ACGT and vicilin core sequences in a promoter domain required for seed-specific expression of a 2S storage protein gene are recognized by the opaque-2 regulatory protein

M. Vincentz[1,5], A. Leite[3], G. Neshich[1], G. Vriend[4], C. Mattar[1], L. Barros[2], D. Weinberg[1], E.R. de Almeida[1], M. Paes de Carvalho[1], F. Aragão[2] and E.S. Gander[1,*]
[1]*Laboratório de Biologia Molecular and* [2]*Laboratório de Biologia Celular, Centro Nacional de Recursos Genéticos e Biotecnologia, SAIN, Parque Rural, CP 02372 CEP 70.770, Brasilia-DF, Brazil (*author for correspondence);* [3]*Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, 1308, Campinas, SP, Brazil;* [4]*EMBL, Meyerhoftrasse 1, Postfach 10 22 09, 69012 Heidelberg, Germany;* [5]*present address: Departamento de Genética, ESALQ-USP, Av. Pádua Dias 11, CEP 13400 970, Piracicaba, SP, Brazil*

Abstract

The expression of Brazil nut storage albumin genes is highly regulated during seed development. Several sequences in the promoter of one of these genes show homologies with the target sites of the maize O2 bZIP regulatory protein. We therefore asked whether the O2 protein would recognize these promoter sequences. We show that the O2 protein binds to three different sequences (F1, F2 and F3). F1 and F3 are hybrid C/G and A/G boxes, respectively, that are homologous to the O2-binding site of a maize α-zein gene. F2 is a new O2-binding sequence related to the O2 target sites of the *Coix* α-coxin, the maize *b-32* genes and the AP-1 pseudopalindrome. Molecular modelling showed that an Asn and a Ser in the 02 DNA binding domain make different base-specific contacts with each operator. 5' Promoter deletions of the be2S1 gene showed that the domain containing the O2 target sites F1 and F2 is required for detectable reporter gene expression in transgenic tobacco seeds. Moreover, the homologous *coix* O2 protein was shown to *in situ* transactivate the promoter region encompassing the three O2-binding sites F1, F2 and F3. Thus, these sites may be *in vivo* regulatory sequences mediating activation by bZIP regulatory proteins.

Introduction

Seeds of *Bertholletia excelsa* (Brazil nut tree) accumulate a methionine-rich (17%) 2S storage albumin which is encoded by a multigene family. Synthesis of the 2S albumin was shown to be restricted to the latest stages of seed development and was correlated with the accumulation of the corresponding mRNA [20].

The expression of seed storage protein genes is specifically induced during seed development and restricted to the embryo and/or the endosperm [22]. Thus, these gene represent an interesting model for

studying the mechanisms of tissue- and development-specific expression. Regulatory domains required to establish the specific expression pattern of these genes were identified in their promoter by deletion and gain of function experiments in transgenic plants [12, 16, 28, 33, 46]. For instance, the bean β-phaseolin promoter was shown to contain several activating and inhibitory regulatory domains. The interplay of these regulatory elements confers temporal and organ- and/or tissue-specific expression in transgenic tobacco seeds [8, 9].

One well described regulatory protein known to induce expression of seed storage protein genes is Opaque-2 in maize (O2). The regulatory *locus opaque-2 (o2)* promotes endosperm-specific expression of the 22 kDa and the 19 kDa α-zein storage protein genes [31] and the *b-32* gene [18]. The *o2 locus* encodes a

---

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X78286 (*Coix* 02 cDNA) and X78287 (*Coix* 02 gene).

bZIP transcriptional activator [35, 42, 48, 54] which binds to a hybrid G/A box (TCCACGTAGA) in the promoter of a 22 kDa α-zein gene [43]. Palindromic DNA bZIP protein binding sites based on a ACGT core flanked in 3′ by a guanine, adenine, cytidine or thymine (ACGTG/A/C/T) are called G, A, C and T boxes, respectively [26].

The O2 binding motif in the promoters of the *b-32* and the 22 kDa α-like coixin genes of the maize-related grass *Coix* is different from the α-zein target sequence. It was suggested to be the GACATGTC palindrome [54]. More recently O2 was shown to bind also to the pseudopalindrome AP-1 site (ATGACTCAT) of the pea lectin promoter [15]. The basis of this broad DNA-binding specificity remains poorly understood however. The DNA-binding part of bZIP proteins consists of a basic region that contacts DNA and a leucine zipper which causes dimerisation. The X-ray structure of the GCN4 bZIP motif complexed to DNA target sites (AP-1 pseudopalindrome ATGACTCAT [19] and ATF/CREB palindrome ATGACGTCAT [32]) showed that bZIP is completely α-helical. The two leucine zippers are packed in a coiled-coil region for dimerization and the basic regions of the dimer fits into the major groove of the half-sites of the target DNA. Five amino acids of the basic region make direct contacts with four bases of the half-site [19, 32]. These residues were shown to be critical for DNA-binding specificity [30, 45, 47].

As part of a project aimed at understanding seed-specific transcriptional control, we studied the promoter of the 2S albumin gene be2S1 of Brazil nut tree. Within this promoter, nine motifs based on the ACGT core sequence were identified. These motifs were tested for O2 binding in order to identify putative regulatory sequences. Three different sequences were recognized by O2. Two target sites contain an ACGT core (F1, TCCACGTCGA; F3, TCCACGTACT) and the third, F2 (GCCACCTCAT), is part of the large vicilin box [21]. We show that a promoter region containing F1 and F2 is required for detectable seed-specific expression in transgenic tobacco. We also show that the promoter domain containing the three O2 target sites is transactivated by the *Coix* O2 protein. Finally, the three O2 F1/F2/F3 complexes were modelled to try to understand O2 binding specificity. The models indicate that within the O2 α-helical basic domain, the invariable Asn of bZIP proteins [19] and a highly conserved Ser in plant bZIP proteins [25] make different contacts with different bases in the operator DNAs. We propose

that these two residues are probably responsible for part of the observed differences in DNA binding.

Materials and methods

Constructs

All DNA manipulation were performed according to standard protocols [39]. 5′ Deletions of the be2S1 promoter were generated by nuclease BAL-31 digestion and the end-points of the deletions determined by dideoxy sequencing. Deletion mutants containing the 55 bp complete untranslated leader sequence were cloned as transcriptional fusions at the ATG initiation codon of the CAT gene plus the poly(A) signals of the octopine synthase gene (*ocs*) of *Agrobacterium tumefaciens* into pGEM7 (Promega). The constructs were cloned as *Sph*I Klenow-repaired *Bgl*II or *Bam*HI fragments into the *Xba*I Klenow-repaired/*Bgl*II or *Bgl*II sites of the binary plant transformation vector pGSC1703A, respectively [13]. The cassette with the CAT gene under the control of the CaMV 35S promoter and the 3′ *ocs* poly(A) signal (Plant Genetic Systems) was cloned into the pGSC1703A vector for plant transformation. The strategy to obtain the 210S1GUS construct (Figure 6) was as follows: the $-210/-34$ promoter fragment (without TATA box) was obtained by PCR amplification and was cloned as an *Eco*RI fragment into Bluescript KS$^+$ (Stratagene). This clone was digested with *Eco*RV, *Bam*HI linkers were ligated and subsequently digested by *Bam*HI to liberate the $-210/-34$ fragment. This fragment was cloned as a dimer into the *Bam*HI site in front of the -46 core promoter of the CaMV 35S promoter fused to GUS [4]. The orientation was determined by dideoxy sequencing. pTJEN used for footprint analysis was obtained by subcloning an *Eco*RI-*Nco*I Klenow-repaired fragment (-210 to $+1$ of the be2S1 promoter) into the *Eco*RI/*Eco*RV sites of KS$^+$. pTJ2 is a *Hin*dIII-*Pst*I fragment of the be2S1 gene cloned into pUC19. The P285α-CGUS construct was described previously [54].

Plant transformation, CAT activity assay

Tobacco (cv. Xanthi XHFD8) transformation was via *Agrobacterium tumefaciens* C58C1 [17]. Plant regeneration and selection of the transformants with kanamycin was as previously described [24]. Extraction of total proteins from mature seeds (50 mg) after

grinding in liquid nitrogen and the CAT assay and CAT activity quantification were done as described [38]. Protein concentrations were determined by the Bradford assay [6].

Transient assay

Leaves of tobacco plants which were grown under shade conditions in a green house were surface-sterilized by rapid immersion in a 50% ethanol solution followed by a 20 min incubation in a 1% hypochlorite solution and finally rinsed three times with sterile distilled water. Discs 2 cm in diameter that were punched out of these sterilized leaves were used for bombardment. A 0.5 μg portion of reporter plasmid alone or together with 1 μg of the effector plasmid was bound to 1.2 μm diameter tungsten particles (M10, Sylvania) as described [3, 49]. Bombardment was conducted utilizing a high-pressure helium-driven particle acceleration device [3, 40, 49]. Helium pressure used was 4.1 MPa and other conditions for bombardment were as described previously [3]. Bombarded leaf discs were kept on solid MS medium [37] supplemented with 1% sucrose for 24 h. Histochemical assay of GUS was then performed overnight according to established procedures [27]. After chlorophyll extraction with ethanol, blue spots were counted under a microscope.

DNA-binding assay and DNAse I footprint experiments

The DNA-binding assay and DNAse I footprint experiment were performed with a β-galactosidase-maize O2 (β-Gal::O2) fusion protein [42]. The fusion protein was immunoprecipitated with a β-galactosidase antibody and this β-Gal::O2 immunocomplex was then treated with protein A-Sepharose beads for further use [42]. For the DNA-binding assay the DNA restriction fragments from vector pTJEN (previous section) were 5′ end-labelled using the Klenow fragment. The ^{32}P-end-labelled fragments were then incubated with 15 μl of protein A-Sepharose beads containing the β-Gal::O2 immunocomplex for 90 min at room temperature with agitation. Washes, processing and agarose gel electrophoresis of the selected fragments was as described earlier [42, 54]. For DNase I footprinting, the *Eco*RI-*Xho*I $-210/+1$ promoter fragment of pTJEN (previous section) was used. The upper strand was labelled at the *Xho*I site and the lower strand at the *Eco*RI site using the Klenow fragment. 35 000 cpm of end-labelled DNA fragments were incubated with 0,
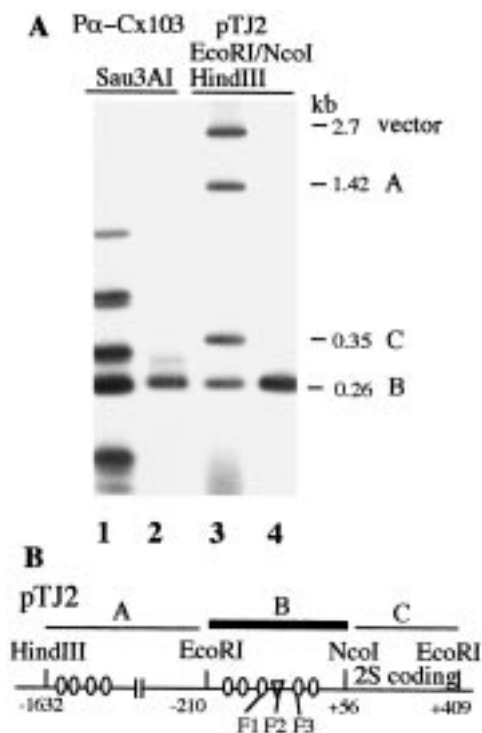

Figure 1. Selective binding of the O2 protein to the -210 to $+56$ promoter fragment. A. Plasmids containing the be2S1 entire promoter and part of the coding sequence (pTJ2) or the α-coixin $\alpha - 3B$ promoter (Pα-Cx103) were digested by the restriction enzymes indicated and were end-labelled (lane 1 and 3). Lanes 2 and 2 show the fragments specifically selected by the β-Gal::O2 fusion protein. The 2.7 kb fragment is the pUC vector. A, B, C are the restriction fragments shown in B. B. Map position of the restriction enzymes used to digest pTJ2. The solid line above the map represents the fragment selected by the β-Gal::O2 fusion protein. Positions are given with respect to the transcription start. The *Nco*I site covers the start codon. Ovals represent ACGT core sequences and the footprinted sequences F1, F2 and F3 shown in Figure 2 are indicated.

6, 12 and 24 μl of the protein A-Sepharose beads containing the β-Gal::O2 immunocomplex. Incubation, DNAse I digestion and sequencing gel analysis was as described previously [54].

Molecular modelling

The Opaque-2 protein [42] was modelled based on the GCN4 structure [32]. Modelling was done as described [51] using the program WHAT IF [50]. Side-chains that had to be modelled were appended to the template main chain in the most preferred conformations [14], as obtained from the WHAT IF relational database. DNAs were modelled using the coordinates of the GCN4 bound to the ATF/CREB DNA site in the
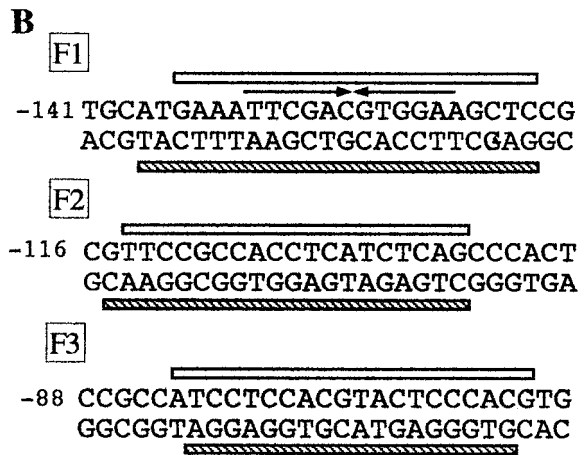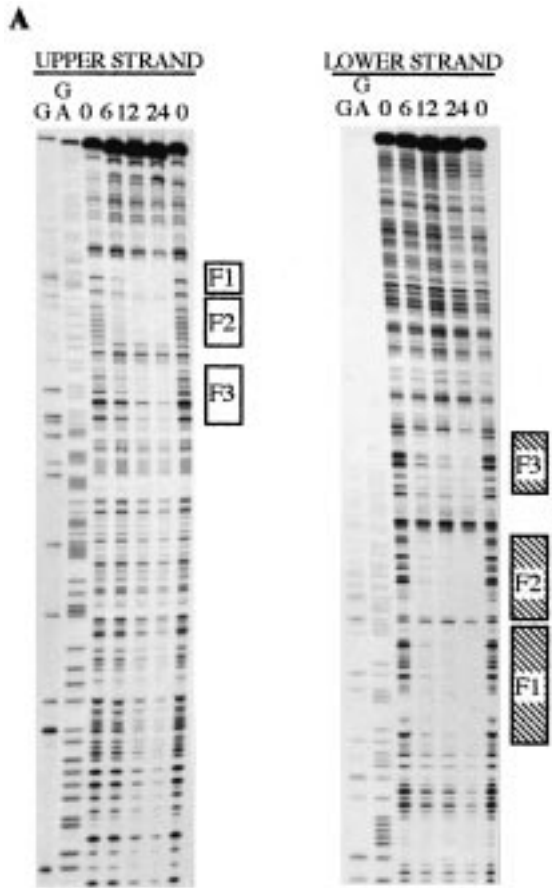
A

UPPER STRAND

G
GA 0 6 12 24 0

LOWER STRAND

G
GA 0 6 12 24 0

F1
F2
F3

F3
F2
F1

B

F1

−141 TGCATGAAATTCGACGTGGAAGCTCCG
ACGTACTTTAAGCTGCACCTTCGAGGC

F2

−116 CGTTCCGCCACCTCATCTCAGCCCACT
GCAAGGCGGTGGAGTAGAGTCGGGTGA

F3

−88 CCGCCATCCTCCACGTACTCCCACGTG
GGCGGTAGGAGGTGCATGAGGGTGCAC

Figure 2. Localization of O2 fusion protein binding sites in the −210 to +56 promoter fragment by footprinting analysis. A. DNAse I protection of the −210/ + 1 promoter domain by the β-Gal::O2 fusion protein. Both strands were 3′ end-labelled and incubated with 0, 6, 12 and 24 μl (lanes 0, 6, 12 and 24, respectively) of immunopurified O2 fusion protein prior to DNAse I digestion. Lanes G and GA are the Maxam and Gilbert G and G+A sequencing ladder generated on the labelled fragments. The footprints obtained are shown by boxes. B. Nucleotide sequence of the three footprinted boxes F1, F2 and F3 shown in A. Bars show the protected sequences identified in A. Arrows indicate an almost perfect palindrome.
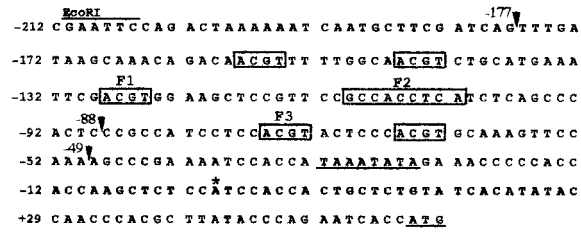

Figure 3. Nucleotide sequence of the be2S1 −210 to +56 promoter fragment selected by the O2 fusion protein. The TATA box and the start codon are underlined. The star indicates the transcription start. The ACGT cores, and the three footprints F1, F2 and F3 are boxed. The vertical arrows indicate the 5′ end points of some of the promoter deletions analyzed in Figure 7.

crystal structure [32]. The phosphate backbone and the ribose rings were kept constant in the model. The bases were placed in the plantes of the template bases. Conserved bases were not modified in the model. All three models were subjected to 500 steps of energy minimization, a very short molecular dynamics run to get out local energy minima of high energy, and a long energy minimization using the Xplor program [7]. Molecular dynamics runs consisted of 50 steps of 2 fs, and were performed at 100 K in vacuum. The RMS deviation between the initial and the refined models typically was about 0.11 nm.

Results

The maize bZIP regulatory protein Opaque-2 binds to the promoter of the be2S1 2S albumin gene

We identified within the be2S1 promoter nine sequences containing the ACGT core (Figure 1B). One of these sequences (TTCCACGTCGA) is almost identical to the O2 protein target sequence (TTCCACGTA-GA) of the maize α-zein gene [43, 48]. We did a DNA binding selection assay using a β-Gal::O2 fusion protein produced in *Escherichia coli* [42, 54] to determine whether O2 would bind to the be2S1 promoter. As shown in Figure 1A, only the restriction fragment B, which encompasses five ACGT core sequences (Figure 1B), was specifically bound by the O2 fusion protein. The promoter of the α-coixin gene (α-Cx103) which is recognized by O2 [54], was used as a control (Figure 1A).

We performed a DNAse I protection experiment to determine the precise sequence(s) recognized by the O2 fusion protein. Two strong footprints F1 and F2,

and a weaker one, F3, were identified (Figures 2A and 1B). The exact localization of these binding sites within the promoter domain recognized by O2 is shown in Figure 3. The sequence F1 contains a hybrid C/G-box, **TTCCACGT**CGAA (Figure 2B), which is almost a palindrome and is nearly identical to the O2 binding site of a α-zein gene *22Z-4* (Table 1). The sequence **TCCACGT** of the F1 C/G box was also found in the weaker footprint F3 (Figure 2B and Table 1). Interestingly, the footprint F2 does not contain an ACGT core but encompasses the sequence CGCCACCTCA (Figure 2B) which is part of the large vicilin box found in 7S storage globulins from legumes [21]. We will refer to the F2 sequence GCCAC**TCAT** as the 'vicilin core'. The motif **TCAT** in the 'vicilin core' is identical to the half-site of the AP-1 pseudopalindrome ATGAC**TCAT** (Table 1). It has been shown that O2 binds *in vitro* to the AP-1 site present in a pea lectin promoter [15]. The TCAT sequence is also present in the O2 target site identified in the promoters of the α-3B coixin gene [54] and the maize *b-32* gene [35]. Therefore, it seems likely that the TCAT AP-1 half-site found in F2 is part of the recognition sequence of the O2 fusion protein. On the other hand, the GCCAC motif in the F2 'vicilin core' is a G-box half-site which was shown to be an optimal O2 binding site [26]. Thus, we suggest that the F2 O2 target site is a hybrid C/G box centred on an ACCT core. Based on the sequence alignment shown in Table 1, we further propose that the O2 protein can bind to hybrid G/C boxes with one mutation in the central dinucleotide of the ACGT core (ACCT, ATGT and ACAT), thus forming at least one 'optimal' (GCCAC and GTCAT) or 'good' (TCCAC) O2 binding half-site [26]. In addition, as shown by the example of box B2 in the b-32 gene, the mutation of the central CG dinucleotide to TA (Table 1) does not impair O2 binding.

Molecular modelling

In order to better understand the basis of the broad O2 DNA-binding specificity, models were built of the complexes of O2 with the three operator DNAs F1, F2 and F3. An overview of an O2-DNA complex is shown in Figure 4. Binding of bZIP proteins to DNA involves the entire scale of non-covalent interactions which makes the task of dissecting the importance of individual contacts very complex. Additionally, the α-helical structure of the bZIP domain is sufficiently flexible to allow it to adopt slightly different conformations on different operator sequences [1, 30]. The conformation and the stability of each complex results from the

adjustment required to optimize the protein-DNA contracts [1, 29, 30]. These small differences between the bZIP – DNA complexes could constitute a signal to the transcriptional apparatus [29]. This concept of flexibility must also apply to the O2 bZIP protein since the same protein binds well to DNA sequences that have different residues in contact with the bZIP. A summary of the results of our modelling assays is shown in Figure 5. Several differences between the complexes can be observed. For instance, Arg-72 forms a hydrogen bond only with the F1 operator base T (W+1). Other, smaller variations are also observed. For example, many conserved hydrogen bonds have slightly different geometries in the different models. However, a consistent picture emerges in which two residues, Asn-9/64 and Ser-12/67, in both helices are the main determinants for recognition. Only if the base (C-4/W-4) is a thymidine can a favourable hydrophobic interaction between its carbon C5M and the C-β of Ser-12/67 be observed. Asn-9/64 shows a wide range of possible interactions with DNA that have been observed in the GCN4 structures and in our models (Figure 5). In all cases the bases presented to Asn are different and lead to differences in hydrogen bond formation. These results point to the importance of Asn and Ser as the determinant of the binding constant.

Biological function of the O2-binding sites present in the be2S1 promoter

We investigated the biological significance of the DNAse I protection experiments described above. The transactivation of the be2S1 promoter by the *Coix* O2 protein was evaluated using an *in situ* transient expression assay. Since the *Coix* and the maize O2 sequences are a 100% identical in the basic region and are 93% identical in the leucine zipper, it seems likely that the two proteins have very similar binding properties.

The effector plasmid which contains the *Coix* O2 cDNA under control of the CaMV 35S promoter (construct A in Figure 6) and one reporter gene were bombarded together on tobacco leaves. The reporter constructs consist of the *uidA* gene, which encodes the enzyme β-glucuronidase (GUS) under control of different promoters (Figure 6). Construct B consists of the be2S1 $-210/-34$ promoter domain which contains the three O2 binding sites (F1, F2, F3), duplicated and joined to the TATA box of the -46 minimal promoter of the CaMV 35S. We chose to duplicate the be2S1 promoter domain because it was shown in several instances that the combination or the duplication

Table 1. Alignment of O2-binding sites.

Box	Gene	Sequence	O2 binding	Ref.
Positions		-4 -3 -2 -1 -0 +0 +1 +2 +3 +4		
B4$_{(R)}$	b-32	A G T C A T G T C A T C	+	1
B3$_{(R)}$	b-32	T T C C A C A T C A T C	+	1
B2$_R$	b-32	A T C C A T A T C A T C	+	1
B1$_R$	b-32	A G C C A T G T C A T C	+	1
O2C	α-3A	A G C C A T G T C A T C	+	2
O2C	α-3B	A G A C A T G T C A T C	+	2
F2	BE2S1	C G C C A C C T C A T C	+	3
Vic.	phaseolin	C G C C A C C T C A A T	nd	4
AP-1	lectin	G A A T G A G T C A T C	+	5
With ACGT core				
F1(R)	BE2S1	T T C C A C G T C G A A	+	3
F3	BE2SI	C T C C A C G T A C T C	+	3
02	22Z-4	T T C C A C G T A G A T	+	6

References: 1, maize *b-32* gene, [35]; 2, coix α-coixin 3A and 3B [54]; 3, this study; 4, vicilin (vic) box of the bean β-phaseolin gene [21]; 5, pea lectin [15]; 6, maize α-zein *22Z-4* [43]. (R), reverse orientation; nd, not determined.
Numbers at the top indicate the position of the bases from the centre of the binding site $(-0, +0)$.

of *cis*-regulatory elements can interact synergistically (i.e. more than additive) [5, 34, 52]. Construct C is the α-3B coixin promoter and D consists of the -46 minimal promoter of the CaMV 35S. Transactivation of the reporter plasmids by O2 was measured by histochemical staining of the product of GUS and estimated by counting the blue spots. Table 2 shows that the be2S1 proximal promoter is activated 7-fold by the *Coix* O2 protein. The promoter of the *Coix* α-3B coixin gene was used here as a positive control since it has been shown to be transactivated by the maize O2 protein [54]. In our transient assay this promoter domain was activated 2.5-fold (Table 2). The low activation of the be2S1 and the α-3B coixin promoter constructs in the absence of the O2 protein may reflect the binding of leave transcriptional factors to these promoter fragments (Table 2). As expected, no activation by O2 was observed from the -46 CaMV promoter which is devoided of regulatory sequences other than the TATA box (Table 2). Taken together, these results confirm that the *Coix* O2 protein is a transcriptional activator, and that the *in vitro* O2-binding sites F1, F2 and F3 are likely to be functional regulatory sequences *in vivo*.

We then verified the involvement of the F1 and F2 O2 target sites in seed expression. 5′ deletions of

Table 2. In situ transactivation by the *Coix* O2 protein.

Plasmids	Blue spots	Activation
B	64	–
B+ A	447	7×
C	254	–
C+ A	640	2.5×
D	0	–
D+	0	–

Tobacco leaves were bombarded with particles coated with the reporter plasmid (0.5 μg) alone or together with the effector pRT101-O2 plasmid (1 μg). Gus activity was revealed by *in situ* staining and estimated by counting the blue spots. For each plasmid, the total number of blue spots from 9 independent bombardments with one leaf disc each were counted.

the promoter were cloned as transcriptional fusions in front of the chloramphenicol acetyl transferase (CAT) reporter gene. The constructs were cloned into a binary vector in order to transform tobacco using *Agrobacterium tumefaciens*. The effect of the promoter 5′ deletions on CAT activity in mature seeds of primary tobacco transformants is shown in Figure 7. For all 5′ dele-
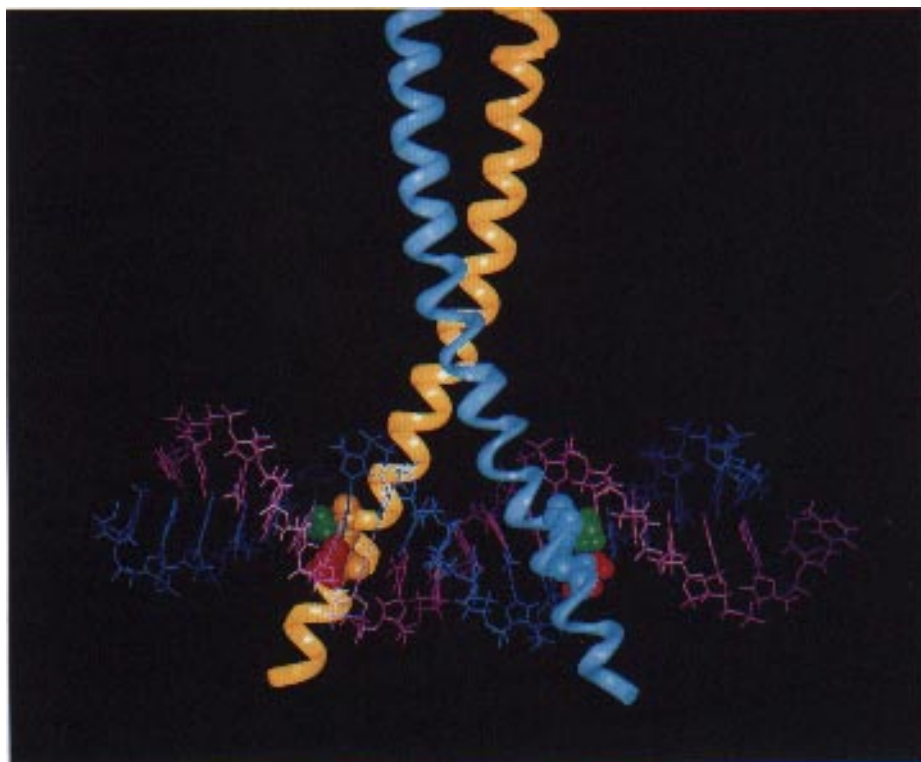
Figure 4. Model of the O2-F1 complex. Model of Opaque-2 interacting with the operator F1 of the Brazil nut 2S gene promoter. The two operator DNA strands are shown in pink and dark blue. The dimerized protein residues are presented as two α-helical ribbons (yellow and blue). The side chains of Asn-9/64 and Ser-12/67 are in red and green CPK, respectively.

tions CAT activity was exclusively detected in seed extracts, while no activity could be detected in leaves (result not shown). As shown in Figure 7, the promoter domain −177 to −88 which covers the F1 and F2 O2 targets is required for significant seed expression of the CAT gene. Indeed, promoter sequences downstream of position −88 which covers the F3 O2-binding site confer a much lower (26-fold) CAT gene expression than deletion −177 (Figure 7), but still allow a 5-fold higher CAT activity accumulation than control WT plants (not shown). These results are in agreement with previous studies which showed that the −210 to +56 promoter domain still confers seed-specific expression [23]. This data suggests that, at least in part, the F1 and F2 O2-binding sites are involved in seed expression of the be2S1 promoter.

Discussion

We have identified three different response elements (F1, F2, F3) in the promoter of the 2S albumin gene be2S1 that are recognized *in vitro* by the maize bZIP regulatory protein O2 (Figure 2B). The sequences F1 (TCC**ACGT**CGA) and F3 (TCC**ACGT**ACT) contain the **ACGT** core motif shared by all plant bZIP proteins [2, 26, 41, 53]. They are hybrid motifs consisting of a C box and G box half-site in the case of F1, and of an A box and G box half-site for F3. The F2 sequences (GCC**ACCT**CAT) can be considered as a hybrid C/G box centered on an **ACCT** core (Table 1). The O2-binding affinity for these assymetric motifs most likely is determined by the relative binding affinity for the two half-sites [26]. Nucleotides flanking the ACGT core at positions +2/−2, +3/−3 and +4/−4 were shown to affect bZIP protein-binding activity [26]. Based on the sequence alignment of Table 1 we suggest that O2 binding to operators with one or two mutations in the CG dinucleotide of the ACGT core can be effective as long as at least one half-site contains optimal flanking sequences GCC, CAT and TCC (positions +2/−2, +3/−3 and +4/−4). Based on our modelling data (Figure 5), we also suggest that the interactions of Ser-12/67 and Asn-9/64 with the
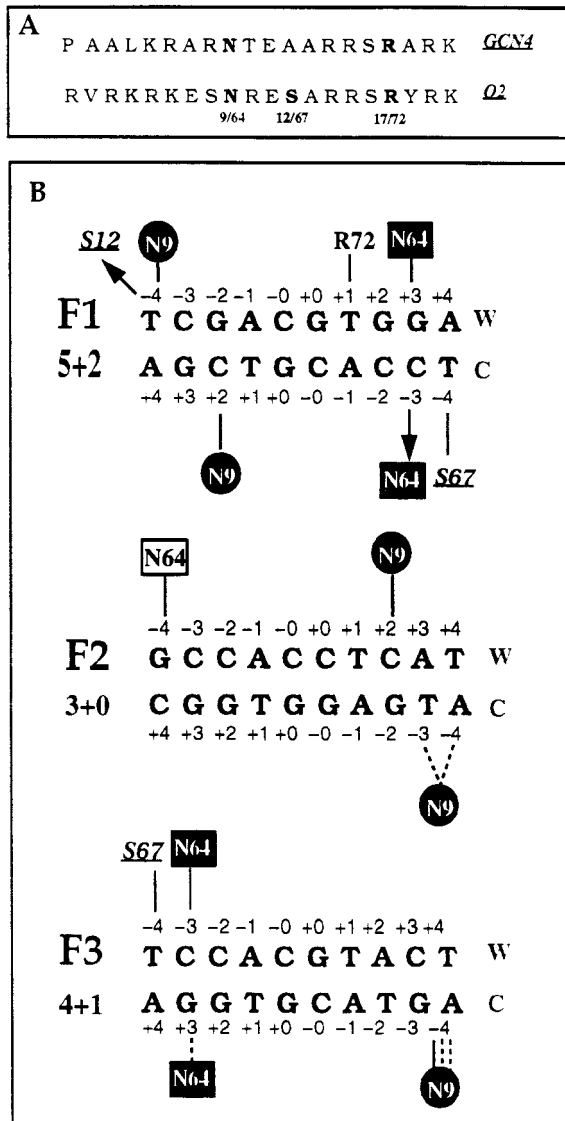
886

Figure 5. Schematic representation of the major interactions in the DNA-O2 complexes. A. Alignment of the O2 and GCN4 basic region amino acid sequence used in the modelling study. Asn-9/64 and Arg-17/72 are conserved in all known bZIP proteins [16]. Ser-12/67 is conserved in various plant bZIPs [23]. B. Major interactions in the O2-DNA complexes. Bases are numbered from the center of the binding site (−0, +0). Filled squares/circles indicate that the hydrogen bounding potential of Asn-9/64 is fully exploited (both Nδ and Oδ are involved in the hydrogen bond); open squares indicate that only one of the two possible hydrogen bonds has been established. Dotted lines indicate a 'hybrid' hydrogen bond formation: one donor and two acceptors or vice versa. 'S' pointed out by an arrow shows that a favourable hydrophobic interaction between the Ser-12/67, β-carbon and the thymidin C5M atom has been observed. The numbers below F1, F2 and F3 represent the sums of hydrogen bonds and hydrophobic interactions involving-Asn 9/64 and Ser-12/67, respectively.
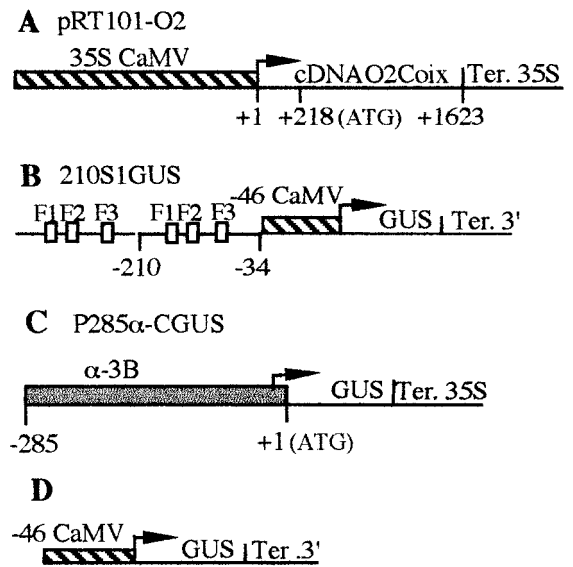

Figure 6. Constructs used for the transactivation assay. A. pRT101-O2, effector plasmid carrying the complete Coix O2 cDNA under control of the 35S CaMV promoter. B. Duplicated −210/−34 be2S1 promoter fused to the −46 CaMV TATA box. C. −285 promoter fragment of the Coix Cx103 α-coixin gene. D. The −46 CaMV promoter containing only a TATA box. Ter 3′: poly(A) signal from the pea rbcsS3C. Ter 35S: poly(A) signal of the CaMV 35S promoter.

bases (+2/−2, +3/−3 and +4/−4) flanking the core sequence are essential for successful binding. Contrary to Ser-12/67 which interacts exclusively if base −4 is a thymidine, in the case of Asn-9/64 a broader interaction spectrum is observed (Figure 5B). This amino acid can form hydrogen bonds with different bases at different positions in the flanking sequence. We suggest that small changes in the relative orientation of the bZIP helix pairs, combined with the possibility of Asn to adopt several different energetically favourable side-chain conformations [42] provides the possibility to adjust to different DNA patterns [30, 44].

The geometry and thus the energetics of the interactions made by Asn-9/64 and Ser-12/67 may be a strong determinant of the binding constants. Indeed, no quantitative correlation could be established between the apparent binding observed on the footprints (Figure 2) and the number of energetically favourable contact revealed through modelling (Figure 5). The contribution of individual contacts made by Asn-9/64 and Ser-12/67 to the binding affinity needs further studies aimed at determining the dissociation constants of O2-operator complexes. On the other hand, Asn-9/64 and Ser-12/67 are conserved in several plant bZIP-proteins [25] yet these proteins differ in their binding
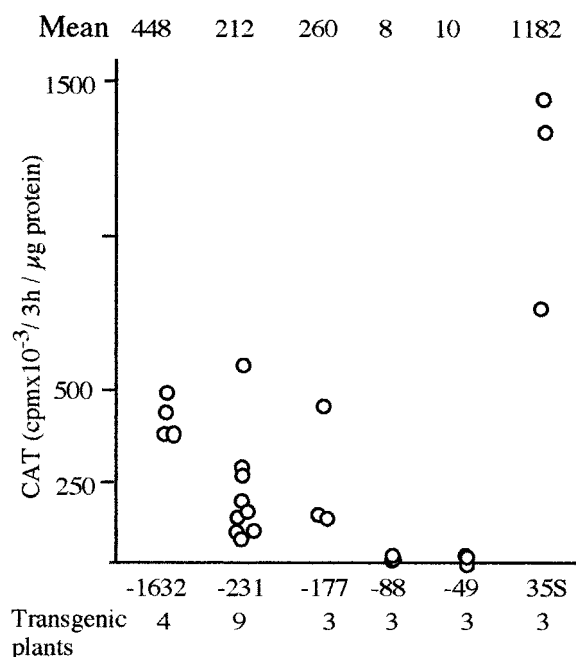
| Mean | 448 | 212 | 260 | 8 | 10 | 1182 |

Figure 7. Functional analysis of the be2S1 promoter in transgenic tobacco. 5′ Deletions of the be2S1 promoter and the CaMV 35S promoter (35S) were fused to the CAT gene at the ATG initiation codon. CAT activity in extracts of mature seeds from independent transgenic plants carrying the 5′ promoter mutants constructs and the 35S CaMV construct were determined. The 5′ deletion end-points and the number of transgenic plants analysed for each mutant are indicated. Each point represents the CAT activity from seeds of a single transgenic plant. Numbers shown above the columns are the mean values of CAT activity. The exact position of the 5′ deletions −177, −88 and −49 are shown in Figure 3.

specificity [26]. Thus, the leucine zipper and/or others non conserved amino acids in the basic domain must be involved in DNA binding specificity probably by indirectly affecting Asn-9/64 and Ser-12/67 [30, 45]. *In vivo*, binding specificity may also be influenced by the interaction with other DNA-binding protein [36].

The sequence GCCACCTCA, which is part of the vicilin core (F2), is also found in the UAS1 activating sequence of the β-phaseolin promoter [9] and in the activating domain c of the promoter of the β-conglycinin α' subunit gene [33]. Our results suggest that the vicilin core may be the target of a bZIP transcription factor and are in agreement with the recent identification of two bean bZIP factors (ROM1 and ROM2) that specifically bind to the same vicilin core sequence [10, 11]. The *in situ* transactivation of the −210/ − 34 promoter domain of be2S1 by the *Coix* O2 protein (Table 2) suggests that the *in vitro* binding of the maize O2 to F1, F2 and F3 sequences reflects an *in vivo*

situation. The importance of the sequences F1 and F2 for seed expression was confirmed by the be2S1 5′ promoter deletion analysis in transgenic tobacco (Figure 7). However, the exact function of the F1, F2 and F3 sequences for seed expression remains to be elucidated. Our data suggest that bZIP proteins are involved in dicot seed gene expression and work is in progress to identify such proteins.

Acknowledgements

References

1. Alber T: How GCN4 binds DNA. Curr Biol 3: 182–184 (1993).
2. Armstrong GA, Weisshaar B, Hahlbrock K: Homodimeric and heterodimeric leucin zipper proteins and parsley nuclear factors recognize diverse promoter elements with ACGT cores. Plant Cell 4: 525–537 (1992).
3. Aragão FJL, Barros LMG, Brasileiro AC, Ribeiro SG, Smith FD, Sanford JC, Faria JC, Rech EL: Inheritance of foreign genes in transgenic bean (*Phaseolus vulgaris* L.) cotransformed via particle bombardment. Theor Appl Genet 93: 142–150 (1996).
4. Benfey P, Chua NH: Tissue-specific expression from CaMV enhancer subdomains in early stages of plant development. EMBO J 9: 1677–1684 (1990).
5. Benfey PN, Ren L, Chua N-H: Combinatorial and synergistic properties of CaMV 35S enhancer subdomains. EMBO J 9: 1685–1696 (1990).
6. Bradford MM: A rapid and sensitive method for the quantitation of protein-dye binding. Anal Biochem 72: 248–254 (1976).
7. Brünger AT: X-FLOR: a system for X-ray crystallography and NMR. Yale University Press, New Haven/London (1987).
8. Burow MD, Sen P, Chlan CA, Murai N: Developmental control of the β-phaseolin gene requires positive, negative and temporal seed-specific transcriptional regulatory elements and

a negative element for stem and root expression. Plant J 2: 537–548 (1992).

9. Bustos M, Begum D, Kalkan F, Battraw M, Hall TC: Positive and negative *cis*-acting DNA domains are required for spatial and temporal regulation of gene expression by a seed storage protein promoter. EMBO J 10: 1469–1479 (1991).

10. Chern M-S, Bobb AJ, Bustos M: The regulator of MAT2 (ROM2) protein binds to early maturation promoters and represses PvALF-activated transcription. Plant Cell 8: 305–321 (1996).

11. Chern M-S, Eiben HG, Bustos MM: The developmentally regulated bZIP factor ROM1 modulates transcription from lectin and storage protein genes in bean embryos. Plant J 10: 135–148 (1996).

12. da Silva Conceição A, Krebbers E: A cotyledon regulatory region is responsible for the different spatial expression patterns of *Arabidopsis* 2S albumin gene. Plant J 5: 493–505 (1994).

13. De Clercq A, Vaudewiele M, De Rycke R, Van Damme J, Van Montagu M, Krebbers E, Vandekerckhove J: Expression and processing of an *Arabidopsis* 2S albumin in transgenic tobacco. Plant Physiol 92: 899–907 (1990).

14. De Fillipis V, Sander C, Vriend G: Predicting local structural changes that result from point mutations. Prot Eng 7: 1203–1208 (1994).

15. de Pater S, Katagiri F, Kijne F, Chua NH: 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 (1994).

16. de Pater S, Pham K, Chua NH, Memelink J, Kijne J: A 22-bp fragment of the pea lectin promoter containing essential TGAC-like motifs confers seed-specific gene expression. Plant Cell 5: 877–886 (1993).

17. Deblaere R, Bytbier B, De Greve H, Deboeck F, Schell J, Van Montagu M, Leemans J: Efficient octopin Ti-derived vectors for *Agrobacterium* mediated gene transfer to plants. Nucl Acids Res 13: 4777–4788 (1985).

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

19. Ellenberger TE, Brandl CJ, Struhl K, Harrison ST: The GCN4 basic region leucin zipper binds DNA as a dimer of uninterrupted helices: crystal structure of the protein-DNA complex. Cell 71: 1223–1237 (1992).

20. Gander ES, Holmstroem GR, De Paiva GR, De Castro LAB, Carneiro M, Grossi de Sá MF: Isolation, characterization and expression of a gene coding for a 2S albumin from *Bertholletia excelsa* (Brazil nut). Plant Mol Biol 16: 437–448 (1991).

21. Gatehouse JA, Evans IM, Croy R, Boutler D: Differential expression of genes during legume seed development. Phil Trans R Soc Lond B314: 367–384 (1986).

22. Goldberg RB, Barker SJ, Perez-Grau L: Regulation of gene expression during plant embryogenesis. Cell 56: 149–160 (1989).

23. Grossi de Sá MF, Weinberg DF, Rech EL, Barros LMG, Aragão F, Gander E: Functional studies of a seed-specific promoter from Brazil nut 2S gene. Plant Sci 103: 189–198 (1994).

24. Horsch RB, Fry JE, Offman NL, Eichholtz D, Rogers SG, Fraley R: A simple method for transferring genes into plants. Science 227: 1229–1231 (1985).

25. Hurst H: Transcription factor 1: bZIP proteins. Protein Profile 2: 105–168 (1995).

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

27. Jefferson RA: Assaying chimaeric genes in plants: the GUS gene fusion system. Plant Mol Biol Rep 5: 387–405 (1987).

28. Kawagoe Y, Campell RR, Murai N: Synergism between CACGTG (G-box) and CACCTG *cis*-element is required for activation of the bean seed storage protein β-phaseolin gene. Plant J 5: 885–890 (1994).

29. Kerpola TA, Curran T: Fos-Jun heterodimers bend DNA in opposite orientations: implication for transcription factor cooperativity. Cell 66: 317–326 (1991).

30. Kim J, Tzamarias D, Ellenberger T, Harrison ST, Struhl K: Adaptability at the protein-DNA interface is an important aspect of sequence recognition by bZIP proteins. Proc Natl Acad Sci USA 90: 4513–4517 (1993).

31. Kodrzycki R, Boston RS, Larkins BA: The opque-2 mutation of maize differentially reduced zein gene transcription. Plant Cell 1: 105–114 (1989).

32. König P, Richmond TJ: The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility. J Mol Biol 233: 139–154 (1993).

33. Lessard PA, Allen RD, Fujiwara T, Beachy RN: Upstream regulatory sequences from two β-conglycinin gene. Plant Mol Biol 22: 873–885 (1993).

34. Lin Y-S, Carey M, Ptashne M, Green MR: How different eukaryotic transcriptional activators can cooperate promiscuously. Nature 345: 359–361 (1990).

35. Lohmer S, Maddaloni M, Motto M, Di Fonzo N, Hartings H, Salamini F, Thompson RD: 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 (1991).

36. Menskens AE, Schindler U, Cashmore AR: The G-box: a ubiquitous regulatory DNA element in plants bound by the GBF family of bZIP proteins. Trends Biochem Sci 20: 506–510 (1995).

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

38. Pröls M, Töfer R, Schell J, Steinbiss HH: Transient gene expression in tobacco protoplasts: time course of CAT appearance. Plant Cell Rep 11: 323–328 (1988).

39. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).

40. Sanford JC, Devit MJ, Russel JA, Smith FD, Harpending PR, Roy MK, Johnston SA: An improved helium-driven biolistic device. Technique 1: 3–16 (1991).

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

42. Schmidt RJ, Burr FA, Aukerman MJ, Burr B: Maize regulatory opaque-2 encodes a protein with a 'leucine-zipper' motif that binds to zein DNA. Proc Natl Acad Sci USA 87: 46–50 (1990).

43. Schmidt RJ, Ketudat M, Aukerman MJ, Hoschek G: Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-KD zein genes. Plant Cell 4: 689–700 (1992).

44. Smith DL, Johnson AD: Operator-constitutive mutations in a DNA sequence recognized by a yeast homeodomain. EMBO J 13: 2378–2387 (1994).

45. Suckow M, von Wilcken-Bergmann B, Müller-Hill B: Identification of three residues in the basic regions of the bZIP proteins GCN4, C/EBP and TAF-1 that are involved in specific DNA binding. EMBO J 12: 1193–1200 (1993).

46. Thomas MS, Flavell RB: Identification of an enhancer element for the endosperm-specific expression of high molecular weight glutenin. Plant Cell 2: 1171–1180 (1990).

47. Tzamarias D, Pu WT, Struhl K: Mutation in the bZIP domain of yeast GCN4 that alter DNA-binding specificity. Proc Natl Acad Sci USA 89: 2007–2011 (1992).

48. Udea T, Waverczak W, Ward K, Sher N, Ketudat M, Schmidt RJ, Messing J: Mutation in the 22 and 27-kD zein promoters affect transactivation by the Opaque2 protein. Plant Cell 4: 701–709 (1992).

49. Vainstein MH, Alves SA, de Lima BD, Aragão F, Rech E: Stable DNA transfection in a flagellate trypanosomatid by microparticle bombardment. Nucl Acids Res 22: 3263–3264 (1994).

50. Vriend G: WHAT IF: a molecular modeling and drug design program. J Mol Graph 8: 52–56 (1990).

51. Vriend G, Eijsink VGH: Prediction and analysis of structure, stability and unfolding of thermolysin-like proteases. J Comp-Aided Mol Design 367–396 (1993).

52. Weisshaar B, Armstrong GA, Block A, de Costa e Silva G, Hahlbrock K: Light-inducible and constitutively expressed DNA-binding proteins recognizing a plant promoter element with functional relevance in light responsiveness. EMBO J 10: 1777–1786 (1991).

53. Williams ME, Foster R, Chua NH: Sequences flanking the hexameric G-box core CACGTG affect the specificity of protein binding. Plant Cell 4: 485–496 (1992).

54. Yunes JA, Cordo Neto G, da Silva MJ, Leite A, Ottoboni LM, Arruda P: The transcriptional activator Opaque2 recognizes two different target sequences in the 22 kD-like a-prolamin genes. Plant Cell 6: 237–249 (1994).