RNA splicing permits expression of a maize gene with a defective Suppressor-mutator transposable element insertion in an exon
(bronze-1 gene/splice-site sequence/intron/Ze a mays)

HWA-YEONG KIM, JOHN W. SCHIEFELBEIN, VICTOR RABOY, DOUGLAS B. FURTENK*, AND OLIVER E. NELSON, JR.

Laboratory of Genetics, University of Wisconsin, Madison, WI 53706

Contributed by Oliver E. Nelson, Jr., May 14, 1987

ABSTRACT The bz-m13CS9 allele of the bronze-1 gene in maize contains a 902-base-pair defective Suppressor-mutator (dSpm) transposable element in the second exon. Nevertheless, 40-50% of the enzymatic activity conditioned by a nonmutant allele at the bronze-1 locus is recovered in crude extracts prepared from plants carrying bz-m13CS9 in the absence of an autonomous Suppressor-mutator element. Analyses of RNAs produced by such plants show that transcription proceeds through the dSpm. The dSpm sequence of the messenger RNA precursor is then removed by RNA splicing using the donor site of the single bronze-1 intron and an acceptor site within the inverted terminal repeat of the dSpm. This results in a messenger RNA with the proper reading frame that could produce a functional enzyme. These data demonstrate that this dSpm insertion in an exon of a structural gene has produced a functional allele with a novel intron consisting, in part, of the dSpm. This mechanism appears to allow dSpm elements to reduce the impact of their insertions on gene expression.

The maize defective Suppressor-mutator (dSpm) transposable elements are nonautonomous members of the Suppressor-mutator (Spm; also known as Enhancer-Inhibitor, En-I) family (1-3). They can transpose or excise only in the presence of an autonomous Spm element. One of the interesting features of dSpm elements, first noted by McClintock (2, 4), is that association of these elements with structural genes in some cases gives rise to nonmutant phenotypes in the absence of Spm. In order to understand the underlying mechanism of this phenomenon, we have analyzed an allele of this type at the maize bronze-1 locus.

The bronze-1 locus is one of the many loci involved in the anthocyanin biosynthetic pathway of maize. The gene encodes UDP-glucose:flavonoid O3-p-glucosyltransferase (UGFTase, EC 2.4.1.91; refs. 5 and 6). Recessive alleles at the locus result in the bronze coloration of the aleurone layer and the brown coloration of many other plant parts as compared with the wild-type purple coloration. The mutant allele analyzed in the present study has a dSpm insertion in Bz, a nonmutant allele at the bronze-1 locus (7). This mutant allele has been designated bz-m13CS9 (abbreviated as CS9). In the presence of an active Spm, expression of CS9 is suppressed in most cells except for infrequent revertant cell lineages. As a result, one observes small colored sectors on a bronze background. In the absence of an active Spm, CS9 conditions full anthocyanin pigmentation.

In this paper, we report that CS9 conditions a relatively high level of UGFTase activity in the absence of Spm despite the presence of the dSpm element in the second exon. Analyses of RNAs produced by CS9 in the absence of Spm show that the dSpm element contains an acceptor site for RNA splicing within its inverted terminal repeat. This enables RNA splicing to efficiently remove the dSpm sequence from the mRNA precursor, allowing for gene function.

MATERIALS AND METHODS

Plant Materials. The isolation of the CS9 and Bz'-3 alleles has been described (7, 8). CS9 arose by a large internal deletion of the dSpm element in the initial insertion mutation bz-m13 (7). The bz-m13 allele, in turn, resulted from a 2.2-kilobase (kb) dSpm insertion into a progenitor Bz allele (9). This progenitor Bz allele has been designated Bz-McC2, since sequence analyses have indicated that Bz-McC2 is different from previously identified Bz alleles (ref. 10; J.W.S., unpublished data). Bz'-3 is a revertant allele of bz-m13 that arose by the precise excision of the dSpm and the host duplication (11). Bz'-3 conditions UGFTase activity equal to Bz-McC2. Since Bz-McC2 was not in an appropriate background when the present study was carried out, the Bz'-3 allele was used as the wild-type control. In addition to other color factors, CS9 and Bz'-3 carry dominant alleles at the b (booster) and pl (purple plant) loci, which are necessary for maximal Bz expression in husk tissue.

Protein Analysis. To prepare crude extracts, samples of husk tissue were harvested 15 days after silk emergence, pulverized in liquid N2, and extracted for 1 hr in a buffer (4 ml per gram of tissue), consisting of 100 mM Hepes (pH 7.5), 100 mM NaCl, and 20 mM 2-mercaptoethanol, in which AGI-X2 ion-exchange resin (250 mg/ml) was suspended. After centrifugation (10,000 x g for 10 min), aliquots of the supernatant fractions were assayed for UGFTase activity as described by Klein and Nelson (12). UGFTase activity in husk tissue used for RNA preparation was as follows: Bz'-3, 5.37 units per gram fresh weight; CS9, 2.19 units per gram fresh weight (41% of Bz'-3). A unit of UGFTase activity is defined as one μmol of product (isoquercitrin) formed per hour. Thermal stability assays of ammonium sulfate-fractionated UGFTase were conducted as described by Dooner and Nelson (13). We found that UGFTases isolated from husk tissue are significantly more stable than those isolated from endosperm tissue (14). The reason for this tissue-specific difference in thermal stability is not currently understood.

RNA Blot Hybridization Analysis. Total cellular RNA was extracted as described by Furtek (10) from the husk tissue of plants homozygous for the specified alleles at the bronze-1 locus in the absence of Spm. Husks were harvested 15 days after silk emergence. RNA samples (7 μg) enriched for the polyadenylated [poly(A)+] RNA fraction were fractionated

Abbreviations: Spm, Suppressor-mutator; dSpm, defective Suppressor-mutator; UGFTase, UDP-glucose:flavonoid O3-p-glucosyltransferase; pre-mRNA, mRNA precursor; Bz, nonmutant allele at the bronze-1 locus.

*Present address: Institut für Genetik, Universität Köln, D-5000 Köln, Federal Republic of Germany.
in 1.0% agarose/2.2 M formaldehyde gels, transferred to nitrocellulose filters, and probed with 32P-labeled DNA probes (15).

**Genomic Cloning.** The genomic cloning of the Bz'-3 allele has been described (7). The CS9 allele was cloned by partial Bgl II digestion of genomic DNA and ligation to BamHI arms of bacteriophage λEMBL3 (16). Recombinant phages were selected by plaque hybridization to the plasmid pMBzP17, which contains a unique 2.1-kb DNA fragment from a Bz allele (10).

**Complementary DNA (cDNA) Cloning.** Double-stranded cDNA was made as described by Furtek (10). Since only the sequence of the cDNA around the intron and the dSpm insertion was necessary, the double-stranded cDNA was digested with restriction enzymes (Xma I and Sst I) that have cleavage sites outside of this region (for the structure of CS9 see Figs. 1 and 2a). The digested cDNA was ligated to pUC18 plasmid that had been digested with the same restriction enzymes. Recombinant plasmids containing partial cDNAs of CS9 RNAs were selected by colony hybridization to the probe spanning the internal transcription unit of the bronze-1 gene. The probe was isolated from the plasmid pMBzP17 (10).

**DNA Sequencing.** DNA sequence was determined by the dyeoxy chain-termination procedure (17) using [α-35S]thiol-dATP (Amersham).

**RESULTS**

DNA sequence analysis of the CS9 genomic clone shows that the dSpm element of CS9 is 902 base pairs (bp) long and is inserted in the second exon 38 bp downstream from the single intron of the bronze-1 gene (Figs. 1 and 2b). Forty to fifty percent of the wild-type (Bz'-3) level of UFGTase activity is routinely recovered in crude extracts prepared from the husk tissue of plants homozygous for CS9 in the absence of Spm.

Blot hybridization analyses of the poly(A)+ RNAs isolated from the husk tissue of plants homozygous for CS9 in the absence of Spm reveals two bands of bronze-1-specific RNAs (Fig. 1). One has a size (about 1.8 kb) similar to that of the Bz'-3 mRNA, and the other is larger. These RNAs hybridize to probes from Bz sequences both 3' and 3' to the dSpm insertion site. However, only the larger RNA also hybridizes to the dSpm probe. Its size is consistent with that expected from the presence of the dSpm sequence in the unspliced mRNA precursor (pre-mRNA). The 1.8-kb RNA hybridizes to Bz probes but not to the dSpm probe. It is much more abundant than the larger RNA. These results suggest that in the absence of Spm, transcription proceeds through the dSpm, and that most or all of the dSpm sequence is removed from the mRNA.

Sequence comparison of the CS9 genomic DNA clone and a cDNA clone representing a CS9 RNA shows that this RNA does not contain the intron, the sequence between the intron and the dSpm, or the dSpm sequence except for the terminal two nucleotides (TG) at the 3' end (Fig. 2). We therefore suggest that this cDNA represents a 1.8-kb RNA. The sequence removed from the pre-mRNA ends with an AG dinucleotide within the 13-bp inverted terminal repeat of the dSpm (Figs. 2b and 3a). The border sequences of the pre-mRNA thus conformed to the GT/AG rule of pre-mRNA splicing (23, 24). In addition, the inverted terminal repeat contains a sequence that conforms to the consensus sequences of acceptor sites for pre-mRNA splicing (18, 19), as shown in Fig. 3a. These data indicate that the dSpm sequence is removed by pre-mRNA splicing using the donor site of the bronze-1 intron and the acceptor site within the inverted terminal repeat of the dSpm.

Results from the cDNA analysis indicate that the 1.8-kb RNA represented by the cDNA is a processed mRNA. The

Fig. 1. Blot hybridization analyses of poly(A)+ RNAs with Bz and dSpm probes. Poly(A)+ RNAs were isolated from the husk tissue of plants homozygous for the indicated alleles at the bronze-1 locus in the absence of Spm; Bz refers to Bz'-3. Poly(A)+ RNAs were fractionated in formaldehyde-containing 1% agarose gels, transferred to nitrocellulose filters, and probed with the indicated 32P-labeled DNA probes. The diagram below the autoradiogram describes the DNA structure of CS9. Probes are indicated by open boxes: probe A contains a Bz sequence 5' to the dSpm, probe B contains most of the dSpm, and probe C contains a 3' Bz sequence. Blots probed with A and B were made by fractionating the poly(A)+ RNAs of Bz'-3 and CS9 in duplicate lanes in the same gel, transferring to a nitrocellulose filter, and dividing the filter into two parts each containing the RNAs from the two genotypes. Ribosomal RNAs (28S and 18S) provided size markers.

sequence between the intron and the dSpm is not present in this mRNA. To test for splicing that would produce a 1.8-kb mRNA or RNAs containing the sequence between the intron and the dSpm, we probed a blot of electrophoretically fractionated RNA with a DNA fragment including a portion of the intron and the 38 nucleotides between the intron and the dSpm (for the probe see the upper diagram of Fig. 4). One possible way to produce a 1.8-kb mRNA containing the 38 nucleotides would be the independent removal of the bronze-1 intron and the dSpm sequence as shown in the lower diagram of Fig. 4. If this hypothetical splicing event were to occur, some of the 1.8-kb mRNA produced by CS9 should hybridize to the probe. The Bz'-3 mRNA should also hybridize to the probe, since it contains the 38 nucleotides. The probe hybridized to the Bz'-3 mRNA as expected but not to the 1.8-kb mRNA produced by CS9 (Fig. 4). This suggests that there is no splicing event that produces a detectable level of a 1.8-kb mRNA containing the 38 nucleotides.

The RNA blot analysis shown in Fig. 4 also suggests that the larger RNA of CS9 is the unspliced pre-mRNA. The RNA blot
Accordingly, the larger RNA shows a much reduced signal as compared with that of the Bz'-3 mRNA on the blot of Fig. 4 if its signal were solely due to the presence of the 38 nucleotides between the intron and the dSpm. However, the larger RNA shows a signal similar to that of the Bz'-3 mRNA, suggesting that the signal by this RNA is not only due to the 38 nucleotides but also due to the intron left unspliced. The presence of the unspliced pre-mRNA might imply that the splicing of CS9 RNA is less efficient than that of Bz'-3 RNA.

The 1.8-kb mRNA revealed by the cDNA does not contain the 38 nucleotides between the intron and the dSpm and does include the two nucleotides (TG) from the 3' end of the dSpm and the trinucleotide (CCG) duplicated upon the insertion of the dSpm (Fig. 2b). This results in a net loss of 33 nucleotides and maintains the proper reading frame. Translation of this mRNA could thus produce a polypeptide for a functional UFGTase if the loss of 11 amino acids and change of alanine (GCC) to valine (GTG) does not destroy UFGTase activity (Fig. 2b).

Thermal-stability profiles indicate that CS9 husk UFGTase activity has a half-life at 55°C of 6.5 min as compared with 42 min for Bz'-3 husk UFGTase activity. This suggests that CS9 UFGTase protein is altered as compared with wild-type UFGTase protein. In addition, no detectable level of the 1.8-kb mRNA contains the 38 nucleotides between the intron and the dSpm. These lines of evidence suggest that the 1.8-kb mRNA represented by the cDNA is responsible for the functional UFGTase encoded by CS9.

**DISCUSSION**

Several authors have proposed that some introns could be the remnants of transposable-element insertions (26--28) despite the claim that no known transposable element has splice-site sequences at its borders (29). The observation that the intron of the yeast mitochondrial 21S rRNA gene encodes enzymes
Fig. 4. Blot hybridization analysis of poly(A)+ RNAs to test for the presence of a 1.8-kb CS9 mRNA or mRNAs containing the sequence between the intron and the dSpm. Experimental procedures and genotypes were the same as those of Fig. 1. The probe includes a portion of the intron (open region of the bar representing the probe) and the 38 nucleotides between the intron and the dSpm element (filled region of the bar). Poly(A)+ RNA from a deletion stock (bz-x2; ref. 25) lacking the bronze-1 locus (indicated by X2) was included as a negative control for the intron sequence. The lower diagram illustrates the splicing event being tested, in which the intron and the dSpm are removed independently. Restriction enzymes: X, Xma I; C, Sca I; S, Sst I.

necessary for its own splicing has suggested a relationship between introns and transposable elements (30, 31). The present study clearly shows that a dSpm transposable element in maize contains an acceptor site for pre-mRNA splicing within its inverted terminal repeat. Insertion of this element into an exon has resulted in a splicing event that uses this acceptor site. Some acceptor sites in maize might thus have been introduced by dSpm or Spm element insertions followed by the loss or divergence of sequences required for transposition.

In the absence of Spm, the acceptor-site sequence within the inverted terminal repeat of the dSpm of CS9 enables RNA splicing to remove the dSpm sequence from the pre-mRNA, allowing for Bz expression. However, expression would not be expected if the dSpm were inserted in the opposite orientation. CS9 has resulted from an internal deletion of the dSpm element of bz-m13 without changing the insertion position (7). The dSpm of bz-m13 is virtually the same as the Spm-I8 element of wx-m8 (an insertion mutation of the maize waxy gene), but the two elements are inserted in the opposite orientation in terms of transcription of the two genes (ref. 32; J.W.S., unpublished data). Thus, the dSpm of CS9 is oriented in the direction opposite to that of the dSpm in wx-m8. In the absence of Spm, wx-m8 produces chimeric mRNAs that are prematurely polyadenylated due to the polyadenylation signals within the element (32). Presumably, Bz mRNAs would also be prematurely polyadenylated if the dSpm of CS9 were inserted in the opposite orientation, provided that the deletion which produced the dSpm in CS9 from that of CS9 in bz-m13 did not remove the polyadenylation signals.

The insertion position of the dSpm is also important for the expression of CS9. The 1.8-kb mRNA of CS9 represented by the cDNA does not contain the 38 nucleotides (12 codons and 2 extra nucleotides) between the intron and the dSpm. Thus, the reading frame of the 1.8-kb mRNA would be shifted if the 2 nucleotides from the 3′ end of the dSpm were not added. Accordingly, the number of nucleotides between an intron and a dSpm insertion must represent a multiple of 3 and an addition of 2 (i.e., 3N + 2) in order to maintain the proper reading frame of a spliced mRNA. In addition, a loss of or change in amino acids (determined by insertion position) may destroy protein function even though the proper reading frame is maintained. Therefore, the expression of a gene with a dSpm insertion in an exon depends on the orientation and insertion position of the element. Moreover, we do not know how far 3′ to the intron a dSpm can be inserted and still be removed from a pre-mRNA by the type of splicing event observed with CS9. Variation in the length of the intron preceding this dSpm might also affect the splicing event observed. The consequences on Bz expression of a dSpm insertion 5′ to the single intron are also unknown.

Our results demonstrate that a transposable-element insertion in an exon of a structural gene has produced a functional allele with a modified intron consisting in part of the transposable element. As shown in Fig. 3b, the DNA sequences of 13-bp inverted terminal repeats of Tam1, a transposable element in snapdragon (20), and Tgm1, a presumed transposable element in soybean (21), are very similar to that of Spm (22, 33). These two elements might also provide an acceptor site when inserted into exons in the same fashion as the dSpm of CS9. Perhaps, as discussed by Doolittle and Sapienza (28), this structural feature allows these transposable elements in widely divergent plant species to take advantage of RNA splicing to reduce the impact of their insertions on gene expression.

We thank Patrick Masson and Nina Fedoroff for communicating unpublished data, William Engels and William Eggleston for critical reading of the manuscript, Stephen Sicker for technical support, and Russell Huseth for assistance with the field work. This research was supported by the College of Agriculture and Life Sciences, University of Wisconsin, by National Science Foundation Grant DCOB-8507895, and by National Institutes of Health Predoctoral Training Grant T32-GM07133. H.-Y.K. was supported by a fellowship from the Rockefeller Foundation, and J.W.S. was supported by a fellowship from the AgriGenetics Corporation. This is paper no. 2935 from the Laboratory of Genetics, University of Wisconsin-Madison.