RAP-1 is an Arabidopsis myc-like R protein homologue, that binds to G-box sequence motifs

Pater, B.S. de; Pham, K.; Memelink, J.; Kijne, J.W.

Plant Molecular Biology **34:** 169–174, 1997.

Short communication

RAP-1 is an *Arabidopsis* MYC-like R protein homologue, that binds to G-box sequence motifs

Sylvia de Pater[1,2,*], Khanh Pham[1], Johan Memelink[2,3] and Jan Kijne[1,3]
[1]*Center for Phytotechnology and* [3]*Institute of Molecular Plant Sciences, Leiden University, Wassenaarseweg 64, 2333 AL Leiden, Netherlands (*[*]*author for correspondence);* [2]*Laboratory of Plant Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA*

Received 10 September 1996; accepted in revised form 21 January 1997

Key words: Arabidopsis thaliana, bHLH domain, DNA binding, MYC-like protein, R gene family

Abstract

An *Arabidopsis* cDNA clone encoding a DNA-binding protein, RAP-1, was isolated by southwestern screening of an *Escherichia coli* cDNA expression library. The protein contains a bHLH DNA-binding domain and is homologous to R proteins, regulating anthocyanin biosynthesis. RAP-1 binds to the sequence CACNTG. It is encoded by a single gene, which is expressed to high levels in root and stem and to low levels in leaf and flower. No expression could be detected in siliques. *Rap-1* does not correspond to one of the known loci involved in anthocyanin biosynthesis, since it is located at a different map position. In contrast to the maize R protein Lc, RAP-1 did not induce anthocyanin biosynthesis in pea cotyledons. Thus, RAP-1 is a novel member of the bHLH class of DNA-binding proteins.

Sequence-specific DNA-binding proteins regulate gene expression. Several different classes of plant DNA-binding proteins have been identified [14]. One of these classes consists of proteins containing a basic region/helix-loop-helix structure (bHLH), and is sometimes referred to as the myc homology family. The HLH region is required for dimerization and the adjacent basic region is required for DNA binding [3, 4]. In plants, genes involved in the regulation of anthocyanin biosynthesis were found to encode myc homologues. The so-called R genes of maize were first identified [12] followed by R homologues of petunia and snapdragon [8]. Together with another class of transcription factors, with homology to myb proto-oncogene-type proteins, they control expression of structural genes that are required for anthocyanin biosynthesis in the different parts of the plant [8]. Most of the plant bHLH proteins characterized to date are from maize, because of the long history of genetic analysis of anthocyanin biosynthesis in this plant. R proteins in maize are encoded by a gene family of several different genes, which are differentially expressed, thereby pigmenting a specific set of tissues [12]. The various R proteins are functionally equivalent and anthocyanin biosynthesis can be induced by ectopically expressing one of the R proteins, not only in maize [11] but also in tobacco and *Arabidopsis* [10].

An *Arabidopsis thaliana* (ecotype C24) cDNA expression library was screened for DNA-binding proteins with the DNA sequences W1 and W2. The W1 element is derived from the pea lectin promoter and confers high expression in tobacco seeds [16]. W2 consists of a tetramer of the 12 bp odd-base C-box present in the 22 bp W1 element [15]. Three cDNA clones were obtained corresponding to the same gene encoding a MYC-like bHLH protein. The longest cDNA is 2677 bp and contains a poly-(A) tail of 8 bp (Figure 1). It contains an open reading frame encoding a protein of 623 amino acids with a predicted molecular mass of 67.9 kD. The cDNA clone is close to full-length, because there is an in frame stop codon preceding the predicted start codon. The encoded protein is 63% identical with a recently isolated bHLH protein (PG1)

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X99548 (*Rap-1*).

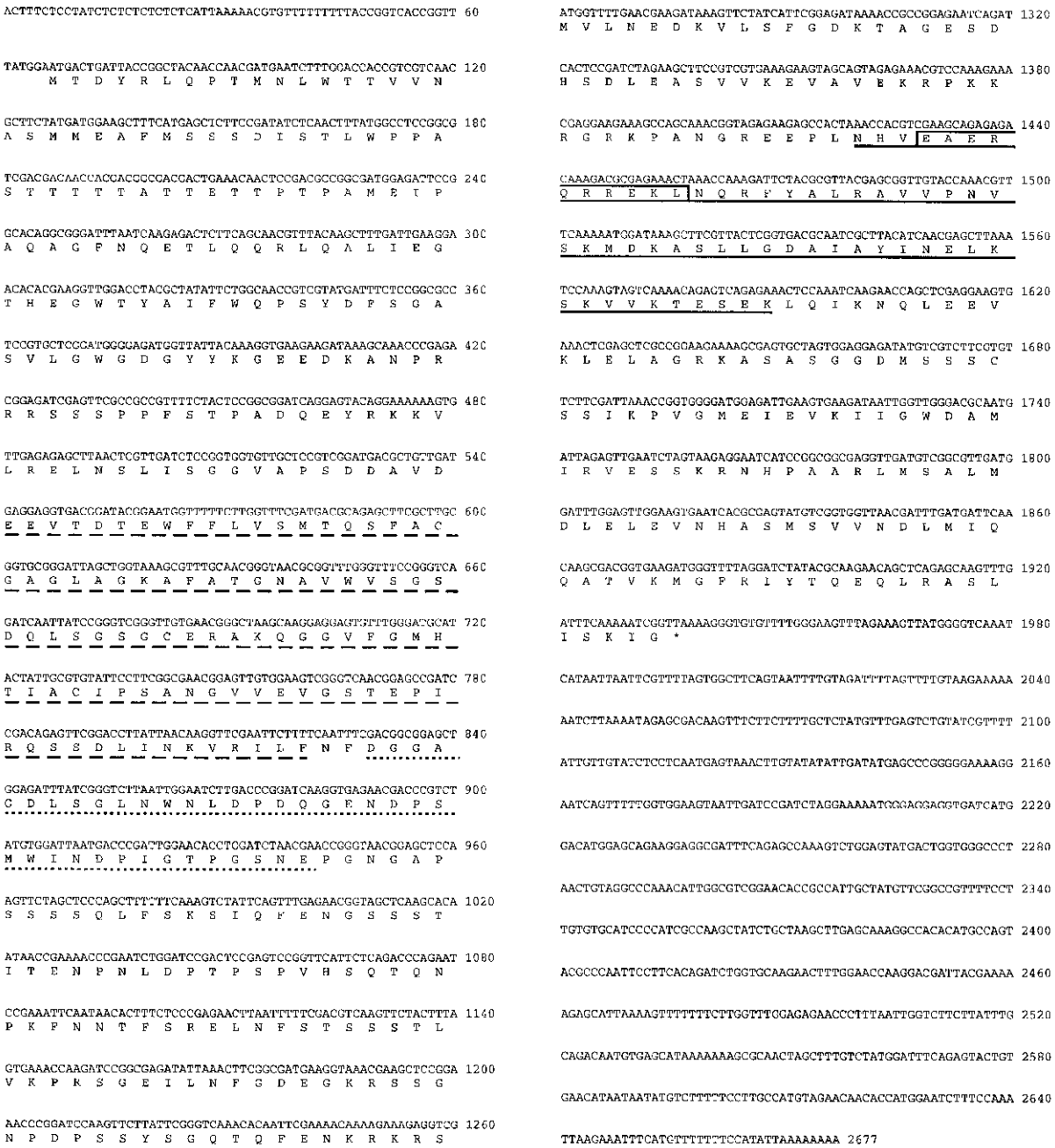

Figure 1. Nucleotide sequence of the *Arabidopsis* cDNA encoding RAP-1 and derived amino acid sequence. The bHLH domain is underlined with a solid line, the region homologous to various R proteins with a broken line and the acidic domain with a dotted line. The region homologous to one of the Lc nuclear localization signals is boxed.

from bean [7] (Figure 2). The major homology with other MYC-like proteins was restricted to the bHLH domain. However, there was another region in the protein with homology to R proteins from various plant species (Figures 1 and 2). In R proteins this conserved region extends to the N-terminus [18] and was found to be necessary for transactivation of anthocyanin structural gene promoters via protein-protein interaction

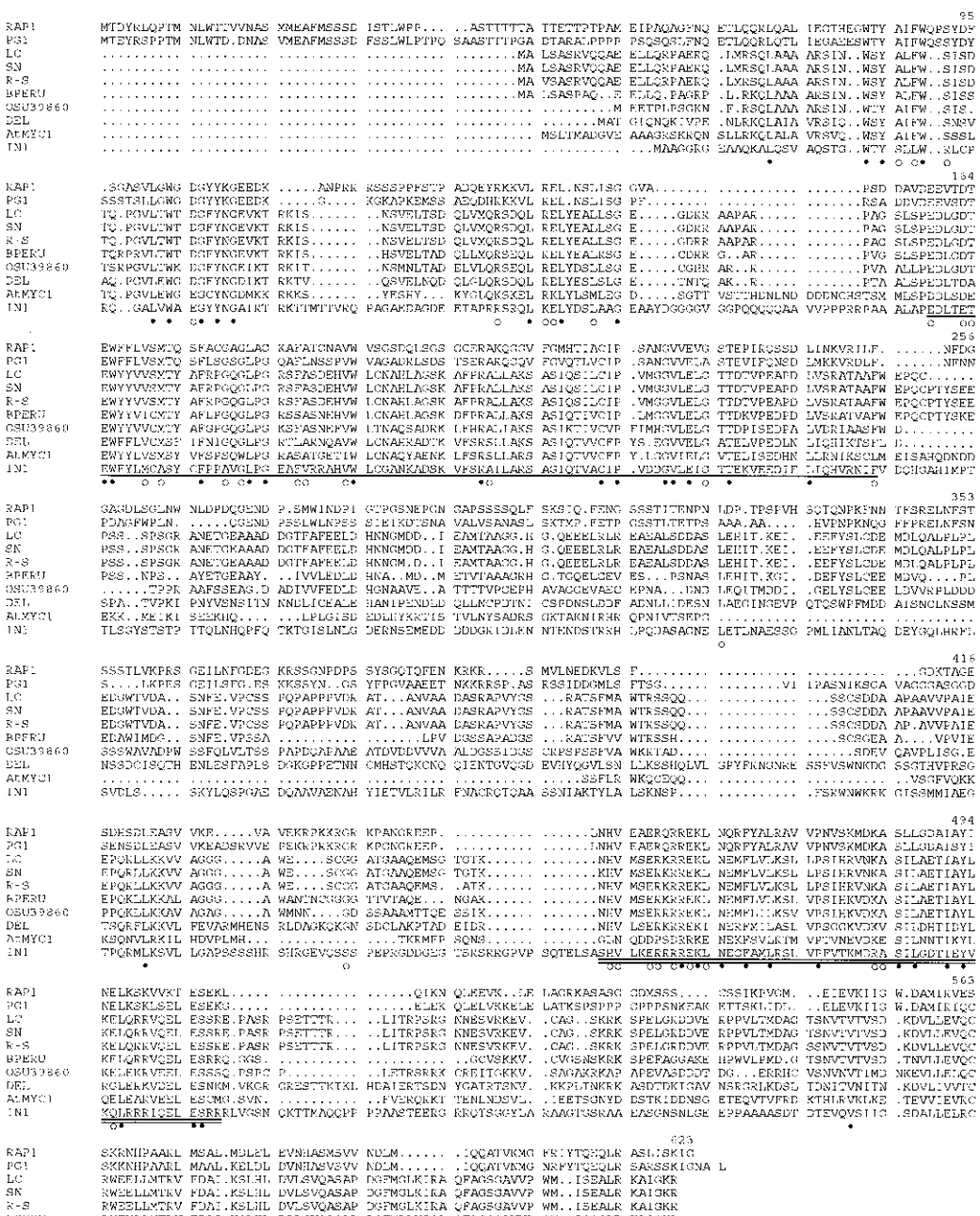

Figure 2. Sequence comparison of RAP-1 with other plant bHLH proteins, that were found to be homologous to RAP-1 using the blast program [1]. Alignment of amino acid sequences of RAP-1 (X99548) from *Arabidopsis*, PG1 (U18348) from bean, Lc (M26227) from maize, Sn (X60706) from maize, R-S (X15806) from maize, Bperu (X57276) from maize, OSU39860 (U39860) from rice, DEL (M84913) from snapdragon, AtMYC1 (D83511) from *Arabidopsis* and IN1 (U57899) from maize was performed using the pileup program with standard settings [5]. The numbers in parenthesis are the database accession numbers. Dots indicate gaps introduced to maximize alignment. The N-terminal homologous region and the bHLH domain are underlined by single and double lines, respectively. Positions with identical amino acids in all sequences are indicated by closed circles and positions with at least seven sequences identical to RAP-1 are indicated with open circles.

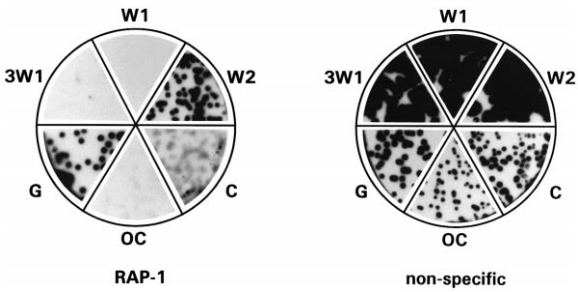

Figure 3. DNA binding specifity of RAP-1. Filters containing *Arabidopsis* RAP-1 protein or a non-specific DNA-binding protein from tobacco (BAD) were incubated with the following probes: W1 (TCGAGGACACGTAGAATGAGTCATCACGTCGA), 3W1, W2 (4× ATGAGTCATCAC), G-box (AGCTTAGACACGTGTCACTCGA), odd-base C-box (AGCTTAATGAGTCATACTCGA) and C-box (AGCTTAATGACGTCATACTCGA). Nitrocellulose filter preparation, binding and washing were done as described previously [17].

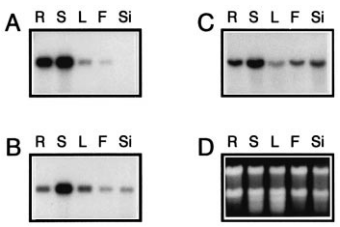

Figure 4. Rap-1 expression in *Arabidopsis* organs. Samples of 600 ng of poly(A)$^+$ RNA (A and B) or 20 μg of total RNA (C) from root (R), stem (S), leaf (L), flower (F) and siliques (Si) of *Arabidopsis* (ecotype Columbia) loaded on a 1.5% formaldehyde agarose gel were blotted and hybridized as described [13] with *Rap-1* (A) or ubiquitin (B and C) cDNA. The blots were washed with 0.1× SSPE, 0.1% SDS at 42 °C. Total RNA blotted in (C) is stained with ethidium bromide (D) to show equal loading. mRNA size was estimated using an RNA size marker (Gibco-BRL, 0.24-9.5 range) run in parallel (not shown).

with myb proteins [6]. The similarities between this region in the *Arabidopsis* protein and the corresponding regions in the maize proteins Lc and Bperu were 64% and 61%, respectively. This region is followed by an acidic domain (Figure. 1), that may serve as a transcriptional activation domain. The nuclear localization signal identified in the basic region of Lc [19] is conserved in the *Arabidopsis* protein (Figure 1). The *Arabidopsis* protein will be further referred to as R-homologous *Arabidopsis* Protein-1 (RAP-1).

The DNA-binding specificity was analysed by filter binding assays (Figure 3). Binding of RAP-1 was observed with an oligonucleotide containing a G-box and with W2. No binding was observed with a monomer or trimer of W1. Oligonucleotides containing a C-box or the W1 odd-base C-box also did not bind RAP-1. A control cDNA clone encoding a non-specific DNA-binding protein from tobacco (BAD; [17]) showed binding to all probes. From these results it can be concluded that RAP-1 does not bind to W1 or parts of W1. RAP-1 very likely binds to the junctions between the W1 odd-base C-box monomers present in the W2 tetramer. The sequence at the junctions in W2 and the G-box oligonucleotide have the sequence CACNTG in common, which is the core sequence of binding sites for animal bHLH proteins [2] and for PG1 of bean [7].

The tissue-specific expression patterns of different members of the maize R gene family are distinct from each other. The expression pattern of *Rap-1* was analysed by northern blot hybridization. Hybridization with poly(A)$^+$ RNA from root, stem, leaf, flower and siliques (Figure 4A) showed that *Rap-1* is expressed mainly in vegetative tissues, with the highest expression in root and stem. No expression was detected in siliques. In those tissues analysed, the expression pattern resembled the expression of PG1 of bean [7]. The hybridizing mRNA has the same length as the cDNA clone, corroborating the notion that the cDNA clone is close to full-length. The blot was reprobed with a cDNA encoding ubiquitin (Figure 4B), to show the presence of intact mRNA in each lane. The lane with mRNA from stem contained a higher signal compared to the other lanes. Since this was also observed on a blot containing equal amounts of total RNA from the same tissues (Figure 4C and D), we conclude that the ubiquitin mRNA is more abundant in stem tissue.

We then performed genomic Southern blot hybridization to determine the number of genes detectable under our experimental conditions. Figure 5 shows that there are 1 to 3 bands in each lane. Multiple bands are observed with restriction enzymes that have recognition sites within the cDNA. A single major band is observed after digestion with *Eco*RV, which has only one recognition site in the *Rap-1* cDNA located very close to the 5′ end. The faint band of about 5 kb in this lane is either the result of weak hybridization with a distantly related (bHLH) gene or with the 5′ end of *Rap-1*. Thus, RAP-1 may be encoded by a single gene.

Genetic map positions of several structural and regulatory loci involved in anthocyanin biosynthesis in *Arabidopsis* have been determined [20]. To determine whether *Rap-1* corresponds to one of these loci, its genetic map position was determined by RFLP mapping in the *Arabidopsis* recombinant inbred line pop-

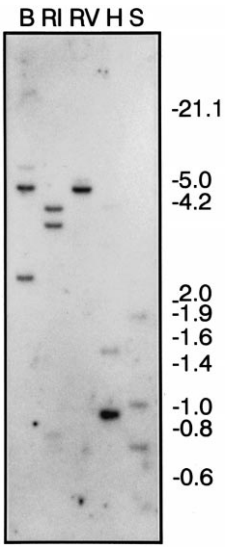

Figure 5. Southern blot analysis of *Rap-1*. One μg of *Arabidopsis* DNA (ecotype Columbia) was digested with *Bam*HI (B), *Eco*RI (E), *Eco*RV (RV), *Hin*dIII (H), or *Sac*I (S), electrophoresed on a 0.8% agarose gel, blotted and hybridized as described [13] with *Rap-1* cDNA. The blot was washed with 0.1× SSPE, 0.1% SDS at 42 °C. Positions and sizes in kb of *Eco*RI- and *Hin*dIII-digested Lambda DNA fragments are indicated.

ulation [9]. *Rap-1* is located on the upper region of chromosome 4, at a different position than *tt8*, the only anthocyanin locus mapped to chromosome 4 so far [20]. Recently, a gene encoding a MYC-related protein (AtMYC1) with unknown function was also mapped to the upper region of chromosome 4, near the position of *Rap-1* [22]. However, *Rap-1* and *Atmyc1* have limited sequence homology (Figure 2) and different expression patterns [22], indicating that they are different genes. Thus, *Rap-1* does not correspond to one of the previously described loci.

R proteins are responsible for determining the temporal and spatial pattern of anthocyanin pigmentation. Expression of an R protein in tissues, where it is normally not present, may cause these tissues to become pigmented [11]. Particle bombardment was used to introduce *Rap-1*, fused to the CaMV 35S promoter, into pea cotyledons, a convenient seed system for particle bombardment. As a control *35S-Lc* [11] was used. Lc is a member of the maize R protein family and was shown previously to induce pigmentation in tissues that are normally not pigmented by the *Lc* gene [11]. Pink spots were observed in pea cotyledons bombarded with *35S-Lc* (results not shown). However, no spots were found after bombardment with the *35S-Rap-1* chimeric gene or with this gene in combination with *35S-C1*, encoding a maize myb homologue that interacts with Lc [6]. This indicates that RAP-1 is not functionally equivalent to Lc.

Our results do not rule out the possibility that RAP-1 is a regulator of anthocyanin biosynthesis. The presence of a region homologous to the N-terminal conserved domain of R proteins indicates that RAP-1 needs to interact with myb proteins for its activity. RAP-1 should have interacted either with C1 from maize or with a myb homologue from pea to induce anthocyanin biosynthesis in our transient expression assays. These proteins may have a different structure compared to the putative *Arabidopsis* myb homologue which interacts with RAP-1 *in vivo*, preventing correct interaction.

RAP-1 was isolated by southwestern screening and binds G-box motifs. So far, binding of R proteins to DNA *in vitro* has not been reported. Since neither the bHLH region in the R protein [6] nor the bHLH consensus DNA binding site [21] is essential for transactivation of the analysed anthocyanin structural promoters, it had been suggested that R proteins interact with DNA via their myb-homologous partners. If this is true, it is impossible to recover R genes by southwestern screening of *E. coli* expression libraries. The PG1 protein from bean is highly homologous to RAP-1 and also binds to G-box motifs [7]. One can argue that RAP-1 and PG1 are functionally different from the R proteins based on (1) their DNA binding properties, (2) the somewhat different sequence of RAP-1 and PG1 compared to the R proteins (Figure 2), and (3) the inability of RAP-1 to induce pigmentation in pea cotyledons. Ectopic expression of *Rap-1* in *Arabidopsis* may answer the question whether RAP-1 is a regulator of anthocyanin biosynthesis or involved in another biological process.

Acknowledgements

We thank Andy Pereira and his group for mapping *Rap-1*, Peter Hock and the Biology Photographic Service, Leiden University, for preparation of the figures. S.d.p and K.P. were supported by the Netherlands Organization for Applied Scientific Research (TNO) and J.M. by a Fulbright fellowship.

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 215: 403–10 (1990).
2. Blackwell TK, Weintraub H: Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science 250: 1104–1110 (1990).
3. Ellenberger T, Fass D, Arnaud M, Harrison SC: Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. Genes Devel 8: 970–980 (1994).
4. Ferré-D'Amaré AR, Prendergast GC, Ziff EB, Burley SK: Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature 363: 38–45 (1993).
5. Genetics Computer Group: Program Manual for the Wisconsin Package, version 8 (1994).
6. Goff SA, Cone KC, Chandler VL: Functional analysis of the transcriptional activator encoded by the maize B gene: evidence for a direct functional interaction between two classes of regulatory proteins. Genes Devel 6: 864–875 (1992).
7. Kawagoe Y, Murai N: A novel basic region/helix-loop-helix protein binds to a G-box motif CACGTG of the bean seed storage protein β-*phaseolin* gene. Plant Sci 116: 47–57 (1996).
8. Koes RE, Quattrocchio F, Mol JNM: The flavonoid biosynthetic pathway in plant: function and evolution. BioEssays 16: 123–132 (1994).
9. Lister C, Dean C: Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana*. Plant J 4: 745–750 (1993).
10. Lloyd AM, Walbot V, Davis RW: *Arabidopsis* and *Nicotiana* anthocyanin production activated by maize regulators *R* and *C1*. Science 238: 1773–1775 (1992).
11. Ludwig SR, Bowen B, Beach L, Wessler SR: A regulatory gene as a novel visible marker for maize transformation. Science 247: 449–450 (1990).
12. Ludwig SR, Wessler SR: Maize R gene family: tissue-specific helix-loop-helix proteins. Cell 62: 849–851 (1990).
13. Memelink J, Swords KMM, Staehelin LA, Hoge JHC: Southern, Northern and Western blot analysis. In: Gelvin SB, Schilperoort RA (eds) Plant Molecular Biology Manual, F1: 1–23. Kluwer Academic Publishers, Dordrecht (1994).
14. Meshi T, Iwabuchi M: Plant transcription factors. Plant Cell Physiol 36: 1405–1420 (1995).
15. de Pater S, Katagiri F, Kijne J, Chua N-H: bZIP proteins bind to a palindromic sequence without an ACGT core located in a seed-specific element of the pea lectin promoter. Plant J 6: 133–140 (1994).
16. de Pater S, Pham K, Chua N-H, Memelink J, Kijne J: A 22-bp fragment of the pea lectin promoter containing essential TGAC-like motifs confers seed-specific gene expression. Plant Cell 5: 877–886 (1993).
17. de Pater S, Pham K, Memelink J, Kijne J: Binding specificity and tissue-specific expression pattern of the *Arabidopsis* bZIP transcription factor TGA2. Mol Gen Genet 250: 237–239 (1996).
18. Purugganan MD, Wessler SR: Molecular evolution of the plant *R* regulatory gene family. Genetics 138: 849–854 (1994).
19. Shieh MW, Wessler SR, Raikhel NV: Nuclear targeting of the maize R protein requires two nuclear localization sequences. Plant Physiol 101: 353–361 (1993).
20. Shirley BW, Kubasek WL, Storz G, Bruggemann E, Koornneef M, Ausubel FM, Goodman HM: Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. Plant J 8: 659–671 (1995).
21. Tuerck JA, Fromm ME: Elements of the maize *A1* promoter required for transactivation by the anthocyanin *B/C1* or phlobaphene *P* regulatory genes. Plant Cell 6: 1655–1663 (1994).
22. Urao T, Yamaguchi-Shinozaki K, Mitsukawa N, Shibata D, Shinozaki K: Molecular cloning and characterization of a gene that encodes a MYC-related protein in *Arabidopsis*. Plant Mol Biol 32: 571–576 (1996).