The EMBO Journal vol.8 no.11 pp.3177-3185, 1989

The putative transposase of transposable element *Ac* from *Zea mays* L. interacts with subterminal sequences of *Ac*

Reinhard Kunze and Peter Starlinger

Institut für Genetik, Universität zu Köln, D-5000 Köln 41, FRG

Communicated by P.Starlinger

The *Ac*-specific ORFa protein, overexpressed in a baculovirus system, specifically binds to several subterminal fragments of *Ac*. The 11 bp long inverted repeats of the transposable element are not bound by the ORFa protein. Major ORFa protein-binding sites were delineated on 60 and 70 bp long sequence segments that lie 100 bp inside of the 5′ *Ac* terminus and 40 bp inside of the 3′ terminus respectively. Within all strongly bound fragments, and particularly in these 60 or 70 bp long segments, the hexamer motif AAACGG is repeated several times in direct or inverted orientation. The ORFa protein binds to synthetic concatemers of this motif, whereas the mutant motif AAAGGG is not complexed. Methylation of the cytosine residues in the AAACGG motif and/or its complementary strand has pronounced effects: whereas one of the two hemimethylated sequences has a higher affinity to the ORFa protein than both unmethylated and holomethylated DNAs, the other hemimethylated DNA is virtually not complexed at all. The native ORFa protein binding sites are more complex than the AAACGG sequence: certain *Ac* and *Ds1* fragments devoid of AAACGG motifs (but containing several similar sequences) are weakly bound by the ORFa protein.

Key words: Activator (*Ac*)/transposase/DNA- binding protein /DNA methylation

Introduction

For more than 40 years, the transpositional behaviour of transposable elements in maize has been studied genetically in great detail (McClintock, 1947, 1951). Greenblatt and Brink (1962) found evidence that transposition may be coupled to replication. Based on the structures of several transposable element-induced mutants and their revertants, Saedler and Nevers (1985) proposed a molecular model for transposition in plants. Meanwhile, the DNA structure of several plant transposable elements is known (for reviews see Döring and Starlinger, 1986; Fedoroff, 1989a), and the expression of autonomous elements belonging to the two most extensively studied families has been analysed (Pereira *et al.*, 1986; Kunze *et al.*, 1987; Cuypers *et al.*, 1988). However, the transposition mechanism of plant transposable elements is still unknown.

The *Ac* element was cloned and sequenced (Müller-Neumann *et al.*, 1984; Pohlman *et al.*, 1984a,b; Dooner *et al.*, 1988). It is 4565 bp long, has 11 bp long inverted repeats and creates 8 bp duplications in the target DNA upon insertion. Transcription analyses and cDNA cloning led to the detection of only one, 3.5 kb long *Ac*-specific mRNA which contains an 807 amino acid long open reading frame (ORFa) (Kunze *et al.*, 1987; Finnegan *et al.*, 1988). The assumption that this transcript is the only *Ac* product required for transposition is corroborated by the observation that expression of the *Ac* cDNA in transgenic tobacco plants promotes the excision of a *Ds* element (Coupland *et al.*, 1988). Furthermore, in maize carrying a transiently inactivated *Ac* the 3.5 kb mRNA is not detectable (Kunze *et al.*, 1988), whereas in transgenic tobacco plants with an active *Ac* element a transcript indistinguishable from the *Ac* transcript in maize is synthesized (Kunze, 1987; Kunze *et al.*, 1987; Finnegan *et al.*, 1988).

In maize nuclear extracts of lines carrying an active *Ac* a single, rare *Ac*-specific protein (ORFa) is detectable by immunostaining (Fußwinkel *et al.*, 1988). In the presence of one or two copies of *Ac*, different transposition patterns are observed in the aleurone of the developing endosperm. Transpositions occur at a later time during kernel development (McClintock, 1951). If these regulatory properties are mediated by the only known *Ac* protein, this may be achieved by binding of this protein to different sites.

As an initial study of this problem, we have employed

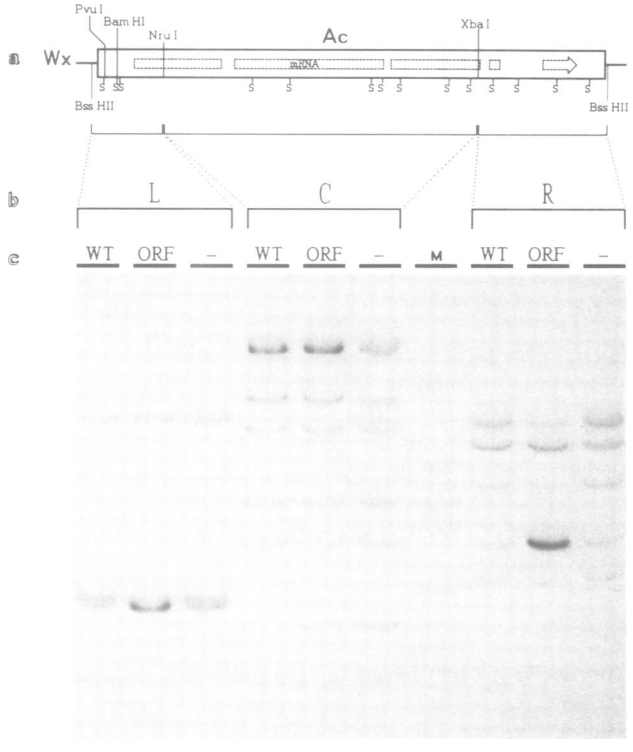

Fig. 1. Immunoprecipitation of *Ac* fragments. **Line a** schematically shows the *Ac* element with the location of its five exons (mRNA) and some restriction sites (S = *Sau*3A). The three sets of labelled *Ac* fragments (L, C and R; **line b**) were incubated with no protein (lanes –), *WT*-extract (lanes WT) or *ORFa*-extract (lanes ORF) (**line c**).

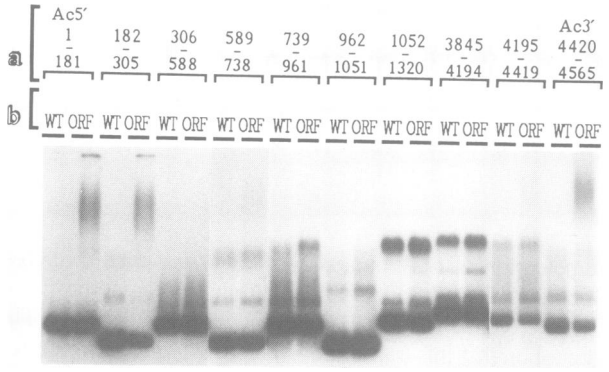

Fig. 2. Gel-retardation assays with *Ac* fragments. The pairs of numbers in **line a** denote the first and last *Ac* nucleotide of the labelled probe fragment. The protein extract added (**line b**) was either *WT*-extract (WT) or *ORFa*-extract (ORF).

Table I. *Ac* fragments tested in gel-retardation assays

Ac fragment		Binding to ORFa protein
1–75	(Ac5'o)	+
76–182	(Ac5'i)	+++
1–182	(Ac5')	+++
182–251		++
182–305		++
252–305		–
306–588		–
589–738		+
739–961		–
962–1051		–
1052–1320		–
1321–1785		–
1786–3390		ND
3391–3556		–
3557–3844		–
3845–4194		–
4195–4419		+
4420–4565	(Ac3')	+++

The pairs of numbers denote the *Ac* nucleotides present in the fragment. In parentheses are the names given to some fragments.
+ weak binding; ++ moderate binding, +++ strong binding; – no binding; ND, not determined.

ORFa protein overproduced in insect cells (Hauser *et al.*, 1988) in DNA binding experiments with different parts of the *Ac* sequence. These studies complement *in vivo* investigations showing that deletion of subterminal fragments at either end of *Ac* decreases or abolishes excision in the presence of an active *Ac* element (Coupland *et al.*, 1988; G.Coupland, C.Plum, S.Chatterjee, A.Post and P.Starlinger, in preparation).

Results

The ORFa protein binds to terminal DNA fragments of Ac

The *Ac* element was dissected into three sets of fragments (L, C and R; Figure 1). These fragments were terminally labelled and incubated with nuclear protein extracts from insect cells infected either with the recombinant baculovirus or with wild-type baculovirus (subsequently abbreviated as *ORFa*- and *WT*-extract respectively). Finally, the samples were immunoprecipitated (see Materials and methods). Two fragments, containing the *Ac* nucleotides 73–181 near the 5' end and 146 bp from the 3' terminus respectively, were enriched in the precipitate of the sample incubated with the *ORFa*-extract (Figure 1). This result suggests that the ORFa protein has a higher affinity to these two DNA fragments than to nearly all other *Ac* fragments. A 21 bp long fragment containing *Ac* nucleotides 182–202 is not represented in this assay.

For further characterization of specific binding of the ORFa protein to *Ac*, gel-retardation assays were performed with a variety of *Ac* fragments. Seventeen fragments covering 1785 bp from the 5' terminus of the element and 1175 bp from the 3' terminus were individually tested for formation of complexes with the ORFa protein. The results obtained with some of them are shown in Figure 2, and a summary of all binding experiments is given in Table I. A slowly migrating, diffuse complex was formed upon incubation of eight *Ac* fragments with the *ORFa*-extract, but not with the *WT*-extract. No such complex is formed between the *ORFa*-extract and the other nine *Ac* fragments or a pUC19 plasmid fragment (Figure 6) respectively. The electrophoretic mobilities of the diffuse complexes are indistinguishable, whereas their intensities vary. Those fragments forming only a low amount of diffuse complex (fragments containing *Ac* nucleotides 1–75, 589–738 and 4195–4419) were not detected in the immunoprecipitation assay. The results of the gel-retardation assays are summarized in Table I.

To test for the specificity of the diffuse complex, unlabelled homologous and non-homologous competitor DNA fragments were included in the binding reactions with the Ac5' and Ac3' fragment. Addition of an ~50-fold excess of either unlabelled Ac5' or Ac3' fragment competed efficiently with the formation of the homologous complex as well as with the complex formed with the other terminus of *Ac*, whereas the admixture of a similar excess of unlabelled pUC19 fragment did not interfere (data not shown). These results support the assumption that the ORFa protein in the *ORFa*-extract binds specifically to sequences in these fragments.

In addition to the *ORFa*-extract-specific diffuse complex, other complexes were observed with nearly all target DNAs tested. However, the faster-migrating complexes producing distinct bands were formed both by ORFa and WT proteins and, therefore, are products of baculovirus-derived or cellular proteins.

The ORFa protein does not bind to the isolated inverted repeats of Ac

The experiments described above demonstrate that the preferred binding sites of the ORFa protein are not the terminal inverted repeats (IR) of the *Ac* element, but they do not rule out a weak binding to them. To test this possibility, the Ac5' fragment was cleaved with *Pvu*I into an outer half (Ac5'o), including the 75 terminal nucleotides of *Ac*, and a 106 nucleotide long inner half (Ac5'i) (Figure 5). In gel-retardation assays the Ac5'i fragment formed a diffuse complex with *ORFa*-extract of similar electrophoretic mobility as the Ac3'/ORFa and the Ac5'/ORFa complexes (Figure 6A). The Ac5'o fragment which contains the IR sequence was also specifically, but weakly complexed by the *ORFa*-extract (Figure 6A). However, this binding was

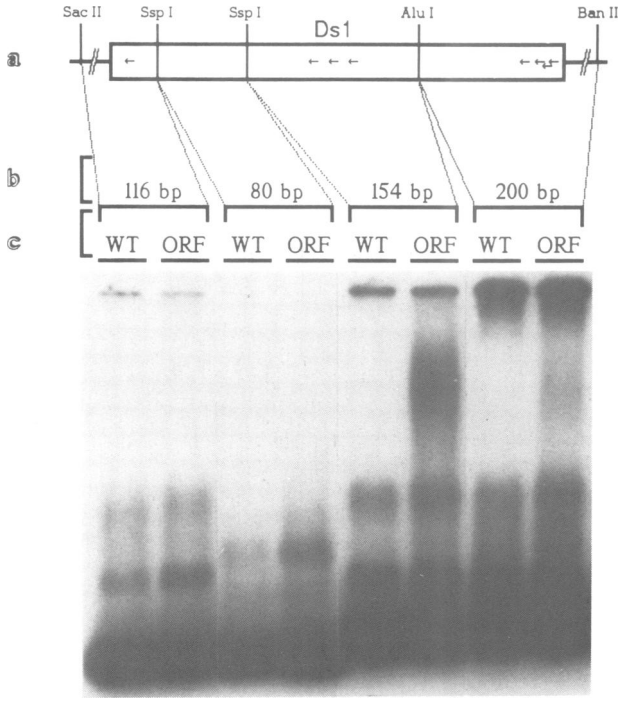

Fig. 3. Gel-retardation assays with *Ds1* fragments. (**Line a**) The *Ds1* element was digested with the indicated enzymes into smaller subfragments, which were employed as probe sequences (**line b**). The samples were mixed either with *WT*-extract (WT) or *ORFa*-extract (ORF) (**line c**). The small arrows within *Ds1* show the positions of hexamer motifs similar to the AAACGG sequence.

insufficient to detect enrichment of this fragment in the immunoprecipitation experiment.

To examine if the IR sequence is responsible for the weak binding of the Ac5′o fragment, DNA fragments containing one to six copies of the 11 bp motif, separated by the four nucleotides GATC from each other and arranged in either direct or inverted orientation, were tested as targets for the ORFa protein in gel-retardation assays. No binding of any of the six DNA fragments by the ORFa protein was detectable (data not shown). We conclude that the ORFa protein does not, at least not *in vitro*, bind to the isolated IRs of the element. Instead, binding site(s) for the ORFa protein are located on several terminal and subterminal restriction fragments of *Ac*.

The ORFa protein binds to Ds1 fragments

The 0.4 kb long *Ds1* element has only 13 bp and 19 bp at either end in common with *Ac* (Sutton *et al.*, 1984). Therefore, we were interested to test whether the DNA of this element interacts with the ORFa protein.

Four *Ds1* fragments were tested as targets for the ORFa protein in gel-retardation assays (Figure 3). Only two fragments, 154 and 200 bp in length, are complexed by the ORFa protein, the longer one with a similarly low efficiency as the Ac5′o fragment. The 154 bp fragment is located internally in *Ds1*: it spans nucleotides 122−276 counted from the left end of *Ds1*. The other fragment includes the 19 bp homologous to the 3′ terminus of *Ac*. No binding of the ORFa protein to the 116 bp fragment, consisting of 74 bp from the *Adh1* gene and 42 terminal nucleotides of *Ds1*, and to the 80 bp fragment is detected.

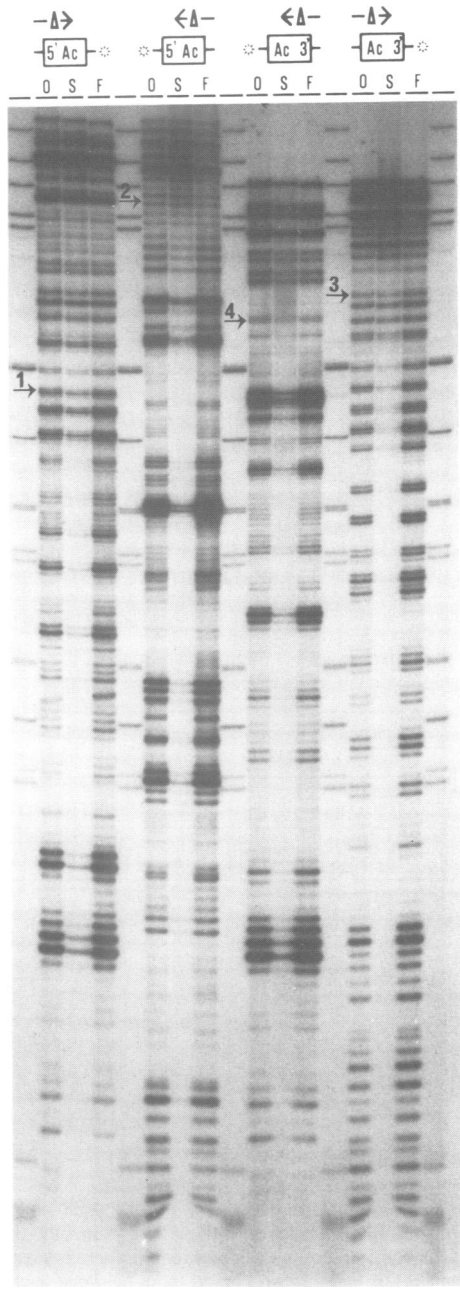

Fig. 4. *Bal*31 deletion analysis of the binding sites of ORFa protein on Ac5′ and Ac3′ fragments. Above each block of three lanes (**O**, **S** and **F**) the radioactively labelled DNA fragments are schematically drawn. The orientation is such that the 5′ terminus of *Ac* in the Ac5′ fragment is located to the left and the 3′ terminus of *Ac* in the Ac3′ fragment is located to the right. The ^{32}P label at one end of each fragment is indicated by a star. The direction of the *Bal*31 digestion extending into each fragment is indicated by the arrows. An aliquot of each digestion was electrophoresed in **lanes O**. The DNA eluted from the diffuse complexes (see Materials and methods) was applied in **lanes S**. The eluted unbound DNA was applied to **lanes F**. All remaining lanes contain pBR322−*Hae*III fragments as size markers; the largest fragment visible on the gel is 267 bp long. The numbers (1−4) beside each lane O corresponds to the borders of the sequence boxes shown in Figure 5.

Determination of the ORFa protein binding sites within the Ac5′ and Ac3′ fragments by Bal31 deletion

We tried to delineate the binding site(s) for the ORFa protein in the Ac5′ and Ac3′ fragments both by DNase I footprinting

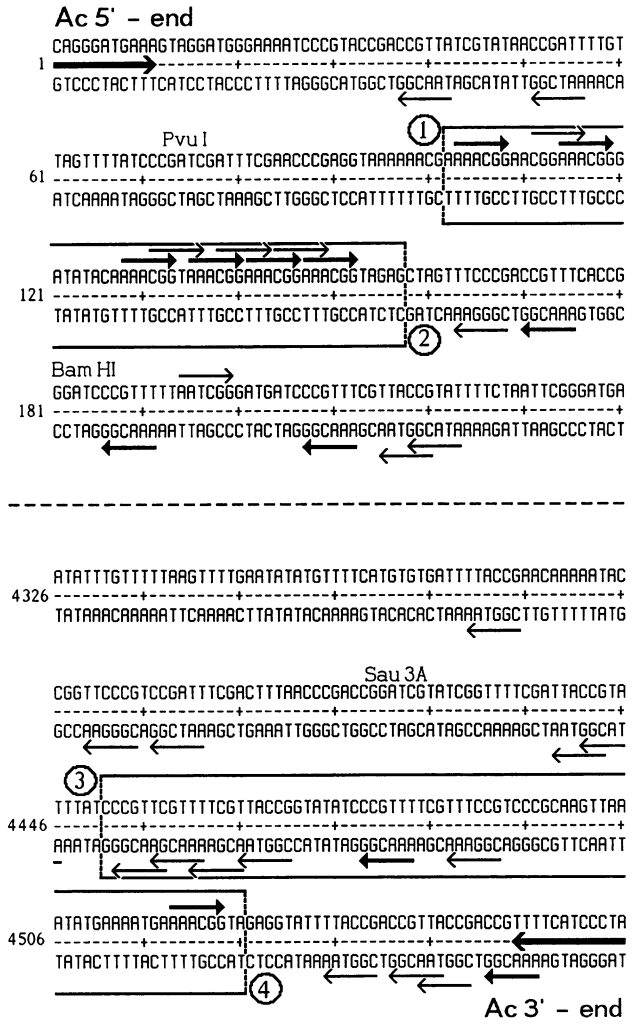

Fig. 5. The 240 5′- and 3′-terminal nucleotides of *Ac* are shown in the upper and lower panels respectively. The Ac5′ fragment employed in the *Bal*31 deletion analysis extends to the *Bam*HI site a position 181. The Ac3′ fragment contains the sequences distal to the *Sau*3A site. The 11 bp inverted repeats of *Ac* are indicated by the two large arrows between the DNA strands. The boxed segments enclose the major ORFa protein binding sequences within the Ac5′ and Ac3′ fragments respectively. Arrows with filled heads show the occurrence of AAACGG motifs. Thin lined arrows indicate hexamers containing the CG dinucleotide and differing by no more than one nucleotide from the AAACGG motif or the permutated version CGGAAA.

and by methylation interference experiments. However, with neither technique did we detect any protectin (interference). Therefore we performed a *Bal*31 deletion analysis (Østergaard Jensen *et al.*, 1988) to characterize binding motifs of the ORFa protein.

Both the Ac5′ and the Ac3′ fragments were labelled at either end and deleted with *Bal*31 from the other end to varying extents (see Materials and methods). The mixtures were used to perform gel-retardation assays, followed by elution of the diffuse complexes and the unbound DNA from the gel and analysis on a sequencing gel (Figure 4). The intensity distribution of the DNA fragments eluted from the complexes does not show an 'all or nothing' effect: fragments shortened below a critical length are progressively more weakly bound. In three of the four assays even very short fragments (<36 bp) were eluted in small amounts from the retarded protein−DNA complexes; however, in these cases we cannot distinguish between specific low efficiency-

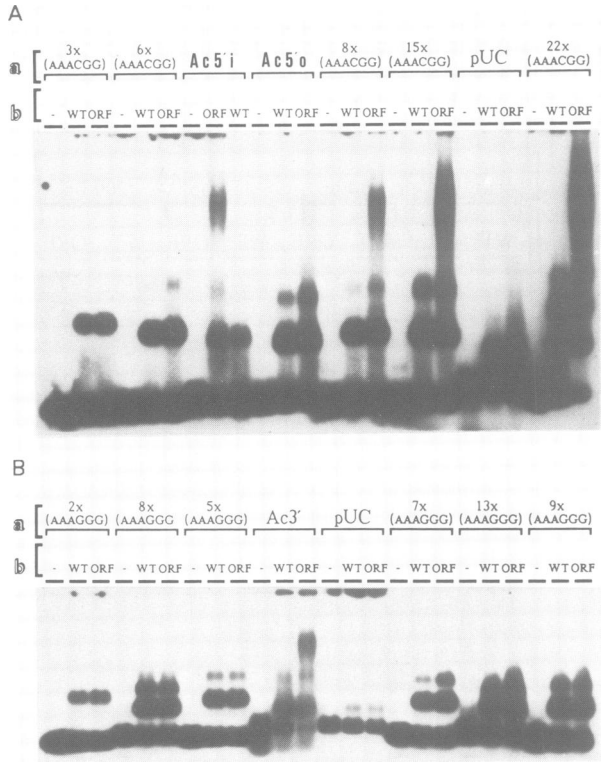

Fig. 6. Gel-retardation experiments with concatemers of the synthetic oligonucleotides AAACGG (**A**) and AAAGGG (**B**). **Line a** on top of each autoradiogram displays the ^{32}P-labelled target DNAs. The Ac3′, Ac5′i, Ac5′o and pUC fragments were included as controls. The factors above the hexamer sequences indicate their copy number in the fragments. The incubation mixture contained either *ORFa*-extract (ORF), *WT*-extract (WT) or no protein (−) (**line b**).

retention and contamination of the complexes with the short fragments. Judged from the exposures of several gels, the positions were determined below which the band intensities of the DNA fragments eluted from the diffuse complexes begin to decrease relative to the unbound fragments (arrows 1−4 in Figure 4).

Figure 5 shows the summary of the experiment: the numbers at the borders of the two boxes in the terminal sequences of *Ac* correspond to the positions indicated in Figure 4. Exonucleolytic degradation beyond these borders precludes or at least strongly decreases inclusion of the truncated fragments in the retarded complex. Thus sequences inside the borders are necessary for binding but their extent remains undetermined. In particular, it cannot be stated whether the left and the right border represents the same binding site. The size of these boxes makes it likely that they represent more than one site and leaves open the function of the central region of the box.

The ORFa protein can bind to a tandemly repeated AAACGG motif

Within all the restriction fragments bound by the ORFa protein and the boxed sequences identified by the *Bal*31 experiments, the hexamers AAACGG or CGGAAA and derivatives of them occur several times, particularly at the 5′ end (Figure 5). In order to determine the possibility of a participation of these sequences in the binding reaction, we tested concatemers of the AAACGG hexanucleotide for their ability to complex with ORFa protein.

Figure 6A displays a gel-retardation experiment with

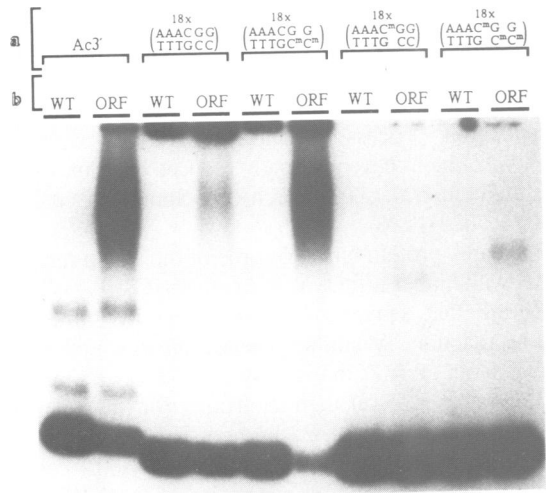

Fig. 7. Gel-retardation experiments with C-methylated versions of the AAACGG motif. (**Line a**) Oligomers composed of 18 hexamer motifs were employed as target DNAs. C-methylation in the upper and/or lower strand is indicated as C^m. The Ac3′ fragment was included as control (**Line b**) The target DNAs were incubated either with *WT*-extract (WT) or *ORFa*-extract (ORF).

oligomers of the AAACGG sequence as targets. The ORFa protein can bind to this sequence motif: *ORFa*-extract, but not *WT*-extract, forms complexes with oligomers of the AAACGG sequence. These complexes appear indistinguishable in electrophoretic mobility from the ones formed between the Ac5′ and Ac3′ probes and the ORFa protein (Figure 6A). Fragments containing three to six copies of the motif in tandem are, however, rather poor targets for the protein. This is noteworthy since the Ac3′ fragment, which contains only three dispersed copies as well as several derivatives of the AAACGG motif, is a much better target. Since the Ac5′o fragment, which does not contain an AAACGG sequence, is also complexed by the ORFa protein, we conclude that tandemly repeated perfect copies of the AAACGG motif are sufficient but not absolutely required for DNA binding of the ORFa protein. The natural ORFa protein binding sites must be more complex in structure than AAACGG oligomers.

With increasing copy number of the AAACGG motif in the target fragments the intensity of the complexes increases, too. If roughly equal molar amounts of the DNA fragments are included in the binding reactions, only traces of the fragment with three, but a large proportion of the fragments containing 22 AAACGG copies are bound by the ORFa protein. Since all binding reactions contain the same quantities of protein extract and unspecific competitor nucleic acid, we conclude that with increasing copy number of the AAACGG sequence the avidity of the ORFa protein binding to the target DNA or the stability of the protein−DNA complexes rises.

As one test for the sequence specificity of the DNA recognition by the ORFa protein, its ability to bind to a derivative of the AAACGG motif in which the central cytosine is replaced by a guanine residue was tested. DNA fragments containing 2−13 tandem copies of the AAAGGG motif were prepared and employed in gel-retardation assays as described above. Figure 6B shows that none of these fragments is complexed by the ORFa protein.

As a second test for sequence-specific binding of ORFa protein to the AAACGG motif, competition experiments with the Ac5′ and Ac3′ fragments were performed. An unlabelled DNA fragment containing 15 tandemly repeated AAACGG motifs interferes with the formation of the diffuse complexes between the ORFa protein and both the Ac5′ and the Ac3′ fragment. The ORFa protein has a higher affinity to the *Ac* termini than to the $(AAACGG)_{15}$ insertion, although they contain fewer copies of the AAACGG motif: an ~100-fold molar excess of unlabelled $(AAACGG)_{15}$ insert is required to reduce complex formed with labelled Ac3′ fragment to ~50%, whereas only a 15-fold excess of unlabelled Ac3′ fragment was sufficient to achieve an equivalent reduction in complex formation with the labelled $(AAACGG)_{15}$ insert (data not shown).

Binding of the ORFa protein to hemi- and holomethylated AAACGG motifs

A peculiarity of the *Ac* element is its capacity to switch from the regular transpositional activity to a transiently inactivated state (McClintock, 1964, 1965). It was found that in the inactive state the *Ac* DNA is modified at particular restriction sites (Schwartz and Dennis, 1986; Chomet *et al.*, 1987), some of which are located close to the ORFa protein binding fragments at the 5′ end of *Ac*. Each AAACGG repeat or the derivatives within this region (Figure 5) contain potential methylation sites (CpG and CpNpG). We therefore tested if the ORFa protein has equal affinities to the four possible hemi-, holo- and unmethylated forms of the AAACGG motif.

The 36mer $(AAACGG)_6$ and its complementary strand was synthesized with either cytosine or with 5-methylcytosine (C^m). Complementary strands were annealed in all four possible combinations. To enhance complex formation, the oligonucleotides were ligated to oligomers prior to incubation with protein extracts. The result of a gel-retardation assay with trimerized 36mers (18xAAACGG) is shown in Figure 7.

The complex formed between the ORFa protein and the unmethylated $(AAACGG)_{18}$ sequence is shown in the fourth lane in Figure 7. Unexpectedly, one of the two hemimethylated AAACGG oligomers is very efficiently complexed by the ORFa protein: when the upper strand is unmodified (5′-AAACGG-3′) and the lower one methylated at both cytosines (5′-C^mC^mGTTT-3′), complex formation increases significantly above the level achieved with unmethylated DNA. In the reciprocal combination, i.e. with the cytosine in the upper strand methylated (5′-AAAC^mGG-3′), but not in the lower strand (5′-CCGTTT-3′), virtually no diffuse complex is formed. When cytosine in both strands is replaced by C^m, a very small amount of the diffuse complex appears again. In each of the two cases where the upper strand (5′-AAAC^mGG-3′) is methylated in the target DNAs, a new protein−DNA complex is formed upon incubation of the DNA with *ORFa*-extract, which migrates as a weak, but distinct band (Figure 7).

We conclude that methylation of the upper strand (5′-AAAC^mGG-3′) manifests two effects: it leads (i) to a strong reduction of formation of the characteristic diffuse complex, and (ii) to the appearance of weak, but defined complexes migrating faster than the diffuse complex. Methylation of the lower strand (5′-C^mC^mGTTT-3′) results in a drastic increase in formation of the diffuse complex. The most likely interpretation of these results is that

the ORFa protein itself has different affinities to the $(AAACGG)_{18}$ sequence which depend on the methylation status of the DNA. However, our experiments do not rule out the participation of other methylation-sensitive protein(s) in formation of the diffuse complex.

Discussion

The ORFa protein binds to subterminal sequences of Ac required in cis for transposition

We have demonstrated in this study by gel-retardation assays that the ORFa protein of the maize *Ac* element, synthesized in insect cells, binds *in vitro* specifically to several *Ac* restriction fragments. The ORFa protein−DNA complexes exhibit some unusual properties: they do not migrate into 3.5% polyacrylamide gels (data not shown), but can be resolved in agarose gels. Here they do not form a defined band but rather a diffuse smear. The electrophoretic mobilities of these complexes are indistinguishable, regardless of which target DNA the ORFa protein is bound to. Therefore we speculate that the complexes are large but of undefined size, and that the reason for this phenomenon is some as yet unknown property of the ORFa protein in the *ORFa*-extract.

The ORFa protein binds *in vitro* primarily to restriction fragments that contain the nucleotides 75−251 near the 5′ end of *Ac* and the 146 3′-terminal nucleotides respectively. By *Bal*31 deletion analysis major recognition sites were localized within these regions between positions 103 and 157, and 4451 and 4525 respectively (Figure 5). No ORFa protein binding of the central *Hin*dIII fragment of *Ac* was detected in the immunoprecipitation assay. This fragment does not contain any sequences required *in cis* for transposition (Coupland *et al.*, 1988).

At the 5′ terminus of *Ac*, a region including the major ORFa protein binding segment (bp 103−157) is absolutely required for transposition: in transgenic tobacco plants a deletion of *Ac* sequences near the 5′ terminus between positions 75 and 181 abolishes or strongly reduces transposability of the *Ac* derivative, indicating that *cis*-acting sequences required for transposition are located in this region (Coupland *et al.*, 1988). Our results are compatible with the assumption that the six copies of the AAACGG motif in the major ORFa protein binding segment, repeated in direct orientation, are (part of) those proposed *cis*-acting sequences (Figure 5): synthetic concatemers of this hexanucleotide are bound by the ORFa protein, mutation of C to G abolishes binding of this sequence, and methylation of the C strongly alters its binding behaviour.

The *Drosophila* transposable element *foldback* (*FB4*) contains near the distal termini of its inverted repeats multiple imperfect copies of the 10 bp sequence CGTTTGCCCA (Potter, 1982). Five nucleotides of this sequence are homologous to the inverted AAACGG (= CCGTTT) motif. The same five nucleotides are found in the repeated 22 bp transposase recognition sites in the ends of phage Mu, and in the ends of the non-viral class II bacterial transposons Tn*3*, gammadelta, Tn*951* and IS101 (Heffron, 1983; Reed *et al.*, 1979), whereas no obvious homologies are found in the sequences of the *Drosophila* transposons *hobo* (Streck *et al.*, 1986) and *P* (O'Hare and Rubin, 1983) and the maize transposable element *En/Spm* (Pereira *et al.*, 1986).

In other respects, however, similarities exist between *En/Spm* and *Ac*: the TnpA protein from *En/Spm* binds to a 12 bp motif which is repeated several times in direct or inverted orientation in both termini of *En/Spm* (Gierl *et al.*, 1988). These regions are likely to be involved in the transposition process, because derivatives with partial deletions of them have a strongly reduced excision frequency (Schiefelbein *et al.*, 1985; Schwarz-Sommer *et al.*, 1985; Tacke *et al.*, 1986).

The TnpA protein binds to some of the 12 bp recognition motifs with higher affinity than to others (Gierl *et al.*, 1988), indicating that sequences in their vicinity influence the binding reaction. A similar observation was made with the TnsB protein of Tn*7*: the affinity of TnsB for a single 22 bp binding motif is lower than its affinity for intact Tn*7* ends, which contains several 22 bp motifs repeated in direct orientation (McKown *et al.*, 1987). Both cases are reminiscent of the binding properties of the ORFa protein to the AAACGG sequence, where a high copy number of the AAACGG motif enhances the affinity of the ORFa protein to the DNA.

The ORFa protein recognizes more than one sequence motif

The three *Ac* fragments which are only weakly complexed by the ORFa protein, among them the Ac5′o fragment (Table I), and the two *Ds1* fragments (Figure 3) do not contain perfect AAACGG motifs, but several derivatives of it retaining the CGG trinucleotide. Thus, the ORFa protein obviously can bind to more than one sequence motif. Possibly, it is able to recognize a family of related sequences carrying a central CpG dinucleotide. Our mutational analysis demonstrated an essential role for the central C residue and the influence of C methylation indicates the importance of the neighbouring G. However, so far no mutational analysis of the other nucleotides has been done, and further investigations will be required to solve this question.

There are some proteins known which can bind families of related sequences or even unrelated sequences. An example is the Oct factors which recognize with only one protein binding domain in addition to the canonical octamer sequence several distantly related sequence motifs (Garcia-Blanco *et al.*, 1989; Kemler *et al.*, 1989). A different situation was described for the transposase of bacteriophage Mu: Leung *et al.* (1989) showed that the N-terminal region of the MuA protein contains two distinct DNA binding domains which recognize different DNA sequences.

Methylation of the AAACGG motif alters the binding characteristic of the ORFa protein

Methylation of either of the two DNA strands of the AAACGG sequence characteristically alters complex formation with the ORFa protein. Replacement of the cytosine residue by 5-methylcytosine in the upper strand ($AAAC^{m}GG$) reduces the amount of the diffuse complex, whereas replacement in the lower strand ($C^{m}C^{m}GTTT$) results in an increase of diffuse complex.

Genetic analyses of transposition events of an *Ac* element inserted in the *P* locus was interpreted by the hypothesis that transposition occurs during replication of the DNA (Greenblatt and Brink, 1962, 1963; Greenblatt, 1984). No biochemical studies of this phenomenon are known. It is possible that the chromatin structure is altered during replication and that this facilitates transposition. It is also

conceivable that recently replicated and yet hemimethylated DNA can be preferentially recognized by the transposase, as was discussed by Fedoroff (1989a).

Preferential transposition of one hemimethylated species of transposable element IS10 of *Escherichia coli* was reported by Roberts *et al.* (1985). If a similar mechanism were operative in *Ac*, some or all of the C residues in its *cis*-acting sequences should always be methylated, as has been described for three *Hpa*II sites located near the 3′ end of *Ac* and the *Pvu*I site close to the 5′ end (Schwartz and Dennis, 1986; Schwartz, 1989).

The *Ac* element can switch from the regular transpositional activity to a transiently inactivated state (McClintock, 1964, 1965) during which it is not transcribed (Kunze *et al.*, 1988) and behaves like a *Ds* element (McClintock, 1964, 1965; Schwartz and Dennis, 1986). Methylation of *Pvu*II and *Hpa*II sites located near the transcription start and mostly within the untranslated leader region of the *Ac* mRNA has been found to be correlated with this temporary inactivation (Schwartz and Dennis, 1986; Chomet *et al.*, 1987). The same is true for the *Bam*HI site at position 181 of *Ac* (Schwartz, 1989). These sites are unmethylated when *Ac* is active and thus no hemimethylated state can be found after their replication. Therefore we consider the possibility that a permanently methylated state of the ORFa protein binding sites is necessary for the cell cycle dependence of transposition, whereas the temporary methylation of restriction sites located downstream from the binding sites is correlated to a lack of transcription. Genomic sequencing experiments will be needed to determine the *in vivo* situation (Church and Gilbert, 1984; Saluz and Jost, 1989).

The phenomenon of inactivation of plant transposable elements correlated to increased levels of C methylation is also known from the *En/Spm* element (McClintock, 1961; Banks *et al.*, 1988; Fedoroff, 1989b), from *Mu1* (Chandler and Walbot, 1986; Bennetzen, 1987) and from *Tam3* (Martin *et al.*, 1989). Gierl *et al.* (1988) demonstrated that the TnpA protein of *En/Spm* binds to a 12 bp motif that contains a CGG trinucleotide. In contrast to the binding properties of the *Ac* ORFa protein, however, the binding efficiency of the TnpA protein to either hemimethylated and the holomethylated form is reduced. It is intersting to note that no genetic studies linking *En/Spm* transposition to replication have been reported.

Do cellular protein factors participate in the transposition reaction?

An *Ac* derivative with a deletion of the *Ac* positions 44−92 was transposition-negative in transgenic tobacco (Coupland *et al.*, 1988). The sequence motifs in the deleted segment are apparently indispensable for transposition, but different from the AAACGG motif, which does not occur between positions 44 and 92. A possible explanation for the deletion derivatives' inability to transpose could be that the altered spacing between the ORFa recognition sequences and the *Ac* 5′ terminus in the deletion derivative interferes with the transposition reaction.

Alternatively, the deleted region could contain a binding site for a cellular factor that is absolutely required for the transposition reaction. A participation of cellular factors in the transposition process was observed with several prokaryotic transposons (Craigie *et al.*, 1985; Phadnis and Berg, 1987; Yin and Reznikoff, 1987; Wiater and Grindley, 1988), and Rio and Rubin (1988) isolated a cellular protein that binds to the terminal inverted repeats of the *Drosophila P* element.

We do not know whether one or several cellular factors are required for *Ac* transposition. If such factors exist in maize, functionally equivalent proteins have to be present also in tobacco (Baker *et al.*, 1986, 1987), tomato (Yoder *et al.*, 1987, 1988), *Arabidopsis*, carrot (Van Sluys *et al.*, 1987), potato (Knapp *et al.*, 1988) and *Petunia* (N.Houba-Herin, personal communication), as in all these plants *Ac* is able to transpose.

Materials and methods

DNA manipulation and plasmid construction

Ac DNA fragments employed as target sequences in gel-retardation assays were either directly excised from plasmid pJAC (Kunze *et al.*, 1987) or subcloned into the *Sma*I site or pUC19 after creating blunt ends with T4 DNA polymerase. *Ds1* fragments were derived from plasmid pDS.2A (Sutton *et al.*, 1984). pUC19 controls were either the 213 bp *Nar*I−*Hin*dIII or the 214 bp *Nde*I−*Eco*RI restriction fragments.

Synthetic oligodeoxynucleotides were purified on denaturing polyacrylamide gels. The 5′-AAACGG-3′ and 5′-AAAGGG-3′ oligodeoxynucleotides were annealed with 5′-GTTTCC-3′ and 5′-CTTTCC-3′ oligodeoxynucleotides respectively, polymerized to oligomers and ligated into the *Sma*I site of pUC19. The oligodeoxynucleotides 5′-GATCTAG-GGATGAAA-3′ and 5′-GATCTTTCATCCCCTA-3′, containing the 11 bp inverted repeats of *Ac*, were annealed and ligated into the *Bam*HI site of pUC19. All inserts were excised from the plasmids with *Eco*RI and *Hin*dIII.

The oligodeoxynucleotide $(AAACGG)_6$ and its complementary strand were synthesized either with the normal cytosine nucleotide or with 5-methylcytosine. The complementary strands were annealed to all four possible combinations, terminally labelled and ligated to oligomers. Trimers (consisting of 18 AAACGG motifs) were gel-purified and employed as target sequences in gel-retardation experiments.

DNA sequence analysis

All plasmid insertions mentioned above were sequenced directly on the double-stranded template by the dideoxy chain termination method (Sanger *et al.*, 1977) using modified T7 DNA polymerase.

Expression of the ORFa protein in baculovirus vectors and preparation of protein extracts

The cloning and expression of the ORFa protein in a baculovirus expression system was described by Hauser *et al.* (1988). Crude nuclear extracts of ORFa-containing and WT baculovirus-infected insect cells were prepared essentially as described by Dignam *et al.* (1983). During the cloning of the *Ac* cDNA into the baculovirus expression vectors a clone was isolated which has a small deletion at the C terminus of ORFa (Hauser *et al.*, 1988). In the protein synthesized in this clone the four authentic C-terminal amino acids were replaced by six different amino acids. We have employed nuclear extracts containing either the ORFa protein with the C-terminal alteration or the authentic ORFa protein respectively, and find that both extracts behave indistinguishably in the gel-retardation assays.

Antisera

The generation of antiserum 'I-3' directed against the ORFa protein was described by Müller-Neumann (1985).

Immunoprecipitation

The *Ac* element was excised as a *Bss*HII fragment from plasmid pJAC and cleaved with *Nru*I, *Xba*I and *Sau*3A into three sets of fragments which were terminally labelled with ^{32}P using Klenow polymerase. Approximately 60 000 c.p.m. per subfragment were incubated in a standard binding reaction (see below) with 4 μg crude nuclear extract for 10 min at room temperature. After addition of 10 μl antiserum 'I-3' the incubation was continued for 30 min on ice. Subsequently, 50 μl of *Staphylococcus aureus* cell walls (Sigma) were added and the incubation continued for another 30 min at 0°C. After centrifugation the pellet was washed, phenol extracted and subsequently the immunoprecipitated nucleic acids were electrophoresed on a 6% native polyacrylamide gel.

Gel-retardation assays

Gel-retardation assays were carried out essentially as described by Garner and Revzin (1981) and Fried and Crothers (1981). Approximately 0.1–0.5 ng labelled DNA fragment (~5000–30000 c.p.m.) was incubated with 1–1.5 μg unspecific competitor nucleic acid [poly(dI-dC) or poly(dA)poly(dT), GRA-buffer (15 mM HEPES, 60 mM NaCl, 3 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 4% glycerol; pH 7.6] and ~1.5 μg crude nuclear extract for 10 min at room temperature. Successively, the mixture was loaded on a 1.2% agarose gel and electrophoresed at 7 V/cm. After electrophoresis the gel was dried onto filter paper and autoradiographed.

Bal31 deletion analysis

*Bal*31 deletion analysis was performed similarly as described by Østergaard Jensen *et al.* (1988). Twenty micrograms of plasmids pRK5 and pRK3, containing the Ac5′ and Ac3′ fragments in the *Sma*I site of pUC19 respectively, were linearized with *Eco*RI (or *Hin*dIII). The DNA was precipitated with isopropanol and resuspended in 200 μl *Bal*31 incubation buffer. Immediately after the addition of 1.5 U of *Bal*31 (Boehringer Mannheim) the solution was incubated at 30°C. Every 20 s 5 μl aliquots were removed and combined with 20 μl 0.5 M EDTA, 0°C. The combined aliquots were adjusted to 0.5 M Nacl, 50 mM MOPS, pH 7.0, and applied to a Quiagen Tip 20 column (Diagen). The column was washed five times with 1 ml buffer A (0.4 M NaCl, 50 mM MOPS, 15% ethanol, pH 7.0), eluted three times with 200 μl buffer F (1.5 M NaCl, 50 mM MOPS, 15% ethanol, pH 7.5) and the eluate precipitated with 480 μl isopropanol. The resuspended DNA was blunt-ended by incubation with T4 DNA polymerase in the presence of 2 mM dNTP and isopropanol precipitated. Subsequently the DNA was digested with *Hin*dIII (or *Eco*RI). One-tenth of the DNA was terminally labelled by Klenow polymerase and the fraction of small molecules separated and purified from the vector sequences by electrophoresis on a native 10% polyacrylamide gel. The mixture was employed in a gel-retardation assay with a low melting point agarose gel and the diffuse complexes were excised. The DNA was eluted from the agarose by melting and successive phenol and chloroform extraction. The solution was adjusted to 100 mM NaCl, 50 mM MOPS, pH 7.0, and applied to a Quiagen Tip 5 column. The column was washed four times with 0.5 ml buffer G (0.1 M NaCl, 50 mM MOPS, 15% ethanol, pH 7.0), the DNA was eluted three times with buffer H (1 M LiCl, 50 mM MOPS, 15% ethanol, pH 7.0) and, after the addition of 4 μg glycogen as carrier, precipitated with 1 vol isopropanol. The DNA pellet was washed with 85% ethanol, dissolved in 96% formamide, 20 mM EDTA and electrophoresed on a 6% sequencing gel.

Acknowledgements

We thank Ulrike Courage for invaluable help during many experiments and Heidi Fusswinkel for the communication of unpublished data. This work was supported by the Deutsche Forschungsgemeinschaft through SFB 274.

References

Baker,B., Schell,J., Lörz,H. and Fedoroff,N. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4844–4848.
Baker,B., Coupland,G., Fedoroff,N., Starlinger,P. and Schell,J. (1987) *EMBO J.*, **6**, 1547–1554.
Banks,J.A., Masson,P. and Fedoroff,N. (1988) *Genes Dev.*, **2**, 1364–1380.
Bennetzen,J.L. (1987) *Mol. Gen. Genet.*, **208**, 45–51.
Chandler,V.L. and Walbot,V. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1767–1771.
Chomet,P.S., Wessler,S. and Dellaporta,S.L. (1987) *EMBO J.*, **6**, 295–302.
Church,G.M. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
Coupland,G., Baker,B., Schell,J. and Starlinger,P. (1988) *EMBO J.*, **7**, 3653–3659.
Craigie,R., Arndt-Jovin,D.J. and Mizuuchi,K. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7570–7574.
Cuypers,H., Dash,S., Peterson,P.A., Saedler,H. and Gierl,A. (1988) *EMBO J.*, **7**, 3653–3659.
Dignam,J.D., Lebowitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1975–1989.
Döring,H.-P. and Starlinger,P. (1986) *Annu. Rev. Genet.*, **20**, 175–200.
Dooner,H.K., English,J. and Ralston,E.J. (1988) *Mol. Gen. Genet.*, **211**, 485–491.
Fedoroff,N. (1989a) In Howe,M.M. and Berg,D.E. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 377–411.
Fedoroff,N. (1989b) *Genetics*, **121**, 591–608.
Finnegan,E.J., Taylor,B.H., Dennis,E.S. and Peacock,W.J. (1988) *Mol. Gen. Genet.*, **212**, 505–509.
Fried,M. and Crothers,D.M. (1981) *Nucleic Acids Res.*, **9**, 6505–6525.
Fußwinkel,H., Müller-Neumann,M., Both,C., Doerfler,W. and Starlinger,P. (1988) *Maize Genet. Coop. Newslett.*, **62**, 47.
Garcia-Blanco,M.A., Clerc,R.G. and Sharp,P.A. (1989) *Genes Dev.*, **3**, 739–745.
Garner,M.M. and Revzin,A. (1981) *Nucleic Acids Res.*, **9**, 3047–3060.
Gierl,A., Lütticke,S. and Saedler,H. (1988) *EMBO J.*, **7**, 4045–4053.
Greenblatt,I.M. (1984) *Genetics*, **108**, 471–485.
Greenblatt,I.M. and Brink,R.A. (1962) *Genetics*, **47**, 489–501.
Greenblatt,I.M. and Brink,R.A. (1963) *Nature*, **197**, 412–413.
Hauser,C., Fußwinkel,H., Li,J., Oellig,C., Kunze,R., Müller-Neumann,M., Heinlein,M., Starlinger,P. and Doerfler,W. (1988) *Mol. Gen. Genet.*, **214**, 373–378.
Heffron,F. (1983) In Shapiro,J.A. (ed.), *Mobile Genetic Elements*. Academic Press, New York, pp. 223–260.
Kemler,I., Schreiber,E., Müller,M.M., Matthias,P. and Schaffner,W. (1989) *EMBO J.*, **8**, 2001–2008.
Knapp,S., Coupland,G., Uhrig,H., Starlinger,P. and Salamini,F. (1988) *Mol. Gen. Genet.*, **213**, 285–290.
Kunze,R. (1987) Ph.D. Thesis, Köln, FRG.
Kunze,R., Stochaj,U., Laufs,J. and Starlinger,P. (1987) *EMBO J.*, **6**, 1555–1563.
Kunze,R., Starlinger,P. and Schwartz,D. (1988) *Mol. Gen. Genet.*, **214**, 325–327.
Leung,P.C., Teplow,D.B. and Harshey,R.M. (1989) *Nature*, **338**, 656–658.
Martin,C., Prescott,A., Lister,C. and MacKay,S. (1989) *EMBO J.*, **8**, 997–1004.
McClintock,B. (1947) *Carnegie Inst. Washington Yearbk*, **46**, 146–152.
McClintock,B. (1951) *Carnegie Inst. Washington Yearbk*, **50**, 174–181.
McClintock,B. (1961) *Carnegie Inst. Washington Yearbk*, **60**, 469–476.
McClintock,B. (1964) *Carnegie Inst. Washington Yearbk*, **63**, 592–602.
McClintock,B. (1965) *Carnegie Inst. Washington Yearbk*, **64**, 527–536.
McKown,R.L., Waddell,C.S., Arciszewska,L.K. and Craig,N.L. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7807–7811.
Müller-Neumann,M. (1985) Ph.D. Thesis, Köln, FRG.
Müller-Neumann,M., Yoder,J.I. and Starlinger,P. (1984) *Mol. Gen. Genet.*, **198**, 19–24.
Østergaard Jensen,E., Marcker,K.A., Schell,J. and de Bruijn,F.J. (1988) *EMBO J.*, **7**, 1265–1271.
O'Hare,K. and Rubin,G.M. (1983) *Cell*, **34**, 25–35.
Pereira,A., Cuypers,H., Gierl,A., Schwarz-Sommer,Z. and Saedler,H. (1986) *EMBO J.*, **5**, 835–841.
Phadnis,S.H. and Berg,D.E. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 9118–9122.
Pohlman,R.F., Fedoroff,N. and Messing,J. (1984a) *Cell*, **37**, 635–643.
Pohlman,R.F., Fedoroff,N. and Messing,J. (1984b) *Cell*, **39**, 417.
Potter,S.S. (1982) *Nature*, **297**, 201–204.
Reed,R., Young,R., Steitz,J.A., Grindley,N. and Guyer,M.(1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4882–4886.
Rio,D.C. and Rubin,G.M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8929–8933.
Roberts,D., Hoopes,B.C., McClure,W.R. and Kleckner,N. (1985) *Cell*, **43**, 117–130.
Saedler,H. and Nevers,P. (1985) *EMBO J.*, **4**, 585–590.
Saluz,H. and Jost,J.-P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2602–2606.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schiefelbein,J.W., Raboy,W., Fedoroff,N.V. and Nelson,O.V. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4783–4787.
Schwartz,D. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2789–2793.
Schwartz,D. and Dennis,E. (1986) *Mol. Gen. Genet.*, **205**, 476–482.
Schwarz-Sommer,Z., Gierl,A., Cuypers,H., Peterson,P.A. and Saedler,H. (1985) *EMBO J.*, **4**, 491–597.
Streck,R.D., MacGaffey,J.E. and Beckendorf,S.K. (1986) *EMBO J.*, **5**, 3615–3623.
Sutton,W.D., Gerlach,W.L., Schwartz,D. and Peacock,W. (1984) *Science*, **223**, 1265–1268.
Tacke,E., Schwarz-Sommer,Z., Peterson,P.A. and Saedler,H. (1986) *Maydica*, **31**, 83–91.

Van Sluys,M.A., Tempe,J. and Fedoroff,N. (1987) *EMBO J.*, **6**, 3881–3889.
Wiater,L.A. and Grindley,N.D.F. (1988) *EMBO J.*, **7**, 1907–1911.
Ying,J.C.P. and Reznikoff,W.S. (1987) *J. Bacteriol.*, **169**, 4637–4645.
Yoder,J., Belzile,F., Alpert,K., Palys,J. and Michelmore,R. (1987) In *Tomato Biotechnology*. Alan R.Liss, New York.
Yoder,J.I., Palys,J., Alpert,K. and Lassner,M. (1988) *Mol. Gen. Genet.*, **213**, 291–296.

Received on June 5, 1989; revised on July 31, 1989