*The Plant Journal* (1992) **2**(2), 233–239

# *Cis*-analysis of a seed protein gene promoter: the conservative RY repeat CATGCATG within the legumin box is essential for tissue-specific expression of a legumin gene

**Helmut Bäumlein*, Istvan Nagy†, Raimundo Villarroel[1], Dirk Inzé[1] and Ulrich Wobus**
*Institut für Genetik und Kulturpflanzenforschung, Corrensstraße 3, Gatersleben, O-4325, Germany, and [1]Laboratorium voor Genetika, Rijksuniversiteit Gent, Ledegankstraat 35, Gent, B-9000, Belgium*

## Summary

**A 2.4 kb fragment containing the 5′-flanking region and the 5′-noncoding sequence of the *Vicia faba* legumin gene *LeB4* mediates high level seed-specific expression in transgenic tobacco plants. Deleted derivatives of this legumin upstream sequence were fused to the *npt-II* reporter gene to determine the tissue-specific activity of the chimeric constructs in stably transformed tobacco plants. The results indicate the presence of positive regulatory, enhancer-like *cis* elements within 566 bp of the upstream sequence. Most importantly, however, these elements are only fully functional in conjunction with the core motif CATGCATG of the legumin box around position −95, since destruction of the motif by a 6 bp deletion in an otherwise intact 2.4 kb upstream sequence drastically reduces expression in seeds. At the same time, low level expression in leaves is observed. The occurrence of similar CATGCATG consensus *cis* elements with alternating purine and pyrimidine base pairs in front of several other plant genes suggests a functional role of the motif in a wider range of plant promoters.**

## Introduction

Spatially and temporarily regulated gene expression programmes are the basis for development and morphology. The strictly seed-specific and development-dependent expression of seed storage protein genes provides a suitable experimental system to study differential gene activation in plants.

It is generally accepted that the seed specificity of storage protein gene expression is primarily regulated at the transcriptional level, although post-transcriptional processes can modulate the final amount of translational products widely (Goldberg *et al.*, 1989). Current ideas imply complex interactions between specific *trans*-acting transcription factors with their *cis*-acting target DNA sequences as the principal mechanism for transcription regulation. Several DNA fragments derived from the 5′-flanking regions of different seed protein genes have been shown to bind defined nuclear protein factors (Allen *et al.*, 1989; Bustos *et al.*, 1989; Chen *et al.*, 1988; Jofuku *et al.*, 1987; Jordano *et al.*, 1989). However, in most cases a causal relationship connecting *trans* factor binding with regulated promoter activity has not been demonstrated. The availability of extensive sequence data from 5′ flanking regions of storage protein genes isolated from several different species has prompted the search for conserved sequence motifs, assuming that these elements might be involved in *trans* factor binding and therefore in the regulation of seed protein gene expression. Thus several sequence conservative, putative regulatory DNA elements have been identified (for review see Okamuro and Goldberg, 1989); among them the legume 12S globulin gene-specific legumin box (Bäumlein *et al.*, 1986) with the internal, highly conserved RY core motif CATGCATG (Dickinson *et al.*, 1988).

Recently we have shown that about 1.2 kb of the legumin B4 (*LeB4*) gene upstream sequence is sufficient for strong seed-specific activity and that deletion derivatives with only 193 bp and 91 bp of upstream sequence are approximately 10 times less active (Bäumlein *et al.*, 1991a). For a more precise localization of the *cis* elements which might be responsible for this reduction in activity we have constructed and analysed a series of new deletions.

In this paper we present data extending our knowledge of functionally important DNA sequences in the 5′-flanking region of the gene *LeB4*. In particular, we demonstrate that strong legumin promoter activity and probably also strict tissue specificity depend on the integrity of the short conserved CATGCATG sequence motif within the legumin box.

## Results

### *Delineation of* cis*-acting elements by 5′ deletion analysis*

Earlier experimental data (Bäumlein *et al.*, 1991a) demonstrate the presence of functionally important elements in

Received 5 April 1991; revised 26 November 1991.
*For correspondence (fax +37 45592 280).
†Present address: Agricultural Biotechnology Centre, H-2101 Gödöllö, Hungary.

the legumin *LeB4* gene upstream sequence between about 1200 bp (*Cla*I site) and 193 bp (*Eco*RV site) in front of the transcription start site. For a more precise characterization of those elements we have analysed the effect of progressive and internal deletions within about 1 kb of the *LeB4* upstream sequence (see Figure 1) on NPT-II reporter enzyme levels in seeds of stably transformed tobacco plants. As a first approximation we interprete the changes in enzyme activity as a reflection of changes in promoter strength.

As shown in Figure 2 the *LeB4* upstream sequence can be deleted to position −701 without an obvious loss of NPT-II activity. The average enzyme activity seems to drop when the sequence between −701 and −566 is removed. However, this transition is not statistically significant, and neither is the increase in activity between construct −844 and −701. A significant (at the 5% level) reduction in expression level can be detected when the promoter is shortened to −471 bp. The 95 bp sequence between −566 and −471 is AT-rich (73%) and includes the motif ATTAATT which partly satisfies the ATT A/T AAT consensus rule (Jofuku *et al.*, 1987). The *Ppu*MI site at position −492 used for the construction of the two internal deletions PC and PR (see Figure 1) is also located within this sequence. This restriction site overlaps a so-called GC element present in all legumin gene upstream sequences surveyed (Rerie, 1989; Rerie *et al.*, 1991).

Another extremely AT-rich (82%) region was removed to obtain construct −407. The enzyme levels produced by this construct are on average less than 10% compared to those produced by constructs −701 or −844. Construct −333 lacks part of a DNA motif with a 20 out of 25 bp homology (see Figure 1) to a promoter sequence of the mainly seed-specifically expressed *USP* gene of *Vicia faba* (Bäumlein *et al.*, 1991b) with no obvious effect. Another significant (at the 1% level) reduction in the expression level is shown by construct −232 in comparison to construct −279. The removed sequence does not show any obvious peculiarity apart from an 11-bp purine stretch.

The question of whether a minimal promoter completely lacking the conservative legumin box is still functional was addressed by the analysis of construct −68. Construct −68 leads to significantly (at the 1% level) reduced but still measurable NPT-II activities in comparison to construct −151. The sequence between position −151 and −68 bp includes the total legumin box and an imperfect direct repeat (TGTCACACACGTtcTGTCACACGT) between position −83 and −60 with similarity both to a motif reported to be present in front of several plant genes (Memelink *et al.*, 1987) and to the CACA motif often found in the upstream regions of seed protein genes (Okamuro and Goldberg, 1989). The effects of even shorter promoter constructs on NPT-II activities in seeds have been compared in a separate experimental series. A 45-bp long

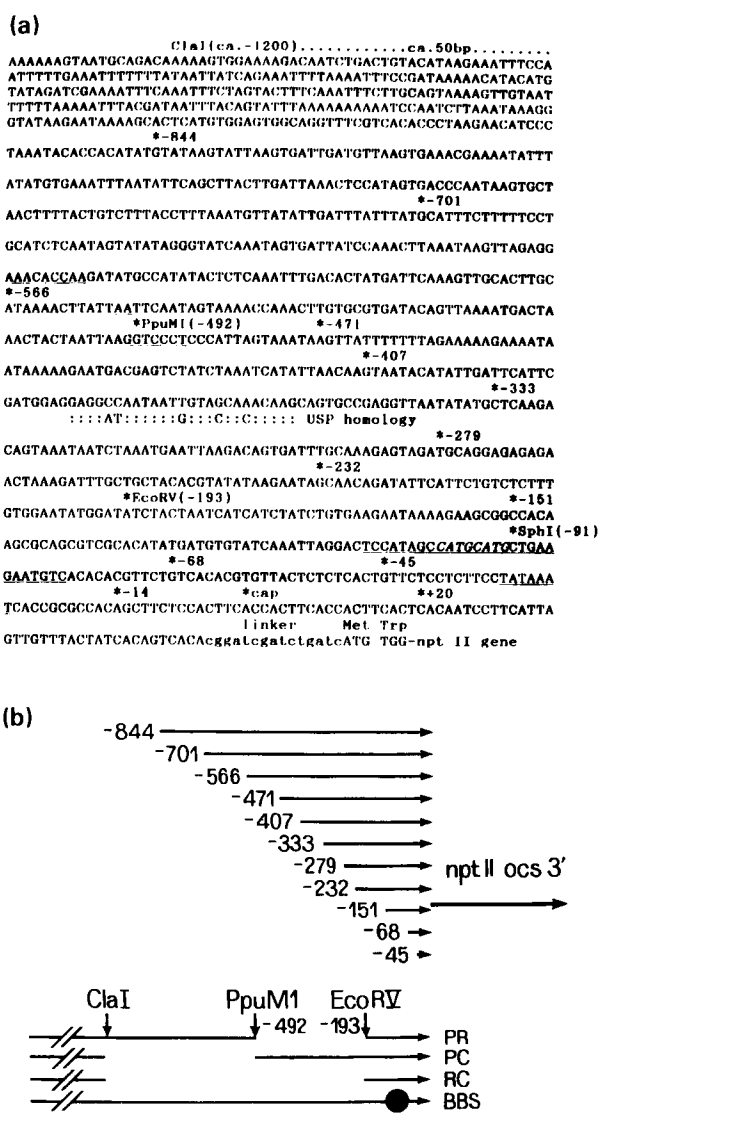


**Figure 1.** Sequence of the 5′-flanking region of *LeB4* and structure of the *LeB4* promoter deletion constructs.

(a) Sequence of the 5′-flanking region of the legumin gene *LeB4* fused to the *npt-II* coding region in the Ti plasmid pGV180 by a linker region. The start points for deletion derivatives are indicated by * above the sequence and the position number is supplemented by a restriction enzyme symbol in case the respective site was used to create the deletion construct. The *Cla*I site indicated at the top has been mapped to about 50 bp upstream of the given sequence but not sequenced itself. Sequence motifs discussed in the text are marked by underlining and the CATGCATG motif within the legumin box is denoted by italics. The linker region between the last nucleotide of the *LeB4* 5′-noncoding region and the first two codons of the *npt-II* reporter gene are printed in lower case letters. The sequence between positions −689 and +56 has already been published by Bäumlein *et al.* (1986).

(b) Schematic structure of the *LeB4* promoter deletion constructs used in this study. The arrow at the right labelled nptII ocs 3′ symbolizes the neomycinphosphotransferase-II reporter gene terminated by the polyadenylation region of the octopine synthase gene. The other arrows represent *LeB4* sequences upstream of the *npt-II* fusion point indicated in (a) and labelled by the respective deletion end-points. Constructs denoted PR, PC, RC and BBS were created by deletions within the total 2.4 kb *LeB4* upstream region by removing the indicated restriction fragments or, in the case of the BBS construct, by deleting 6 bp of the legumin box core motif CATGCATG, as specified in Figure 3.

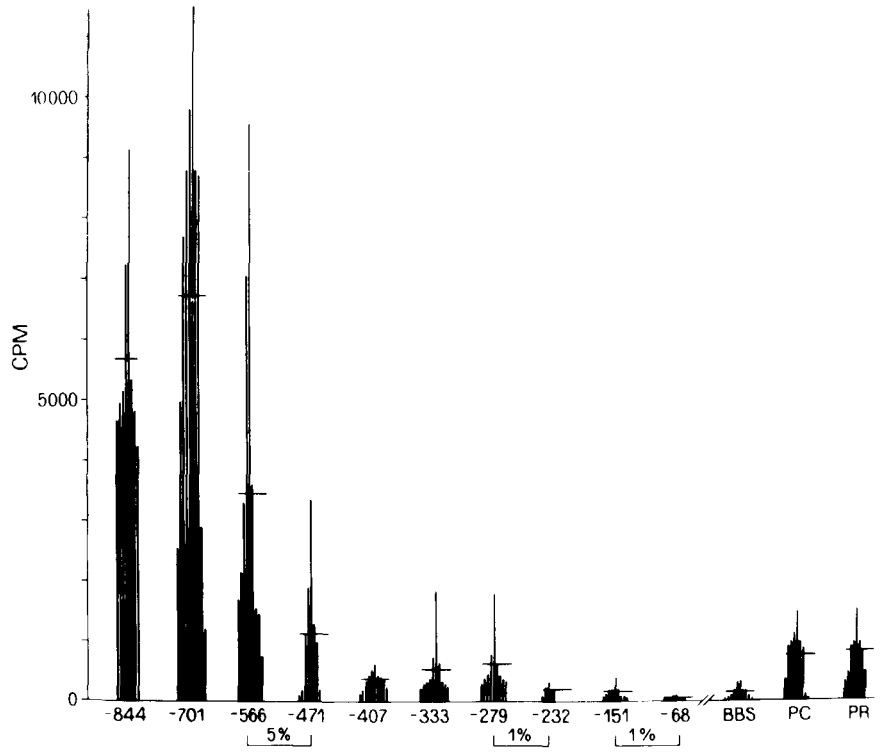


**Figure 2.** NPT-II activity levels measured by the NPT-II gel assay in mature seeds of independent tobacco plant transformants.
The average value from all plants transformed with the constructs indicated below the columns is denoted by a horizontal line. The constructs are defined in Figure 1. Statistically significant differences at the 1% or 5% significance level between consecutive constructs are indicated by brackets. Fifty micrograms of protein were used in each assay. To keep experimental variability as low as possible, all values given were estimated in a single slot-blot experiment.

promoter (construct −45) still causes low NPT-II activity comparable to that of construct −68. Only the removal of the TATA box in construct −14 completely extinguishes the promoter activity. In addition, a cap site deletion (construct +20) is also inactive, as expected (data not shown).

The progressively shortened promoter constructs described above necessarily change the spatial relationship between the transcription start site and the flanking vector-derived sequences as well as the sequences adjacent to the genomic integration site. To reduce the potential influence of spatial changes we have created and analysed deletions within the 2.4 kb 5′ flanking region (see Figure 1). As shown in Figure 2, both the PR and PC constructs show strongly decreased expression. The low activity of PR confirms the data obtained with the progressive deletion constructs −471, −407, −333, −279 and −232 and demonstrates the necessity of sequence elements between −492 (*Ppu*MI site) and −193 (*Eco*RV site) for optimal promoter function. Moreover, the reduced expression of the PC construct indicates that additional sequence elements at or upstream of the *Ppu*MI site quantitatively affect the expression of the legumin promoter. Considering that sequences upstream of position −566 can be deleted without a significant effect (Figure 2), we conclude that those additional sequence elements are localized closely upstream of or even overlapping the *Ppu*MI site at −492.

*The legumin box core motif CATGCATG is essential for seed-specific promoter activity*

Assuming that sequence conservation is an indication of functional importance, it has been suggested that the legumin box and its core motif CATGCATG are crucial for legumin gene expression (Bäumlein *et al.*, 1986; Dickinson *et al.*, 1988). To test this hypothesis experimentally we have used a suitable unique *Sph*I site overlapping the CATGCATG core element of the legumin box to specifically remove 6 bp out of the 8 bp core motif (see Figure 3) in the 2.4 kb *LeB4* upstream sequence (BBS deletion). All of the 10 individual transformants analysed show low NPT-II activity in mature seeds, comparable in intensity to the enzyme levels caused by construct −151 (Figure 2). Surprisingly, seven out of the 10 plants transformed with the BBS construct also showed low NPT-II activity in leaves. Examples are given in Figure 4. In contrast, leaf activity is not found in plants carrying constructs with at least 700 bp proximal to the *LeB4* transcription start site (data not shown). To exclude additional unintended changes within the mutated fragment as a cause for the low and tissue-specifically relaxed NPT-II levels, we have confirmed the overall integrity of all BBS constructs by Southern hybridization. Moreover, the removal of the former *Sph*I site was proven by the resistance to *Sph*I treatment of a legumin box containing PCR fragment amplified from genomic DNA of BBS-transformed tobacco

```
GmGly      TCCATAGCCATGCATACTGAAGAATG
PsLegA     TCCATAGCCATGCAAGCTGCAGAATG
PsLegJ     TCCATAGCCATGCATGCTGAACAATG
VfLegB     TCCATAGCCATGCATGCTGAAGAATG
BBS        TCCATAGCCA******CTGAAGAATG
GmßCG          AGCCATGCA
                 CCATGCATG
Asglo5           TCAT-CATG
ZmC1           TCCATGCATGCAC
               TGCATGCATGCAC
ZmRAB17    TCCACTCATGCAT
               CTCATGCATGCCC
OsRAB16    TCCACCCATGCCG
TsEm           TGCATGCATGCAA
Gmaux22          CATGCAT
SV40           AAGCATGCATCTC
               AAGTATGCA
```

**Figure 3.** CATGCATG-like motifs are present in front of several plant genes as well as in the *Sph*I element of the SV40 enhancer.
Abbreviations: GmGly, *Glycine max*, glycinin gene; PsLegA, *Pisum sativum* legumin A gene; PsLegJ, *P. sativum*, legumin J and K genes (Thompson *et al.*, 1991); VfLegB, *Vicia faba*, legumin B gene (Bäumlein *et al.*, 1986); BBS, 6 bp deletion within the legumin box (this paper); GmβCG, *G. max*, β-conglycinin gene (Harada *et al.*, 1989); Asglo5, *Avena sativa*, 12S globulin gene (Schubert *et al.*, 1990); ZmC1, *Zea mays*, *C1* regulator gene of anthocyan synthesis (Paz-Ares *et al.*, 1987); ZmRAB17, *Z. mays*, abscisic acid-induced gene (McCarty, personal communication); OsRAB16, *Oryza sativa*, abscisic acid-responsive gene (Mundy *et al.*, 1990); TsEm, *Triticum aestivum*, abscisic acid-induced wheat gene (McCarty, personal communication); Gmaux22, *G. max*, auxin-regulated gene (Ainley *et al.*, 1988); SV40, *Sph*I element in the simian virus 40 enhancer (Zenke *et al.*, 1986).

plants. Thus experimental data clearly demonstrate that the destruction of the conservative RY motif CATGCATG within the 2.4 kb upstream region strongly disturbs the function of the legumin B4 promoter.

### *The AT-rich RC fragment enhances the activity of a truncated foreign promoter*

Earlier experimental data (Bäumlein *et al.*, 1991a) demonstrate that the generally AT-rich region between positions

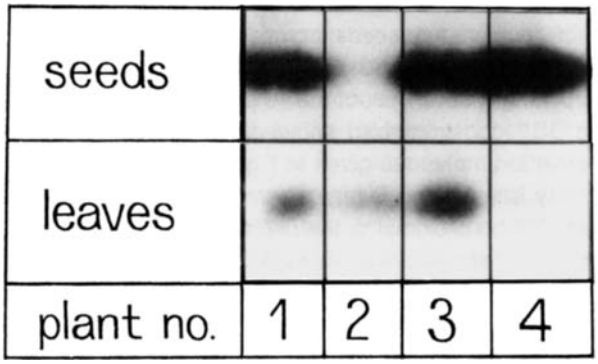


**Figure 4.** NPT-II activity in seeds and leaves of four individual plants transformed with the BBS construct.
All extracts containing 50 μg of protein were assayed on the same gel and the autoradiogram exposed for 3 days. Note that there is no obvious correlation in activity between seeds and leaves in each construct.

−1200 (*Cla*I site) and −193 (*Eco*RV site) exhibit a clear-cut quantitative effect on the basic *LeB4* promoter. The same promoter region can also co-operate with the truncated *nos* promoter in the Ti-plasmid-derived vector pGV300 which contains 148 bp of upstream sequence still including the b, a, z and reversed b element configuration (Ebert *et al.*, 1987). As shown in Table 1, the RC fragment in constructs RCD+ and RCD− enhances the activity of the truncated *nos* promoter in leaves more than 25-fold, independent of its orientation. Surprisingly, in seed tissue, the enhancing effect is only two- to fourfold; the difference in NPT-II activity between the two orientations is not statistically significant. In contrast to the RC fragment, a legumin box containing *Mbo*II fragment (positions −155 to −77 in Figure 1) in constructs LBL+ and LBL− does not (or only weakly) interact with the *nos* promoter in seeds, whereas in leaves the reverse but not the natural orientation increases *nos* promoter activity about sixfold (Table 1).

## Discussion

### *Several upstream elements quantitatively influence the legumin promoter activity*

The functional analysis in transgenic tobacco plants of a series of deletions covering about 1 kb of *LeB4* gene upstream sequences specifies further earlier conclusions about *LeB4* promoter regulation (Bäumlein *et al.*, 1991a). As shown in Figure 2 there is a highly significant decrease in activity when constructs −701 and −471 are compared. The data suggest that the region distal of −566, but certainly distal of −701, is of little importance for high promoter activity in contrast to the region downstream of bp −566. This region up to bp −407 is rich in AT base pairs. Similar AT-rich sequences have been described as being involved in the regulation of genes coding for seed and other plant proteins. These sequences preferentially interact with high mobility group (HMG) proteins which seemingly recognize certain structural features instead of specific primary sequences (reviewed by Weising and Kahl, 1991). Within these AT-rich sequences lies the *Ppu* MI site-overlapping, evolutionary conserved GC element AAGGTCCCT (Rerie, 1989; Rerie *et al.*, 1991). We take its sequence conservation and the reduced NPT-II activity of the PR and PC constructs (Figure 2), in which either the 5′ or the 3′ part of the *Ppu*MI site are removed, as an indication of the functional importance of the GC element.

Another significant transition in activity, although already at a low level, occurs when the fragment between positions −151 and −68, containing the legumin box, is removed (Figure 2). Whereas the BBS deletion clearly reveals the importance of the legumin box core motif CATGCATG (see below) the role, if any, of the additional box sequences

**Table 1.** Effect of *LeB4* upstream region fragments on a truncated *nos* promoter in transgenic tobacco plants

| | Mean (± SEM) value of NPT-II activity (c.p.m.) | | Total number of plants |
|---|---|---|---|
| | Seed | Leaf | |
| pGV300 | 120 ± 12 | 343 ± 111 | 9 |
| LB14 | 179 ± 17* | 344 ± 46 | 2 |
| LB11 | 139 ± 11 | 2038 ± 743** | 10 |
| RCD37 | 270 ± 64** | 9790 ± 3975** | 4 |
| RCD2 | 500 ± 195** | 9202 ± 3298** | 4 |

pGV300 is the control Ti plasmid described in Experimental procedures containing the truncated *nos* promoter fused to the *npt-II* gene. We fused either the legumin box containing promoter fragment −156 to −77 in the natural (LBL+) or the inverse orientation (LBL−), or the RC fragment (−1200 to −193), again in either the natural (RCD+) or the inverse orientation (RCD−) in front of the truncated promoter. Fifty micrograms of protein extracted from leaves or mature seeds of a total of 29 transgenic plants were analysed by the NPT-II gel assay. Significant difference, at the *5% or **1% level, between a given construct and the control pGV300 in either seeds or leaves as calculated by the Mann–Whitney U test.

around the core motif remains undefined. The low but significant NPT-II activities in seeds of plants transformed with the *LeB4* promoter constructs −232, −151 and −68 are in contrast to results reported by Shirsat *et al.* (1989) and Rerie *et al.* (1991). These authors tested promoter deletions of the pea legumin gene *Leg1* by estimating Leg1 protein levels in transgenic tobacco seeds and were unable to detect any expression when upstream sequences of only 97 bp, 124 bp and 237 bp control legumin expression. The difference between these and our results may be explained by the lower detection sensitivity of the immunological technique used by Shirsat *et al.* (1989) and Rerie *et al.* (1991), although differences due to the constructs used (intact gene versus chimeric gene) cannot be excluded. Presently we cannot fully explain the results of *nos* promoter stimulation by *LeB4* promoter fragments (see Table 1) but we initially conclude that (i) the AT-rich RC fragment contains sequences which meet the criteria for enhancers (Müller *et al.*, 1988) in stimulating the foreign minimal *nos* promoter in an orientation-independent manner, especially in leaves, and (ii) there is no element within the RC fragment acting as a *seed-specific* enhancer in the given construct.

### *The CATGCATG motif – a key element of the legumin gene promoter*

The sequence motif CATGCATG is conserved among legume seed protein genes (Dickinson *et al.*, 1988) and is part of the 28 bp legumin box found in front of genes coding for 12S legume seed globulins (Bäumlein *et al.*, 1986). The exclusive deletion of 6 out of 8 bp of the CATGCATG motif within the 2.4 kb *LeB4* upstream sequence in front of the *npt-II* reporter gene leads to a dramatic reduction of NPT-II enzyme levels (see Figure 2). However, since similar reductions are caused by progressive deletions (−232, −151) leaving the CATGCATG motif intact, we conclude that this motif is necessary but not sufficient for optimal promoter function. These data also explain why we were unable to demonstrate the functional importance of the legumin box using progressive deletions only (Bäumlein *et al.*, 1991a) and imply that the legumin box core element CATGCATG can only function properly in co-operation with additional upstream elements.

Destruction of the CATGCATG motif also causes low NPT-II activity in leaves of BBS plants (Figure 4). Such leaf activity has been already observed in plants carrying the RC deletion construct (see Figure 1) as well as −193 and −91 constructs (Wobus *et al.*, 1989). Relaxed tissue specificity was also reported for shortened patatin-1 promoter constructs (Jefferson *et al.*, 1990) and for a truncated anonymous root-specific promoter (Koncz *et al.*, 1989). We favour the idea that the *LeB4* promoter loses its tissue specificity when the promoter is turned down by the removal or destruction of important *cis* elements. However, we have still not rigorously excluded other explanations, such as an unknown role of the *npt-II* coding sequence, as described for mammalian cells by Artelt *et al.* (1991).

### *The CATGCATG motif also occurs in other plant gene promoters*

Although originally described as an element specific for legume seed protein genes, here we suggest that the CATGCATG motif acts as a functional module in a wider range of plant promoters. Figure 3 shows its physical

presence within the upstream regulatory sequences of several plant genes as well as the SV40 *Sph*I enhancer motif. At least for the maize *C1* gene it was shown that the CATGCATG sequence is crucial for its regulation by the *viviparous* gene product Vp1 (McCarty and Carson, 1991; McCarty, personal communication). We presently favour the idea that either a CATGCATG-binding transcription factor or a structural peculiarity due to the alteration of purine and pyrimidine bases, or both, are involved in the integration of a functional transcription complex in seed tissue.

## Experimental procedures

### *Plasmid constructs*

Standard cloning, construction and sequencing techniques have been performed following the guidelines given in Ausubel *et al.* (1987) and Sambrook *et al.* (1989). The starting point for the generation of progressively deleted promoter fragments was the plasmid p4/12BB, described previously (Bäumlein *et al.*, 1991a). p4/12BB contains, beside pUC18 vector sequences, a 2.4 kb upstream region with unique restriction sites for *Cla*I (around −1200), *Ppu*MI (−492), *Eco*RV (−193) and *Sph*I (−91) plus the complete 56 bp 5′-untranslated region of gene *LeB4*. The whole fragment is flanked by an upstream *Eco*RI/*Bgl*II/*Sma*I linker sequence and a downstream *Bam*HI site. After cleavage at the *Cla*I site, p4/12BB was partially digested with *Bal* 31, re-cut with *Sma*I and recircularized. The deletion end-points were determined by the Sanger sequencing technique.

To create the BBS construct, plasmid p4/12BB was cut at the *Sph*I site overlapping the CATGCATG sequence motif, treated with T4 DNA polymerase to resect the 3′ protruding ends, and recircularized. The cloned products were sequenced to analyse the extent of the deletion.

The *Ppu*MI site, dividing the *Eco*RV/*Cla*I fragment (RC) into a distal and a proximal part was used to create the two internal deletions, PC (removing the distal *Cla*I/*Ppu*MI fragment) and PR (removing the proximal *Ppu*MI/*Eco*RV fragment).

The deleted promoter fragments were isolated as *Bgl*II/*Bam*HI fragments and cloned in the right orientation into the *Bgl*II site of the intermediate vector pGV180 containing a promoterless *npt-II* gene (see Bäumlein *et al.*, 1991a). Another strategy was applied to create the three promoter constructs −45, −14 and +20. In this case, the unique *Cla*I site of the plasmid pGV180/legP FL (Bäumlein *et al.*, 1991a), containing the same *LeB4* sequences as plasmid p4/12BB described above, was used as the start point for the partial *Bal* 31 digestion. Again the digestion products were cut with *Sma*I to remove the upstream sequences, gel-purified, recircularized, transformed and the deletion end-points determined by sequence analysis.

To test the influence of several *LeB4* promoter fragments on a truncated foreign promoter, we used the enhancer trap vector pGV300, originally designed by Allan Caplan, Rijksuniversiteit Gent. In this plasmid, which was derived from the pGV180 vector (Bäumlein *et al.*, 1991a; Herman *et al.*, 1986) the *npt*-II reporter gene is driven by a truncated *nos* promoter. Using a suitable *Sst*II site the *nos* promoter was shortened to a length of 148 bp, still including the b, a, z and reversed b sequence elements described to be important for (albeit reduced) promoter activity (Ebert *et al.*, 1987). Both the *Eco*RV/*Cla*I fragment (RC) and the legumin box-containing *Mbo*II fragment (LBL) spanning from position −156 to −77 have been cloned in either orientation in front of this truncated *nos* promoter.

### *Plant transformation*

The intermediate plasmids were transferred into the *Agrobacterium* strain pGV2260 by triparental mating and used for leaf disc transformation of *Nicotiana tabacum* cv. Havana as described previously (Bäumlein *et al.*, 1991a). The integrity of all constructs was checked both in *Agrobacterium* and in the plants using Southern hybridization and PCR techniques.

### *NPT-II assays*

NPT-II activity was detected in 100 mg of tissue. Equal amounts of protein determined by the Bradford assay, were assayed for NPT-II activity either by the gel test (Reiss *et al.*, 1984) or the dot technique (Platt and Yang, 1987). For quantification, the radioactivity of cut filter spots was counted. Seed NPT-II activity was determined from each individual transformant and the grouped values compared by the Mann–Whitney U test. In another experiment, equal amounts of seeds (100 mg each) of all transformants harbouring the same construct were mixed, extracted and analysed on a single gel. The prinicipal results (not shown) did not deviate from those shown in Figure 2 for individual transformants.

## Acknowledgements

We thank Andreas Czihal, Elke Liemann, Angela Stegmann, Hannelore Weiß and Thea Tölle for excellent technical assistance, Drs Allan Caplan and Diana Jofuku for helpful discussions, Dr Ute Heim for assistance with DNA sequencing, and Dr Kirk Schnorr for critical reading of the manuscript. UW acknowledges the support given by the 'Fond der Chemischen Industrie'. HB would also like to thank Professor Marc Van Montagu for his hospitality and advice.

## References

**Ainley, W.M., Walker, J.C., Nagao, R.T. and Key, J.L.** (1988) Sequence and characterization of two auxin regulated genes from soybean. *J. Biol. Chem.* **263**, 10658–10666.

**Allen, R.D., Bernier, F., Lessard, P.A. and Beachy, R.N.** (1989) Nuclear factors interact with a soybean β-conglycinin enhancer. *Plant Cell*, **1**, 623–631.

**Artelt, P., Grannemann, R., Stocking, C., Friel, J., Bartsch, J. and Hauser, H.** (1991) The prokaryotic neomycin-resistance-encoding gene acts as a transcriptional silencer in eukaryotic cells. *Gene*, **99**, 249–254.

**Ausubel, F.N., Brent, R., Kingston, R.E., Moore, D.D., Seichmann, J.G., Smith, J.A. and Struhl, K.** (1987) *Current Protocols in Molecular Biology*. New York: John Wiley and Sons.

**Bäumlein, H., Wobus, U., Pustell, J. and Kafatos, F.C.** (1986) The legumin gene family: structure of a B-type gene of *Vicia faba* and a possible legumin gene specific regulatory element. *Nucl. Acids Res.* **14**, 2707–2720.

**Bäumlein, H., Boerjan, W., Nagy, I., Panitz, R., Inzé, D. and Wobus, U.** (1991a) Upstream sequences regulating legumin gene expression in heterologous transgenic plants. *Mol. Gen. Genet.* **225**, 121–128.

**Bäumlein, H., Boerjan, W., Nagy, I., Bassüner, R., Van Montagu, M., Inzé, D. and Wobus, U.** (1991b) A novel seed protein gene

from *Vicia faba* is developmentally regulated in transgenic tobacco and *Arabidopsis* plants. *Mol. Gen. Genet.* **225**, 459–467.

**Bustos, M.M., Guiltinau, M.J., Jordano, J., Begum, D., Kalkan, F.A. and Hall, T.C.** (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.

**Chen, Z.L., Naito, S., Nakamura, I. and Beachy, R.N.** (1988) Regulated expression of genes encoding soybean β-conglycinin in transgenic plants. *Devel. Genet.* **10**, 112–122.

**Dickinson, C.D., Evans, R.P. and Nielsen, N.C. (1988)** RY repeats are conserved in the 5′-flanking region of legume seed protein genes. *Nucl. Acids Res.* **16**, 371.

**Ebert, P.R., Ha, S.B. and An, G.** (1987) Identification of an upstream element in the nopaline synthase promoter by stable and transient assays. *Proc. Natl Acad. Sci. USA*, **84**, 5745–5749.

**Goldberg, R.B., Barker, S.J. and Perez-Grau, L.** (1989) Regulation of gene expression during plant embryogenesis. *Cell*, **56**, 149–160.

**Harada, J.J., Barker, S.J. and Goldberg, R.B.** (1989) Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell*, **1**, 415–425.

**Herman, L.M.F., Van Montagu, M. and Depicker, A.G.** (1986) Isolation of tobacco DNA segments with plant promoter activity. *Mol. Cell. Biol.* **6**, 4486–4492.

**Jefferson, R., Goldbrough, A. and Bevan, M.** (1990) Transcriptional regulation of a patatin-1 gene in tobacco. *Plant Mol. Biol.* **14**, 995–1006.

**Jofuku, K.D., Okamuro, J.K. and Goldberg, R.B.** (1987) Interaction of an embryo DNA binding protein with a soybean lectin gene upstream region. *Nature*, **328**, 734–737.

**Jordano, J., Almoguera, C. and Thomas, T.L.** (1989) A sunflower helianthinin gene upstream sequence ensemble contains an enhancer and sites of nuclear protein interaction. *Plant Cell*, **1**, 855–866.

**Koncz, C., Martini, N., Mayerhofer, R., Koncz-Kalman, Z., Körber, H., Redei, G.P. and Schell, J.** (1989) High-frequency T-DNA-mediated gene tagging in plants. *Proc. Natl Acad. Sci.* **86**, 8467–8471.

**McCarty, D.R. and Carson, C.B.** (1991) The molecular genetics of seed maturation in maize. *Physiol. Plant.* **81**, 267–272.

**Memelink, J., de Pater, B.S., Hoge, J.H.C. and Schilperoort, R.A.** (1987) T-DNA hormone biosynthetic genes: phytohormones and gene expression in plants. *Devel. Genet.* **8**, 321–337.

**Müller, M.M., Gerster, T. and Schaffner, W.** (1988) Enhancer sequences and the regulation of gene transcription. *Eur. J. Biochem.* **176**, 485–495.

**Mundy, J., Yamaguchi-Shinozaki, K. and Chua, N.H.** (1990) Nuclear proteins bind conserved elements in the abscisic acid responsive promoter of a rice rab gene. *Proc. Natl Acad. Sci. USA*, **87**, 1406–1410.

**Okamuro, J.K. and Goldberg, R.B.** (1989) Regulation of gene expression: general principles. In *The Biochemistry of Plants*, Volume 15. San Diego: Academic Press, pp. 1–82.

**Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P.A. and Saedler, H.** (1987) The regulatory *C1* locus of *Zea mays* encodes a protein with homology to *myb* proto-oncogene products and with structural similarities to transcriptional activators. *EMBO J.* **6**, 3553–3558.

**Platt, S.G. and Yang, N.S.** (1987) Dot assay for neomycin phosphotransferase activity in crude cell extracts. *Anal. Biochem.* **162**, 529–535.

**Reiss, B., Sprengel, R., Will, H. and Schaller, H.** (1984) A new sensitive method for quantitative and qualitative assay of neomycin phosphotransferase in crude cell extracts. *Gene*, **30**, 211–218.

**Rerie, W.G.** (1989) *The structure of pea legumin genes and their expression in transgenic tobacco.* Ph.D. Thesis. Canberra: Australian National University.

**Rerie, W.G., Whitecross, M. and Higgins, T.J.V.** (1991) Developmental and environmental regulation of pea legumin genes in transgenic tobacco. *Mol. Gen. Genet.* **225**, 148–157.

**Sambrook, J., Fritsch, E.F. and Maniatis, T.** (1989) *Molecular Cloning – a Laboratory Manual.* New York: Cold Spring Harbour Laboratory Press.

**Schubert, R., Bäumlein, H., Czihal, A. and Wobus, U.** (1990) Genomic sequence of a 12 S seed storage protein gene from oat (*Avena sativa* L. cv. 'Solidor'). *Nucl. Acids Res.* **18**, 377.

**Shirsat, A.H., Wilford, N., Croy, R.R.D. and Boulter, D.** (1989) Sequences responsible for the tissue specific promoter activity of a pea legumin gene in tobacco. *Mol. Gen. Genet.* **15**, 326–331.

**Thompson, A.J., Bown, D., Yaish, S. and Gatehouse, J.** (1991) Differential expression of seed storage protein genes in the pea legJ subfamily, sequence of the gene legK. *Biochem. Physiol. Pflanzen*, **187**, 1–12.

**Weising, K. and Kahl, G.** (1991) Towards an understanding of plant gene regulation: the action of nuclear factors. *Z. Naturforsch.* **46c**, 1–11.

**Wobus, U., Bäumlein, H., Inzé, D. and Nagy, I.** (1989) *Vicia faba* storage protein genes and their promoter activity in transgenic tobacco plants. In *Applied Plant Molecular Biology* (Galling, G., ed.). Braunschweig: Technical University, pp. 98–103.

**Zenke, M., Grundström, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. and Chambon, P.** (1986) Multiple sequence motifs are involved in SV40 enhancer function. *EMBO J.* **5**, 387–397.

EMBL Data Library accession number X03677.