The EMBO Journal vol.9 no.2 pp.315-321, 1990

Molecular analysis of the C1-I allele from *Zea mays*: a dominant mutant of the regulatory C1 locus

Javier Paz-Ares, Debabrota Ghosal and Heinz Saedler

Max-Planck-Institut für Züchtungsforschung, Carl-von-Linne-Weg 10, 5000 Köln 30, FRG

Communicated by H.Saedler

The C1 locus of *Zea mays* (maize) controls the expression of genes involved in anthocyanin biosynthesis in aleurone and scutellar tissue and encodes a protein with the features of a transcriptional activator. C1-I is a dominant negative mutant which inhibits pigment formation. The structure of the C1-I allele was determined by cloning and sequencing of this allele and of two distinct C1-I derived cDNAs. C1-I has two major and several minor sequence differences with respect to the wild-type C1 allele. Transcription initiation occurs at the same position as in wild-type but transcription yields two different products, one major RNA of 1.3 kb and one minor RNA of 1.45 kb in length, encoding two proteins of 252 and 108 amino acids respectively. The longer 252 amino acid C1-I protein differs from the 273 amino acid wild-type C1 protein at several positions but most prominently at its carboxy terminus, resulting in reduced acidity of the C1-I protein. A similar change in acidity of the Gal4 protein of yeast converted this transcriptional activator into a repressor protein. We discuss the dominant phenotype of C1-I with respect to its possible repressor function in contrast to the activator function of the C1 gene product.

Key words: C1 locus/C1-I allele/*Zea mays*

Introduction

Anthocyanins are plant pigments whose synthesis is both environmentally and developmentally controlled (Dixon, 1986; Harborne, 1986). The anthocyanin pathway of *Zea mays* represents an excellent system for studying the molecular basis of gene regulation in plants, because it is genetically well characterized and a number of genes involved in the pathway have been cloned (Fedoroff *et al.*, 1984; O'Reilly *et al.*, 1985; Cone *et al.*, 1986; Paz-Ares *et al.*, 1986; Wienand *et al.*, 1986; Theres *et al.*, 1987).

Several maize regulatory genes (B, C1, P1, R, Vp1, Dek) which mediate developmental and/or environmental control of the anthocyanin structural genes (C2, A1, A2, Bz1, Bz2 and Pr) have been genetically and biochemically described (Dooner and Nelson, 1979; Dooner, 1983). In particular, the function of the regulatory C1 locus is required for the synthesis of anthocyanins in the aleurone and scutellar tissues of maize kernels (Coe and Neuffer, 1977). The C1 locus has been cloned by transposon tagging (Cone *et al.*, 1986; Paz-Ares *et al.*, 1986), and molecular anlaysis has shown that a C1-encoded protein shares partial homology with *myb* proto-oncogene products and probably represents a transcriptional activator (Paz-Ares *et al.*, 1987).

A number of recessive C1 mutants have been identified which are colourless if homozygous, and coloured if heterozygous in the presence of the C1 wild-type allele. In contrast to these mutants, another colourless allele, C1-I, has been described (Emerson *et al.*, 1935; Coe, 1962) which suppresses the wild-type C1 allele in the heterozygous condition. In order to understand the molecular basis of the C1-I transdominant effect, genomic DNA and two cDNAs corresponding to this allele were cloned and analysed. Several differences between the C1-I and C1 alleles could be detected. The possible involvement of the major protein encoded by the C1-I allele in the interference with C1 function is discussed within the framework of positive control of gene regulation.

Results

Sequence comparison of the C1-I and wild-type alleles

A λEMBL4 genomic library was prepared from *Mbo*I partially digested DNA of a plant homozygous for C1-I. Using the central 1 kb *Eco*RI fragment present in the genomic clone of the wild-type C1 allele (clone no. 5, Paz-Ares *et al.*, 1986, see Figure 1) as a probe, a homologous clone (no. 62I) containing a 13 kb insert was isolated.

Heteroduplex analysis between phage DNAs from the C1 wild-type clone no. 5 and the C1-I clone no. 62I revealed two single-stranded substitution loops and one single-stranded insertion/deletion loop in an otherwise double-standed molecule (Figure 1A and B). The single-stranded loop 1 in Figure 1(B) reflects the fact that clone no. 62I extends beyond the 3′ end of clone no. 5. Substitution loop 2 is a replacement of 370 bp of wild-type sequences by a 3.8 kb insert in the C1-I allele. The terminal substitution loop 3 indicates sequence divergence between the 5′ part of both alleles as confirmed by Southern hybridization analysis (Figure 1A and B). The cause of the third sequence divergence was not investigated further since it is located far outside (~2.2 kb) of the C1 transcription unit. The relevant portion of the restriction map of the C1-I clone no. 62I is compared with that of the C1 wild-type clone no. 5 in Figure 1(C). Southern blot anlaysis of DNA from phage no. 62I and of genomic DNA isolated from homozygous C1-I maize lines confirmed the size and relative positions of the two DNA sequence alterations (hatched boxes in Figure 1C), and indicated that no major rearrangements occurred during the cloning step. The restriction map of C1-I is also in agreement with data previously reported by Cone *et al.* (1986). The substitution of 370 bp in the C1 allele for a 3.8 kb DNA sequence in C1-I was interpreted by Cone *et al.* (1986) as a simple insertion of 3.5 kb, a size that corresponds well with the difference in size of the segments that are in fact exchanged.

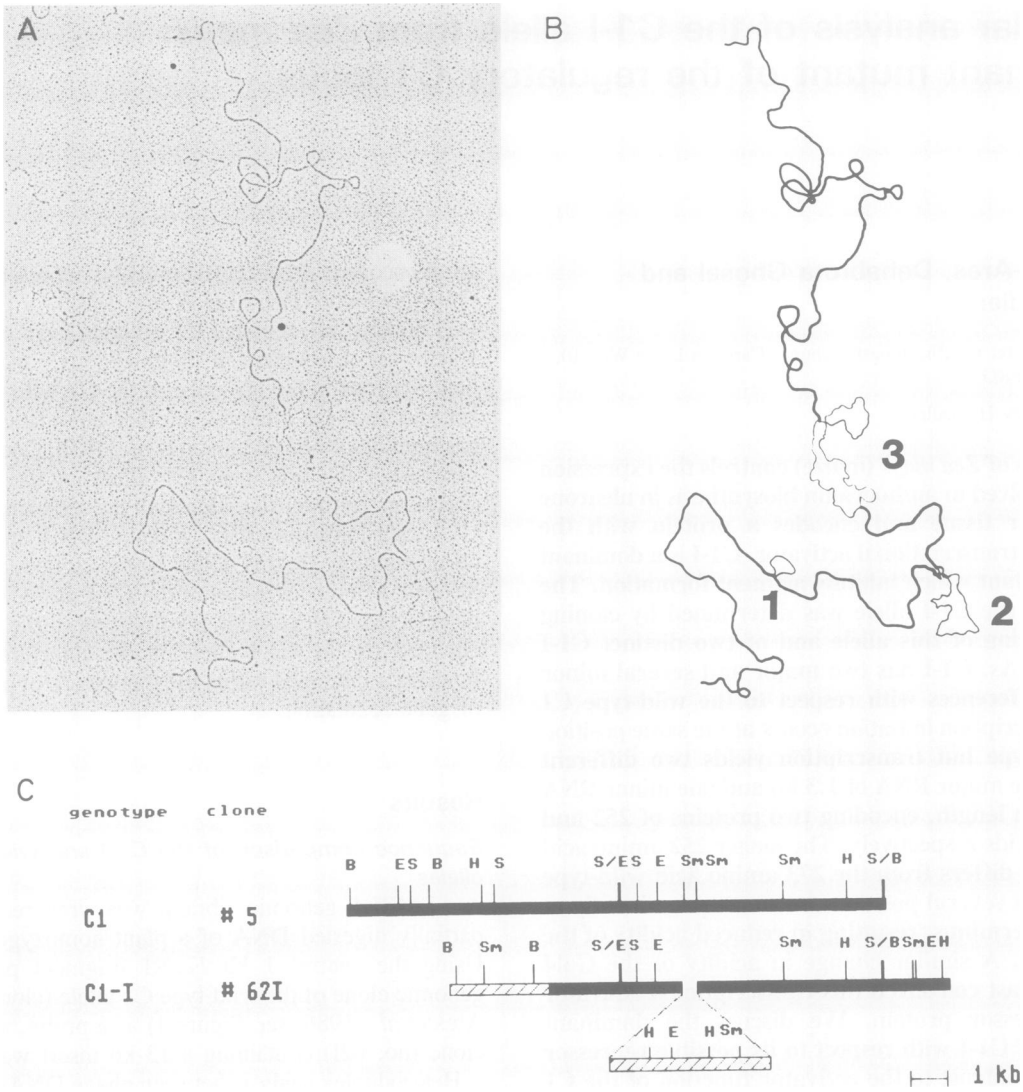

Fig. 1. Comparison of the C1-I and wild-type C1 allele DNA sequences by heteroduplex analysis and restriction mapping. **(A)** Electron micrograph of heteroduplex between DNAs from recombinant phages no. 5 and no. 62I, representing the C1 and C1-I alleles, revealing the presence of three loops. **(B)** Schematic drawing of the heteroduplex. The three loops are numbered 1–3. **(C)** Restriction map of the DNA inserts present in no. 5 and no. 62I. Solid bars in both maps represent regions of homology with the wild-type C1 locus DNA, as detected by Southern hybridization experiments. For restriction enzymes the following abbreviations were used: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *Sal*I; Sm, *Sma*I.

To get further insights into the structure of the C1-I allele, a 2.2 kb fragment was sequenced. This fragment extends from the promoter proximal region (position −613) to 500 bp into the 3.8 kb substitution sequence (position 1590). Comparison of this sequence with that of the wild-type C1 allele (Figure 2) revealed several alterations, including both transitions and transversions as well as small insertions and deletions. A total of 18 positions are altered. Twelve are 1–2 bp alterations, the remainder affect 3 bp or more. Except for the larger 3.8 kb alteration, four insertions/deletions reflect typical footprints suggestive of excision of a transposable element (Schwarz-Sommer *et al.*, 1984).

The most striking of these is an insertion of 8 bp, at position 992 in Figure 2 which would result in a translational frameshift (see below).

Transcripts of the C1-I allele differ from those of the C1 allele

Two C1-I-specific RNAs had been previously observed by Cone *et al.* (1986) and are shown in Figure 3. To investigate the C1-I transcripts, a cDNA library was prepared in the vector λNM1149 from poly(A)$^+$ RNA obtained from developing C1-I kernels and screened using the 1 kb *Eco*RI fragment (from −612 to +460, Figure 2) as a probe. Two cDNAs, cCI5 and cCI49, 1.1 and 1.2 kb in length respectively, were identified and sequenced. The two sequences were similar. However, they differed by an internal 142 bp insertion in cCI49, and length heterogeneity between their 3′ ends as well as by a sequence divergence between the 5′ ends (data not shown). The last difference seemed to be the result of artefactual ligation of an unrelated cDNA with C1-I cDNA in the cloning procedure. In fact, additional sequence located 5′ in cCI49 differed from the corresponding genomic sequence and included a poly(A) track at its 5′ end. Primer extension experiments, similar to those conducted previously with the wild-type C1 allele (Paz-Ares *et al.*, 1987) (except for the use of two different primers: see Materials and methods), have shown that the start of transcription is at the same position in both C1 and C1-I alleles (Figure 4).

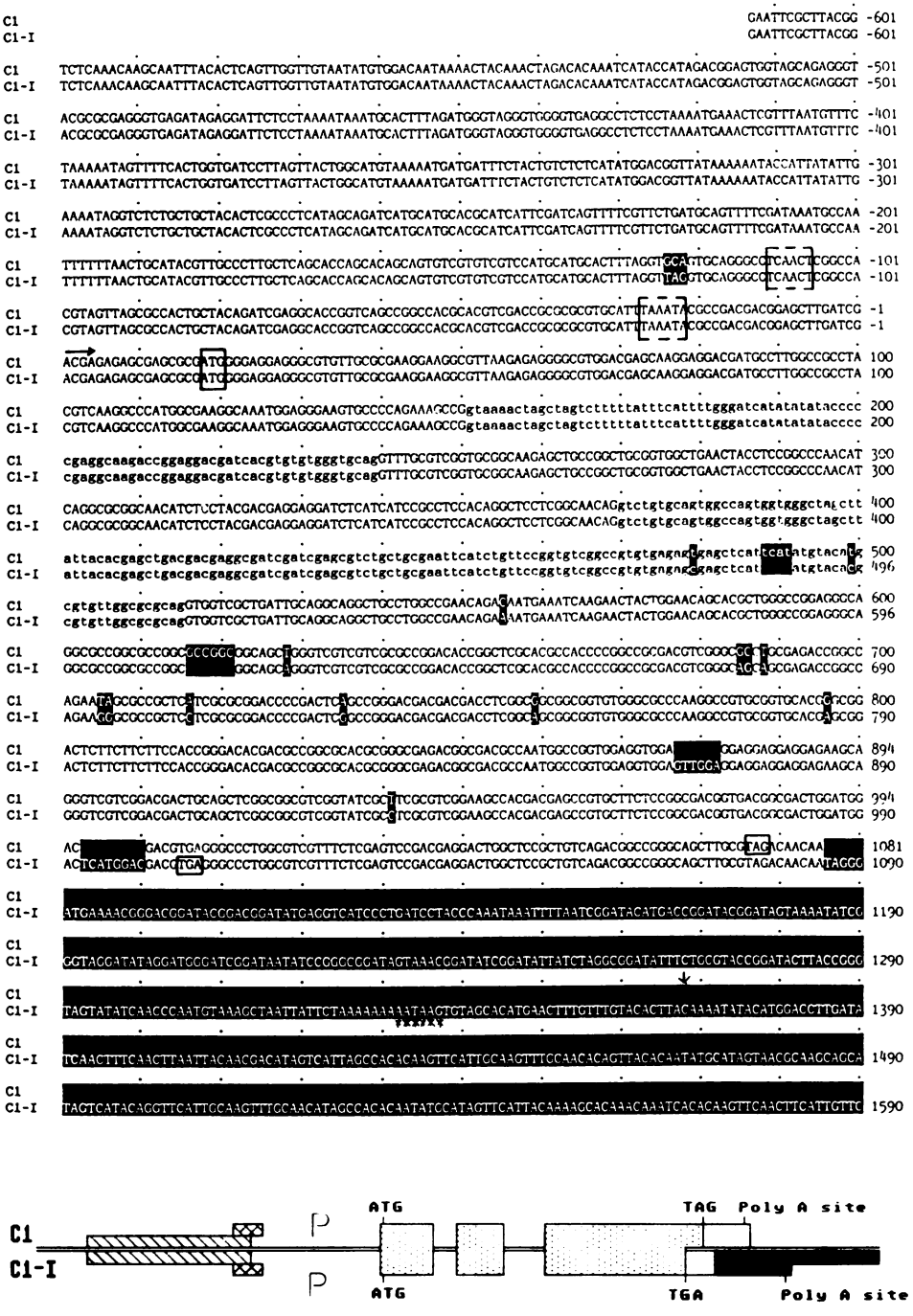

Fig. 2. Sequence and structural comparison between the C1 and the C1-I alleles. The upper part of the figure shows the sequence of a 2.2 kb C1-I DNA fragment, from 613 bp upstream of the start of transcription to 200 bp downstream of the end of the C1-I RNAs inclusive, compared with homologous sequences of the wild-type locus (Paz-Ares *et al.*, 1987). Sequence differences between both wild-type and mutant alleles are boxed in black. Genomic C1-I sequences are compard with the major C1-I transcript sequences represented by cDNA clone cCI5 (see text) and the genomic wild-type C1 allele. The resulting intron sequences are given in lower-case letters. The horizontal arrow at position +1 indicates the start of transcription. CAAT and TATA signals (at position −111 and −27 respectively) are boxed with broken lines. Translation start (at position +18 in both alleles) and stop codons of the major C1-I transcript and of the C1 transcripts (at positions +1005 and 1067 respectively) are boxed with continuous lines. The polyadenylation signal of the major C1-I RNA (at position +1331) is underlined with stars. The arrow at position +1368 indicates the end of the major C1-I transcript. Not shown in the figure is the stop codon (at position 1365), the polyadenylation signal (at position 1371) and the end of the minor C1-I transcript (at position +1393), represented by clone cCI49 (see text). The C1-I sequence from position +1085 on represents part of the 3.8 kb sequence unique for the C1-I allele (see Figure 1 and text). The lower part of the figure shows a schematic drawing of the intron−exon structure of the C1-I allele compared with that of the C1 wild-type allele (Paz-Ares *et al.*, 1987). For reasons of clarity, only the major transcript of C1-I (represented by clone cCI5) and the 1.2 kb transcript of C1 (represented by cDNA clone cLC6, Paz-Ares *et al.*, 1987) are considered. The hatched and cross-hatched boxes upstream from the promoter (P) show the region of homology of the C1 locus with the C2 locus and with the small RNAs hybridizing with the C1 locus respectively (Paz-Ares *et al.*, 1986, 1987). The boxes downstream of the promoter indicate exons. The shadowed parts of the boxes indicate the translated part of the transcript. The black area of the lower box indicates the 3.8 kb sequence unique for the C1-I allele, part of which is present in its transcripts.

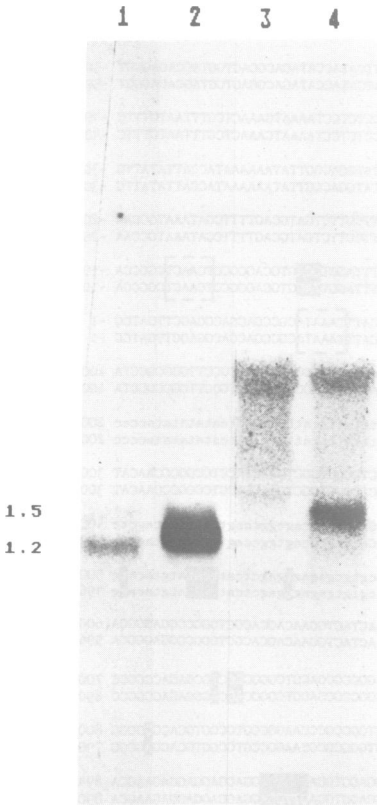

Fig. 3. Transcript analysis of C1-I and C1 alleles. Northern blots of poly(A)$^+$ RNA (5 μg) from 30 day old developing endosperms of C1 wild-type (**lanes 1** and **3**) and C1-I mutant (**lanes 2** and **4**) hybridized to the *Eco*RI−*Xho*I fragment +12 to +1020 derived from the wild-type cLC28 cDNA clone (**lanes 1** and **2**) and to the *Eco*RI−*Bss*HII probe (from +445 to +508, Figure 2) derived from the second intron (**lanes 3** and **4**).

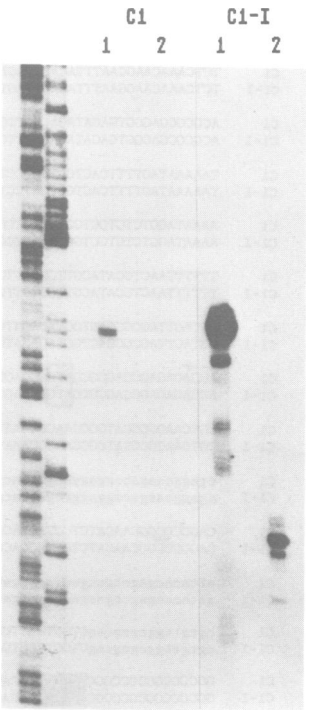

Fig. 4. Determination of the transcription start of C1-I compared with C1 transcripts. Poly(A)$^+$ RNA from kernels homozygous for C1 and homozygous for C1-I were primed with two 30mer synthetic oligonucleotides, one exending from position +241 to +271 (**lanes 2**) and the other from +85 to +115 (**lanes 1**). The reverse transcriptase generated fragments were run, together with the G- and A-specific reactions of a fragment of known sequence, in an 8% sequencing gel.

Due to the higher amount of the C1-I versus C1 transcripts, we noticed an extra band two bases ahead of the major one which might suggest that the start of transcription is at position −2. Comparison with the C1 and C1-I genomic sequences indicated that the 5′ end points of the cCI5 and cCI49 cDNAs are at positions 100 and 117 respectively and their 3′ ends are at positions 1378 and 1392 respectively, extending 200 bp and 214 bp into the 3.8 kb sequence unique for the C1-I allele. The 142 bp insert in cCI49 corresponds to an intron of the C1 locus which is spliced in the C1-I transcript corresponding to cCI5. To verify whether clone cCI49 corresponded to one of the steady-state transcripts, Northern analysis of poly(A)$^+$ RNA from C1 (Figure 3, lanes 1 and 3) and C1-I (lanes 2 and 4) homozygous kernels was performed using two probes, a 750 bp *Eco*RI−*Xho*I fragment from wild-type cDNA clone cLC28 (Paz-Ares *et al.*, 1987) and an intron-specific probe (from +445 to +510, Figure 2). The first probe lit up two transcripts in the C1-I poly(A)$^+$ RNA, a major 1.3 kb and a minor 1.45 kb transcript (Figure 3, lane 2), while the 1.2 and 1.5 kb band are seen in wild-type RNA (lane 1). Hybridization with the intron-specific probe revealed only the larger 1.45 kb transcript (Figure 3, lane 4), the minor component in C1-I RNA. A larger component, seen in both the C1 and the C1-I poly(A)$^+$ RNAs hybridized with the intron-specific probe, very likely corresponds to the 28S RNA. The non-stringent hybridization conditions used in the experiment (because of the small size of the probe) probably resulted in its non-specific detection. From this experiment, it can be concluded that cCI49 corresponds to a minor transcript originating in the C1-I allele in which the second intron is not spliced out.

C1 is expressed in heterozygous (C1/C1-I) kernels

One possible explanation for the dominant effect of C1-I over the wild-type allele could be that C1-I prevents the expression of C1. The 3′ end of the wild-type and mutant C1 alleles differ considerably (Figure 2) due to the 3.8 kb substitution in C1-I. It should therefore be possible to distinguish between transcripts of each allele in S1 mapping experiments, thus allowing the study of the expression of both alleles in the heterozygous condition. Poly(A)$^+$ RNA isolated from both homozygous C1-I and C1 lines as well as from heterozygous (C1/C1/C1-I) kernels was annealed with an *Xho*I−*Sna*BI (from +1021 to −1576) fragment from the C1 locus (Paz-Ares *et al.*, 1987). A diagrammatic representation of this experiment is given in Figure 5. As shown in Figure 5, specific transcripts for either the C1 or the C1-I allele are present in the respective homozygotes, while the heterozygote (C1/C1/C1-I) contains both transcripts in roughly similar amounts to those expected for the relative contribution of each allele. Hence C1-I apparently does not prevent expression of the wild-type C1 allele.

Features of the C1-I proteins

The amino acid sequence of the putative proteins encoded by the C1-I allele was deduced from clones cCI5 and cCI49. The putative protein (Figure 6) corresponding to the major transcript (represented by cCI5) would be 252 amino acids

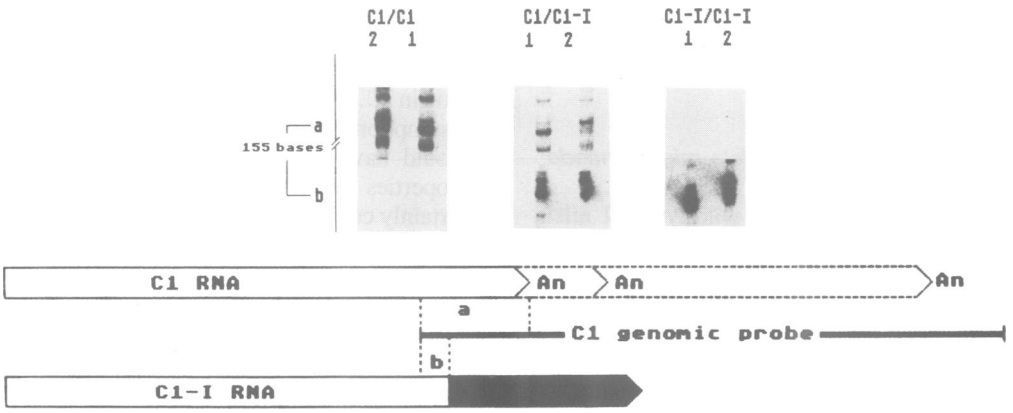

Fig. 5. Analysis of C1 and C1-I transcription in the heterozygote by S1 mapping experiments. An *Xho*I−*Sna*BI fragment (from +1021 to +1576, Paz-Ares *et al.*, 1987) of C1 genomic sequences was hybridized with poly(A)$^+$ RNA from developing kernels homozygous for C1 (10 μg), homozygous for C1-I (5 μg) or heterozygous (C1, C1, C1-I) for both alleles (15 μg). After treatment with S1 nuclease for 30 (**lanes 1**) and 10 (**lanes 2**) min, the protected fragments were run together with the A- and G-specific reactions of a fragment of known sequence, in a 5% sequencing gel (not shown). The lower part of the figure represents a schematic drawing of the S1 protection experiment. The continuous lines in the C1 RNA represent the transcript of 1.2 kb (corresponding to clone cLC6, Paz-Ares *et al.*, 1987). The broken lines indicate that there is heterogeneity at the 3′ end of the C1RNAs (Paz-Ares *et al.*, 1987; A.Hudson, unpublished observations). Due to this size heterogeneity, C1 RNAs protect several fragments of the C1 genomic probe as observed in the upper part of the figure. The size of the fragment protected by the 1.2 kb long C1 RNA is 215 bases (a). The black part of the C1-I RNA indicates the divergence of C1-I transcript at the 3′ with respect to wild-type DNA sequences (see text). Due to this divergence, the C1-I RNAs only protects a small portion of the C1 genomic probe (b = 60 bases).

long (27 kd). The most remarkable difference with respect to its wild-type counterpart is its shorter size, 252 instead 273 amino acids. This is due to the 8 bp insertion mutation at position 993, leading to premature termination of the open reading frame at position 1005 (Figure 2), of the C1-I transcript.

In addition, the other sequence differences between mutant and wild-type proteins are mostly scattered over the carboxy-terminal half. Only one conservative replacement (D for E) at amino acid position 101 was found within the first 114 amino acids, corresponding to the basic domain with homology to the myb proto-oncogene proteins and probably representing the putative DNA-binding domain of the wild-type protein (Paz-Ares *et al.*, 1987). The changes observed in the C1-I protein result in a loss of negative charge in two segments of the protein (Figure 6, lower part).

The second protein encoded by the C1-I allele, corresponding to the minor, incompletely processed 1.45 kb transcript, is very small (108 amino acids) due to the presence of an in-frame termination codon in the second unspliced intron. Only the first 88 amino acids are conserved with respect to the wild-type protein (not shown).

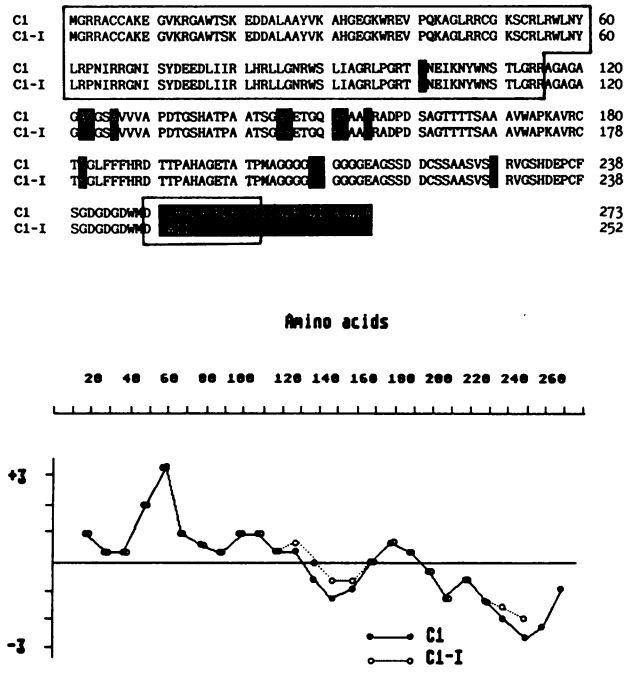

Fig. 6. Comparison between the amino acid sequence of the putative C1 and the 252 amino acid C1-I proteins. Amino acids shown in black represent differences between both proteins. The boxed-in area, from positions 1 to 115, shows the basic domain which shares homology to the myb proto-oncogene products (Paz-Ares *et al.*, 1987) and the boxed-in region, from positions 248 to 260, represents the amphipathic α-helix structured acidic domain present in the wild-type protein (see text). The lower portion of the figure shows the charge distribution in the putative C1 and C1-I proteins. The average net charges over successive 30 amino acids, measured at 10 amino acid intervals are given.

Discussion

Structure−expression relationships in the C1-I allele

We have shown that the C1-I allele contains two large and several minor DNA sequence alterations with respect to the wild-type C1 allele. The effect of each of these alterations on the C1-I transcription and translation products could be different. The large DNA rearrangement in the 5′ part of C1-I is 2 kb upstream of the start of transcription (see Figure 1). Hence this insert might be irrelevant for C1-I expression, especially since the start of transcription is the same in both C1-I and C1 alleles. The 3.8 kb sequence in C1-I, which replaces 370 bp of the corresponding wild-type sequences, is partialy transcribed and the first 200−230 bp of this sequence is present in both C1-I transcripts. However, this alteration does not affect the coding regions of either transcript, because the translation termination codons are upstream of the first nucleotide of the substitution sequence. Therefore, this DNA rearrangement would not play a role in post-translational events concerning the C1-I allele, but it could affect transcription, splicing of increased stability of C1-I RNAs.

The small sequence alterations observed in C1-I with respect to C1 could be more significant, since they alter the protein structure (see below).

Mechanism of inhibition: the C1-I allele might encode a transcriptional repressor

Differences between the C1-I and the wild-type C1 allele occur at the DNA, RNA and protein levels. Since C1-I does not affect the expression of the C1 allele in the heterozygous condition (Figure 5), the negative interference of C1-I with C1 might be conditioned by the protein structure of C1-I itself.

Based on genetic, biochemical and molecular data, we previously proposed that the C1 locus encodes a transcriptional activator. (Paz-Ares *et al.*, 1987). This hypothesis is further supported by our observations on the DNA-binding properties of the C1 protein (Wienand *et al.*, 1989). Structural data on transcriptional activators mainly in yeast (Struhl, 1987a), led to our proposal that the C1 protein has a basic DNA-binding domain at its N terminus with 40% sequence similarity to the DNA-binding domain of myb proto-oncogene proteins (Klempnauer and Sippel, 1987; Biedenkamp *et al.*, 1988). This assumption is supported by the fact that the DNA-binding domains of Jun and GCN4 proteins, which have the same degree of sequence similarity as is observed between C1 and myb (Vogt *et al.*, 1987), are functionally exchangeable (Struhl, 1987b).

On the other hand, the transcription activating domain of these regulatory proteins was previously thought merely to require acidity (Struhl, 1987a). However, present evidence indicates that there are some structural requirements in addition to acidity for a domain to function as an activator. Most activator domains are arranged in an amphipathic α-helix (Giniger and Ptashne, 1987; Ptashne, 1988). In order to define the possible activator domain(s) of the C1 protein, a computer-assisted search of ampipathic α-helical structures was conducted among the acidic regions of this protein. A region of this protein with high probability of being amphipathic if in an α-helical conformation (α-helix hydrophobic moment >0.4) was detected between amino acids 246 and 260, using the PEPPLOT program (Gribskov *et al.*, 1984). This region overlaps fairly well with that predicted to be in α-helix configuration (amino acids 248–263) by the two methods used (Chou and Fasman, 1974; Garnier *et al.*, 1978). These results strongly suggest that the 273 amino acid C1 protein indeed contains an activator domain at its carboxy terminus, the core of which is limited by amino acids 248–260. Other types of activator domains (Courey and Tjian, 1988) are not observed in the C1 protein.

In the case of the C1-I-encoded functions we consider only the large 252 amino acid protein, since the smaller (108 amino acid) protein is encoded by a very minor transcript and could very well be a product of an insufficiently spliced minor RNA product. The suggestion that this species of RNA exists in C1-I because the altered 3′ sequences have an influence on splicing of introns is attractive but requires further substantiation.

The 252 amino acid C1-I protein might lack the putative activator domain. It is noteworthy that the putative DNA-binding domain is practically identical in the wild-type and in this mutant protein, whereas in the rest of the C1-I protein there are several alterations. The only observed change in the first domain is a conservative (D for E) exchange and hence no drastic changes in its DNA-binding properties are expected. Preliminary DNA-binding studies with C1-I protein (U.Wienand, unpublished observations) support this assumption. Changes outside the DNA-binding domain should have little or no influence on the DNA-binding properties of the protein (Ptashne, 1988), although they certainly could be relevant in connection with the proposed activator function of the second domain. At the carboxy terminus of C1-I, only one out of the 13 amino acids of the core of the putative activator domain of the wild-type protein is retained in the mutant. This remaining amino acid is unlikely to contribute to the α-helix as judged by the two above-mentioned prediction methods. This then suggests that the putative activator function would be missing in the mutant. A second acidic domain is present both in the wild-type and in the mutant (see Figure 6), but it is less acidic in the mutant. Therefore, if it also has some role in activation, this domain would be impaired in the mutant. It was shown in the Gal4 system of yeast that a transcriptional activator depleted of the activator domain is converted into a transcriptional repressor (Hope *et al.*, 1988). Through this analogy, it appears likely that C1-I encodes a transcriptional repressor.

The dominance of the C1-I allele is best understood in relation to the situation in the yeast Gal4 system (Keegan *et al.*, 1986). In maize, the C1 activator protein binds to several sites near the genes (Wienand *et al.*, 1989) involved in anthocyanin biosynthesis, i.e. the A1 or the C2 locus. Through its acidic domains, C1 is believed to interact with the DNA–RNA polymerase complex as a prerequisite for activation of transcription. The C1-I protein lacking the activating acidic domains none the less binds to the above DNA sites (U.Wienand, unpublished observations) and if these sites were occupied by C1-I protein this would prevent activation of the C1 controlled genes even in the presence of one or two doses of wild-type protein.

However, if higher amounts of C1 protein are present in the cell, as with the higher doses of C1, the inhibiting effect of C1-I is overcome (Schwartz, 1956).

An alternative explanation to account for the dominance of the C1-I alleles is a mixed dimer model (Peterson and Leleji, 1974) like that given for the bacterial lack i^{-d} alleles. Because of the fact that C1-I protein still binds to the same DNA sites as C1 protein, we consider the mixed dimer model unlikely.

In order to determine more precisely which of the protein domains are responsible for DNA binding and activation, as well as to study the dominance of the C1-I mutant protein, several experiments are currently under way.

The excision of a transposable element (TE) from genes predominantly results in altered sequences of the flanking DNA. Therefore if a TE is integrated into an exon, an altered protein will frequently be formed upon excision of the element. Since numerous TE-containing C1 mutations are available, isolation of revertants with altered phenotypes will allow us to correlate small changes in protein structure to changes in function.

In addition, *in vitro* modification of C1 followed by microprojectile injection of the altered product into aleurone cells to monitor its effect on pigment biosynthesis will be attempted.

We expect that these studies will provide details of C1 action and distinguish between the models proposed.

Materials and methods

Plant strains

Maize line C was described previously (Shepherd *et al.*, 1982), while a line containing the C-I allele was obtained from E.H.Coe. Both lines were used as a source of the wild-type C1 and mutant C1-I alleles respectively.

Heteroduplex analysis

Heteroduplex analyses were performed essentially according to Davis *et al.* (1971) and Davis and Hyman (1971).

Standard molecular procedures

All methods, except those indicated here, were performed either as published earlier (Schwarz-Sommer *et al.*, 1984, 1985; Paz-Ares *et al.*, 1986, 1987), or as described in Maniaits *et al.* (1982). Vectors for cloning were λEMBL4, (Frischauf *et al.*, 1983) for genomic DNA, λNM1149 for cDNA (Murray, 1983) and pUC18/19 for subcloning of DNA (Yanisch-Perron *et al.*, 1985).

Conditions for hybridization and washing of Northern blots were the same as for Southern blot hybridization reported in Paz-Ares *et al.* (1986) when using the *Eco*RI−*Xho*I fragment of cDNA clone cLC28 (from +12 to +1020 in Paz-Ares *et al.*, 1987). However, when the second intron-specific DNA fragment *Eco*RI−*Bss*HII (from +445 to +510, Figure 2) was used as a probe, the hybridization temperture was decreased to 60°C. Washing was performed at 60°C in 2 × SSPE, 1% SDS. Radioactive labelling of probes was performed using the general primer pd(N)6 (Feinberg and Vogelstein, 1984).

Primer extension experiments

The conditions of hybridization, removal of excess primer and reverse transcriptase reactions were as in Paz-Ares *et al.* (1987), using 5 μg of poly(A)$^+$ RNA from C1 and C1-I developing kernels. Two 30mer synthetic oligonucleotides, one from position +85 to +115 and the other from position +241 to +271 (Figure 2), were used as primers in independent experiments.

S1 nuclease protection experiments

A 555 bp *Xho*I−*Sna*BI fragment from the C1 genomic sequence (position +1021 to +1576, Paz-Ares *et al.*, 1987) was 3′ end labelled at the *Xho*I site with Klenow polymerase and used as a probe. Conditions for annealing and S1 nuclease digestion were as in Paz-Ares *et al.* (1987), except that after annealing, excess probe was removed by affinity chromatography on oligo(dT)−cellulose.

Sequence analysis

Most of the genomic DNA sequence was determined by the dideoxy chain termination method (Sanger *et al.*, 1977) after subcloning in the M13 vector mp18 and mp19 (Norrander *et al.*, 1983). For deletion subcloning, the method of Henikoff (1984) was followed. The cDNAs and minor parts of the genomic DNAs were sequenced by the chemical method (Maxam and Gilbert, 1980). The sequence given in Figure 2 was read at least once from each strand.

Computer program for protein structure analysis

The amphipathic regions present in C1 and in C1-I proteins were predicted by calculating the α-helix hydrophobic moment (Eisenberg, 1984) with the PEPPLOT program (Gribskov *et al.*, 1984) from the UWGCG program library (Devereux *et al.*, 1984). Secondary structure prediction of C1 and C1-I proteins was performed according to the methods of Chou and Fasman (1974) and Garnier *et al.* (1978), with the UWGCG program library.

Acknowledgements

We thank Margot Küsters for DNA heteroduplex analysis in the electron microscope. We wish to thank Brian Scheffler and Monika Frey for help in preparing the figures and Udo Wienand in preparing the manuscript. We thank Alfons Gierl, Sarah Grant, F.Garcia Olemdo and Alicia de la Pena for critical reading of the manuscript.

References

Biedenkamp,H., Borgenmeyer,U., Sippel,A. and Klempnauer,K.H. (1988) *Nature*, **335**, 835−837.
Chou,P.Y. and Fasman,G.D.(1974) *Biochemistry*, **53**, 595−623.
Coe,E.H.,Jr (1962) *Genetics*, **47**, 779−783.
Coe,E.H., Jr and Neuffer,M.G. (1977) In Sprage,G.F. (ed.), *Corn and Corn Improvement*. American Society of Agronomy, Madison, WI, pp. 111−123.
Cone,K., Burr,F.A. and Burr,B. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9631−9635.
Courey,A.J. and Tjian,R. (1988) *Cell*, **55**, 887−898.
Davis,R.W. and Hyman,R.W. (1971) *J. Mol. Biol.*, **62**, 287−301.
Davis,R.W., Simon,M. and Davidson,N. (1971) *Methods Enzymol.*, **21**, 413−428.
Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387−295.
Dixon,R.A. (1986) *Biol. Rev.*, **61**, 239−291.
Dooner,H.K. (1983) *Mol. Gen. Genet.*, **189**, 136−141.
Dooner,H.K. and Nelson,O.E. (1979) *Genetics*, **91**, 309−315.
Eisenberg,D. (1984) *Annu. Rev. Biochem.*, **53**, 595−623.
Emerson,R.A., Beadle,G.W. and Fraser,A.C. (1935) Cornell Univ. Agric. Exp. Station Memo 180.
Fedoroff,N.V., Furtek,D.B. and Nelson,O.N.,Jr (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3825−3829.
Feinberg,A.P. and Vogelstein,B. (1984) *Anal. Biochem.*, **137**, 266−267.
Frischauf,A.M., Lehrach,H., Poutska,A. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827−842.
Garnier,J., Osguthorpe,D. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97−120.
Giniger,E. and Ptashne,M. (1987) *Nature*, **330**, 670−672.
Gribskov,M., Devereux,J. and Burgess,R.R. (1984) *Nucleic Acids Res.*, **12**, 539−549.
Harborn,J.B. (1986) In Harborne,J.B. and Mabry,T.J. (eds), *Plant Flavonoids in Biology and Medicine*. Alan R.Liss, New York, pp. 15−24.
Henikoff,S. (1984) *Gene*, **28**, 351−359.
Hope,I.A., Mahaderan,S. and Struhl,K. (1988) *Nature*, **333**, 635−640.
Keegan,L., Gill,G. and Ptashne,M. (1986) *Science*, **231**, 699−704.
Klempnauer,K.M. and Sippel,A.P. (1987) *EMBO J.*, **6**, 2719−2725.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Labortory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499−560.
Murray,N.E. (1983) In Hendreix,R.W., Roberts,J.W., Stahl,F.W. and Weisberg,R.A. (eds), *Lambda II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 395−432.
Norrander,J., Tempe,J. and Messing,J. (1983) *Gene*, **26**, 101−106.
O'Reilly,C., Shepherd,N.S., Pereira,A., Schwartz-Sommer,Zs., Gierl,A., Bertram,I., Robertson,D.S., Peterson,P.A. and Saedler,H. (1985) *EMBO J.*, **4**, 877−882.
Paz-Ares,J., Wienand,U., Peterson,P.A. and Saedler,H. (1986) *EMBO J.*, **5**, 829−833.
Paz-Ares,J., Ghosal,D., Wienand,U., Peterson,P.A. and Saedler,H. (1987) *EMBO J.*, **6**, 3553−3558.
Peterson,P.A. and Leleji,O. (1974) *Can. J. Genet. Cytol.*, **16**, 389−398.
Ptashne,M. (1988) *Nature*, **335**, 683−689.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463−5464.
Schwartz,D. (1956) *Maize Gen. Coop. Newslett.*, **30**, 106.
Schwarz-Sommer,Zs., Gierl,A., Cuypers,H., Peterson,P.A. and Saedler,H. (1984) *EMBO J.*, **3**, 1021−1028.
Schwarz-Sommer,Zs., Gierl,A., Klösgen,R.B., Wienand,U., Peterson,P.A. and Saedler,H. (1985) *EMBO J.*, **4**, 591−597.
Shepherd,N.S., Schwarz-Sommer,Zs., Wienand,U., Sommer,H., Deumling,B., Peterson,P.A. and Saedler,H. (1982) *Mol. Gen. Genet.*, **188**, 266−271.
Struhl,K. (1987a) *Cell*, **49**, 295−297.
Struhl,K. (1987b) *Cell*, **50**, 849−846.
Theres,N., Scheele,T. and Starlinger,P. (1987) *Mol. Gen. Gent.*, **209**, 193−197.
Vogt,P.R., Bos,T.J. and Doolittle,R.R. (1987) *Proc. Natl. Acad. Sci. USA*, **89**, 3316−3319.
Wienand,U., Weydeman,U., Niesbach-Klösgen,U., Peterson,P.A. and Saedler,H. (1986) *Mol. Gen. Genet.*, **203**, 202−207.
Wienand,U., Paz-Ares,J., Scheffler,B. and Saedler,H. (1989) In Lamb,C. and Beachy,R. (eds), *Plant Gene Transfer*, UCLA Symp. Mol. Cell. Biol., 129. Alan R.Liss, New York, in press.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103−119.

Received on September 26, 1989; revised on November 9, 1989