26. P. K. Chanda, M. Ono, M. Kuwano, H. F. Kung, *J. Bacteriol.* **161**, 446 (1985).
27. T. M. Shinnick *et al.*, *Infect. Immun.* **56**, 446 (1988).
28. R. W. Hendrick, unpublished data.
29. C. P. Georgopoulos *et al.*, *J. Mol. Biol.* **76**, 645 (1973); M. Sternberg, *ibid.*, p. 1.
30. J. Kochan and H. Murialdo, *Virology* **131**, 100 (1983).
31. T. Takano and T. Kakefuda, *Nature New Biol.* **239**, 34 (1972).
32. T. W. McMullin and R. L. Hallberg, *Mol. Cell Biol.* **8**, 371 (1988).
33. R. L. Hallberg, *Semin. Cell Biol.* **1**, 37 (1990).
34. S. M. Hemmingsen *et al.*, *Nature* **333**, 330 (1988).
35. P. Goloubinoff *et al.*, *ibid.* **342**, 884 (1989).
36. P. V. Viitanen *et al.*, *Biochemistry* **29**, 5665 (1990).
37. R. J. Ellis, *Nature* **345**, 484 (1990).
38. C. B. Anfinsen, *Science* **181**, 223 (1973).
39. R. T. Morimoto, A. Tissieries, C. Georgopoulos, Eds., *Stress Proteins in Biology and Medicine* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1990).
40. S. M. Dilworth and C. Dingwall, *BioEssays* **2**, 44 (1988).
41. S. Lecker *et al.*, *EMBO J.* **8**, 2703 (1989).
42. C. Georgopoulos and D. Ang, *Semin. Cell Biol.* **1**, 19 (1990).
43. D. S. Reading *et al.*, *Nature* **337**, 655 (1989).
44. D. B. Young, *Semin. Cell Biol.* **1**, 27 (1990).
45. R. Martel, L. P. Cloney, L. E. Pelcher, S. M. Hemmingsen, *Gene*, in press.
46. T. H. Lubben *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
47. T. H. Lubben *et al.*, *Plant Cell* **1**, 1223 (1989).
48. M. Y. Cheng *et al.*, *Nature* **337**, 620 (1989); J. Ostermann *et al.*, *ibid.* **341**, 125 (1989); A. L. Horwich *et al.*, *Trends Biotechnol.* **8**, 126 (1990); F.-U. Hartl *et al.*, *Science* **247**, 930 (1990).
49. K. Willison *et al.*, *Cell* **57**, 621 (1989).
50. G. M. F. Watson, N. H. Mann, R. J. Ellis, unpublished data.
51. R. S. Gupta, *Biochem. Int.* **4**, 833 (1990).
52. R. A. Johnson, thesis, University of Warwick, United Kingdom (1987).
53. D. J. Lipman and W. R. Pearson, *Science* **227**, 1435 (1985).
54. S. M. van der Vies, thesis, University of Warwick, United Kingdom (1989).

The Cauliflower Mosaic Virus 35*S* Promoter: Combinatorial Regulation of Transcription in Plants

PHILIP N. BENFEY AND NAM-HAI CHUA

Appropriate regulation of transcription in higher plants requires specific cis elements in the regulatory regions of genes and their corresponding trans-acting proteins. Analysis of the cauliflower mosaic virus (CaMV) 35*S* promoter has contributed to the understanding of transcriptional regulatory mechanisms. The intact 35*S* promoter confers constitutive expression upon heterologous genes in most plants. Dissection into subdomains that are able to confer tissue-specific gene expression has demonstrated that the promoter has a modular organization. When selected subdomains are combined, they confer expression not detected from the isolated subdomains, suggesting that synergistic interactions occur among cis elements. The expression patterns conferred by specific combinations of 35*S* subdomains differ in tobacco and petunia. This indicates that a combinatorial code of cis-regulatory elements may be interpreted differently in different species.

RECENT STUDIES OF GENES EXPRESSED IN A TISSUE-SPECIFIC manner (*1*) and genes that control specific developmental pathways (*2*) have revealed the importance of transcriptional mechanisms in the control of development in higher plants. In some cases, the DNA sequence elements that are necessary for transcriptional regulation and the protein factors that interact with these sequences have been identified. What has emerged is a complex picture in which DNA sequence elements that are important for regulation are scattered over thousands of base pairs (bp), and these elements interact with factors that can be either ubiquitous or highly restricted in their distribution. A simple model in which transcriptional regulation is mediated solely by the presence or absence of a particular trans-acting factor now seems inadequate. Rather, transcriptional regulation may be accomplished through combinatorial mechanisms (*3, 4*) by which diverse expression patterns are achieved through different combinations of a limited number of regulatory elements and trans-acting factors.

Laboratory of Plant Molecular Biology, The Rockefeller University, New York, NY 10021.

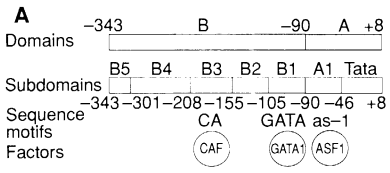
If combinatorial processes control transcription, we should be able to identify the basic components and generate different transcription patterns when the components are combined in different ways. In order to observe the effects of different combinations of the basic components, gene expression must be monitored in a variety of tissues and throughout development. Plants are particularly well suited to this type of analysis, because transgenic plants can be rapidly and easily generated, and reporter gene expression can be monitored in most cells and at various stages of development. We have used a viral promoter that is able to confer expression in most plant cells as a model system to dissect some of the combinatorial properties of transcriptional control in plants. In this article, we review our current knowledge of the components within this promoter and the results of combining these components in different ways. We contrast expression patterns conferred by various combinations of cis elements in flowers from two different species. We also compare these findings with evidence for combinatorial mechanisms of gene regulation obtained from plants and other organisms.

Expression Modules Within the CaMV 35S Promoter

The 35S promoter of cauliflower mosaic virus (CaMV) confers a high amount of expression in most cells when transferred into plants (*5–7*). In addition, expression is not dependent on viral trans-acting factors (*5*). For these reasons it is one of the most commonly used promoters in plant biotechnology. Deletion analyses of the 35S promoter have revealed the existence of variously regulated expression modules (*5, 8–10*). In tobacco callus and leaf tissue, deletion from −943 to −343 bp upstream of the transcription initiation site does not decrease expression. However, deletion to −105 decreases transcription to one-third of control values, and further deletion to −46 decreases transcriptional activity to one-twentieth of the amount observed with the −105 deletion (*5*). These results suggest that regions of the CaMV promoter between −343 and −105 and −105 and −46 are important for transcriptional activation. A second study employed progressive 5′ and 3′ deletions to better define important cis-regulatory elements. Four regions that affect transcription in tobacco leaf tissue were identified (*10*). When the region between −208 and −46 is deleted, transcription is decreased to one-twentieth of the control value. This region activates transcription in both orientations, indicating that it has the properties of an enhancer. In addition, synergistic effects are observed with combinations of cis elements found in the 35S promoter. The region from −90 to +8 gives no detectable expression in leaf tissue, as is the case for the fragment from −343 to −208, joined to a minimal 35S promoter (−46 to +8). However, when the −343 to −208 fragment is fused to the −90 to +8 fragment, expression in leaf is reproducibly detected (*10*).

These deletion analyses of the 35S promoter suggest that more than one region of the promoter contributes to expression. When the −90 to +8 region is fused to the chloramphenicol acetyltransferase (CAT) structural gene, CAT activity is detected in roots but not in leaves (*11*). Therefore, distinct regions of the 35S promoter might confer expression in different tissues. In order to determine the cell specificity of expression of different regions of the 35S promoter, we used the β-glucuronidase (GUS) reporter gene fused to various promoter segments to permit histochemical localization (*6*). We initially dissected the promoter into two regions termed domain A (−90 to +8) and domain B (−343 to −90) (Fig. 1A). Expression from domain B was analyzed with a fusion to a minimal promoter (−72 to +8) that alone gives no detectable expression at any stage of development. Expression was monitored in seeds, seedlings, and mature transgenic tobacco plants. Histochemical localization of expression shows that domain A confers expression principally in root or embryonic tissue destined to become root (*12*). The expression in the root is strongest in the meristematic region. Some expression from A is also detected in the meristematic region of the stem (particularly in seedlings) and in the vascular tissue of the stem and leaf. In contrast, domain B confers expression principally in the aerial portions of the plant. Expression is particularly strong in vascular, epidermal, and mesophyll cells of the embryonic cotyledon and leaf, as well as in vascular tissue of the stem. In the root, expression from domain B is detected primarily in vascular tissue and in the root cap (*12*). The combination of domain B with domain A results in expression in most cell types at all stages of development analyzed.

The conclusion from this analysis is that the two domains have complementary expression patterns with little overlap. This suggests that a fairly simple additive relationship exists among cis elements. However, evidence from deletion analyses (*9, 10*) in which synergistic interactions were observed, indicate that more complicated combinatorial relations might exist. In addition, expression from domain B is observed in numerous cell types, suggesting that domain B may be a combination of regulatory elements. To investigate whether domain B is composed of different modules, each able to confer a specific expression pattern, we further dissected domain B into five subdomains (Fig. 1A). We used as sites for our dissection the boundaries defined in the deletion analyses (*10*), as well as one additional site. We reasoned that the quantitative effect on transcription in leaves that results from progressive deletion of these regions might be due to reduction of expression in specific sets of cells within the leaf. We also wished to analyze the effect on expression of combinations of cis elements. To this end, we made combinations of each of the subdomains with domain A. The control constructs consisted of four copies of the subdomains fused to the minimal TATA region of the 35S promoter (−46 to +8). We also examined the combination of subdomains B4 and B5 (−343 to −208), which interacts synergistically with domain A in deletion analyses (*10*). A fragment that contained subdomains B4 and B5 was placed upstream of the minimal TATA region and fused to domain A (*13, 14*).



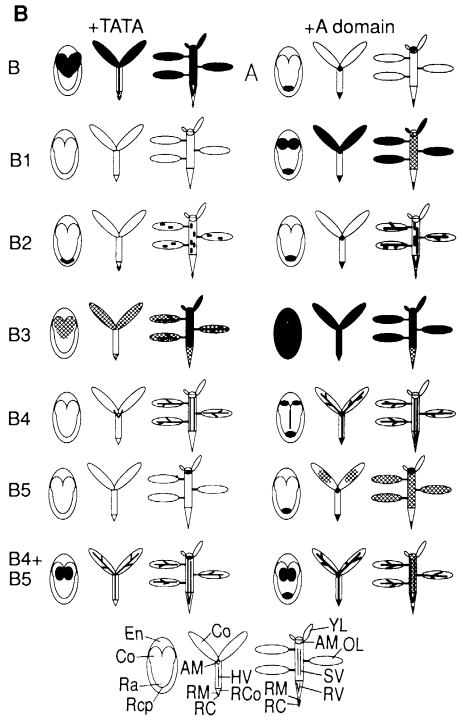


Fig. 1. Regulation of gene expression from the CaMV 35S promoter. (**A**) Analysis of the CaMV 35S promoter. Progressively smaller regions of the 35S promoter have been analyzed for their ability to confer cell-specific gene expression. This has served to define two domains, domain A from −90 to +8, and domain B from −343 to −90, and five subdomains, B1 through B5, whose end points are indicated. The A1 subdomain is between −90 and −46. Biochemical analysis has identified three sequence motifs within the promoter (as-1, GATA, and CA), which are the binding sites for three trans-acting factors (ASF-1, GATA1, and CAF, respectively). (**B**) Schematic representation of expression patterns conferred by domain B (top left) and the individual B subdomains placed upstream of the minimal TATA region (−46 to +8). For the B3 subdomain, only one of two expression patterns is shown (*13, 14*). Expression patterns conferred by domain A alone (top right) or the B subdomains in combination with domain A are shown on the right side of the figure. Crosshatching represents low amounts of expression. Expression is depicted in (from left to right) seeds, seedlings, and mature plants. Expression in the root cap is indicated by an additional set of lines for seeds and seedlings, and by an additional filled triangle in mature plants. The tissues represented schematically are identified in the last row. AM, apical meristem; Co, cotyledon; En, endosperm; HV, hypocotyl vascular tissue; OL, older leaf; Ra, radicle; Rcp, root cap primordia; RC, root cap; RCo, root cortex; RM, root meristem; RV, root vascular tissue; SV, stem vascular tissue; YL, younger leaf.

A schematic summary of the results of this analysis in tobacco seeds, seedlings, and mature plants is presented in Fig. 1B. Four of the five subdomains, when placed upstream of the minimal TATA region, confer GUS expression in at least one cell type. The B2 subdomain confers expression in a single cell type within the phloem elements of stem and leaf, as well as in root cap and root hair cells. The B3 subdomain has two expression patterns, each detected in four independent transgenic plants. Except for a difference in the endosperm staining pattern in seeds, the two patterns differ quantitatively rather than qualitatively. In one set of plants, the B3 subdomain conferred strong expression in endosperm tissue at the radicle pole of the seed, weak expression in the base of the cotyledons of the seedling, and weak expression in leaf primordia emerging from the meristem. A second set of plants that contained the B3 subdomain showed expression in the cotyledon of the embryo and seedling, as well as in root cortex of the seedling, and in most cells of the mature stem. The B4 subdomain confers expression in vascular parenchyma cells of the leaf and stem. The B5 subdomain confers weak expression only in a region below the stem apex and in leaf buds emerging from the stem. Except for a single plant among the 14 independent transgenic plants analyzed, the B1 subdomain gives no expression at any stage of development (*13, 14*).

These results from the individual subdomains placed upstream of a minimal promoter indicate that there are at least four identifiable components or modules within domain B, each able to confer a different expression pattern. Both complementary and redundant expression patterns are conferred by the B subdomains. However, the sum of the expression patterns of the individual modules does not equal the expression pattern of the intact B domain. One possible explanation for this is that additional cis elements exist that are interrupted when the subdomains are isolated. Until we know the precise sites of interaction of all trans-acting factors with the 35*S* promoter, it will be difficult to rule out this possibility. However, our analysis of combinations of the B subdomains and domain A suggest another possibility. At least a part of the missing expression may be conferred by combinations of modules within domain B.

Combinatorial Properties of 35*S* Subdomains

The combination of each of the five B subdomains with domain A results in an expression pattern that differs from that of the individual subdomain or domain A. A striking example is that of subdomain B1, which, when placed upstream of the minimal TATA region, confers no reproducible expression in seeds and seedlings. When subdomain B1 is fused to domain A, expression is easily detectable in the cotyledons of both seeds and seedlings (*13*). The synergistic effect of other combinations is sometimes only observable at one stage of development. The combination of subdomain B2 with domain A does not reproducibly change expression in seeds or seedlings (as compared to that of domain A alone), but results in strong expression in phloem elements in leaf, stem, and root of mature plants. For B4 fused to domain A, strong vascular expression is evident in seedlings, and phloem expression is detected in the root of mature plants (Fig. 1B). When B4 and B5 are placed upstream of the TATA region, expression is detected in the vascular tissue of seedlings, a pattern not observed with either B4 or B5 alone. Expression from B4 and B5 is also observed in leaf, stem, and root vascular tissue of the mature plant (*14*).

Unlike in other organs, the 35*S* promoter is not highly active in all cells of tobacco flowers (*15*). The intact 35*S* promoter confers expression principally in vascular tissue and in trichomes of the petal, with weaker expression observed in epidermal tissue. As a means of determining which subdomain or combination of subdomains is responsible for this expression pattern, we analyzed representative transgenic tobacco plants that contained constructs with subdomains fused to either the minimal (−46 to +8) promoter or to domain A.

Transgenic tobacco plants that contained domain A alone showed essentially no expression in mature petal (Fig. 2A) (occasionally very low vascular expression was observed). In contrast, the combination of domain B fused to A conferred a similar expression pattern to that of the intact 35*S* promoter. Expression in mature petal was observed in trichomes and in vascular and epidermal tissue. We rarely detected expression in mesophyll tissue (Fig. 2B) (mesophyll expression was observed occasionally at the upper edge of the corolla). In the developing flower bud, domain A conferred weak expression in the floral stalk. The combination of domain B and A gave strong expression in the floral stalk as well as in most tissues of the developing flower. When subdomain B2 was fused to either the minimal promoter (−46 to +8) (Fig. 2C) or to domain A (Fig. 2D), there was no striking difference in the expression pattern in mature petal tissue. Both gave expression only in vascular tissue. However, in the developing bud, subdomain B2 fused to the minimal promoter conferred expression only in isolated cells that appear to be part of the phloem. When B2 was combined with domain A, expression was observed in the vascular tissue of the floral stalk and, in the plant with the strongest expression, in tissue of the stamen.

Subdomain B3 conferred two different expression patterns in

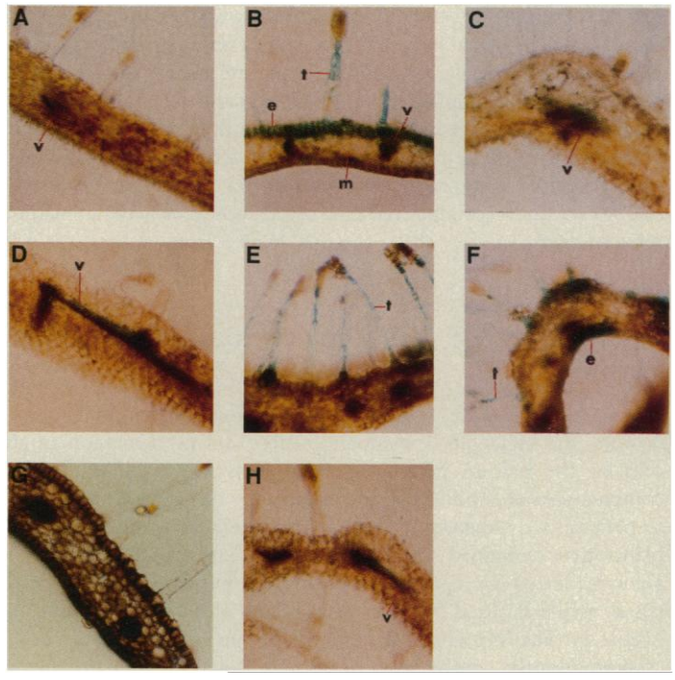


Fig. 2. Expression conferred by combinations of 35*S* subdomains in flowers from tobacco. (**A**) Domain A in mature petal; (**B**) domain B + domain A in mature petal; (**C**) four copies of subdomain B2 fused to the minimal promoter (4 × B2 + TATA) in mature petal; (**D**) 4 × B2 + A in mature petal; (**E**) 4 × B3 + TATA in mature petal; (**F**) 4 × B3 + A in mature petal; (**G**) 4 × B4 + TATA in mature petal; (**H**) 4 × B4 + A in mature petal. Three to five representative R1 or R2 plants that contained the constructs described in (*13*) were grown to maturity in the greenhouse and allowed to flower. Mature petals were removed and hand sectioned. Histochemical analysis was performed as described (*6, 15*). Except where noted in the text, there was no significant qualitative variation among independent transformants with the same construct, although variability in the intensity of staining was observed. Abbreviations: e, epidermis; m, mesophyll; t, trichome; v, vascular tissue.

floral tissue. In the plants with strong expression in stem and in the lamina of young leaves (*14*), expression in mature petal was detected in trichomes and, to a lesser extent, in epidermal and vascular tissue (Fig. 2E). The combination of subdomain B3 and domain A consistently gave a similar expression pattern, with staining detectable in trichomes and epidermal layers, and weaker staining observed in vascular tissue (Fig. 2F). Mature vegetative plants that showed expression principally in the emerging leaves in stem apical sections (*14*) showed no expression in mature petal tissue. In the flower bud of these plants, expression was restricted to the developing ovary. Plants that gave strong expression in the stem displayed expression in the floral stalk of the flower bud as well as in the developing ovary. The combination of subdomain B3 and domain A consistently gave expression throughout the floral stalk as well as in other tissues of the developing flower.

We were unable to detect any expression in mature petal from subdomains B1, B4 (Fig. 2G), or B5 fused to the minimal promoter. When B1 was combined with domain A, we occasionally detected weak expression in trichomes. B4 combined with A conferred weak expression in vascular tissue (Fig. 2H). In the flower bud, B4 fused to the minimal TATA region conferred expression principally in the vascular tissue of the floral stalk. There was no striking difference in expression when B4 was fused to A. We were unable to detect any expression from B5 fused to A in the plants we analyzed.

We conclude that the expression pattern of the intact 35*S* promoter in mature tobacco petals can be attributed principally to additive effects. In the developing flower bud, synergistic effects seem to function, as seen with the combination of B2 and A. That is, expression is observed in tissues that exhibit no expression with A or B2 alone. The vascular expression in mature petal may arise from contributions of domain A and subdomain B2, with a lesser contribution from B4 in combination with domain A. The epidermal and trichome expression appear to be principally from subdomain B3 with a minor contribution from B1. Thus, domain A may not be essential to produce the expression pattern detected in mature tobacco petal.

Analysis of 35*S* Subdomain Combinations in Petunia

In contrast to the cell-specific expression pattern in tobacco flowers, the intact 35*S* promoter confers strong expression in all cells of the mature petal of petunia flowers (*15*). To identify combinations of subdomains responsible for this expression pattern in petunia, we measured GUS activity in mature petal tissue from plants that contained various subdomain constructs. Domain A alone conferred no reproducible expression in mature petunia petal tissue, while B fused to A conferred high amounts of expression (Table 1). The histochemical analysis of expression in mature petal conferred by the combination of domain B and domain A showed an identical expression pattern to that of the intact promoter. Expression was detected in epidermal and vascular tissue and, in contrast to tobacco, there was also strong expression in mesophyll tissue (Fig. 3A). No expression was detected in the histochemical analysis of plants that contained domain A alone (Fig. 3B). In combination with domain A, only B1 and B3 consistently gave expression above background amounts (Table 1). One plant that contained B2 fused to A also gave expression that was significantly above background. To test whether these were additive or synergistic effects, we analyzed expression from these three subdomains placed upstream of the minimal promoter. None of the three was able to confer expression above background amounts (Table 1). The histochemical analysis confirmed this finding. No expression was detected with subdomains B1, B2, or B3 in combination with the minimal promoter in mature petal. However, when fused to domain A, B1 (Fig. 3C) conferred expression that appeared to be strongest in mesophyll tissue, while B3 (Fig. 3D) gave expression in all tissues of the mature petal. The one plant with B2 and A that gave significant expression in the enzymatic assay showed only vascular expression.

These results suggest that the expression pattern of the 35*S* promoter in petunia petals can be principally attributed to the B1 and B3 subdomains. However, these subdomains need to act in concert with domain A in order to confer detectable expression in petals. In contrast to the results in tobacco, expression in petunia petal appears to result from synergistic interactions among cis elements. Furthermore, the expression pattern of subdomain B3 with or without domain A differs significantly between the two plant species. A schematic summary of the salient features of the petal expression patterns conferred by subdomain combinations in the two species is presented in Fig. 4.

Combinatorial Properties of Plant Promoters

Analysis of other plant promoters for sequences that control tissue-specific expression in vivo has generally been performed with progressive deletions from an arbitrary site upstream of the promoter. In most cases, more than one region was identified that affected transcription (*15–17*). These regions were often identified as having either quantitative or tissue-specific effects, suggesting that combinations of these different regions are necessary to achieve the normal expression pattern.

More detailed analyses have been performed on several genes including the small subunit of ribulose-1, 5-*bis*-phosphate carboxylase (*rbcS*). For the *rbcS* 3A gene from pea, there is evidence for at least three regions that confer light-responsive expression (*17*). Trans-acting factors have been identified that bind within these regions (*17*). Although these promoter elements appear to have redundant functions, more than one of the defined light-responsive regions is necessary for high-level expression in newly formed leaves. However, a promoter with one of these regions deleted still confers high-level expression in older leaves (*18*). Another light-responsive gene, *chalcone synthase* from parsley, has two separable ultraviolet

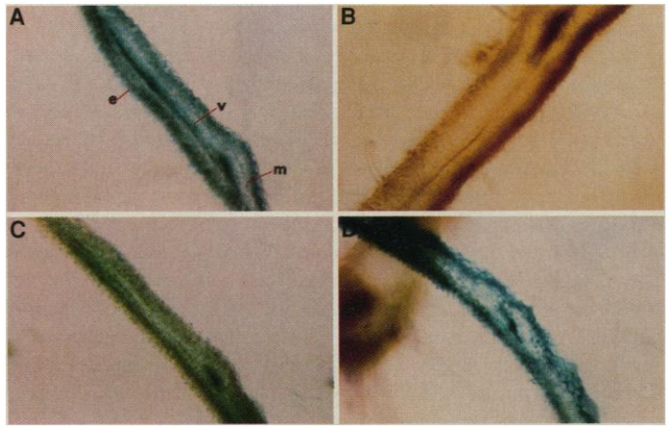


Fig. 3. Expression conferred by combinations of 35*S* subdomains in flowers from petunia. (**A**) Domain B + domain A in mature petal; (**B**) domain A in mature petal; (**C**) 4 × B1 + A (Fig. 2, legend) in mature petal; (**D**) 4 × B3 + A in mature petal. Primary transformants of petunia, cultivar Mitchell diploid, were analyzed for expression by histochemical localization as described (*6, 15*). Abbreviations are as in Fig. 2.

(UV) light-responsive cis elements. These cis elements were shown by in vivo footprinting to be bound by factors in a UV light-dependent fashion (*19*). Moreover, in addition to these cis elements, there is an upstream region that is unable to confer expression when the UV light-responsive elements are deleted, but significantly enhances expression when the light-responsive cis elements are present (*19*). Because expression was analyzed only in parsley protoplasts, it is possible that the upstream region on its own is able to confer expression in another cell type, in addition to acting in combination with the light-responsive elements.

More direct evidence for combinatorial control of gene expression comes from studies of genes that regulate the maize anthocyanin biosynthetic pathway. Genetic analyses have implicated at least two genes, *C1* and *B*, in the regulation of the structural gene, *Bronze1*. Co-delivery by particle bombardment of both *C1* and *B* is necessary to achieve activation of *Bronze1* in cells in which this gene is not normally expressed (*20*). The region upstream of the gene that encodes the α′ subunit of the soybean seed storage protein, β-conglycinin, binds at least three distinct factors (*21*). One of the factors specifically binds to a region that has seed-specific enhancer activity. The binding activity of this factor is developmentally regulated, with the peak of binding activity correlating with the peak of β-conglycinin gene expression. However, regulation may not depend solely on this binding activity, as a second factor binds specifically within the enhancer region. Moreover, another distinct region that is required for high-level tissue-specific activity does not detectably bind any of the factors characterized thus far (*21*). Therefore, combinatorial mechanisms appear to control gene expression that is highly tissue-specific and developmentally regulated.

The regulatory region upstream of the structural gene for 5-enolpyruvylshikimate-3-phosphate synthase (*EPSPS*) is able to confer strong gene expression in mature petals (*15*) and in seedling root cortex tissue of petunia (*22*). A detailed analysis of cis elements showed that expression in both tissues is conferred by an ~1 kb region located between −1800 and −800 bp upstream of the transcriptional start site. When this region is dissected into two 500-bp fragments and each fragment is fused to a heterologous promoter, both regions are able to confer petal- and root cortex–specific expression. However, the amount of expression is drastically reduced in comparison to the intact 1-kb fragment. A fragment that spans the junction between the two 500-bp fragments was also analyzed to ascertain whether an important cis element was disrupted. No expression is detected from this fragment in petal, but a low amount is observed in root cortex. When each of the 500-bp fragments is dissected into two ~250-bp segments, no expression is detected with three of the four segments, and very low expression is observed in petals with the fourth segment (*22*). One interpretation of these results is that high amounts of petal-specific expression can be attained only when cis elements on both of the 500-bp fragments are combined.

Combinations of Plant and Viral Cis-Regulatory Elements

The ability of domain A to interact synergistically with other cis-regulatory elements from the 35*S* promoter suggests that new expression patterns might result when this domain is combined with cis elements from other promoters. Indeed when the region upstream of the maize gene that encodes alcohol dehydrogenase is combined with domain A, expression becomes constitutive instead of responsive to anaerobic stress (*23*). In addition, a region (−1038 to −93 bp) of the *rbc*S-8B gene from *Nicotiana plumbaginifolia* when placed upstream of a 35*S* promoter fragment (−105 to +8) gives high amounts of expression in leaf but no detectable expression in root. This indicates that a negative interaction may occur between cis elements in the plant gene and sequences of the 35*S* promoter that normally confer expression in root (*11*). Another example of interactions between plant and viral cis elements is that of a tetramer of the Box II sequence from the pea *rbc*S 3A promoter, which does not confer detectable expression when fused to the 35*S* minimal promoter (−46 to +8) but is able to confer light-responsive expression when fused to domain A (*24*).

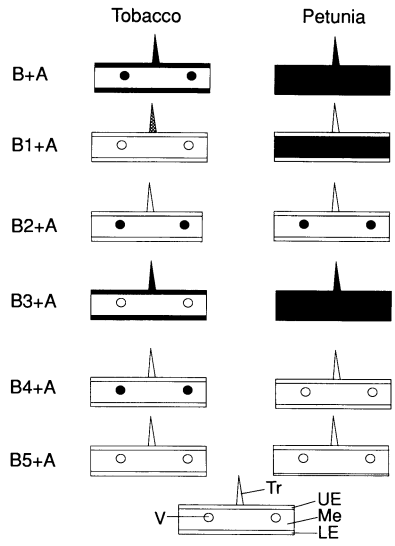


Fig. 4. Schematic representation of expression patterns conferred by 35*S* subdomains in tobacco and petunia petals. Expression conferred by the combinations of 35*S* subdomains listed on the left are shown in schematic sections through mature petals of tobacco and petunia. Only the salient features of the expression patterns are indicated. For 4 × B2 + A in petunia, the expression pattern of the single high expressing plant is shown. The cell types represented are indicated in the last row. LE, lower epidermis; Me, mesophyll; Tr, trichome; UE, upper epidermis; V, vascular tissue.

To further analyze the effect on tissue-specific expression of combinations of viral and plant cis elements, we combined domain A with the upstream region of *EPSPS*. In petunia and tobacco petals, there was no difference in expression if *EPSPS* sequences were placed upstream of domain A or upstream of the 35*S* minimal TATA region (−46 to +8). In tobacco, weak expression was occasionally detected with either construct only in the upper epidermal layer (Fig. 5, A and B). In petunia, expression was detected with both constructs in nearly all cells of the mature petal (*15, 22*) and in a gradient of expression from the top to the bottom of the tube with strongest expression in the upper epidermis (Fig. 5, C and D). Therefore, although combinations of domain A and other cis elements from the 35*S* promoter can have dramatic effects on

Table 1. Analysis of GUS activity in transgenic petunia petal tissue. The numbers of independent transgenic petunia plants, which contained the 35*S* subdomain constructs shown in the first column, with GUS activity in the ranges given at the top of the subsequent columns are shown. Tissue is from mature petals of primary transformants. The GUS activity was measured by fluorimetry as described (*6*). Activity is in picamoles of 4-methylumbelliferone per minute per milligram of protein. For the purpose of this analysis, GUS activity below 2000 units was considered to be not significantly above background levels.

Construct	Units of GUS activity (× 10^3)			
	>20	10 to 20	2 to 10	0 to 2
Domain A				18
Domain B + A	6			3
4 × B1 + A	1	2	1	6
4 × B2 + A		1		8
4 × B3 + A	8			2
4 × B4 + A				10
4 × B5 + A				11
4 × B1 + TATA				13
4 × B2 + TATA				11
4 × B3 + TATA				12

expression in petunia and, to a lesser extent in tobacco flowers (Figs. 2 and 3), we detected no such effect with cis elements derived from a normal plant gene. However, we did observe an effect at other developmental stages. In tobacco seedlings, a fragment from *EPSPS* placed upstream of the 35*S* TATA region conferred weak expression only in root cortex cells (Fig. 5E). Domain A alone confers expression principally in the root meristematic region (*12*). However, the combination of the *EPSPS* fragment and domain A gave strong expression in the cotyledons and hypocotyl (Fig. 5F), as well as in the root (Fig. 5G). In petunia, the *EPSPS* fragment placed upstream of the minimal 35*S* promoter conferred expression principally in root cortex (*22*) (Fig. 5H). Expression from domain A in petunia seedlings was more variable than in tobacco. We detected no expression in five plants, weak expression in the root tip of one plant, and expression in both root and cotyledons in one plant. However, the combination of the *EPSPS* fragment and domain A consistently conferred high expression in both root and cotyledons of petunia seedlings (Fig. 5I).

We conclude that cis elements within the upstream region of the cellular gene, *EPSPS*, can interact with virally derived domain A. This was not unexpected, as transcription from the 35*S* promoter occurs in the absence of viral trans factors, and thus, must utilize cellular factors for regulation. Our results suggest that domain A does not modify the expression pattern when combined with all cis elements. The specificity of enhancement or modification suggests that there are strict constraints on the types of interactions that can occur. These constraints are likely to be mediated by the trans factors that bind to the cis elements.

Cellular Trans-Acting Factors That Bind to the 35*S* Promoter Elements

Our initial analysis of the trans-acting factors that interact with the 35*S* promoter focused on the region of domain A that appears to be responsible for the synergistic interactions with other cis elements. Within the region between −90 and −46, there are two putative CCAAT box sequences. A deoxyribonuclease I (DNase I) footprint of this region obtained from whole-cell extracts from pea and nuclear extracts from tobacco covers the two putative CCAAT boxes, but is centered over a tandem repeat of the sequence TGACG (*25, 26*). Site-directed mutagenesis of the CCAAT boxes had no effect on footprint formation, while mutation of the TGACG sequences abolishes the protection from DNase I. In transgenic plants, the same mutations in the TGACG motifs, within the context of the full promoter (−343 to +2), has a dramatic effect on expression in root, but little effect on expression in leaf. Mutations in the CCAAT boxes have no significant effect on expression in root or leaf. The 21-bp region that encompasses the two TGACG sequences was named activating sequence–1 (as-1), and the factor that binds in this region was named activating sequence factor–1 (ASF-1). The 21-bp as-1 site was inserted into the *rbc*S 3A promoter, which normally only confers expression in photosynthetic tissue (*26*). When the as-1 site is added, the promoter gives high expression in root in addition to the normally high amounts seen in leaf (*26*). When an as-1 site with mutations in the TGACG motifs is inserted in the same position, expression is detected in leaves but not in roots (*26*).

An oligonucleotide that binds ASF-1 from nuclear extracts was used to probe an expression complementary (cDNA) library made from tobacco leaf mRNA. Several clones were isolated that encode proteins that bind specifically to the as-1 sequence. One of these encodes a protein named TGA1a (*27*), which is a member of the basic region–leucine zipper (bZip) family of transcription factors (*28*). Within the basic DNA-binding region, TGA1a is similar to the mammalian cyclic adenosine monophosphate (cAMP) responsive element binding (CREB) factor, whose binding site is also TGACG. Three criteria suggest that TGA1a is a positive regulator of transcription and interacts with domain A: (i) the binding specificities of TGA1a and ASF-1 are very similar (*27*); (ii) purified TGA1a produced in *Escherichia coli* stimulates transcription in both plant- (*29*) and human- (*30*) derived in vitro systems, from a construct that contains two as-1 sites; and, (iii) microinjection of purified TGA1a into the cotyledons of seedlings that contain domain A–GUS constructs results in GUS expression (*31*). In addition, TGA1a mRNA is five- to tenfold more abundant in root than in leaf (*27*). Although these results strongly implicate TGA1a in as-1–mediated transcriptional regulation, we cannot rule out the possibility that other factors may bind to this region and regulate gene expression.

We have also characterized factors that bind specifically to subdomains B1 and B3. The factor that binds to subdomain B1 is found in nuclear extracts from leaf but not from root (*32*). By DNase I footprint analysis, this factor binds to the region between −106 and −85. This factor is also able to bind to conserved sequences found upstream of the genes that encode chlorophyll a/b binding proteins (Cab) from various species (*32*). Because the core conserved sequence found in these Cab-encoding genes contains the motif GATA, the factor has been named GATA1 (Fig. 1A). The fact that the GATA motif is found in at least 12 different Cab genes from various species as well as in the 35*S* promoter (*32*) is a further indication of how a particular transcriptional component can be used in different combinations by different promoters.

With gel-retardation and footprint analyses, a factor was found in nuclear extracts from tobacco leaf that binds specifically to sequences within the B3 subdomain (*33*). We have named this factor CAF (CA-rich region factor) (Fig. 1A), and are currently analyzing its tissue-specific distribution. We have yet to detect sequence-specific interactions of factors from leaf nuclear extracts with the B2 and B4 subdomains. This observation, combined with the fact that the abundance of TGA1a mRNA is higher in root than leaf, suggests that the expression patterns correlate with the abundance of the transcription factors that interact with these sequences. Therefore, the expression patterns of the isolated subdomains may be a useful guide for tissue sources in which to find high concentrations of the factors that bind to these regions.

Combinatorial Properties of Cis-Regulatory Elements in Promoters of Other Organisms

Among the first examples of combinatorial control of gene transcription in eukaryotes was the synergistic effect of different mammalian hormones on gene regulation (*3*). Combinatorial control of tissue-specific expression was also demonstrated when combinations of transcription factor binding sites within the simian virus 40 (SV40) enhancer gave varying amounts of expression in different cell lines (*34*).

In *Drosophila* there are several examples of cell-specific expression determined by combinatorial mechanisms. Two yolk protein genes, *yp1* and *yp2*, are divergently transcribed only in specific follicle cell types. Dissection of the region between *yp1* and *yp2* reveals that distinct fragments confer different cell-type specificities. One region confers expression in anterior pole cells and border cells, while another region, when combined with the first, represses expression in border cells and confers expression in columnar main body cells (*35*). Positive and negative combinatorial mechanisms are also invoked to explain the periodic expression of the segmentation gene, *even-skipped* (*eve*). From genetic and biochemical studies, it has been

suggested that expression of *eve* is controlled by the binding of at least three distinct transcription factors. These factors can act either as repressors or activators, depending on their concentration in the cell. Because the factors are not uniformly distributed throughout the developing embryo, it is proposed that their relative concentrations determine the precise stripe-like expression pattern of *eve* (*36*).

A Model for Combinatorial Properties of Cis-Regulatory Elements

The finding that combinations of cis elements can produce expression patterns not generated when the cis elements are in isolation suggests that a combinatorial code directs expression throughout development. From our analysis of the 35*S* promoter, it appears that, if a code exists, it is degenerate for expression in a particular tissue at a single time point. For example, three different combinations of cis elements confer expression in root vascular tissue (*14*). Another level of complexity is indicated by the presence of both negative and positive interactions [for example, the *rbc*S-8B interaction with domain A (*11*)]. Thus, we can begin to make testable models for certain aspects of the code.

A simple case is that of two cis elements that on their own confer expression in a particular cell type, yet when combined they confer expression in a third cell type. This can be explained by two types of cooperative interactions. The first involves cooperative interactions among heterologous factors to mediate binding to DNA (Fig. 6A). When the concentration of active factor is high (Fig. 6A, I and II), the factor will consistently occupy its cis element and activate transcription in those cells. When the factor is not present in high enough concentrations to remain bound to its cis element (Fig. 6A, III), cooperative interaction with a second factor can allow binding and subsequent activation of transcription. The second model involves cooperative interactions with a target factor. This is a modification of a model originally proposed as an explanation for the ability of the yeast transcriptional activator, GAL4, to interact synergistically with mammalian activators (*37*). In this model, the target factor must interact with more than one DNA-bound factor at a time (Fig. 6B). Synergism arises because a minimum number of factors must be bound simultaneously to the target factor and to their cognate cis elements to obtain a productive complex. When the concentration of active factor is high enough to allow binding to the correct number of sites on the DNA, activation of transcription occurs (Fig. 6B, I and II). However, when there is an insufficient concentration of active factor to consistently occupy the sites or bind

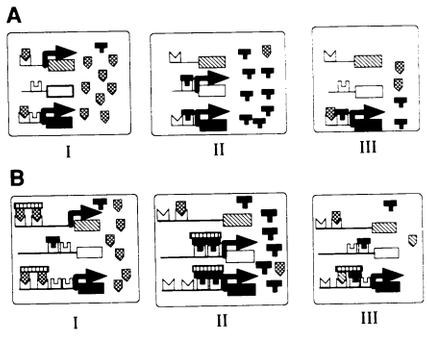


Fig. 6. Models of interactions among cis-regulatory elements. (**A**) Cooperative interactions among heterologous trans-acting factors. High concentrations of a trans-acting factor in cell I (crosshatched polygons) or a second trans factor in cell II (solid T-shapes) allow transcriptional activation of promoters with the cognate binding sites. In cell III, when there are lower concentrations of both factors, a cooperative interaction between the factors allows both to bind and activate transcription only from the promoter that has sites for both factors. (**B**) Synergism mediated by cooperative interactions with a target factor. In cell I or II, active factor concentration is sufficiently high to fill the number of sites needed for productive interaction with the target factor (striped bar). Transcriptional activation occurs from promoters that have enough occupied sites. In cell III, when factor concentrations are lower, transcriptional activation only occurs when the target factor interacts with heterologous factors that bind independently to their cognate sites.

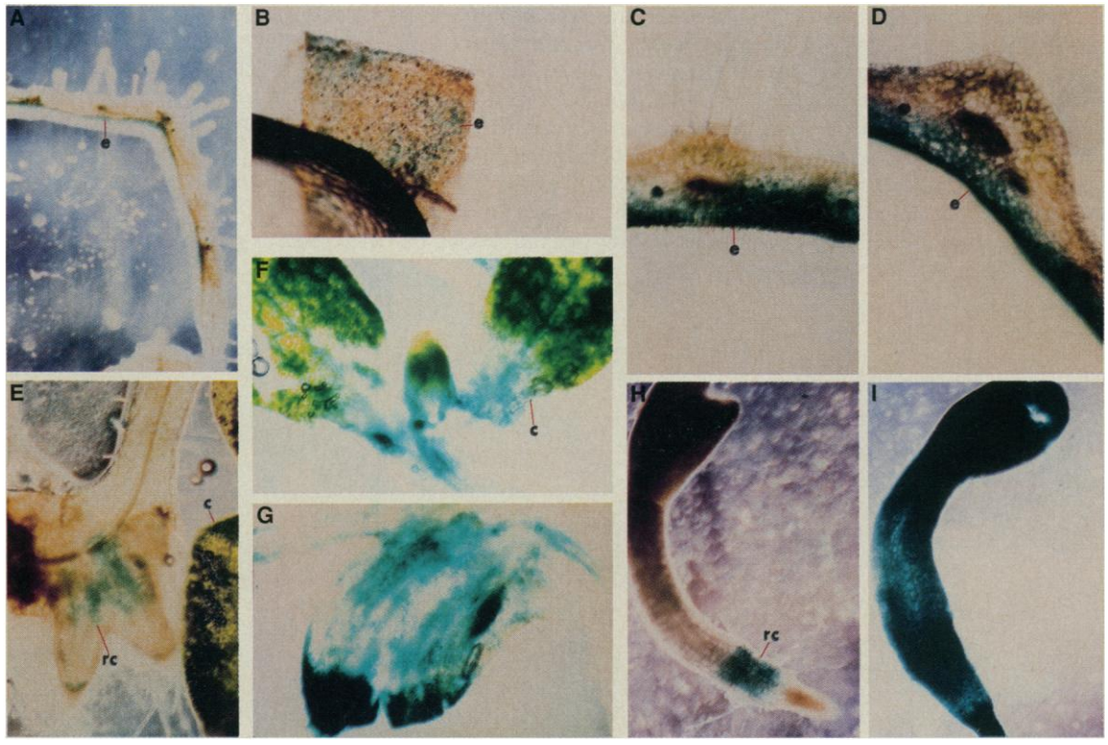


Fig. 5. Expression conferred by the upstream region (−1761 to −268) of *EPSPS* combined with either the TATA region (EPSPS + 35*S*-TATA) or domain A (EPSPS + A) of the 35*S* promoter. (**A**) EPSPS + 35*S*-TATA in mature tobacco petal; (**B**) EPSPS + A in mature tobacco petal (the section is twisted to visualize the small amount of epidermal staining); (**C**) EPSPS + 35*S*-TATA in the upper tube of a mature petunia flower; (**D**) EPSPS + A in the upper tube of a mature petunia flower; (**E**) EPSPS + 35*S*-TATA in a 17-day-old tobacco seedling; (**F**) EPSPS + A in the hypocotyl and cotyledon of a 17-day-old tobacco seedling; (**G**) EPSPS + A in the root of a 17-day-old tobacco seedling; (**H**) EPSPS + 35*S*-TATA in a 7-day-old petunia seedling; and, (**I**) EPSPS + A in a 7-day-old petunia seedling. Histochemical analysis was performed as described (*6, 15, 22*). Abbreviations are as in Fig. 2; c, cotyledon; rc, root cortex.

to the target factor, then no expression occurs from promoters that contain only those cis elements (Fig. 6B, III, upper two promoters). However, a promoter that contains a combination of different binding sites can confer expression because two heterologous factors bound to different cis elements can interact with the target factor to promote a productive interaction (Fig. 6B, III, lowest promoter). In order to have a large synergistic effect, the target factor must have a higher affinity for the bound heterologous factors than for homologous factors.

Both models provide means of regulating transcription by modulating amounts of active transacting factors in the cell. The first model predicts that factors interact with each other to mediate cooperative binding. This should be observable in vitro once the factors are purified. A key feature of the second model is that cooperativity is mediated through interactions with the target factor and not through direct contact of two trans factors. If interaction with the target factor occurs prior to factor binding to the DNA, it could mediate cooperative binding to DNA (*37*). Evidence for the existence of target factors or adapter factors that may mediate an interaction between bound factors and the basic transcription machinery has been reported for other organisms (*38*).

A prediction of both models is that an increase in the number of binding sites for either of the two factors should produce expression in the cells in which the combination confers expression. This is the case for the as-1 element. A construct that contains four copies of the as-1 site upstream of domain A (producing five as-1 sites) confers high expression in leaf, while the same construct with mutations in the TGACG motifs gives an expression pattern identical to that of domain A alone (*24*).

Identification of controlling aspects of a combinatorial code is important, both for the understanding of plant development and for plant biotechnology, in cases where it is essential to have a gene expressed in the correct tissue at the proper time. Whether it will be possible to predict the expression pattern of a gene solely from the knowledge of its sequence is an unknown. However, some of the parameters that determine the transcriptional regulation of a particular gene are now apparent: (i) the number and type of factor binding site; (ii) the affinity of the factor for the site; and (iii) the concentration of active factor in the cell. An understanding of how a cell uses the information contained in DNA-bound proteins to produce regulated gene expression will require empirically derived data on the expression patterns generated by specific combinations of cis elements. As we identify the factors that bind to the 35*S* promoter and determine the sequences with which they interact, we will be able to introduce new combinations into the plant to define precisely the combinatorial properties of cis-regulatory elements that determine developmentally regulated expression.

REFERENCES AND NOTES

1. J. W. Edwards and G. M. Coruzzi, *Annu. Rev. Genet.*, in press.
2. Z. Schwarz-Sommer, P. Huijser, W. Nacken, H. Saedler, H. Sommer, *Science* **250**, 931 (1990).
3. K. R. Yamamoto, *Annu. Rev. Genet.* **19**, 209 (1985).
4. W. S. Dynan, *Cell* **58**, 1 (1989).
5. J. T. Odell, F. Nagy, N.-H. Chua, *Nature* **313**, 810 (1985).
6. R. A. Jefferson, T. A. Kavanagh, M. W. Bevan, *EMBO J.* **6**, 3901 (1987).
7. J. S. Jensen, K. A. Marcker, L. Otten, J. Schell, *Nature* **321**, 669 (1986); R. Kay, A. Chan, M. Daly, J. McPherson, *Science* **236**, 1299 (1987); P. R. Sanders *et al.*, *Nucleic Acids Res.* **4**, 1543 (1987).
8. D. W. Ow *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4870 (1987).
9. J. T. Odell, S. Knowlton, W. Lin, C. J. Mauvais, *Plant Mol. Biol.* **10**, 263 (1988).
10. R.-X. Fang, F. Nagy, S. Sivasubramaniam, N.-H. Chua, *Plant Cell* **1**, 141 (1989).
11. C. Poulsen and N.-H. Chua, *Mol. Gen. Genet.* **214**, 16 (1988).
12. P. N. Benfey, L. Ren, N.-H. Chua, *EMBO J.* **8**, 2195 (1989).
13. ———, *ibid.* **9**, 1677 (1990).
14. ———, *ibid.*, p. 1685.
15. P. N. Benfey and N.-H. Chua, *Science* **244**, 174 (1989).
16. R. B. Goldberg, S. J. Barker, L. Perez-Grau, *Cell* **56**, 149 (1989).
17. P. M. Gilmartin, L. Sarokin, J. Memelink, N.-H. Chua, *Plant Cell* **2**, 369 (1990).
18. C. Kuhlemeier *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4662 (1988).
19. P. Schulze-Lefert, M. Becker-Andre, W. Schulz, K. Hahlbrock, J. L. Dangl, *Plant Cell* **1**, 707 (1989).
20. S. A. Goff *et al.*, *EMBO J.* **9**, 2517 (1990).
21. R. D. Allen, F. Bernier, P. A. Lessard, R. N. Beachy, *Plant Cell* **1**, 623 (1989).
22. P. N. Benfey, H. Takatsuji, L. Ren, D. Shah, N.-H. Chua, *Plant Cell*, in press.
23. J. Peacock *et al.*, *Plant Gene Systems and Their Biology*, vol. 62 of *UCLA Symposium on Mol. Cell. Biol.* (Liss, New York, 1987), pp. 263–277.
24. E. Lam and N.-H. Chua, *Science* **248**, 471 (1990).
25. S. Prat, L. Willmitzer, J. J. Sanchez-Serrano, *Mol. Gen. Genet.* **217**, 209 (1989).
26. E. Lam, P. N. Benfey, P. Gilmartin, R.-X. Fang, N.-H. Chua, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 7890 (1989).
27. F. Katagiri, E. Lam, N.-H. Chua, *Nature* **340**, 727 (1989).
28. W. H. Landschulz, P. F. Johnson, S. L. McKnight, *Science* **240**, 1759 (1988).
29. K. Yamazaki, F. Katagiri, H. Imaseki, N.-H. Chua, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
30. F. Katagiri *et al.*, *Genes Dev.*, in press.
31. G. Neuhaus, personal communication.
32. E. Lam and N.-H. Chua, *Plant Cell* **1**, 1147 (1989).
33. T. Conner, personal communication.
34. S. Schirm, J. Jiricny, W. Schaffner, *Genes Dev.* **1**, 65 (1987); B. Ondek, A. Shepard, W. Herr, *EMBO J.* **6**, 1017 (1987); C. Fromental, M. Kanno, H. Nomiyama, P. Chambon, *Cell* **54**, 943 (1988).
35. S. K. Logan and P. C. Wensink, *Genes Dev.* **4**, 613 (1990).
36. D. Stanojevic *et al.*, *Nature* **341**, 331 (1989); S. B. Carroll, *Cell* **60**, 9 (1990).
37. M. Carey, Y.-S. Lin, M. R. Green, M. Ptashne, *Nature* **346**, 361 (1990); Y.-S. Lin, M. Carey, M. Ptashne, M. R. Green, *ibid.* **345**, 359 (1990).
38. S. L. Berger, W. D. Cress, A. Cress, S. J. Triezenberg, L. Guarente, *Cell* **61**, 1199 (1990); B. F. Pugh and R. Tjian, *ibid.*, p. 1187.

39. We thank L. Ren for expert technical assistance; members of the Chua laboratory for critical reading of the manuscript and many useful discussions; T. Conner and G. Neuhaus for communicating results prior to publication; J. Ma for discussion of models; D. Shah for the EPSPS promoter; and K. Richards for the cloned CaMV genome. Supported by a fellowship from the Helen Hay Whitney Foundation (to P.N.B.) and by a grant from the Monsanto Company.