Plant Physiol. (1992) 98, 387–391
0032-0889/92/98/0387/05/$01.00/0

Received for publication August 9, 1991
Accepted September 5, 1991

Communication

5′-CATGCAT-3′ Elements Modulate the Expression of Glycinin Genes

Jean-Marc Lelievre[1], Luiz O. Oliveira, and Niels C. Nielsen*

United States Department of Agriculture, Agricultural Research Service, Department of Agronomy, Purdue University, West Lafayette, Indiana 47907

ABSTRACT

Promoters of most seed proteins in legumes contain one or more 5′-CATGCAT-3′ elements. To test if these elements have a function in the expression of these genes, the 2.3 kilobase pairs *Gy2* glycinin promoter was ligated to a β-glucuronidase reporter sequence and transformed into tobacco. Elimination of a 5′-CATGCAT-3′ element 101 base pairs upstream from the transcription start site in the construction caused about a 10-fold reduction in the amount of β-glucuronidase activity compared with when the element was present in the gene. Elimination of 1.9 kilobase pairs from the 5′-end of the promoter caused a two- to threefold reduction in activity. The results show that the 5′-CATGCAT-3′ element plays a role in regulating the amount of expression from the gene, but that there are also other factors farther upstream from the gene that affect the level of expression.

A DNA element with the sequence 5′-CATGCAT-3′ has been implicated in the regulation of seed protein genes in legumes (6). It was first identified as part of a sequence that was absent from the 5′-flanking region of a lectin null-allele from *Phaseolus vulgaris* (8). Subsequently, sequences in which 7 of the 8 base pairs in the element were conserved were identified in the first 400 base pairs of promoters in 20 of 21 legume seed protein genes but were absent in the same region of nonseed protein genes. Because of the wide distribution of the 5′-CATGCAT-3′ element among seed protein genes, a more general role for the elements in gene expression was proposed (6).

The participation of 5′-CATGCAT-3′ elements in gene expression is not based on a demonstration of function. Indeed, a number of unsuccessful efforts to demonstrate the binding of nuclear proteins to 5′-CATGCAT-3′ elements have been reported (13, 15). The purpose of the experiments described in this communication was to establish that the 5′-CATGCAT-3′ element affected the expression of a glycinin gene. For this purpose, tobacco was transformed with fusion genes that contained a β-glucuronidase reporter driven by the promoter from a *Gy2* soybean glycinin gene. With an identical fusion gene in which 5′-CATGCAT-3′ was destroyed, it was possible to compare expression from promoters with and without the element. Evidence is presented that modification of the element results in a large decrease in β-glucuronidase activity in mature seeds from transformed tobacco plants.

[1] Current address: ENSAT, Institute National Polytechnique de Toulouse, 145 Avenue de Muret, 31076 Toulouse, France.

MATERIALS AND METHODS

DNA Clones

The promoter from the *Gy2* glycinin gene from soybean was used for these studies. The gene containing it was derived from the genomic clone λDA28–30 (14). *Gy2* was subcloned into pUC8 and sequenced as described by Thanh *et al.* (18).

Tobacco Transformation

Plasmid constructions were purified and used to transform the LB4404 strain of *Agrobacterium tumefaciens* as described by An *et al.* (1). Minipreparations of *Agrobacterium* DNA were isolated from the transformants and checked for rearrangements by restriction mapping. The DNA sequence across the borders where ligation occurred was determined to assure that small deletions or insertions had not occurred. Tobacco of the variety SRI (gift from R. Schilperoort) was transformed by the leaf disc transformation method of Horsch *et al.* (9). Shoots and roots were selected on medium containing 200 and 100 μg/mL kanamycin, respectively. A single leaf from each primary transformant of tobacco was collected and the DNA from it purified. Polymerase chain reactions programmed with this DNA were carried out to ensure that the promoters remained the proper size after transformation. Pairs of DNA primers, one complementary to the β-glucuronidase reporter and the other complementary to the 5′ end of each promoter fusion, were used for this purpose. Seeds from each M1 tobacco transformant were analyzed for kanamycin resistance by germinating and growing them on agar plates containing 150 μg/mL kanamycin.

GUS Activity Measurements

Activity measurements were performed essentially as described by Jefferson (10). Ten milligrams of mature tobacco seeds were extracted per sample, with two replications per plant. Approximately 2.5 μg of protein from the extract was included in each 1 mL reaction mixture. The substrate concentration recommended by Jefferson was increased to 1 mg 4-methylumbelliferyl glucuronide/mL to ensure that changes

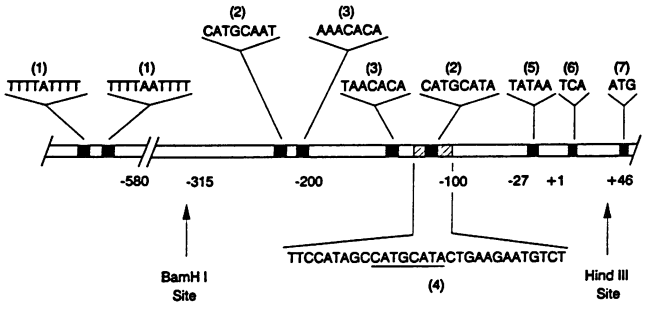

Figure 1. Location of putative regulatory sequences in the promoter of the *Gy2* glycinin gene. 1, A/T rich, *cis*-acting sequences (4); 2, CATGCAT elements (6); 3, CACA elements (7); 4, legumin box (3); 5, TATAA recognition site; 6, transcription initiation; 7, initiation codon.

in fluorescence remained linear over at least a 1-h time interval. Results are expressed in pmol methylumbelliferone/min·μg protein. The protein concentration in seed extracts was determined by the method of Bradford (2).

Statistical Analysis

The differences between means of β-glucuronidase activity measurements were analyzed by a *t* test for samples with unequal variances and an effective number of degrees of freedom (17).

RESULTS AND DISCUSSION

Construction of Fusion Genes

Figure 1 presents a schematic representation of the promoter region from the *Gy2* glycinin gene. In addition to a TATAA region at −27 bp,[2] the promoter contained a perfect 5′-CATGCAT-3′ element at −101 and an imperfect one at −202. The perfect 5′-CATGCAT-3′ element was located centrally in the legumin box (−89 to −119), a 30 bp nucleotide sequence originally described by Baumlein *et al.* (3). A TAACACA element was located immediately upstream from the legumin box at position −124, and a second CACA element was found at position −193, immediately downstream from the 5′-CATGCAA-3′ element. The CACA elements, which were found in promoters from several seed protein genes and preferentially bind nuclear proteins isolated from soybean embryos (7), were originally described in soybean lectin promoters by Jofuku *et al.* (11).

We used a fusion gene consisting of the *Gy2* promoter ligated to a β-glucuronidase reporter element to assess the role of the 5′-CATGCAT-3′ element in glycinin gene expression. The strategy to produce the fusion gene is outlined in Figure 2. The 367 bp *Bam*HI-*Hin*dIII restriction fragment from pGD30H2.3 (Fig. 1) was isolated and introduced into the polylinker of a pUC8 plasmid. This construction yielded plasmid pG30HB400. The *Bam*HI-*Hin*dIII fragment contained 317 bp of DNA 5′ to the translation initiation site (+1). Downstream from the initiation site, the transcript encoded 43 bases of untranslated leader and 12 bases from the *Gy2* coding region, as well as a 17 base leader prior the region encoding β-glucuronidase.

A unique *Nsi*I restriction site was used to eliminate the 5′-CATGCAT-3′ element that was located within the legumin box. The overhang resulting from *Nsi*I digestion was removed with mung bean nuclease, and the resulting blunt ends were ligated to form plasmid pG30HB400Δ. As indicated in Figure 2, nucleotide sequence analysis revealed that 7 bp were removed from the 5′-CATGCAT-3′ element by this procedure. To introduce this construction into the binary vector pBi101.2, plasmids pG30HB400 and pG30HB400Δ were digested with *Hin*dIII, exposed briefly to mung bean nuclease, and then digested with *Bam*HI. Treatment with mung bean nuclease was to place the initial ATG codon from *Gy2* in frame with the β-glucuronidase reporter sequence. After the *Bam*H1-*Hin*dIII fragment was inserted into the binary vector, nucleotide sequence analysis was used to identify the desired clones, and designated pBi400 and pBi400Δ. This construction scheme generated normal and defective *Gy2* promoters that were fused to a β-glucuronidase reporter sequence. In addition to the amino acids in the β-glucuronidase gene, the

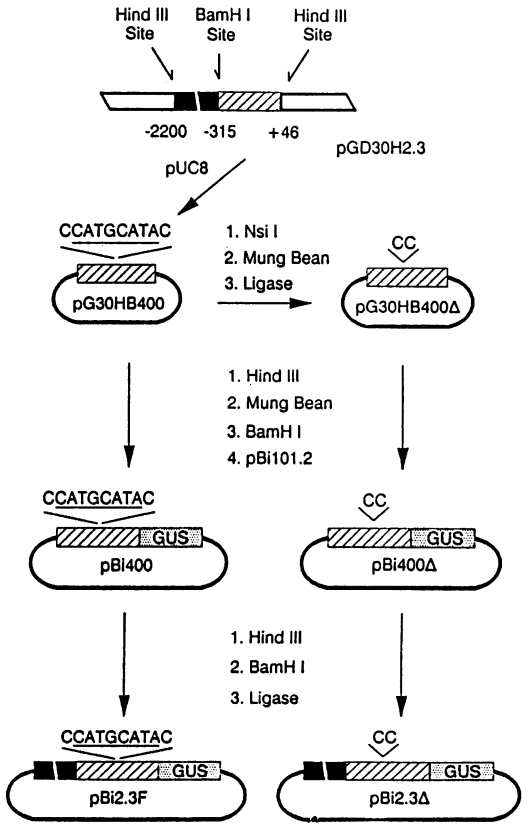

Figure 2. Strategy used to produce promoters with and without the 5′-CATGCAT-3′ element. These manipulations resulted in the elimination of the eight core nucleotides of the CATGCAT element (underlined). DNA sequence analysis was used to demonstrate that nucleotides in the element were eliminated, and that the junctions formed during cloning had the expected sequences. Mung bean, mung bean nuclease.

[2] Abbreviations: bp, base pairs; kb, kilobase pair.

NH_2-terminal ends of polypeptides from the fusion genes contained the first two amino acids from *Gy2* plus six amino acids from linker DNA used to place the *Gy2* promoter in frame to the reporter element. Because regulatory information required for efficient expression of the *Gy2* glycinin gene could be located farther downstream from the transcription initiation site than in the original fusion genes, additional 5′-flanking DNA was included in pBi23F and Bi2.3Δ. This was accomplished by adding the 1.9 kb *Hin*dIII-*Bam*HI fragment that is normally found in the gene immediately 5′ to the 0.4 kb *Hin*dIII-*Bam*HI fragment (Fig. 2). The binary vectors containing these four fusion genes, each of which also contained a kanamycin resistance gene, were then used for transformation of tobacco. Progeny from the primary transformants were tested to identify those that segregated genetically for either single or multiple loci that conferred kanamycin resistance to the plants. Table I summarizes the data obtained. The χ-square test showed that many of the transformants contained a single genetic locus for kanamycin resistance, although several contained two or more such loci.

The β-glucuronidase activities in mature seeds from the transformed tobacco plants were determined. Table I presents the activities associated with each of the transformed plants used in this study. We considered that multiple kanamycin resistance genes would reflect additional copies of the fusion gene, and that these might cause additional β-glucuronidase to be made in seeds from these plants. The presence of additional β-glucuronidase genes in some plants could complicate interpretation of the data. However, examination of the data shown in Table I reveals that this was not the case. Although the presence of additional gene copies may have influenced the magnitude of β-glucuronidase activity present in the seed, these changes were small by comparison with the changes in activity that were caused by elimination of the 5′-CATGCAT-3′ elements.

Table II summarizes a statistical analysis of the average amount of β-glucuronidase activity in seeds from all plants transformed with each of the four fusion genes, and compares these activities with those in seeds from nontransformed tobacco. Because the activity was determined in seeds from several independently transformed plants for each treatment, fluctuations in the level of gene expression that accompanied insertion of genes into different locations within the tobacco genome could be taken into account when the data were analyzed. The results showed that seeds from all plants transformed with fusion genes had more β-glucuronidase activity than seeds from nontransformed plants. Moreover, the activity in seeds from plants transformed with pBi400 and pBi2.3F, which had intact 5′-CATGCAT-3′ elements, not only differed from one another, but also from the lower activities in seeds

Table I. *Characterization of Seeds from Tobacco Transformed with Four Fusion Genes*

Kanamycin resistance was determined as described by Derol and Gardner (5). Plants were scored as green or yellow and analyzed for goodness of fit to a 3:1 (Kan^r:Kan^s) or 15:1 segregation ratio using a χ-square test ($\chi^2 = 6.63$).

Plant	Green	Yellow	Total	Observed Ratio	χ^2	Ratio Tested	β-Glucuronidase Activity[a]
Bi2.3F-1	164	58	222	2.83	0.1501	3:1	37.8133
Bi2.3F-2	47	9	56	5.22	2.3809	3:1	46.7233
Bi2.3F-3	46	16	62	2.87	0.0210	3:1	33.5532
Bi2.3Δ-A1	208	76	284	2.73	0.4694	3:1	2.8124
Bi2.3Δ-C1	>150	<5	>150			≥15:1[b]	1.3991
Bi2.3Δ-E1	113	29	142	3.89	1.5868	3:1	1.3494
Bi2.3Δ-F1	>150	<5	>150			≥15:1[b]	1.8799
Bi2.3Δ-F2	>150	<5	>150			≥15:1[b]	3.2443
Bi400-1	65	24	89	2.70	0.1798	3:1	9.3414
Bi400-2	161	45	206	3.58	1.0938	3:1	3.2011
Bi400-3	171	62	233	2.76	0.3218	3:1	17.1659
Bi400-5	190	9	199	21.11	3.5094	15:1	16.6976
Bi400-6	50	21	71	2.38	0.7933	3:1	7.0982
Bi400Δ-1	119	3	122	39.66	2.9922	15:1[c]	0.7294
Bi400Δ-2	133	53	186	2.51	1.1215	3:1	0.5985
Bi400Δ-3	145	3	148	48.33	4.5042	15:1[c]	1.3436
Bi400Δ-4	183	13	196	14.07	0.0489	15:1	0.6787
Bi400Δ-5	178	9	187	19.77	0.6588	15:1	1.5421
Bi400Δ-6	146	40	186	3.65	1.2115	3:1	0.9559
NT-1	0	193	193				0.0028
NT-2	0	207	207				0.0070
NT-3	0	201	201				0.0063
NT-4	0	190	190				0.0036
NT-5	0	187	187				0.0050

[a] MU/min·mg protein. [b] Estimated, technical problems due to thickness of planting. [c] May be >15:1, but require minimum of 291 seeds to test 63:1 ratio.

Table II. *Statistical Analysis of Means of β-Glucuronidase Activities from Mature Tobacco Seeds Transformed with Four Fusion Genes*

The relationships among the means (Y_i) were analyzed using a t' test for samples with unequal variances (18). Each of the means is different from the next smaller one at either the 0.01* or 0.05** level.

Treatment	Promotor Length	CATGCAT Element	Y_i	s_i^2/ni	df[a]	edf[b]	sig
	kb						
1	2.3	+	39.3633	15.0549	2	4	**
2	0.4	+	10.7008	7.4416	4	4	*
3	2.3	−	2.1370	0.1457	4	5	*
4	0.4	−	0.9747	0.0249	5	5	**
5	0.0	−	0.0049	0.00000062	4		

[a] Degrees freedom. [b] Effective degrees freedom.

of transformants with defective elements. Similarly, the β-glucuronidase activities in seeds containing fusion genes with defective elements were also different from one another.

Several conclusions can be drawn from these data. First, it is apparent that deletion of the 5′-CATGCAT-3′ element from the legumin box results in a 10- to 15-fold reduction in β-glucuronidase activity accumulated in the seeds. This order of magnitude reduction in activity is considered to reflect a similar reduction in the expression of the respective fusion genes. The data also reveal that the activity in seeds of transformants with the defective element is greater than in seeds from nontransformed controls. This result shows that the elimination of the 5′-CATGCAT-3′ element in the legumin box substantially reduces but does not eliminate expression from the fusion gene. Hence, the function of the element appears to modulate expression rather than to enable expression *per se*. Another conclusion from these data is that the amount of β-glucuronidase activity is approximately threefold greater when the 2.3 kb promoter is used instead of the 0.4 kb promoter. This result confirms that some factors important for maximum expression from the gene are located more than 0.4 kb upstream from the site of transcription initiation in *Gy2* (16). A/T-rich motifs, similar to the ones that have been demonstrated to function as general transcriptional activators in the promoters of phaseolin (4, 12), are present in this portion of the *Gy2* promoter and possibly are responsible for the stimulatory effect observed when the 1.9 kb fragment is added to the construction.

The decrease in β-glucuronidase activity that accompanies deletion of the 5′-CATGCAT-3′ element in the fusion genes might be explained in several ways. Nuclear factors that are required for maximal expression from the gene may interact with the element. In this regard, we have been unable to detect binding of nuclear proteins to 5′-CATGCAT-3′ elements by gel displacement experiments, and similar results have been reported by others (13, 15). In any event, because transcription complexes consist of multiple components, interactions within this region of the gene may require more than one kind of protein. The formation of such a complex *in vitro* may be difficult and may account for the lack of binding that we and others have observed. Alternatively, deletion of the element could affect expression indirectly. As an example, conformational changes within the promoter region that are elicited by deletion of the element could cause a down-regulation of expression.

Multiple 5′-CATGCAT-3′ elements are frequently located in promoters from seed protein genes of legumes (6). The promoter from *Gy2* contains two such sequences, one in the legumin box approximately 100 bp from the transcription initiation site, and another approximately 200 bp farther upstream. We refer to these elements as being proximal and distal from the coding region, respectively. The functional relationship, if any, between the two elements is unknown. It is clear from the results, however, that positional effects must be involved, because a high level of gene activity is not maintained when the distal 5′-CATGCAT-3′ element is intact and the proximal one is destroyed. It might be argued that the gene activity remaining when the proximal element is destroyed reflects the presence of the second more distally located one. It should be pointed out, however, that the location of the proximal 5′-CATGCAT-3′ element in a legumin box is well conserved among homologous 11S storage protein genes in legumes, whereas the location of distal elements varies considerably (6). The variation in location of the distal elements suggests that they play a less important role than the 5′-CATGCAT-3′ elements in controlling gene expression.

In summary, the results from these experiments establish a functional role for the 5′-CATGCAT-3′ elements in the 11S storage protein genes from legumes as was predicted earlier (6). They also describe a system by which the regulation of expression for these genes can be investigated in more detail.

LITERATURE CITED

1. **An G, Ebert PR, Mitra A, Ha SB** (1988) Binary vectors. *In* S Gelvin, and R Schilperoort, eds, Plant Molecular Biology Manual A3. Kluwer Academic Publishers, Dordrecht, pp 1–19
2. **Bradford MM** (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem **72:** 248–254
3. **Baumlein H, Wobus U, Pustell J, Kafatos FC** (1986) The legumin gene family: structure of a B type gene of *Vicia faba* and a possible legumin gene specific regulatory element. Nucleic Acids Res **14:** 2707–2720
4. **Bustos M, Guiltinan MJ, Jordano J, Begum D, Kalkan FA, Hall TC** (1989) Regulation of β-glucuronidase expression in transgenic tobacco plants by an A/T-rich, cis-acting sequence found

upstream of a French bean β-phaseolin gene. Plant Cell **1:** 839–853

5. **Derol S, Gardner R** (1988) Expression and inheritance of kanamycin resistance in a large number of transgenic petunias generated by *Agrobacterium* mediated transformation. Plant Mol Biol **11:** 355–364
6. **Dickinson CD, Evans RP, Nielsen NC** (1988) RY-repeats are conserved in the 5′-flanking regions of legume seed-protein genes. Nucleic Acids Res **16:** 371
7. **Goldberg RB, Barker SJ, Perez-Grau L** (1989) Regulation of gene expression during plant embryogenesis. Cell **56:** 149–160
8. **Hoffman LM, Donaldson DD** (1985) Characterization of two *Phaseolus vulgaris* phytohemagglutin genes closely linked on the chromosome. EMBO J **4:** 883–889
9. **Horsch RB, Fry J, Hoffmann N, Neidermeyer J, Rogers SG, Fraley RT** (1988) Leaf disc transformation. *In* S Gelvin, R Schilperoort, eds, Plant Molecular Biology Manual A5. Kluwer Academic Publishers, Dordrecht, pp 1–9
10. **Jefferson RA** (1987) Assaying chimeric genes in plants: the GUS gene fusion system. Plant Mol Biol Rep **5:** 387–405
11. **Jofuku KD, Okamuro JK, Goldberg RB** (1987) Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. Nature **328:** 734–737
12. **Jordano J, Almoguere C, Thomas TL** (1989) A sunflower helianthinin gene upstream ensemble contains an enhancer and sites of nuclear protein interaction. Plant Cell **1:** 855–866
13. **Meakin PJ, Gatehouse JA** (1991) Interaction of seed nuclear proteins with transcriptionally-enhancing regions of the pea (*Pisum sativum* L.) *leg A* gene promoter. Planta **183:** 471–477
14. **Nielsen NC, Dickinson CD, Cho TJ, Thanh VH, Scallon BJ, Fischer RL, Sims TL, Drews GN, Goldberg RB** (1989) Characterization of the Glycinin gene family in soybean. Plant Cell **1:** 313–328
15. **Riggs CD, Voelker TA, Chrispeefs MJ** (1989) Cotyledon nuclear proteins bind to DNA fragments harboring regulatory elements of phytohemagglutinin genes. Plant Cell **1:** 609–621
16. **Shirsat A, Wilford N, Croy R, Boulter D** (1989) Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. Mol Gen Genet **215:** 326–331
17. **Steel RGO, Torrie JH** (1980) Principles and Procedures of Statistics, Ed 2. McGraw-Hill, New York, p 106
18. **Thanh VH, Tumer NE, Nielsen NC** (1989) The glycinin *Gy2* gene from soybean. Nucleic Acids Res **17:** 4387