4624–4631 *Nucleic Acids Research, 1996, Vol. 24, No. 23*

Characterization of a zinc-dependent transcriptional activator from *Arabidopsis*

Sylvia de Pater[1,2,*], Valentina Greco[1,+], Khanh Pham[1], Johan Memelink[2,3] and Jan Kijne[1,3]

[1]Center for Phytotechnology and [3]Institute of Molecular Plant Sciences, 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

Received September 12, 1996; Revised and Accepted October 18, 1996 DDBJ/EMBL/GenBank accession no. X92976

ABSTRACT

The C_2-H_2 zinc-finger is a widely occurring DNA binding motif, usually present as tandem repeats. The majority of C_2-H_2 zinc-finger proteins that have been studied are derived from animals. Here, we characterize a member of a distinct class of plant C_2-H_2 zinc-finger proteins in detail. A cDNA clone encoding a DNA binding protein from *Arabidopsis* was isolated by SouthWestern screening. The protein, termed ZAP1 (Zinc-dependent Activator Protein-1), is encoded by a single copy gene, which is expressed to similar levels in root and flower, to a somewhat lower level in stem and to low levels in leaf and siliques. The optimal binding site was determined by random binding site selection, and the consensus sequence found is CGTTGACCGAG. The homology between ZAP1 and other DNA binding proteins is restricted to a repeated region of a stretch of 24 highly conserved amino acids followed by a zinc-finger motif (C-X_4-C-$X_{22–23}$-H-X_1-H). The C-terminal zinc-finger region is essential for DNA binding, whereas deletion of the N-terminal one resulted in 2.5-fold reduced binding affinity. Binding of ZAP1 to DNA was abolished by metal-chelating agents. The activation domain as determined in yeast is adjacent to and possibly overlapping with the DNA binding domain. Particle bombardment experiments with plant cells showed that ZAP1 increases expression of a *gus*A reporter gene that is under control of ZAP1 binding sites. We conclude that ZAP1 is a plant transcriptional activator with a C_2-H_2 zinc-finger DNA binding domain.

INTRODUCTION

Transcriptional regulation of gene expression is determined by the interaction of transcription factors with sequences in the promoter region. Besides general transcription factors that assemble at the TATA-box (1), a variety of sequence-specific DNA binding proteins have been studied, that are necessary for inducible or high levels of transcription. Transcription factors can be classified according to their DNA binding domain (2). A widely occurring DNA binding motif is the so-called zinc-finger (3). The term zinc-finger applies to a rather diverse set of protein motifs. These motifs have in common that zinc-ions interact with histidines and/or cysteines to stabilize a small functional protein domain or 'finger'. In addition to DNA binding, zinc-fingers may play a role in protein–protein interactions, as observed for the LIM domain (4).

Several cDNA clones encoding different classes of zinc-finger DNA binding proteins have been isolated from plants (5). Recently a novel class of DNA binding proteins has been found, that contain a DNA binding domain predicted to form a special type of C_2-H_2 zinc-finger structure. This motif, only found in plants so far, is present in SPF1 from sweet potato (6) and ABF1 and ABF2 from wild oat (7). A stretch of 40 amino acids of the DNA binding domain is duplicated in the N-terminal part of SPF1 and ABF1. SPF1 binds to promoter sequences of the sporamin and β-amylase genes and is thought to play a role in sucrose-inducible gene expression (6). ABF1 and ABF2 bind to α-amylase 2 promoters and may play a role in the induction of α-amylase expression during germination (7).

For many different plant transcription factors the exact sequence requirement for DNA binding have been determined. The most extensively studied binding sites are for bZIP proteins (8). They contain a core ACGT (8) or AC/GT (9) sequence with different flanking sequences. For one group of bZIP proteins, the C-box binding proteins, the binding sites contain TGAC half sites. This sequence can also be found in binding sites for an unrelated plant zinc-finger protein (10) and for animal nuclear hormone receptors (11). Thus, the TGAC sequence is present in binding sites for different classes of DNA binding proteins and forms part of DNA elements that are likely to confer different expression patterns.

For transcriptional activation, the activation domain of a sequence-specific DNA binding protein needs to interact directly or indirectly via co-activators with the general transcription machinery (12). Several different types of activation domains have been identified and are classified as acidic, glutamine rich or proline rich. The mechanisms by which these different activation domains function involve protein–protein interaction with one or more components of the basal transcription machinery, thereby recruiting the basal machinery to the promoter or increasing the rate of a kinetically slow step (13,14). Several plant DNA binding proteins were shown to act as transcriptional activators (5). Only for very few of these proteins have the activation domains been

*To whom correspondence should be addressed. Tel: +31 71 527 4803; Fax: +31 71 527 4863; Email: sbtnbp@rulsfb.leidenuniv.nl

+Present address: Istituto di Ricerca sulle Biotecnologie Agroalimentari, CNR, Via Prov. le Lecce-Monteroni, 73100 Lecce, Italy

localized. The transcriptional activator C1 from maize (15), regulating anthocyanin biosynthesis, and a myb protein from snapdragon (16), regulating flavonoid biosynthesis, contain acidic activation domains.

Here, we describe the characterization of a DNA binding protein from *Arabidopsis*, termed ZAP1, that is homologous with SPF1 from sweet potato and ABF1 and ABF2 from wild oat. The optimal binding site was determined and the activation and DNA binding domains were mapped. For binding to its optimal sequence, CGTTGACCGAG, ZAP1 requires metal-ions. Since the DNA binding domain contains a C_2-H_2 motif, we postulate that it forms a zinc-finger. Furthermore, we show that ZAP1 can function as a sequence-specific transcriptional activator in plant cells.

MATERIALS AND METHODS

SouthWestern screening

A cDNA expression library in λZAP (Stratagene), made on flowers and 0 to 4-day-old siliques of *Arabidopsis thaliana* (ecotype C24), was screened with a mixture of three probes: 3W1 (17), W2 (9) and IWT (18) (Table 1). Lambda phages (20 000 p.f.u./150 mm plate) were grown on *E.coli* XL-1 blue for 4 h. Nitrocellulose filters saturated with 10 mM IPTG were placed on the plates and phages were allowed to grow for another 4 h. Filters were blocked for 1 h in binding buffer [20 mM HEPES/KOH (pH 7.2), 40 mM KCl, 1 mM EDTA, 0.5 mM DTT, 10% glycerol] supplemented with 5% non-fat dry milk at room temperature. After washing twice in binding buffer the filters were incubated in binding buffer containing 2 ng/ml probe and 5 µg/ml sonicated calf thymus DNA for 1 h. Probes were end-labeled using the Klenow fragment of DNA polymerase I and [α-^{32}P]dNTPs. Finally, the filters were quickly washed 3 times and dried before autoradiography.

Table 1. Nucleotide sequence of DNA binding sites

Name	Repeat sequence	Number of repeats
3W1	TCGACGTGA**TGACTCA**TTC**TACGTG**TCC	3
W2	GTGA**TGACTCA**T	4
IWT	G**TACGTG**GCG	4
IMU	**TGAC**TGTTCT**TGAC**TGTTCT	2
4A1	C**TGAC**GTAAGGGA**TGAC**GCAC	4

Odd base C-boxes are underlined, G/A-boxes double underlined and TGAC motifs are shown in bold type.

Sequence analysis

The *Zap1* cDNA insert was subcloned and sequenced with a Pharmacia T7 sequencing kit. Nucleotide sequence data were collected, assembled and analysed with a VAX computer using the Genetics Computer Group Sequence Analysis Software Package (19).

Blotting and hybridization

For genomic Southern hybridization 1 µg DNA was digested, electrophoresed on a 0.8% agarose gel, blotted and hybridized as described (20). For Northern hybridization 10 µg total RNA were electrophoresed on a 1.5% formaldehyde gel, blotted and hybridized as described (20). As probes randomly labeled cDNA inserts were used. Blots were washed with 0.1× SSPE, 0.1% SDS at 42°C.

Gel shift assays and binding site selection

Gel shifts were performed essentially as described (9). DNA (10–20 fmol), 3′-end labeled using the Klenow fragment of DNA polymerase I, was incubated with 50 ng of crude extract from *E.coli* expressing ZAP1.

Random binding site selection was essentially done as described (21). A pool of oligonucleotides containing a stretch of 20 random nucleotides (TCTAGAACTAGTGGATCC-N_{20}-CGATACCGT-CGACCTCG) were annealed with KS primer (CGAGGTCGA-CGGTATCG) and extended with the Klenow fragment of DNA polymerase I. One µg ds-probe and 5 µg crude extract from *E.coli* expressing ZAP1 was allowed to form complexes. Five ng of [α-^{32}P]dCTP-labeled probe was included in the binding reaction in order to follow the complex during further procedures. After binding, the complex was separated from the free probe on a 5% non-denaturing polyacrylamide gel. The gel was dried and exposed for 2 h to X-ray film. Complexes were cut out from the dried gel and DNA was eluted overnight in 0.5 M NH_4Ac (pH 8.0), 1 mM EDTA. DNA was phenol/chloroform extracted and 0.1 ng was used for PCR amplification for 15 cycles in the presence of [α-^{32}P]dCTP using SK (TCTAGAACTAGTGGATC) and KS primers. In the following rounds of binding 25, 2.5 and 0.5 ng PCR amplified DNA were used, respectively. After the fourth round DNA was amplified, digested with *Bam*HI and *Sal*I and cloned in pBluescript SKII. The inserts of 96 different plasmids were amplified by PCR using SK and KS primers and were used as competitors in gel shift assays. About 50% of the analysed inserts were found to compete and were used as gel shift probes after removal of most of the flanking sequences by digestion with *Bam*HI and *Sal*I.

In vitro protein synthesis

For functional analysis of ZAP1, deletion derivatives were produced using selected restriction sites present in the cDNA (Fig. 8). In D2 the *Nco*I site (accession number X92976, position 819) was fused to the *Sac*I site (position 1301) after generation of blunt ends by the Klenow fragment of DNA polymerase I and T4 DNA polymerase, respectively. In D3 the *Hin*fI site (position 344) was fused to the *Bam*HI site (position 776), separated by polylinker sequence. In D4 the *Hin*fI site (position 344) was fused to the *Sac*I site (position 1301) after generation of blunt ends by the Klenow fragment of DNA polymerase I and T4 DNA polymerase, respectively. In D5 the *Kpn*I site (position 1047) was blunt-ended with T4 DNA polymerase and fused to the *Rsa*I site (position 1639) present in the 3′ non-coding region, in order to retain the polyA tail. The D5 encoded protein contains nine additional amino acids at its C-terminus, from the *Rsa*I site to the first stop codon. In D6 the region 5′ to the *Kpn*I site was deleted.

^{35}S-labeled ZAP1 and deletion derivatives were synthesized using the TnT system of Promega. Products were analysed on a 10% SDS-polyacrylamide gel and quantified with a phosphorimager (Molecular Dynamics). Equimolar amounts of protein were used for gel shifts.

Determination of activation domain

N-terminal deletions of *Zap1* were cloned in pAS1ΔNco, a derivative of pAS1 (22), and introduced into yeast strain Y153 (22). Yeast transformants were grown on nitrocellulose filters, frozen in liquid N_2 and stained for β-gal activity by incubating the filters on paper impregnated with 0.3 mg/ml X-gal in 0.1 M Na-phosphate-

NaOH (pH 7), 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercapto-ethanol.

Particle bombardment

The synthetic binding site 65 (BS 65) was multimerized. The monomer, dimer, trimer and hexamer were cloned in plasmid GusSH-47 (23). *Zap1* cDNA missing the first 150 bp of the leader sequence was cloned under control of the CaMV 35S promoter and the *Rbc*S-3C terminator.

Catharanthus roseus cell suspensions were grown in LS medium containing 2 mg/l NAA and 0.2 mg/l kinetin at 25°C with light/dark intervals of 16/10 h and subcultured weekly. Four days after subculture, 1 ml of suspension was collected on Whatman filter paper using a Büchner funnel. Paper discs were transferred to Petri dishes containing solidified medium and left overnight in the growth chamber.

Particles were prepared using a procedure modified from (24). The precipitation mixture included 2 mg gold particles (Aldrich), 25 µg plasmid DNA, 8% PEG 1500, 5 mM spermidine in a total volume of 575 µl. After each addition the suspension was mixed by sonication. After 10 min whirl mixing at 4°C, gold particles were collected by centrifugation at 1000 r.p.m. for 5 min at 4°C and resuspended in 20 µl supernatant. The suspension was sonicated for 2 s and 2 µl was pipetted on the support screen of the loader of a home-made helium gun. A steel mesh was used at a distance of 3 cm below the loader. Petri dishes containing the paper filter with the cells were placed 9.5 cm below the loader and bombardment was performed at 35 mbar with 2,5 bar helium pressure. After 16–20 h the cells were overlaid with 250 µl GUS staining solution (25) containing 6 mM 5-bromo-4-chloro-3-indolyl glucuronide (X-gluc) and incubated for 1 h at 37°C, followed by an overnight incubation at 28°C.

For each bombardment experiment constructs were delivered in triplicate. This was repeated once or twice depending on the construct. GUS activities obtained by one representative bombardment experiment were quantified on black and white video images of the plates by measuring dark staining of a fixed area, using Image Quant^{tm} software. The value obtained after bombardment with the control plasmid GusSH-47 was subtracted from all other values. The value obtained after bombardment with the GusSH-47 derivative containing the BS65 monomer was set at 100.

RESULTS

Isolation of *Zap1*

In a SouthWestern screening of an *Arabidopsis thaliana* (ecotype C24) cDNA expression library with TGAC half site-containing binding sites (C-boxes and G/A box hybrids), different classes of DNA binding proteins were obtained. The sequences of these clones were compared with databases. One clone was homologous to previously isolated cDNAs from sweet potato (6), wild oat (7) and cucumber (accession number L44134) and to EST sequences of *Arabidopsis* (accession numbers T44598 and T45479) and rice (accession number D38923). The clone from sweet potato, termed *SPF1*, was described to encode a novel type of DNA binding protein. The *Arabidopsis* cDNA clone is 1888 bp long and contains a polyA-tail of 62 bp. It contains an open reading frame encoding a protein of 463 amino acids in length with a predicted molecular mass of 51.2 kDa. The most striking feature of the putative polypeptide is the presence of two regions of 58

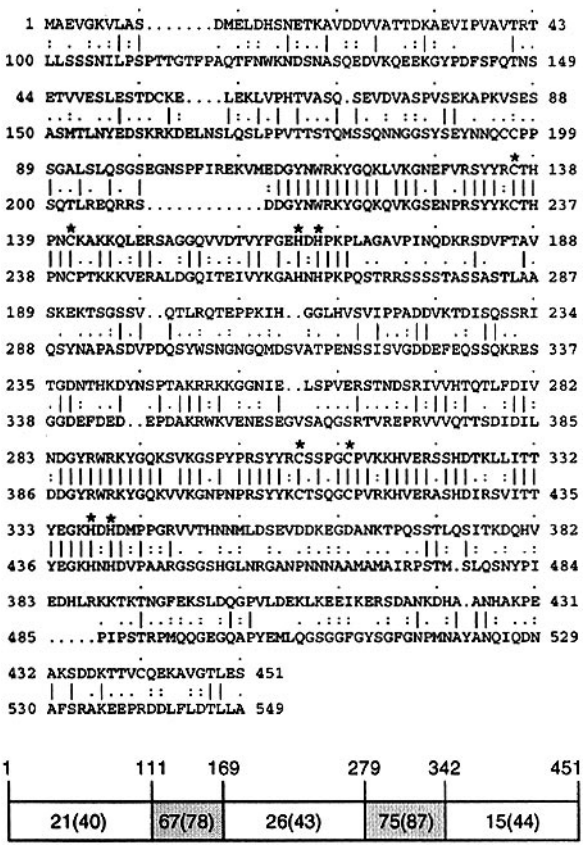

Figure 1. Comparison of ZAP1 from *Arabidopsis* with SPF1 from sweet potato. Identical amino acids are indicated by vertical lines and similar amino acids by dots. Conserved cysteine and histidine residues possibly involved in the formation of zinc-fingers are indicated by asterisks. The bar at the bottom of the figure shows the percentage identity and similarity (between brackets) of different parts of the proteins. The two repeated regions are shaded.

amino acids, which are 56% identical. The homology with SPF1 is highest in these repeated regions (Fig. 1). They are 67 and 75% identical, with the highest homology in the N-terminal part of each conserved sequence. The protein encoded by the cDNA clone isolated from sweet potato is 99 amino acids longer at the N-terminus. The *Arabidopsis* cDNA is full-length since there are in frame stop codons preceding the predicted start codon. We have designated the *Arabidopsis* protein ZAP1 (Zinc-dependent Activator Protein 1).

Expression pattern of *Zap1*

To explore the expression of *Zap1* in different *Arabidopsis* organs, we performed Northern blot analysis. Hybridization with total RNA from root, stem, leaf, flower and siliques showed expression mainly in root and flower (Fig. 2A). The length of the detected transcript is ~2.0 kb. The expression in stem was somewhat lower and in leaf and siliques the expression was hardly detectable. The blot was reprobed with ubiquitin cDNA to show the presence of intact RNA in all lanes (Fig. 2B).

We performed a Southern blot analysis to determine the number of genes that are homologous to *Zap1*. Figure 3 shows one band in each lane except with *Bam*HI which has a recognition site within the gene. This indicates that the observed RNA levels are derived from a single gene.

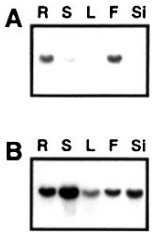

Figure 2. *Zap1* expression in *Arabidopsis* organs. A gel blot containing total RNA isolated from root (R), stem (S), leaf (L), flower (F) and siliques (Si) of *Arabidopsis* was hybridized with *Zap1* (**A**) or ubiquitin cDNA (**B**).

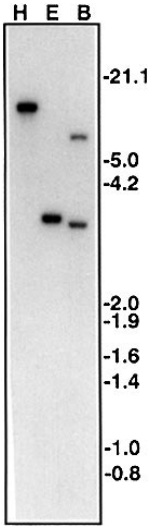

Figure 3. Southern blot with *Zap1* cDNA. A gel blot of *Arabidopsis* DNA digested with *Hin*dIII (H), *Eco*RI (E) or *Bam*HI (B) was hybridized with *Zap1* cDNA. Positions and sizes in kb of *Eco*RI and *Hin*dIII digested lambda DNA fragments are indicated.

DNA binding site specificity of ZAP1

The cDNA clone encoding ZAP1 was isolated by SouthWestern screening with a mixture of odd base C-boxes and G/A-boxes [3W1 (17), W2 (9) and IWT (18)], which are medium to high affinity binding sites for bZIP proteins. To determine the DNA binding specificity of ZAP1, these probes and some related ones (Table 1) were used in a gel shift experiment, using crude extract from *E.coli* expressing ZAP1. Figure 4 shows that ZAP1 binds to 3W1 and the shorter derivative W2. These probes both contain an odd base C-box (9). Lower affinity was observed for 4A1 (26), a tetramer of the as-1 sequence of the CaMV 35S promoter. No binding to IWT was observed, but the affinity for its mutant derivative IMU (18) was relatively high. The sequences that show binding to ZAP1 have the C-box half site sequence TGAC in common.

Although IMU was the best binding site used in this gel shift experiment, it was not clear whether it was an optimal binding site for ZAP1. We therefore performed a random binding site selection assay that is based on the selection of specific DNA binding sites from a pool of randomized oligonucleotides. A probe containing 20 random base pairs inserted in the center of a synthetic oligonucleotide pool was subjected to binding to ZAP1

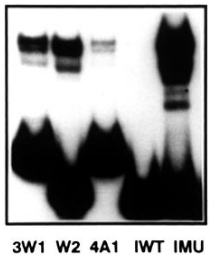

Figure 4. DNA binding specificity of ZAP1. Crude *E.coli* extract expressing ZAP1 was incubated with the following binding sites: 3W1, W2, 4A1, IWT and IMU (Table 1).

followed by gel shift analysis. After four rounds of selection the oligonucleotides were cloned and individual clones were sequenced and tested in gel shifts as competitors as well as probes. The sequences of 52 binding sites are shown in Figure 5. These sequences were divided in high affinity binding sites (>30% of the best binding site), medium affinity binding sites (10–30%) and low affinity binding sites (<10%). From these binding site preferences, the consensus sequence CGTTGACCGAG was deduced. The consensus sequence indeed contains the C-box half site TGAC, present in 3W1, W2, 4A1 and IMU. Otherwise no resemblance with bZIP protein binding sites was found. Most of the high affinity binding sites contain two sequences with high similarity to the consensus sequence.

To compare the relative affinities of ZAP1 for a high affinity binding site (BS 65) and for IMU, we performed competitive gel shift experiments. Two other selected binding sites were included in the experiment. IMU probe was incubated with ZAP1 in the presence of increasing amounts of competitor (Fig. 6). The affinities for the selected binding sites BS1 and BS105 (10–20%; Fig. 5) were ~2-fold higher compared to IMU. The affinity for BS65 was at least 10-fold higher. Since IMU is a tetramer, the affinity of ZAP1 for the corresponding monomer is probably lower. The affinities of ZAP1 for the probes that were used for the SouthWestern screening (3W1 and W2) are ~5-fold lower than the affinity for IMU (Fig. 4) and more than 50-fold lower than the affinity for BS65.

Zinc requirement for DNA binding of ZAP1

ZAP1 protein contains a repeated sequence (56% identity), that is also conserved in ABF1 and ABF2 of wild oat (7) and in SPF1 from sweet potato (6). The C-terminal part of these repeated regions contain putative zinc-finger motifs ($C\text{-}X_4\text{-}C\text{-}X_{22-23}\text{-}H\text{-}X_1\text{-}H$) somewhat different from previously described zinc-fingers. Binding of ABF1 and ABF2 is abolished by the metal-chelating agent 1,10-*o*-phenanthroline and by EDTA (7). We tested whether these agents also abolish binding of ZAP1 (Fig. 7). No complex formation was observed after addition of 10 mM 1,10-*o*-phenanthroline (in 10% ethanol final concentration) or 50 mM EDTA to the binding reaction, whereas addition of 10% ethanol or 50 mM EGTA (calcium-chelating agent) did not abolish ZAP1 binding. As a control we tested the effect of these metal-chelating agents on binding of proteins that do not require metal-ions. Binding of the bZIP proteins RITA-1 (27) to 3W1 and TGA2 (28) to 4A1 was not affected by 1,10-*o*-phenanthroline or EDTA (results not shown). These results show that ZAP1 requires metal-ions for its binding activity. The presence of zinc-finger motifs in ZAP1 suggests that this metal is zinc.

#	sequence	affinity
65	atcgTTGACCGAGTTGACTTTTTAggatcc	100.0
59	atcgTTGACCGAGCTTTGGACTTTggatcc	83.4
46	atcgTTGACCGAGTTTGACCGCCGggatcc	75.8
35	atcgTTGACCGAGTCCCGCGTCATggatcc	60.4
94	atcgTTGACCGAGCCGGTCGTCTAggatcc	57.2
5	atcgTTGACCGAGTTGACAGGCCTggatcc	52.3
100	atcgTTGACCGAGTCAAACTGTCTggatcc	51.2
78	atcgTTGACCGAGCGGCTGACCCTggatcc	50.5
80	atcgTTGACCGAGTCTTCCCCCCTggatcc	45.8
15	atcgTTGAGCTCAAAGTGCAACA.ggatcc	45.3
66	atcgTAGACCGCGTTGACCGATCGggatcc	45.2
82	atcgTTGACCGAGCCTGGTCGCCTggatcc	44.9
12	atcgCGTTGACCGAGCATCTAGTTggatcc	39.4
113	atcgTTGACCGAGTCAGCAGGCCTggatcc	37.9
110	atcgTTGACCGAGCTTCATGGACGggatcc	34.9
115	atcgCGTTGACCGAGCGTGTTTCTggatcc	28.4
53	atcgTCGAGCGAGCGTTGACTATTggatcc	24.9
48	atcgTTGACCGAGCACCCAAGCCAggatcc	24.7
57	atcgTGTCGTTGACCGGTCCCCCGggatcc	22.9
98	atcgTTGACCGAGCACCCCACGCCggatcc	22.8
83	atcgTTGACCGAGCACATCTTAATggatcc	22.7
41	atcgCCCGTTGACCGACTCGCCCTggatcc	22.4
1	atcgTTGACCGAGCCGGTCGTCTAggatcc	21.5
89	actgTTGACCGTTGTGTTCCCCCAggatcc	21.2
72	atcgTTGACCGAGCCCCTCACTTCGggatcc	21.0
92	atcgTTGACCGAGCAGCCCCCTCAggatcc	20.4
111	atcgTTGACCGAGTATCCACCTCCggatcc	19.4
71	atcgTTGACCGAGCCCCCCTGTTTggatcc	18.8
26	atcgTTGACCGCGCTGCCACCCCTggatcc	18.6
61	atcgTTGACCGAGGTGATCTCTATggatcc	18.4
64	atcgTTGACCCGCGTCTGCTCCCGggatcc	17.9
141	atcgTTGACCGAGTACCTACACCGggatcc	17.3
32	atcgTTGACCGAGCCTCACCCCCCggatcc	16.9
144	atcgTTGACCGAGTTGGCGAATCTggatcc	16.6
68	atcgACGTTGACCCGCATCCGCTTggatcc	16.1
97	atcgTTGACCGAGCCCCCCTCACAggatcc	15.4
81	atcgCTCGTTGACCTTGTGACTGTggatcc	14.9
140	atcgTTGACCGAGTGGAACAGCGTggatcc	14.0
143	atcgTTGACCGCGCCGCTTCCGTGggatcc	13.5
131	atcgCGTTGACCGAGCTTCCACCCggatcc	12.8
112	atcgTACACCGTTGACCGCCCTCCggatcc	12.6
52	atcgTTGACCCCACGGACGCTAATggatcc	12.1
104	atcgTTGACCGAGCCCCACACCCGggatcc	11.9
114	atcgCAACGTTGACCGCGCCCCCggatcc	10.8
105	atcgTTGACCGAGCCTCTCAGCCCggatcc	10.3
95	atcgAAAGATACCCGCGCTCCACGggatcc	9.9
130	atcgTTGACCACGCGCGCACCCTGggatcc	9.9
56	atcgAACCAACGACTTTGACGCCCggatcc	8.8
55	atcgCGTTGACCCCCCCACATCC.ggatcc	6.9
119	atcgTTGACATCGTCTAGCTGATTggatcc	4.0
23	atcgGTTGACCGGCGCCCCCCCCTggatcc	3.6
107	atcgGTGCAGACTTTTTTACGTTTggatcc	2.1
CONSENSUS	CGTTGACCGAG	

Figure 5. Compilation of ZAP1 DNA binding sites identified by random binding site selection. The sequences are aligned to show the homology with the consensus sequence shown at the bottom. Nucleotides derived from the random part of the initial probe are given in upper case, flanking nucleotides in lower case and nucleotides homologous to the consensus sequence are bold. Clone numbers are indicated at the left and the relative affinities at the right. The sequences are arranged from high affinity to low affinity.

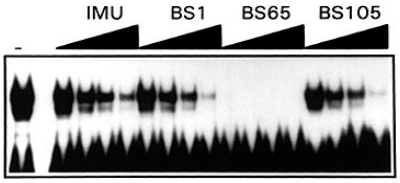

Figure 6. Relative binding affinities of ZAP1 for IMU, BS1, BS65 and BS105. Crude extract from *E.coli* expressing ZAP1 was incubated with labeled IMU and increasing amounts of the unlabeled competitors IMU, BS1, BS65 or BS105. The competitors were added in 12.5-, 25-, 50- and 100-fold molar excess.

Localization of DNA-binding domain and activation domain

The C_2-H_2 zinc-finger motif is usually present as tandem repeats (2). Most of the members of the class of DNA-binding proteins to which ZAP1 belongs also contain a repeated zinc-finger motif.

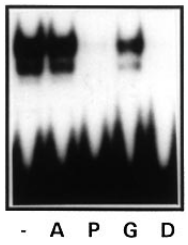

Figure 7. Effect of metal-ion chelators on DNA binding activity of ZAP1. Crude extract from *E.coli* expressing ZAP1 was preincubated with 10% ethanol (A), 10 mM 1,10-*o*-phenanthroline in 10% ethanol (P), 50 mM EGTA (G) or 50 mM EDTA (D) for 10 min at 30°C before binding to the labeled BS65 binding site.

ABF2 from wild oat, however, contains only one zinc-finger motif (7). SPF1 contains two motifs but only the second one is required for DNA binding (6). We tested whether both motifs of ZAP1 are necessary for DNA binding, employing *in vitro* generated proteins. The wild-type and deletion constructs (D2-D6) shown in Figure 8 were introduced in an *in vitro* transcription/translation system and labeled proteins were analysed on a 10% SDS–polyacrylamide gel (Fig. 9A). In all cases proteins of the correct size were produced. With the shortest construct (D6) an additional polypeptide with lower mobility was visible, which was also produced in low amounts in the transcription/translation system programmed with the other *Zap1* constructs and with unrelated genes and is not encoded by the cDNA insert.

Equimolar amounts of these proteins were used in a gel shift assay. The optimal binding site for ZAP1 (BS65; Fig. 5) was used as probe. The amount of complex formed with the full-length protein (Fig. 9B) was set at 100% (Fig. 8). The only other protein that gives a visible complex is encoded by D3. It forms 42% of the wild-type amount of complex. In D3 the N-terminal conserved sequence is deleted and the C-terminal one is intact. The D2 protein, in which the C-terminal repeated sequence is deleted but the N-terminal one is intact, forms only 3% of the wild-type complex and D5 protein, also missing the C-terminal repeated sequence, forms 2% of the wild-type complex. These values are in the range of background levels. The other constructs (D4 and D6), which are missing both conserved regions also do not produce proteins that are able to bind to the probe. From these results we conclude that the C-terminal conserved region is essential for DNA binding and that the N-terminal one increases the binding 2.5-fold. The N-terminal conserved region alone does not allow complex formation above background level.

Having determined the DNA binding domain of ZAP1, we next investigated whether ZAP1 is able to activate transcription in yeast cells. Examination of the amino acid sequence of ZAP1 does not reveal obvious homology to known activation domains found in other transcription factors. N-terminal truncated proteins (D7–D9; Fig. 8) were coupled to the DNA-binding domain of GAL4 (22), and activation of a *Gal4–LacZ* reporter gene was tested in yeast. The D7 and D8 constructs containing the second half of the ZAP1 protein were able to give activation (Fig. 8). Further deletion of the first few amino acids of the second conserved sequence (D9) abolished activation of the reporter gene. We conclude that the activation domain is located just N-terminally of or in the second repeat and might extend to the C-terminal end.

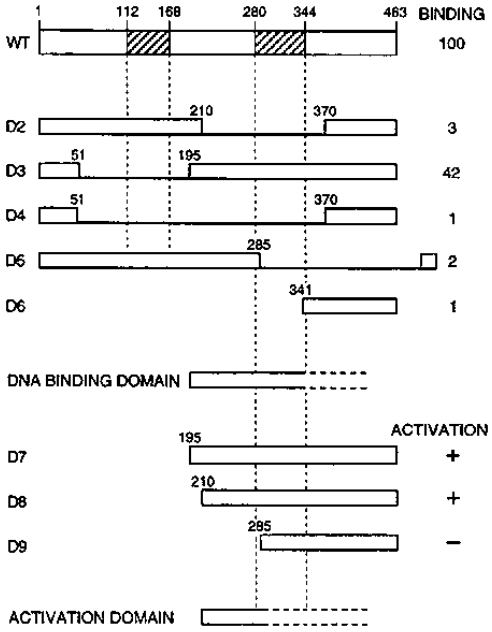

Figure 8. Effect of deletions on DNA binding and transcriptional activation of ZAP1. Deletion constructs used to map the DNA binding and activation domains of ZAP1 are shown. Bars represent the parts of the protein that are present in the constructs, thin lines the parts that are deleted. The two conserved regions are hatched. The numbers indicate the positions of the repeated sequences and the end points of the deletions. The relative amounts of DNA–protein complexes formed between proteins WT and D2 to D6 and the binding site BS65 (Fig. 9B) are shown at the right. The constructs D7, D8 and D9 contain the indicated parts of ZAP1 coupled to the C-terminus of GAL4 DNA binding domain. The presence (+) or absence (–) of transcriptional activation following introduction of these constructs in yeast is indicated at the right. The conclusions about the positions of the DNA binding domain and activation domain are indicated.

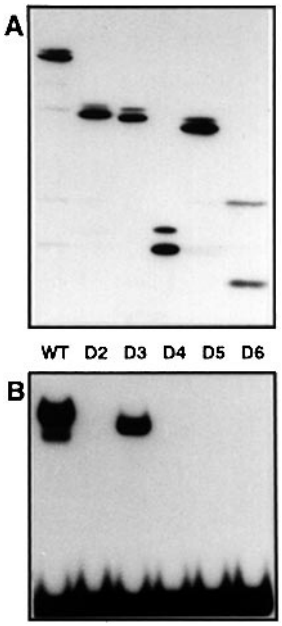

Figure 9. DNA binding of deletion mutants of ZAP1. (**A**) Coupled transcription-translation extracts were programmed with the constructs indicated. The ^{35}S-labeled proteins were analysed on an SDS polyacrylamide gel. (**B**) Equimolar amounts of protein were used for gel shift analysis with ^{32}P-labeled BS65 probe.

Thus, the activation domain is located proximal to and possibly overlaps with the DNA binding domain.

Transcriptional activation by ZAP1 in plant cells

ZAP1 binds DNA *in vitro* and activates gene expression *in vivo* in yeast. To test whether ZAP1 can activate gene expression in plant cells, particle bombardment experiments were done with *C.roseus* cell suspensions. The reporter constructs (Fig. 10A) contained a monomer or multimers of BS65 (Fig. 5) coupled to the TATA box of the CaMV 35S promoter and the *gus*A reporter gene (GusSH-47). GusSH-47 and 35S-GUS were used as negative and positive control constructs, respectively. The reporter constructs were introduced in *C.roseus* cells together with *ZAP1* cDNA under control of the CaMV 35S promoter or together with an empty vector in order to keep the amount of DNA constant.

The negative control construct GusSH-47 gives low background staining (Fig. 10C). This value was subtracted from all other values (Fig. 10B). One copy of BS65 on GusSH-47 increased the basal level of expression (Fig. 10B). This value was taken as reference for the other expression levels and set at 100. The basal expression level remained low after multimerization of the BS65 binding site. Simultaneous introduction of 35S–ZAP1 and the BS65–GusSH-47 reporter gene increased the GUS expression level 2.4-fold compared to BS65–GusSH-47 alone. This transactivation by ZAP1 was dependent on the copy number of the BS65 binding site and was most clear with three copies, resulting in a 6.3-fold increase (Fig. 10B and C). Further increase of the copy number did not have much effect. No significant increase was observed with 35S–ZAP1 and GusSH-47 compared to GusSH-47 alone, indicating that the transactivation depended on the presence of BS65 binding sites. Bombardment with the positive control construct 35S–GUS (Fig. 10C) resulted in 2.3-fold higher GUS expression compared with ZAP1-activated expression of 3BS65–GusSH-47. Thus ZAP1 activates expression via its DNA binding site almost to the level of the positive control.

DISCUSSION

This paper describes the characterization of the DNA binding protein ZAP1 from *Arabidopsis*. The *Zap1* cDNA clone was isolated by SouthWestern screening using a mixture of bZIP protein binding sites. ZAP1 is not a bZIP protein but a putative member of a distinct class of DNA binding proteins: the C_2-H_2 zinc-finger proteins. The optimal binding site for ZAP1 has a 4 bp homology with the probes used for the screening of the library. Low affinity of ZAP1 for these probes made it possible to isolate the *Zap1* cDNA during this SouthWestern screening experiment.

The sequence of the ZAP1 DNA binding domain and the requirement of transition metal ions for DNA binding suggests that a zinc-finger is formed. Two zinc-finger motifs are found in ZAP1 and several other members of the class of proteins to which ZAP1 belongs. Nevertheless, only one motif is required for binding [Fig. 8 and (6)]. DNA binding activity was also observed for a truncated derivative of the *Petunia* EPF1 protein containing only the C-terminal zinc-finger (10). In this respect, these proteins differ from other zinc-finger proteins, in which single base-pair mutations in one of three zinc-finger modules present in the DNA-binding domain abolished binding to their cognate binding sites (29,30,31). Whereas zinc-fingers generally are

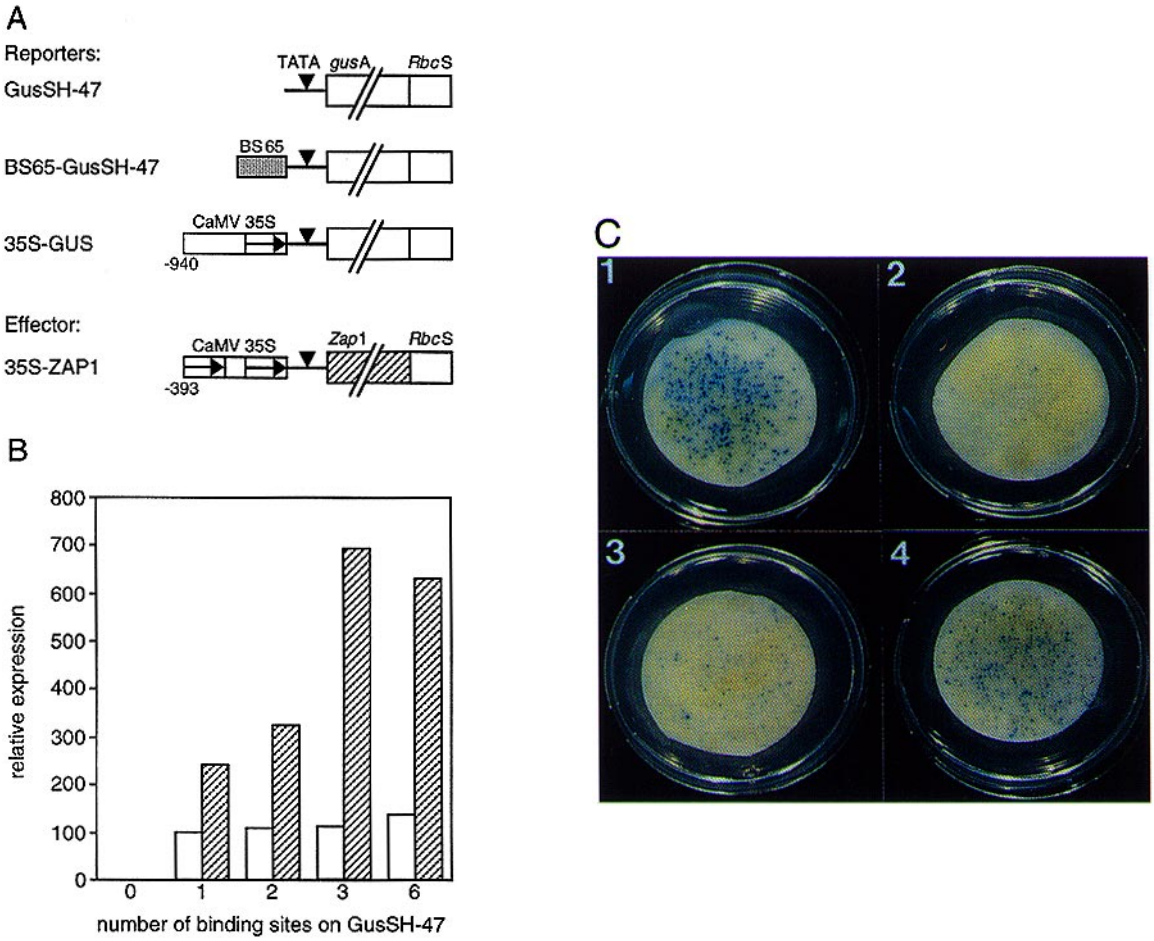

Figure 10. Transcriptional activation by ZAP1 in plant cells. (**A**) Structure of GUS reporter and ZAP1 effector constructs. BS65–GusSH-47 derivatives contain 1–6 copies of the BS65 binding site (grey box). The ZAP1 cDNA (hatched box) is under control of a CaMV 35S promoter containing a double enhancer. TATA-box is indicated with an arrowhead. The terminator is derived from the pea *RbcS*-3C gene. GusSH-47 derivatives and 35S–GUS specify identical mRNAs. (**B**) Relative GUS expression in *C.roseus* suspension cells after introduction by particle bombardment of the reporters GusSH-47 or BS65–GusSH-47, together with pBluescript SKII (open bars) or the effector 35S–ZAP1 (hatched bars). (**C**) Histochemical GUS staining of bombarded *C.roseus* suspension cells. Thin layers of cells on filter paper were bombarded with 35S–GUS (1), GusSH-47 and pBluescript SKII (2), 3BS65–GusSH-47 and pBluescript SKII (3) and 3BS65–GusSH-47 and 35S–ZAP1 (4) and stained with X-gluc.

adjacent to each other, in ZAP1, SPF1 and EPF1 they are separated by long stretches of 140, 145 and 61 amino acids, respectively, and therefore possibly function as independent DNA binding units. The regions N-terminal of the zinc-finger motifs of ZAP1 and SPF1 are highly conserved and are probably also involved in DNA binding.

For both SPF1 and ZAP1 the C-terminal conserved region is essential for DNA binding, whereas deletion of the N-terminal one only has a minor effect. Apparently, the C-terminal region or its flanking regions contain some extra information that is missing in the other one. Examination of the structural features of the conserved regions according to the Chou–Fasman (32) and Garnier–Osguthorpe–Robson (33) methods reveal similar patterns. Both contain a stretch of amino acids that probably forms an α-helix followed by a stretch that probably forms a β-sheet. These structures are conserved in the repeats of SPF1. However, outside the repeat the three dimensional structure diverges. Thus the difference might reside in the flanking regions of the repeats. On the other hand a single amino acid difference might cause the first conserved sequence to be non-functional. This has also been found for DNA-binding regions of bZIP proteins (34). Exchanging (parts of) both conserved sequences may answer the question whether the functional difference is located in the flanking regions or within the conserved sequences themselves.

The region between 210 and 285 contains (part of) the activation domain, since its deletion from a GAL4–ZAP fusion resulted in loss of activation (Fig. 8). The activation domain of ZAP1 is not acidic (pI 7.9), proline-rich or glutamine-rich, as is often observed for activation domains of other proteins. Possibly the secondary structure determines its activation ability. It has been suggested for the activation domain of GAL4 that it forms a β-sheet structure (35). This putative structure rather than the negative charge is thought to be important for its function (36). The region in ZAP1 that activates transcription contains a stretch of 13 amino acids (270–282), which might form a β-sheet (32,33). In analogy with the model for GAL4, these β-sheet structures in ZAP1 could form the activation domain.

The homology between the DNA binding domains of ZAP1 and related proteins suggests that the proteins bind to similar DNA sequences. The optimal binding site for ZAP1 is CGTTGACCG-

AG. This sequence shows similarity to the binding site for ABF1 and ABF2 in the α-amylase promoter (ATTGACTTGACCG-ATCGG) (7) and to the binding site of SPF1 in the sporamin promoter (TGAAGATTTAACCTACACAGTGATTGACCGGA-TC) (6). However, mutations in the SPF1 binding site outside the TTGACCG sequence decreases the binding affinity (6). Since the optimal binding site has not been determined, conclusions about the precise sequence requirements for SPF1 binding sites cannot be drawn. Many sequences selected for ZAP1 contain a C-rich region in addition to the consensus sequence (Fig. 5). This suggests that for ZAP1 the sequences outside the consensus sequence contribute to the binding affinity.

Although ZAP1 and related proteins recognize similar DNA sequences, they may be involved in different biological processes, since the parts of the proteins outside the DNA-binding domains are not homologous and may have different functions. Furthermore, their expression patterns are different. *ZAP1* is mainly expressed in root and flower (Fig. 2). *ABF1* and *ABF2* mRNAs are accumulating in germinating seed (7), where the encoded proteins are thought to control α-amylase expression. Although SPF1 binds to the promoters of sweet potato tuber-specific genes encoding β-amylase and sporamin genes, it is ubiquitously expressed (6).

Comparison of the ZAP1 consensus binding sequence with plant promoters revealed a striking homology with an elicitor-responsive element (AATTGACC) of a maize *PRms* gene (37). A similar sequence is present in promoters of several other genes that are expressed in response to fungal infection and elicitor treatment (37). This could mean that ZAP1 or a related protein is involved in the expression of elicitor inducible genes. An indication about the biological process in which ZAP1 plays a role might be obtained by ectopic expression or inactivation of the gene.

The characterization of ZAP1 described in this paper contributes to the knowledge of this novel class of zinc-dependent DNA binding proteins. Since *Arabidopsis* contains at least two other genes with regions homologous to the conserved zinc-finger motif (EST sequences) it would be interesting to study these as well as homologous genes from other organisms in more detail.

Note added in proof

While this paper was in press, three cDNAs from parsley were reported encoding members of the class of zinc-finger DNA binding proteins described in this paper. The parsley proteins bind to elicitor response elements in the promoters of PR1 genes and their mRNA levels show rapid changes upon elicitor treatment [Rushton *et al.* (1996) *EMBO J.* **15**, 5690–5700].

ACKNOWLEDGEMENTS

We thank Annemarie Meijer for pAS1ΔNco and other control plasmids and Peter Hock and the Biological Photographic Service, Leiden University, for preparation of the figures. SdP and KP were supported by the Netherlands Organization for Applied Scientific Research TNO, VG by a fellowship of the Consiglio Nazionale delle Ricerche (CNR) and JM by a Fulbright fellowship.

REFERENCES

1 Conaway,R.C. and Conaway,J.W. (1993) *Annu. Rev. Biochem.*, **62**, 161–190.
2 Struhl,K. (1989) *Trends Biochem. Sci.*, **14**, 137–140.
3 Klug,A. and Schwabe,J.W.R. (1995) *FASEB J.*, **9**, 597–604.
4 Sánchez-García,I. and Rabbitts,T.H. (1994) *Trends Genet.*, **10**, 315–320.
5 Meshi,T. and Iwabuchi,M. (1995) *Plant Cell. Physiol.*, **36**, 1405–1420.
6 Ishiguro,S. and Nakamura,K. (1994) *Mol. Gen. Genet.*, **244**, 563–571.
7 Rushton,P.J., MacDonald,H., Huttley,A.K., Lazarus,C.M. and Hooley,R. (1995) *Plant Mol. Biol.*, **29**, 691–702.
8 Foster,R., Izawa,T. and Chua,N.-H. (1994) *FASEB J.*, **8**, 192–200.
9 de Pater,S., Katagiri,F., Kijne,J. and Chua,N.-H. (1994) *Plant J.*, **6**, 133–140.
10 Takatsuji,H., Mori,M., Benfey,P.N., Ren,L. and Chua,N.-H. (1992) *EMBO J.*, **11**, 241–249.
11 Kaptein,R. (1991) *Curr. Opin. Struct. Biol.*, **2**, 109–115.
12 Tjian,R. and Maniatis,T. (1994) *Cell*, **77**, 5–8.
13 Buratowski,S. (1995) *Science*, **270**, 1773–1774.
14 Sharp,P.A. (1991) *Nature*, **351**, 16–18.
15 Goff,S.A., Cone,K.C., and Fromm,M.E. (1991) *Genes Dev.*, **5**, 298–309.
16 Sablowski,R.W.M., Moyano,E., Culianez-Macia,A., Schuch,W., Martin,C. and Bevan,M. (1994) *EMBO J.*, **13**, 128–137.
17 de Pater,S., Pham,K., Chua,N.-H., Memelink,J. and Kijne,J. (1993) *Plant Cell*, **5**, 877–886.
18 Salinas,J., Oeda,K. and Chua,N.-H. (1992) *Plant Cell*, **4**, 1485–1493.
19 Genetics Computer Group (1994) *Program Manual for the Wisconsin Package*, Version 8.
20 Memelink,J., Swords,K.M.M., Staehelin,L.A. and Hoge,J.H.C. (1994) In Gelvin,S.B. and Schilperoort,R.A. (eds), *Plant Molecular Biology Manual.* Kluwer Academic Publishers, Dordrecht, chapter F, pp 1–23.
21 Blackwell,T.K. and Weintraub,H. (1990) *Science*, **250**, 1104–1110.
22 Durfee,T., Becherer,K., Chen,P.-L., Yeh,S.-H., Yang,Y., Kilburn,A.E., Lee,W.-H. and Elledge,S.J. (1993) *Genes Dev.*, **7**, 555–569.
23 Pasquali,G., Ouwerkerk,P.B.F. and Memelink,J. (1994) *Gene*, **149**, 373–374.
24 Gordon-Kamm,W.J., Spencer,T.M., Mangano,M.L., Adams,T.R., Daines,R.J., Start,W.G., O'Brien,J.V., Chambers,S.A., Adams,W.R.,Jr., Willetts,N.G., Rice,T.B., Mackey,C.J., Krueger,R.W., Kausch,A.P. and Lemaux,P.G. (1990) *Plant Cell*, **2**, 603–618.
25 Jefferson,R.A. (1987) *Plant Mol. Biol. Rep.*, **5**, 387–405.
26 Lam,E., Benfey,P.N., Gilmartin,P.M., Fang,R.-X. and Chua,N.-H. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7890–7894.
27 Izawa, T., Foster, R., Nakajima, M., Shimamoto, K. and Chua, N.-H. (1994) *Plant Cell*, **6**, 1277–1287.
28 de Pater,S., Pham,K., Memelink,J. and Kijne,J. (1996) *Mol. Gen. Genet.*, **250**, 237–239.
29 Desjarlais,J.R. and Berg,J.M. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 7345–7349.
30 Desjarlais,J.R. and Berg,J.M. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 2256–2260.
31 Nardelli,J., Gibson,T.J., Vesque,C. and Charnay,P. (1991) *Nature*, **349**, 175–178.
32 Chou,P.Y. and Fasman,G.D (1978) *Adv. Enzymol.*, **47**, 45–147.
33 Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97–120.
34 Koldin,B., Suckow,M., Seydel,A., von Wilcken-Bergmann,B. and Müller-Hill,B. (1995) *Nucleic Acids Res.*, **23**, 4162–4169.
35 Van Hoy,M., Leuther,K.K., Kodadek,T. and Johnston,S.A. (1993) *Cell*, **72**, 587–594.
36 Leuther,K.K., Salmeron,J.M. and Johnston,S.A. (1993) *Cell*, **72**, 575–585.
37 Raventós,D., Jensen,A.B., Rask,M.-B., Casacuberta,J.M., Mundy,J. and San Segundo,B. (1995) *Plant J.*, **7**, 147–155.