© 1994 Oxford University Press

Nucleic Acids Research, 1994, Vol. 22, No. 19 3983–3989

Plant activating sequences: positively charged peptides are functional as transcriptional activation domains

Juan J.Estruch*, Lyle Crossland and Stephen A.Goff
CIBA Agricultural Biotechnology, Research Triangle Park, NC 27709-2257, USA

Received May 3, 1994; Revised and Accepted August 5, 1994 GenBank accession nos+

ABSTRACT

Plant sequences that act as transcriptional activation domains in yeast as well as in plants have been isolated by genetic selection in yeast. The selection was based on the reconstitution of a functional GAL4 transcriptional activator. Since the peptides show no homology with reported activation domains, they represent a new class of activating sequences. The sequence P1, which is 10 amino acids long, is the shortest functional activation domain reported. A cDNA that encodes the P14 class (peptides P14–P18) activating sequence have been cloned. The protein exhibits strong homology (higher than 50% amino acid identity) with the BBC1-related sequences, a highly conserved family of basic proteins containing nuclear localization signals. The P14 and P15 peptides are the most effective plant activating sequences. The P14 and P15 peptides are highly hydrophilic, positively charged and mostly unstructured. These properties are at odds with the ones usually found in known activation domains.

INTRODUCTION

Research on transcription regulation has been centered on determining how transcriptional factors modulate transcription of their target genes. Transcription factors possess a DNA-binding domain which mediates recognition and binding to specific upstream activation sequences (UAS) of the target promoter. Transcription factors also have a second and independent domain called activation domain, required for activation. Activation domains are believed to bridge transcription factors with the transcriptional complex, but how they exert their function remains unknown. A direct interaction between the activation domain of the herpes simplex virus protein, VP16, with TATA-box binding proteins has been reported (1). Other studies suggest that the interaction between the activators and general transcription factors is mediated by other proteins termed adaptors (2).

The structural and sequence requirements for activation domains are unclear. Activation domains are generally though to be amphipathic alpha helices or unstructured regions with high density of negative charges (3,4). It has been shown that small *Escherichia coli* DNA fragments can encode activation domains which are functional in yeast (5). The peptides encoded by these *E.coli* DNA fragments have no similarities other than a net negative charge. In addition, a synthetic peptide which is negatively charged, and forms an amphipathic alpha helix can activate transcription both *in vivo* and *in vitro* (6). A detailed study performed on the VP16 activation domain showed that negative charge contributes to, but is not sufficient for transcriptional activation in mouse cells (7). Two recent reports (8,9) provide genetic and functional evidence that the secondary structure of the GAL4 activation domain is indeed relevant while net charge plays a secondary role. GAL4 activation domain does not adopt an amphipathic alpha helix as suggested previously, but most likely an antiparallel β-sheet.

In recent years, the characterization of transcriptional factors has been extended to plants. Plant transcriptional factors have been isolated through mutant studies, *cis*-acting element analysis, and homology to known animal transcriptional factors (for review see reference 10). Functional analysis of the B and C1 transcriptional activators demonstrate that the C1 gene contains an activation domain and can activate GAL4 promoter-regulated genes when fused to the GAL4 DNA-binding domain (11,12). The activation domain of the C1 gene has a 70% amino acid similarity with the VP16 activation domain (S.Goff, unpublished results). The maize B and C1 genes are involved in the regulation of the anthocyanin biosynthetic pathway (13).

We have used a genetic strategy to isolate plant activation domains. We fused tobacco cDNAs to the GAL4 DNA-binding domain and screened for the reconstitution of a functional GAL4 transcriptional activator. We isolated 18 clones representing six different activation domains. We report that sequences as short as 10 aa act effectively as activation domains. The stronger activation domains isolated from tobacco plants are hydrophilic peptides with a net positive charge. Our results suggest that the activator regions may have very limited sequence requirements.

MATERIALS AND METHODS

Strains and media

The yeast strains used were GGY1::171 (Δgal4, Δgal80, ura3, his3-200, leu2-3)(14) and CTY1 (Δgal4, Δgal80, his3-200,

*To whom correspondence should be addressed

+L31415, L31416 and L33874–L33877 (incl.)

leu2-3, lys 2-801, trp1-901)(15). GGY1::171 bears a GAL1-*lacZ* fusion integrated in URA3 and CTY1 harbors, in addition, a GAL1-*his3* fusion integrated in LYS2. Yeast were grown in YPD rich-medium (1% yeast extract, 1% bactopeptone, 1% dextrose) or SD minimal medium (0.67% yeast nitrogen base without aa, 1×aa minus leucine or histidine), and the carbon sources of minimal media were 2% glucose, or 2% galactose or 2% galactose, 2% glycerol, and 2% ethanol. β-galactosidase activity was assayed on SDG plates which contained 0.67% yeast nitrogen base without aa, 1×aa minus leucine, 2% bactoagar, 20 mg of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal), 0.1 M potassium phosphate and 2% sucrose as carbon source.

Construction of GAL4 activation domain plasmids

Plasmid pSKGAL2 (11) contains the GAL4 DNA-binding domain (aa1−147). Plasmid pGALC1 (11) contains the GAL4 DNA-binding domain from pSKGAL2 fused to the acidic region of C1 (aa173−273). Oligonucleotides containing *Bam*HI sites and termination codons in all three frames were inserted into the *ClaI*−*KpnI* sites of the pSKGAL2. The GAL4 DNA-binding domain together with the oligonucleotides were taken as *Eco*RI−*KpnI* fragments and cloned into the *Eco*RI−*KpnI* sites of the yeast shuttle vector pSGBCU11 to generate the three pPAS vectors (Fig. 1A).

Construction of GAL4 protein fusions with plant cDNAs

RNA was purified from 3 week old tobacco (*Nicotiana tabacum*, cv Xanthi) plants as described in (16). Poly(A$^+$) RNA was purified by two cycles of chromatography on oligo(dT)-cellulose and cDNAs were generated using oligo(dT) as primers and reverse transcriptase from avian myeloblastosis virus (16). After the addition of *Bam*HI linkers, the resulting cDNAs were ligated to the pPAS vectors that have been previously digested with *Bam*HI and dephosphorylated with calf intestinal alkaline phosphatase. Ligation mixes were used to transform *E.coli* DH5alpha by electroporation. About 360 000 independent clones were obtained for each of the pPAS vectors. After growth on ampicillin-containing plates overnight, replica-plates were performed and colonies were washed off the plates with LB liquid. Plasmid isolation on these colonies was done using a Promega Magic Miniprep kit and used to transform yeast.

Screening of plant activating sequences

The GAL4 protein fusion library was transformed by electroporation into the Leu$^-$ yeast strain GGY1::171 (17). After 3 days growth on SD−Leu medium, transformants were replica-plated to SDG plates and incubated until blue colonies appeared (from 1 to 8 days). Blue colonies were grown in liquid SD−Leu medium for 2 days at 28°C. Plasmid DNA was isolated (18), and used to retransform GGY1::171 and CTY1 yeast strains. Transformed GGY1::171 were assayed for β-galactosidase activity, and ability to grow on SD−Leu using galactose as the only carbon source. Transformed CTY1 were plated on SD−Leu−His plates and scored for ability to grow.

Northern blot analysis

Poly(A$^+$) RNA was isolated from 3 week old tobacco plants as described in (16). Two μg of poly(A$^+$) RNA were electrophoresed in 1.2% agarose gels containing 16% formaldehyde and blotted on to nitrocellulose membranes (Micron Separation Inc.). The blotted membranes were hybridized in 5×SSC, 10×Denhardt's solution, 0.1% SDS, 0.1 mg/ml of denaturated salmon sperm DNA and 50% (v/v) deionized formamide at 42°C for 16 h. ^{32}P-random labeled plant activating sequences P1, P5, P8, P11, P12 and P14 were used as probes. The membranes were washed twice in 1×SSC, 0.1% SDS at 50°C for 15 min, and once in 0.1×SSC, 0.1% SDS at 45°C for 5 min.

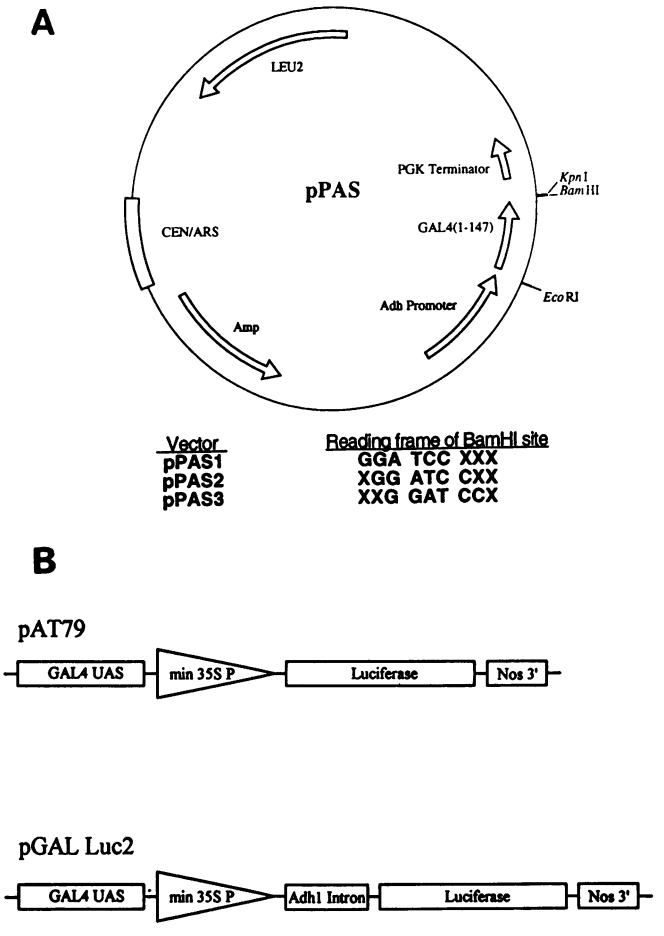

Figure 1. Scheme of pPAS vectors and reporter genes. (**A**) pPAS is a 6.8 kb vector which expresses the DNA-binding domain of the GAL4 (residues 1−147) from the ADH1 promoter; an oligonucleotide sequence containing a *Bam*HI site and stop codons was placed adjacent to the coding sequence of the GAL4. Plant cDNAs were inserted at this *Bam*HI site which is unique on pPAS vectors. The GAL4 DNA-binding domain plant cDNAs fusions were recovered as *Eco*RI−*KpnI* double digests. The three vectors differ in the reading frame of the *Bam*HI cloning site. pPAS vectors also contain a replication origin of M13, ampicillin drug resistance, yeast LEU2 selection marker, and yeast CEN/ARS. (**B**) pAT79 and pGALLUC2 are two pUC18-derivative plasmids containing the luciferase as reporter gene. Both plasmids contain a GAL4 UAS composed of 10 GAL4 binding sites (28) placed upstream of a minimal ^{35}S promoter (−1, −35) and the *nos* terminator. In addition, pGALLUC2 posses the ADH1 intron positioned between the promoter and the coding region of the luciferase.

Random mutagenesis by polymerase chain reaction (PCR)

The plant activating sequece P14 was mutagenized by PCR amplification with limiting dATP and substitution of Mn^{2+} for Mg^{2+} as described in (19). The following oligonucleotide primers were used: 5′-CGGAAGAGAGTAGTAACAAT; 5′-GCCGCGGCCGCTGCTGTTCAAATTAGGT. The resulting

PCR products were subcloned into pPAS vectors as *Bam*HI−*Not*I. Yeast transformation was performed as described previously (17).

Assay of β-galactosidase

β-galactosidase activity was assayed as described in (20). Exponentially growing cultures were permeabilized with sodium dodecyl sulfate and chloroform. Protein concentrations were determined prior to reactions using a Bio-Rad protein assay. Assays were run in triplicate and the standard error was below 20%.

Plant vectors

The plasmids pAT79 and pGALLUC2 (Fig. 1B) which contain luciferase as reporter gene were used in the bombardments of tobacco leaf discs and maize suspension cultures respectively. GAL4 DNA-binding domain plant activating sequence fusions were cut out of pPAS vectors as *Eco*RI−*Kpn*I fragments and cloned into pCIB710 (21) for tobacco bombardment or into pMF6 (22) for maize culture cells bombardment.

Microprojectile bombardment and luciferase assay

Two μg of CsCl-purified plasmid DNA was precipitated on to microprojectiles by the addition of 50 μl of gold particles (1.0 μm diameter, 60 mg/ml), followed by the addition of 50 μl of 2.5 M $CaCl_2$ and 20 μl of 1.0 M spermidine free base. Approximately 0.3 μg of plasmid DNA per sample was loaded on to each macrocarrier. Leaf discs (2 cm in diameter) were cut from young (10−15 cm long) tobacco plants and placed on Murashige−Skoog media in Petri dishes and bombarded within 4 h. Maize embryogenic callus tissue (line L6) was prepared for gene transfer by spreading a thin lawn of tissue on agarose plates containing N6 media (23). Luciferase assays were performed on the bombarded tissue about 36 h after bombardment. Luciferase activity was measured using a Promega Kit and is expressed as light units detected by an Analytical Luminiscence for 10 s at 25°C.

Isolation of the BBC1-tobacco homologue

cDNAs obtained during the construction of the GAL4 protein fusions with plant cDNAs were used to make a tobacco cDNA library. After addition of *Eco*RI linkers, the resulting cDNAs were ligated to lambda gt10 arms and packaged as described by Stratagene protocols. About 6×10^5 independent clones were obtained. 2×10^5 plaques were screened with either the 165 bp *Bam*HI fragment corresponding to the P8 or the 80 bp *Bam*HI fragment of the P14. One positive clone was isolated with each probe. After three rounds of purification, the inserts were subcloned into Bluescript SK(+) and sequenced using the *Taq* cycle sequencing (Applied Byosystems). All sequence analysis and comparison were performed with the University of Wisconsin Genetic Computer Group Program.

RESULTS AND DISCUSSION

Experimental design

Genetic selection in yeast has been used to isolate an olfactory neuronal transcription factor (24) and a component of the yeast origin recognition complex (25). These strategies are based upon fusion of random DNA segments to the activation domain of the GAL4 gene. Clones expressing functional DNA binding regions are identified using a reporter gene whose expression is driven by a promoter with specific DNA-binding sequences. We utilized a modification of this strategy to identify potential plant transcriptional activation domains. We constructed a hybrid expression library by fusing the GAL4 DNA binding domain (aa1−147) to random DNA sequences obtained from tobacco cDNAs. Fusion proteins containing plant sequences that act as activation domains will activate a reporter gene with GAL4 binding sites in its promoter. The hybrid expression library was constructed in a yeast shuttle vector (Fig. 1A) and transformed into *E.coli* by electroporation. The plasmids were then rescued and retransformed into the yeast strain GGY1::171 that lacks a functional *GAL4* gene but contains an integrated *GAL1*−*lacZ* reporter gene with the normal UAS (5). Hybrid GAL4−plant protein fusions which activate the GAL1 promoter result in expression of β-galactosidase. This confers a distinctive blue color to the yeast colonies.

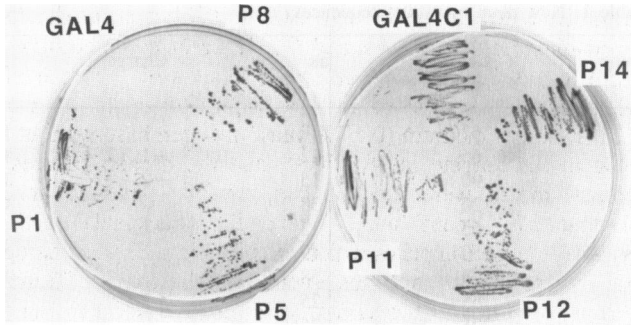

Figure 2. Growth of yeast cells in galactose. Plasmids expressing various plant activating sequences, a GAL4 DNA-binding domain or a GAL4C1 fusion were transformed into a yeast strain lacking functional GAL4 (GGY1::171). The yeast cells were first grown in liquid culture with glucose as carbon source, and after 2 days they were streaked on to SD−Leu plates supplemented with galactose as the sole carbon source. The plates were incubated for 2 days at 28°C. Only those cells expressing all the chromosomal GAL genes could grow in this test. Growth was detected after less than 1 day for yeast cells bearing GAL4C1 and GAL4P8 fusions. The rest of the fusions took 2 days to grow. GAL4 bearing yeast cells did not grow even after 4 days when the picture was taken.

Using the technique described above, we detected 18 blue colonies among about 1.5×10^6 yeast transformants. The plasmids were isolated from all blue colonies and retransformed into yeast. β-galactosidase activity was associated with presence of the plasmid in each case. The following additional tests were performed. The plasmids enable the yeast strain GGY1::171, which lacks functional *GAL4* to grow in medium containing galactose as the only carbon source (Fig. 2). This indicates that the hybrid protein fusions were able to activate all chromosomal *GAL* genes required for galactose metabolism. The plasmids were also introduced into the yeast strain CYT1, which is auxotrophic for histidine but contains the *GAL1*−*his3* gene fusion integrated in its chromosome. The transformed yeast were able to grow in minimal medium lacking histidine (data not shown). The removal of the GAL4 DNA-binding domain from the fusion constructs abolished all phenotypes described above.

We isolated and sequenced the 18 DNA fragments that act as activation domains. The predicted amino acid sequences for the 18 DNA fragments are shown in Table 1 (26). The length of the activating sequences varied from 10 aa (P1) to 104 aa (P18). Although the clones were recovered independently, some of them represent polypeptides of different lengths but originating from the same transcript. Indeed, the 18 sequences can be clustered

Table 1. New plant activating sequences

Name	Length bp	aa	Charge	Amino acid sequence
P1	62	10	−1	EQYQPLSWLL
P2	65	12	−1	FADIF-$_{P1}$
P3	90	15	−2	IF-$_{P1}$
P4	120	38		QLKGRFPPKAARTVADIAQRCLQ-$_{P2}$
P5	65	13	−1	ILPMSWLLEMGVS
P6	150	49		SSKLTSLKGTKAGTGPPLEVRGIDKLDIEFLQPGG-$_{P5}$
P7	174	58		ETAMPMITP-$_{P6}$
P8	165	25	+2	KVDIISHSMGGLLIKCFISLYSDVF
P9	225	75		IDMLVKCGYKKGTTLFGLGYDFRQSNRIDKAMNDLKAKLETAYKASGGR-$_{P8}$
P10	260	85		HLTGVYHFHDM-$_{P9}$
P11	135	26	0	ISHSMGGLLVLLVKCFVDYHCHSFSS
P12	196	62	0	FRRVRSSLPAKYMVDIDVQIPSAFDP FAEAKDSGAPGAKEYVHIRIQQRNGK KSLTTVQGPD
P13	250	83		LLKWCCFKSCCVTKFEGLV-$_{P12}$
P14	80	25	+4	PKKLAPTIGIAVDHRRRNRSLEGLQ
P15	109	37	+3	PKKLAPTIGIAVDHRRRNRSLEGLQTNVQRLEDLQGQ
P16	261	87		IARQKKAVKISPRPTAGTLRPIVHGQ TLKYNMKVRSGRGFSLEELKAA-$_{P15}$
P17	309	103		YVKTWFNQPARKTRRR-$_{P16}$
P18	312	104		N-$_{P17}$

Shown are predicted plant activating sequences represented by the standard one-letter code. The sequence of the shortest peptide of each class is disclosed. The other peptides are represented by X and the number of amino acids is shown in subscripts. The charge reflects the balance between the positive charge aa (R and K) and the negative charge aa (D and E). All sequences but P14, P15, P16, P17 and P18 contained their own termination codon. The amino acids that link the DNA-binding domain of GAL4 with the plant sequences are G (pPAS1), GD (pPAS2) and GI (pPAS3). All sequences were obtained either in pPAS1 or pPAS3.

in six different classes which represent expressed gene products in 3 week old tobacco plants (Fig. 3). It is interesting to note that, Ma and Ptashne (5) also recovered the same *E.coli* sequences several times when they use a similar approach. The strengh of the selection strategy applied to recover functional activation domains could explain why the same class of peptides is recovered as independent clones. Within the same class, the longer plant activating sequences do not have stronger activation properties than their shorter versions (see Table 2), so the short sequences have all information required for activation. It is difficult to make an estimation of how frequently activation domains occur in tobacco plants. Although the sequences fused to the GAL4 DNA-binding domain must be in frame and the selection strategy is strong, the number of clones screened to find these six different classes of activating sequences represent 10 times the number required to obtain a complete representation of low abundance mRNA sequences (16). We conclude that strong activation domains which are functional in yeast are not very frequent in tobacco plants.

The level of activation produced by each sequence was quantitated by using a β-galactosidase assay (Table 2). Yeast cells carrying a GAL4-C1 containing plasmid (11) were used as a positive control. The GAL4 DNA-binding domain fused to the C1 transactivation domain activates transcription of GAL4-regulated promoters in yeast as well as in plants (11). The levels of activation achieved by GAL4-C1 are up to 40% higher than with GAL4-VP16 which is considered a powerful transcriptional activator (11). Yeast cells carrying P14 (and P15) effector plasmids contained levels of β-galactosidase activity similar to the GAL4-C1 (Table 2). Therefore, P14 (and P15) is one of the strongest transcriptional activation domains reported. Sequences as short as 10−15 amino acids (Table 1) also act as effective activating sequences in all experimental conditions used. Ma and Ptashne (5) described series of *E.coli* genomic DNA

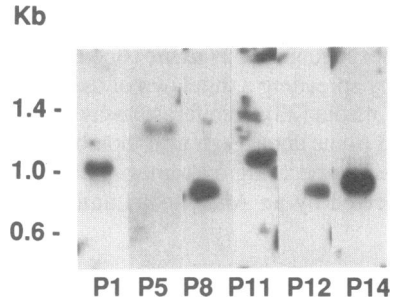

Figure 3. Analysis of the expression of the plant activating sequences by Northern blot. Each lane was loaded with 2 μg of poly(A$^+$) RNA isolated from 3 week old tobacco plants. The plant activating sequences indicated at the bottom of the blot were used as probes.

fragments that function as activating sequences when fused to the GAL4 DNA-binding domain. Some of them encode 12, 13 and 15 amino acids peptides. However, the levels of β-galactosidase induced by the *E.coli* sequences were very low (less than 1.6% in relation to the wild-type GAL4) and they did not rescue a GAL4$^-$ phenotype. Instead, the presence of the shorter plant activating sequences (P1, P2, and P3) activate the β-galactosidase to levels between 16 and 32% of GAL4-C1 levels and the yeast carrying them grew utilizing galactose as the carbon source. Therefore, P1 constitutes the shortest sequence reported to act effectively as activation domain.

Plant activating sequences identified in yeast are functional in plants

We tested whether the sequences P1, P5, P8, P11, P12 and P14 (and P15) function as activation domains in plants. The sequences

Table 2. Activities of plant activating sequences in yeast

Activator	β-galactosidase activity
P1	661
P2	580
P3	672
P4	646
P5	372
P6	340
P7	325
P8	104
P9	95
P10	92
P11	256
P12	125
P13	115
P14	2040
P15	2246
P16	2150
P17	2090
P18	2180
GAL4	4
GAL4-C1	2369

Plasmids expressing either plant activating sequences, GAL4-C1 fusion, or the GAL4 DNA-binding domain were introduced into GGY1::171 as described previously (17). β-galactosidase activities were measured from yeast cells grown in 2% galactose, 2% glycerol and 2% ethanol as carbon sources. Values of β-galactosidase activity were calculated and expressed according to (20).

Table 3. Activities of the plant activating sequences in plants

Activator	Luciferase activities ($LU \times 10^{-3}$) Tobacco	Maize
P1	52.4 ± 4.0	68.5 ± 5.5
P5	23.1 ± 3.2	12.5 ± 2.8
P8	30.8 ± 3.2	18.3 ± 1.9
P11	30.2 ± 2.6	17.5 ± 3.5
P12	32.5 ± 4.0	20.4 ± 3.0
P14	315.7 ± 24.5	480.0 ± 32.0
P15	382.5 ± 31.2	495.8 ± 29.4
GAL4	0.3 ± 0.1	0.6 ± 0.1
GAL4-C1	379.5 ± 36.0	570.0 ± 48.0

The plasmids delivered to tobacco leaf discs or maize embryogenic callus contained the sequences described in Table 1 as activating sequences. A plasmid containing the GAL4 DNA binding domain was used as negative control, while the activation domain of the C1 gene fused to GAL4 DNA-binding domain was a positive control (11). Activities are expressed in arbitrary light units (LU) and they represent three independent experiments.

were subcloned into the plant vectors pCIB710 and pMF6 (see experimental procedures) and transformed into tobacco leaf discs and maize culture cells by microprojectile bombardment. A GAL4−UAS containing promoter driving a luciferase gene was co-delivered with vectors bearing the plant activating sequences. Luciferase activity was measured 36 h after bombardment (Table 3). All activating sequences are functional in tobacco and maize, and their relative activations of the reporter gene are similar to the activation observed in yeast (compare Tables 2 and 3).

In a data bank search, only the sequences P8 (P9, P10) and P14 (P15−P18) show significant homologies with reported proteins. P8 and P14 sequences were used as probes to clone the corresponding cDNAs encoding these proteins (26). A 450 bp partial cDNA for the P8 protein was isolated. The predicted P8 protein (Fig. 4A) corresponds to a hydrophobic peptide

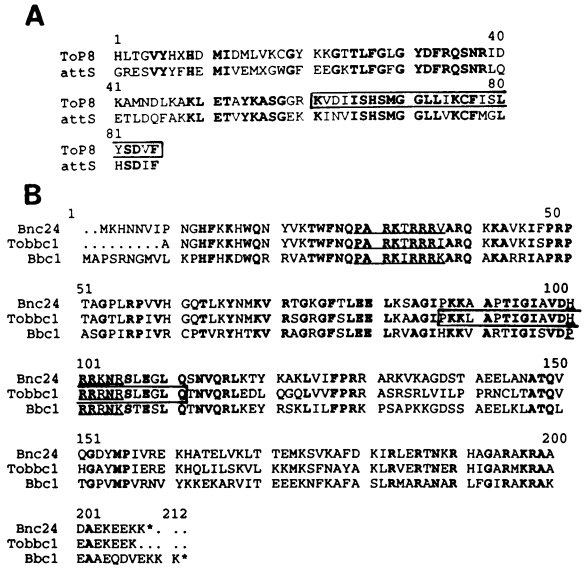

Figure 4. (**A**) Predicted amino acid sequence of P8 protein (ToP8; accession no. L31415) and alignment with an *Arabidopsis thaliana* transcribed sequence (attS). (**B**) Predicted amino acid sequence of BBC1-tobacco homolog (Tobbc1; accession no. L31416) and alignment with BnC24 protein from *Brassica napus* and BBC1 protein from mammalian cells. Identical amino acids are indicated in bold. Potential nuclear localization signals are underlined. The boxed sequence represents the portion of the protein that has been characterized as an activation domain when fused to the GAL4 DNA-binding domain.

exhibiting a 50% amino acid identity with an *Arabidopsis thaliana* transcribed sequence (GenBank accession number Z18107). The P14 protein (Fig. 4B) is 72% identical to recently described low-temperature induced genes from *Brassica* (27) and 51% identical to the BBC1 human protein (27). The BBC1 protein was identified by differential screening of a cDNA library generated from a breast carcinoma (27). The levels of the BBC1 mRNA are down-regulated in carcinomas as compared with fibroadenomas. BBC1-related sequences have been observed throughout eukaryotes and constitute a highly conserved family of basic, hydrophilic proteins containing nuclear localization signals. We have shown that the tobacco homologue for the BBC1 protein contains a strong transcriptional activating sequence. These results indicate that BBC1 proteins could interact with nucleic acids modulating transcriptional activities.

Structural patterns of the plant activating sequences

Distinct classes of activation domains have been identified on the basis of their amino acid composition; acidic-rich activation domains represented by GAL4 (29); glutamine-rich activation domains represented by Sp1 (30); metal-binding cysteine-containing activation domains of adenovirus E1a protein (31); and proline-rich activation domains represented by CTF1 (32,33). The activation sequences described in this report are not rich in acidic amino acids, glutamines, cysteines, or prolines. Thus, they cannot be classified within any of these established activation domain classes. Since, the activation domains described here show no homology with any other activation domain reported, they represent a new class of activating sequences.

Our study shows that sequences with a net positive charge can act as powerful activation domains. Indeed, the plant sequences

Table 4. Limited mutational analysis of the plant activating sequence P14 and their effects on the activation properties

	Yeast assay[a]	Tobacco assay[b]
P14wt PKKLAPTIGIAVDHRRRNRSLEGLQ	100	100
m1.....G.....	82	78
m2.....N.....	98	95
m3.....G.....G.....	90	85
m4.....L.....Q.....	85	82
m5.....G.....G.....	68	62
m6.....P.....	92	87
m7.....G.....	117	107
m8.....F.....	102	105
m9.....G.....H.....	86	81

[a] Values are given in % of β-galactosidase activity observed with the wild-type P14 (P14 wt) plant activating sequence.
[b] Values are given in % of arbitrary light units observed with the P14 wt plant activating sequence.

P14 and P15 have a net charge of +4 and +3, respectively, and activate transcription as strongly as the GAL4-C1 fusion (Tables 2 and 3). A similar observation has been reported by Leuther *et al.* (8) following mutational analysis of the GAL4 activation domain. However, the constructs analyzed by Leuther *et al.* (8) contained a non-mutagenized region of GAL4 activation domain which is placed N-terminal to the 21 amino acid region mutagenized. Since this non-mutagenized region can still function as activation domain (5), it may have contributed to the activation observed in the mutants with a net positive charge. Our approach is based on the fusion of peptides to a GAL4 DNA-binding domain free of any activation domain sequences. Thus, our results show that positively charged peptides function as activation domains.

In recent reports, Leuther *et al.* (8) and Van Hoy *et al.* (9) suggest that a β-sheet structure may be required for proper function of the GAL4 activation domain. Whether the β-structure plays a role in activation or represents active or inactive states of the protein is not known. We analyzed the predicted secondary structure of the plant activating sequences by applying the Chou and Fasman (33) program. While an unstructured domain composed of hydrophilic amino acids is predicted for P1, P5 is predicted to form an amphipathic α-helix. P8 and P11 are structured mainly as β-sheets rich in hydrophobic amino acids, and P12 and P14 are highly hydrophilic sequences exhibiting α- as well as β-structured domains. These results indicate that a certain secondary structure could not be assigned to these plant activating sequences.

In order to assess what feature(s) of the plant activating sequences is important for activation, a limited mutational analysis was performed on the P14 activation domain (Table 4). Most of the P14 mutants generated retain activation properties higher than 80% of the P14 wild type in both yeast and tobacco plants. The P14m5 shows a significantly lower level of activation (32–38% less than P14wt) and it contains two substitutions (G for R_{15}, and G for R_{19}). P14m3 also has two similar mutations (G for D_{13}, and G for E_{22}) but its level of activation is just 10–15% lower than P14wt. The effects on the secondary structrure of P14m3 and m5 caused by these substitutions are similar (data not shown). However, while the net charge of P14m3 remains the same, the one of P14m5 has change from +4 to +2. These data are consistent with the importance of the positive charges for the function of the plant activating sequences.

It has been proposed that activation domains employ the same pathway to achieve their function. Indeed, acidic activation domains function in all eukaryotic organisms (35). They have been postulated to work recruiting transcriptional factors by virtue of their net negative charge (3,4). This may not be the case considering the great variety of activation domains described in this report and in the literature. In plants, functional studies on transcription factors have been performed only with the B and C1 genes (11,12). The gene C1 contains an acidic-type activation domain. We describe a pool of tobacco sequences that function as activation domains in yeast as well as in plants. These sequences do not seem to follow any consistent pattern. So, the variability found in yeast and mammalian activation domains may also exist in plants.

ACKNOWLEDGEMENTS

The authors thank Janet Walters for media supplies; Dawn Miner for greenhouse support; Cindy Boyce and Roseanne Novitzky for their technical support involving sequencing and process of the sequences. We also thank Dr Brian Carr for comments on the manuscript.

REFERENCES

1. Lin, Y-S., Ha, I., Maldonado, E., Reinberg, D. and Green, M.R. (1991) *Nature* **353**, 569–571.
2. Berger, S.L., Pina, B., Silverman, N., Marcus, G.A., Agapite, J., Regier, J.L., Triezenberg, S.J. and Guarente, L. (1992) *Cell* **70**, 251–265.
3. Ptashne, M (1988) *Nature* **335**, 683–689.
4. Sigler, P.B. (1988) *Nature* **333**, 210–212.
5. Ma, J. and Ptashne, M. (1987) *Cell* **51**, 113–119.
6. Giniger, E. and Ptashne, M. (1987) *Nature* **330**, 670–673.
7. Cress, W.D. and Triezenberg, S.J. (1991) *Science* **251**, 87–90.
8. Leuther, K.K., Salmeron, J.M. and Johnston, S.A. (1993) *Cell* **72**, 575–585.
9. Van Hoy, M., Leuther, K.K., Kodadek, T. and Johnston, S.A. (1993) *Cell* **72**, 587–594.
10. Katagiri, F. and Chua, N-H. (1992) *Trends Genet.* **6**, 22–26.
11. Goff, S.A., Cone, K.C. and Fromm, M.E. (1991) *Genes Dev.* **5**, 289–309.
12. Goff, S.A., Cone, K.C. and Chandler, V. (1992) *Genes Dev.* **6**, 864–875.
13. Paz-Ares, J., Ghosal, D., Wienand, P.A., Peterson, P.A. and Saedler, H. (1987) *EMBO J.* **6**, 3553–3558.
14. Gill, G. and Ptashne, M. (1987) *Cell* **51**, 121–126.
15. Bartel, P.L., Chien, C., Sternglanz, R. and Fields, S. (1993) In Hartley, D. (ed.), *Cellular Interactions in Development: A Practical Approach.* IRL Press, Oxford.
16. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd edn. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.
17. Estruch, J.J., Goff, S.A. and Crossland, L. (1994) *BioTechniques* **16**, 610–612.
18. Kaiser, P. and Auer, B. (1993) *BioTechniques* **14**, 552.
19. Leung, D.W., Chen, E. and Goeddel, D.V. (1989) *Technique* **1**, 11–15.

20. Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.
21. Rothstein, S.J., Lahners, K.N., Lotstein, R.J., Carozzi, N.B., Jayne, S.M. and Rice, D.A. (1987) *Gene* **53**, 153–161.
22. Callis, J., Fromm, M.E. and Walbot, V. (1987) *Genes Dev.* **1**, 1183–1200.
23. Lowe, K.S., Taylor, D.B., Ryan, P.L. and Paterson, K.P. (1985) *Plant Sci.* **41**,125–132.
24. Wang, M.M. and Reed, R.R. (1993) *Nature* **364**, 121–126.
25. Li, J.J. and Herskowitz, I. (1993) *Science* **252**, 1870–1874.
26. Estruch, J.J. (unpublished). EMBL accession numbers: P1, L33874; P5, L33877; P8, L31415; P11, L33875; P12, P33876; and P14, L31416.
27. Saez-Vasquez, J., Raynal, M., Meza-Basso, L. and Delseny, M. (1993) *Plant Mol. Biol.* **23**, 1211–1221.
28. Adams, S.M., Helps, N.R., Sharp, M.G.F., Brammar, W.J., Walker, R.A. and Varley, J.M. (1992) *Hum. Mol. Gen.* **1**, 91–96.
29. Giniger, E., Varnum, S.M. and Ptashne, M. (1985) *Cell* **40**, 767–774.
30. Courey, A.J. and Tjian, R. (1988) *Cell* **55**, 887–898.
31. Martin, K.J., Lillie, J.W. and Green, M.R. (1990) *Nature* **346**, 147–152.
32. Mermod, N., O'Neill, E.A., Kelly, T.J. and Tjian, R. (1989) *Cell* **58**, 741–753.
33. Santoro, C., Mermod, N., Andrews, P.C. and Tjian, R. (1988) *Nature* **334**, 218–224.
34. Chou, P.Y. and Fasman, G.D. (1978) *Annu. Rev. Biochem.* **41**, 251–256.
35. Struhl, K. (1988) *Annu. Rev. Biochem.* **58**, 1051–1077.