The EMBO Journal vol.4 no.3 pp.579–583, 1985

Molecular interactions between the components of the En-I transposable element system of *Zea mays*

Alfons Gierl, Zsuzsanna Schwarz-Sommer and Heinz Saedler

Max-Planck-Institut für Züchtungsforschung, 5000 Köln 30, FRG

Communicated by H.Saedler

The sequence of the Inhibitor element Spm-I8 isolated from the *wx-m8* allele has been determined. The element is 2242 bp in length. Its ends can be folded into long stem and loop structures. In a line containing the autonomous En element (*wx-m8+En*) we have detected a 2.5-kb transcript hybridizing to Spm-I8. A cDNA copy of this En-specific transcript containing 1.2 kb of the 3′ end was cloned and its DNA sequence was determined. The 3′ half of the cDNA is homologous to Spm-I8 and the region of homology is interrupted by intervening sequences. In the absence of an autonomous En element two chimeric transcripts are observed in the *wx-m8* line which are probably initiated at the *wx* promoter and terminate in the Spm-I8 insertion. In the presence of the En element, these transcripts are suppressed, possibly by a *trans*-acting function of En, inhibiting transcription readthrough into Spm-I8.

Key words: receptor element/*waxy* locus/DNA sequence/cDNA cloning/chimeric transcript

Introduction

The En (Enhancer) (Peterson, 1953) and the Spm (Suppressor-Mutator) (McClintock, 1954) transposable element systems of *Zea mays* have been shown to be homologous both genetically (Peterson, 1965) as well as molecularly (Pereira *et al.*, 1985). Therefore the names can be used synonomously. The autonomous element can give rise to a two component system. One component, the regulator (McClintock, 1961) is autonomous and encodes the functions required for transposition. The second component, the Inhibitor (Peterson, 1953) is non-autonomous with respect to transposition, but inhibits the expression of a gene simply by virtue of its presence at the locus. Mutations induced by the inhibitor are stable unless *trans*-activated by the En regulator element. Since the inhibitor component I responds to En signals it also has been termed *receptor* (Peterson, 1965; Fincham and Sastry, 1974).

In the Ac/Ds as well as in the En/I system of *Zea mays* the receptor components are deletion derivatives of the regulator components (Fedoroff *et al.*, 1983; Döring and Starlinger, 1984; Pereira *et al.*, 1985). We have previously isolated the Spm (En) receptor element Spm-I8 from the *wx-m8* allele (Schwarz-Sommer *et al.*, 1984). Spm-I8 was shown to be integrated in an exon of the *waxy* (*wx*) gene (Schwarz-Sommer *et al.*, see accompanying paper).

In this paper we report the DNA sequence of Spm-I8 as well as the cDNA sequence of an En encoded transcript. Furthermore, Northern blotting experiments show that two chimeric transcripts are seen in the *wx-m8* mutant, both of which are suppressed by the introduction of the regulatory component En.

Results

DNA sequence of the Spm (En) receptor element Spm-I8

The Spm-I8 element was isolated from a recombinant λEMBL4 clone carrying the *wx-m8* allele (Schwarz-Sommer *et al.*, 1984) by subcloning into plasmid vectors. The DNA sequence was determined by the chemical degradation method (Maxam and Gilbert, 1980). Both DNA strands have been sequenced by labelling appropriate restriction fragments.

The nucleotide sequence of the 2242 bp long receptor element Spm-I8 appears in Figure 1. The terminal 200-bp of each border contain various direct and inverted repeats. Because of the repetitiveness of the inverted repeats many secondary structures are possible. Figure 2 shows two examples of stem and loop structures. With the exception of the terminal 13-bp inverted repeat, all other stems contain at least part of the 13-bp motif ACCGACACTCTTA. This 13-mer nucleotide (or close derivatives of it) is repeated eight times in the left arm and 10 times in the right arm in direct or inverse orientation.

Another feature observed is a distinct G+C-rich domain in the left arm of the element. The G+C content between positions 300 and 550 (Figure 1) is 82% compared with 48% of the whole element. In this region two direct repeats of 12 bp and 13 bp (Figures 1a, b, and 2a, b) are found.

No significant homologies between the termini of Spm-I8 and the DNA sequence around the integration site in the *wx* gene have been detected.

Characterisation of an En-specific transcript

Previously we showed that Spm-I8 is a deletion derivative of the autonomous En element (Pereira *et al.*, 1985). In poly $(A)^+$RNA preparations of a line (*wx-m8+En*) that contains an autonomous En element in addition to the Spm-I8 insertion at the *wx* locus, a 2.5-kb transcript is seen that hybridizes to Spm-I8 (Figure 3A). Lower mol. wt. RNA is also detected with this probe (Figure 3, probe B). This could be explained by the repetitive distribution of defective parts of Spm (En) in the genome (Schwarz-Sommer *et al.*, 1984), which seem to be co-transcribed with other genes.

The 2.5-kb transcript, however, is only found in the presence of En. To characterize further this En-specific transcript, we have cloned cDNA from *wx-m8+En* poly$(A)^+$RNA into the λNM1149 vector (Murray, 1983). Two cDNA clones, pEnc1 and pEnc2 were obtained, which contained cDNA inserts of 1.2 kb and 0.5 kb, respectively. Both correspond to the 3′ end of the transcript including the poly(A) tail.

DNA sequence comparison revealed that the 3′-terminal 570 bp of pEnc1 are homologous from position 862 to 1827 of Spm-I8 (Figure 1) and that Spm-I8 contains two intervening sequences in this region. No homology is found between Spm-I8 and the first 588 bp of pEnc1 (data not shown). The cDNA contains one open reading frame that ends at a UAG stop codon shown at position 1603 in Figure 1. Four single base pair substitutions relative to Spm-I8 are found in the last exon, three of which are in the 3′-non-coding region (Figure 1). The fourth one at position 1588 would change the amino acid encoded. The total length

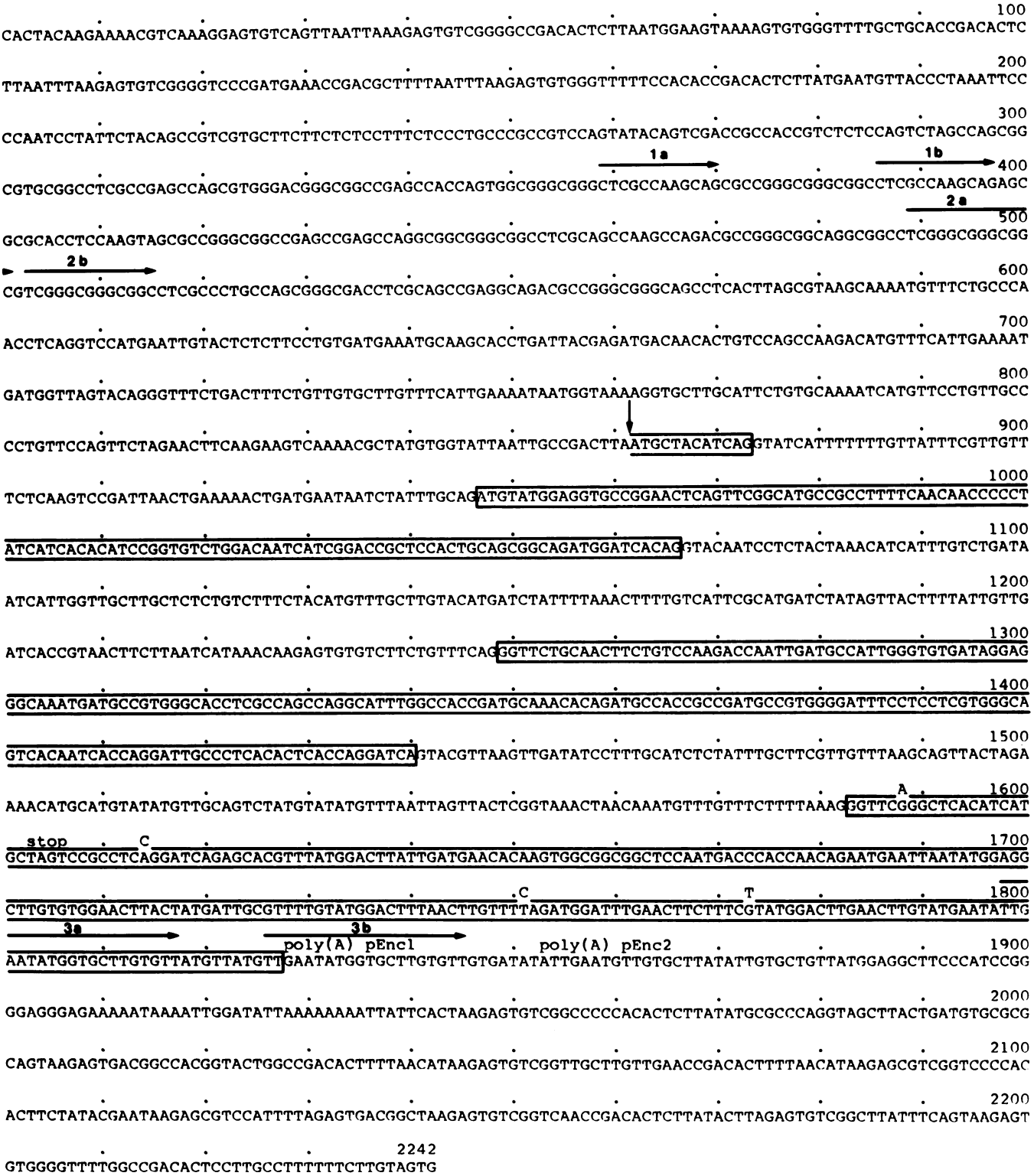

Fig. 1. DNA sequence of Spm-I8. Only one strand is shown. Regions homologous to the cDNA pEnc1 are indicated by the boxes. Deviations are shown above the sequence, as are the translational stop codon and the poly(A) addition sites. The vertical arrow indicates the point at which Spm-I8 is deleted relative to En1. A few direct repeats outside the highly structured termini are indicated by horizontal arrows.

of the 3′-non-coding region is 222 bp and does not contain the usual AATAAA polyadenylation signal. This point is considered further in the Discussion.

Transcription of the wx-m8 allele in the presence and absence of En

In the *wx-m8* allele, Spm-I8 is inserted in an exon of the *wx* gene (Schwarz-Sommer *et al.*, see accompanying paper) ~1.5 kb upstream of the *wx* poly(A) signal (Klösgen *et al.*, unpublished; the detailed structure of the *wx* locus will be published elsewhere). Figure 3B shows a schematic comparison of the *wx* wild-type and *wx-m8* locus.

The size of the *wx* wild-type transcript is 2.4 kb (Figure 3A) (Shure *et al.*, 1983). In poly(A)$^+$ RNA preparations from *wx-*

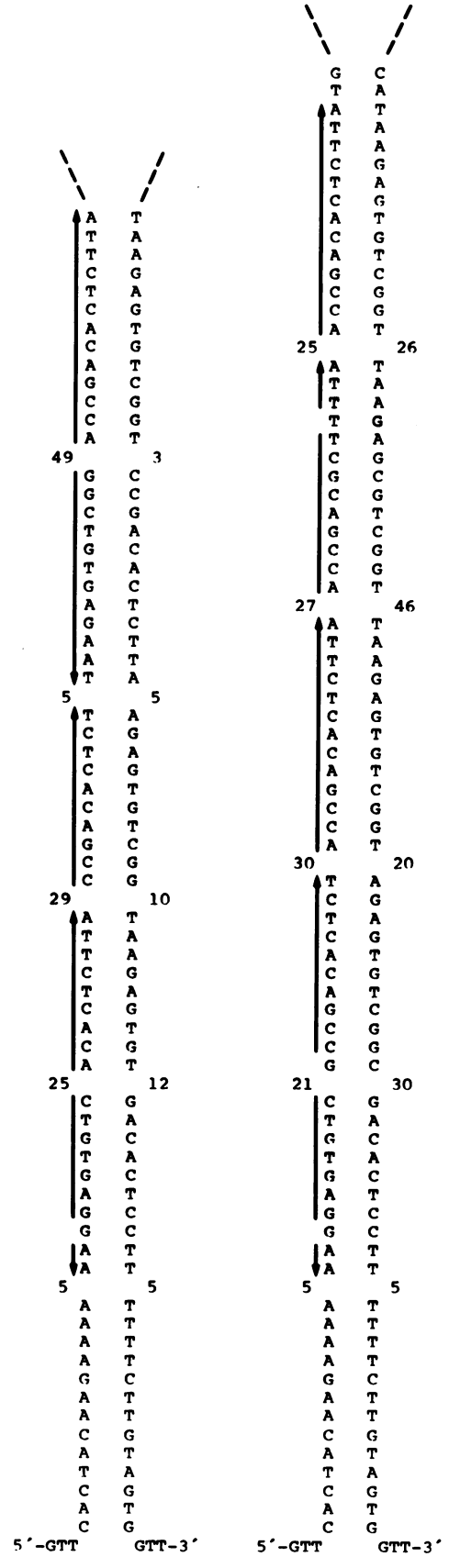

Fig. 2. Two possible stem and loop structures of the termini of Spm-I8. The target site duplication is shown at the bottom of each. The arrows indicate sequences which contain all or part of the common motif described in the text. The numbers indicate the size (in bases) of the single stranded loops.

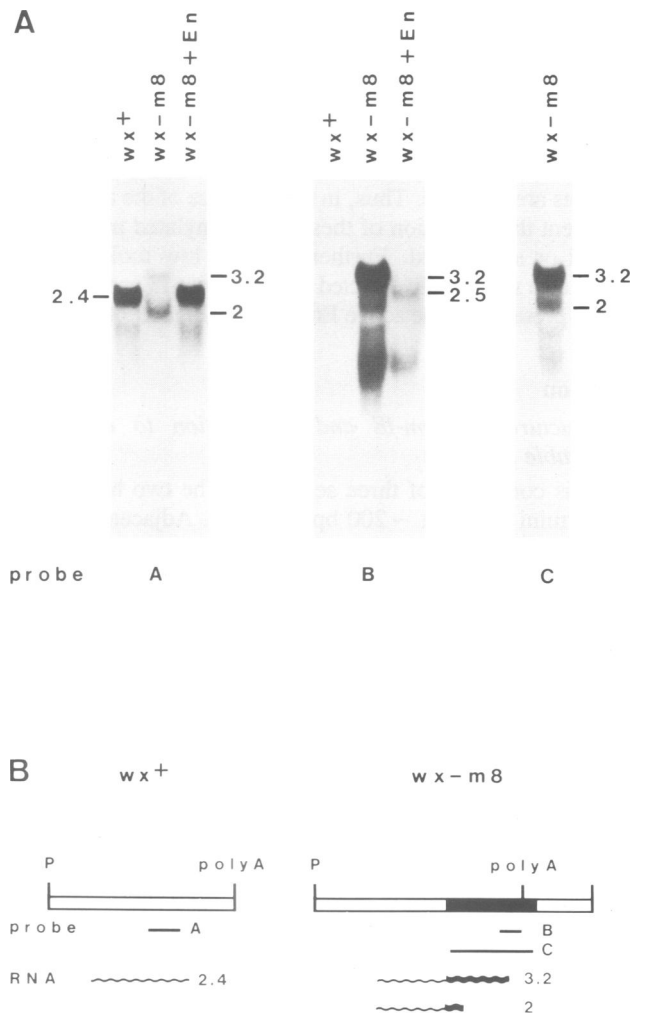

Fig. 3. (**A**) Northern experiment including RNA from three different plant lines. The orgin of the RNA is indicated on top. 1 μg of poly(A)$^+$ RNA from wx^+ was compared with 5 μg poly(A)$^+$ from *wx-m8* and *wx-m8+En* each. The three panels were probed with probes A, B and C whose origins are indicated in **part B**. The size is indicated in kilobases. The exposure time of the plot hybridized with probe B was four times longer than that with probe C. (**B**) Schematic drawing showing wx^+ *and wx-m8* alleles. The open box represents the wild-type allele, while the filled box indicates the Spm-I8 insertion. The fragments used as probes are indicated. Probe A corresponds to a *wx* genomic *SalI* 0.75-kb fragment including the Spm-I8 insertion site, probe B corresponds to position 1047−1842 and probe C to position 266−2157 of Spm-I8 (Figure 1). The wavy lines represent our interpretation of the transcripts observed with each probe.

m8, two transcripts are found instead, 3.2 kb and 2 kb in size. These hybridize to *wx* gene probes derived from the region upstream of the Spm-I8 integration site (Figure 3A, probe A), but not to probes from the downstream regions (data not shown). The larger of the two transcripts is also detected using probes of Spm-I8 for hybridization (Figure 3A, probes B, C). The size of this transcript is compatible with the assumption that a transcript is initiated at the *wx* promoter and terminates at the same site in Spm-I8 as the En-specific transcript (Figure 3B). The smaller transcript hybridizes to probes from the left (Figure 3A, probe C) but not to probes from the right half (Figure 3A, probe B) of the Spm-I8 element. It could, therefore, have been terminated further upstream in Spm-I8. [A possible poly(A) addition signal is found at position 750 in Spm-I8.]

Upon introduction of an active En element in the genome of a plant carrying the *wx-m8* allele (*wx-m8+En*) the excision fre-

quency of the Spm-I8 receptor element is ~20% (Schwarz-Sommer *et al.*, see accompanying paper). As a consequence of this, when poly(A)$^+$ RNA prepared from *wx-m8+En* tissue is probed with *wx* gene probes, a band corresponding in size to the wild-type is detected (Figure 3A). Although Spm-I8 is still integrated in 80% of the *wx* transcription units, no chimeric transcripts are detected. Thus, in the presence of the autonomous En element the formation of these polyadenylated transcripts is inhibited or suppressed. Furthermore the low mol. wt. RNA which is also specifically detected with the Spm-I8 probe is reduced in the presence of the active En element (Figure 3A, probe B).

Discussion

The structure of Spm-I8 and its relation to other plant transposable elements

Spm-I8 is composed of three segments. The two highly structured termini are each ~200 bp in length. Adjacent to the left end there is a 250 bp long stretch of 82% G+C. The remainder of the element contains the 3′ part of a mosaic gene. Besides the terminal 13-bp inverted repeat there are patches of other inverse sequence repetitions, which have a common motif. The 13-mer nucleotide ACCGACACTCTTA, or close derivatives of it, is repeated eight times in the left and 10 times in the right arm in direct or inverse orientation. Whether or not these structures play a role in the transposition process remains to be determined. However, the isolation of 'states' of the element (McClintock, 1955; Reddy and Peterson, 1984) with altered excision properties may shed light on this question.

Similar patches of inverse sequence repetitions have been observed in Tam1 (Bonas *et al.*, 1984) and Tam2 (Upadhyaya *et al.*, 1985) of *Antirrhinum majus* and Tgm1 (Rhodes and Vodkin, 1984) of *Glycine max*. Because of the preserved terminal five bases in all these elements they have been grouped into the CACTA family; their 13-bp terminal repeats display extensive sequence homology, and each generates a 3-bp duplication of the target site upon integration (Bonas *et al.*, 1984; Schwarz-Sommer *et al.*, 1984).

Spm-I8 contains the 3′ part of an En gene

Sequence comparison of an En-specific cDNA (pEnc1) and Spm-I8 revealed the presence of a remnant of that En gene on the receptor element. From the available sequence data one can conclude that the gene is organized in the usual eukaryotic exon-intron structure. Whether or not all functions encoded by plant transposable elements are organized in mosaic genes is unknown. Previously a large open reading frame (Pohlman *et al.*, 1984) in the Ac system of maize was thought to encode the transposase, but independent sequencing of Ac revealed two open reading frames instead (Müller-Neumann *et al.*, 1984), so that the possibility of a spliced gene in the Ac system still remains.

pEnc1 does not contain the usual poly(A) signal (AATAAA). We consider the sequence AATATG 21 bp upstream of the poly(A) tail to be the most likely candidate for the polyadenylation signal. This hexanucleotide resembles other variant poly(A) signals of plants (Messing *et al.*, 1983); plant poly(A) signals tend to show more variation than those of animal genes. Furthermore, a second cDNA clone, pEnc2, has been sequenced. Its length is 480 bp, it is virtually identical in sequence to pEnc1, but it is polyadenylated 25 p further downstream. 18 bp upstream from this site the same hexanucleotide AATATG is again found. These two possible poly(A) signals each are part of a 20-bp direct repeat as indicated in Figure 1 (3a, b). The extensive sequence homology between the En transcript and the Spm receptor element again demonstrates the close relationship of both elements. The function of the En-specific transcript is unknown. There are at least two *trans*-active 'signals' provided by the autonomous element, an S-function (suppressor) and an M-function (mutator) (McClintock, 1954). Whether pEnc1 corresponds to one of these functions can be tested by analysing En mutants (McClintock, 1954; Peterson, 1981) that are defective in either the M or the S-function.

It is known from heteroduplex analysis that Spm-I8 is a deletion derivative of the autonomous element (Pereira *et al.*, 1985). Therefore it is not unexpected that the sequence homology to the En-specific cDNA comprises only the 3′-terminal part of the cDNA. From preliminary sequence comparison with the autonomous element En1 (Pereira *et al.*, 1985; Pereira, personal communication) we know that the point of sequence divergence at the 5′ end of the cDNA from Spm-I8 is identical with the deletion endpoint on that element (Figure 1).

We intend to clone cDNA fragments into expression vectors to raise antibodies against the putative En protein. It is known that 20 – 50 Spm (En) homologous sequences are found in the maize genome (Schwarz-Sommer *et al.*, 1984). Therefore one has to ensure that a particular cDNA is a copy of a functional messenger by sequence comparison with an autonomous element. In case of pEnc1 these criteria seem to be fulfilled, as judged by preliminary sequence comparison with En1. Furthermore pEnc1 contains an open reading frame.

The presence of En suppresses transcription readthrough into Spm-I8 in the wx-m8 allele

Genetically it is observed that some integrations of the Inhibitor component into a locus still allow some residual expression of that locus in the absence of Spm (McClintock, 1954). If, however, Spm is crossed into the line, this residual activity is suppressed. In *wx-m8* tissue a 3.2-kb and a 2-kb transcript containing *wx* and Spm-I8 sequences are observed. Both chimeric transcripts are assumed to be initiated at the *wx* promoter but either terminate early or late in the Spm-I8 component. The size of the larger transcript is in good agreement with the data available on the transcription of both the *wx* gene and Spm-I8. As outlined in the schematic drawing in Figure 3B, Spm-I8 interrupts the *wx* transcription unit after the first 1600 nucleotides on the mature mRNA level. Assuming that splicing of the RNA within Spm-I8 occurs at the positions revealed by sequencing of pEnc1, the chimeric transcript would terminate close to the poly(A) sites shown (Figure 1). After processing, the total length of the larger chimeric transcript would be ~3.1 kb. This length, together with a poly(A) tail, would agree precisely with the observed size of 3.2 kb. The shorter transcript may terminate early in the Spm-I8 segment. A possible poly(A) site is located at position 750 in Spm-I8. We are aware that the above assumptions must be verified by cDNA cloning.

Even more interesting is the observation that the chimeric transcripts are suppressed in the presence of an active En element. Furthermore, the transcripts of other I-homologous sequences, which do not hybridize with a *wx*-specific probe, but are obviously co-transcribed with other genes, are also suppressed by En. This could reflect En's suppressor activity abolishing transcription. Alternatively it could be due to the En-encoded transposase (mutator-function) binding at the ends of Spm-I8 and thus sterically blocking transcriptional readthrough into the element. This transcript would probably not be polyadenylated and would thus have escaped detection.

The latter could perhaps explain McClintock's statement that M-action requires some S-activity. Since transposition of the

receptor component requires the binding of the transposase to the ends of the element, this would always result in at least partial blocking of transcriptional readthrough, and hence suppressor activity. Therefore, the isolation of suppressor and mutator mutations does not necessarily mean that both functions have to be encoded by separate genes. The suppressor activity could be another property of the transposase (mutator) protein. Analysis of S- and M-negative mutations will clarify this situation.

Materials and methods

Genetic stocks

All genetic stocks were the same as previously described (Schwarz-Sommer *et al.*, 1984).

Preparation of recombinant plasmids

From the recombinant λEMBL4 phage carrying a 19-kb insert containing the Spm-I8 element of the *wx-m8* allele (Schwarz-Sommer *et al.*, 1984) two overlapping fragments (a 1.3-kb *Pst*I and a 2.5-kb *Sal*I fragment) were subcloned into pUC9 (Vieira and Messing, 1982). They contained Spm-I8 and the flanking *wx* gene sequences. DNA from these clones was used as starting material for DNA sequencing, either by isolating appropriate restriction fragments or by further subcloning into pUC9.

For DNA sequence analysis of pEnc1 and pEnc2, the corresponding 1.2-kb and the 0.5-kb *Eco*RI fragments from the recombinant NM1149 clones were subcloned into pUC9.

cDNA cloning

Endosperm RNA from *wx-m8+En* kernels 18 days after pollination was prepared as already reported (Schwarz-Sommer *et al.*, 1984). cDNA synthesis was performed following standard protocols (Land *et al.*, 1981; Maniatis *et al.*, 1982). a more detailed description will be published elsewhere (Schwarz-Sommer *et al.*, see accompanying paper). cDNA fragments were ligated into the *Eco*RI site of the λNM1149 phage (Murray, 1983; Scherer *et al.*, 1981). Packaging was done as described previously (Wienand *et al.*, 1982).

Recombinant phage were selected by plating the packaged phage onto the hf1 *Escherichia coli* host POP 13 (obtained from V.Pirrotta, EMBL Heidelberg), which is a derivative of the strain POP 101 (Murray, 1983).

In a typical cloning experiment we obtained $2-5 \times 10^5$ p.f.u.s per $10-20$ ng of double-stranded, sized cDNA.

Sequence analysis

For DNA sequence analysis the chemical degradation procedure described by Maxam and Gilbert (1980) was followed. Electrophoresis was carried out on 0.25 mm thick polyacrylamide gels of $5-16\%$.

Acknowledgements

We thank R.Berndtgen for skilful technical assistance. A.G. was supported by a post-doctoral scholarship from the Deutsche Forschungsgemeinschaft.

References

Bonas,U, Sommer,H. and Saedler,H. (1984) *EMBO J.*, **3**, 1015-1019.
Döring,H.P. and Starlinger,P. (1984) *Cell*, in press.
Fedoroff,N., Wessler,S. and Shure,M. (1983) *Cell*, **35**, 235-242.
Fincham,J.R.S. and Sastry,G.R.K. (1974) *Am. Rev. Genet.*, **8**, 15-50.
Land,H., Grez,M., Hauser,H., Lindenmaier,W. and Schütz,G. (1981) *Nucleic Acids Res.*, **9**, 2251-2266.
Maniatis,T., Fritsch,E.P. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
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. (1961) *Am. Nat.*, **95**, 265-277.
Messing,J., Geraghty,D., Heidecker,G., Hu,N., Kridl,J. and Rubenstein,I. (1983) in Hollaender,A. (ed.), *Genetic Engineering of Plants*, Plenum Press, NY, pp. 211-227.
Müller-Neumann,M., Yoder,J.I. and Starlinger,P. (1984) *Mol. Gen. Genet.*, in press.
Murray,N.E. (1983) in Hendrix,R.W., Roberts,J.W., Stahl,F.W. and Weisberg,R.A. (eds.), *Lambda II*, Cold Spring Harbor Laboratory Press, NY, pp. 395-432.
Pereira,A., Schwarz-Sommer,Zs., Gierl,A., Bertram,I., Peterson,P.A. and Saedler,H. (1985) *EMBO J.*, **4**, 17-23.
Peterson,P.A (1953) *Genetics*, **38**, 682-683.
Peterson,P.A. (1965) *Am. Nat.*, **99**, 391-398.
Peterson,P.A. (1981) *Cold Spring Harbor Symp. Quant. Biol.*, **45**, 445-455.
Pohlman,R.F., Fedoroff,N.V. and Messing,J. (1984) *Cell*, **37**, 635-643.
Rhodes,P.R. and Vodkin,L.O. (1984) *Proc. Natl. Acad. Sci. USA*, in press.
Reddy,L.V. and Peterson,P.A. (1984) *Mol. Gen. Genet.*, **194**, 124-137.
Scherer,G., Telford,J., Baldari,C. and Pirrotta,V. (1981) *Dev. Biol.*, **86**, 438-447.
Schwarz-Sommer,Zs., Gierl,A., Klösgen,R.B., Wienand,U., Peterson,P.A. and Saedler,H. (1984) *EMBO J.*, **3**, 1021-1028.
Shure,M., Wessler,S. and Fedoroff,N. (1983) *Cell*, **35**, 225-233.
Upadhyaya,K.C., Sommer,H., Krebbers,E. and Saedler,H. (1985) *Mol. Gen. Genet.*, in press.
Wienand,U., Sommer,H., Schwarz,Zs., Shepherd,N., Saedler,H., Kreuzaler,F., Ragg,H., Fautz,E., Hahlbrock,K., Harrison,B. and Peterson,P.A. (1982) *Mol. Gen. Genet.*, **187**, 195-201.
Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259-268.

Received on 6 December 1984