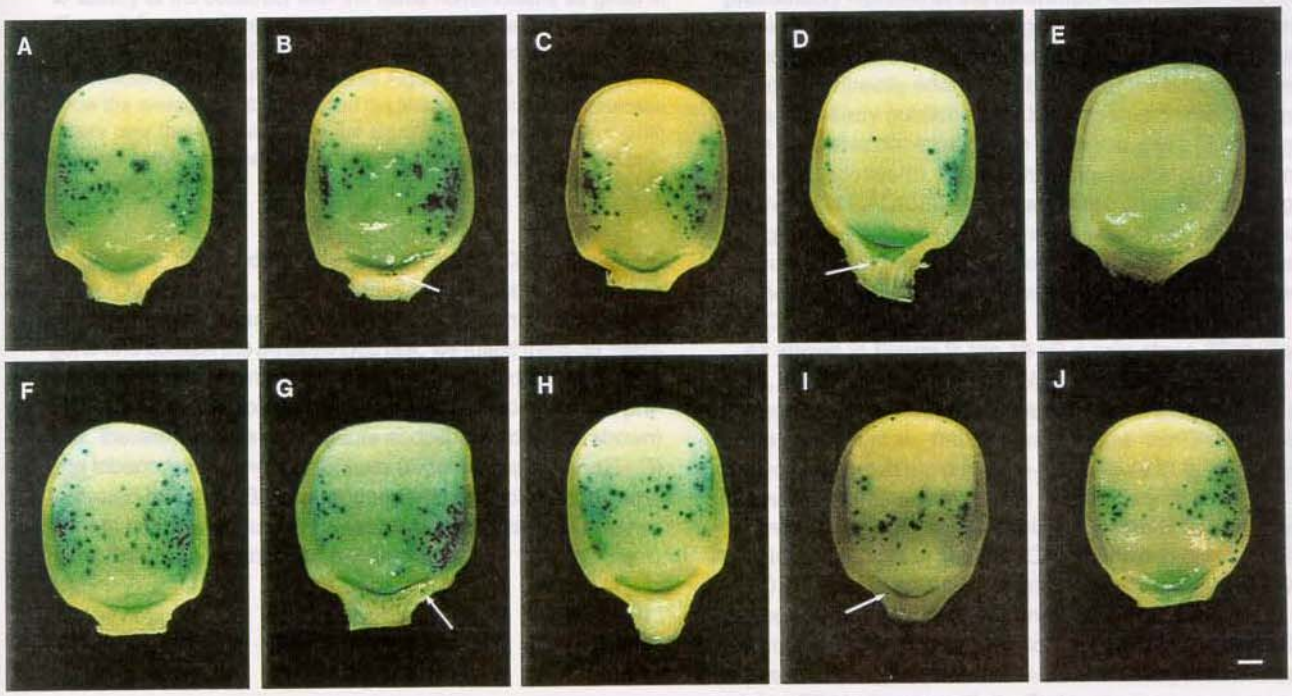


Figure 7. Spatial Distribution of GUS Activity in Immature Maize Endosperms Following Microprojectile Bombardment with the 22-kD–like α-Coixin and 22-kD α-Zein Constructs.

(A) and **(B)** Expression of the Pα-CxGUS (see Figure 5) reporter construct.
(C) and **(D)** Expression of the P285α-CGUS.
(E) Expression of Pα-Cs(−)GUS. The α-coixin promoter in a reverse orientation fails to drive GUS expression.
(F) and **(G)** Expression of Pα-Z4GUS.
(H) and **(I)** Expression of P276α-Z4GUS.
(J) Expression of pAct1-F (McElroy et al., 1990). The constitutive rice *actin1* promoter, as expected, showed nontissue-specific activity. Bar = 1 mm.
(B), **(D)**, **(G)**, and **(I)** Reporter construct expression upon cotransfection with the effector plasmid pRT101-O2. Note that in this case the pericarp showed blue spots (arrows).

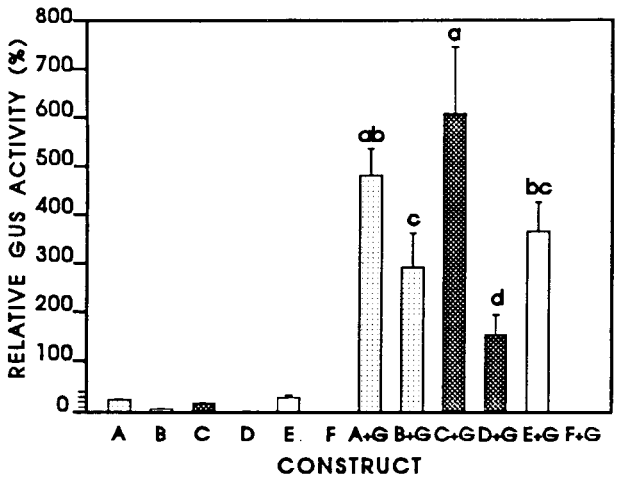


Figure 8. Transactivation of Whole and Truncated 22-kD–like α-Coixin and 22-kD α-Zein Promoters by O2 Protein in Tobacco Protoplasts.

GUS activity was normalized according to the expression derived from the pRT103GUS vector (average 165 nmol of 4-methylumbelliferone per minute per milligram of protein = 100%). The bars correspond to activity of the construct with the same nomenclature as given in Figure 5. Bars A to F represent the activity of reporter plasmids transfected alone. Bars A+G to F+G show activity upon cotransfection with the effector plasmid pRT101-O2 (see Figure 5). The values indicated show the average GUS activity and the standard deviation. Lowercase letters over the bars represent the significant differences ($P < 0.05$) between activities as determined by analysis of variance and SNK multiple range tests.

"premature" stop codon, was isolated from a *Coix* genomic library (Ottoboni et al., 1993). Because the promoter sequences of both pseudogenes are nearly identical to the normal *α-3B* gene and contain the putative O2c box, we fused the *α-3A* pseudogene promoter with the *GUS* gene (Pα-PCxGUS in Figure 5) and tested its propensity to be transactivated by the O2 protein. Transient assays in immature endosperm (data not shown) and tobacco mesophyll protoplasts (Figure 8, bars E and E+G) indicated that the *α-3A* pseudogene promoter drives the same level of expression and is transactivated in a similar manner with respect to the normal *α-3B* gene.

DISCUSSION

Comparisons of the 22-kD–like α-prolamin genes from maize, *Coix*, and sorghum have shown that they share highly conserved protein structure and putative *cis*-acting regulatory elements. The presence of *O2* homologous sequences in the *Coix* and sorghum genome (Ottoboni et al., 1993) indicates that the O2 protein may represent a conserved regulatory *trans*-acting factor. In fact, we have isolated an *O2*-like clone from a *Coix* cDNA library (A. Vettore, unpublished data). However, a perfect counterpart to the recently identified O2 target sequence 5′-TCCACGTAGA-3′ present in the 22-kD α-zein clone *22Z-4* (Schmidt et al., 1992) could not be found in the homologous α-prolamin promoters. When all of the 22-kD–like α-prolamin promoters were compared (Ottoboni et al., 1993), two homologous O2 putative target sites were found in each gene. In the α-coixin promoter, such target sites are represented by the sequences 5′-TCCTCATGAA-3′ (box O2a) and 5′-TCCAATAGA-3′ (box O2b), located at positions −301 to −291 and −281 to −273, respectively (Figure 1).

The Maize O2 Protein Recognizes a *b-32*-like Target Site in the 22-kD α-Coixin Promoter

In this study, we used DNA binding assays and DNase I footprint analysis to demonstrate that the maize O2 protein recognizes neither box O2a nor O2b in the α-coixin promoter but rather a third sequence, which we designated the O2c box. The overlapped region protected by the O2 protein on both strands of the α-coixin promoter comprises 23 nucleotides and has an inner palindromic sequence 5′-GACATGTC-3′, which presumably represents the most important requirement for O2 binding effectiveness. Although we did not perform a mutational analysis of the O2c box, further support for its importance could be indirectly addressed because (1) it is highly homologous to already published O2 binding sites in the *b-32* gene; (2) in an in vivo dimethylsulfate footprinting of a glutenin gene in wheat endosperm, it was demonstrated that a region including the motif GACATGTA is protected at the time when the gene is highly transcribed (Hammond-Kosack et al., 1993); and (3) the naturally occurring mutations in the GACATGTC-like sequence of the 22-kD α-zein promoter were sufficient to prevent the O2 protein binding.

As shown in Figure 9, the 19-kD α-zein clone *ZE19* (Spena et al., 1982) contains three highly homologous GACATGTC sequences. Such sequences are also found in other 19-kD α-zein clones (Brown et al., 1986a). One of these, the CACATGTG sequence located at the prolamin-box (Figure 9), is also present in several 22-kD α-zein genes (see Figure 1).

However, in spite of the exhibited homology, no apparent O2 binding occurs in such GACATGTC-like sequences of the 19-kD or in that of the 22-kD α-zein clones (Schmidt et al., 1992). Because the CATG core is conserved in all sequences that are not recognized by O2, we concluded that the nucleotides at positions −3, −2, +2, and +3 are also important in determining O2 binding specificity. Strikingly, even a variation that preserves the palindromic trait, as in the case of the CACATGTG sequence mentioned above, can be sufficient to abolish O2 binding. Paradoxically, a mutation that disrupts the palindrome, as occurred with the O2c box of the *α-3A* coixin gene (Figure 9), does not prevent O2 binding (Figure 2A). Therefore, it may be assumed that not only the palindromic trait but also some nucleotides at certain positions play an important role in the DNA–O2 protein interaction.

Interestingly, the protected sequences flanking the GACATGTC palindrome in the *α-3B* O2c box comprise two other smaller motifs (Figure 9). The first is TGACT, which resembles the GCN4 binding site (Hill et al., 1986), and the second is CATCTCTA, a sequence common to the 5′-flanking region of prolamin genes from several cereals, called CATC box (Kreis et al., 1986). The relevance of such motifs on the O2 protein binding is under investigation.

Surprisingly, the GACATGTC sequence showed high homology with the O2 binding sites B1 and B4 identified in the promoter of the *b-32* gene (Lohmer et al., 1991). Such a finding helps to resolve the conflicting results concerning the O2 target site reported by two different laboratories (Lohmer et al., 1991; Schmidt et al., 1992). By using the same experimental procedures and conditions reported for the identification of the O2 box in the α-zein promoter (Schmidt et al., 1992), we found a *b-32*–like target site in the α-coixin promoter. This clearly denotes that the apparently conflicting results cannot be attributed to the different experimental conditions (Schmidt et al., 1992; Izawa et al., 1993), and it is more likely that all the binding sites identified thus far can be recognized by the O2 protein even though they have different binding affinities.

The fact that in the 22-kD prolamin genes of maize and *Coix*, two different sequences can be recognized by the O2 protein raises a very interesting evolutionary issue. Because the 22-kD class of prolamin-coding genes are present in multicopy in both plants, one would suppose that the duplication events giving rise to these genes would have occurred before the events that separated the two plant species. Therefore, the regulatory system between the O2 and its prolamin-dependent genes would be the same in the two plant species. Our data support the fact that they are not. In addition, the *pGK1* 22-kD α-prolamin gene of sorghum exhibits an α-coixin–like O2 box. The issue could be solved if in the 22-kD α-prolamin multigene family both kinds of O2 target sites were present in the ancestor of maize, sorghum, and *Coix*. Increasing the number of sequences of 22-kD–like α-prolamin promoters could provide more information to confirm our hypothesis. Interestingly, the *ψgZ22.8* α-zein clone seems to be an intermediate between the *α-3B* α-coixin and the *22Z-4* α-zein genes because it contains an *α-3B*–like O2a box and a *22Z-4*–like O2c box (see Figure 1). In contrast, the contiguous *gZ22.8* α-zein clone (Thompson et al., 1992) presents the O2a and the O2c target sites similar to those of the *22Z-4* α-zein clone (Figure 1).

```
                        |  Footprinted region   |
α-3B   -188  CAAAATTGACTAGGA GACATGTC ATCTCTAGCTTA
α-3A   -190  CAAAATCGACTAGGA GCCATGTC ATCTCTAGCTTA
ZE19   -185  CAAAACCAACTAGAT ACCATGTC ATCTCTACCTTA
       -309  GTCACAACATTGTCA CCCATGTA TTTGGACAATAC
       -344  ATGTGGCTATCGTTA CACATGTG TAAAGGTATTGC
22Z-4  -340  TCATGTTAAGGTTGT CACATGTG TAAAGGTGAAGA
```

Figure 9. Comparison of the α-Coixin O2c Site with Homologous Sequences of 22- and 19-kD Zein Promoters from *22Z-4* and *ZE19*.

Positions are relative to the translation start. The O2 target sequence footprinted in the α-coixin promoter is indicated. The GCN4-like motif, the O2c palindromic core, and the CATC box are indicated by underline, boldface italic letters, and a double underline, respectively. The sources are *α-3A*, *α-3B* (Ottoboni et al., 1993), *ZE19* (Spena et al., 1982), and *22Z-4* (Schmidt et al., 1992).

The O2 Protein Binds Target Sites without ACGT Core

Without losing binding affinity, the O2 protein supports a relatively high degree of degeneracy at the nucleotides flanking the ACGT core (Izawa et al., 1993). However, the replacement of the ACGT core by ACAT was sufficient to abolish binding in vitro (Schmidt et al., 1992) and transactivation in vivo (Ueda et al., 1992) of the α-zein promoter by the O2 protein. Nevertheless, this conflicts with the fact that the B1, B2, B3, and B4 sites to which O2 is reported to bind in the promoter of the *b-32* gene (Lohmer et al., 1991) and the O2c target site of the α-coixin do not contain an ACGT core. Thus, although most of the characterized plant bZIP proteins recognize the ACGT core, it does not imply that they necessarily require this core. For example, the TGACGT/C binding proteins also interact with the TGACGC motif of the cauliflower mosaic virus (CaMV) 35S promoter (Katagiri et al., 1989). Moreover, in most cases, the importance of the ACGT core was analyzed by mutations that create asymmetric cores (Oeda et al., 1991; Schindler et al., 1992; Izawa et al., 1993), whereas in the O2c α-coixin box, the CATG core remains perfectly symmetric. This does not mean, however, that the binding will remain unaffected by any alteration to the ACGT core that preserves symmetry. For example, the substitution of the ACGT core by GTAC in boxes II and III of the chalcone synthase parsley promoter prevents the recognition by the common plant regulatory factor bZIP proteins (Armstrong et al., 1992).

Can Low-Affinity Binding Sites in Vitro Mediate Transcriptional Activation in Vivo?

Using immature maize endosperm and tobacco protoplast, we demonstrated that the transiently coexpressed maize O2 protein enhanced the activity of the heterologous 22-kD–like α-coixin promoter at the same level as it does for the maize 22-kD α-zein promoter. Furthermore, even when the region upstream from the EcoRI site was deleted, thus preserving only the O2c box, the O2-mediated enhancement of expression was maintained on both promoters. Nevertheless, the resulting expression level was lower than that obtained with the nondeleted promoters. In this case, however, the truncated α-coixin promoter was transactivated to a higher level than the truncated

α-zein. From this data, we proposed that both the O2a box in the α-zein promoter and the O2c in the α-coixin promoter do not represent the unique *cis* elements (albeit being the most important) responsible for the resulting O2 transactivation.

In the binding assay in vitro, however, only the O2a box of the α-zein and O2c box of the α-coixin strongly interacted with the O2 protein. Because the transactivation experiments performed in maize endosperm and tobacco protoplasts in vivo gave similar results, an artifact seemed unlikely. This discrepancy could be explained if the conditions of the binding assay in vitro were too restrictive to detect weak but selective interaction of the protein to the target fragment. In fact, when we performed a footprint analysis of the α-coixin promoter without the DNA competitor poly(dI-dC), the sequence CATGCATG, located just downstream of the prolamin-box (Figure 1), was weakly protected by the β-Gal::O2 protein (data not shown).

It is also possible that the protein–DNA complex was stabilized in vivo by other proteins not present in the assay in vitro. In this instance, binding affinity may not be a good indication of transcriptional rates, the biologically relevant parameter. For example, in the ethyl methanesulfonate–induced *o2* mutant *o2-676* (Aukerman et al., 1991), members of the 22-kD α-zein class are expressed, while the rescued O2-676 protein is unable to bind the 22-kD α-zein promoter in vitro.

The results presented here showed that the O2 protein recognizes and binds to alternative sequences, even if they lack the ACGT core. This corroborates the hypothesis that one factor may have different affinities for multiple sites and points to the necessity of combining binding experiments in vitro with expression assays in vivo to dissect the biologically relevant parameters involved in the protein–DNA interaction. Moreover, we stress the importance of finding among related genes the occurrence of naturally preserved variants of functional *cis* elements to obtain insights into the understanding of the evolution of gene regulation mechanisms.

METHODS

Construction of Plasmids

All sequence position numbers refer to the ATG initiation codon. The pACT1-F plasmid (McElroy et al., 1990), which was used as a positive control in the endosperm transient expression assay, contained the β-glucuronidase (*GUS*) gene under the control of the strong constitutive rice *actin1* promoter.

P285α-CGUS is a 0.3-kb fragment that was amplified from an EcoRI-StuI *α-3B* promoter subclone (Ottoboni et al., 1993) using the universal sequencing primer and the primer Alfacoix–ATG 5′-TGCTGGATCCATGGCGACAAGATGCTGCT-3′ (New England BioLabs, Inc., Beverly, MA), which is complementary to the −18 to +8 fragment and creates NcoI and BamHI restriction sites over the ATG. The amplified fragment was digested with EcoRI and BamHI and inserted into pBluescript KS+ (Stratagene). The resulting clone *P285α-C* was verified by sequencing of both strands. The cloned fragment was then recovered by digestion with HincII and NcoI and inserted into the pRT103 plasmid (Töpfer et al., 1987), which was digested with the same enzymes, replacing the cauliflower mosaic virus (CaMV) 35S promoter. The *GUS* coding region was excised from pJII140 (Gallie et al., 1987) by digestion with NcoI and BamHI and fused in the corresponding sites among the 285-bp *α-3B* promoter and the CaMV 35S poly(A) signal.

P285ΔO2b is the 285-bp promoter fragment of P285α-C that was amplified using the universal sequencing primer and the primer Alfacoix-272 5′-CTGCGAATTCAAGCGACACTTTTTTCTT-3′ (New England BioLabs, Inc.), which is complementary to the −285 to −262 and creates a deletion of the CAATAGA sequence in the O2b box. The amplified fragment was digested with EcoRI and BamHI and inserted into pBluescript KS+. The resulting clone was verified by sequencing both strands.

Pα-CxGUS is the *α-3B* promoter fragment that spans residues from −1084 to −285 and was excised from the *α-3C* coixin clone (Ottoboni et al., 1993) by digestion with HindIII and EcoRI. The 0.8-kb fragment was ligated into the corresponding sites of P285α-C, giving rise to the whole *α-3B* promoter clone Pα-Cx. Afterward, this promoter was recovered by HincII and NcoI digestion and inserted at the same sites of pRT103 replacing the CaMV 35S promoter, which resulted in the *Pα-Cx103* clone. The *GUS* gene was inserted as given above.

Pα-Cs(−)GUS is an EcoRI-Sau3AI fragment spanning nucleotides −285 to −50 of the *α-3B* (Ottoboni et al., 1993) promoter. It was subcloned in the EcoRI-BamHI fragment of the pBluescript KS+ vector. Subsequently, the upstream promoter fragment from −1165 to −285, which was obtained by digesting the *α-3C* clone (Ottoboni et al., 1993) with HincII and EcoRI, was joined, resulting in the *α-3B* promoter clone Pα-Cs. The whole fragment was excised with HincII and SstI, blunt ended with T4 DNA polymerase, and ligated with pRT103, which was linearized with HincII and XhoI and blunt ended with T4 DNA polymerase. Recombinant plasmids were screened by replica plate hybridization, and those with the promoter in the reverse orientation with respect to the CaMV 35S poly(A) signal were identified by restriction mapping. The *GUS* gene was then inserted as mentioned previously.

Pα-PCxGUS is an XbaI clone containing the pseudogene *α-3A* (Ottoboni et al., 1993) from positions −962 to +1081. It was used as a template for polymerase chain reaction amplification of the promoter region. The primers used were the reverse sequencing primer and the Alfacoix-ATG oligonucleotide. The amplified product was cloned using the TA cloning system K2000-01 (Invitrogen, San Diego, CA) and sequenced on both strands. After digestion with XbaI and treatment with T4 DNA polymerase, the promoter fragment was excised with NcoI and used to replace the CaMV 35S promoter of pRT103 linearized with HincII and NcoI, resulting in the *Pα-PCx103* clone. Subsequently, the *GUS* gene was cloned into the NcoI-BamHI sites as described above.

P276α-Z4GUS is the fragment spanning nucleotides −276 to +1 of the *22Z-4* α-zein gene (Schmidt et al., 1992) that was excised as an EcoRI-NcoI fragment from the vector PZ4Luc (kindly provided by R.J. Schmidt, University of California at San Diego) and used to replace the α-coixin promoter of P285α-CGUS.

Pα-Z4GUS was derived from the PZ4Luc plasmid that was digested with AccI, treated with the Klenow fragment of DNA polymerase I, and then restricted with NcoI to obtain an 840-bp *22Z-4* promoter sequence spanning nucleotides −841 to +1. Subsequently, in a tripartite ligation, this fragment was fused to pRT103 digested with HincII-BamHI and the NcoI-BamHI *GUS* coding sequences. The *Pα-Z4103* clone used in the DNA binding assay was obtained by fusing the above mentioned 840-bp *22Z-4* fragment to pRT103 digested with HincII and NcoI.

pRT103GUS was obtained by ligating the NcoI-BamHI *GUS* fragment into the corresponding sites of pRT103.

pRT101-O2 is the complete maize *Opaque2* (*O2*) cDNA from clone *O2cDNA1-4* (Schmidt et al., 1990) obtained by partial digestion with EcoRI and subsequently inserted into pRT101 (Töpfer et al., 1987) linearized with EcoRI and treated with calf intestinal alkaline phosphatase (New England BioLabs, Inc.) according to the manufacturer's instructions. A clone in which the *O2* cDNA entered in the correct orientation was identified by restriction mapping.

Transient Expression Assays in Immature Maize Endosperm

The commercial maize hybrid F-352 from Sementes Agroceres S/A (São Paulo, Brazil) was grown at the experimental field of University of Campinas and self-pollinated. Ears were harvested at 15 days after pollination, surface sterilized for 15 min with 5% commercial bleach, and rinsed four times in distilled water. Seeds were dissected from the cob and sectioned longitudinally; 16 sections were flattened on 100-mm-diameter Petri dishes containing 20 mL of MS medium (Murashige and Skoog, 1962), with the sliced surface facing upward. Five micrograms of CsCl-purified DNA alone or coprecipitated with 10 μg pRT101-O2 was used to coat 3 mg of 1- to 3-μm-diameter gold particles according to the method of Ye et al. (1990). The endosperms were bombarded twice with 0.5 μg DNA (or 1.5 μg when cotransfected) using a homemade helium particle delivery device (Ye et al., 1990). After bombardment, the samples were incubated for 48 hr in the dark at 27 ± 1°C. The endosperms were then stained for GUS activity according to the method of Jefferson (1987). To minimize experimental errors, all constructs were evenly analyzed using seeds of the same ear.

Transient Expression Assays in Tobacco Mesophyll Protoplasts

Nicotiana tabacum cv SR1 plants were grown in axenic conditions on MS salts (Murashige and Skoog, 1962). Mesophyll protoplasts were prepared from young leaves as described by Saul et al. (1988). Samples of 0.5 × 10^6 protoplasts were used for PEG transformation (Saul et al., 1988; Negrutiu et al., 1990). Ten micrograms of Pα-CxGUS DNA, or equimolar amounts of the other constructs, plus sonicated salmon sperm DNA to a final amount of 43 μg were used in the transformation of each aliquot of protoplast. After incubation for 24 hr, the protoplasts were pelleted, lysed in GUS extraction buffer (Jefferson, 1987), and centrifuged; the supernatant was used for protein estimation (Bradford, 1976). One microgram of protein was used in the fluorometric GUS assay (Jefferson, 1987).

DNA Binding Assay

The DNA binding assay was performed with a β-galactosidase–O2 (β-Gal::O2) fusion protein according to the method of Schmidt et al. (1992). Approximately 40 ng of ^{32}P-labeled Sau3AI, EcoRI-Sau3AI, or EcoRI-XbaI digests of plasmids Pα-Cx103, Pα-Z4103, Pα-PCx103, P285α-C, P285ΔO2b, and pBluescript KS+ (used as a negative control) was incubated with 15 μL protein A–Sepharose beads containing the β-Gal::O2 immunocomplex for 90 min at room temperature with intermittent agitation. Washes, processing, and agarose gel electrophoresis of the retained fragment were performed as described by Schmidt et al. (1992).

DNase I Footprint

The 260-bp Sau3AI restriction fragment from *α-3B*, which contained all the putative O2 target sites, was cloned into the BamHI site of pBluescript KS+. Either a HindIII-NotI or an EcoRV-XbaI fragment containing the 260-bp sequence and flanking polylinker sites was isolated from an agarose gel and 3′ end labeled by partial fill-in with α-^{32}P-dATP, the other deoxynucleotide triphosphates (when required), and the Klenow fragment of DNA polymerase I, thereby labeling the bottom or the top strand, respectively. Note that in contrast to the T4 polynucleotide kinase labeling procedure in the Klenow method that was just described, the 3′ end of the strand is labeled rather than the 5′ end. The labeled fragments were then purified from a 7.5% polyacrylamide gel, and for each reaction, 35,000 cpm of DNA was incubated with 0, 10, 20, and 30 μL of the protein A–Sepharose beads containing the immunoprecipitated β-Gal::O2 fusion plus 30, 20, 10, and 0 μL of the protein A–Sepharose beads, respectively, in a total volume of 50 μL of binding buffer containing 12.5 mM Hepes, pH 7.5, 50 mM KCl, 10% glycerol, 0.05% Nonidet P-40, 0.5 mM DTT, and 0.7 μg poly(dI-dC). Reaction mixtures were incubated for 15 min at 4°C and then digested with 0.3 units of DNase I according to recommendations of the Sure Track Footprinting Kit (No. 27-9101-01; Pharmacia LKB Biotechnology). The resulting fragments were separated by electrophoresis on a 5 or 6% sequencing gel. The same end-labeled DNA was treated with the G+A and G chemical cleavage reaction, according to the Sure Track Footprinting Kit or the Maxam-Gilbert Sequencing Kit (No. SEQ-1; Sigma) manufacturer's recommendations, respectively, and run on lanes adjacent to the footprinting reactions.

ACKNOWLEDGMENTS

We thank Edson L. Kemper for technical help. We are very grateful to Dr. Robert J. Schmidt, University of California at San Diego, for supplying pZ4Luc and *O2cDNA1-4* and for instruction in the DNA binding assays; to Dr. Reinhard Töpfer, Max Planck Institut, Cologne, Germany, for supplying pRT101 and pRT103 vectors; to Diatech Ltd., London, and Dr. T. Michael A. Wilson, Scottish Crop Research Institute, Invergowrie, Dundee, Scotland, for supplying the pJII140 plasmid; and to Dr. Ray Wu, Cornell University, Ithaca, NY, for providing the pACT1-F vector. We also thank Dr. Francisco G. Nobrega, Universidade de São Paulo, São Paulo, Brazil, for instruction in the Maxam-Gilbert reactions. This work was financed by grants to P.A. from Financiadora de Estudos e Projetos (FINEP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)/ Projeto Temático de Equipe (PTE)-90/3808-5. P.A., A.L., and L.M.M.O. received research fellowships from CNPq. J.A.Y. and G.C.N. received postgraduate fellowships from CNPq and Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Received October 4, 1993; accepted December 6, 1993.

REFERENCES

Armstrong, G.A., Weisshaar, B., and Hahlbrock, K. (1992). Homodimeric and heterodimeric leucine zipper proteins and nuclear factors

from parsley recognize diverse promoter elements with ACGT cores. Plant Cell **4**, 525–537.

Aukerman, M.J., Schmidt, R.J., Burr, B., and Burr, F.A. (1991). An arginine to lysine substitution in the bZip domain of an *opaque-2* mutant in maize abolishes specific DNA binding. Genes Dev. **5**, 310–320.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**, 248–254.

Brown, J.W.S., Wandelt, C., and Feix, G. (1986a). The upstream regions of zein genes: Sequence analysis and expression in the unicellular green alga *Acetabularia*. Eur. J. Cell Biol. **42**, 161–170.

Brown, J.W.S., Wandelt, C., Feix, G., Neuhaus, G., and Schweiger, H.G. (1986b). The upstream regions of zein genes: Sequence analysis and expression in the unicellular alga *Acetobacteria*. Eur. J. Cell. Biol. **42**, 161–170.

Burr, B., and 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.

Burr, B., Burr, F.A., John, T.P., Thomas, M., and Davis, R.D. (1982). Zeins: Storage protein family of maize. J. Mol. Biol. **154**, 33–49.

DeRose, R.T., Ma, D.P., Kwon, I.S., Hasnain, S.E., Klassy, R.C., and Hall, T.C. (1989). Characterization of the kafirin gene family from sorghum reveals extensive homology with zein from maize. Plant Mol. Biol. **12**, 245–256.

Di Fonzo, N., Mandocchi, L., Salamini, F., and Soave, C. (1986). Purification and properties of an endospermic protein of maize associated with the *opaque-2* and *opaque-6* genes. Planta **167**, 587–594.

Entwistle, J., Knudsen, S., Müller, M., and Cameron-Mills, V. (1991). Amber codon suppression: The *in vivo* and *in vitro* analysis of two C-hordein genes from barley. Plant Mol. Biol. **17**, 1217–1231.

Esen, A. (1986). Separation of alcohol-soluble proteins (zeins) from maize into three fractions by differential solubility. Plant Physiol. **80**, 623–627.

Gallie, D.R., Sleat, D.E., Watts, J.N., Turner, P.C., and Wilson, T.M.A. (1987). A comparison of eukaryotic viral 5′-leader sequences as enhancers of mRNA expression *in vivo*. Nucl. Acids Res. **15**, 8693–8711.

Hagen, G., and Rubenstein, I. (1981). Complex organization of zein genes in maize. Gene **13**, 239–249.

Halford, N.G., Forde, J., Shewry, P.R., and Kreis, M. (1989). Functional analysis of the upstream regions of a silent and an expressed member of a family of wheat seed protein genes in transgenic tobacco. Plant Sci. **62**, 207–216.

Hammond-Kosack, M.C.U., Holdsworth, M.J., and Bevan, M.W. (1993). *In vivo* footprinting of a low molecular weight glutenin gene (LMWG-1D1) in wheat endosperm. EMBO J. **12**, 545–554.

Hartings, H., Maddaloni, M., Lazzaroni, N., Di Fonzo, N., Motto, M., Salamini, F., and Thompson, R. (1989). The *O2* gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. EMBO J. **8**, 2795–2801.

Hill, D.E., Hope, I.A., Macke, J.P., and Struhl, K. (1986). Saturation mutagenesis of the yeast *his3* regulatory site: Requirements for transcriptional induction and for binding by GCN4 activator protein. Science **234**, 451–457.

Izawa, T., Foster, R., and Chua, N.-H. (1993). Plant bZIP protein DNA binding specificity. J. Mol. Biol. **230**, 1131–1144.

Jefferson, R.A. (1987). Assaying chimeric genes in plants: The *GUS* gene fusion system. Plant Mol. Biol. Rep. **5**, 387–405.

Katagiri, F., Lam, E., and Chua, N.-H. (1989). Two tobacco DNA-binding proteins with homology to the nuclear factor CREB. Nature **340**, 727–730.

Kodrzycki, R., Boston, R.S., and Larkins, B.A. (1989). The *opaque-2* mutation of maize differentially reduces zein gene transcription. Plant Cell **1**, 105–114.

Kreis, M., Williamson, M.S., Forde, J., Schmutz, D., Clark, J., Buxton, B., Pywell, J., Harris, N., Shewry, P.R., Forde, B.G., and Miflin, B.J. (1986). Differential gene expression in the developing barley endosperm. Phil. Trans. R. Soc. Lond. **B314**, 355–365.

Leite, A., Ottoboni, L.M.M., Targon, M.L.P.N, Silva, M.J., Turcinelli, S.R., and 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.

Liu, C.-N., and Rubenstein, I. (1992). Genomic organization of an α-zein gene cluster in maize. Mol. Gen. Genet. **231**, 304–312.

Lohmer, S., Maddaloni, M., Motto, M., Di Fonzo, N., Hartings, H., Salamini, F., and 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.

McElroy, D., Zhang, W., Cao, J., and Wu, R. (1990). Isolation of an efficient actin promoter for use in rice transformation. Plant Cell **2**, 163–171.

Motto, M., Maddaloni, M., Ponziani, G., Brembilla, M., Marotta, R., Di Fonzo, N., Soave, C., Thompson, R., and Salamini, F. (1988). Molecular cloning of the *o2-m5* allele of *Zea mays* using transposon marking. Mol. Gen. Genet. **212**, 488–494.

Motto, M., Di Fonzo, N., Hartings, H., Maddaloni, M., Salamini, F., Soave, C., and Thompson, R.D. (1989). Regulatory genes affecting maize storage protein synthesis. In Oxford Surveys of Plant Molecular and Cell Biology, Vol. 6, B.J. Miflin, ed (Oxford: Oxford University Press), pp. 87–114.

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue culture. Physiol. Plant. **15**, 473–497.

Negrutiu, I., Dewulf, J., Pietrzak, M., Botterman, J., Rietveld, E., Wurzer-Figurelli, E.M., De Ye, and Jacobs, M. (1990). Hybrid genes in the analysis of transformation conditions: II. Transient expression vs. stable transformation—Analysis of parameters influencing gene expression levels and transformation efficiency. Physiol. Plant. **79**, 197–205.

Oeda, K., Salinas, J., and Chua, N.-H. (1991). A tobacco bZip transcription activator (TAF-1) binds to a G-box-like motif conserved in plant genes. EMBO J. **10**, 1793–1802.

Oliphant, A.R., Brandl, C.J., and 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.

Osborne, T.B., and Mendel, L.B. (1914). Nutritive properties of proteins of the maize kernel. J. Biol. Chem. **18**, 1–6.

Ottoboni, L.M.M., Leite, A., Targon, M.L.P.N., Silva, M.J., and Arruda, P. (1990). Heterogeneity of *Coix*, maize and teosinte prolamins detected by isoelectric focusing. Rev. Bras. Genet. **13**, 313–322.

Ottoboni, L.M.M., Leite, A., Yunes, J.A., Targon, M.L.P.N., Souza Filho, G.A., and 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.

Pysh, L.D., Aukerman, M.J., and Schmidt, R.J. (1993). OHP1: A maize basic domain/leucine zipper protein that interacts with Opaque2. Plant Cell **5,** 227–236.

Rafalski, J.A. (1986). Structure of wheat *gamma*-gliadin genes. Gene **43,** 221–229.

Roth, B.A., Goff, S.A., Klein, T.M., and 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.

Saul, M.W., Shillito, R.D., and Negrutiu, I. (1988). Direct DNA transfer to protoplast with and without electroporation. In Plant Molecular Biology Manual, S.B. Gelvin, R.A. Schilperoort, and D.P.S. Verma, eds (Dordrecht: Kluwer Academic Publishers), pp. A1.1–A1.16.

Schindler, U., Beckmann, H., Kadesch, T., and Cashmore, A.R. (1992). DNA binding site preferences and transcriptional activation properties of the *Arabidopsis* transcription factor GBF1. EMBO J. **11,** 1275–1289.

Schmidt, R.J., Burr, F.A., and 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., and 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., and 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.

Shotwell, M.A., and Larkins, B.A. (1989). The biochemistry and molecular biology of seed storage proteins. In The Biochemistry of Plants: A Comprehensive Treatise, Vol. 15, A. Marcus, ed (New York: Academic Press), pp. 296–345.

Soave, C., Righetti, P.G., Lorenzoni, C., Gentinetta, E., and Salamini, F. (1976). Expressivity of *opaque-2* gene at the level of zein molecular components. Maydica **21,** 61–75.

Spena, A., Viotti, A., and Pirrota, V. (1982). A homologous repetitive block structure underlies the heterogeneity of heavy and light chain zein genes. EMBO J. **1,** 1589–1594.

Steel, R.G.D., and Torrie, J.H. (1980). Principles and Procedures of Statistics: A Biometrical Approach, 2nd ed (New York: McGraw-Hill).

Targon, M.L.N., Ottoboni, L.M.M., Leite, A., Ludevid, D., Puigdomènech, P., and Arruda, P. (1992). Synthesis and deposition of coixin in seeds of *Coix lacryma-jobi*. Plant Sci. **83,** 169–180.

Thompson, G.A., Siemieniak, D.R., Sieu, L.C., Slightom, J.L, and Larkins, B.A. (1992). Sequence analysis of linked maize 22 kDa α-zein genes. Plant Mol. Biol. **18,** 827–833.

Töpfer, R., Matzeit, V., Gronenborn, B., Schell, J., and Steinbiss, H.H. (1987). A set of plant expression vectors for transcriptional fusions. Nucl. Acids Res. **15,** 5890.

Ueda, T., Waverczak, W., Ward, K., Sher, N., Ketudat, M., Schmidt, R.J., and M[illegible]ssing, J. (1992). Mutations of 22- and 27-kD zein promoters a[illegible]t transactivation by the Opaque-2 protein. Plant Cell **4,** 701–709.

Viotti, A., Caioro, G., Vitale, A., and Sala, E. (1985). Each zein gene class can produce polypeptides of different sizes. EMBO J. **4,** 1103–1110.

Wandelt, C., and Feix, G. (1989). Sequence of a 21-kD gene from maize containing an in-frame stop codon. Nucl. Acids Res. **17,** 2354.

Ye, G.-N., Daniell, H., and Sanford, J.C. (1990). Optimization of delivery of foreign DNA into higher-plant chloroplasts. Plant Mol. Biol. **15,** 809–819.

Running title: Cooperative DNA binding by Opaque2

The Opaque2 transcription factor binds to two adjacent GCN4-like motifs in a cooperative manner

José A. Yunes[a], André L. Vettore[a], Marcio J. da Silva[a], Adilson Leite[a], and Paulo Arruda[a,b,1]

[a]Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, 13083-970, Campinas, SP, Brazil

[b]Departamento de Genética e Evolução, IB, Universidade Estadual de Campinas, 13083-970, Campinas, SP, Brazil

[1]To whom correspondence should be addressed

Mailing address:

Paulo Arruda

Centro de Biologia Molecular e Engenharia Genética,

Universidade Estadual de Campinas (UNICAMP),

Cidade Universitária "Zeferino Vaz"

Cx.P. 6109

13083-970 - Campinas - SP

Brasil

FAX: (55-19) 239 4717

PHONE: (55-19) 239 8351 or (55-19) 239 7030

E-mail: PARRUDA@TURING.UNICAMP.BR

Summary

The maize Opaque2 (O2) protein is a bZIP factor, which controls the expression of distinct classes of endosperm genes through recognition of different *cis*-acting elements in their promoters. In this paper we demonstrate that the O2 target sequences in the promoter of the α-coixin gene comprise two closely adjacent binding sites related in sequence to the GCN4 binding site. Quantitative DNase footprint analysis using a series of DNA templates having a mutation at one or both sites, or disjointing sites, indicated that O2 binding to the α-coixin target sites is best described by a cooperative model. Simultaneous numerical analysis of footprint titration data from the normal and mutant templates, were used to determine relative thermodynamic constants for local site binding and cooperativity. This analysis indicated that O2 interacts with similar intrinsic affinity with both sites. Some mutations affected cooperativity, while others did not, indicating that cooperative binding is based on other properties in addition to primary sequence recognition. Transient expression assays indicated that the two adjacent sites act synergistically. Comparisons of the transcriptional activities with the resolved relative binding free energies for the various templates, suggest that this synergy is mediated both by cooperative DNA binding and by a specific interaction of O2-DNA complexes with another factor.

In plants, two adjacent binding sites are a distinctive feature of ocs-elements. The α-coixin O2 target sequence resembles an ocs-element, both in sequence and in some functional features of the resulting DNA-protein complex responsible for transcriptional enhancement.

Keywords: Opaque2; Cooperativity; bZIP; Protein-DNA interactions; seed storage protein.

Introduction

The major class of seed storage proteins in maize, *Coix lacryma-jobi* and sorghum is a group of prolamins called zeins, coixins and kafirins, respectively. Based on their solubility prolamins can be grouped into four classes, α-, β-, γ- and δ-, that can be further separated into polypeptides of distinct molecular mass by SDS-PAGE (Esen, 1986; DeRose et al., 1989; Leite et al., 1990; Ottoboni et al., 1990).

Several mutations are known to affect the level of zeins in the maize endosperm (Motto et al., 1989). One of the first characterized mutant, the *opaque2* (*o2*), affects a DNA-binding protein belonging to the basic/leucine zipper (bZIP) class of transcriptional activators (Hartings et al., 1989; Schmidt et al., 1990). Albeit no *o2* mutants have been described for *Coix* and sorghum, both species contain homologous *O2* genes (Ottoboni et al., 1993, Pirovano et al., 1994, Vettore et al., in preparation), indicating that this regulatory protein was conserved among these species.

The Opaque2 protein (O2) has been shown to activate the transcription of several maize and *Coix* endosperm genes through binding to significantly different *cis*-acting elements (Lohmer et al., 1991; Schmidt et al., 1992; Yunes et al., 1994b; Cord Neto et al., 1995; Maddaloni et al., 1996; Muth et al., 1996). Other functional O2-binding sites have also been found in the seed-specific genes of pea lectin (de Pater et al., 1994), wheat prolamin (Holdsworth et al., 1995), and French bean phytohemaglutinin (Chern et al., 1996).

It is very difficult to draw a consensus sequence through the alignment of all described O2 target sites. While we do not yet understand how can a protein recognize such diversity of sequences, it is possible to infer that specificity has not been maximized during evolution.

In a previous work we have shown that the O2 binding sequence in the 25-kD α-coixin promoter spans approximately 30 nucleotides. It was early apparent that over the entire sequence, the palindrome GACATGTC was the

most important requirement for O2 binding effectiveness (Yunes et al., 1994b). However, the large size of the footprinted sequence leads us to ask for the existence of an adjacent O2 binding site. In fact, based on sequence similarity we identified at the 5' side of the protected site the sequence TGACTA, which resembles the GCN4 binding site (Hill et al., 1986). The relevance of this GCN4-like motif was addressed by site-directed mutagenesis. A single mutation converting the TGACTA motif to TAACTA was sufficient to abolish the O2 binding under conditions where binding to the wild site is apparent (Yunes et al., 1994a).

In plants, numerous examples exist where gene regulation is mediated by DNA target sites bearing two neighboring DNA motifs related in sequence. The characterization of putative discrimination factors for such targets show, at least in some cases, that the same protein is able to bind both motifs (Tokuhisa et al., 1990; Fromm et al., 1991; Foley et al., 1993; Zhang et al., 1993; Feldbrügge et al., 1994; Feltkamp et al., 1994; Lam & Lam, 1995; Chern et al., 1996; Lu et al., 1996).

In this study we show that for O2 binding at the α-coixin promoter, rather than the palindrome previously identified, the most important features of the footprinted region are two adjacent motifs related in sequence to the GCN4 binding site. By using gel mobility-shift assays and footprint titration experiments we determined that the sequence-specific binding of O2 to α-coixin promoter occurs via cooperative binding of two O2 dimers to the two adjacent sites. Relative microscopic binding constants that describe the intrinsic affinity of the O2 protein ligand for each of the adjacent motifs and the cooperative interaction between liganded motifs were resolved by simultaneous numerical analysis of binding data for the wild-type and mutant sites. Furthermore, the same mutants were analyzed in transient expression assays, providing interesting insights about the overall architecture of the functional nucleoprotein complex.

Results

Two O2 homodimers simultaneously bind to the α-coixin O2 target region

The large size of the O2 footprinted sequence in the α-coixin α-3B promoter and the presence of putative bZIP recognition motifs in each half of this region (Yunes et al., 1994a, 1994b), led us to ask whether the O2 protein binds one or another half of this region, or simultaneously to both halves.

To address this question, an N-terminal deleted *Coix* O2-homologous protein comprising 63 residues of the N-terminal region, the bZIP domain, and the C-terminal extension, was expressed in *Escherichia coli* and partially purified by heat treatment according to Izawa et al. (1993).

Binding of this O2 protein to oligonucleotide probes containing the α-coixin O2 target sequence was studied, using standard electrophoretic mobility shift assays. Figure 1a shows the two oligonucleotides used. Oligo-A, which contains a deletion of three nucleotides in the upstream side of the footprinted region (Yunes et al., 1994b), and oligo-B, that contains all the footprinted region plus five flanking nucleotides and a restriction enzyme site on each end. Incubation of oligo-A with increasing amount of the O2 protein resulted in a single retarded band (Figure 1b, lanes 2 to 5). Incubation with oligo-B resulted, in addition to the single shifted complex observed with oligo-A, in an upper retarded band (Figure 1b, lanes 7 to 10). When using oligo-B, the fraction of total probe shifted and the ratio between the bands does not change appreciably above 125 ng protein. Contrarily, the free probe does continues to decrease with increasing amount of protein. This happened because an increasing amount of probe did not enter the gel (data not shown). Since this effect was not observed with oligo-A, it is likely that the probe arrested in the gel's well corresponds to a protein-DNA complex where the oligo is bound by two O2 dimers.

These complexes were stable in the presence of 2 ng/μl salmon sperm DNA, included in the binding reaction as a nonspecific competitor. In addition, formation of the two complexes was sequence-specific, since binding was clearly

competed by addition of identical unlabeled oligo-B (Figure 1c, lanes 3 to 8), but not by a 50-fold molar excess of the 63-mer Oct-1 oligonucleotide (Pharmacia Band Shift Kit) (Figure 1c, lane 10). No binding was detected when the same assays were performed using protein extract of a control *E. coli* strain carrying the pET vector without the O2 coding sequence (data not shown).

Since bZIP proteins dimerise prior to binding to DNA (Landschulz et al., 1989), the lower retarded band resulted from binding of one O2 homodimer to the oligonucleotide, and the upper band was due to the binding of a second O2 homodimer to the same oligonucleotide molecule. Since no upper band was evident with oligo-A, we concluded that the second O2 homodimer was unable to bind due to the upstream truncation.

These data are consistent with a conclusion that the O2 footprinted region in the α-coixin promoter covers indeed two adjacent binding sites, where two O2 homodimers bind simultaneously.

O2 binds cooperatively to two adjacent GCN4-like sites

We named the upstream site, *O2u*, and the downstream, *O2d*. As shown in Figure 2a, both sites present homologous GCN4 binding motifs. Previous works indicate that O2 is able to bind GCN4-like sites (Mauri et al., 1993, de Pater et al., 1994, Holdsworth et al., 1995).

By dividing the footprinted region at the middle, and considering the 5'-TGAC-3' GCN4-like motif as a half-site for an O2 dimer, we roughly depicted the boundaries of each of the sites (Figure 2b). The *O2d* site covers the palindromic sequence GACATGTC that was previously supposed to be the most important requirement for O2 binding effectiveness (Yunes et al., 1994b). The complementary strand of *O2d*, however, covers the sequence GATGACATGT, which is much more similar to a GCN4 binding site (Arndt & Fink, 1986; Hill et al., 1986), and is identical to the consensus O2 binding sequence described for the

b-32 gene (Lohmer et al., 1991). The *O2u* site shares lesser degree of similarity with any published O2 binding sequence.

A series of mutations was generated in the two sites (Figure 2c) and the ability of these mutants to bind the *Coix* O2 protein was analyzed (see bellow). In mutant *m1*, the TGAC motif of site *O2u* was modified to TAAC. Mutation *m2* disrupted the palindromic sequence GACATGTC, but preserved in the complementary strand the b-32-type O2 target. Contrasting, mutation *m4* preserved the palindrome but disrupted the b-32 complementary motif. The *m3* mutant is a combination of the *m1* and *m2* mutations.

An integral *Coix* O2 protein was expressed in *E. coli*, and partially purified by heat treatment. This protein was used in the footprint assays of Figures 3, 4 and 5. The footprinted region obtained with the *Coix* O2 protein (Figure 3) was identical to that previously reported for the maize β-gal::O2 fusion (Yunes et al., 1994b). The finding that in the wild-type promoter *O2u* and *O2d* filled up together, suggests either O2 protein interacts with similar affinities at the individual sites or O2 binds cooperatively to *O2u* and *O2d*.

All the mutations decreased the binding efficiency, as higher concentration of O2 protein was necessary to obtain complete DNase I protection (Figure 3). Mutation *m2* was far less detrimental than *m4*, indicating that rather than the GACATGTC palindrome, the most important motif of *O2d* site is the GCN4-like sequence. Similarly, the decreased binding efficiency observed with mutant *m1* indicates that the GCN4-like motif in *O2u* site is also required for accurate binding, as we previously reported (Yunes et al., 1994a).

In contrast to the simultaneous occupancy of sites *O2u* and *O2d* in the normal promoter, in the *m1* and *m4* mutants, the half-side region corresponding to the mutated site is lately protected (Figure 3). These results are in accordance with the gel retardation results, clearly indicating that the single footprint region observed for the normal promoter, is due to the combination of two adjacent binding events.

The high detrimental effect of mutation *m3* can not be simply explained by the sum of effects caused by *m1* and *m2*. Moreover, a complete protection of *m1* and *m4* required higher O2 concentrations than the wild-type promoter, not only for the mutated site but also for the nonmutated one. Since it is unlikely that the mutation of one or two nucleotides at one site might be affecting the intrinsic binding to the adjacent site, the observed results are a good indication for cooperative interaction, where occupancy of one site facilitates occupancy of another, and contrarily, the mutation of one site results in a decrease of the overall binding affinity.

The O2 binding at the adjacent sites show differential sequence specificity and a strict spacing constraint

From the preceding results it is clear that both sites work jointly in a cooperative manner, however, no indication of the intrinsic binding affinity for each site is given. In order to understand the coupling process of the O2 interaction with the α-coixin target sequence it is important to test whether both sites display similar affinity for O2 or whether there is a hierarchy of importance. To answer these questions we analyzed four mutant templates, in which the O2 binding sites were differentially spaced.

In mutant *m5* (Figure 2c), ten nucleotides encompassing around one helical turn, were inserted between *O2u* and *O2d*, thus leaving the sites in the same phase of the DNA helix. The boundaries of the intervening sequence mimic the pre-existing nucleotides, thus preserving the natural flanking nucleotides for each site. In mutant *m6* (Figure 2c) fourteen nucleotides were inserted, thus locating *O2u* and *O2d* sites on opposite faces of the DNA helix.

As shown in Figure 4a, reproducible DNase I footprints were recorded at *O2u* and *O2d* sites on the *m5* mutant, demonstrating once more that the large O2 target sequence in the α-coixin promoter is due to the presence of two adjacent binding sites. The protection pattern for *O2u* and *O2d* in *m5* was much

less complete and required higher O2 concentration than for the wild-type promoter. In addition, a greater degree of protection is apparent at *O2d* site when compared with *O2u* site. These results indicate that spacing between both binding sites is critical, and also that the O2 protein interacts with higher affinity with *O2d*, in relation to *O2u*. However, the last result must be interpreted with caution. In fact, we can not assure whether the 10 bp insertion did significantly alter the primary recognition sequence at the individual sites, mainly at site *O2u*.

In the *m6* mutant template, O2 still binds *O2d* site, however, no clear protection was observed in *O2u* site by visual inspection (Figure 4a). It is not possible to infer whether the loss of binding at site *O2u* is due to the increased distance between the two sites or derives from the opposite stereospecific alignment. Since, in both *m5* and *m6* mutants the flanking sequences of *O2u* and *O2d* sites were identically altered by the intervening sequence, the difference in the O2 binding results among these mutants, probably reflect the loss of cooperative interactions between O2 bound to adjacent sites rather than differences in intrinsic binding. Thus, it is possible that in *m5* mutant some degree of cooperativity is still possible.

To be sure that the binding results obtained with the *m6* mutant are not due to an anti-cooperative interaction, where the O2 bound at the stronger *O2d* site hinders the binding at the weaker *O2u* site, we decided to analyze the binding affinity of detached *O2u* and *O2d* sites.

From clone *m5*, we deleted either the *O2u* or *O2d* site, giving rise to *m7* or *m8*, respectively (Figure 2c). The detached *O2d* site was able to bind O2, whereas for the detached *O2u* site no clear protection was visually observed (Figure 4a). This is in keeping with the *m6* footprint results, and excludes the anti-cooperativity hypothesis.

By scanning densitometry (see bellow) and gel retardation assay (Figure 4b) it was possible to demonstrate that O2 is able to specifically recognize the detached *O2u* site, although with very weak affinity. This does not mean, however, that *O2u* contributes weakly for the binding affinity observed at the

whole double site. In fact, *m1* mutant clearly shows the importance of an intact *O2u* site, and suggests that O2 binds *O2u* site of the normal promoter with a higher affinity than was observed in the mutants *m5*, *m6*, and *m8*. However, this interaction probably requires much more sequence specificity than the interaction with the *O2d* site, and could explain why any variation at the recognition or flanking DNA sequence severely reduces O2 binding at *O2u* site.

Estimation of relative O2-α-coixin intrinsic binding energies and cooperativity energy by simultaneous numerical analysis of wild-type and reduced valency mutant footprint titration data

Cooperative interactions between regulatory proteins bound to multiple DNA sites are often crucial in the regulation of transcription. In these cases, a quantitative understanding of transcriptional regulation requires elucidation of the structural and thermodynamic features of the individual interactions and the mechanism of coupling of these interactions.

As initial steps to examine the O2/α-coixin assembly into its constituent individual macromolecular interactions and to understand the role that each of those local interactions plays in the regulation of the entire assembly we performed quantitative DNase I footprint titrations. In a DNase I footprint experiment conducted over a suitable titration of protein concentrations, profiles of fractional protection at the individual sites can be simultaneously measured, and these can be used to resolve not only the intrinsic binding constants but also those for cooperative interactions between the various sites (Brenowitz et al., 1986; Senear et al., 1986).

The individual-site binding equations for *O2u* and *O2d* sites, constructed according to a statistical-thermodynamic model (Ackers et al., 1982, 1983), are:

$$Y_u = (k_u [P] + k_u k_d k_{ud} [P]^2) / (1 + (k_u + k_d)[P] + k_u k_d k_{ud} [P]^2) \quad (1a)$$

and

$$Y_d = (k_d [P] + k_u k_d k_{ud} [P]^2) / (1 + (k_u + k_d)[P] + k_u k_d k_{ud} [P]^2) \quad (1b)$$

where Y_u and Y_d represent the fractional saturation of site *O2u* and *O2d*, respectively, [P] is the free protein concentration, and k_u and k_d are the microscopic or intrinsic equilibrium constants for binding of O2 to sites *O2u* and *O2d*, respectively. k_{ud} is the equilibrium constant for the cooperative interaction, which is defined as the extra free energy of binding two sites simultaneously, compared with the sum of the free energies of binding to each site separately. The Gibbs free energies are obtained from the relation $\Delta G_i = -RT \ln k_i$, where R is the gas constant, and T is the absolute temperature.

Non-linear least-squares parameter estimations were used (see Methods) to determine the best-fit model parameters that yield a minimum in the variance. Because of the high numerical correlation between the constants to be resolved, it is necessary to conduct the regression process simultaneously for different but related experiments, whose modeling functions share one or more parameters (Brenowitz et al., 1986; Senear et al., 1986). Thus, we included in our numerical analysis titration data from the mutated promoters *m1* to *m8*.

The results from the footprint experiments allowed us to write the individual-site binding equations for the mutant templates. We assumed that the alternated mutation of *O2u* or *O2d* in *m1*, *m2*, and *m4* templates disturbed solely the intrinsic binding to the mutated site, preserving the adjacent nonmutated site with the same affinity encountered in the wild-type template. Since such mutations reduced, but do not eliminated, the specific O2 binding to the given mutant site, a new parameter for intrinsic binding to each of these mutated sites was used, as described by Senear & Ackers (1990) and Scholl & Nixon (1996). A new parameter for cooperative interaction was also added. These new parameters correspond to the microscopic binding constants shown in Table I, and were substituted for the corresponding microscopic constants of equations 1a and 1b. For *m3* template the equation combines the intrinsic binding

constants of the mutated sites of *m1* and *m2*, plus a new cooperative constant (Table I).

It was also assumed that deletion of *O2u* site in template *m7* did not affect the intrinsic affinity to site *O2d*. This is consistent with our footprint results, and also with the fact that the O2 binding to b-32-like target sites supports high degree of variability at the surrounding, and even at the recognition sequence (Hartings et al., 1990; Lohmer et al., 1991). The individual site isotherm for the *m7* template is:

$$Y_d = (k_d [P]) / (1 + k_d[P]) \quad (2)$$

For *m5* and *m6* templates, we assumed that the 10 bp and 14 bp intervening sequences produced an equivalent reduction in the intrinsic binding affinity of *O2u* site, without disturbing the intrinsic binding at *O2d* site. Therefore, the same new parameter for intrinsic binding to the *O2u* site of *m5* and *m6* was used. For the cooperative interaction, a new parameter was used for each of the templates (Table I).

Since *m8* derives from *m5* (see Methods), the intrinsic binding affinity of the *m5*'s *O2u* site was used to write the equation below:

$$Y_{um5} = (k_{um5} [P]) / (1 + k_{um5}[P]) \quad (3)$$

Using a more extensive protein dilution series than shown in Figures 3 and 4, it was possible to determine the fractional protection of *O2u* and *O2d* sites for the normal and *m1* to *m8* DNA templates. A subset of the bands in each site was used for optical density measurement (Figure 3 and 4). We wrote an if/then logic program to associate each data set with the proper equation, thus allowing simultaneous least-squares parameters estimation. Because constraints exist for a given parameter in one experiment that eliminates some of the possible values

for that parameter in the second experiment, and vice versa, it was possible to determine the best-fit values of the parameters.

Determination of absolute equilibrium binding constants requires that the protein concentration is known. Thus, the protein must be chemically pure. Since we do not know the dimerization constant or the fraction of active molecules in our O2 protein preparation, all these calculated intrinsic binding and cooperative free energies are based on total O2 concentration, and can only be viewed as relative free energy terms (Brenowitz et al., 1989).

Figure 5 shows titration data for each of the templates included in the analysis, together with the resolved individual-site binding isotherms. For some templates, the range of protein concentration employed was not sufficient to obtain the lower limit of fractional protection for one or both binding sites (*m2* and *m6* in Figure 5). Likewise, we were unable to saturate the mutant binding sites in *m3*, *m5*, *m6*, and *m8* (Figure 5), because a systematic, ligand concentration-dependent variation in intensity of bands outside the binding site, precludes fractional saturation computing at higher protein concentrations. Nevertheless, the simultaneous analysis of these truncated data, and in conjunction with the complete data of the other templates, resulted in bounds being present on both extremes of the experimental phenomena, thus allowing resolution of their relative free energy values.

The estimated relative intrinsic binding and cooperative free energie terms, are summarized in Table II. Inspection of these relative free energy values clearly show that O2 interacts with similar affinity and in a highly cooperative manner to *O2u* and *O2d* sites of the normal promoter.

For the mutant sites of *m1*, *m2*, *m3*, and *m4*, the relative O2 binding affinity is reduced by ca. 1.0 to 1.3 kcal/mol in relation to the wild type site, which corresponds to 6- to 8-fold at the conditions studied. The relative intrinsic energy of *O2u* site in *m5*, *m6*, and *m8* is reduced by ca. 2.3 kcal/mol (49-fold). The m8 isotherm plays an important role in the determination of this intrinsic free energy

(ΔG_{um5}), and therefore, the 2.3 kcal/mol reduction is consistent with the fact that O2 recognizes the detached *O2u* site with very weak affinity (Figure 4b). As we already pointed out, this reduction is likely to be a consequence of the 10 bp insertional mutagenesis.

The relative cooperative free energy remained the same for *m2*, decreased ca. 1 kcal/mol (6-fold) for *m1* and *m4*, and decreased ca. 0.5 kcal/mol (2-fold) for *m5*. Evaluation of differences between the cooperative free energies is less certain due to larger errors in the ΔG_{ud} terms. The confidence intervals for these parameters reflect all systematic differences between separate experiments, and between the different templates, as well as the imprecision of individual experiments. Therefore, the small anticooperative effect observed for *m3*, and the cooperative free energy of -0.7 kcal/mol for *m6*, are probably meaningless within the confidence limits of the data. In fact, in separate analysis of all individual templates, where different values for the cooperative free energy were fixed and used to estimate the two intrinsic Gibbs free energy terms (Brenowitz et al., 1986), we observe a minimum variance for $\Delta G_{ud}=0$ when analyzing *m3* and *m6* (data not shown).

The resolved microscopic Gibbs energies observed are in good accordance with our initial footprint conclusions. However, since these results rely in the crucial assumptions we used to define which parameters are common to each isotherm and which forms of the binding equations are to be applied (see above), we do not know how well the data for each of the different templates are described by the simultaneous analysis. In this regard, the individual-site loading free energies, $\Delta G_{L,i}$ (Ackers et al., 1983), provide an important criterion in judging the goodness of fit and completeness of the model (Koblan et al., 1992; Senear & Bolen, 1992). As shown in Table III, there is a close agreement between $\Delta G_{L,i}$ values obtained from isotherms separately fitted to each template and those from isotherms resolved by the model-dependent (simultaneous) least-squares analysis, thus providing an important verification that the assumptions

established correctly describe the macromolecular interactions between O2 and both wild-type and mutant templates.

Transient expression assay

As initial steps to determine the biological significance of the free energy values above calculated, we conducted transient expression assays in tobacco mesophyll protoplasts. The reporter plasmids expressing the β-glucuronidase (*GUS*) gene under the control of either the normal 285-bp α-coixin promoter fragment, or the mutant version *m1* to *m7*, and the *Coix* O2-protein-expression vector used in the cotransfection experiments, are shown in Figure 6a.

The wild type vector treatment was included in all experiments as an internal standard. .For each independent protoplast preparation, the GUS activity was normalized to the average expression directed by the wild-type construct cotransfected with the O2 expression vector. For comparisons between treatments we employed the least significant difference (LSD) method, which is a very effective test for detecting true differences in means (Carmer & Swanson, 1973).

Figure 6b shows the GUS activity of the various templates alone and after cotransfection with the O2 effector vector. A clear difference in the responses of the different constructs was observed upon cotransfection with the pRTO2Coix effector plasmid. Similar results were obtained when using different amounts of the O2 effector plasmid (data not shown), indicating that there is no interference of a tobacco transacting factor.

The *m1* mutation causes a ca. 50% loss of activity, while the *m2* and *m4* mutations reduced the level of O2 transcriptional activation by 5% and 25%, respectively (Figure 6B). Considering that the total free energies of O2 interaction with *m1* and *m4* are very similar (Table II), these results suggest that in the context of our assay the *O2u* site is more important to enhance activity than the *O2d* site.

In mutant *m3*, which contains both *m1* and *m2* mutations, the reduction in transactivation is much higher than the simple sum of effects caused by *m1* and *m2* (Figure 6b). Therefore we can infer that a synergistic effect on transcription occurs in the normal situation. Taken these results together with the free energy values (Table II), suggest that the cooperative O2 binding feature is likely to contribute to the transcriptional synergism, by providing a more stable residence at the enhancers, and increasing the affinity of the protein for relatively weak binding sites.

Mutants *m5*, *m6*, and *m7* retained poor transactivation activity (Figure 6b). According to the statistics analysis these treatments show no difference at 5% of significance. Looking at the free energy values, the *m5* mutation, which still permits certain degree of cooperative interaction of O2 bound at *O2u* and *O2d* sites, showed higher binding affinity than *m6* and *m7* (Table II). These data indicate that the functional properties of a protein-DNA interaction do not always correlate with binding affinities. In case of *m5*, the interaction between the two O2 dimers bound at *O2u* and *O2d* sites probably precludes either formation of a domain competent to activate transcription, or effective interaction with the transcription complex, perhaps due to steric constraints.

Discussion

We demonstrated that two linked sites, which we named *O2u* and *O2d*, are responsible for the O2 binding and transactivation of the α-coixin gene. The *O2u* and *O2d* sites show a considerable degree of sequence divergence. While O2d site is almost identical to the B1 and B4 binding sites of the *b-32* gene (Lohmer et al., 1991), the *O2u* site is poorly homologous to any O2 binding site described so far. Nevertheless, both sites contain a TGAC motif, which is also part of the CREB/ATF and GCN4/AP1 binding sites.

Recently, it has been shown that O2 recognizes two linked binding sites in the promoter of the 22-kD α-zein (Muth et al., 1996). In fact, it was earlier shown

that the stronger TCCACGTAGA O2 target region in the 22-kD α-zein promoter, is flanked by the weakly footprinted sequence comprising the domain TGCATGTCA (Schmidt et al., 1992), which was recently characterized as an O2 binding site called Z2 (Muth et al., 1996). The sequence alignment of the α-coixin and α-zein O2 binding sites is shown in Figure 7.

At the time, the characterization of two closely spaced plant bZIP binding sites is mainly reported for the ocs-elements. These elements are a group of extensively studied *cis*-acting sequences that have been identified in the promoters of some *Agrobacterium* and caulimoviruses genes (Bouchez et al., 1989). A considerable degree of sequence divergence occurs between the ocs-element sequences described, and rare are the cases where an ocs-element is constituted by two identical adjacent motifs (Bouchez et al., 1989; Fox et al., 1992; Ulmasov et al., 1994). As shown in Figure 7, the double O2 binding sites of α-coixin and α-zein promoters show sequence similarities with some ocs-elements.

O2 binds cooperatively to the adjacent target sites

By a simultaneous numerical analysis of footprint titration data we were able to estimate relative free energy components involved in the O2-DNA interaction, and demonstrated that O2 binds to *O2u* and *O2d* sites with a similar intrinsic affinity and in a cooperative manner. The GCN4 protein seems also to bind cooperatively to very close repeated sequences in the promoters of *his3* and *his4* genes (Arndt & Fink, 1986). Recently, crude evidences suggest that TGA2 or TGA5, two Arabidopsis bZIP factors, are able to bind cooperatively to the two tandem TGACG motifs in the *as-1* site, an ocs-element found in the CaMV 35S promoter (Lam & Lam, 1995).

As far as we know, our results represent the first clear demonstration of cooperative binding of a bZIP transcriptional activator. However, some of the assumptions we made during our analysis might influence the cooperativity

constant measured. For example, it seems plausible that O2 binds more weakly to *O2d* in a construct that lack the adjacent *O2u* (mutant *m7*) than to the same site in the natural context. Since we assumed that the intrinsic binding constant is the same for both constructs, the intrinsic affinity of *O2d* site could be underestimated, which in turn leads to an overestimation of cooperativity.

On the other hand, several different factors might lead to an underestimation of cooperativity. For example, our estimation of cooperativity was based on the assumption that 100% of the O2 protein in the preparation is active and predominantly in the dimer form at the concentration of our assays. However, as discussed by Carlson & Little (1993), in case of either weak dimerization of O2 or if the fraction of active O2 was considerably less than 100%, an underestimation of cooperativity may have occurred.

In order to obviate this concern, we simultaneously analyzed data from several other templates, where a point mutation probably disturbed the intrinsic binding to the mutant site but did not affect the intrinsic binding to the adjacent, nonmutant site. Moreover, our results satisfied the three criteria used to judged the goodness of fit and completeness of the model (Koblan et al., 1992; Senear & Bolen, 1992), (i) the favorable comparison between the individual-site loading free energies (Table III), and (ii) variances (data not shown), calculated from the simultaneous analysis for each of the separate templates and those from the best phenomenological fits to each of the templates, and (iii) the apparently random distribution of residuals for each of the different templates (data not shown).

It is never possible to eliminate all inaccuracy in global parameter estimation. However, despite any uncertainty on the resolution of the binding constants, there are no doubt regarding the cooperative DNA binding of O2. In a separate analysis, where different values of the cooperative free energy were fixed (Brenowitz et al., 1986), we obtained better values of the square root of the variance when assuming values of $\Delta G_{ud}<0$ (data not shown).

The cooperativity might result from direct protein-protein interactions or be the product of an altered DNA conformation induced by O2 binding. Aside the uncertainties in the values of ΔG_{ud} terms, it is difficult to explain the fact that different mutations, while altering at the same extent the intrinsic binding affinity have different effects in the cooperative free energy terms (compare *m2* and *m4*, Table II). However, in some cases global protein folding transitions are coupled to DNA binding, then it is likely that the final protein conformation will be a function of the DNA sequence (Spolar & Record Jr., 1994), and therefore, it is possible that the same binding free energy drives folding of residues to create or not key parts of the protein responsible for the protein-protein interaction, depending on the DNA-sequence. A similar rationale can be envisioned considering cooperativity being the product of a DNA conformational change. In this case, differential sequence-dependent DNA distortions are induced by protein binding, giving rise to DNA structures that facilitates or not the O2 binding at an adjacent site. We had no evidence of hypersensitivity within or between the O2 binding sites that might indicate such DNA conformational change. However, DNase I footprint is not of good utility in the study of DNA structure itself.

Evolutionary significance and biological role for cooperativity of Opaque2

To date, O2 is the only plant bZIP factor to which a regulatory function could be assigned. Surprisingly, there is a large variability in the functional O2 binding sites so far identified (Lohmer et al., 1991; Schmidt et al., 1992; Yunes et al., 1994b; Cord Neto et al., 1995; Maddaloni et al., 1996; Muth et al., 1996). This implies that specificity has not been maximized during evolution, which is in accordance to the statistical-mechanical selection theory of Berg & von Hippel (1987). This theory assigns that the recognizer protein is not designed for maximum specificity because of an evolutionary requirement of flexibility.

In order to reduce the competitive binding to pseudosites in the genome to appropriate levels, a tendency towards larger DNA sites is required. An effective

increase in site size, without the undesirable enlargement of the discrimination factor, is then achieved by dimerization or cooperative binding of the same protein to two neighboring binding sites, so that the effective recognition sequence consists of the two sites taken together (Berg & von Hippel, 1987).

Weaker discrimination factors permits a fine tuning or modulation of the binding (or activity) at different specific sites. This could represent a means for the O2 protein to control, with different intensities, the expression of the different classes of endosperm genes (Kodrzycki et al., 1989; Yunes et al., 1994a; Cord Neto et al., 1995;. Maddaloni et al., 1996).

The two adjacent binding sites and transcriptional synergy

By transient expression assays in tobacco protoplats we observed that a mutation on either of the adjacent sites or on both sites simultaneously, affects the transactivation by O2 in such a manner that can not be simply explained by an additive mechanism. Therefore, it is plausible that *O2u* and *O2d* sites act synergistically to achieve a high level of expression of the α-coixin gene.

Recently, Muth et al. (1996) demonstrated that the maize O2 protein binds to three sites in the α-zein promoter, and that one of such sites, which has the lowest affinity of all three, confers no detectable O2-dependent promoter activation alone, but significantly increases activation in combination with either one of the other sites. Similarly, a sinergistic effect on transcription is observed when both binding motifs in the ocs-element are occupied (Bouchez et al., 1989; Singh et al., 1989; Ellis et all., 1993; Kim et al., 1994; Ulmasov et al., 1994; van der Zaal et al., 1996).

Cooperative DNA binding that leads to a synergistic effect on transcription is a property of some eukaryotic activators, including the steroid hormone receptor (Tsai et al., 1989), the immunoglobulin enhancer binding factor Oct-2 (Lebowitz et al., 1989), the ubiquitous octamer-binding protein Oct-1 (Bruggemeier et al., 1991), the Drosophila heat shock factor (Xiao et al., 1991),

the HeLa cell protein TEF-1 (Davidson et al., 1988), and the yeast GAL4 (Giniger & Ptashne, 1988). However, other eukaryotic activators do not appear to bind cooperatively to DNA, although activate transcription synergistically on templates bearing two or more recognition sites. These suggest that there must be an additional mechanism, that together with cooperative binding, contributes for the synergistic action (Okuda et al., 1990; Oliviero & Struhl, 1991; Emami & Carey, 1992).

Cooperative binding of two or more activators to DNA, can in principle, generate a synergistic effect on transcription because the activator's affinity for DNA increases, allowing a template bearing multiple sites to become saturated at lower concentration of activator than a template bearing a single site. Since for the different analyzed mutants, the total binding free energy is correlated with the expression activity, it is likely that cooperative binding of O2 to the linked sites contributes to the observed sinergistic effect on transcription.

But the cooperative O2 binding must not be the only mechanism contributing for the synergistic action. Some templates, while having significantly different transcriptional activity, present an identical total energy of saturation upon binding of O2 (compare *m1* and *m4*). This demonstrates that the global binding affinity may not be always a secure indicative of transcriptional activity. By analyzing how is this total binding energy partitioned between events occurring at the two sites (Table III) it is possible to infer that, in the context of our assay, binding at *O2u* site is more important to enhance activity. Interestingly, the upstream site of the ocs-element was also reported to be more important for activity (Singh et al., 1989).

Therefore, each of the *O2u* and *O2d* sites might have different roles in determining the arrangement of the nucleoprotein complex and the interaction with the transcriptional machinery. In this regard, we can suggest that the O2 molecules which cooperatively bind to the 10 bp distant sites of *m5*, might form such a deviate protein (or DNA) structure, that precludes or hinders the domain competent to activate transcription, as was reported for the thyroid hormone

receptor (Glass et al., 1988) and the λ repressor (Hochschild & Ptashne, 1988). A similar circumstance, of normal in vitro binding with no enhancer effect in vivo was detected for the ocs-elements (Singh et al., 1989; Ellis et al., 1993).

Materials and Methods

Construction of plasmids with mutant O2 target sequences

The pBluescript KS (Stratagene) derived clone P285α-C (Yunes et al., 1994b), which contains the EcoRI/BamHI 285-bp α-coixin promoter fragment spanning the O2 target sequence, was the template for the site-specific mutagenesis performed by asymmetric PCR (Perrin & Gilliland, 1990) using the sequence primers and one of the following mutant primers: O2c-m1 5′-TGACAAAATTAACTAGGAGAC-3′,
O2c-m2 5′-TAGGAGACATATCATCTCTAG-3′,
O2c-m3 5′-TGACAAAATTAACTAGGAGACATATCATCTCTAG-3′,
O2c-m4 5′-GAGACATGTCCGCTCTAGCTTA-3′, and
O2c-m5 5′-AAATTGACTAGGAGACTCGAGGAGACATGTCATCTC-3′ (National Biosciences, Inc., Plymouth, MN). Note that primer O2c-m5 creates a XhoI site (underlined). The amplified fragments were subsequently digested with EcoRI and BamHI and inserted into KS vector, resulting in the clones P285m1, P285m2, P285m3, P285m4, and P285m5, respectively. The P285m6 mutant was obtained by subcloning the 285-bp EcoRI/BamHI fragment of P285m6GUS (see bellow) into KS. We cloned a detached *O2d* site, P285m7, by digesting plasmid P285m7GUS (see bellow) with HindIII and NcoI, recovering the 175-bp fragment, filling the recessed ends with Klenow, and cloning into the SmaI site of KS. A clone with an isolated *O2u* site, P285m8, was obtained by subcloning the 200-bp EcoRI-NcoI fragment of P285m8GUS (see bellow) in place of the 285-bp EcoRI-NcoI fragment of P285α-C. All the clones were verified by sequencing.

Overproduction of *Coix* O2-homologous protein in *Escherichia coli*

An incomplete *Coix* O2 coding sequence was cloned in pET3d (Studier et al., 1990). The incomplete O2 cDNA was excised from clone pCO2-1 (Vettore et al., in preparation) by digestion with BamHI and XhoI, the XhoI recessed end was partially filled with dTTP and dCTP using Klenow, and the resulting fragment was cloned into the BamHI site of pET3d, giving rise to pETO2iCx.

We also cloned the entire *Coix* O2 coding sequence into the pET3a vector (Studier et al., 1990). Using the reverse sequencing primer and the oligonucleotide O2ATG 5′-CAATCGATCATATGGAGCACGTCATCTCAATGG-3′ (National Biosciences, Inc., Plymouth, MN), which creates a NdeI restriction site (underlined) over the ATG, we amplified the O2 coding sequence from ATG to +382, using clone KSO2-212/374 (see bellow) as template. The amplified fragment was digested with ApaI and cloned into KS linearized with EcoRV and ApaI, giving rise to clone KCRO2C. Subsequently, in a tripartite ligation, the 382-bp NdeI-AccI fragment from KCRO2C, plus the 1024-bp AccI-BamHI fragment from pRTO2Coix (see bellow), were inserted into the NdeI and BamHI sites of the pET3a vector, resulting in the pETO2Cx.

BL21(DE3)/pLysS transformants were incubated in 50 ml LB medium at 36°C until an A600 of 0.4. Logarithmic phase cultures were induced with 0.4 mM-IPTG for 3 h. Bacteria were collected by centrifugation, washed with BBF buffer (12.5 mM HEPES/KOH (pH 7.5), 50 mM KCl, 0.5 mM DTT, 10% glycerol), resuspended in 2 ml BBF buffer, and lyzed by freeze-thawing until the suspension became translucent. The extract was boiled for 3 min and then incubated on ice for 30 min (Izawa et al., 1993). Insoluble proteins were removed by centrifugation. Protein concentration was determined by the Bradford assay (Bradford, 1976). The O2 protein used was partially purified to approximately 40%, as judged by electrophoresis on SDS gels. Molar concentrations were based on the predicted molecular weight deduced from the sequence of the *Coix* O2-homologous gene (Vettore et al., in preparation).

Electrophoretic mobility shift assays

DNA probes oligo-A and oligo-B were prepared from the synthetic pairs of oligonucleotides (National Biosciences, Inc., Plymouth, MN, and GIBCO BRL):

oligo-A

5′-ATTGACTAGGAGACATGTCATCT-3′
3′-CTCTGTACAGTAGAGATC-5′ and,

oligo-B

5′-GCTCTAGATGACAAAATTGACATGGAGACATGTCATCTCTAGCTTA-3′
3′-ACTGTTTTAACTGATCCTCTGTACAGTAGAGATCGAATCTTAAGG-5′

The above oligonucleotides were annealed (50μg/ml) in buffered 150 mM NaCl by heating at 65°C for 5 min and gradual cooling to room temperature. For radiolabeling, 250 ng of annealed oligonucleotides were extended with Klenow in 50 mM Tris-HCl, 10 mM $MgCl_2$ (pH 7,6), [α-^{32}P]dATP, and the other unlabeled dNTPs.

For the assays with the insert fragment of P285m7 and P285m8, preparation of DNA was the same as for Footprint (see bellow).

Protein-DNA complexes were formed by mixing different amounts of proteins with 2 μg/ml of sonicated salmon sperm DNA in 50 μl of binding buffer (10 mM HEPES/KOH (pH 7.9), 50 mM KCl, 1 mM EDTA, 10 mM DTT, 2 mg/ml BSA, 10% Glicerol). After a pre-incubation of 5 min at 25°C, the labeled probe was added (2-10 fmol, 5 000-30 000 cpm), and the mixture was incubated for 20 min at 25°C. The mixture was fractionated onto a 5% acrylamide (30:1 polyacrylamide/bisacrylamide) gels in 0.25(x) TBE (50 mM Tris, 42 mM boric acid, 0.5 mM EDTA), at 12°C.

DNase I Footprint titrations

Preparation of DNA. The P285α-C, or P285m1 to P285m8 plasmid DNA was digested with EcoRI and NotI, and the binding-site-containing fragments were isolated using agarose gel electrophoresis. After visual quantitation of DNA concentration, we labeled 300 ng of each fragment by Klenow fill-in reaction of

the 3´-recessed EcoRI end, using [α-^{32}P]dATP plus the other unlabeled dNTPs. The labeled fragments were then purified by PAGE, followed by "crush and soak", standard procedures.

DNase I footprinting. Reactions were carried out at 25ºC in a mixture containing 12.5 mM HEPES/KOH (pH 7.5), 50 mM KCl, 0.5 mM DTT, 10% glycerol, 0.05% Nonidet P-40, 1μg poly(dI-dC), and 20 000 - 60 000 cpm of end-labeled DNA, in a total volume of 50 or 200 μl. The molar concentration of DNA was measured based on the starting amounts of DNA and recovery of counts. The estimated concentration of the specific DNA fragment in each experiment was then used to calculate the free protein ligand concentration during data analysis (see bellow). The reactions were gently mixed every 2-3 min for 35 min. $MgCl_2$ and $CaCl_2$ were then added to a final concentration of 1mM and 0.5 mM, respectively. After 1 min of incubation, 0.1 units of DNase I (HPLC pure, Pharmacia) was added and, after mixing, the reaction was incubated another minute. The reaction was stopped by addition of DNase I stop solution to final concentrations of 137 mM sodium acetate (pH 5.2), 22.8 mM EDTA (pH 8.0), 0.1% SDS, and 10 μg tRNA, followed by vigorous vortexing. The samples were extracted once with an equal volume of a 25:24:1 phenol / CHCl3 / isoamyl alcohol mixture, and the DNA was ethanol precipitated from the aqueous phase, ressuspended in formamide loading buffer, and separated by electrophoresis on an 6% sequencing gel. The G+A sequencing ladder was prepared according to the Sure Track Footprinting Kit (Sigma) manufacturer´s recommendations. Autoradiography of the dried gels was conducted at -70ºC, using preflashed X-ray film and an intensifying screen.

Data analysis. Two dimensional optical scans of the footprint autoradiographs were obtained using a Biorad GS-700 imaging densitometer. We set a spatial resolution of 56 μm (450 dpi) and a pixel depth of 12 (12 bit precision). The optical density was processed by using the Molecular Analyst ™/PC program. The fractional protection (*P*) of the sites was determined by the

method of Brenowitz et al. (1986). Energetic parameters for the interaction of the O2 protein with the α-coixin binding sites, and the end points for each transition curve, were estimated by simultaneous nonlinear least-squares analysis, using the multivariate secant (DUD) iterative method of the NLIN SAS program. The convergence criterion used, specifies the program to continue the iterative process until the decrease in SSE (square sum error) is 10^{-8} and the maximum change among the parameters is 10^{-6}. The G4SINGULAR option, which specifies that a g4 or Moore-Penrose inverse is used in parameter estimation if the Jacobian is (or become) of less than full rank, was used. The NLIN SAS program uses a linearized form of a nonlinear model for the analysis of confidence contours, consequently it predicts a symmetrical confidence interval which is an underestimate of a true confidence interval. The reliability of the parameter estimation process was evaluated by the cross-correlation coefficient, considering a critical limit of ~0.96. In order to evaluate the uniqueness of the estimated parameters, multiple least-squares parameter estimations were performed with a variety of different starting values for the parameters, such that all of these converged to the same set of parameter estimates.

An if/then logic was used to assign the data point to the correct equation described in the Results. To treat quantitative footprinting data as transition curves (Brenowitz et al., 1986), the final equation used in fitting the fractional protection data (*P*) and the appropriate fractional saturation function (*Y*) was:

$$P = L + (U - L) * Y$$

where L and U are fitted lower and upper end points to the observed transition curves.

At each iteration of the nonlinear least-square procedure used to fit the *Y* (fractional saturation) value observed, the fitted *Y* was used to correct each total protein concentration, [Pt], to the free protein concentration, [Pf], according to the relationship:

[Pf] = [Pt] - [P-DNA], and

[P-DNA] = *Y* * [DNA]

where [P-DNA] represents DNA-bound protein.

Construction of plasmids for the transient expression assay

All sequence position numbers refer to the ATG initiation codon.

The reporter vectors P285m1GUS, P285m2GUS, P285m3GUS, P285m4GUS, and P285m5GUS were constructed by excising the 285-bp (295-bp for m5) EcoRI-NcoI fragment from the counterpart clones P285m1, P285m2, P285m3, P285m4, and P285m5 and inserted into the P285α-CGUS vector (Yunes et al., 1994b), which was digested with the same enzymes, replacing the normal 285-bp α-coixin promoter fragment. Vector P285m6GUS was obtained by digesting P285m5GUS with XhoI, filling the recessed ends with klenow, and recircularizating with T4 DNA ligase. Vector P285m7GUS was obtained by deleting a 100-bp fragment spanning the *O2u* site, by digesting P285m5GUS with EcoRI and XhoI, filling ends with klenow, and recircularizating with T4 DNA ligase. Vector P285m8GUS was obtained by deleting a 100-bp fragment spanning the *O2d* site, by digesting P285m5GUS with XhoI and ApaI, making blunt ends with T4 DNA polymerase, and recircularizating with T4 DNA ligase. Since the putative TATA box was also deleted, P285m8GUS was not assayed in the protoplasts transfections.

pRTO2Coix is an effector vector containing the *Coix* O2-homologous cDNA sequence spanning from -212 to 1406, placed under control of the CaMV35S promoter of pRT101 (Töpfer et al., 1987). From the incomplete *Coix* O2 cDNA clone pCO2-1 (Vettore et al., in preparation), the sequence spanning from +382 to +1406 was amplified by PCR and inserted in the AccI and KpnI sites of KS, giving rise to the clone KSO2+382/1406. The primers for PCR were the reverse sequence primer and the oligonucleotide O2C382 5′-TTCGGCACGTCGACCTCGTGGCCTTC-3′ (National Biosciences, Inc.,

Plymouth, MN), which creates an AccI restriction site (underlined) at position +382, without changing the amino acid sequence.

The upstream +1 to +382 coding region, plus the untranslated leader sequence, were obtained from the *Coix* O2 genomic clone (Vettore et al., in preparation). By comparison with the maize O2 cDNA clone 6-1 (Schmidt et al., 1990), the putative site for transcription initiation in the *Coix* O2 genomic clone was identified at position -212. The *Coix* O2 sequence from -212 to +382 was amplified and inserted in the EcoRI and AccI sites of KS, giving rise to the clone KSO2-212/374. The oligonucleotides for the PCR were O2C-212 5′-GCTCGAATTCTGCACGCACAAGGCCTCA-3′ and O2C+374 5′-GCCACGAGGTCGACCTCCAATT-3′ (National Biosciences, Inc., Plymouth, MN), which creates an EcoRI and AccI restriction sites (underlined), respectively. All the PCR-derived clones were verified by sequencing

Subsequently, in a tripartite ligation the whole *Coix* O2 cDNA sequence was assembled. The 594-bp EcoRI-AccI fragment from KSO2-212/374, plus the 1024-bp AccI-KpnI fragment from KSO2+382/1406, were inserted in the EcoRI and KpnI sites of the pRT101 vector, resulting in the pRTO2Coix.

Transient expression assays in tobacco mesophyll protoplasts

Transient expression assays in tobacco mesophyll protoplasts were performed essentially as described previously (Yunes et al., 1994b). Because of the limiting amount of protoplasts it was not possible to include all treatments (the different reporter vectors) in a single protoplast preparation. Five independent protoplasts preparations were used. The overall treatments were analyzed in a partially balanced incomplete block design, where each block corresponds to a protoplast preparation. The P285α-CGUS:pRTO2Coix treatment was included in all protoplasts preparations, as a control. For each independent protoplast preparation, the GUS activity was normalized to the average expression directed by this treatment.

Ten micrograms of the reporter vector alone or with 20 μg of the pRTO2Coix effector vector, plus sonicated salmon sperm DNA to a final amount of 43 μg were used in the transformation of each aliquot of protoplast. The total number of replicas per treatment was 11, 5, 5, 5, 3, 3, 3, 4 and 18, 11, 8, 8, 9, 6, 11, 4 for the normal, *m1*, *m2*, *m3*, *m4*, *m5*, *m6*, and *m7* templates alone and cotransfected with O2, respectively. Ten micrograms of protein was used in the fluorimetric GUS assay (Jefferson, 1987).

Acknowledgments

We thank Pedro Mauricio for careful technical help. We are very grateful to MsC. Nilda D. Montes Villanueva for the excelent statistics assistance. We wish to thank also MsC. Lurdes Contreras Castillos for helpful instructions in SAS programing; Drs Marcos Alcocer and Franco M. Lajolo, Universidade de São Paulo, São Paulo, Brazil, for providing the densitometer. We also thank Dr Tracy Nixon, Pennsylvania State University, PA, USA, for helpful discussion and for providing preprinted manuscript; Dr. Reinhard Töpfer, Max Planck Institut, Cologne, Germany, for supplying pRT101 and pRT103 vectors. This work was financed by grants to P.A. from FINEP, CNPq and FAPESP / PTE-90/3808-5. P.A. and A.L. received research fellowships from CNPq. J.A.Y. received postgraduate fellowship from CNPq. A.L.V. received postgraduate fellowship from FAPESP.

References

Ackers, G.K., Johnson, A.D. & Shea, M.A. (1982). Quantitative model for gene regulation by lambda phage repressor. *Proc. Natl. Acad. Sci., USA,* **79,** 1129-1133.

Ackers, G.K., Shea, M.A. & Smith, F.R. (1983). Free energy coupling within macromolecules. The chemical work of ligand binding at the individual sites in co-operative systems. *J. Mol. Biol.* **170,** 223-242.

Arndt, K. & Fink, G.R. (1986). GCN4 protein, a positive transcription factor in yeast, binds general control promoters at all 5' TGACTC 3' sequences. *Proc. Natl. Acad. Sci., USA,* **83,** 8516-8520.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72,** 248-254.

Berg, O.G. & von Hippel, P.H. (1987). Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. *J. Mol. Biol.* **193,** 723-750.

Bouchez, D., Tokuhisa, J.G., Llewellyn, D.J., Dennis, E.S. & Ellis, J.G. (1989). The ocs-element is a component of the promoters of several T-DNA and plant viral genes. *EMBO J.* **8,** 4197-4204.

Brenowitz, M., Senear, D.F and Kingston, R.E. (1989). Dnase I footprint analysis of protein-DNA binding. In Current Protocols in Molecular Biology, Vol 2, F.M. Ausubel,, ed (John Wiley & Sons, Inc.), pp. 12.4.12.

Brenowitz, M., Senear, D.F., Shea, M.A. & Ackers, G.K. (1986). "Footprint" titrations yield valid thermodynamic isotherms. *Proc. Natl. Acad. Sci., USA,* **83,** 8462-8466.

Bruggemeier, U., Kalff, M., Franke, S., Scheidereit, C. & Beato, M. (1991). Ubiquitous transcription factor OTF-1 mediates induction of theMMTV promoter through synergistic interaction with hormone receptors. *Cell,* **64,** 565-572.

Carlson, N.G. & Little, J.W. (1993). Highly cooperative DNA binding by the coliphage HK022 repressor. *J. Mol. Biol.* **230,** 1108-1130.

Carmer, S.G. & Swanson, M.R. (1973). Evaluation of ten pairwise multiple comparison procedures by Monte Carlo methods. *Journal of the American Statistical Association,* **68,** 66-74.

Chern, M.-S., Bobb, A.J. & Bustos, M.M. (1996). The regulator of *MAT2* (RO*M2*) protein binds to early maturation promoters and represses pvALF-activated transcription. *Plant Cell,* **8,** 305-321.

Cord Neto, G., Yunes, J.A., Silva, M.J., Vettore, A.L., Arruda, P. & Leite, A. (1995). The involvement of Opaque 2 on β-prolamin gene regulation in maize and *Coix* suggests a more general role for this transcriptional activator. *Plant Mol. Biol.* **27,** 1015-1029.

Davidson, I., Xiao, J.H., Rosales, R., Staub, A. & Chambon, P. (1988). The HeLa cell protein TEF-1 binds specifically and cooperatively to two SV40 enhancer motifs of unrelated sequence. *Cell,* **54,** 931-942.

de Pater, S., Katagiri, F., Kijne, J. & Chua, N.-H. (1994). bZIP proteins bind to a palindromic sequence without an ACGT core located in a seed-specific element of the pea lectin promoter. *Plant J.* **6,** 133-140.

DeRose, R.T., Ma, D.P., Kwon, I.S., Hasnain, S.E., Klassy, R.C. & Hall, T.C. (1989). Characterization of the kafirin gene family from sorghum reveals extensive homology with zein from maize. *Plant Mol. Biol.* **12,** 245-256.

Ellis, J.G., Tokuhisa, J.G., Llewellyn, D.J., Bouchez, D., Singh, K., Dennis, E.S. & Peacock, W.J. (1993). Does the ocs-element occur as a functional component of the promoters of plant gene? *Plant J.* **4,** 433-443.

Emami, K.H. & Carey, M. (1992). A synergistic increase in potency of a multimerized VP16 transcriptional activation domain. *EMBO J.* **11,** 5005-5012.

Esen, A. (1986). Separation of alcohol-soluble proteins (zeins) from maize into three fractions by differential solubility. *Plant Physiol.* **80,** 623-627.

Feldbrügge, M, Sprenger, M., Dinkelbach, M., Yazaki, K., Harter, K. & Weisshaar, B. (1994). Functional analysis of a light-responsive plant bZIP transcriptional regulator. *Plant Cell,* **6,** 1607-1621.

Feltkamp, D., Masterson, R., Starke, J. & Rosahl, S. (1994). Analysis of the involvement of *ocs*-like bZip-binding elements in the differential strength of the bidirectional *mas*1'2' promoter. *Plant Physiol.* **105,** 259-268.

Foley, R.C., Grossman, C., Ellis, J.G., Llewellyn, D.J., Dennis, E.S., Peacock, W.J. & Singh, K.B. (1993). Isolation of a maize bZIP protein subfamily: candidates for the ocs-element transcription factor. *Plant J.* **3,** 669-679.

Fox, P.C., Vasil, V., Vasil,, I.K. & Gurley, W.B. (1992). Multiple *ocs*-like elements required for efficient transcription of the mannopine synthase gene of T-DNA in maize protoplasts. *Plant Mol. Biol.* **20,** 219-233.

Fromm, H., Katagiri, F. & Chua, N.-H. (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.

Giniger, E. & Ptashne, M. (1988). Cooperative DNA binding of the yeast transcriptional activator GAL4. *Proc. Natl. Acad. Sci., USA,* **85,** 382-386.

Glass, C.K., Holloway, J.M., Devary,, O.V. & Rosenfeld, M.G. (1988). The thyroid hormone receptor binds with opposite transcriptional effects to a common sequence motif in thyroid hormone and estrogen response elements. *Cell,* **54,** 313-323.

Hartings, H., Lazzaroni, N., Marsan, P.A., Aragay, A., Thompson, R., Salamini, F., Di Fonzo, N., Palau, J. & Motto, M. (1990). The b-32 protein from maize endosperm: characterization of genomic sequences encoding two alternative central domains. *Plant Mol. Biol.* **14,** 1031-1040.

Hartings, H., Maddaloni, M., Lazzaroni, N., Di Fonzo, N., Motto, M., Salamini, F. & Thompson, R. (1989). The *O2* gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. *EMBO J.* **8,** 2795-2801.

Hill, D.E., Hope, I.A., Macke, J.P. & Struhl, K. (1986). Saturation mutagenesis of the yeast *his3* regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. *Science,* **234,** 451-457.

Hochschild, A. & Ptashne, M. (1988). Interaction at a distance between λ repressors disrupt gene activation. *Nature,* **336,** 353-357.

Holdsworth, M., Muñoz-Blanco, J., Hammond-Kosack, M., Colot, V., Schuch, W. & 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.

Izawa, T., Foster, R. & Chua, N.-H. (1993). Plant bZIP protein DNA binding specificity. *J. Mol. Biol.* **230,** 1131-1144.

Jefferson, R.A. (1987). Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Mol. Biol. Rep.* **5,** 387-405.

Kim, Y., Buckley, K., Costa, M.A. & An, G. (1994). A 20 nucleotide upstream element is essential for the nopaline synthase (*nos*) promoter activity. *Plant Mol. Biol.* **24,** 105-117.

Koblan, K.S., Bain, D.L., Beckett, D., Shea, M.A. & Ackers, G.K. (1992). Analysis of site-specific interaction parameters in protein-DNA complexes. *Methods Enzymol.* **210,** 405-425.

Kodrzycki, R., Boston, R.S. & Larkins, B.A. (1989). The *opaque-2* mutation of maize differentially reduces zein gene transcription. *Plant Cell,* **1,** 105-114.

Lam, E. & Lam, Y.K.-P. (1995). Binding site requirements and differential representation of TGA factors in nuclear ASF-1 activity. *Nucleic Acids Res.* **23,** 3778-3785.

Landschulz, W.H., Johnson, P.F. & McKnight, S.L. (1989). The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. *Science,* **243,** 1681-1688.

Lebowitz, J.H., Clerc, R.G., Brenowitz, M. & Sharp, P.A. (1989). The Oct-2 protein binds cooperatively to adjacent octamer sites. *Genes Develop.* **3,** 1625-1638.

Leite, A., Ottoboni, L.M.M., Targon, M.L.P.N, Silva, M.J., Turcinelli, S.R. & 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.

Lohmer, S., Maddaloni, M., Motto, M., Di Fonzo, N., Hartings, H., Salamini, F. & 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.

Lu, G., Paul, A.-L., McCarty, D.R. & Ferl, R.J. (1996). Transcription factor veracity: Is GBF3 responsible for ABA-regulated expression of Arabidopsis *Adh*? *Plant Cell,* **8,** 847-857.

Maddaloni, M., Donini, G., Balconi, C., Rizzi, E., Gallusci, P., Forlani, F., Lohmer, S., Thompson, R., Salamini, F. & Motto, M. (1996). The transcriptional activator Opaque-2 controls the expression of a cytosolic form of pyruvate orthophosphate dikinase-1 in maize endosperms. *Mol. Gen. Genet.* **250,** 647-654.

Mauri, I., Maddaloni, M., Lohmer, S., Motto, M., Salamini, F., Thompson, R. & Martegani, E. (1993). Functional expression of the transcriptional activator *Opaque-2* of *Zea mays* in transformed yeast. *Mol. Gen. Genet.* **241,** 319-326.

Motto, M., Di Fonzo, N., Hartings, H., Maddaloni, M., Salamini, F., Soave C. & Thompson, R.D. (1989). Regulatory genes affecting maize storage protein synthesis. In Oxford Surveys of Plant Molecular and Cell Biology, Vol 6, B.J. Miflin, ed (Oxford: Oxford University Press), pp. 87-114.

Muth, J.R., Müller, M., Lohmer, S., Salamini, F. and Thompson, R.D. (1996). The role of multiple binding sites in the activation of zein gene expression by opaque-2. *Mol. Gen. Genet.* **252,** 723-732.

Okuda, A., Imagawa, M., Sakai, M. & Muramatsu, M. (1990). Functional cooperativity between two TPA responsive elements in undifferentiated F9 embryonic stem cells. *EMBO J.* **9,** 1131-1135.

Oliviero, S. & Struhl, K. (1991). Synergistic transcriptional enhancement does not depend on the number of acidic activation domains bound to the promoter. *Proc. Natl. Acad. Sci., USA,* **88,** 224-228.

Ottoboni, L.M.M., Leite, A., Targon, M.L.P.N., Silva, M.J. & Arruda, P. (1990). Heterogeneity of *Coix*, maize and teosinte prolamins detected by isoelectric focusing. *Rev. Bras. Genet.* **13,** 313-22.

Ottoboni, L.M.M., Leite, A., Yunes, J.A., Targon, M.L.P.N., Souza Filho, G.A. & 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.

Perrin, S. & Gilliland, G. (1990). Site-specific mutagenesis using asymmetric polymerase chain reaction and a single mutant primer. *Nucl. Acids Res.* **18,** 7433-7438.

Pirovano, L., Lanzini, S., Hartings, H., Lazzaroni, N., Rossi, V., Joshi, R., Thompson, R.D., Salamini, F. & Motto, M. (1994). Structural and functional analysis of an Opaque-2-related gene from sorghum. *Plant Mol. Biol.* **24,** 515-523.

Schmidt, R.J., Burr, F.A., Aukerman, M.J. & 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. & 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.

Scholl, D. & Nixon, B.T. (1996). Cooperative binding of DctD to the *dctA* UAS of *Rhizobium meliloti* is enhanced in a constitutively active truncated mutant. *J. Biol. Chem.,* **271,** 26435-26442.

Senear, D.F. & Ackers, G.K. (1990). Proton-linked contributions to site-specific interactions of λ cI repressor and OR. *Biochemistry,* **29,** 6568-6577.

Senear, D.F. & Bolen, D.W. (1992). Simultaneous analysis for testing of models and parameter estimation. *Methods Enzymol.* **210,** 463-481.

Senear, D.F., Brenowitz, M., Shea, M.A. & Ackers, G.K. (1986). Energetics of cooperative protein-DNA interactions: comparison between quantitative deoxyribonuclease footprint and filter binding. *Biochemistry,* **25,** 7344-7354.

Singh, K., Tokuhisa, J.G., Dennis, E.S. & Peacock, W.J. (1989). Saturation mutagenesis of the octopine synthase enhancer: correlation of mutant phenotypes with binding of a nuclear protein factor. *Proc. Natl. Acad. Sci., USA,* **86,** 3733-3737.

Spolar, R.S. & Record Jr., T. (1994). Coupling of local folding to site-specific binding of protein to DNA. *Science,* **263,** 777-784.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185,** 60-89.

Tokuhisa, J.G., Singh, K., Dennis, E.S. & Peacock, W.J. (1990). A DNA-binding protein factor recognizes two binding domains within the octopine synthase enhancer element. *Plant Cell,* **2,** 215-224.

Töpfer, R., Matzeit, V., Gronenborn, B., Schell, J., & Steinbiss, H.H. (1987). A set of plant expression vectors for transcriptional fusions. *Nucl. Acids Res.* **15,** 5890.

Tsai, S.Y., Tsai, M. & O'Malley, B.W. (1989). Cooperative binding of steroid hormone receptors contributes to transcriptional synergism at target enhancer elements. *Cell,* **57,** 443-448.

Ulmasov, T., Hagen, G. & 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.

Xiao, H., Perisic, O. & Lis, J.T. (1991). Cooperative binding of Drosophila heat shock factor to arrays of a conserved 5 bp unit. *Cell,* **64,** 585-593.

Yunes, J.A., Cord Neto, G., Leite, A., Ottoboni, L.M.M. & Arruda, P. (1994a). The role of the Opaque2 transcriptional factor in the regulation of protein accumulation and amino acid metabolism in maize seeds. *An. Acad. Bras. Ci.* **66,** 227-238.

Yunes, J.A., Cord Neto, G., Silva, M.J., Leite, A., Ottoboni, L.M.M. & Arruda, P. (1994b). The transcriptional activator Opaque2 recognizes two different target sequences in the 22-kD-like α-prolamin genes. *Plant Cell,* **6,** 237-249.

van der Zaal, B.J., Droog, F.N.J., Pieterse, F.J. & Hooykaas, P.J.J. (1996). Auxin-sensitive elements from promoters of tobacco *GST* genes and a consensus *as-1*-like element differ only in relative strength. *Plant Physiol.* **110,** 79-88.

Zhang, B., Foley, R.C. & Singh, K.B. (1993). Isolation and characterization of two related *Arabidopsis* ocs-element bZIP binding proteins. *Plant J.* **4,** 711-716.

Figure legends

Figure 1. Binding of *Coix* O2-homologous protein to oligonucleotides containing normal and truncated O2 target sites. (a) A portion of the α-coixin promoter sequence spanning the O2 target region (boxed) is shown. The palindromic and GCN4-like sequences are depicted (inner boxes). The sequences of the oligonucleotides used in the mobility shift assay are given bellow. The flanking sequence (small caps) of oligo-B, which do not align with the α-coixin promoter, are part of the newly created restriction sites. (b) Mobility shift assay using oligo-A or oligo-B and increasing amounts of the protein extract containing the N-terminal truncated *Coix* O2 protein at final concentrations of 25 ng (lanes 2 and 7), 125 ng (lanes 3 and 8), 250 ng (lanes 4 and 9), and 500 ng (lanes 5 and 10). In lanes 1 and 6, no protein was added. The position of the free probe (fp), lower band (lw), and upper band (up) is shown. (c) Mobility shift assay using the oligo-B probe mixed with different fold molar excess of the unlabeled oligo-B (lanes 3 to 8) or the 63-mer Oct-1 oligonucleotide (lanes 9 and 10) as specific or nonspecific competitors for the binding activity, respectively. In lane 1 no protein was added. In lanes 2 to 10 the amounts of protein extract used corresponds to 25 ng of the N-terminal truncated *Coix* O2 protein.

Figure 2. Mapping and site directed mutagenesis of the O2 target sequence in the α-coixin promoter. (a) The O2 target region (shaded) in the α-coixin promoter is shown. The GCN4- and b-32-like sequences, together with the palindrome are depicted (inner boxes), and the 5′-TGAC-3′ motifs are indicated by arrows. (b) Considering the upstream and downstream inverse TGAC repeats as half-sites (arrow), the linked binding sites *O2u* and *O2d*, respectively, were outlined. (c) Sequence of the O2 target region in the eight different mutant templates obtained by site directed mutagenesis. The mutagenized nucleotides (small caps) are circled. In *m5* and *m6*, the inserted nucleotides (small caps) are shown

ladder should be read C+T when referring to the sequence shown in Figure 2. (b) Mobility shift assay using the same labeled fragments (*m7* and *m8*) as in (a), and increasing N-terminal truncated *Coix* O2 protein concentrations: 0.25 ng (lanes 1 and 6), 0.5 ng (lanes 2 and 7), 2.5 ng (lanes 3 and 8), and 5 ng (lanes 4 and 9). In lane 5, no protein was added. The position of the free probe (fp) and the protein-DNA complex (arrow) is shown. The diamond indicates the position of a band corresponding to the vector plasmid, which derives from an incomplete purification of the insert fragment.

Figure 5. Individual-site titration data for the interaction of the *Coix* O2-homologous protein with the normal and mutant α-coixin promoter templates. Panels show titration data for the binding of the integral *Coix* O2 protein to the adjacent *O2u* (red circles) and *O2d* (green squares) sites. The fractional saturation is plotted against the log of the total O2 protein concentration. The solid curves are the result of simultaneous nonlinear least-square fit to all of the data. The dashed line in the mutant template panels (*m1* to *m8*) represents the isotherm for the O2 binding to site *O2d* in the *normal* template, and was included to facilitate comparisons. Estimated relative interaction energies are in Table 2.

Figure 6. Transactivation of normal and mutant α-coixin promoters by the *Coix* O2-homologous protein in tobacco protoplasts. (a) Schematic representation of chimeric reporter and effector constructs used to analyze the interaction between the O2 protein and the α-coixin promoter. P285α-CGUS, *GUS* gene under the control of the normal α-coixin promoter. P285m1GUS to P285m7GUS, *GUS* gene under the control of the mutant promoters *m1* to *m7* described in Figure 2. The promoter is represented as a shaded arrowhead, and sites *O2u* and *O2d* as open ovals. pRTO2Coix, *Coix O2* gene under the control of the CaMV 35S promoter (arrowhead). The translated sequence of the O2 cDNA is represented as a box. Restriction enzyme sites are A, AccI; Ap, ApaI; B, BamHI; E, EcoRI;

over the site's sequence. In *m7* and *m8*, the sequence resulted from the deletion of one of the binding sites is shown in small caps.

Figure 3. Footprint of the *Coix* O2-homologous protein to the normal and mutant α-coixin promoter templates. The DNA templates, normal, and *m1* to *m4* are indicated above each panel. The positions of the *O2u* and *O2d* binding sites are indicated to the left of the upper (*normal*) panel. For the sequence of the *m1* to *m4* mutant templates refer to Figure 2. Bracketed on the right are blocks of bands used to calculate the extent of saturation of each site. The asterisk indicates the position of a mutated nucleotide. The concentrations of the integral *Coix* O2 protein added to each lane were: lanes 2, 9, 16, 23, and 30, no addition; lanes 3, 10, 17, 24, and 31, 18.7 nM; lanes 4, 11, 18, 25, and 32, 37.5 nM; lanes 5, 12, 19, 26, and 33, 56.2 nM; lanes 6, 13, 20, 27, and 34, 74.9 nM; lanes 7, 14, 21, 28, and 35, 93.7 nM. Lanes 1, 8, 14, 22, and 29 show the G+A sequencing ladder. Note that the G+A sequencing ladder should be read C+T when referring to the sequence shown along the left of the upper panel.

Figure 4. Binding of the *Coix* O2-homologous protein to different mutant templates bearing separated or detached *O2u* and *O2d* sites. (a) Dnase I protection experiment. The DNA templates *m5* to *m8* are indicated above each panel. For the sequence of the *m5* to *m8* mutant templates refer to Figure 2. Bracketed on the right are blocks of bands used to calculate the extent of saturation of each site. The concentrations of the integral *Coix* O2 protein added to each lane were: lanes 2, 9, 16, and 23, no addition; lanes 3 and 10, 30.9 nM; lanes 4 and 11, 93.7 nM; lanes 5 and 12, 156.4 nM; lanes 6 and 13, 218.3 nM; lanes 7 and 14, 281.0 nM. Lane 17, 7.5 nM; lane 18, 9.4 nM; lane 19, 18.7 nM; lane 20, 56.2 nM; lane 21, 93.7 nM. Lane 24, 56.2 nM; lane 25, 93.7 nM; lane 26, 234.2 nM; lane 27, 374.7 nM; lane 28, 515.2 nM. Lanes 1, 8, 14, and 22 show the G+A sequencing ladder. For *m5*, *m6*, and *m8* the G+A sequencing

H3, HindIII; K, KpnI; N, NcoI; P, PstI; S, SalI; Sm, SmaI; Sp, SphI; X, XbaI; Xh, XhoI. (b) GUS activities observed in the transiently transformed tobacco protoplasts. Ten micrograms of the reporter vectors, P285α-CGUS (normal) or P285m1GUS to P285m7GUS (*m1* to *m7*), were transfected alone (white bars) or together with 2-fold molar excess of the effector plasmid pRTO2Coix (shaded bars). GUS activity was normalized according to the expression from the P285α-CGUS + pRTO2Coix (normal, shaded bar) co-transformation. The values represent the mean normalized GUS activity, and standard deviation (δ) is represented by an error bar. Lowercase letters over the bars represent the significant differences ($P<0.05$) between activities as determined by analysis of variance and LSD test, using the Statistica program (StatSoft)

Figure 7. Comparison of the O2 double binding sites with the ocs-elements. Homologous *O2u* and *O2d* site sequences are outlined by open and shaded boxes, respectively. The homologous 5´-TGAC-3´ motifs are indicated by arrows. Note that sequence divergence observed both within the ocs-elements, and between ocs-elements and the O2 binding sites, are comparable. The putative binding sites of the O2 target sequences are 2 bp closely allocated, if compared with the ocs-elements. The sources are α-zein 22-kD (Muth et al., 1996), mas b, mas e (Fox et al., 1992), ocs, nos, ags, CaMV35S (Bouchez et al., 1989).

Table 1. Microscopic binding constants parameters used in writing equations for each template DNA.

Template DNA	Intrinsic binding constants		Cooperative binding constants
	Upstream site	Downstream site	
normal	k_u	k_d	k_{ud}
m1	k_{um1}	k_d	k_{udm1}
m2	k_u	k_{dm2}	k_{udm2}
m3	k_{um1}	k_{dm2}	k_{udm3}
m4	k_u	k_{dm4}	k_{udm4}
m5	k_{um5}	k_d	k_{udm5}
m6	k_{um5}	k_d	k_{udm6}
m7	-	k_d	-
m8	k_{um5}	-	-

Table 2. Relative microscopic Gibbs free energies of O2/α-coixin promoter interactions resolved by footprint titration of normal and mutant binding sites[a].

Template DNA	Intrinsic binding energies		Cooperative energies	Total binding energies[b]
	Upstream site	Downstream site		
normal	ΔG_u	ΔG_d	ΔG_{ud}	$\Delta G_{tnormal}$
	-9.96±0.06	-10.24±0.02	-2.52±0.05	-22.72
m1	ΔG_{um1}	ΔG_d	ΔG_{udm1}	ΔG_{t1}
	-8.91±0.18	-10.24±0.02	-1.59±0.18	-20.74
m2	ΔG_u	ΔG_{dm2}	ΔG_{udm2}	ΔG_{t2}
	-9.96±0.06	-9.13±0.24	-2.64±0.18	-21.73
m3	ΔG_{um1}	ΔG_{dm2}	ΔG_{udm3}	ΔG_{t3}
	-8.91±0.18	-9.13±0.24	0.36±0.34	-17.68
m4	ΔG_u	ΔG_{dm4}	ΔG_{udm4}	ΔG_{t4}
	-9.96±0.06	-8.98±0.32	-1.49±0.34	-20.44
m5	ΔG_{um5}	ΔG_d	ΔG_{udm5}	ΔG_{t5}
	-7.65±0.04	-10.24±0.02	-2.05±0.08	-19.94
m6	ΔG_{um5}	ΔG_d	ΔG_{udm6}	ΔG_{t6}
	-7.65±0.04	-10.24±0.02	-0.73±0.06	-18.62
m7	-	ΔG_d	-	ΔG_{t7}
		-10.24±0.02		-10.24
m8	ΔG_{um5}	-	-	ΔG_{t8}
	-7.65±0.04			-7.65

[a] Relative free energies (in Kilocalories per mole) are presented with 67% confidence intervals (see Methods) according the two site cooperativity model.
[b] The total free energy of binding (ΔG_t) corresponds to the sum of ΔG_i and ΔG_{ij} values.
[c] Relative free energies are related to the corresponding microscopic equilibrium constants, k_i (Table I), by the relationship $\Delta G_i = -RT \ln k_i$.
[d] Data from normal and mutant DNA templates were analyzed simultaneously. They are shown separately in Table only to facilitate comparisons. The square root of variance for this simultaneous analysis was 0.066.

Table 3. Individual site loading free energies ($\Delta G_{L,i}$) from isotherms separately fitted to each template and those from isotherms resolved by the simultaneous least-squares analysis[a].

Template DNA	Isotherm resolution	$\Delta G_{L,u}$	$\Delta G_{L,um1}$	$\Delta G_{L,um5}$	$\Delta G_{L,d}$	$\Delta G_{L,dm2}$	$\Delta G_{L,dm4}$
normal	separately	-11.27			-11.24		
	simultaneous	-11.34			-11.39		
m1	separately		-10.19		-10.44		
	simultaneous		-10.10		-10.64		
m2	separately	-11.01				-10.85	
	simultaneous	-10.94				-10.79	
m3	separately		-8.63			-9.06	
	simultaneous		-8.70			-8.97	
m4	separately	-10.37					-10.03
	simultaneous	-10.42					-10.02
m5	separately			-9.65	-10.24		
	simultaneous			-9.48	-10.47		
m6	separately			-8.53	-10.40		
	simultaneous			-8.33	-10.29		
m7	separately				-10.53		
	simultaneous				-10.24		
m8	separately			-7.61			
	simultaneous			-7.65			

[a] $\Delta G_{L,i}$ values (in Kilocalories per mole) were calculated by numerical integration according to Ackers et al. (1983).

Figure 1

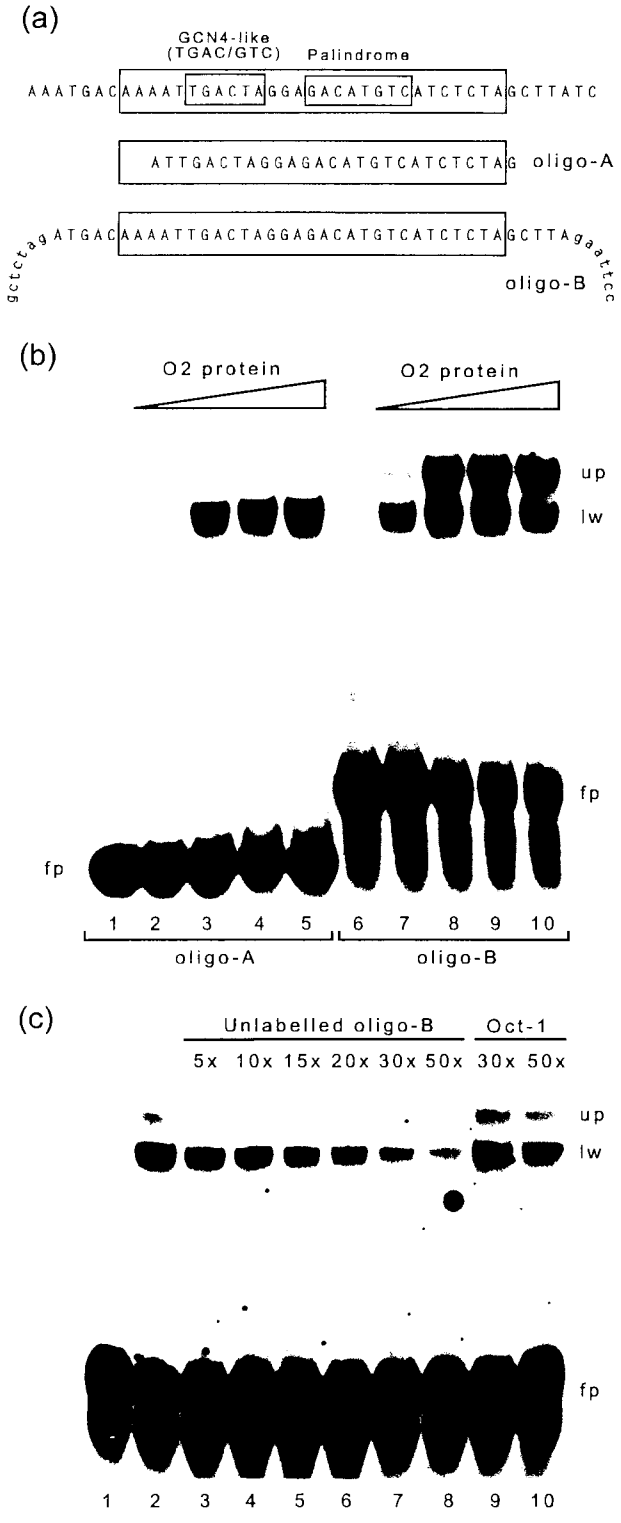

(a)
GCN4-like
(TGAC/GTC)
Palindrome
AAATGACAAAATTGACTAGGAGACATGTCATCTCTAGCTTATC
ATTGACTAGGAGACATGTCATCTCTAG
oligo-A
gctctagATGACAAAATTGACTAGGAGACATGTCATCTCTAGCTTAgaattcc
oligo-B
(b)
O2 protein
O2 protein
up
lw
fp
fp
1 2 3 4 5 6 7 8 9 10
oligo-A
oligo-B
(c)
Unlabelled oligo-B
Oct-1
5x 10x 15x 20x 30x 50x 30x 50x
up
lw
fp
1 2 3 4 5 6 7 8 9 10

Figure 2

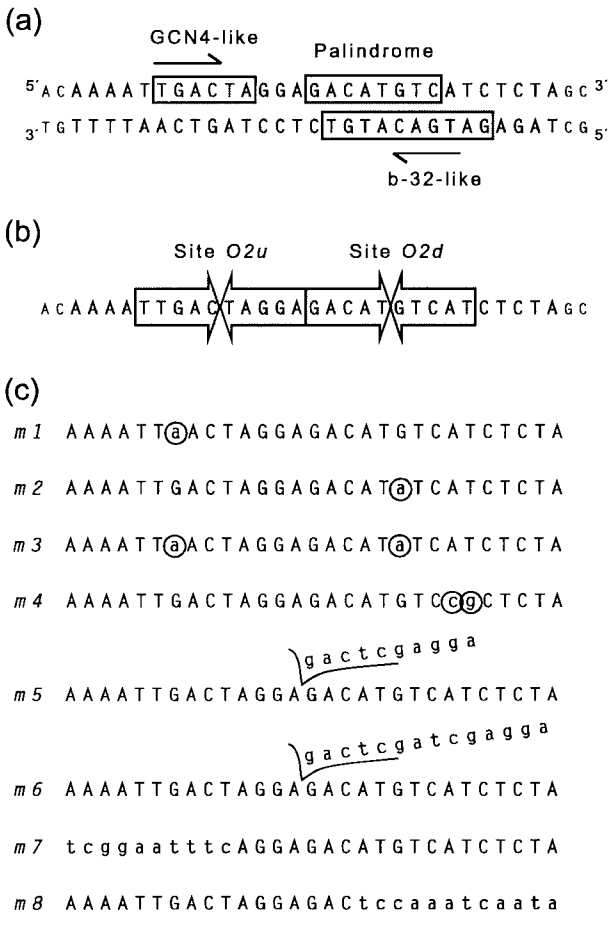

(a)
GCN4-like
Palindrome
5′ACAAAATTGACTAGGAGACATGTCATCTCTAGC 3′
3′TGTTTTAACTGATCCTCTGTACAGTAGAGATCG 5′
b-32-like
(b)
Site O2u
Site O2d
ACAAAATTGACTAGGAGACATGTCATCTCTAGC
(c)
m1 AAAATTaACTAGGAGACATGTCATCTCTA
m2 AAAATTGACTAGGAGACATaTCATCTCTA
m3 AAAATTaACTAGGAGACATaTCATCTCTA
m4 AAAATTGACTAGGAGACATGTCcgCTCTA
gactcgagga
m5 AAAATTGACTAGGAGACATGTCATCTCTA
gactcgatcgagga
m6 AAAATTGACTAGGAGACATGTCATCTCTA
m7 tcggaatttcAGGAGACATGTCATCTCTA
m8 AAAATTGACTAGGAGACtccaaatcaata

Figure 3

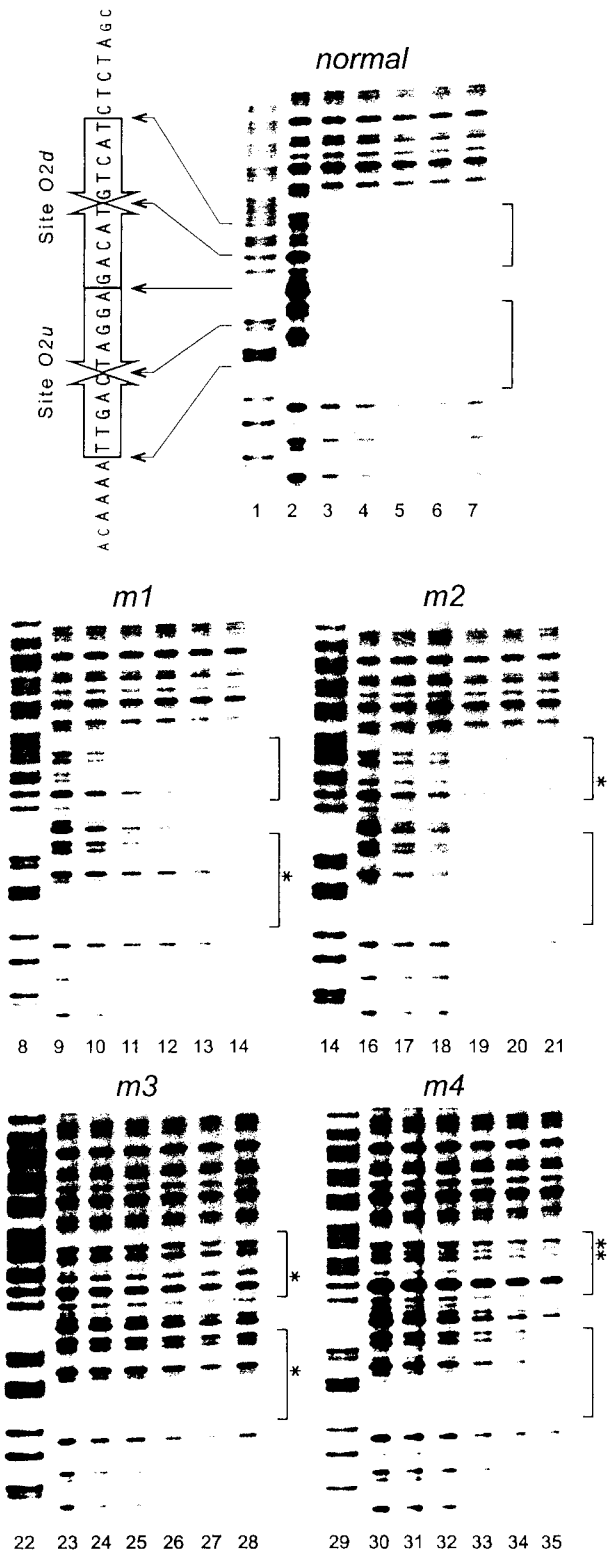

normal
Site O2d
Site O2u
ACAAAATTGACTAGGAGACATGTCATCTCTAGC
1 2 3 4 5 6 7
m1
m2
8 9 10 11 12 13 14
14 16 17 18 19 20 21
m3
m4
22 23 24 25 26 27 28
29 30 31 32 33 34 35

Figure 4

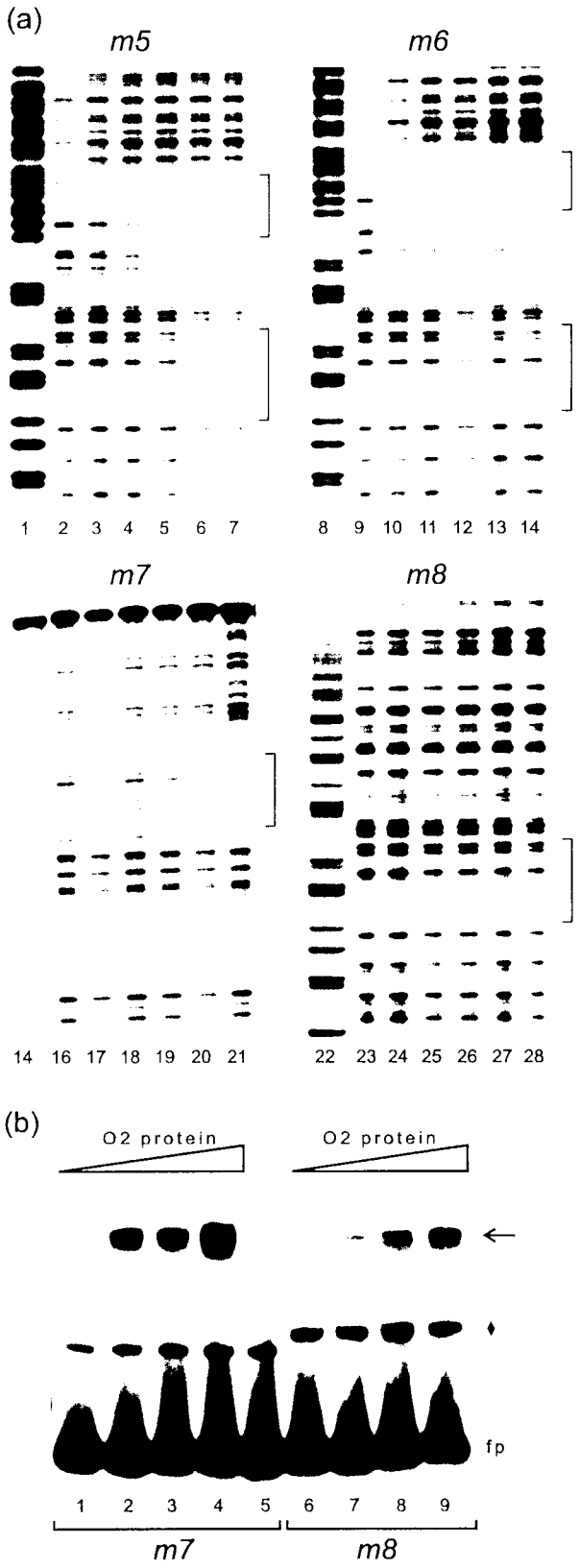

(a)
m5
m6
1 2 3 4 5 6 7
8 9 10 11 12 13 14
m7
m8
14 16 17 18 19 20 21
22 23 24 25 26 27 28
(b)
O2 protein
O2 protein
fp
1 2 3 4 5 6 7 8 9
m7
m8

Figure 5

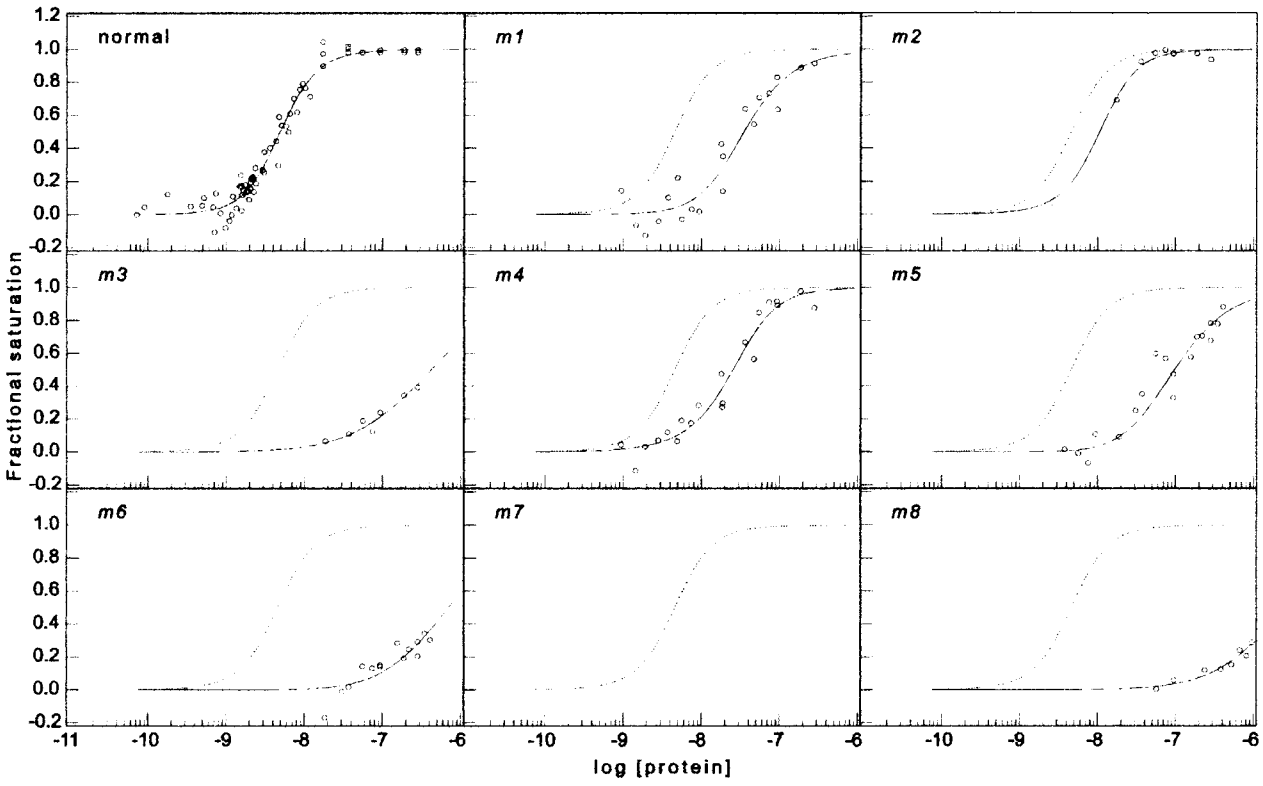

normal
m1
m2
m3
m4
m5
m6
m7
m8
Fractional saturation
log [protein]
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-11
-10
-9
-8
-7
-6

Figure 6

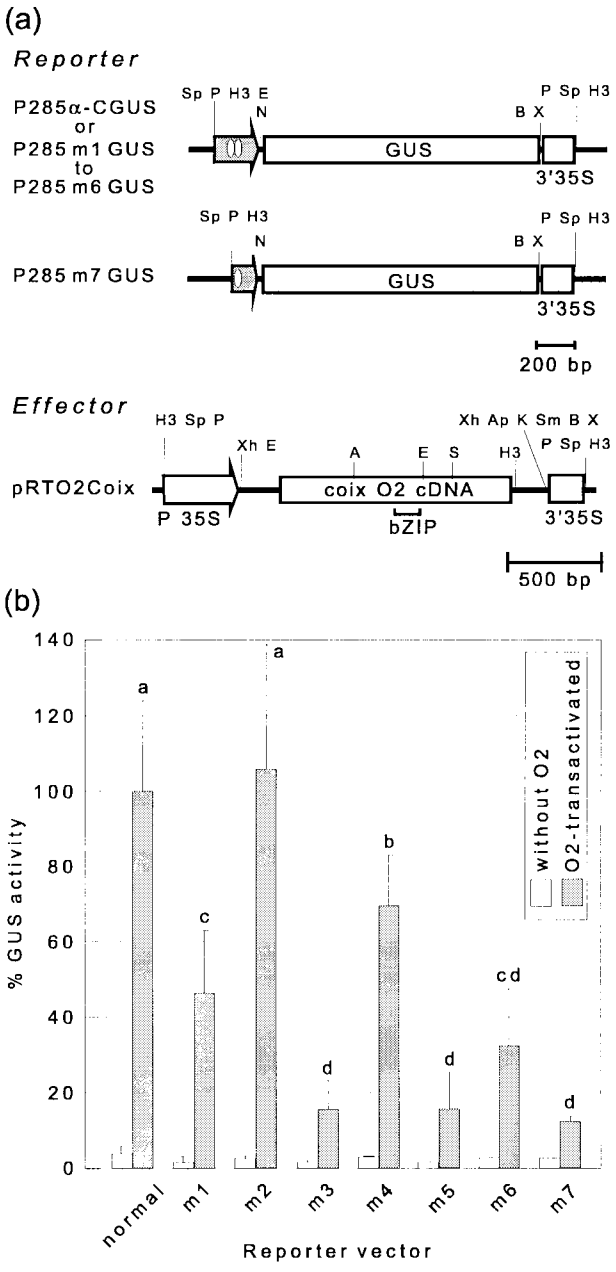
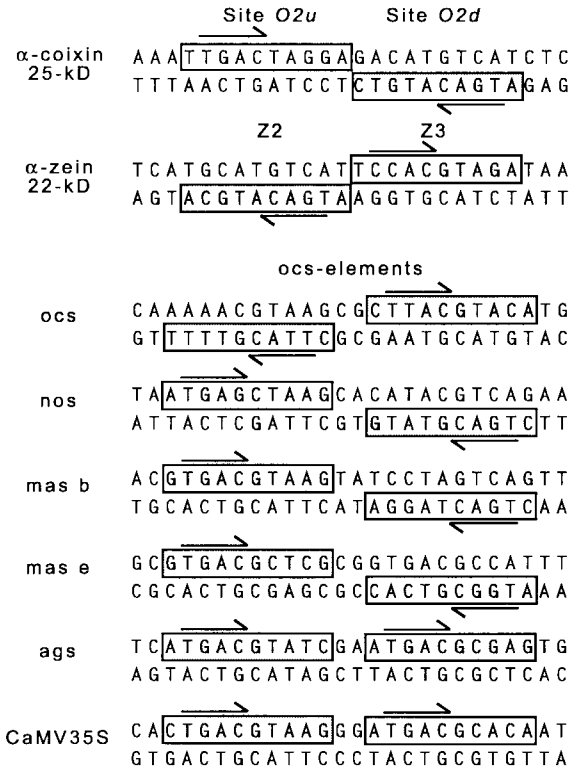

(a)
Reporter
P285α-CGUS
or
P285 m1 GUS
to
P285 m6 GUS
Sp P H3 E
N
P Sp H3
B X
GUS
3'35S
P285 m7 GUS
200 bp
Effector
pRTO2Coix
H3 Sp P
Xh E
A
E S
H3
Xh Ap K Sm B X
P Sp H3
coix O2 cDNA
P 35S
bZIP
3'35S
500 bp
(b)
% GUS activity
without O2
O2-transactivated
normal
m1
m2
m3
m4
m5
m6
m7
Reporter vector

Figure 7


Site O2u
Site O2d
α-coixin 25-kD
AAATTGACTAGGAGACATGTCATCTC
TTTAACTGATCCTCTGTACAGTAGAG
Z2
Z3
α-zein 22-kD
TCATGCATGTCATTCCACGTAGATAA
AGTACGTACAGTAAGGTGCATCTATT
ocs-elements
ocs
CAAAAACGTAAGCGCTTACGTACATG
GTTTTTGCATTCGCGAATGCATGTAC
nos
TAATGAGCTAAGCACATACGTCAGAA
ATTACTCGATTCGTGTATGCAGTCTT
mas b
ACGTGACGTAAGTATCCTAGTCAGTT
TGCACTGCATTCATAGGATCAGTCAA
mas e
GCGTGACGCTCGCGGTGACGCCATTT
CGCACTGCGAGCGCCACTGCGGTAAA
ags
TCATGACGTATCGAATGACGCGAGTG
AGTACTGCATAGCTTACTGCGCTCAC
CaMV35S
CACTGACGTAAGGGATGACGCACAAT
GTGACTGCATTCCCTACTGCGTGTTA

DISCUSSÃO*

Após a formulação da hipótese do "operon" por Jacob e Monod (1961) e desde que foi estabelecido que a regulação da expressão gênica era mediada por proteínas, a tentativa da compreensão molecular das interações destas proteínas com os promotores, operadores e "enhancers" correspondentes aos seus alvos de regulação tem sido um dos tópicos centrais na Biologia Molecular. Isto porque o entendimento do processo de regulação gênica, requer a elucidação de questões relacionadas à especificidade das interações proteína-DNA, ou seja, de como uma proteína pode reconhecer e discriminar uma sequência particular de DNA no contexto do genoma como um todo.

Em plantas, a grande maioria dos fatores reguladores da transcrição descritos pertence à classe das bZIPs. Inicialmente, todas as bZIPs de plantas cuja especificidade de ligação ao DNA foi estudada mostraram-se capazes de se ligar a sequências contendo o motivo central ACGT (Izawa *et al.*, 1993; Foster *et al.*, 1994). Isto levou à conclusão generalizada de que as bZIPs de planta reconheciam apenas elementos *cis* com ACGT, variando a afinidade da ligação - e portanto a especificidade - em função das sequências flanqueadora deste núcleo (Williams *et al.*, 1992; Izawa *et al.*, 1993).

Entretanto, todas as bZIPs estudadas, com exceção da Opaco2 de milho e da PosF21 de *Arabidopsis* (Schmidt *et al.*, 1987; Aeschbacher *et al.*, 1991), haviam sido isoladas por seleção em bibliotecas de expressão, utilizando-se como sonda uma sequência de DNA contendo o motivo ACGT (Meshi e Iwabuchi, 1995). Portanto, era esperado que as proteínas codificadas pelos genes de tais clones apresentassem afinidade por sequências contendo ACGT. Logo, a inferência de que o núcleo ACGT é imprescindível aos elementos *cis*

---

* Uma vez que cada um dos dois artigos apresenta uma ampla discussão dos resultados, esta sessão consta de uma discussão geral, onde foi abordada a contribuição de nosso trabalho para o entendimento da regulação gênica em plantas.

reconhecidos pelas bZIPs em plantas era incorreta, visto que a amostra de bZIPs estudada não era representativa.

Outra questão a ser discutida é o fato de as várias bZIPs serem capazes de reconhecer sítios distintos com ACGT central, ou seja, o fato de apresentarem uma especificidade "relaxada" por sequências com ACGT (Izawa *et al.*, 1993). Este resultado poderia levar a duas interpretações: ou estas bZIPs estariam realmente envolvidas na expressão de uma ampla variedade de genes e por isso apresentariam baixa especificidade para permitir a transativação dos genes; ou então nem todas as sequências reconhecidas pela proteína *in vitro* teriam uma função *in situ*. A última hipótese foi mais aceita, e portanto postulou-se que somente aquelas sequências com maior afinidade pela proteína regulatória seriam fisiologicamente importantes (Izawa *et al.*, 1993).

Entretanto, os estudos de especificidade de ligação de bZIPs de planta que levaram a esta premissa eram feitos com o uso de sequências curtas de DNA (oligonucleotídeos) aleatoriamente desenhadas, em que se variavam dois ou três nucleotídeos em torno de um núcleo ACGT (Izawa *et al.*, 1993). Isto porque, exceção feita ao caso da proteína O2, não se conhecem quais genes estão sob o controle de quaisquer das bZIPs, e portanto, não se conhecem os elementos *cis* naturalmente reconhecidos por elas, de tal modo que os resultados obtidos nestes experimentos devem ser interpretados com certa cautela.

Trabalhando com a Opaco2, tinha-se a vantagem de poder analisar os genes que estão sob seu controle; ou seja, havia a possibilidade de analisar a especificidade de ligação de um fator bZIP de plantas com seus sítios naturais de interação com o DNA. Ao mesmo tempo, pela comparação dos resultados obtidos em milho e *Coix*, poder-se-ia vislumbrar tendências evolutivas com respeito a tal especificidade de ligação.

Ao contrário do proposto na hipótese generalizada que julgava imprescindível a presença do motivo ACGT (Schmidt *et al.*, 1992), e apesar do descrédito dado por Izawa *et al.* (1993) ao trabalho referente à determinação do

sítio de ligação da proteína O2 no gene *b-32* (Lohmer *et al.*, 1991), verificou-se que o sítio de ligação da O2 no gene da α-coixina não apresentava o citado motivo ACGT (Yunes *et al.*, 1994b). Resultados semelhantes foram obtidos do estudo da interação da proteína O2 e outra bZIP, a TGA1a, com o promotor de um gene de lectina de ervilha (de Pater *et al.*, 1994), e mais recentemente da análise da interação do fator O2 com um gene de proteína de reserva de trigo (Holdsworth *et al.*, 1995). Embora não se saiba ainda da existência de fatores regulador tipo Opaco2 em ervilha ou trigo, estes resultados corroboram a idéia de que o motivo ACGT não é fundamental, pelo menos para algumas das bZIPs.

O panorama que surgia dos estudos de afinidade de ligação das bZIPs a oligonucleotídeos artificialmente desenhados sugeria que a regulação exercida por um fator de transcrição sobre um dado gene-alvo era função da presença de um sítio de ligação altamente específico, cuja sequência primária de nucleotideos permitiria uma alta afinidade de ligação. Paradoxalmente, os resultados obtidos do estudo dos sítios naturais de ligação da O2 mostravam que um mesmo fator de transcrição seria capaz de regular genes relacionados, como os da α-zeína de 22 kDa e α-coixina de 25 kDa, através de sítios de ligação bastante diferentes.

Esta última constatação, conjuntamente com dados obtidos da análise dos genes da β-coixina e β-zeína (Cord Neto *et al.*, 1995), onde uma nova sequência de ligação da proteína O2 foi encontrada, indica que a especificidade de ligação de um fator ao DNA, e por conseguinte o padrão específico de expressão espacial e temporal de seus genes-alvo, não parece ser devido unicamente à presença de uma sequência específica e conservada de nucleotídeos nos elementos *cis*-regulatórios.

Nenhum dos sítios naturais descritos para a proteína O2 (Lohmer *et al.*, 1991; Schmidt *et al.*, 1992; Yunes *et al.*, 1994b; Cord Neto *et al.*, 1995) corresponde à sequência de maior afinidade por O2 descrita nos estudos de ligação a oligonucleotídeos (Izawa *et al.*, 1993). Resultado semelhante foi descrito para o caso da proteína GCN4 (Hill *et al.*, 1986), e mesmo para o

operador *lac* de *Escherichia coli* (Simons *et al.*, 1994). Este fato sugere que durante o processo evolutivo o aumento na especificidade de interação proteína/DNA não foi maximizado, o que está de acordo com predições teóricas feitas por Berg e von Hippel (1987) ao abordar, por um método mecano-estatístico, a questão da seleção de sítios de ligação ao DNA pelas proteínas regulatórias.

A teoria descrita por Berg e von Hippel (1987) parte do pressuposto de que sequências específicas têm sido positivamente selecionadas no promotor dos genes, de forma a possibilitar uma certa afinidade de ligação ou atividade biológica. Por outro lado, em oposição a esta tendência tem-se as mutações fortuitas introduzindo desordem no sistema.

A evolução tenderia então a atuar tanto no aumento da complexidade estrutural da proteína regulatória, aperfeiçoando o processo de discriminação do DNA pela proteína e vice-versa, como no tamanho da sequência-alvo no DNA, de forma a diminuir a probabilidade de que haja interação proteína-DNA ao acaso. Ambas as estratégias fariam com que a ligação fosse o suficientemente seletiva para evitar a interação do fator com pseudosítios no DNA.

Espera-se então que quanto maior for a complexidade estrutural da proteína regulatória, menor poderá ser o sítio alvo; entretanto, neste caso, menor poderá ser também a variabilidade permitida em ambos os componentes, proteína e sequência de DNA. Diz-se então que a interação proteína-DNA, por ser altamente precisa, atingiu um grau máximo de especificidade. A alta especificidade de ligação é vantajosa na medida em que permite que a célula reduza seu investimento na produção da proteína necessária para obter níveis apropriados de interação com as sequências-alvo específicas. Entretanto, tal vantagem é contrabalanceada pelos efeitos muito mais danosos que as eventuais mutações causariam numa situação como esta, bem como, pelos problemas decorrentes de pequenas flutuações na concentração da proteína (Berg, 1978; von Hippel e Berg, 1986).

Ao contrário, como mostram os resultados obtidos no estudo da especificidade de ligação da proteína O2, evolutivamente parece ser mais vantajosa uma condição onde a interação proteína-DNA seja flexível. Tal flexibilidade deve ser suficiente para que a maioria das eventuais mutações pontuais, tanto na proteína quanto no sítio do DNA, tenham uma pequena influência no processo regulatório sem todavia serem letais. Desta forma, ao invés de maximizar a especificidade da interação, a tendência evolutiva é no sentido de *minimizar a máxima perda de especificidade* (Berg e von Hippel, 1987).

Em decorrência disto, nem a proteína regulatória, nem a sequência no DNA serão os melhores ligantes, e por este motivo, os resultados de medição da afinidade de ligação da proteína a oligonucleotídeos acima citados devem ser interpretados com cautelosamente.

Uma análise mais minuciosa do local de ligação da O2 no promotor da α-coixina mostrou que o mesmo era composto por dois sítios adjacentes. É importante lembrar que este não é um fato isolado, já que tanto para o gene *b-32* (Lohmer *et al.*, 1991) quanto para o da α-zeina de 22 kDa (Muth *et al.*, 1996) também foram descritos múltiplos sítios de O2, alguns dos quais localizados em posições adjacentes.

Através de experimentos de "footprint" quantitativo demonstrou-se que a ligação da proteína O2 a estes sítios adjacentes do promotor da α-coixina ocorre de maneira cooperativa, uma característica também prevista no modelo evolutivo de Berg e von Hippel (1987): um fator regulatório menos discriminatório requer um sítio maior no DNA, cuja chance de ocorrer ao acaso no genoma seja menor, de forma a reduzir a um nível aceitável a existência de pseudosítios de alta afinidade; por sua vez, um sítio maior requer um domínio protéico de ligação ao DNA maior, maximizando assim a especificidade da interação proteína-DNA. Ao contrário, a formação de dímeros e/ou a cooperatividade permitem que a proteína regulatória se torne capaz de reconhecer um sítio maior sem no entanto maximizar a especificidade de

ligação. Obviamente, o custo por este não-aumento na especificidade da interação proteína-DNA determina um novo requisito, como a necessidade de interações proteína-proteína. Em suma, o que ocorre é que a combinação de processos específicos permite um relaxamento no requerimento pela especificidade das reações individuais.

Desta forma, a regulação da expressão gênica não depende somente da especificidade de ligação de um determinado fator regulatório, mas é decorrente em grande parte da combinação de interações entre várias proteínas regulatórias. Em consequência disso, duas observações podem ser feitas:

(i) Contrariamente a algumas das concepções atuais, é evidente que um elemento *cis* qualquer, apresentando forte ou fraca afinidade por um determinado fator regulatório, não pode ser analisado em termos de sua importância fisiológica quando separado do contexto molecular como um todo. Prova disto é o relato de Muth *et al.* (1996) de que um sítio de baixa afinidade pela proteína O2, isoladamente incapaz de mediar à ativação da transcrição por O2, possibilita níveis significativamente mais elevados de transcrição quando em combinação com qualquer um dos outros dois sítios de ligação da O2 encontrados no promotor da α-zeína de 22 kDa.

(ii) A afinidade da interação proteína-DNA não é fator determinante do processo, já que o fato de a proteína estar ligada ao DNA não é garantia de que ocorrerá a ativação da transcrição. De fato, conforme demonstrado em nosso segundo trabalho, a estrutura tridimensional do complexo proteína-DNA formado é tão importante quanto a afinidade da interação *per se* para a expressão final do gene, pois é o que em definitivo possibilita ou não a combinação dos vários fatores que interagem e possibilitam a transcrição dos genes.

CONCLUSÕES

1. O gene da α-coixina de 25 kDa contém no seu promotor uma região de 29 pares de base que é especificamente reconhecida pela proteína regulatória Opaco2 (O2) de milho e *Coix*. A sequência deste elemento *cis* é 5'-AAAATTGACTAGGAGACATGTCATCTCTA-3' e se encontra localizada entre os nucleotideos -159 e -187 em relação ao A (+1) do codon ATG que marca o início da tradução.
2. Este elemento *cis* é constituído por dois sítios adjacentes, aos quais a proteína O2 é capaz de se ligar simultaneamente.
3. O sítio mais distante do ínicio de transcrição foi chamado *O2u* e compreende a sequência 5'-TTGACTAGGA-3'. O sítio mais próximo foi chamado *O2d* e compreende a sequência 5'-GACATGTCAT-3'. Nenhum destes sítios contém o motivo central ACGT, descrito por vários autores como sendo importante para a ligação de proteínas bZIP de plantas aos seus respectivos sítios no DNA.
4. Os sítios *O2u* e *O2d* apresentam similaridade com a sequência reconhecida pela proteína bZIP de levedura GCN4. Dentre todos os sítios de ligação da proteína O2 descritos na literatura, estes asemelham-se mais aos encontrados no promotor do gene *b-32*.
5. Os sítios *O2u* e *O2d* apresentam afinidade semelhante pela proteína O2. A ligação da proteína O2 a um destes sítios facilita em aproximadamente 70 vezes a ligação ao segundo, o que caracteriza uma interação cooperativa da proteína O2 com tais sítios.
6. A presença deste elemento *cis* é fundamental para a obtenção de altos níveis de transcrição mediada pela proteína O2. Os sítios *O2u* e *O2d* interagem de maneira sinergística na transcrição mediada por O2. Parte do efeito sinergístico na transcrição é devido à cooperatividade da ligação.
7. Considerando os sítios individualmente, a ligação da proteína O2 ao sítio *O2u* tem maior efeito na taxa final de transcrição do que a ligação ao sítio *O2d*.

8. A atividade transcricional não depende unicamente da energia total de ligação da proteína O2 ao promotor, mas também da correta conformação espacial do complexo proteína-DNA.

BIBLIOGRAFIA

Aeschbacher, R.A., Schrott, M. Potrykus, I. and Saul, M.W. (1991). Isolation and molecular characterization of PosF21, an Arabidopsis thaliana gene which shows characteristics of a b-Zip class transcription factor. *Plant J.* 1: 303-316.

Apel, K., Batschauer, A., Dehesh, K., Kreuz, K. and Steinmuller, K. (1986). The complexity of gene expression in higher plants. *Phil. Trans. R. Soc. Lond.* B314: 481-492.

Bass, H.W., Webster, C., OBrian, G.R., Roberts, J.K. and Boston, R.S. (1992). A maize ribosome-inactivating protein is controlled by the transcriptional activator opaque-2. *Plant Cell* 4: 225-234.

Benfey, P.N. and Chua, N.-H. (1989). Regulated genes in transgenic plants. *Science* 244: 174-181.

Berg, O.G. (1978). A model for the statistical fluctuations of protein numbers in a microbial population. *J. Theor. Biol.* 71: 587-603.

Berg, O.G. and von Hippel, P.H. (1987). Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters. *J. Mol. Biol.* 193, 723-750.

Bianchi, M.W. and Viotti, A. (1988). DNA methylation and tissue-specific transcription of the storage protein genes of maize. *Plant Mol. Biol.* 11: 203-214.

Bonven, B.J., Nielsen, A.L., Norby, P.L., Pedersen, F.S. and 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.

Breathnach, R. and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Ann. Rev. Biochem.* 50: 349-383.

Buratowski, S. (1994). The basics of basal transcription by RNA polymerase II. *Cell* 77: 1-3.

Burr, F.A. and Burr, B. (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.

Chomet, P.S. (1991). Cytosine methylation in gene-silencing mechanisms. *Curr. Opin. Cell Biol.* 3: 438-443.

Cord Neto, G., Yunes, J.A., Silva, M.J., Vettore, A.L., Arruda, P. and Leite, A. (1995). The involvement of Opaque 2 on β-prolamin gene regulation in maize and *Coix* suggests a more general role for this transcriptional activator. *Plant Mol. Biol.* 27, 1015-1029.

Curran, T. and Franza Jr., B.R. (1988). Fos and Jun: the AP-1 connection. *Cell* 55: 395-397.

de Pater, S., Katagiri, F., Kijne, J. and Chua, N.-H. (1994). bZIP proteins bind to a palindromic sequence without an ACGT core located in a seed-specific element of the pea lectin promoter. *Plant J.* 6, 133-140.

Du, W., Thanos, D. and Maniatis, T. (1993). Mechanisms of transcriptional synergism between distinct virus-inducible enhancer elements. *Cell* 74: 887-898.

Dure, L.S.III. (1975). Seed formation. *Ann. Rev. Plant Physiol.* 26: 259-278.

Dynan, W.S. and Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA binding proteins. *Nature* 316: 774-778.

Dynan,W.S. (1989). Understanding the molecular mechanism by which methylation influences gene expression. *Trends Genet.* 5: 35-36.

Ehrlich, M. and Ehrlich, K.C. (1993). Effect of DNA methylation on the binding of vertebrate and plant proteins to DNA. *EXS* 64: 145-168.

Ellenberger, T.E., Brandl, C.J., Struhl, K. and Harrison, S.C. (1992). The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. Cell 71: 1223-1237.

Esen, A. (1986). Separation of alcohol-soluble proteins (zeins) from maize into three fractions by diferential solubility. *Plant Physiol.* 80: 623-627.

Feix, G. and Quayle, T. (1993). Structure and expresssion of zein genes of maize. *Critical Reviews in Plant Sciences* 12: 111-127.

Finnegan, E.J., Brettell, R.I. and Dennis, E.S. (1993). The role of DNA methylation in the regulation of plant gene expression. *EXS* 64: 218-261.

Folkers, N.S., Borrelli, E. and Sassone-Corsi, P. (1991). CREM gene: use of alternative DNA-binding domains generates multiple antagonists of cAMP-induced transcription. *Cell* 64: 739-749.

Foster, R., Izawa, T. and Chua, N.-H. (1994). Plant bZIP proteins gather at ACGT elements. *FASEB J.* 8: 192-200.

Gallie, D.R. (1993). Posttranscriptional regulation of gene expression in plants. *Plant Mol. Biol.* 44: 77-105.

Garratt, R., Oliva, G., Caracelli, I., Leite, A. and 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. and 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.

Giese, K., Kingsley, C., Kirshner, J.R. and Grosschedl, R. (1995). Assembly and function of a TCRα enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev.* 9: 995-1008.

Gille, H., Sharrocks, A. and 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. and Chua, N.-H. (1990). Molecular light switches for plant genes. *Plant Cell* 2: 369-378.

Gonzales, G.A. and Montminy, M. (1989). Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. *Cell* 59: 675-680.

Goodrich, J.A., Cutler, G. and Tjian, R. (1996). Contacts in context: promoter specificity and macromolecular interactions in transcription. *Cell* 84: 825-830.

Grosschedl, R. (1995). High-order nucleoprotein complexes in transcription: analogies with site-specific recombination. *Curr. Opin. Cell Biol.* 7: 362-370.

Hartings, H., Maddaloni, M., Lazzaroni, N., Di Fonzo, N., Motto, M., Salamini, F. and Thompson, R. (1989). The *O2* gene which regulates zein deposition in maize endosperm encodes a protein with structural homologies to transcriptional activators. *EMBO J.* 8, 2795-2801.

Hepburn, A.G., Clarke, L.E., Pearson, L. and White, J. (1983). The role of methylation in the control of nopalin synthase gene expression in a plant tumor. *J. Mol. Appl. Genet.* 2: 315-329.

Higgins, T.-J.V. (1984). Synthesis and regulation of major proteins in seeds. *Ann. Rev. Plant Physiol.* 35:191-221.

Hill, D.E., Hope, I.A., Macke, J.P. and Struhl, K. (1986). Saturation mutagenesis of the yeast *his3* regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. *Science,* 234, 451-457.

Holdsworth, M., Muñoz-Blanco, J., Hammond-Kosack, M., Colot, V., Schuch, W. and 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.

Inamdar, N.M., Ehrlich, K.C. and Ehrlich, M. (1991). CpG methylation inhibits binding of several sequence-specific DNA-binding proteins from pea, wheat, soybean and cauliflower. *Plant Mol. Biol.* 17: 111-123.

Ingelbrecht, I., Van Houdt, H., Van Montagu, M. and Depicker, A. (1994). Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation. *Proc. Natl. Acad. Sci. USA* 91: 10502-10506.

Izawa, T., Foster, R. and Chua, N.-H. (1993). Plant bZIP protein DNA binding specificity. *J. Mol. Biol.* 230, 1131-1144.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3: 318-356.

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.

Kerstetter, R., Vollbrecht, E., Lowe, B., Veit, B., Yamaguchi, J. and Hake, S. (1994). Sequence analysis and expression patterns divide the maize *knotted1*-like homeobox genes into two classes. *Plant Cell* 6: 1877-1887.

Keshet, I., Lieman-Hurwitz, J. and Cedar, H. (1986). DNA methylation affects the formation of active chromatine. *Cell* 44: 535-543.

Kirihara, J.A., Husperger, J.P. and Mahoney, J.W. (1988). Differential expression of a gene for a methionine-rich storage protein. *Mol. Gen. Genet.* 211: 477-484.

Kodrzycki, R., Boston, R.S. and Larkins, B.A. (1989). The *opaque-2* mutation of maize differentially reduces zein gene transcription. *Plant Cell* 1: 105-114.

Kreil, G. (1981). Transfer of proteins across membranes. *Ann. Rev. Biochem.* 50: 317-348.

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. and Miflin, B.J. (1986). Differential gene expression in the developing barley endosperm. *Phil. Trans. R. Soc. Lond.* B314: 355-365.

Kuhlemeier, C. (1992). Transcriptional and post-transcriptional regulation of gene expression in plants. *Plant Mol. Biol.* 19: 1-14.

Kuhlemeier, C., Green, P.J. and Chua, N.-H. (1987). Regulation of gene expression in higher plants. *Ann. Rev. Plant Physiol.* 38: 221-257.

Landry, J. and 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.

Larkins, B.A. (1981). Seed storage proteins: characterization and biosynthesis. *Biochemistry of. Plants*, 6: 449-489.

Leite, A., Ottoboni, L.M.M., Targon, M.L.P.N., Silva, M.J., Turcinelli, S.R. and 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., Yunes, J.A., Turcinelli, S.R. and Arruda, P. (1992). Cloning and characterization of a cDNA encoding a sulfur-rich coixin. *Plant Mol. Biol.* 18: 171-174.

Lending, C.R. and Larkins, B.A. (1989). Changes in the zein composition of protein bodies during maize endosperm development. *Plant Cell* 1: 1011-1023.

Lohmer, S., Maddaloni, M., Motto, M., Di Fonzo, N., Hartings, H., Salamini, F. and 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.

Madison, J.T., Thompson, J.F., Muenster, A.E. (1981). Turnover of storage protein in seeds of soya bean and pea. *Ann. Bot.* 47: 65-73.

Maldonado, E. and Reinberg, D. (1995). News on initiation and elongation of transcription by RNA polymerase II. *Curr. Opin. Cell Biol.* 7: 352-361.

Maniatis,T., Goodbourn,S. and Fischer,J.A. (1987). Regulation of inducible and tissue-specific gene expression. *Science* 236: 1237-1245.

Mauri, I., Maddaloni, M., Lohmer, S., Motto, M., Salamini, F., Thompson, R. and Martegani, E. (1993). Functional expression of the transcriptional activator *Opaque-2* of *Zea mays* in transformed yeast. *Mol. Gen. Genet.* 241, 319-326.

Meshi, T. and Iwabuchi, M. (1995). Plant transcription factors. *Plant Cell Physiol.* 36: 1405-1420.

Meyer, P. (1995). DNA methylation and transgene silencing in Petunia hybrida. *Curr. Top. Microbiol. Immunol.* 197: 15-28.

Motto, M., Di Fonzo, N., Hartings, H., Maddaloni, M., Salamini, F., Soave, C. and Thompson, R.D. (1989). Regulatory genes affecting maize storage protein synthesis. *Oxf. Surv. Plant Mol. Cell Biol.* 6: 87-114.

Muntz, K., Bassuner, R., Lichtenfeld, C., Scholz, G. and Weber, E. (1985). Proteolytic cleavage of storage proteins during embryogenesis and germination of legume seeds. *Physiol. Vég.* 23: 75-94.

Muth, J.R., Müller, M., Lohmer, S., Salamini, F. and Thompson, R.D. (1996). The role of multiple binding sites in the activation of zein gene expression by *Opaque-2*. *Mol. Gen. Genet.* 252: 723-732.

Natesan, S. and Gilman, M.Z. (1993). DNA bending and orientation-dependent function of YY1 in the c-fos promoter. *Genes Dev.* 7: 2497-2509.

O'Neil, K.T., Hoess, R.H. and DeGrado, W.F. (1990). Design of DNA-binding peptides based on the leucine zipper motif. *Science* 249: 774-778.

Osborne, T.B. and 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., Turcinelli, S.R. and Arruda, P. (1990). 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 Filho, G.A. and Arruda, P. (1993). Sequence analysis of 22 kDa-like α-coixin genes and their comparison with homologous zein and kafirin genes reveals higly conserved protein structure and regulatory elements. *Plant Mol. Biol.* 21: 765-778.

Pernollet, J.-C. (1985). Biosynthesis and accumulation of storage proteins in seeds. *Physiol. Vég.* 23: 45-59.

Pernollet, J.-C. and Mossé,J. (1983). Structure and location of legume and cereal seed storage proteins. In *Seed Proteins*, J. Daussant, J. Mossé and J. Vaughan, ed., Academic Press, London, 155-191.

Pirovano, L., Lanzini, S., Hartings, H., Lazzaroni, N., Rossi, V., Joshi, R., Thompson, R.D., Salamini, F. and Motto, M. (1994). Structural and functional analysis of an Opaque-2-related gene from sorghum. *Plant Mol. Biol.* 24: 515-523.

Purugganan, M.D., Rounsley, S.D., Schmidt, R.J. and Yanofsky, M.F. (1995). Molecular evolution of flower development: diversification of the plant MADS-box regulatory gene family. *Genetics* 140: 345-356.

Ramachandran, S., Hiratsuka, K. and Chua, N.-H. (1994). Transcription factors in plant growth and development. *Curr. Opin. Genet. Dev.* 4: 642-646.

Schleif, R. (1988). DNA binding by proteins. *Science* 241: 1182-1187.

Schmidt, R.J., Burr, F.A. and 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. and 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. and 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.

Shewry, P.R. and Tatham, A.S. (1990). The prolamin storage proteins of cereal seeds: structure and evolution. *Biochem. J.* 267: 1-12.

Shuai, K., Stark, G.R., Kerr, I. and Darnell, J.E.J. (1993). A single phosphotyrosine residue of Stat91 required for gene activation by interferon-gamma. *Science* 261: 1744-1746.

Simons, A., Tils, D., von Wilcken-Bergmann, B. and Muller-Hill, B. (1984). Possible ideal lac operator: *Escherichia coli* lac operator-like sequences from eukaryotic genomes lack the central G X C pair. *Proc. Natl. Acad. Sci. USA* 81:1624-1628.

Soave, C. and Salamini, F. (1984). Organization and regulation of zein genes in maize endosperm. *Phil. Trans. R. Soc. Lond.* B304: 341-347.

Sørensen, M.B., Müller, M., Skerritt, J. and Simpson, D. (1996). Hordein promoter methylation and transcriptional activity in wild-type and mutant barley endosperm. *Mol. Gen. Genet.* 250: 750-760.

Takatsuji, H., Mori, M., Benfey, P.N., Ren, L. and Chua, N.-H. (1992). Characterization of a zinc finger DNA-binding protein expressed specifically in Petunia petals and seedlings. *EMBO J.* 11: 241-249.

Targon, M.L.P.N., Ottoboni, L.M.M., Leite, A., Ludevid, D., Puigdomenech, P. and Arruda, P. (1992). Synthesis and deposition of coixin in seeds of *Coix lacryma-jobi*. *Plant Sci.* 83: 169-180.

Tjian, R. (1995). Molecular machines that control genes. *Scientific American* 272: 38-45.

Tjian, R. and Maniatis, T. (1994). Transcriptional activation: a complex puzzle with few easy pieces. *Cell* 77: 5-8.

Ueda, T., Waverczak, W., Ward, K., Sher, N., Ketudat, M., Schmidt, R.J. and Messing, J. (1992). Mutations of the 22- and 27-kD zein promoters affect transactivation by the Opaque-2 protein. *Plant Cell* 4: 701-709.

Vanyushin, B.F. and Kirnos, M.D. (1988). DNA methylation in plants. *Gene* 74: 117-121.

Vettore, A.L. (1994). Clonagem e caracterização do fator de transativação Opaco2 de *Coix lacryma-jobi*. Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas para obtenção do título de Mestre em Ciências.

Vinson, C.R., Sigler, P.B. and McKnight, S.L. (1989). Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* 246: 911-916.

von Hippel, P.H. and Berg, O.G. (1986). On the specificity of DNA-protein interactions. *Proc. Natl. Acad. Sci. USA* 83: 1608-1612.

Wagner, S. and Green, M.R. (1994). DNA-binding domains: targets for viral and cellular regulators. *Curr. Opin. Cell Biol.* 6: 410-414.

Weising, K., Schell, J. and Kahl, G. (1988). Foreign genes in plants: transfer, structure, expression, and applications. *Ann. Rev. Genet.* 22: 421-477.

Williams, M.E., Foster, R. and Chua, N.-H. (1992). Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding. *Plant Cell* 4: 485-496.

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.

Yunes, J.A., Cord Neto, G., Leite, A., Ottoboni, L.M.M. and Arruda, P. (1994a). The role of the Opaque2 transcriptional factor in the regulation of protein accumulation and amino acid metabolism in maize seeds. *An. Acad. Bras. Ci.* 66, 227-238.

Yunes, J.A., Cord Neto, G., Silva, M.J., Leite, A., Ottoboni, L.M.M. and Arruda, P. (1994b). The transcriptional activator Opaque2 recognizes two different target sequences in the 22-kD-like α-prolamin genes. *Plant Cell,* 6, 237-249.