The EMBO Journal vol.9 no.10 pp.3051 – 3057, 1990

The En/Spm transposable element of *Zea mays* contains splice sites at the termini generating a novel intron from a dSpm element in the *A2* gene

Adriane Menssen, Susanne Höhmann, William Martin[1], Patrick S.Schnable[2], Peter A.Peterson[3], Heinz Saedler and Alfons Gierl

Max-Planck-Institut für Züchtungsforschung, Molekulare Pflanzengenetik, Carl-von-Linné-Weg 10, 5000 Köln 30, FRG

[2]Iowa State University, Genetic Department, Ames, IA 50011 and [3]Iowa State University, Agronomy Department, Ames, IA 50011, USA

[1]Present address: Institut für Genetik, Technische Universität Braunschweig, Postfach 3329, 3300 Braunschweig, FRG

Communicated by H.Saedler

The *A2* locus of *Zea mays*, identified as one of the genes affecting anthocyanin biosynthesis, was cloned using the transposable elements rcy and dSpm as gene tags. The *A2* gene encodes a putative protein of 395 amino acids and is devoid of introns. Two *a2-m1* alleles, containing dSpm insertions of different sizes, were characterized. The dSpm element from the *original state* allele has perfect termini and undergoes frequent transposition. The element from the *class II state* allele is no longer competent to transpose. It has retained the 13 bp terminal inverted repeat but has lost all subterminal sites at the 5′ end, which are recognized by tnpA protein, the most abundant product of the En/Spm transposable element system. The relatively high *A2* gene expression of one *a2-m1* allele is due to removal of almost all dSpm sequences by splicing. The slightly altered A2 enzyme is still functional as shown by complementation of an *a2* mutant with the corresponding cDNA. The 5′ and 3′ splice sites are constituted by the termini of the dSpm element; it therefore represents a novel intron of the *A2* gene.

Key words: A2 gene cloning/*a2-m1* (*class II state*)/ anthocyanin pathway/En/Spm excision/RNA splicing

Introduction

The Enhancer (En) or Suppressor–mutator (Spm) transposable element system of *Zea mays* comprises autonomous and non-autonomous elements. The autonomous elements En (Peterson, 1953) and Spm (McClintock, 1954) were independently identified genetically. En and Spm were later shown to be virtually identical genetically (Peterson, 1965) as well as molecularly (Pereira *et al.*, 1986; Masson *et al.*, 1987), therefore, the autonomous element is called En/Spm. En/Spm encodes functions required for transposition. The non-autonomous elements represent deletion derivatives of En/Spm and have been termed defective (dSpm) elements (Masson *et al.*, 1987) or Inhibitor elements (Peterson, 1953). These elements cannot promote their own transposition, rather they transpose only in the presence of En/Spm. The term dSpm will be used here to designate the non-autonomous elements.

The interaction of functions encoded *in trans* by En/Spm with dSpm elements has led to the genetic characterization of En/Spm-encoded functions (McClintock, 1961) and provided insights into the substrate requirements for transposition. In several cases, further deletion derivatives of a given dSpm have been isolated, which differ in the somatic and germinal excision frequencies in the presence of En/Spm (for review, see Fedoroff, 1989). McClintock (1968) has termed these derivatives 'changes in state'. Molecular analysis of such 'states' of a dSpm element has revealed that the 13 bp terminal inverted repeat (TIR) and subterminal repetitions of a 12 bp motif are *cis*-acting determinants for excision (for review, see Gierl *et al.*, 1989), which is mediated by the 'mutator' function of En/Spm.

The 'suppressor' function of En/Spm can only be observed with certain alleles bearing dSpm insertions. In these 'suppressible' alleles, in the absence of En/Spm, the dSpm insertion reduces but does not abolish expression of the gene into which it is inserted. Suppressible alleles of the *A1* gene (*a1-m1*; Schwarz-Sommer *et al.*, 1985b) and of the *bronze1* gene (*bz-m13*; Schiefelbein *et al.*, 1988) have been analyzed molecularly. The dSpm insertion is located within the transcribed regions of these genes and in the case of *bz-m13* it was shown that almost all of the dSpm sequence is efficiently removed by splicing of the primary transcript, thus restoring the function of the *bz-m13* gene product. In the presence of the autonomous En/Spm element, however, gene expression of suppressible alleles is fully repressed (suppressed). According to the model of Gierl *et al.* (Gierl *et al.*, 1985, 1988), the suppressor function resembles a negative regulatory circuit in which an En/Spm-encoded protein acts as a repressor. The repressor binds to a defined *cis*-element located within the dSpm element. The bound protein is thought to hinder progression of RNA polymerase sterically, resulting in prematurely terminated transcripts. The suppressor function of En/Spm has been attributed to tnpA protein by reconstruction of the suppressor system in transgenic tobacco (Grant *et al.*, 1990). tnpA is the most abundant En/Spm-encoded product that has been shown to bind to the 12 bp motif which is present in the subterminal regions of En/Spm (Gierl *et al.*, 1988).

The *A2* gene encodes an enzyme involved in the synthesis of the anthocyanin pigments in the aleurone layer of the maize kernel and other parts of the plant. The enzyme function, however, is as yet unknown. McClintock (1957, 1958) has described two alleles of the *A2* gene, containing dSpm insertions, which seem to be intriguing in view of element excision and suppression. The *a2-m1* (*original state*) allele represents the original isolate, which conditions a very low pigmentation in the absence of En/Spm (see Figure 4). In the presence of En/Spm, this residual gene expression is abolished; however, deeply pigmented sectors appear on the colorless background due to frequent excision of the

© Oxford University Press

dSpm element and restoration of the structure of the *A2* gene. The *a2-m1* (*class II state*) allele, a derivative of the original isolate, conditions deeply pigmented kernels in the absence of En/Spm, almost indistinguishable from wild-type. In the presence of En/Spm, gene expression is abolished, resulting in almost homogeneously colorless kernels. The dSpm element at this allele is no longer competent to undergo excision (McClintock, 1958). The infrequent occurrence of small pigmented spots indicates inactivity or loss of the autonomous En/Spm element and hence lack of suppressor function.

In order to analyze the structure of dSpm insertions, these two *a2-m1* alleles were cloned using the dSpm elements and rcy, another transposable element insertion, as gene tags. The *A2* transcription unit was characterized. Comparison with the genomic clones shows that *A2* is a gene devoid of introns. The dSpm element of the *a2-m1* (*original state*) allele has perfect termini; in contrast, the dSpm of the *class II state* allele lacks all of the tnpA binding motifs at one end, but retains the 13 bp terminal inverted repeat. The relatively high *A2* expression level seen in this allele in the absence of En/Spm is due to the fact that this dSpm element efficiently mimics an intron and almost all of the element's sequence is removed by processing of the primary transcript. The 5′ and the 3′ splice sites are provided by the element's termini. As a result of mRNA processing, seven new amino acids are inserted into the *A2* protein. The functionality of this modified protein was directly confirmed by transient expression studies. The evolutionary implications of these findings are discussed.

Results

Molecular cloning of A2 alleles

The rcy element is a non-autonomous member of the Cy transposable element system of *Z.mays* (Schnable and Peterson, 1988). One particular rcy element has been cloned and sequenced (Schnable *et al.*, 1989) and was termed rcy:Mu7. Only 2−6 copies of this element are present in the maize genome, so rcy is very suitable as a gene tag. For cloning of the *A2* locus we used a rcy-induced *a2* mutant and the dSpm insertions at the *a2-m1* alleles as tags.

The mutant *a2-m668291* has been shown genetically to contain a rcy element at the *A2* locus (Peterson, 1988). DNA from plants segregating for this mutable allele was digested with *Bam*HI and analyzed by Southern blots. Using rcy-specific sequences as a probe, a 3.4 kb *Bam*HI fragment was detected that co-segregated with the mutable phenotype (data not shown). The 3.4 kb fragment was cloned from genomic DNA and partially sequenced. It contains a 2.2 kb rcy element (Figure 1) that is similar if not identical to the rcy:Mu7 element which has been described previously (Schnable *et al.*, 1989).

If the rcy element is inserted at the *A2* locus one would expect that sequences flanking the insertion could be used to identify the dSpm insertions at the *a2-m1* alleles. Therefore a fragment adjacent to the rcy element was used for the identification of genomic clones of the *a2-m1* alleles. In the case of *a2-m1* (*original state*), a clone was identified that contained a 2241 bp dSpm element (Figure 1). It represents an internal deletion derivative of the autonomous En/Spm element. The deletion extends from position 861 to position 6906 (positions refer to the 8287 bp long En-1 element, Pereira *et al.*, 1986) of En/Spm. Identical dSpm elements were isolated previously from the *waxy* gene (Gierl *et al.*, 1985; as modified by the elimination of a thymine residue in Pereira *et al.*, 1986), the *A1* gene (Schwarz-Sommer *et al.*, 1985b) and the *Bronze-1* gene (Schiefelbein *et al.*, 1988). In the case of *a2-m1* (*class II state*), a 1341 bp dSpm

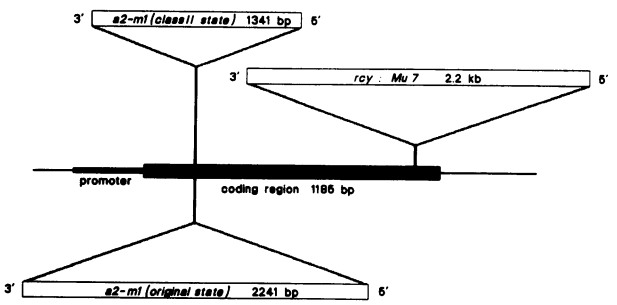


Fig. 1. Position of transposable element insertions within the *A2* gene. The *A2* coding region and the promoter region are represented by black bars. The insertion sites of the dSpm elements and the rcy element are marked by vertical lines. The orientations of the insertions with respect to the published sequences are indicated.

```
                                                                  -161
                                        TGCAGCAGTGCTGATCACCAAATTTTTTTCCCTTGGCACG
                                                                   -81
TGTTCGGGTGGTAGCTGGCTCCTCCATGCAATGCCTATGCCTGTCGTCGCGATCGCAACCACCAGTCAAGACGAATGGCA
                                                                    -1
GGCAGCTAAGTAGCTAACAACAACAGGCTTGTATTGTATGTACAGCAATATATATACATCTCGAGAAGCTCCTAGCAGAG
                                                                    80
ATGCATCCATCTCGACGACGATCCATCCATCTTAAAGCAAGCAGAGACATGGAGTCGTCGCCGCTGCTGCAGCTGCCGGC
                                                                   160
GGCACGCGTGGAGGCTCTGAGCCTCAGCGGCCTCTCCGCCATCCCGCCCGAGTACGTCCGCCCCGCCGACGAGCGAGCCG
                                                                   240
GCCTCGGGGATGCCTTCGACCTGGCGCGCACCCATGCCAACGACCACACCGCGCCGAGGATCCCCGTCGTCGACATCTCC
                                                                   320
CCGTTCCTCGACAGCAGCAGCCAGCAGCAGCAGCGGGACGAATGCGTGGAGGCCGTGCGTGCCGCCGCCGCCGACTGGGG
                                                                   400
CGTCATGCACATCGCCGGCCACGGCATCCCCGCCGAGCTCATGGACCGCCTGCGCGCCGCGAGGAACCGCCTTCTTCGCCC
                                                                   480
TCCCCGTCCAGGACAAGGAGGCCTACGCCAACGACCCCGCCGCCGGCCGCCTGCAGGGCTACGGCAGCCGCCTCGCCACC
                                                                   560
AACACCTGCGGGCAGCGCGAGTGGGAGGACTACCTCTTCCACCTTGTGCACCCCGACGGGCTCGCCGACCACGCGCTCTG
                                                                   640
GCCTGCGTACCCGCCCGACTACATCGCCGCCACCCGCGACTTCGGCCGCCGAACGCGGGACCTGGCCTCCACGCTGCTCG
                                                                   720
CCATCCTCTCCATGGGCCTCCTTGGCACAGACCGTGGCGACGCGCTAGAGAAGGCGCTCACCACCACCACCACCAGGACA
                                                                   800
GCAGCTGACGACGACCTCCTCCTGCAGCTCAAGATCAACTACTACCCGAGGTGCCCGCAGCCGGAGCTGGCCGTCGGCGT
                                                                   880
GGAGGCCCACACGGACGTCAGCGCCCTCTCCTTCATCCTCCACAACGGCGTGCCGGGCCTGCAGGTGCTCCACGGCGCCC
                                                                   960
GCTGGGTGACGGCGCGCCACGAGCCGGGCACCATCATCGTCCACGTGGGGGACGCCCTGGAGATCCTCAGCAACGGCCGC
                                                                  1040
TACACCAGCGTCCTCCACCGCGGCCTCGTCAACCGGGAGGCCGTGCGCATCTCCTGGGTCGTCTTCTGCGAGCCGCCACC
                                                                  1120
AGACTCCGTGCTGCTGCACCCGCTGCCGGAGCTTGTCACGGAAGGCCACCCCGCAAGGTTCACGCCGCGCACATTCAAGC
                                                                  1200
AGCACCTGGATCGCAAGCTTTTCAAGAAGAAACAGCAGCACAAAGCAAAAGCAGAGGAAGAGGATGGCGGCAATGGTGAC
                                                                  1280
CACCACCGCCACGAGCCGCCGCCGCAGACCAACTGATGGGCTGCACATGTCTTTCCATCCGCCCACGCATATCTTCTCTC
                                                                  1360
GCGAAATTAATAAGGATCCATCAGCATTTTCCATATATTTTATATTAGTTTCATGCTCCTACGTTACTACGAGAAAAAAAA
                                                                  1440
TAGTATCTATGATATATAGAATCCATGTGTTAAAATATCACTGTAACACTAATATTATATATGTTGTTACGAAATCAATA
TAATAAAATAAATTGATATACC
```

Fig. 2. DNA sequence of the *A2* gene. Transcription initiation occurs at position 1. The putative TATA-box, the ATG translation start and the TGA stop codon are underlined, as are the target site duplications of the transposable element insertions. The two poly(A) addition sites are indicated.

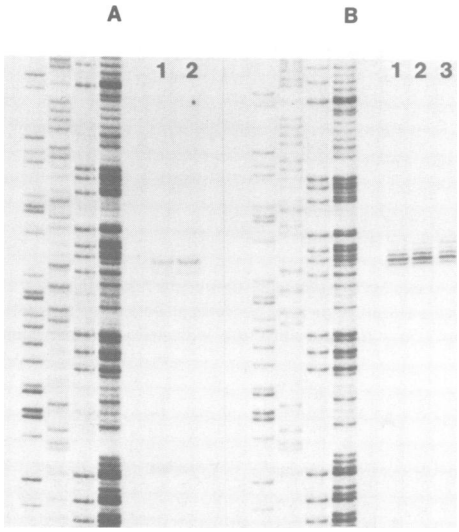


Fig. 3. Determination of the *A2* transcription start site. The primer extension experiment (**A**) was done with *A2* wild-type poly(A) RNA (lane 1) and with poly(A) RNA from *a2-m1* (*class II state*) material (lane 2). The S1 nuclease experiment (**B**) was done with *a2-m1* (*class II state*) poly(A) RNA, incubation with S1 nuclease was for 60, 40 and 20 min, respectively (lanes 1–3). The sequence ladder (G, A, C, T+C) corresponds to the strand complementary to the mRNA.

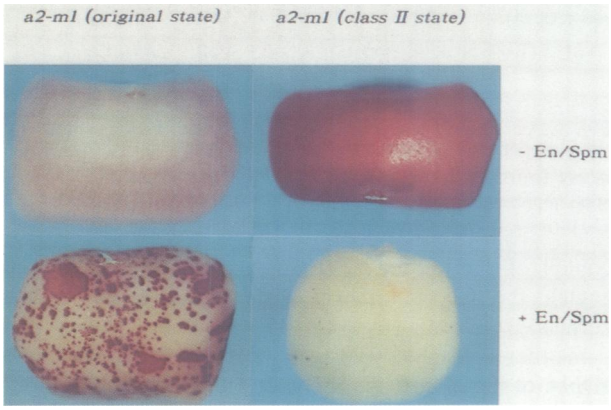


Fig. 4. Phenotypes of the *a2-m1* alleles. Top photographs show the phenotypes of the two alleles in the absence of the autonomous En/Spm element, bottom photographs in the presence of En/Spm, respectively.

element was identified that represents an extended deletion derivative of the *a2-m1* (*original state*) dSpm element (Figure 1). In this case, the deletion includes positions 35–6980. The genomic *A2* DNA sequence was determined from the *a2-m1* (class II state) allele (Figure 2).

Characterization of the A2 transcription unit

In order to determine the *A2* transcription unit, cDNA clones from a wild-type plant were analyzed. Several almost full size clones were obtained. They appeared to be completely colinear with the sequence deduced from the genomic clones (except for the transposable element insertions), indicating the absence of introns in the *A2* gene (Figure 2). The DNA sequence of the wild-type cDNA is identical to the corresponding genomic sequence of the *a2-m1* allele. Two different poly(A) addition sites were detected by the cDNA analysis (Figure 2). The transcription start was determined by S1 nuclease and primer extension experiments (Figure 3). The

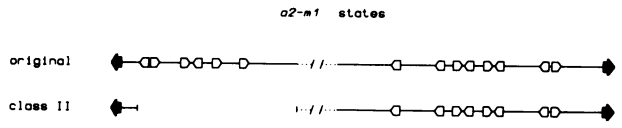


Fig. 5. Schematic representation of the termini of the dSpm elements. 200 bp from the 5′ terminus and 300 bp from the 3′ terminus are indicated. The filled arrows represent the 13 bp terminal inverted repeats, the 12 bp tnpA binding motifs are indicated by open arrows.

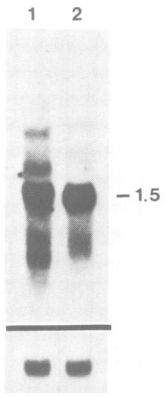


Fig. 6. Northern blot analysis of *A2* transcripts. 5 g of poly(A) RNA were analyzed. Lane 1: *a2-m1* (*class II state*); lane 2: *A2* wild-type. The size of the most abundant transcript is indicated in kb. The upper bands (lane 1) correspond to sizes of 2.8 and 1.8 kb, respectively. A control hybridization with a probe specific for maize NAD-dependent GAPDH proved that almost identical amounts of poly(A) RNA were compared (bottom line).

mRNAs of wild-type and of the *a2-m1* (*class II state*) allele initiate at the same position. The sequence context of the AUG translation start codon (Figure 2) agrees well with the eukaryotic consensus sequence (Kozak, 1987). The open reading frame of the *A2* gene codes for a putative protein of 395 amino acids (Figure 2). The *A2* coding region is relatively GC rich (69.8%), showing a strong codon bias for G or C in the third codon position (frequency 90.9%). This distinctiveness has also been found for other genes of monocotyledonous plants and in particular for genes of the anthocyanin pathway (Niesbach-Klösgen *et al.*, 1987).

The dSpm and the rcy insertions are within the coding region of *A2*. Comparison with the wild-type sequence showed that the dSpm insertion caused a 3 bp duplication of the target sequence and the rcy element a 9 bp duplication, respectively. This has been found to be typical for these element insertions (Schwarz-Sommer *et al.*, 1984; Schnable *et al.*, 1989).

Gene expression of the a2-m1 (class II state) allele

The two *a2-m1* alleles condition quite different phenotypes (McClintock, 1957 and 1958). In the absence of the autonomous En/Spm element, the *a2-m1* (original state) allele produces a uniform but low anthocyanin pigmentation of the maize kernel (Figure 4). In the presence of En/Spm, this residual coloration is suppressed and colored sectors frequently appear on a colorless background (Figure 4). The colored sectors represent somatic excision events of the residing element, restoring wild-type pigmentation in these cells. In contrast, the *a2-m1* (*class II state*) allele produces deep and uniform pigmentation in the absence of En/Spm; this pigmentation is almost as intense as that of wild-type (Figure 4). In the presence of the autonomous element,

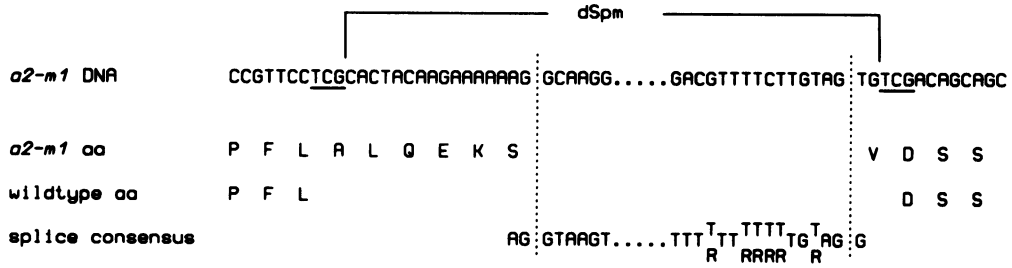


Fig. 7. Splice sites at the termini of dSpm from the *a2-m1* (*class II state*) allele. The DNA sequence around the dSpm insert and that of the termini of dSpm are shown. The target site duplication flanking the dSpm is underlined. The splice sites are indicated by vertical dotted lines. The amino acid sequences as deduced from the cDNA clones of the A2 wild-type gene and the *a2-m1* (*class II state*) allele are indicated, as is the plant splice consensus sequence (Brown, 1986).

however, this coloration is suppressed, resulting in an almost homogeneously colorless kernel (Figure 4). There is genetic evidence (McClintock, 1958) that the dSpm element at this allele is no longer competent to undergo excision in the presence of En/Spm. Therefore, it is thought that the infrequent appearance of small pigmented spots is due to inactivity or loss of En/Spm and hence lack of suppressor function (McClintock, 1958).

How do these genetic characterizations relate to the molecular findings? The difference in the excision behavior of both *a2-m1* elements should be a reflection of differences in the relevant *cis*-acting sequences necessary as substrates for excision. The dSpm element of the original state is an internal deletion derivative of the autonomous En/Spm element but has retained the highly structured ends that have been inferred as substrate sites for transposition (for review, see Gierl *et al.*, 1989). These regions consist of the 13 bp terminal inverted repeat (TIR) and subterminal repetitions of a 12 bp motif (Figure 5), to which tnpA protein binds (Gierl *et al.*, 1988), the most abundant En/Spm encoded function. The dSpm element of the *class II state* allele has retained the 13 bp TIR but lost all of the tnpA binding motifs at the 5′ end (Figure 5).

In order to understand the relatively high level of expression of the *a2-m1* (*class II state*) allele in the absence of En/Spm, poly(A) RNA of this material was compared with poly(A) RNA from wild-type kernels. The size of the wild-type *A2* mRNA is 1.5 kb (Figure 6). In the case of *a2-m1* (*class II state*), the most abundant band is of similar size. In addition, two minor bands with larger size are revealed (Figure 6).

To analyze this finding further, cDNA clones were isolated from *a2-m1* (*class II state*) poly(A) RNA. Two clones were sequenced. Both were found to be polyadenylated and almost full size with respect to the wild-type *A2* transcription unit. One clone corresponds in size to the major RNA band of 1.5 kb and the other to the upper (minor) band (Figure 6). The latter one contained all the 1341 bp of the residing dSpm element in addition to the *A2* sequences, demonstrating that transcription had proceeded throughout the entire dSpm element. The clone corresponding in size to the major 1.5 kb band contained only 18 bp of dSpm, indicating that most of the element sequences had been removed by splicing (Figure 7). The 3′ and the 5′ splice sites are both constituted by the element's termini. As a consequence of this RNA processing event, new amino acids are added at the site of element insertion to the *A2* protein sequence. Including the target site duplication of 3 bp, seven amino acids are added to the wild-type sequence (Figure 7). If this altered protein

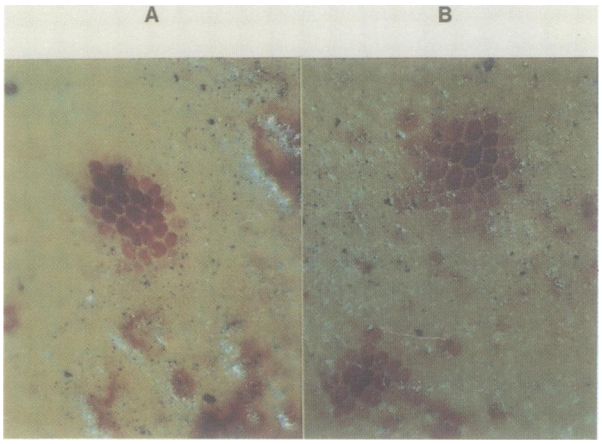


Fig. 8. Anthocyanin production in aleurone tissue lacking a functional *A2* gene following delivery of wild-type *A2* cDNA (**A**) and processed *a2-m1* (*class II state*) cDNA (**B**) by particle bombardment. The purple spots represent clusters of cells in which anthocyanin is produced.

could still catalyze the *A2* reaction, then this finding would explain the deep coloration of the *a2-m1* (*class II state*) kernels.

In order to test this, wild-type *A2* cDNA and the processed cDNA of the *a2-m1* (*class II state*) allele were cloned into the plant expression vector pRT100 (Töpfer *et al.*, 1987). Colorless kernels homozygous for *a2* (recessive) were transiently transformed with the two constructs by the particle bombardment method, developed by Klein *et al.* (1987). After two days, red spots appeared on the kernels transformed either with *A2* wild-type cDNA or with the processed cDNA from the *a2-m1* (*class II state*) allele (Figure 8). The intensity and the frequency of the colored spots was similar in both cases. This directly proves the enzymatic functionality of the protein product resulting from the processed transcript. The seven additional amino acids, remaining after processing, seem not to interfere with the *A2* enzymatic function, at least not at this level of detection.

Discussion

The A2 gene of Zea mays

The product of the *A2* gene is involved in the anthocyanin biosynthetic pathway of *Z.mays*. From inter-tissue complementation assays (Reddy and Coe, 1962), it was concluded that the *A2* enzyme catalyzes a step occurring after the *A1* enzyme reaction (dihydroflavonol 4-reductase) and before glycosylation by the *Bronze-1* gene product (antho-

cyanin 3-O-glycosyltransferase) (for review, see Heller and Forkmann, 1988). According to this concept, the *A2* gene product is involved in the formation of anthocyan, the first colored compound of the pathway. The formation of anthocyan from the respective flavan-3,4-*cis*-diol (the product of the *A1* reaction) involves a dehydration and an oxidation step. The exact reaction mechanism is unknown and no enzyme catalyzing this step in anthocyanin biosynthesis has been isolated to date. The reaction catalyzed by the *A2* gene product can now be analyzed biochemically by expressing the *A2* cDNA. This has been successfully done for the *A1* gene product of *Z.mays*, by *in vitro* transcription and translation of the *A1* cDNA (Rhode *et al.*, 1987). That the *A2* cDNA isolated encodes a functional *A2* protein was confirmed by transient transformation into the aleurone tissue (Figure 8).

The noteworthy features of the *A2* gene structure can be summarized as follows. The *A2* gene contains no intron and encodes a putative protein of 43.5 kd. A homology search revealed a 38% identity of a stretch of 100 amino acids at the carboxy-terminus of the *A2* protein with an adequate region in the pTOM13 cDNA of tomato, the mRNA of which is expressed during fruit ripening (Holdsworth *et al.*, 1987). More significantly, homology was also found with the *Antirrhinum majus Inc* gene which encodes the flavan-3-hydroxylase of the anthocyanin pathway. The products of *A2* and *Inc* share a domain of 110 amino acids with 37% identity (C.Martin and A.Prescott, personal communication). This finding could support the speculation of Heller and Forkmann (1988) that the *A2* product also represents a hydroxylase that however should oxidate the C2 position of the flavan-3,4-*cis*-diol, leading to formation of the aromatic structure in the anthocyanin product.

Pedigree analysis of mutable alleles caused by insertion of the 2.2 kb dSpm element

The dSpm element from the *a2-m1* (*original state*) allele is 2241 bp in size and represents an internal deletion derivative of the autonomous En/Spm element. The deletion extends from position 861 to position 6906 (positions refer to the 8287 bp long En-1 element, Pereira *et al.*, 1986) of En/Spm. Identical dSpm elements have been isolated earlier from other mutable alleles like *wx-m8* (Gierl *et al.*, 1985), *a1-m1* (Schwarz-Sommer *et al.*, 1985b), and *bz-m13* (Schiefelbein *et al.*, 1988). This raises the question of whether the dSpm observed at these four different locations arose independently, or whether a unique dSpm was formed at one location and subsequently transposed to the other genes. This can be answered, if one traces historically how these mutable alleles arose.

In the early fifties, McClintock first isolated the *a2-m1* (*original state*) allele, which originated from self-pollinating a plant showing variegation in chlorophyll content. The *A2* gene is located on chromosome five. Crossing this material by an *a1* recessive plant the *a1-m1* allele on chromosome three was recovered (McClintock, 1957). A similar cross with a recessive *waxy* line, gave rise to the *wx-m8* allele on chromosome nine (McClintock, 1961). *wx-m8* in turn was used to introduce the element into the *Bronze1* gene on the same chromosome, resulting in the *bz-m13* mutation (Nelson and Klein, 1984). This strongly suggests that dSpm elements are transposed throughout the maize genome in the presence of the autonomous element. It also illustrates how efficiently transposable elements can be used for gene tagging.

TnpA binding motifs are required for excision

Although the 2241 bp dSpm element is a deletion derivative of En/Spm, it has retained perfect ends which are required as substrates for excision. These regions consist of the 13 bp terminal inverted repeat (TIR) and 14 subterminal repetitions of a 12 bp motif (Figure 5), to which tnpA protein binds (Gierl *et al.*, 1988), the most abundant En/Spm encoded product, as well as a GC-rich region adjacent to the tnpA binding motifs at the left end (Masson *et al.*, 1987). Therefore it is not surprising that element insertions of this type are characterized by high somatic and germinal excision frequencies (Raboy *et al.*, 1989) and by transposition (see above). In contrast, the 1341 bp dSpm of the *a2-m1* (*class II state*) allele, is completely stable both somatically and germinally (McClintock, 1958). Sequence analysis of this dSpm has revealed that it is a deletion derivative of the *original state* element and that the deletion extends from position 35 to position 6980 of En-1. This dSpm element has retained the 13 bp TIR but has lost all of the tnpA binding motifs at the 5′ end and the GC-rich region, while the 3′ end of the element remains intact (Figure 5). Partial deletions of the 12 bp motifs have been correlated with decreased excision rates (for review, see Gierl *et al.*, 1989). Obviously, the entire deletion of these motifs at one end completely abolishes excision ability. This suggests that tnpA protein might be one of the components involved in excision.

On the other hand, the presence of the tnpA binding motifs at one end is sufficient for the suppressor effect, as indicated by the colorless phenotype of *a2-m1* (*class II state*) kernels in the presence of En/Spm. This effect results from bound tnpA protein that probably prevents progression of RNA polymerase and thus formation of the primary transcript. In a reconstruction of the suppressor system in transgenic tobacco, it was shown that one inverted repeat of the tnpA binding motif is sufficient for suppression to occur via the tnpA protein (Grant *et al.*, 1990).

The 1341 bp dSpm element forms a stable intron of the A2 gene

The analysis of the *a2-m1* alleles documents a sequence of events that lead to the formation of a new and stable intron in the *A2* gene. The *original state* allele arose by a dSpm element insertion of 2241 bp. This element undergoes frequent excision in the presence of En/Spm. In the absence of the autonomous element it conditions a low pigmentation level (Figure 4) indicating a low level of gene activity. Further deletion to a 1341 bp element permits a relatively high level of gene expression and leads to stabilization of the insertion.

A similar sequence of events was reported for the suppressible allele *bz-m13* and some derivatives of it, two of which form almost stable insertions (Schiefelbein *et al.*, 1988; Raboy *et al.*, 1989). Gene expression of these alleles in the absence of En/Spm is restored to a significant extent because all but 2 bp of the dSpm element are removed by splicing. The dSpm insertion in these alleles occurred 38 bp downstream of the 5′ end of the second exon of the *Bronze1* gene. The splicing event utilizes the 5′ site (donor) of the single *Bronze1* intron and the 3′ site (acceptor) located in the 13 bp TIR of the dSpm element. The reading frame is maintained by this alteration. Tissues homozygous for these alleles contained enzymatic activity, suggesting that the modified protein is still able to catalyze the *Bronze1* reaction (Raboy *et al.*, 1989).

Processing of the *a2-ml* (*class II state*) allele is different with respect to the above situation in several aspects. The *A2* gene contains no intron. Here also the 5′ splice site is provided by one terminus of the 1341 bp dSpm insertion (Figure 7). The 3′ site utilized is the same as in the case of *bz-m13*. As a consequence of this RNA processing event, 18 bp of dSpm sequence plus 3 bp from the duplication generated during integration of the element remain within the mRNA. This modified *A2* gene product is functional, as shown directly by the transient expression of its cDNA (Figure 8).

The good accordance of the 3′ splice site with the plant consensus sequence (Brown, 1986) has already been pointed out by Raboy *et al.* (1989). Also, the 5′ site utilized in the case of *a2-ml* agrees well with the consensus sequence, except for one position (Figure 7). This site does not conform to the GT rule but instead reads GC. The same violation of the GT rule was reported for the first intron of the *nodulin-24* gene (Katinakis and Verma, 1985). In this respect it is interesting to note that one of the few sequence differences between En and Spm occurs at this site (Pereira *et al.*, 1986; Masson *et al.*, 1987). In En this sequence fulfils the GT rule perfectly. In spite of this aberrant feature, splicing seems to occur efficiently. According to mRNA analysis of the *a2-ml* (*class II allele*), the majority of the primary transcripts are processed to yield the 1.5 kb mRNA (Figure 6). *A2* gene expression of the *a2-ml* (*original state*) allele is considerably lower than the expression level of the *class II state* allele (Figure 4). Although not tested, it seems likely that correct processing is less efficient, presumably because additional sequences are present in the larger dSpm element, which might reduce the amount of functional product.

In the suppressible alleles *a2-ml*, *bz-m13* (Schiefelbein *et al.*, 1988) and *a1-ml* (Schwarz-Sommer *et al.*, 1985b), the orientation of the dSpm elements with respect to transcription is identical. The 3′ end of the element is promoter proximal (Figure 1). Removal of dSpm sequences by splicing in this orientation seems to be relatively efficient, since these alleles condition relatively high pigmentation levels. The dSpm insertion at the *wx-m8* allele occurred in opposite orientation (Gierl *et al.*, 1985). McClintock (1961) reported a very low level of gene activity for this allele. This might indicate that removal of dSpm in this orientation by splicing might also be possible, though at a lesser level.

The splicing event alters the primary structure of the protein

Splicing does not precisely remove the dSpm sequences from pre-mRNA. As a result, the protein sequence is altered. In the case of the *a2-ml* (*class II state*) allele, seven new amino acids are added to the *A2* protein sequence at the site of dSpm insertion (Figure 7). In case of the *bz-m13* allele a net loss of eleven amino acids occurred (Schiefelbein *et al.*, 1988). The alterations that occur after splicing should depend on the position of the insertion with respect to gene structure, and on the structure of the element itself. These alterations are reminiscent of the sequence variations ('footprints') that occur after plant transposable element excision (Schwarz-Sommer *et al.*, 1985a; Saedler and Nevers, 1985) that have been postulated to contribute to genetic variability.

Reduction of the impact of transposable element insertion

Transposition is tightly linked to mutation. There are 50–100 dSpm elements in the maize genome (Schwarz-Sommer *et al.*, 1984). These elements can potentially transpose, provided that they have retained intact ends, as illustrated by the multiple transpositions of the dSpm of the *a2-ml* (*original state*) allele (see above). The deleterious consequences of insertion mutagenesis are reduced, because the autonomous element promotes the deletion of *cis*-acting sequences that are necessary for excision (McClintock, 1955), perhaps by prematurely aborted attempts at excision. As in the case of the *a2-ml* (*class II state*) allele, this mechanism leads to stabilization of dSpm elements. Another mode of silencing the mutability of the En/Spm system involves methylation of element sequences (Banks *et al.*, 1988).

The consequences of transposable element insertion could be reduced further by the ability of dSpm elements to function as introns. This feature seems to apply more generally, because it has been observed with another transposable element system. The defective transposable elements (Ds) of the Ac family can also be removed by splicing from primary transcripts (Wessler *et al.*, 1987; for review, see Wessler, 1989). In addition it has been shown that a Ds element can 'activate' cryptic splice sites that are close to the insertion site (Simon and Starlinger, 1987). Maybe these features provide a certain selective advantage by mitigating the impact of transposable element insertion.

The dSpm element at the *a2-ml* (*class II state*) allele forms a stable intron, however, this allele is still 'controlled' by the suppressor function of En/Spm (Figure 4). Further deletion of the tnpA binding sites at the intact 3′ end would lead to escape from this control. Such a deletion derivative could be isolated by its predicted phenotype, and if successful, this would complete a chain of events leading to the genesis of a perfect intron.

Materials and methods

Maize stocks

Lines containing the *a2-ml* (*original state*) allele and the *a2-ml* (*class II state*) allele (McClintock, 1957 and 1958) were originally obtained from B.McClintock and were propagated by sibbing and by introduction of En(s) into the genomes. The phenotypes of kernels carrying the *a2-ml* in the absence and presence of En are the same, as described by McClintock (1958) for the absence and presence of Spm. The origin and genetic characterization of the *a2-rcy* line (*a2-m668291*) have been described (Peterson, 1988). A family segregating for this allele was obtained by selecting colored (*A2/a2*) and colorless (*a2-m668291/a2*) kernels from ear87 3113-21/2902 derived from the cross: *A2/a2-m668291* × *a2/a2* (Peterson, 1988).

Molecular cloning

The 3.4 kb *Bam*HI fragment originating from the *a2-rcy* mutant was cloned into the lambda vector NM1149; the two dSpm elements originating from the *a2-ml* alleles were cloned as *Mbo*I partial digestion fragments into the lambda vector EMBL4, as previously described (Sommer *et al.*, 1990). Phages were plated on the host CES200 (rec BC) (Wyman *et al.*, 1985).

DNA sequence analysis

DNA sequence was determined by a combination of the methods described by Maxam and Gilbert (1980) and Sanger *et al.* (1977).

Determination of transcription start site

Essentially the protocols as described in Sambrook *et al.* (1989) were followed. For S1 mapping, a 331 bp *Pvu*I–*Bam*HI fragment (positions −109

to 202, see Figure 2), labeled at the *Bam*HI site was used. For primer extension, a 56 bp *Ava*I–*Bam*HI fragment (positions 167–202), labeled at the *Bam*HI site was used.

Transient transformation by particle bombardment

Essentially the method of Klein *et al.* (1988) was used. Mature maize kernels homozygous for *a2*, devoid of the pericarp layer, were bombarded with tungsten particles (0.7 μm in diameter). For transformation of 10 kernels, 2.5 mg particles, coated with 5 g plasmid DNA were used. For negative controls, plasmid DNA without *A2*-cDNA inserts was applied, transformation frequencies were controlled with a plasmid containing the glucuronidase gene (data not shown).

All other methods applied for the studies in this report were performed as published earlier (Sommer *et al.*, 1990).

Acknowledgements

We thank Cathie Martin and Andrea Prescott for providing the amino acid sequence of the *Inc* gene prior to publication.

References

Banks,J.A., Masson,P. and Fedoroff,N. (1988) *Genes Dev.*, **2**, 1364–1380.
Brown,J.W.S. (1986) *Nucleic Acids Res.*, **14**, 9549–9559.
Fedoroff,N.V. (1989) In Howe,M. and Berg,P. (eds), *Mobile DNA*, Society for Microbiology, DC:ASM Press, pp. 375–411.
Gierl,A., Schwarz-Sommer,Zs. and Saedler,H. (1985) *EMBO J.*, **4**, 579–583.
Gierl,A., Lütticke,S. and Saedler,H. (1988) *EMBO J.*, **7**, 4045–4053.
Gierl,A., Saedler,H. and Peterson,P.A. (1989) *Annu. Rev. Genet.*, **23**, 71–85.
Grant,S.R., Gierl,A. and Saedler,H. (1990) *EMBO J.*, **9**, 2029–2035.
Heller,W. and Forkmann,G. (1988) In Harborne,J.B. (ed.), *The Flavonoids*. Chapman and Hall Ltd, London, pp. 399–425.
Holdsworth,M.J., Bird,C.R., Ray,J., Schuch,W. and Grierson,D. (1987) *Nucleic Acids Res.*, **15**, 731–739.
Katinakis,P. and Verma,D.P.S. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4157–4161.
Klein,T.M., Wolf,E.D., Wu,R. and Sanford,J.C. (1987) *Nature*, **327**, 70–73.
Klein,T.M., Gradziel,T., Fromm,M.E. and Sanford,J.C. (1988) *Biotechnology*, **6**, 559–563.
Kozak,M. (1987) *Nucleic Acids Res.*, **15**, 8125–8133.
Masson,P., Surosky,R., Kingsbury,J. and Fedoroff,N.V. (1987) *Genetics*, **177**, 117–137.
Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
McClintock,B. (1954) *Carnegie Inst. Washington Year Book*, **53**, 254–260.
McClintock,B. (1955) *Carnegie Inst. Washington Year Book*, **54**, 245–255.
McClintock,B. (1957) *Carnegie Inst. Washington Year Book*, **56**, 393–401.
McClintock,B. (1958) *Carnegie Inst. Washington Year Book*, **57**, 415–429.
McClintock,B. (1961) *Carnegie Inst. Washington Year Book*, **60**, 469–476.
Nelson,O.E. and Klein,A.S. (1984) *Genetics*, **106**, 769–779.
Niesbach-Klösgen,U., Barzen,E., Bernhardt,J., Rhode,W., Schwarz-Sommer,Zs., Reif,H.J., Wienand,U. and Saedler,H. (1987) *J. Mol. Evol.*, **26**, 213–225.
Pereira,A., Cuypers,H., Gierl,A., Schwarz-Sommer,Zs. and Saedler,H. (1986) *EMBO J.*, **5**, 835–841.
Peterson,P.A. (1953) *Genetics*, **38**, 682–683.
Peterson,P.A. (1965) *Am. Nat.*, **99**, 391–398.
Peterson,P.A. (1988) In Nelson,O. (ed.), *Proceedings of the International Symposium on Plant Transposable Elements*. Plenum, New York, pp. 43–68.
Raboy,V., Kim,H.-Y., Schiefelbein,J.W. and Nelson,O.E. (1989) *Genetics*, **122**, 695–703.
Reddy,G.M. and Coe,E.H. (1962) *Science*, **138**, 149–150.
Rhode,W., Barzen,E., Marocco,A., Schwarz-Sommer,Zs., Saedler,H. and Salamini,F. (1987) *Barley Genetics*, **5**, 533–541.
Saedler,H. and Nevers,P. (1985) *EMBO J.*, **4**, 585–590.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1990) *Molecular Cloning: a Laboratory Manual*. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schiefelbein,J.W., Raboy,V., Kim,H.-Y. and Nelson,O.E. (1988) In Nelson,O.E. (ed.), *Proceedings of the International Symposium on Plant Transposable Elements*. Plenum, New York, pp. 261–278.
Schnable,P.S. and Peterson,P.A. (1988) *Genetics*, **120**, 587–596.
Schnable,P.S., Peterson,P.A. and Saedler,H. (1989) *Mol. Gen. Gent.*, **217**, 459–463.
Schwarz-Sommer,Zs., Gierl,A., Klösgen,R.-B., Weinand,U., Peterson,P.A. and Saedler,H. (1984) *EMBO J.*, **3**, 1021–1028.
Schwarz-Sommer,Zs., Gierl,A., Cuypers,H., Peterson,P.A. and Saedler,H. (1985a) *EMBO J.*, **4**, 591–597.
Schwarz-Sommer,Zs., Gierl,A., Berntgen,R. and Saedler,H. (1985b) *EMBO J.*, **4**, 2439–2443.
Simon,R. and Starlinger,P. (1987) *Mol. Gen. Genet.*, **209**, 198–199.
Sommer,H., Beltrán,J.-P., Huijser,P., Pape,H., Lönning,W.-E., Saedler,H. and Schwarz-Sommer,Zs. (1990) *EMBO J.*, **9**, 605–613.
Töpfer,R., Matzeit,V., Gronenborn,B., Schell,J. and Steinbiss,H.-H. (1987) *Nucleic Acids Res.*, **15**, 5890.
Wessler,S.R. (1989) *Gene*, **82**, 127–133.
Wessler,S.R., Baran,G. and Varagona,M. (1987) *Science*, **237**, 916–918.
Wyman,A.R., Wolfe,L.B. and Botstein,D. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 2880–2884.

Received on May 14, 1990; revised on July 13, 1990