bZIP proteins bind to a palindromic sequence without an ACGT core located in a seed-specific element of the pea lectin promoter

Pater, B.S. de; Katagiri, F.; Kijne, J.W.; Chua, N.H.

The Plant Journal (1994) **6**(2), 133–140

bZIP proteins bind to a palindromic sequence without an ACGT core located in a seed-specific element of the pea lectin promoter

Sylvia de Pater[1,2,*], Fumiaki Katagiri[2,†], Jan Kijne[1] and Nam-Hai Chua[2]

[1]*Institute of Molecular Plant Sciences and Center for Phytotechnology, Leiden University, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands, and*

[2]*Laboratory of Plant Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA*

Summary

Previously, it has been shown that a trimer of a 22 bp fragment of the promoter of the seed-specific pea lectin gene confers high expression in seed. Here it is reported that this fragment contains a binding site for the cloned basic domain/leucine zipper (bZIP) proteins TGA1a and Opaque-2 (O2). Gel shift assays, DNaseI footprinting and methylation interference assays using purified TGA1a were performed to determine whether additional binding sites are present in the *psl* promoter. Within the 469 bp upstream region only one other TGA1a binding site was found, which is much weaker than the one present in the 22 bp element. Both O2 and TGA1a bound to the odd base palindromic C-box sequence, ATGAGTCAT, present within the 22 bp fragment. The 22 bp fragment also contains the sequence CACGTA, which contains the ACGT core usually found in binding sites for bZIP proteins. However, this sequence did not significantly contribute to bZIP protein binding. The binding affinity of TGA1a for the odd base palindromic sequence was low relative to a high-affinity C-box (ATGACGTCAT). By contrast, O2 strongly bound to the odd base C-box; the affinity was comparable with that for high-affinity G- (GACACGTGTC) and C-boxes. It is concluded that the presence of an ACGT core sequence is not a prerequisite for high-affinity binding of O2.

Received 6 August 1993; revised 14 January 1994; accepted 20 April 1994.

*For correspondence (fax +31 71 275088).

†Present address: Department of Genetics, Harvard Medical School and Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA.

Introduction

The pea lectin (*psl*) gene is highly expressed during seed development, whereas expression declines during germination and is very low in vegetative tissue. Generally, tissue-specific gene expression is controlled by transcription factors that bind to specific DNA sequences in the promoter. Well-documented examples include activation of muscle-specific genes by MyoD (Weintraub *et al.*, 1991) and of liver-specific genes by liver factor-B1 (LF-B1) (Frain *et al.*, 1989) and D-site binding protein (DBP) (Mueller *et al.*, 1990).

The bZIP class of transcription factors contains a basic domain that directly contacts the DNA and a leucine zipper that mediates dimerization (Landschulz *et al.*, 1988). Molecular models propose that the leucine zipper positions a diverging pair of α-helical basic regions to make sequence-specific contacts with a dyad symmetrical DNA target site. Izawa *et al.* (1993) introduced a nomenclature for dyad symmetrical DNA binding sites containing an ACGT core. Previously, sequences containing a guanine 3' of the ACGT core sequence (CACGTG) have been designated G-boxes. Following this terminology, sequences containing cytidine, thymine or adenine were designated C-boxes, T-boxes or A-boxes, respectively. Hybrid boxes consist of two different half-sites.

Several genes encoding bZIP proteins have been cloned from plants (reviewed by Katagiri and Chua (1992) and Brunelle and Chua (1993)). These proteins can be distinguished on the basis of their DNA sequence specificity. A subset including G-box binding factor (GBF) and transcription activator factor-1 (TAF-1) preferentially binds to G-boxes. Proteins which belong to the TGA1 family differ from GBFs in that their optimal binding site is a C-box (Izawa *et al.*, 1993) and that they apparently do not form heterodimers with GBFs (Schindler *et al.*, 1992). A third class consists of proteins that have a relaxed specificity and bind to both G- and C-boxes (e.g. Opaque-2 (O2)) (Izawa *et al.*, 1993). All these boxes contain an ACGT core sequence, which has been postulated to be necessary for efficient binding of bZIP proteins (Izawa *et al.*, 1993; Schmidt *et al.*, 1992). Nucleotides flanking the six central base pairs strongly influence the binding specificity (Izawa *et al.*, 1993).

Several plant bZIP proteins or their binding sites have been shown to be involved in tissue-specific gene expression. O2 is a maize bZIP regulatory protein that plays

an essential role in controlling the expression of seed storage protein genes encoding the 22 kDa zein proteins and a 32 kDa albumin, termed b-32 (reviewed by Motto *et al.*, 1989). TGA1a is a tobacco transcription factor that binds to the A-domain of the cauliflower mosaic virus (CaMV) 35S promoter which confers expression mainly in root (Benfey *et al.*, 1989). The optimal binding site for TAF-1 (G-box) also confers expression in root tissues with a low activity in leaves (Salinas *et al.*, 1992). Another hybrid motif consisting of A-box and G-box half-sites directs developmentally regulated expression in seeds (Salinas *et al.*, 1992).

Here we show that a 22 bp fragment of the *psI* promoter that confers seed-specific gene expression upon a reporter gene in tobacco (de Pater *et al.*, 1993) contains a binding site for bZIP proteins, which does not contain the ACGT core sequence.

Results

Binding of O2 and TGA1a to the 22 bp psI *promoter fragment*

Previously, we have shown that a trimer of a 22 bp fragment of the *psI* promoter confers seed-specific gene expression in transgenic tobacco (de Pater *et al.*, 1993). This fragment (GACACGTAGAATGAGTCATCAC) is located at –56 to –35 with respect to the transcription start site (see below). Our objective was to characterize transcription factors that bind to this sequence. Examination of this fragment revealed the presence of three overlapping TGAC-like motifs at position –45/–36 (one in the coding strand and two in the non-coding strand) together forming a C-box (Izawa *et al.*, 1993), except that one central nucleotide is missing. Hereafter, we will refer to this sequence (ATGAGTCAT) as an odd base C-box. At position –54/–49 a G-box/A-box hybrid is present containing an ACGT core, which can be found in binding sites for many different bZIP proteins (Izawa *et al.*, 1993).

To determine whether the 22 bp fragment is a binding site for bZIP proteins, gel shift assays were performed with the bZIP proteins Opaque-2 (O2) (Schmidt *et al.*, 1990) and TGA1a (Katagiri *et al.*, 1989), which both bind to C-boxes with high affinity (Izawa *et al.*, 1993). O2 also binds to G-boxes with high affinity (Izawa *et al.*, 1993). Figure 1 shows that both proteins form complexes with the 22 bp fragment (W1). Competition with 50-fold molar excess of a trimer of W1 completely abolished complex formation. Competition experiments were also performed with the oligonucleotides 4A1 and 4A3. 4A1 is a tetramer of the *as-1* element, a strong binding site for TGA1a. The *as-1* element is derived from the cauliflower mosaic virus (CaMV) 35S promoter, where it is located between positions –82 and –62. Two base pairs of each binding

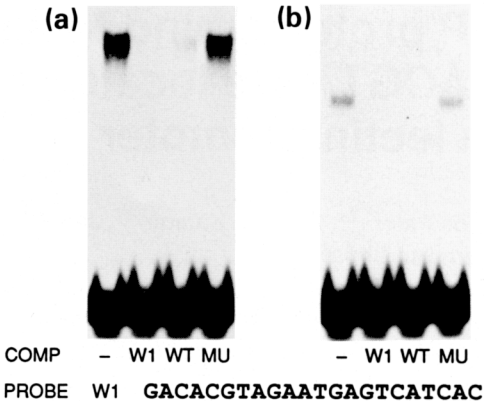

Figure 1. Binding of O2 and TGA1a to the 22 bp *psI* promoter fragment.
(a) The 22 bp fragment (W1) was used as a probe for binding of O2. The following competitors were added in 50-fold molar excess: no competitor (–); a trimer of W1; 4A1 (WT); and 4A3 (MU).
(b) Binding of TGA1a to W1 and competition with the same competitors as in (a).

site present in 4A1 are mutated in 4A3, causing a severe reduction of binding (Lam *et al.*, 1989). Otherwise, the sequences of 4A1 and 4A3 are identical. Oligonucleotide 4A1 competed for binding with both proteins. The affinity of TGA1a for 4A1 was somewhat stronger than the affinity of O2. Mutant oligonucleotide 4A3 did not compete for binding. Thus, W1 contains a binding site for the bZIP proteins TGA1a and O2.

The bZIP binding site on the 22 bp fragment is the strongest one within the psI *promoter*

To determine whether additional bZIP protein binding sites are present in the *psI* promoter, gel shift assays were performed with larger promoter fragments.

The *psI* promoter sequence as determined from a lambda clone isolated from a pea genomic library (variety Feltham First) is shown in Figure 2. It is very similar but not identical to formerly published *psI* promoter

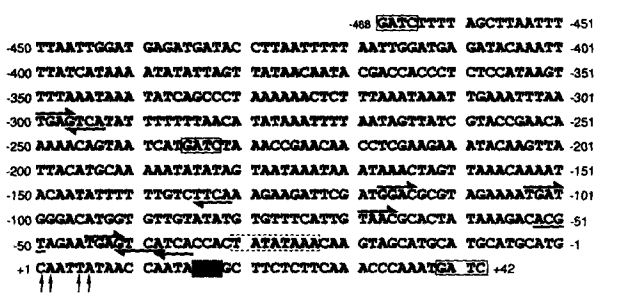

Figure 2. Nucleotide sequence of the *psI* promoter.
*Sau*3AI restriction sites are boxed with solid lines, the 5' ends of the mRNA are indicated by arrows and the ATG start codon is shaded. The putative TATA-box is boxed with dotted lines. TGAC-like motifs are overlined with arrows when present in the coding strand or underlined when present in the non-coding strand. The ACGT core sequence is underlined.

sequences from pea varieties Feltham First (Gatehouse *et al.*, 1987) and Frisson (Kaminski *et al.*, 1987). Although Gatehouse *et al.* isolated the lectin gene from the same variety used by us, the region from –7 to –19 (Figure 2) is absent in their sequence. It was important to know the exact sequence to be able to determine the transcription start site accurately and thereby define the upstream promoter region. Therefore, the region from –57 to +41 was amplified directly from genomic DNA by the polymerase chain reaction and cloned. Several clones were sequenced and they all were identical to the corresponding region in our genomic lambda clone, proving that our sequence is correct.

The transcription start site of the *psl* gene was determined by primer extension analysis, and was identified to be 10–14 bp upstream of the ATG start codon (data not shown). The most upstream start site was designated +1 (Figure 2). Thirty base pairs upstream of the transcription start site a sequence with homology to the TATA box consensus sequence is present.

The bZIP protein TGA1a was purified to homogeneity from *Escherichia coli* as described (Katagiri *et al.*, 1989). Two *Sau*3AI fragments, termed I and II, spanning 468 bp of the *psl* upstream region and 42 bp of the *psl* transcribed region (Figure 2 and 3a), were tested for binding using purified recombinant TGA1a protein. Figure 3 shows that DNA–protein complexes were formed with both fragments. In order to compare the affinity of TGA1a for each of these promoter fragments, competition experiments were performed. Besides fragments I and II, oligonucleotides 4A1 and 4A3 were used as competitors at two concentrations.

Competition experiments with each of the two *psl* promoter fragments showed that TGA1a binding to *psl* promoter fragment II was stronger than binding to fragment I (Figure 3). Even the mutant tetramer 4A3 competed for binding of TGA1a to fragment I (Figure 3b, lanes 9 and 10), indicating that the affinity of TGA1a for fragment I was very low. The affinity of TGA1a for fragment II was much higher than the affinity for fragment I, since 4A3 hardly competed for binding to fragment II (Figure 3, lanes 19 and 20). TGA1a bound more strongly to 4A1 than to fragment II (Figure 3, lanes 13, 14 and 17, 18), but it should be noted that 4A1 is a tetramer containing multiple TGA1a binding sites.

DNase I footprinting was performed to determine the specific sequences within the *psl* promoter fragments I and II which interact with TGA1a. A very weak footprint around –297 was observed on both strands of fragment I (Figure 4a, lanes 3, 4, 8 and 9). The footprint obtained with fragment II was much stronger (Figure 4a, lanes 13 to 15 and 19 to 21) and was located between positions –52 and –31 on both strands. This sequence is almost completely contained within the 22 bp fragment (W1).

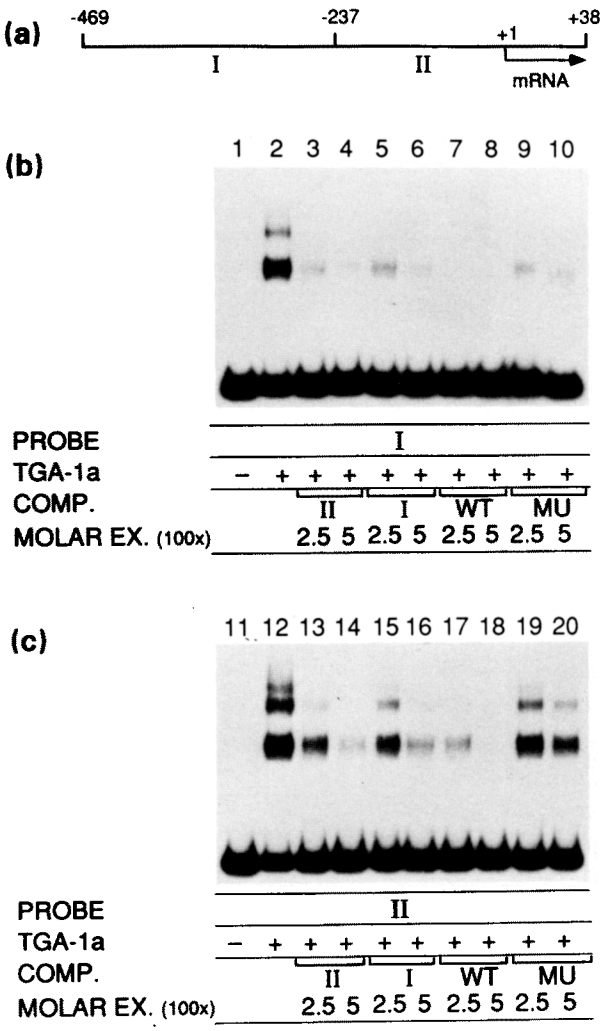

Figure 3. TGA1a binds to the *psl* promoter.
(a) *Sau*3AI fragments (I and II) of the *psl* promoter used as probes for gel shift assays.
(b) Gel shift assays with *psl* promoter fragment I and TGA1a. Competitors were fragment II, fragment I, tetramerized wild-type *as-1* (4A1; WT) and tetramerized mutant *as-1* (4A3; MU). Competitors were added in 250- (2.5) and 500-fold (5) molar excess (molar ex).
(c) Gel shifts with *psl* promoter fragment II as a probe and the same competitors as in (b).

Both footprinted sequences contain an odd base C-box with two and three overlapping motifs homologous to the TGAC half-site, respectively, as indicated by arrows in Figure 4(c).

To investigate which G-nucleotides of the latter footprinted sequence are involved in binding, methylation interference was performed using fragment II and purified TGA1a. Methylation of Gs at positions –45 and –43 in the coding strand and at positions –41 and –38 in the non-coding strand interfered with binding of TGA1a (Figure 4b and c). Methylation of any of the other G residues did not interfere with TGA1a binding. In conclusion, the odd base C-box sequence ATGAGTCAT located within the 22 bp

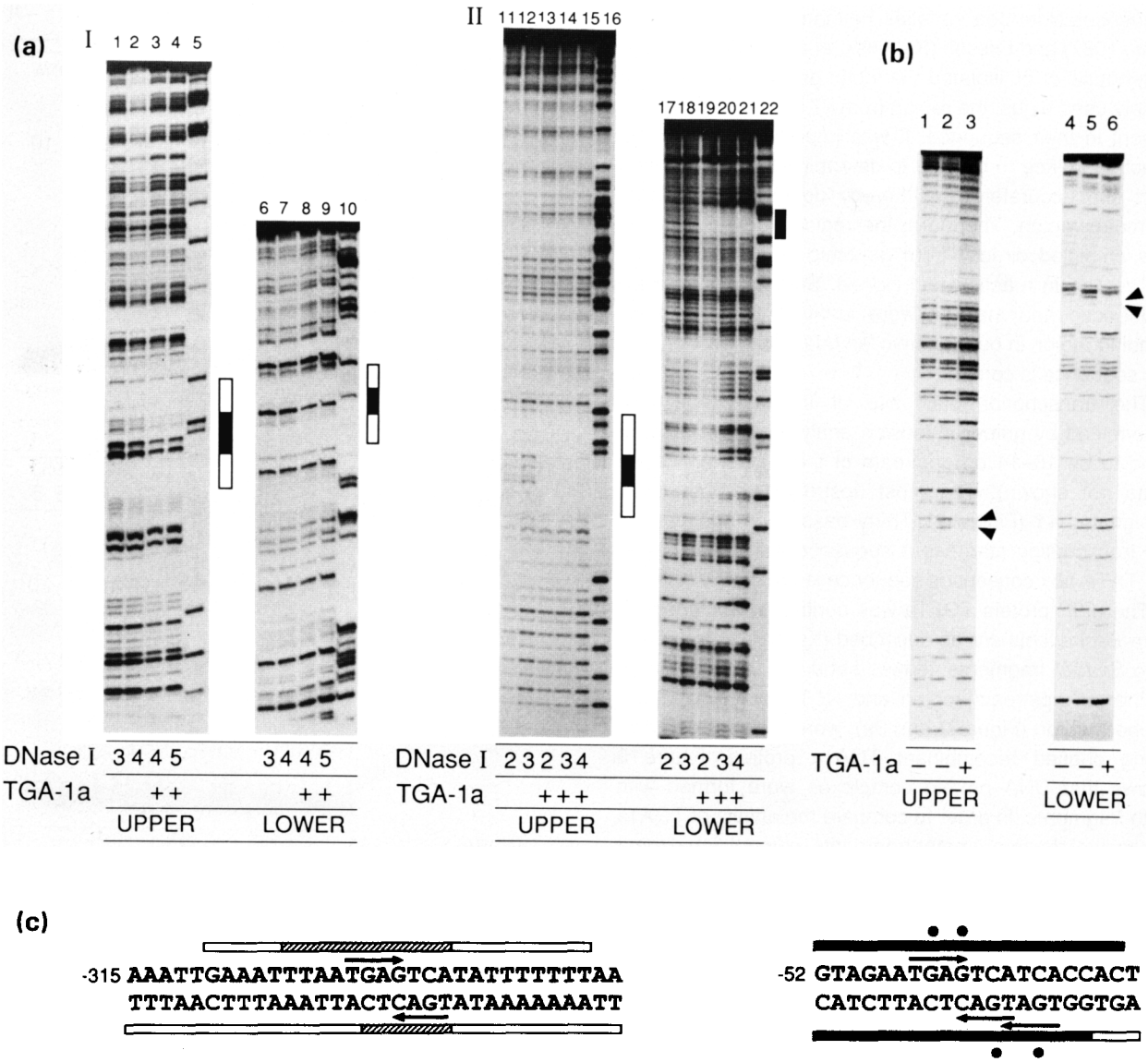

Figure 4. TGA1a binding sites in the *psl* promoter.

(a) DNase I protection of the *psl* promoter by TGA1a. Both strands of fragments I and II were used for DNase I footprinting with purified TGA1a. In lanes 1, 2, 6, 7, 11, 12, 17 and 18 the fragments were incubated with different amounts (in μl; 1 μl = 15 ng) of DNase I in the absence of TGA1a protein. Increasing amounts of TGA1a were used (lanes 3, 8, 13 and 19: 4.5 μg; lanes 4, 10, 14 and 20: 9 μg; lanes 15 and 21: 13.5 μg). Lanes 5, 10, 16 and 22 represent Maxam–Gilbert G-reactions performed on the labeled fragments. The footprints obtained with TGA1a are indicated by bars: open bars represent regions that are not digested by DNase I in the absence of protein and indicate the boundaries of DNaseI protection. Upper and lower strands were 3' end labeled.

(b) Methylation interference with the *psl* promoter and TGA1a. Lanes 1 and 4 are the patterns obtained with the probes in the absence of protein. Lanes 2 and 5 show the non-bound DNA, whereas lanes 3 and 6 represent DNA bound to purified TGA1a; bound and non-bound DNAs were separated on and purified from a gel after the binding reaction. The arrowheads indicate Gs which interfered with binding after methylation. Upper and lower strands of fragment II were 3' end labeled.

(c) Summary of regions and nucleotides of the *psl* promoter bound by TGA1a. The footprinted regions are located just downstream of positions –315 and –52 with respect to the transcription start site. TGAC-like motifs are indicated by arrows and the DNase I footprints by bars. Black bars represent strong footprints, hatched bars represent weak footprints and open bars represent flanking regions that are not digested by DNase I in the absence of TGA1a. Gs that interfere with binding of TGA1a when methylated are indicated by black dots.

promoter fragment (W1) is the only TGA1a binding site in the *psl* promoter downstream of position −469 with relatively high affinity.

Both O2 and TGA1a bind to the odd base C-box

To determine further which nucleotides within the W1 sequence were important for binding of O2 and TGA1a, competition assays were performed. Competitors, shown in Figure 5(a), were included in the binding reaction. M3 is mutated in the right half of W1, containing the odd base C-box (ATGAGTCAT), M4 is mutated in the left half of W1, containing the G/A-box hybrid (ACACGTAG), and M5 is mutated in both sequences. As shown in Figure 5(b), M3 was a poor competitor, whereas M4 competed to the same extent as did the wild-type sequence (W1). The double mutant M5 did not compete at all. A shorter oligonucleotide (W2), containing only the odd base C-box, competed very well, which further demonstrated that O2 bound to the odd base C-box and not to the G/A-box hybrid. The binding specificities of TGA1a and O2 were very similar (Figure 5c). These results show that TGA1a and O2 bind to the odd base C-box present in the W1 oligonucleotide, whereas the ACGT sequence does not significantly contribute to binding. This finding is consistent with the methylation interference assays, which showed that interference with binding of TGA1a occurred only when the four G residues present in the odd base C-box were methylated.

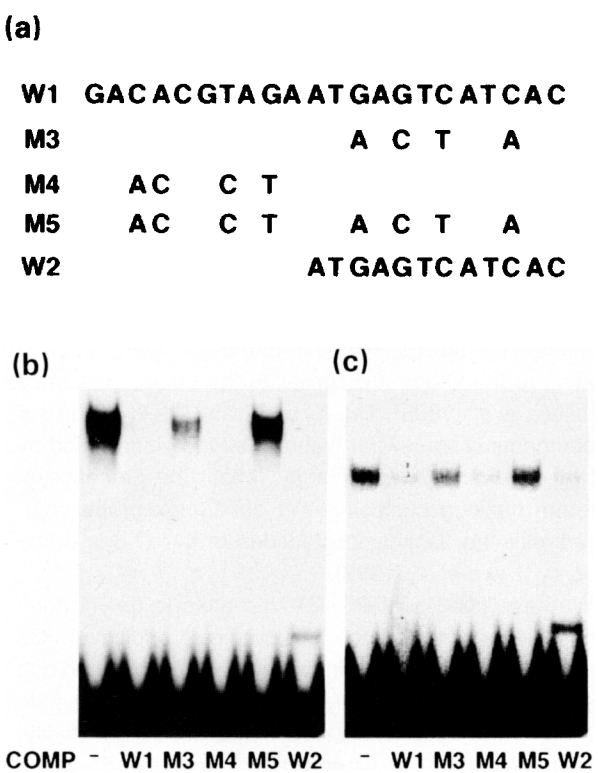

Figure 5. Nucleotide specificity of O2 and TGA1a binding to W1.
(a) Oligonucleotides used as probe (W1) and competitors (W1, M3, M4, M5 and W2).
(b) Gel shifts with O2 and W1, using different competitors (trimers of W1, M3, M4, M5, tetramer of W2) added in 100-fold molar excess.
(c) Gel shifts of TGA1a and W1 and the same competitors as in (b).

Binding affinity of TGA1a and O2 for W1

The binding affinities of TGA1a and O2 for the W1 sequence were determined using saturation binding assays (Cao *et al.*, 1991). For comparison, oligonucleotides containing a high-affinity G-box (GACACGTGTC) or a high-affinity C-box (ATGACGTCAT) (Izawa *et al.*, 1993) were used. The odd base C-box within the W1 oligonucleotide has the same flanking nucleotides as this high-affinity C-box. Equilibrium conditions were first determined at two different salt concentrations. Binding was allowed to proceed from 15 min to 4 or 8 h. Figure 6 shows the binding of O2 to the G-box. At 40 mM KCl, a salt concentration commonly used in gel shift assays, equilibrium was not reached even after 4 h. However, at 100 mM KCl equilibrium binding occurred after approximately 1 h. In general, a higher salt concentration drives the binding reaction to equilibrium more rapidly than lower salt conditions, not only with the probe/protein combination shown in Figure 6, but also with the other probes and with TGA1a. Based upon these experiments, the binding reactions were allowed to reach equilibrium by incubation for 2 h in 100 mM KCl to determine the dissociation constants (K_ds).

An example of a saturation gel shift assay to determine

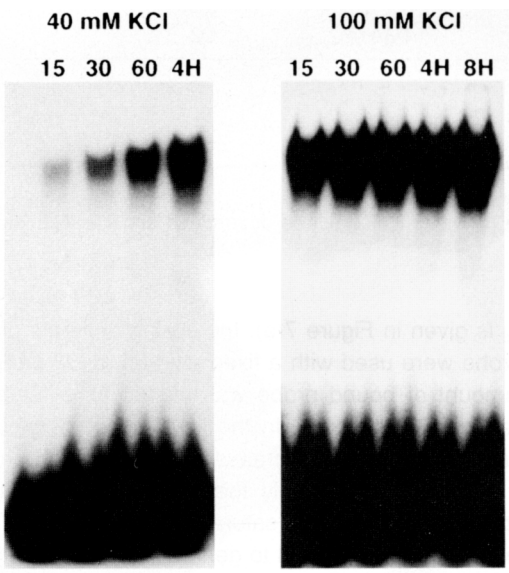

Figure 6. Time course of binding of O2 to the G-box oligonucleotide GACACGTGTC, in either 40 mM or 100 mM KCl.
The binding reactions were performed for 15 min to 8 h.

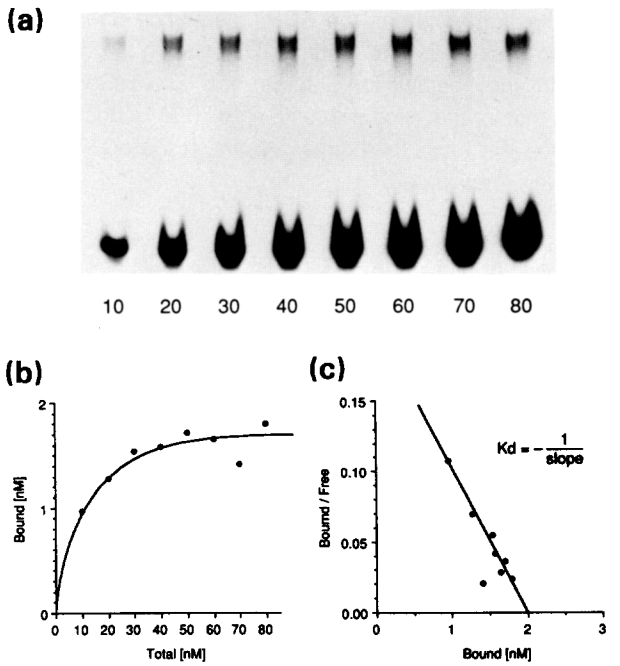

Figure 7. Measurement of the dissociation constant of O2 and the G-box.
(a) Saturation gel shift assay with increasing concentrations (10–80 nM) of ^{32}P-labeled G-box as a probe and a fixed amount of O2. Bound and free label was quantitated.
(b) The amount of bound probe plotted as a function of total probe.
(c) Scatchard plot of the saturation curve shown in (b). The dissociation constant (K_d) can be calculated ($K_d = -1/\text{slope}$) and is 11 nM.

Table 1. Dissociation constants (K_d) of complexes formed between TGA1a or O2 and the C-box (agcttaATGACGTCAT actcga), the G-box (agcttaGACACGTGTCactcga) or the W1 oligonucleotide (tcgagGACACGTAGAATGAGTCATCACgtcga)

	TGA1a	O2
C-box	5	8
G-box	>>>[a]	11
W1	174	23

K_d is given in nM.
[a]Large K_d, which could not be determined under the conditions used.

the K_d is given in Figure 7(a). Increasing amounts of G-box probe were used with a fixed amount of O2 protein. The amount of bound probe was plotted as a function of total input (Figure 7b). In the lower input range, the amount of bound probe increased with increased probe input. At higher input levels the bound probe reached a plateau level, reflecting saturation of the binding reaction. The data were used to generate a Scatchard plot (Figure 7c).

Table 1 presents the K_ds calculated for TGA1a and O2 for the three probes tested. TGA1a exhibited the highest binding affinity for the C-box. The affinity for W1 was 35-fold lower than for the C-box. The affinity for the G-box was very low and the K_d could not be determined with the probe concentrations used in the gel shift assays. In contrast, O2 bound to all three probes with high affinity. There was only a threefold difference in the K_ds for the C-box and the W1 sequence. Thus, O2 which has been shown to bind with high affinity to C-boxes as well as to G-boxes (Izawa *et al.*, 1993) also bound to the odd base C-box sequence in the W1 oligonucleotide with high affinity.

Discussion

Using cDNA-encoded protein expressed in and isolated from *E. coli*, sequence-specific binding of the bZIP proteins TGA1a and O2 was detected to the 22 bp *psI* promoter fragment, which confers seed-specific expression (de Pater *et al.*, 1993). This fragment contains the strongest bZIP protein binding site within the *psI* promoter. The sequence of the binding site is the odd base C-box ATGAGTCATCA. The weaker footprint in fragment I of the *psI* promoter (Figure 4) contains an identical sequence except for the last two nucleotides. It can be concluded that these two and possibly other flanking nucleotides are important for optimal binding of TGA1a to the fragment II sequence.

We have determined binding affinities using *in vitro* saturation binding assays. The relevance of these measurements is based on the assumption that only high-affinity binding sites are likely to function *in vivo*. However, it must be kept in mind that binding affinities determined by *in vitro* assays do not necessarily reflect *in vivo* binding affinities. *In vivo*, binding could be influenced by protein modifications or interaction with accessory proteins. The affinities we have found for TGA1a and O2 binding to the C-box are similar to the values described by Izawa *et al.* (1993). The K_d of O2 and the G-box in our experiments is somewhat higher than the value found by Izawa *et al.* (1993), which may reflect the presence of different flanking nucleotides in our G-box probe compared with the flanking nucleotides in the G-box probe used by Izawa *et al.* (1993).

We have used purified TGA1a protein to detect additional binding sites in the *psI* promoter. However, O2 binds to several different binding sites (G-, C-box and odd base C-box) with relatively high affinity, in contrast to TGA1a, which only binds to C-boxes with high affinity. This suggests that O2, when available as purified protein, would be more suitable than TGA1a to screen promoters for bZIP protein binding sites.

In previous studies, an ACGT core sequence was postulated to be necessary for efficient binding of bZIP proteins (Izawa *et al.*, 1993; Schmidt *et al.*, 1992), since

mutations within the ACGT core reduced binding affinity to a very low level. However, O2 efficiently binds the odd base C-box in which the central C residue in the ACGT core has been deleted. Also, the odd base C-box is the optimal binding site for the AP-1 family of bZIP transcription factors (Hill *et al.*, 1986). Thus, it makes an essential difference whether a mutation in the ACGT core consists of a nucleotide replacement or a deletion. Binding sites for bZIP proteins can now be extended to include sequences containing an AGT core.

Suckow *et al.* (1993) have shown that binding of bZIP proteins to different target sites including the odd base C-box (ATGAGTCAT), C- and G-boxes, requires the presence of specific amino acids in their basic domain. Substitution of residues −15 (alanine) and −17 (threonine) of the basic region of GCN4 (the first leucine of the leucine zipper is +1) by serine and arginine, respectively, broadened the binding specificity to include G-boxes. These amino acids are also present at these positions in the basic domain of O2. This is in agreement with our results, showing that O2 has a high affinity for several different binding sites. In the basic region of TGA1a an alanine is present at position −15, which might explain the low affinity of TGA1a for the G-box.

The odd base C-box is the optimal binding site for GCN4, whereas the C-box is the optimal binding site for ATF/CREB. Substitution of the leucine at position −6 in GCN4 by lysine, which is present at the corresponding position in ATF/CREB proteins, alters the half-site spacing specificity (Kim *et al.*, 1993). This mutation increases the relative affinity for the C-box over the odd base C-box. Thus, the lysine at position −6 in TGA1a might account for the low affinity for the odd base C-box.

A trimer of the W1 oligonucleotide confers high levels of gene expression in seeds and low expression in vegetative tissues of transgenic tobacco (de Pater *et al.*, 1993). The *in vivo* nucleotide requirement (de Pater *et al.*, 1993) correlates very well with the binding specificities of TGA1a and O2. Low expression in vegetative tissues could be conferred by a factor with low binding affinity for W1 like TGA1a and high expression in seed by a factor with high affinity for W1 like O2. Our future research will be aimed at identification of the factor responsible for W1-driven expression in seed. Based on our results we expect it to be a bZIP protein with high affinity for W1 and similar binding specificities as TGA1a and O2.

Experimental procedures

Isolation of the psl *gene and primer extension*

A genomic library of pea (*Pisum sativum*, variety Feltham First) produced by Gatehouse *et al.* (1987) was screened using the PSL encoding cDNA clone pMS2 (Stubbs *et al.*, 1986). The *Bgl*II–*Bam*HI fragment and the *Bam*HI fragment, which together contain the lectin coding region and 469 bp of the 5′ non-coding region and 402 bp of the 3′ non-coding region, were sequenced with the dideoxy method using sequenase (USB). The 5′ ends of *psl* mRNA were determined by primer extension. In brief, 30 μg of total RNA isolated from seeds harvested 20 days after flowering (DAF) were reverse transcribed using the ^{32}P-labeled antisense 19-mer primer 5′-GGTTTCAGTTGAGTTCACC-3′, that hybridizes at positions +96 to +114 and the products were analyzed on a sequencing gel.

Gel retardation, DNase I footprinting and methylation interference

TGA1a protein, encoded by a cDNA clone (Katagiri *et al.*, 1989), was isolated from *Escherichia coli* and purified to homogeneity as described by Katagiri *et al.* (1990). Crude *E. coli* extracts, isolated as described by Izawa *et al.* (1993) were kindly provided by R. Foster. The *Sau*3AI fragments containing the 5′ non-coding region of the *psl* gene were cloned in the *Bam*HI site of pBluescript SKII+. Synthetic oligonucleotides containing *Xho*I and *Sal*I sticky ends were cloned in pBluescript SKII+, as monomer (W1) or multimerized to trimers (W1, M3, M4 and M5). W2 was synthezised as a tetramer. Tetramers of wild-type (4A1, *as-1a*) and mutant (4A3, *as-1c*) ASF-1 binding sites were described by Lam *et al.* (1989). Plasmids were digested with *Xba*I and *Hin*dIII (*psl* promoter fragments I and II) or *Xho*I and *Sal*I (W1) and 3′ end labeled using the Klenow fragment of DNA polymerase I and [α-^{32}P]dCTP at both ends (for gel retardation) or at one end (for DNase I footprinting and methylation interference). For C- and G-box containing probes, complementary oligonucleotides were annealed and 500 ng were labeled by 3′ end labeling. C-box: 5′-AGCTTAATGACGTCATACTCGA-3′; G-box: 5′-AGCTTAGACACGTGTCACTCGA-3′ (Izawa *et al.*, 1993). Labeled fragments were isolated from polyacrylamide gels.

Gel shift experiments (Green *et al.*, 1987), and *in vitro* DNase I footprinting and methylation interference tests were performed essentially as described previously (Green *et al.*, 1989). Binding reactions for gel shift assays were done in a final volume of 10 μl with 0.04–0.1 ng (1 fmol) of probe (fragment I and II) and 0.9 μg of purified TGA1a in the presence of 1–3 μg of poly(d[IC]). For competitive gel shifts with oligomers 0.35-0.7 ng (10–20 fmol) of probe and 0.4–4 μg of *E. coli* proteins were used in the presence of 3 μg of poly(d(IC)). Saturation binding assays to determine the dissociation constants (K_d) were carried out as described by Cao *et al.* (1991). Probe concentrations between 10 and 80 nM were used. Bound and free probe were measured using a phosphorimager (Molecular Dynamics), and K_ds were determined by the method of Scatchard (1949).

Acknowledgments

We thank Bob Schmidt for the maize O2 cDNA clone, Randy Foster for crude *E. coli* extracts of TGA1a and O2, and Peter Hock and the Biology Photographic Service, Leiden University, for preparation of the figures. We are especially indebted to Johan Memelink for fruitful discussions and comments on the manuscript. This work was supported by a grant from the Netherlands Organization for Applied Scientific Research (TNO) to SdP and by a grant from the Rockefeller Foundation to NHC.

References

Benfey, P.N., Ren, L. and Chua, N.-H. (1989) The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. *EMBO J.* **8**, 2195–2202.

Brunelle, A.N. and Chua, N.-H. (1993) Transcription regulatory proteins in higher plants. *Curr. Biol.* **3**, 254–258.

Cao, Z., Umek, R.M. and McKnight, S.L. (1991) Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells. *Genes Devel.* **5**, 1538–1552.

Frain, M., Swart, G., Monaci, P., Nicosia, A., Stämpfli, S., Frank, R. and Cortese, R. (1989) The liver-specific transcription factor LF-B1 contains a highly diverged homeobox DNA binding domain. *Cell*, **59**, 145–157.

Gatehouse, J.A., Bown, D., Evans, I.M., Gatehouse, L.N., Jobes, D., Preston, P. and Croy, R.R.D. (1987) Sequence of the seed lectin gene from pea (*Pisum sativum* L.). *Nucl. Acids Res.* **15**, 7642.

Green, P.J., Kay, S.A. and Chua, N.-H. (1987) Sequence-specific interaction of a pea nuclear factor with light-responsive elements upstream of the rbcS-3A gene. *EMBO J.* **6**, 2543–2549.

Green, P.J., Kay, S.A., Lam, E. and Chua, N.-H. (1989) *In vitro* DNA footprinting. In *Plant Molecular Biology Manual*, Suppl. II (Gelvin, S. and Schilperoort, R.A., eds). Dordrecht: Kluwer Academic Publishers, pp 1–22.

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.

Kaminski, P.A., Buffard, D. and Strosberg, A.D. (1987) The pea lectin gene family contains only one functional gene. *Plant Mol. Biol.* **9**, 497–507.

Katagiri, F. and Chua, N.-H. (1992) Plant transcription factors: Present knowledge and future challenges. *Trends Genet.* **8**, 22–27.

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.

Katagiri, F., Yamazaki, K., Horikoshi, M., Roeder, R.G. and Chua, N.-H. (1990) A plant DNA-binding protein increases the number of active preinitiation complexes in a human *in vitro* transcription system. *Genes Devel.* **4**, 1899–1909.

Kim, J., Tzamarias, D., Ellenberg, T., Harrison, S.C. and Struhl, K. (1993) Adaptability at the protein–DNA interface is an important aspect of sequence recognition by bZIP proteins. *Proc. Natl Acad. Sci. USA*, **90**, 4513–4517.

Lam, E., Benfey, P.N., Gilmartin, P.M., Fang, R.-X. and Chua, N.-H. (1989) Site-specific mutations alter *in vitro* factor binding and change promoter expression pattern in transgenic plants. *Proc. Natl Acad. Sci. USA*, **86**, 7890–7894.

Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science*, **240**, 1759–1764.

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

Mueller, C.R., Maire, P. and Schibler, U. (1990) DBP, a liver-enriched transcriptional activator, is expressed late in ontogeny and its tissue specificity is determined posttranscriptionally. *Cell*, **61**, 279–291.

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

Salinas, J., Oeda, K. and Chua, N.-H. (1992) Two G-box-related sequences confer different expression patterns in transgenic tobacco. *Plant Cell*, **4**, 1485–1493.

Scatchard, G. (1949) The attractions of proteins for small molecules and ions. *Ann. N.Y. Acad. Sci.* **51**, 660–672.

Schindler, U., Beckmann, H. and Cashmore, A.R. (1992) TGA1 and G-box binding factors: Two distinct classes of Arabidopsis leucine zipper proteins compete for the G-box-like element TGACGTGG. *Plant Cell*, **4**, 1309–1319.

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 bind 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.

Stubbs, M.E., Carver, J.P. and Dunn, R.J. (1986) Production of pea lectin in *Escherichia coli*. *J. Biol. Chem.* **261**, 6141–6144.

Suckow, M., von Wilcken-Bergmann, B. and Müller-Hill, B. (1993) 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.

Weintraub, H., Davis, R., Tapscott, S., *et al.* (1991) The *myoD* gene family: Nodal point during specification of the muscle cell lineage. *Science*, **251**, 761–766

EMBL Data Library accession number X66368 (*Pisum sativum*).