LETTER TO THE EDITOR

The APETALA2 Domain Is Related to a Novel Type of DNA Binding Domain

The recently cloned *APETALA2* (*AP2*) gene regulates meristem identity, floral organ specification, and seed coat development in Arabidopsis (Bowman et al., 1989, 1991; Kunst et al., 1989; Irish and Sussex, 1990; Jofuku et al., 1994). The deduced AP2 amino acid sequence is different from that of other floral regulatory proteins, most of which share the same type of DNA binding motif, the MADS domain (for review, see Davies and Schwarz-Sommer, 1994). Although AP2 does not have a MADS domain, it contains two copies of a novel motif, dubbed the AP2 domain, which is also found in other plant genes without known function (Jofuku et al., 1994). It has been suggested that a conserved core region of the AP2 domain forms an amphipathic helix that mediates protein–protein interactions. The additional presence of a putative nuclear localization signal and of an acidic and serine-rich region outside the AP2 domain have been taken as evidence for AP2 being a transcriptional regulator (Jofuku et al., 1994).

Ohme-Takagi and Shinshi (1995) have now identified a small family of tobacco proteins that bind to the ethylene-responsive elements of genes encoding pathogenesis-related proteins. The amino acid sequences of the ethylene-responsive element binding proteins (EREBPs) are rather divergent, except for an ~60–amino acid long domain that is highly conserved among all family members. This conserved domain is coincident with the DNA binding domain of EREBPs, as demonstrated by deletion analyses. Ohme-Takagi and Shinshi (1995) noted that the EREBP domain is also found in a number of plant genes without known function.

Sequence comparison shows that the EREBP DNA binding domain and the AP2 domain are closely related (Figure 1), suggesting that the AP2 domain also binds DNA. In contrast to the four published EREBPs, AP2 would appear to contain two DNA binding domains. However, there are precedents for the presence of more than one DNA binding domain in a single pro-

```
D23002      SSKYRGVT-L HKCGRWEARM GQF-LGKKYV YLGLFDTEEE AARAYDRAAI KCNGKDAVTN FDPSIY
AP2-R2      ********-* ********** ***-****** *******V* ******K*** ********** ******
AP2-R1      **Q*****FY RRT****SHI WDC--**-Q* ***G***AHA ********** *FR*VE*DI* *NIDD*
D15799      A*I*****RR Q*D***Q**I *LV-A*TRDI ***T*K**** **E**GI*** EIR**N**** **R*N*

pPLZ2       QQR***FR-Q RHW*S*VSEI RHS-IL*TRI WQ*T*ESA*D ******E**R LMC*TR*R** *PYNPN
EREBP1      GRH****R-R RPW*KFA*EI RDPAKNGAR* W**TYE*D** **I***K**Y RMR*SK*HL* *PHR*G
EREBP2      GRH****R-Q RPW*KFA*EI RDPAKNGAR* W**TYE*A** **L***K**Y RMR*SK*LL* *PHR*G
EREBP3      EVH****R-K RPW**YA*EI RDP-GK*SR* W**T***A** **K***T**R EFR*PK*K** *PSPTE
EREBP4      KKH****R-Q RPW*KFA*EI RDPNRKGTR* W**T***AI* **K******F *LR*SK*IV* *PLEVA
Cd-ind.     KNV***IR-K RPW*K*A*EI RDP-RKGVR* W**T*N*A** **M***VP*K QIR*DK*KL* *PDLHH
Z27045      RKPH**IR-R R*WRK*V*EI REP-NKRSRL W**SYT*DIA ******V*VF YLR*PSTRL* *PDLLL

                     s  m              y   y
consensus   ...yRG.... ...g.#.a.i .........≡ wlG.f.t... AA.AYd.aa. ...G..a..N F
```

Figure 1. Sequence Alignment of Selected AP2/EREBP Domains.

AP2/EREBP domains are from D23002, a rice EST; AP2-R2 (the second repeat domain); AP2-R1 (the first repeat domain); D15799, a rice EST; pPLZ2, a cDNA isolated from suspension-cultured lupine cells (Perrey et al., 1990); EREBP1; EREBP2; EREBP3; EREBP4; Cd-ind., a Cd-inducible Arabidopsis gene (GenBank accession number Z37504); and Z27045, an Arabidopsis EST. These domains fall into two major subfamilies, as indicated by the blank line separating the top four sequences from the rest. Asterisks indicate amino acids that are identical to those in the first sequence, D23002 from rice. In the consensus sequence, uppercase letters indicate invariant amino acids, lowercase letters indicate conserved amino acids (i.e., where two or three conserved amino acids occupy the same position or where just one of the 11 sequences shows a nonconservative substitution), # indicates a hydrophobic aromatic amino acid, and ≡ indicates a hydrophobic aliphatic amino acid. The region predicted to form an amphipathic helix (Rost and Sander, 1994) is doubly underlined. This region is fully contained within the amphipathic helix proposed by Jofuku et al. (1994). The invariant glycine that is changed in the *ap2-1* allele is indicated in boldface type.

tein. Examples include POU domain transcription factors, which contain a homeodomain and a POU domain, as well as C_2H_2 zinc finger proteins, most of which contain several zinc fingers.

A recent search of the GenBank and Expressed Sequence Tags (EST) data bases (March 1995) uncovered about 50 AP2/EREBP-related genes without known function. AP2/EREBP domains seem to belong to several subgroups, with only a small number of amino acids being identical in all members of the extended family (Figure 1). The AP2/EREBP domains of AP2 and those of the EREBPs appear to belong to two different subdivisions. Curiously, animal or fungal proteins with an obvious AP2/EREBP domain have yet to be found. One EST clone from rice, D23002, aligns almost perfectly with the second AP2 domain in AP2 (Figure 1); the sequence similarity actually extends beyond the AP2 domain (not shown), suggesting that this gene might be an AP2 ortholog.

Use of the aligned sequences to predict secondary protein structure confirms that a 15–amino acid stretch of the AP2/EREBP domain that contains several invariant amino acids has a very high probability of forming an amphipathic helix (Rost and Sander, 1994). In other DNA binding proteins, amphipathic helices often make intermolecular protein–protein contacts that increase DNA binding affinity through oligomerization, one example being helix 5 of the operator binding domain of λ repressor (Pabo and Lewis, 1982; Beamer and Pabo, 1992). One might imagine that this region of the AP2/EREBP domain has a similar function.

The recognition of the AP2 domain as a new type of DNA binding domain will be useful for formulating hypotheses about the functions of proteins containing such sequence motifs. A case in point

is the AP2 gene itself. During the organ specification phase, AP2 has at least two activities in the outer two whorls of the flower: repression of the homeotic gene AGAMOUS (AG) and promotion of sepal and petal fates (Bowman et al., 1989, 1991; Kunst et al., 1989; Drews et al., 1991). These two activities appear to be separable, as evident from the idiosyncratic ap2-1 mutation, which severely reduces the ability of AP2 to promote sepal and petal fates but has only a minor effect on the ability of AP2 to repress AG expression (Bowman et al., 1989; Drews et al., 1991). Interestingly, the ap2-1 mutation causes a nonconservative substitution of a glycine to a serine in the second AP2/EREBP domain in AP2 (indicated in bold in Figure 1) (Jofuku et al., 1994). This glycine is invariant in all AP2/EREBP domains surveyed, and the ap2-1 mutation is thus likely to change dramatically the DNA binding activity or specificity of this domain. It is conceivable that the differential inactivation of one of the two AP2 DNA binding domains is responsible for the differential effects of the ap2-1 mutation on AP2 activity. It will be interesting to determine whether these effects involve a direct interaction of AP2 protein with AG regulatory sequences.

Detlef Weigel
Plant Biology Laboratory
The Salk Institute for Biological Studies
10010 N. Torrey Pines Rd.
La Jolla, CA 92037

REFERENCES

Beamer, L.J., and Pabo, C.O. (1992). Refined 1.8 Å crystal structure of the lambda repressor-operator complex. J. Mol. Biol. 227, 177–196.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1989). Genes directing flower development in Arabidopsis. Plant Cell 1, 37–52.

Bowman, J.L., Smyth, D.R., and Meyerowitz, E.M. (1991). Genetic interactions among floral homeotic genes of Arabidopsis. Development 112, 1–20.

Davies, B., and Schwarz-Sommer, Z. (1994). Control of floral organ identity by homeotic MADS-box transcription factors. In Plant Promoters and Transcription Factors, L. Nover, ed (Berlin: Springer-Verlag), pp. 235–258.

Drews, G.N., Bowman, J.L., and Meyerowitz, E.M. (1991). Negative regulation of the Arabidopsis homeotic gene AGAMOUS by the APETALA2 product. Cell 65, 991–1002.

Irish, V.F., and Sussex, I.M. (1990). Function of the apetala-1 gene during Arabidopsis floral development. Plant Cell 2, 741–751.

Jofuku, K.D., den Boer, B.G.W., Van Montagu, M., and Okamuro, J.K. (1994). Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. Plant Cell 6, 1211–1225.

Kunst, L., Klenz, J.E., Martinez-Zapater, J., and Haughn, G.W. (1989). AP2 gene determines the identity of perianth organs in flowers of Arabidopsis thaliana. Plant Cell 1, 1195–1208.

Ohme-Takagi, M., and Shinshi, H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell 7, 173–182.

Pabo, C.O., and Lewis, M. (1982). The operator-binding domain of λ repressor: Structure and DNA recognition. Nature 298, 443–447.

Perrey, R., Warskulat, U., and Wink, M. (1990). Molecular cloning of a lupin-specific gene from a cDNA library of suspension-cultured cells of Lupinus polyphyllus. Plant Mol. Biol. 15, 175–176.

Rost, B., and Sander, C. (1994). Combining evolutionary information and neural networks to predict protein secondary structure. Proteins 19, 55–72.