The Plant Journal (1993) 3(2), 335–346

PAPER PRESENTED AT THE ANNUAL SEB MEETING 1992

Molecular homology among members of the *R* gene family in maize

Gabriella Consonni, Filippo Geuna, Giuseppe Gavazzi and Chiara Tonelli*
Dipartimento di Genetica e di Biologia dei Microrganismi, Via Celoria 26, I-20133 Milano, Italy

Summary

The *R* gene family determines the timing, distribution and amount of anthocyanin pigmentation in maize. This family comprises a set of regulatory genes, consisting of a cluster of several elements at the *R* locus, on chromosome 10, the *Lc* and *Sn* gene lying about two units *R* distal and *B* on chromosome 2. Each gene determines a tissue-specific pigmentation of different parts of the seed and plant. The proposed duplicated function of *R*, *Sn*, *Lc* and *B* loci is reflected in cDNA sequence similarity. In this paper an extensive analysis of the predicted proteins of the *R*, *Sn*, *Lc* and *B* genes together with a search for putative sites of post-translational modification is reported. A comparison with the PROSITE database discloses several *N*-glycosylation and phosphorylation sites, as well as the basic Helix-Loop-Helix (HLH) domain of transcriptional activators. *Sn*, *Lc*, and *R-S* show a high conservation of these sites, while *B* is more divergent. Analysis of the 5′ leader of mRNA sequences discloses the presence of five ATG triplets with two upstream open reading frames (uORFs) of 38 and 15 amino acids and a loop structure indicating a possible mechanism of control at the translational level. It is conceivable that possible mechanisms acting at the translational and post-translational level could modulate the expression and the activation of these transcription factors. Northern analysis of various tissues of different *R* alleles highlights a strict correlation between pigment accumulation in different tissues and the expression of the regulatory and structural genes suggesting that the pattern of pigmentation relies on a mechanism of differential expression of the members of the *R* family. Analysis of the *Sn* promoter discloses the presence of several sequences resembling binding sites of known transcription factors (as GAGA and GT) that might be responsible for the spatial and light-induced expression of this gene. Two regions include a short sequence homologous to the consensus binding site of the B-HLH domain suggesting a self-regulatory control of the *Sn* gene.

Received 24 April 1992; revised 11 August 1992; accepted 10 September 1992.
*For correspondence (fax +39 2 2664551).

Introduction

Anthocyanins represent a class of secondary metabolites widespread among plants. Pigmentation is controlled by numerous genes that can influence the type, amount and distribution of pigments synthesized. The structural genes encoding the enzymes involved in the biosynthesis are expressed in a co-ordinate way as a result of the activity of several genes that regulate the spatial and temporal distribution of the pigments among the various tissues. Anthocyanin biosynthesis is also modulated by environmental stimuli such as light, temperature, phytopathogens, wounding and nutrient supply (Nagy *et al.*, 1988). Therefore this process presents unique features that make it an appealing model system for the analysis of gene regulation as well as for studies in developmental biology.

In maize the genetic control of anthocyanin accumulation is well documented and more than 20 structural and regulatory genes have been identified (Coe *et al.*, 1988). The *R*, *Sn*, *Lc*, *B* and *C1* loci affect the activities of at least three structural genes in the pathway, *C2*, *A1*, *Bz1* coding respectively for chalcone synthase, dihydroquercetin reductase and UDP glucose: flavanol 3-*O*-glucosyltransferase (Chandler *et al.*, 1989; Cone *et al.*, 1986; Dooner and Nelson, 1977; Ludwig *et al.*, 1989; Paz-Ares *et al.*, 1987; Schwarz-Sommer *et al.*, 1987; Tonelli *et al.*, 1991; Wienand *et al.*, 1986).

Accumulation of pigment in competent tissues requires the presence of either *C1* in the seed or *Pl* in the plant tissue (Coe *et al.*, 1988). Sequence analysis of *C1* reveals that it encodes a protein with a terminal basic domain showing homology to the DNA binding region of the *myb* oncogene family and an acidic domain characteristic of known transcriptional activators (Paz-Ares *et al.*, 1987).

A second class of regulatory loci is that of the *R* family, characterized by an extensive allelic variability in terms of both the developmental timing and tissue specificity of expression (Styles and Ceska, 1977).

Genetic and molecular studies have demonstrated that some *R* alleles are complex and that tissue-specific control is mediated through duplicated elements. For example, in *R-r* one component (S) is responsible for seed pigmentation

and another component (P) is responsible for color in anthers, coleoptile and other plant tissues (Stadler, 1948). Two types of derivatives are recovered from the parental *R-r* allele: those without plant color termed *R-g*, and those without seed color termed *r-r*. Molecular studies have confirmed the dual nature of *R-r* predicted by genetic studies (Dellaporta *et al.*, 1988; Robbins *et al.*, 1991).

In certain accessions, additional members of the *R* gene family are found about two map units *R* distal (on chromosome 10). These genes, called *Sn* and *Lc*, show a tissue-specific pattern of expression different from those of the S and P component of the *R* locus (Dooner and Kermicle, 1976; Gavazzi *et al.*, 1986, 1990).

The *B* gene (on chromosome 2) is required for anthocyanin pigmentation in plant tissues independently from *R*. However a few *B* alleles are known that control anthocyanin formation in the seed (Styles and Ceska, 1977). All these genes appear as duplicates that perform similar regulatory roles in different tissues. Their duplicate function is reflected in their DNA sequence similarity. *R* gene sequences have been used as hybridization probes to identify *B*, *Lc* and *Sn* clones (Chandler *et al.*, 1989; Ludwig *et al.*, 1989; Tonelli *et al.*, 1991). The cDNA sequences of *Lc*, *Sn*, *R-g* (S) and *B* have been obtained (Consonni *et al.*, 1992; Ludwig *et al.*, 1989; Perrot and Cone, 1989; Radicella *et al.*, 1991). Their putative proteins deduced from their cDNA sequence, show a very high homology and contain a basic domain with the Helix-Loop-Helix (HLH) motif similar to that found in several DNA binding proteins in mammals and *Drosophila* and an acidic domain characteristic of known transcriptional activators (Murre *et al.*, 1989). The predicted protein of the *B* gene, on the other hand, shows a lower homology to that of *R* while maintaining the HLH motif (Radicella *et al.*, 1991).

The observation that expression of the structural genes *C2*, *A1* and *Bz1* is mediated by the presence of an active member of the *R* family suggests a regulatory role for these genes and the involvement of a *trans*-acting *R* gene product in the co-ordinate activation of the genes of the anthocyanin pathway.

Accordingly, *R* and related *Sn*, *Lc* and *B* genes might encode functionally related proteins that act as transcriptional activators of the genes encoding the biosynthetic enzymes. The distinct patterns of tissue pigmentation controlled by the *R*, *Sn*, *Lc* and *B* genes may reflect differences in their promoter rather than in their product. Recent experiments of transient transformation (with high velocity microprojectiles) of *r-g* tissues with a vector containing the *Lc*, *B* or *Sn* cDNA fused to CaMV 35S promoter support this interpretation, confirming the functional equivalence of the different members of the *R* family (Dellaporta *et al.*, unpublished results; Goff *et al.*, 1990; Ludwig *et al.*, 1990).

In the present study, we report an extensive analysis of the predicted proteins of the *R*, *Sn*, *Lc* and *B* genes together with a search for putative sites for post-translational modification. Analysis of the 5′ leader of mRNA sequences discloses the presence of five AUG triplets with two open reading frames (ORFs) of 38 and 15 amino acids, suggesting a possible mechanism of control at the translational level. It is conceivable that mechanisms acting at post-transcriptional level could modulate the expression and activity of these transcription factors.

Northern analysis of various tissues of different *R* alleles highlights a strict correlation between pigment accumulation and expression of the regulatory and structural genes suggesting that the diverse patterns of pigmentation are due to the differential expression of the *R* alleles. *Sn* promoter sequence analysis, compared with the transcription factor database discloses the presence of several sequences resembling binding sites of known transcription factors that might be responsible for the spatial and temporal expression of this gene.

Results and discussion

Tissue-specific expression of the genes of the R *family*

As previously mentioned, *R*, *B*, *Lc and Sn* share a regulatory role in anthocyanin biosynthesis suggesting that their action is implicated in the co-ordinate expression of the enzymes of this pathway.

Sn transcripts have tissue- and cell-specific localization, as shown by Northern analysis and by *in situ* hybridization (Tonelli *et al.*, 1991).

To further investigate the type of transcript and the tissue specificity of expression, total RNA was extracted from mesocotyls, pericarps, aleurones and roots of plants differing in their *R* constitution, and the expression of regulatory and structural genes analyzed by Northern blot. The following genotypes were analyzed: *rSn:bol3* which allows pigmentation in mesocotyl and pericarp but not in the aleurone; *R-sc* and *R-g* (S) leading to pigmentation in the aleurone but not in seedling tissues and pericarp; *r-r* (P) leading to red seedling tissues like roots and coleoptiles but not mesocotyls and seed tissues; *r-g* allele as control that does not allow pigmentation in any seed or seedling tissues. All these genotypes are homozygous for the structural genes required for pigmentation.

The filter was hybridized with the *Sn* cDNA probe that detects *Sn* and *R* transcripts and subsequently with probes specific for the structural genes. The RNA load in each lane was controlled using a maize tubulin probe.

The results (Figure 1) show that *Sn* and *R* transcripts are specifically expressed only in the pigmented tissues. Their presence correlates with the induction of the mRNA of the structural genes *A1* and *C2* (the last one is not shown) and with the accumulation of the anthocyanin pigments as

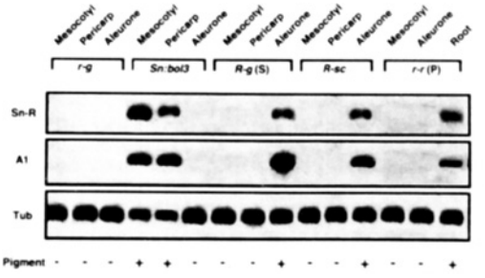


Figure 1. Northern blot analysis showing a differential expression of *Sn* and *R* alleles.
Total RNA was extracted from mesocotyl, pericarps, aleurone and roots of *r-g*, *Sn:bol3*, *R-g* (S), *R-sc* and *r-r* (P) genotypes. The amount of RNA in each lane was about 40 μg except for *Sn:bol3* mesocotyl and pericarp that was about 20 μg. The blot is hybridized with the probe indicated on the left: Sn-R, 1.4 kb *PstI–EcoRI* fragment of *Sn* cDNA (Tonelli *et al.*, 1991); A1, 700 bp *BamHI* fragment of *A1* gene (Schwarz-Sommer *et al.*, 1987); Tub, 1 kb cDNA clone of maize tubulin (Mereghetti *et al.*, 1990) as control of RNA loads.

expected on the assumption of their regulatory role. Colorless tissues of the *r-g* genotype on the other hand do not show any detectable transcripts of either regulatory or structural genes.

Taken together these results confirm that, although all tissues analyzed are competent to accumulate anthocyanin, pigment production is strictly dependent on the specific *R* allele expression.

Sequence comparison of the Sn, Lc, S *and* B-Peru *putative proteins*

In the light of recent reports showing that post-translational modification is an important means for regulating transcription factor activity, we have done an extensive analysis to search for putative sites of post-translational modification of the putative proteins deduced from the cDNA sequences of *Sn*, *Lc*, *S* (referred to as the S component of the *R-g* allele) and *B-Peru* (Consonni *et al.*, 1992; Ludwig *et al.*, 1989; Perrot and Cone, 1989; Radicella *et al.*, 1991). We have compared these four proteins in terms of primary structure and of putative sites of post-translational modification with the belief that a high degree of conservation of these sites might be indicative of functional significance.

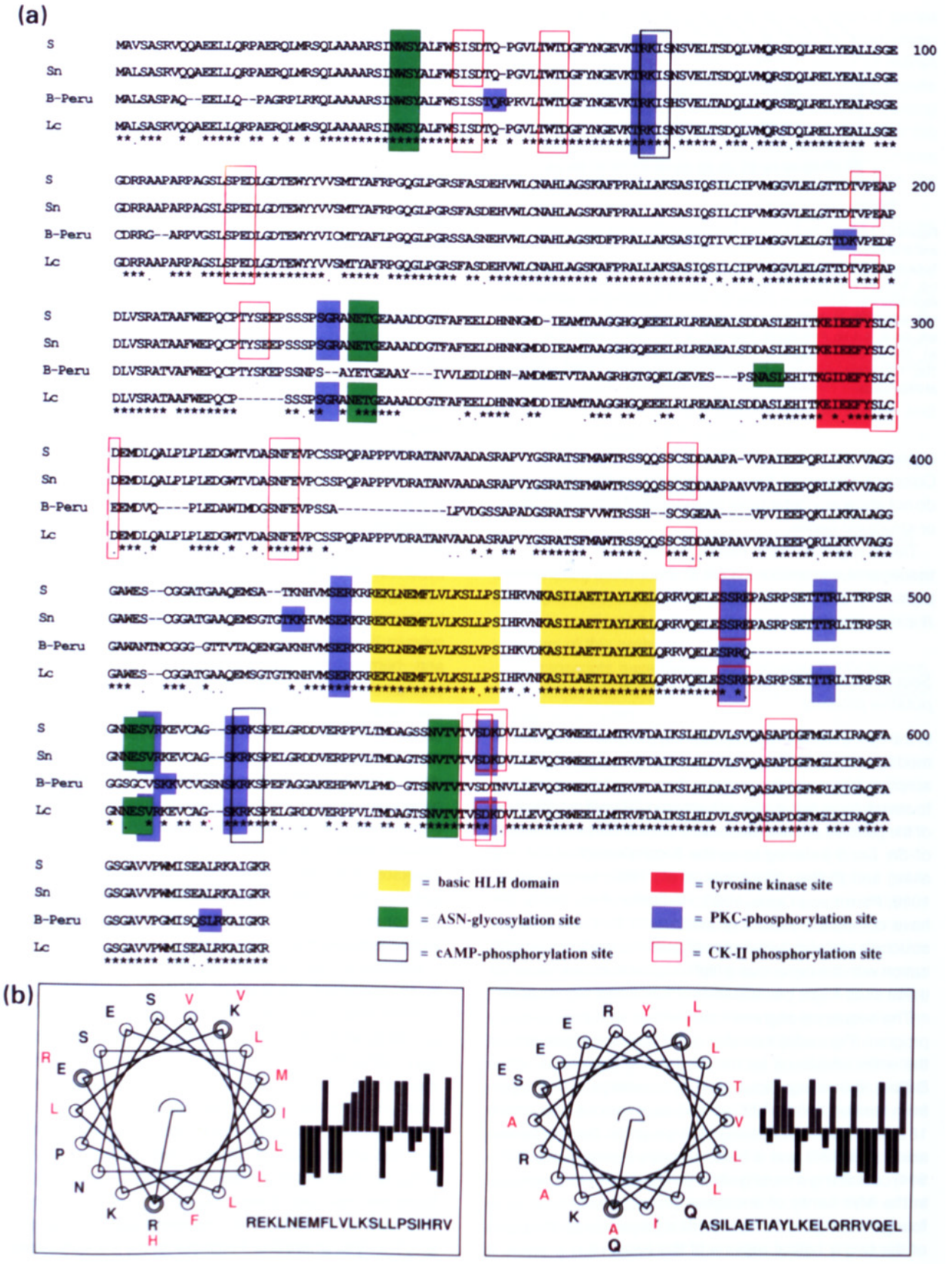
The sequence alignment performed with the CLUSTAL V program (Figure 2a) indicates good homology over almost the entire sequence for the Sn, Lc and S proteins, while B-Peru shows significant lack of homology in the region from residue 238 to 422, as well as at residue 483 where 18 amino acids are missing. A large acidic domain (amino acids 188–324) and a basic domain (amino acids 421–514) containing a motif (yellow boxes) similar to that found in the *Myc* family of oncogene proteins, are found in all four genes. Figure 2b represents a projection on the plane of the two α-helical regions of the basic HLH domain. It should be noticed that the hydrophobic residues tend to occupy one half of the helix, while the hydrophilic ones occupy the other half. This feature, shared by both helices, fits the amino acid distribution described in other HLH-based DNA-binding proteins (Murre *et al.*, 1989) being the basis for the correct reciprocal positioning of the two α-helices and for the steric conformation of the DNA-binding motif.

The analysis of the sequences with the PROSITE database discloses several covalent modification sites (Figure 2a). Potential *N*-glycosylation sites are specific to the consensus sequence N-X-S/T-X where X stands for any amino acid (Marshall, 1972). Presence of the consensus tripeptide in itself does not imply that an asparagine residue is glycosylated, since folding of the protein plays an important role in the regulation of *N*-glycosylation (Pless and Lennarz, 1977). Of the five boxes indicated as presumed glycosylation sites (green boxes), the first and the fifth are conserved in the four sequences, while the second and the fourth boxes are conserved in *S*, *Sn* and *Lc* and are missing in *B-Peru* and the third one is found only in *B-Peru*.

Although glycosylation is often regarded as being restricted to proteins confined to cell surface or within the lumen of intracellular organelles, transcription factors bearing *N*-acetylglucosamine (GlcNAc) residues have been recently detected in the nucleus. Many RNA polymerase II transcription factors isolated from *Drosophila* and human cells contain terminal GlcNAc residues (Jackson and Tjian, 1988). This observation suggests that this modification might have an important, evolutionary conserved role in the mechanism of transcriptional activation. Differential glycosylation might reflect functional differences between transcription factors implying that glycosylation may be employed by the cell to regulate transcription factor activity for example in response to diverse physiological stimuli or developmental stages (Jackson and Tjian, 1988). Glycosylation of plant transcription factors has not yet been reported. In this context the conservation of two putative sites of glycosylation in all four proteins may be relevant.

In animal systems phosphorylation of transcription factors is an important mechanism of gene regulation. DNA binding properties of transcription factors (Yamamoto *et al.*, 1988), their ability to activate transcription (Tanaka and Herr, 1990), the mobility of DNA-bound protein complexes (Sorger *et al.*, 1987) and of the factor itself (Jackson *et al.*, 1990) are affected by phosphorylation.

Four types of phosphorylation sites were detected in the proteins under analysis. *In vivo*, protein kinase C (PKC) exhibits a preference for the phosphorylation of serine or threonine residues close to a C-terminal basic residue (Woodget *et al.*, 1986). The consensus pattern for this site is S/T-X-R/K, where S or T represents the phosphorylation


(a)
S
Sn
B-Peru
Lc
100
200
300
400
500
600
= basic HLH domain
= tyrosine kinase site
= ASN-glycosylation site
= PKC-phosphorylation site
= cAMP-phosphorylation site
= CK-II phosphorylation site
(b)
REKLNEMFLVLKSLLPSIHRV
ASILAETIAYLKELQRRVQEL

site. Of the twelve PKC phosphorylation sites (blue boxes) detected in the alignment only five are shared by all four proteins and three by Sn, Lc and S. The second phosphorylation signal (R/K-X-X-S/T) detected, is typical of the cAMP- or cGMP-dependent kinases that share a preference for the phosphorylation of serine or threonine residues close to at least two consecutive N-terminal basic residues (Fremisco *et al.*, 1980; Glass and Smith, 1983). The two recognition sites found (empty black box) are highly conserved and exhibit partially overlapping regions of recognition for protein kinase C (blue boxes). In addition, a unique tyrosine kinase site (Hunter, 1982) is found that is conserved in S, Sn and Lc, but partially divergent in *B-Peru* (red box). Casein kinase II (CK-2) is a serine-threonine protein kinase acting independently of cyclic nucleotides and of calcium. Of the twelve CK-2 phosphorylation sites (red open boxes), detected according to the rules indicated in the review of Pinna (1990), six are shared by all the four proteins. The conservation of a glycosylation and a CK-2 phosphorylation site, almost coincident, around position 546 might be of some relevance. The localization of phosphorylation sites is highly conserved in Sn, Lc and S protein, but divergent both in the site number and position in B-Peru.

The relative concentration of sites around the region critical for the structure of the inferred HLH domain in these proteins is signficant, since this region is involved in the correct positioning of the transcriptional activator relative to both the DNA sequence and other possible factors.

Interestingly Myc oncoproteins are phosphorylated by CK-2 in two distinct sites, the highly acidic region and the region proximal to the HLH domain (Luscher *et al.*, 1989). Furthermore, studies in mammals have revealed that phosphorylation of the cJun transcription factor is the key step to trigger the activation of the genes under its control (Binetruy *et al.*, 1991). It was also shown that phosphorylation of CREB (a eukaryotic cAMP-responsive transcription factor) by protein kinase A increases the gene activating potential of CREB without altering its affinity for DNA. Thus it is conceivable that this phosphorylation alters the contacts that CREB may make with other DNA-bound factors (Gonzalez and Montminy, 1989). A similar situation is also found in the *D-raf* system of *Drosophila* where the product of the *D-raf* gene, which acts downstream of the receptor tyrosine kinase gene *torso*, is responsible for the activation of a transcription factor that in turn activates the expression of *tailless* and *huckebein* genes at the poles of the embryo (Siegfried *et al.*, 1990).

Also, more recently in plants phosphorylation of the nuclear factor has been reported. The DNA-binding activity of AT-1, a factor that binds to specific AT-rich elements (the AT-1 box) within the promoter of photoregulated genes is modulated by phosphorylation (Datta and Cashmore, 1989). Klimczak *et al.* (1992) identified a protein kinase activity (CK-2-like) that is able to phosphorylate the G-box binding factor (GBF1) in *Arabidopsis.* Phosphorylation of GBF1 resulted in stimulation of G-box binding activity and formation of a slower migrating protein-DNA complex. The 3AF5 and 3AF3 proteins are two other examples of light-regulated proteins required for the transcription of *rbcS3A.* Sarokin and Chua (1992) showed that phosphorylated forms of both proteins might be required for the light-responsive transcription of *rbcS-3A.*

5' Leader sequence analysis

Sn has a transcript of 2572 bp, an estimate based on Northern analysis, cDNA sequencing and primer extension (Consonni *et al.*, 1992; Tonelli *et al.*, 1991). On the transcript five ATG triplets precede the actual start codon for the *Sn* gene ORF (Figure 3), two of them representing the start site for two ORFs of 38 and 15 amino acids, respectively (positions +60 and +297). The triplet at position +60 is the first from the 5' end of the transcription product. According to the ribosome scanning hypothesis (Kozak, 1984) it represents the first start codon whereby translation should start. The nucleotide sequence immediately preceding the start codon fits partially the consensus sequence suggested by Kozak (1984) with only the cytosine at position −1 and, especially, with a guanine at position +4 (immediately following the triplet) enhancing the signal required for a presumptive ribosome recognition. Further evidence supporting the translational potential of this ORF is the presence of two stop codons (TAA) at its end. The second and third start codons at positions +126 and + 162, respectively, are in frame within this ORF. Sequences

Figure 2. Sequence alignment of Sn, Lc, S and B-Peru proteins.
(a) The protein sequences of the four regulatory genes described in the paper aligned with the ClustalV program (see the text): every asterisk marks the full conservation of an amino acid residue over the four protein sequences, whereas a point indicates the substitution of at least one of them to another one with similar physico-chemical properties. Lack of any symbol means either that a substitution occurred with an unrelated amino acid residue in one of the sequences at that position or the complete absence of the residue (deletion). Putative modification sites, as inferred from comparison with the PROSITE database, are also shown inside colored or open boxes.
(b) The planar projections of the helices in the HLH domain conserved in the four proteins. The amino acids with hydrophobic residues are shown with red characters, while those with hydrophilic or neutral residues are shown with black characters. The histogram beside the projection of each helix represents graphically the degree of hydropathy (upward gray bars) or hydrophilicity (downward black bars) of the residues according to the estimation method of Kyte and Doolittle (1982).

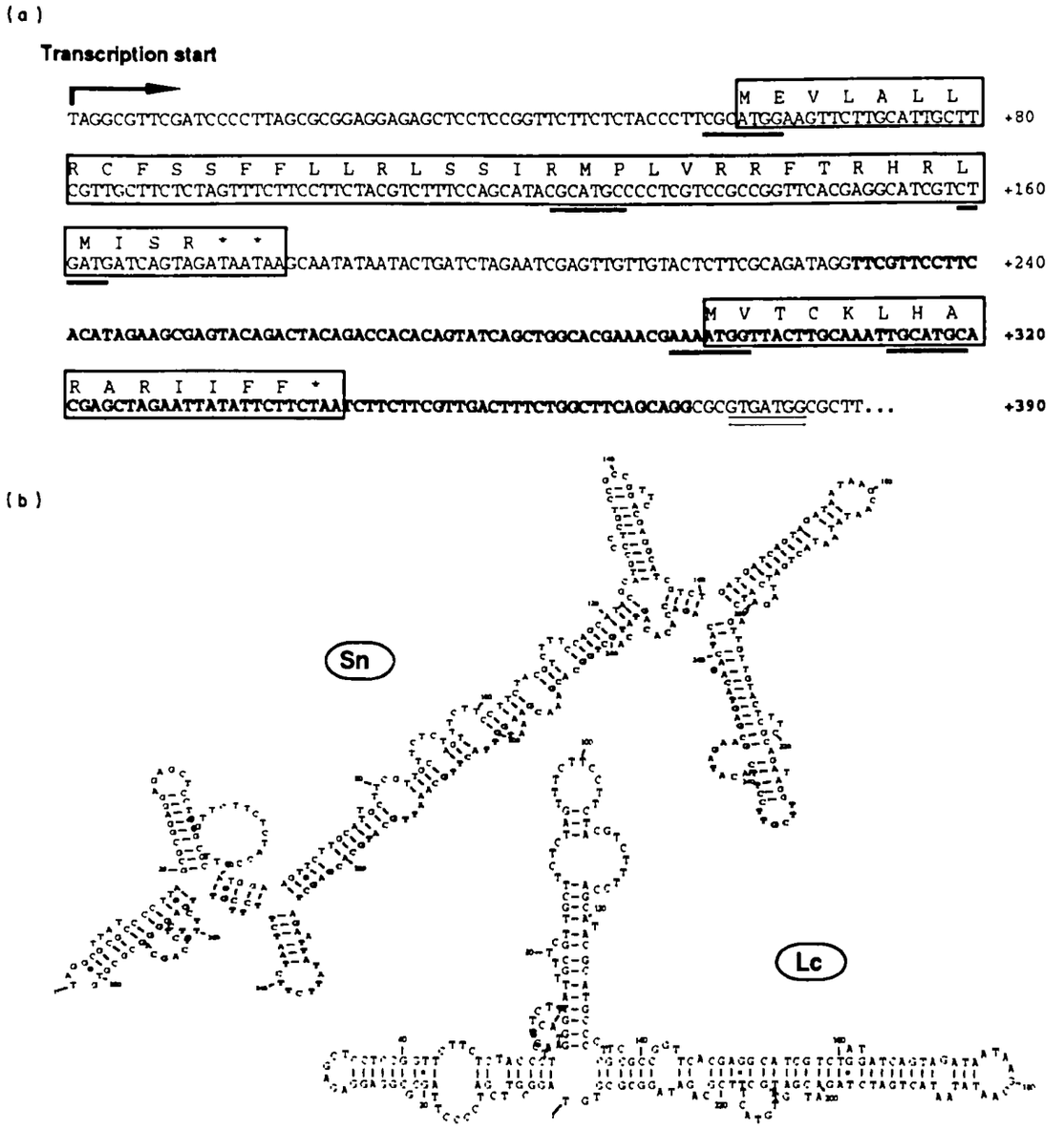


Figure 3. Sequence and putative secondary structure of the *Sn* and *Lc* genes 5′ leader regions.
(a) The *Sn* gene 5′ leader sequence: the insertion region not present in *Lc* is printed in bold characters; the regions comprehending the five ATG triplets upstream from the actual start codon (double-underlined) are marked with a thick single line. The two small ORFs with the highest likelihood of being translated before the *Sn* coding region are enclosed in boxes.
(b) A graphic representation of the looping pattern for the *Sn* gene 5′ leader sequence and, by comparison, of the same region in *Lc* as calculated by the ZUKER program (see the text).

flanking these two ATGs are not in agreement with the consensus of Kozak (1984) and probably are not active. The fourth triplet at +297 from the transcription start marks the start for a peptide of 15 amino acids and is preceded by a purine at −3 and followed by a guanine.

The fifth ATG at position +316 lies inside the previous ORF but is not in frame and would represent the start for a peptide of 13 residues ending at position +355 with the stop codon TGA, but it does not fit the consensus. The start signal relative to the *Sn* gene product is found at position +382: it is preceded by the sequence GTG (with a purine at −3) and is followed by a guanine at +4. The deduced protein sequence for this ORF (616 residues long) discloses a high homology with that of the other regulatory genes described in the *R* family. Thus, both the good agreement of the sequence surrounding this ATG with the consensus suggested and the sequence of the polypeptide starting at this position strengthen the

suggested role of this codon as the start signal in the *Sn* gene translation. The role of the two small ORFs with the highest likelihood of being translated before the *Sn* coding region could be similar to that reported in previous examples (Abastado *et al.*, 1991; Kozak, 1986) where small ORFs lying upstream from the true coding region of the transcript may modulate the translation downstream.

The comparison of the leader sequence of *Sn* cDNA with that of *Lc* discloses a complete homology except for a deletion in *Lc* of 146 nucleotides, from position +230 to +375. The leader sequences of *Sn* and *Lc* cDNA have also been subjected to a computer analysis based on the ZUKFOLD program (Zuker and Stiegler, 1981) aimed at defining secondary structure motives responsible for some kind of post-transcriptional regulation of the transcripts. The comparison of the looping patterns, reported in Figure 3b, indicates that the *Lc* 146 bp deletion is responsible for a quite different theoretical conformation of the molecule and that the 15 amino acids-long putative ORF, found in *Sn*, is missing in *Lc*. The 5′ leader sequence of the *S* gene cDNA starting inside the insertion of *Sn*, shows complete homology with *Sn*, while the same region in *B-Peru* cDNA presents a high degree of divergence (data not shown), with the exception of the nine nucleotides preceding the translation start signal (AGGCGCGTG**ATG**).

Short upstream open reading frames (uORFs) in animals and fungi mRNAs can inhibit translation of downstream coding sequences. This inhibitory effect has been explained by assuming that 40S initiation complexes bind at or near the 5′ end of the mRNA and must traverse the entire leader to reach the AUG start codon (the scanning model). If a uORF is present, initiation occurs preferentially at this site, thus reducing the initiation downstream because reinitiation at internal AUG codons appears to be inefficient (Kozak, 1986, 1991).

The presence of uORFs is known to regulate the expression of *GCN4*, a transcriptional activator of amino acid biosynthetic genes in yeast. Regulation of *GCN4* expression by amino acid availability occurs at the level of translation initiation. Four uORFs in the leader of mRNA restrict the flow of scanning ribosomes to the *GCN4* AUG start codon when amino acids are available while, under starvation, this translational barrier is overcome (Abastado *et al.*, 1991).

This control might also be active in plants. AUGs upstream of the start codon have been found in the maize regulatory gene *opaque 2* although its possible role in regulating translation has not been established (Schmidt *et al.*, 1990). The presence of uORFs and a loop structure in the 5′ leader of the *Sn* transcript might indicate the existence of a translational control mechanism in the *Sn* regulation. Furthermore RACE (rapid amplification of cDNA ends) analysis performed on *Sn* mesocotyls has disclosed two *Sn* messengers differing for the presence of the 146 bp insert in the 5′ leader (unpublished results). This will cause the loss of the second uORF and a change in the secondary structure of the leader making *Sn* and *Lc* similar. It is conceivable that a sort of poise between these two forms may be responsible for a differential modulation of the rate of translation.

Sn *gene promoter*

The first 500 nucleotides upstream from the transcription start site of the *Sn* gene have been obtained by sequencing a 3.8 kb *Hin*dIII *Sn* genomic clone (Tonelli *et al.*, 1991). This promoter sequence (Figure 4) has been compared with the transcription factor database (TFD) compiled by Ghosh (1990). A putative TATA box is located upstream of the transcription start site at position −31 (red box in Figure 4). Two other putative TATA boxes are located further upstream at positions −401 and −424. Evidence, however, for a function of these two putative promoters is lacking. A similar situation has been described in the promoter of the *A1* gene in maize in which a second CAAT box and TATA box-like signals have been found in the region between positions −310 and −390 (Schwarz-Sommer *et al.*, 1987). A true CAAT box is not present in the expected position although a similar sequence (CAAC) is found at position −55 marked with an open box. In the *Sn* promoter the sequence AAGGTT at −137 (marked with a gray box) is similar to an SV40 promoter enhancer (Weiher *et al.*, 1983); this sequence is also present in the *A1* promoter as well as in other plant genes (Schwarz-Sommer *et al.*, 1987). Core sequences of two other SV40 enhancers (Weiher *et al.*, 1983) have also been found at positions −118 (TAGGC) and −440 (GGGGC). Two boxes (CATATG and CACTTG, respectively, shown by blue boxes) at positions −9 and −179 correspond to recognition sites for the DNA binding basic HLH domain typical of some transcriptional activators (Blackwell and Weintraub, 1990). The evidence of the occurrence of this domain in the *Sn* product may be a clue to infer a possible mechanism of self-regulation superimposed upon a light-dependent transcription control system. A striking feature within this portion of the promoter is the presence of a sequence resembling the GAGA transcription factor binding site (shown within a green box in Figure 4) described in the promoters of the transcription activators *Ultrabithorax* and *engrailed* in *Drosophila* (Biggin and Tjian, 1989). A role in spatial and temporal determination of development has been attributed to the genes controlled by the GAGA transcription factor in *Drosophila*; this is in agreement with the tissue-specific and developmental-dependent behavior of the *Sn* gene. Sequences 6–9 bp long, resembling the DNA region bound by GT and related transcription factors (Green *et al.*, 1988; Lawton *et al.*, 1991) are marked with orange boxes at positions −146, −500 and −490; the last

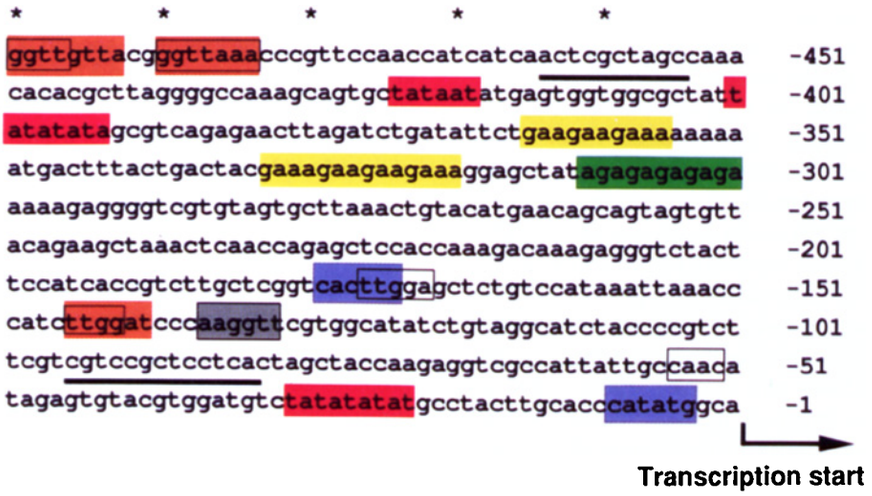


Figure 4. *Sn* gene promoter sequence and relevant features.
The first 500 nucleotides present upstream of the transcription start of the *Sn* gene are shown; putative transcription factor binding sites are evidenced. TATA box-like sequences are marked with red boxes at positions −31, −401 and −424. A sequence resembling the CAAT box (−55) is marked with an open box. A gray box (−137) highlights a sequence similar to an SV40 promoter enhancer also described in the promoter of the *A1* gene as well as of other plant genes. Two boxes (CATATG and CACTTG, respectively, shown by blue boxes) at positions −9 and −179 correspond to recognition sites for the DNA binding basic HLH (B-HLH) domain of certain transcriptional activators. A GAGA transcription factor binding site sequence found in the promoters of the transcription activators *Ultrabithorax* and *engrailed* in *Drosophila* is shown within a green box. Sequences resembling the DNA region bound by GT and related transcription factors are evidenced with orange boxes at positions −146, −490 and −500 (the core GT binding region is marked). Two other GT-rich sequences are evidenced with lines starting at positions −46 and −414, respectively. Two tandemly repeated sequences of 10 bp (GAAGAAGAAA, yellow boxes) are localized at positions −365 and −333, respectively.

one shows a sequence (GGTTAAA) that fits almost completely the core binding region of GT factor in the promoters of several light-regulated genes (Green *et al.*, 1988). Two other GT rich sequences are observed at positions −46 and −414, respectively, while two tandemly repeated sequences of 10 bp (GAAGAAGAAA marked out with yellow boxes) without any apparent role are localized at positions −365 and −333, respectively. A comparison of the first 140 nucleotides of *Sn* with those of *Lc* (Ludwig *et al.*, 1989) shows 100% homology.

The indications of potential regulatory sites as outlined in this computer analysis might be useful for a functional dissection of the *Sn* promoter. Particularly the presence of GAGA and GT factor binding sites might account for the developmental and light-induced expression of the *Sn* gene. Furthermore, sequence comparison of promoters from different members of the *R* family together with deletion analysis experiments might help in elucidating the basis of the differential expression of these genes.

Conclusions

The functional and structural relationship of *R*, *Sn*, *Lc* and *B* may best be explained by postulating a common evolutionary origin from a single ancestral gene. *R* and *B* lie on different chromosomes possibly as the result of a genome or chromosome duplication event. On the other hand, *Sn* and *Lc* mapping on the same chromosome arm as *R* and not ubiquitous in their distribution but apparently restricted to a closely related group of maize races, may represent an intrachromosomal duplication of the *R*-containing region as a result of unequal exchange or translocation. Molecular divergence may then have contributed to the establishment of functionally distinct genes, each one with its own tissue specificity. Since the molecular divergence of *R*, *Lc* and *Sn* is thought to have occurred relatively recently compared with that of *R* and *B*, the analysis of the nucleotide sequence should indicate that *Sn* and *Lc* are more closely related to *R* than *R* to *B*. This expectation is confirmed by a comparison of their cDNA nucleotide sequences. Even though the dependence of tissue pigmentation on an active *R* gene product has been well established, the mechanism accounting for the tissue-specific pattern of pigmentation and for the synthesis of an *R* gene product remain unknown. It might well be that the basis of these events is to be sought at different levels of gene regulation. Existence of control at the transcriptional level is shown by the observation that the messenger of specific *R* alleles is confined to the tissue of competence. So the mRNA of *R-g* is detectable in the aleurone but not in the pericarp (Figure 1).

The existence of mechanisms of regulation operating at post-transcriptional level is suggested by the recent observation (unpublished results) that the mRNA level of *A1* and *C2* in roots and aleurones is increased by light treatments, while that of *R* remains constant. The sequence

analysis reported in this work suggests two possible levels of post-transcriptional regulation, one translational and one post-translational.

Although changes in the synthesis of a given transcription factor in response to physiological signals may provide an effective way of transcriptional regulation, a rapid and potentially more versatile way is its control by post-translational modifications that could operate by channelling the factor to a particular subcellular location, or modulating its stability or activating or repressing its DNA-binding or transcriptional activation function.

The polymorphism observed among the different proteins does not affect the basic protein activity since the products of the genes of the *R* family are functionally equivalent as shown by transient expression experiments (Goff *et al.*, 1990). Putative sites of modifications that are highly conserved might thus be of functional significance. On this basis we are currently investigating the possibility of R protein modifications.

The observation that *R* members, conditioning different patterns of pigmentation, share a conserved product, points to the promoter as the region responsible for tissue specificity. The analysis of the first 500 bp of the *Sn* promoter reported here discloses the presence of putative boxes that might be involved in the temporal and spatial expression and in the light-inducibility. In this context it will be informative to compare the promoter sequences of different genes of the *R* family as well as the dissection of these controlling regions. Another feature that makes this system appealing for the study of gene regulation, is the availability of genetic data pointing to an interaction between members of the *R* family (paramutation).

When two alleles are in heterozygous association, the expression of one allele does not affect the expression of the other. Paramutation contradicts this basic genetic tenet. This phenomenon in fact, as defined by Brink (1973), is an interaction between alleles that leads with high frequency to a heritable change of one of the two alleles within one generation. For example when *R-r* and *R-st* alleles are brought in heterozygous association and then extracted by test-crosses, *R-r* appears changed in its expression in a heritable manner even though it shows a tendency to regain its original level of expression in succeeding generations.

In the *Sn* paramutation events, *R* genes extracted from *R/Sn* heterozygote, exhibit an *Sn* phenotype in their mesocotyls, as if in the heterozygous association *Sn* had imposed an imprinting on the *R* gene leading to its expression in an additional tissue. This imprinting effect is negatively affected by chromosomal rearrangement and varies in its frequency (10–40%) according to the *R* alleles (Gavazzi *et al.*, 1990). The molecular basis of these interactions is still unknown (see Dooner *et al.*, 1991 for a discussion on these interactions). Recent studies on transgenic plants have disclosed similar events taking place between ectopic genes and the resident homologous genes in ways that lead to changes in gene expression at either one or both locations (see Jorgensen 1990 for a review).

It might well be that these interactions between members of gene families and those between ectopic and resident genes have a common basis and that their study will disclose new levels of gene regulation.

Experimental procedures

Genetic stocks

All seed stocks used in this study are in the W22 background and are homozygous for the color factors *A1, A2, C1, C2, Pr, Bz1, Bz2* and *pl* while differing for their *R* or *Sn* constitution.

The R *locus.* Detailed description of the origin, phenotype and structural characteristics of the *R* alleles used in this study can be found in Dooner and Kermicle (1974). *R-r: standard* as a gene complex, consisting of two gene components, symbolized by (P) and (S) controlling plant and seed color, respectively. Properties of the *R* alleles that are pertinent to this presentation are as follows. *R-g*: purple aleurone, green plant, an (S) derivative of *R-r*. *r-r*: colorless aleurone, red plant, it confers pigmentation to roots, coleoptile and leaf tip of seedlings and to anthers of mature plants, a (P) derivative of *R-r*. *R-sc*: self-colored aleurone, green plant, a germinal derivative of *R-st* that confers a homogeneous pigmentation of the aleurone. *r-g*: colorless aleurone, green plant, the null allele.

The Sn *locus.* A factor lying two map units distal to *R*, conferring specific tissue pigmentation (following light exposure) to the scutellar node, mesocotyl, leaf basis and midrib and to the seed integuments (glumes and pericarp). Four independent *Sn* accessions were transferred to a common genetic background by repeated backcrosses into a colored version of the W22 inbred line. All four accessions are borne on an *r* marked chromosome that confers weak red pigmentation to the anthers (Gavazzi *et al.*, 1990, Tonelli *et al.*, 1991).

The *Sn:bol3* accession differs from the others in that it confers (following light exposure) a higher pigmentation level in the mesocotyl, at least six times greater than that conferred by other *Sn* alleles, and a weak color development of seed integuments, mesocotyl and scutellar nodes even in the dark.

RNA isolation and Northern blot analysis

Seeds were allowed to germinate in darkness for 5 days at 25°C and then exposed to white light for 24 h at 21°C. As a light source Power Stars-HQ1-T400 W/DV OSRAM lamps were used with photon fluence rate of 68 W m^{-2}. Total RNA was extracted separately from mesocotyls and roots. Pericarps and aleurones were peeled off from developing seeds (14 and 30 days after pollination) exposed to light for 24 h. Tissues were ground in liquid N_2 using pestle and mortar and RNA was isolated by a previously described method (Van Tunen *et al.*, 1988). For Northern blot analysis 20 or 40 μg total mRNA was loaded on formaldehyde gels (Maniatis *et al.*, 1982) and after electrophoresis blotted on Biodyne B nylon membrane (Pall) according to manufacturer's

instructions. Filters were prehybridized for 2 h at 65°C in 10% dextran sulfate (Pharmacia), 6 × SCP (1 × SCP = 100 mM NaCl, 30 mM Na_2HPO_4, and 1 mM EDTA, pH 6.5), 2% sarcosine and 500 μg ml^{-1} heparin (type II Sigma). Hybridization was performed for 12 h at 65°C in the above buffer containing denatured salmon sperm DNA (final concentration 100 μg ml^{-1}) and denatured probe DNA (final probe concentration = 10 ng ml^{-1}). The filters were washed three times in 2 × SCP, 1% SDS at 65°C for 30 min and then in 0.2 × SCP, 0.1% SDS at 65°C for 15 min as previously described (Tonelli *et al.*, 1991). RNA molecular weight markers were from Bethesda Research Laboratory.

DNA probes: the 1.4 kb *Pst*I–*Eco*RI fragment of *Sn* cDNA (Tonelli *et al.*, 1991), the 700 bp *Bam*HI fragment of *A1* gene (Schwarz-Sommer *et al.*, 1987) and the 1 kb maize tubulin cDNA (Mereghetti *et al.*, 1990) were used. Restriction fragments from plasmid subclones were extracted from low melting point agarose and labeled by random primed method (Boehringer). The specific activities of the probes ranged between 2 to 3 × 10^8 d.p.m. μg^{-1}.

DNA sequencing

The 3.8 kb *Hin*dIII genomic clone (pSn3.8H) containing the promoter and the transcription start (Tonelli *et al.*, 1991), have been subcloned into the plasmid vector pBS (Stratagene cloning system) and sequenced according to Sanger *et al.* (1977) using oligonucleotides as primers.

Computer analysis tools

The alignment of the four protein sequences was performed on a Macintosh IIci computer using the ClustalV package (Higgins *et al.*, 1992) based on the original multiple-sequence-alignments algorithm described by Higgins and Sharp (1989).

The search for modification sites and structural motifs on the protein sequences described was accomplished with the use of the program MacPattern (Fuchs, 1991) and the PROSITE database (Bairoch, 1991) as source of protein sites and patterns.

The projection of the two α-helices of the HLH domain inferred from protein analysis was obtained with the PLOTA/HEL program from the MacProt package (Angela Luttke, Arnimstr. 2 5000, Koln 30, Germany).

The search for transcription-control sites in the sequence of the promoter of *Sn* gene was achieved by comparison against the transcription factor database (TFD; Ghosh, 1990) using the MacPattern program described above (Fuchs, 1991).

The 10 bp long direct repeat found in the promoter sequence was highlighted by the program REPEAT of the GCG package from University of Wisconsin (Devereaux *et al.*, 1984) running on a MicroVax 3900 (Digital Equipment Corporation) under the VMS operating system (version 5.3).

The secondary structures of the 5′ leader regions of *Sn* and *Lc* transcripts (Figure 3b) were calculated with the program ZUKFOLD (Zuker and Stiegler, 1981) running on a MicroVax 3900. The program calculates the thermodynamically most probable conformation of large primary nucleotide sequences: it finds RNA minimum free energy secondary structures by using published values of stacking and loop destabilizing energies. The energy output files generated by the program for the two molecules were translated by the program LoopViewer into the final looping structures (Gilbert, D.G., 1990; LoopViewer, a Macintosh program for visualizing RNA secondary structure; published electronically on the Internet at the address: bio.nih.edu).

Acknowledgments

Research supported by funds of the BRIDGE program of the Commission of the European Communities (Contract BIOT-0164-C EDB) to CT.

References

Abastado, J.P., Miller, P.F., Jackson, B.M. and Hinnebush, A.G. (1991) Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for GCN4 translational control. *Mol. Cell. Biol.* **11**, 486–496.

Bairoch, A. (1991) PROSITE: A dictionary of sites and patterns in proteins. *Nucl. Acids Res.* **19**, 2241–2245.

Biggin, M.D. and Tjian, R. (1989) Transcription factors and the control of *Drosophila* development. *Trends Genet.* **5**, 377–383.

Binetruy, B., Smeal, T. and Karin, M. (1991) Ha-*ras* augments c-Jun activity and stimulates phosphorylation of its activation domain. *Nature*, **351**, 122–127.

Blackwell, T.K. and Weintraub, H. (1990) Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science*, **250**, 1104–1110.

Brink, R.A. (1973) Paramutation. *Ann. Rev. Genet.* **7**, 129–152.

Chandler, V.L., Radicella, P.J., Robbins, T.P., Chen, J. and Turks, D. (1989) Two regulatory genes of the maize anthocyanin pathway are homologous: isolation of *B* utilizing *R* genomic sequences. *Plant Cell*, **1**, 1175–1183.

Coe, E.H., Hoisington, D.A. and Neuffer, M.G. (1988) The genetics of corn. In *Corn and Corn Improvement* (Sprague, G.F. and Dudley, J., eds). Madison, WI: American Society of Agronomy Inc., pp. 81–258.

Cone, K.C., Burr, F.A. and Burr, B. (1986) Molecular analysis of the maize anthocyanin regulatory locus *C1*. *Proc. Natl Acad. Sci. USA*, **83**, 9631–9635.

Consonni, G., Viotti, A., Dellaporta, S.L. and Tonelli, C. (1992) cDNA nucleotide sequence of *Sn*, a regulatory gene in maize. *Nucl. Acids Res.* **20**, 373.

Datta, N. and Cashmore, A.R. (1989) Binding of a pea nuclear protein to promoters of certain photoregulated genes is modulated by phosphorylation. *Plant Cell*, **1**, 1069–1077.

Dellaporta, S.L., Greenblatt, I., Kermicle, J.L., Hicks, J.B. and Wessler, S.R. (1988) Molecular cloning of the maize *R-nj* allele by transposon tagging with *Ac*. In *Chromosome Structure and Function*, 18th Stadler Genetics Symposium (Gustafson, J.P. and Appels, R., eds). New York: Plenum Publishing Co., pp. 263–282.

Devereaux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387–395.

Dooner, H.K. and Kermicle, J.L. (1974) Reconstitution of the *R-r* compound allele in maize. *Genetics*, **78**, 691–701.

Dooner, H.K. and Kermicle, J.L. (1976) Displaced and tandem duplication in the long arm of chromosome 10 in maize. *Genetics*, **82**, 309–322.

Dooner, H.K. and Nelson, O.E. Jr. (1977) Genetic control of UDP-glucose-flavanol 3-*O*-glucosyl transferase in the endosperm of maize. *Biochem. Genet.* **15**, 509–519.

Dooner, H.K., Robbins, T.P. and Jorgensen, R.A. (1991) Genetic and developmental control of anthocyanin biosynthesis. *Ann. Rev. Genet.* **25**, 173–179.

Flemisco, J.R., Glass, D.B. and Krebs, E.G. (1980) Optimal spatial requirements for the location of basic residues in peptide

substrates for the cyclic AMP-dependent protein kinase. *J. Biol. Chem.* **255**, 4240–4245.

Fuchs, R. (1991) MacPattern: Protein pattern searching on the Apple Macintosh. *CABIOS*, **7**, 105–106.

Gavazzi, G., Viani, I., Jacomini, E. and Helm, B. (1986) The genetics of *Sn*, a factor responsible for tissue specific anthocyanin formation in higher plants. In *Gene Structure and Function in Higher Plants*, Proc. Int. Symp. held at Hyderabad (Reddy, G.M. and Coe, E.H. Jr., eds). New Delhi: Oxford & Ibh Publishing Co., pp. 91–103.

Gavazzi, G., Mereghetti, M., Consonni, G. and Tonelli, C. (1990) *Sn*, a light-dependent and tissue-specific gene of maize: the genetic basis of its instability. *Genetics*, **125**, 193–199.

Ghosh, D. (1990) A relational database of transcription factors. *Nucl. Acids Res.* **18**, 1749–1756.

Glass, D.B. and Smith, S.B. (1983) Phosphorylation by cyclic GMP-dependent protein kinase of a synthetic peptide corresponding to the autophosphorylation site in the enzyme. *J. Biol. Chem.* **258**, 14 797–14 803.

Goff, S.A., Klein, T.M., Roth, B.A., Fromm, M.E., Cone, K.C., Radicella, J.P. and Chandler, V.L. (1990) Transactivation of anthocyanin biosynthetic genes following transfer of *B* regulatory genes in maize tissues. *EMBO J.* **9**, 2517–2522.

Gonzalez, G.A. and Montminy, M.R. (1989) Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. *Cell*, **59**, 675–680.

Green, P.J., Yong, M.H., Cuozzo, M., Kano-Murakami, Y., Silverstein, P. and Chua, N.H. (1988) Binding site requirement for pea nuclear protein factor GT-1 correlate with sequences required for light-dependent transcriptional activation of the *rbcS-3A* gene. *EMBO J.* **7**, 4035–4044.

Higgins, D.G. and Sharp, P.M. (1989) Fast and sensitive multiple sequence alignments on a microcomputer. *CABIOS*, **5**, 151–153.

Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1992) CLUSTAL V: improved software for multiple sequence alignment. *CABIOS*, **6**, 189–191.

Hunter, T. (1982) Synthetic peptide substrates for a tyrosine protein kinase. *J. Biol. Chem.* **257**, 4843–4848.

Jackson, S.P. and Tjian, R. (1988) *O*-glycosylation of eukaryotic transcription factors: implications for the mechanisms of transcriptional regulation. *Cell*, **55**, 125–133.

Jackson, S.P., MacDonald, J.J., Lees-Miller, S. and Tjian, R. (1990). GC box binding induces phosphorylation of Sp1 by a DNA-dependent protein kinase. *Cell*, **63**, 155–165.

Jorgensen, R. (1990) Altered gene expression in plants due to *trans* interactions between homologous genes. *Trends Biotech.* **8**, 340–344.

Klimczak, L.J., Schindler, U. and Cashmore, A.R. (1992) DNA binding activity of the *Arabidopsis* G-box binding factor GBF1 is stimulated by phosphorylation by casein kinase II from broccoli. *Plant Cell*, **4**, 87–98.

Kozak, M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucl. Acids Res.* **12**, 857–872.

Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283–292.

Kozak, M. (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* **266**, 1991.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132.

Lawton, M.A., Dean, S.M., Dron, M., Kooter, J.M., Kragh, K.M., Harrison, M.J., Yu, L., Tanguay, L., Dixon, R.A. and Lamb, C.J. (1991) Silencer region of a chalcone synthase promoter contains multiple binding sites for a factor, SBF-1, closely related to GT-1. *Plant Mol. Biol.* **16**, 235–249.

Ludwig, S.R., Habera, L.F., Dellaporta, S.L. and Wessler, S.R. (1989) *Lc*, a member of the maize *R* gene family responsible for tissue-specific anthocyanin production encodes a protein similar to transcriptional activators and contains the *myc*-homology region. *Proc. Natl Acad. Sci. USA*, **86**, 7092–7096.

Ludwig, S.R., Bowen, B., Beach, L. and Wessler, S.R. (1990) A regulatory gene as a novel visible marker for maize transformation. *Science*, **247**, 449–450.

Luscher, B., Kuenzel, E.A., Krebs, E.G. and Eisenman, R.N. (1989) Myc oncoproteins are phosphorylated by casein Kinase II. *EMBO J.* **8**, 1111–1119.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Marshall, R.D. (1972) Glycoproteins. *Ann. Rev. Biochem.* **41**, 673–702.

Mereghetti, M., Consonni, G. and Tonelli, C. (1990) A cDNA clone of the maize α-tubulin. *Maize Gen. Coop. News Lett.* **64**, 87–88.

Murre, C., Schonleber, McCaw, P. and Baltimore, D. (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding Daughterless, MyoD and Myc proteins. *Cell*, **56**, 777–783.

Nagy, F., Kay, S.A. and Chua, N.H. (1988) Gene regulation by phytochrome. *Trends Genet.* **4**, 37–42.

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.

Perrot, G.H. and Cone, K.C. (1989) Nucleotide sequence of the maize *R-S* gene. *Nucl. Acids Res.* **17**, 8003.

Pinna, L.A. (1990) Casein kinase 2: an 'eminence grise' in cellular regulation? *Biochim. Biophys. Acta*, **1054**, 267–284.

Pless, D.D. and Lennarz, W.J. (1977) Enzymatic conversion of proteins to glycoproteins. *Proc. Natl Acad. Sci. USA*, **74**, 134–138.

Radicella, J.P., Turks, D. and Chandler, L.V. (1991) Cloning and nucleotide sequence of a cDNA encoding *B-Peru*, a regulatory protein of anthocyanin pathway in maize. *Plant Mol. Biol.* **17**, 127–130.

Robbins, T.M., Walker, E.L., Kermicle, J.L., Alleman, M. and Dellaporta, S.L. (1991) Meiotic instability of the *R-r* complex arising from displaced intragenic exchange and intrachromosomal rearrangement. *Genetics*, **129**, 271–283.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.

Sarokin, L.P. and Chua, N.H. (1992) Binding sites for two novel phosphoproteins, 3AF5 and 3AF3, are required for rbcS-3A expression. *Plant Cell*, **4**, 473–483.

Schmidt, R.J., Burr, F.A., Aukerman, M.J. and Burr, B. (1990) Maize regulatory gene *opaque 2* encodes a protein with a "leucine zipper" motif that binds to zein DNA. *Proc. Natl Acad. Sci. USA*, **87**, 46–50.

Schwarz-Sommer, Z., Shepherd, N., Tacke, E., Gierl, A., Rohde, W., Leclercq, L., Mattes, M., Berndtgen, R., Peterson, P.A. and Saedler, H. (1987) Influence of transposable elements on the structure and function of *A1* gene of *Zea mays*. *EMBO J.* **6**, 287–294.

Siegfried, E., Ambrosio, L. and Perrimon, N. (1990) Serine/threonine protein kinases in *Drosophila*. *Trends Genet.* **6**, 357–362.

Sorger, P.K., Lewis, M.J. and Pelham, H.R.B. (1987) Heat shock factor is regulated differently in yeast and HeLa cells. *Nature*, **329**, 81–84.

Stadler, L.J. (1948) Spontaneous mutation at the *R* locus in maize. Race differences in mutation rate. *Am. Nat.* **82**, 289–314.

Styles, D.E. and Ceska, O. (1977) The genetic control of flavonoid synthesis in maize. *Can. J. Genet. Cytol.* **19**, 289–302.

Tanaka, M. and Herr, W. (1990) Differential transcriptional activation by Oct-1 and Oct-2: Interdependent activation domains induce Oct-2 phosphorylation. *Cell*, **60**, 375–386.

Tonelli, C., Consonni, G., Faccio Dolfini, S., Dellaporta, S.L., Viotti, A. and Gavazzi, G. (1991) Genetic and molecular analysis of *Sn*, a light-inducible, tissue specific regulatory gene in maize. *Mol. Gen. Genet.* **225**, 401–410.

Van Tunen, A.J., Koes, R.E., Spelt, C.E., van der Krol, A.R., Stuitje, A.R. and Mol, N.M. (1988) Cloning of two chalcone flavanone isomerase genes from *Petunia hybrida*: coordinate, light regulated and differential expression of flavonoid genes. *EMBO J.* **7**, 1257–1263.

Weiher, H., Konig, M. and Gruss, P. (1983) Multiple point mutations affecting the Simian Virus 40 enhancer. *Science*, **219**, 626.

Wienand, U., Weydemann, U., Niesbach-Kloesgen, U., Peterson, P.A. and Saedler, H. (1986) Molecular cloning of *C2* locus of *Zea mays*, the gene coding for chalcone synthase. *Mol. Gen. Genet.* **203**, 202–207.

Woodget, J.R., Gould, K.L. and Hunter, T. (1986) Substrate specificity of protein kinase C. *Eur. J. Biochem.* **161**, 177–184.

Yamamoto, K.K., Gonzalez, G.A., Biggs III, W.H. and Montminy, M.R. (1988) Phosphorylation-induced binding and transcriptional efficacy of a nuclear factor CRE. *Nature*, **334**, 494–498.

Zuker, M. and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucl. Acids Res.* **9**, 133–148.

EMBL Data Library accession number X67619 (Figure 4).