The EMBO Journal vol.10 no.13 pp.4003-4010, 1991

The ORFa protein, the putative transposase of maize transposable element *Ac*, has a basic DNA binding domain

Siegfried Feldmar and Reinhard Kunze

Institut für Genetik, Universität zu Köln, Weyertal 121, W-5000 Köln 41, FRG

Communicated by P.Starlinger

***Ac* encodes the 807 amino acid ORFa protein which binds specifically to multiple AAACGG motifs that are subterminally located in both ends of *Ac*. The wild-type ORFa protein and a number of deletion and amino acid exchange mutants were expressed in *Escherichia coli*, renatured and used for mobility shift assays. At least 136 amino acids from the N-terminus and 537 C-terminal amino acids may be removed from the ORFa protein without destroying the DNA binding domain, whereas a protein starting at amino acid 189 is DNA binding deficient. Certain basic amino acids between positions 190 and 200 are essential for DNA binding, as their substitution with uncharged amino acids leads to the loss of this function. The DNA binding domain of ORFa protein has an overall basic character, but no obvious sequence homology to any other known DNA binding protein. The homologies to the major open reading frames of transposable elements *Tam3* from *Antirrhinum majus* and *Hobo* from *Drosophila* are found between the C-terminal two thirds of the three proteins. The ORFa protein forms discrete complexes with target DNA that appear, depending on the protein concentration, as a 'ladder' of bands on gels, indicating the occupation of target DNA by multiple ORFa protein molecules.**

Key words: Activator/protein renaturation/T7 expression system/transposition/*Zea mays*

Introduction

The maize transposon *Activator* (*Ac*) is an autonomous element that encodes all non-host functions necessary to transpose itself and to mobilize non-autonomous elements *in trans* (McClintock, 1951). Genetic data indicate that it transposes by a non-replicative mechanism and that transposition can be coupled to the replication process (Greenblatt and Brink, 1962). *Ac* has 11 bp long terminal inverted repeats (TIR) and creates 8 bp target site duplications upon insertion (Müller-Neumann *et al.*, 1984; Pohlman *et al.*, 1984). During the excision reaction characteristic footprints are created, whose DNA structures led to the development of molecular models for the transposition mechanism (Saedler and Nevers, 1985; Coen and Carpenter, 1988).

The 4565 bp *Ac* element is transcribed into a 3.5 kb mRNA (Kunze *et al.*, 1987; Finnegan *et al.*, 1988) which encodes an 807 amino acid open reading frame (ORFa). In maize endosperm nuclei a rare protein with an apparent molecular weight of 112 kDa was detected with five antisera directed against separate regions of ORFa. This protein is only found in strains with a biologically active *Ac* (Fusswinkel *et al.*, 1991). Overexpression of ORFa in insect cells results in the synthesis of a protein that is indistinguishable immunologically and in size from the maize protein (Hauser *et al.*, 1988).

The ORFa protein is a likely candidate for the transposase, because expression of the cDNA segment encoding ORFa in transgenic tobacco plants is sufficient to trigger the excision of an *Ac*-dependent, non-autonomous *Ds* element (Coupland *et al.*, 1988). Subsequently, it was shown that the *Ac* ORFa protein, isolated from nuclei of insect cells infected with recombinant baculovirus, can bind specifically to multimeric AAACGG sequence motifs which are located subterminally in both ends of the *Ac* element (Kunze and Starlinger, 1989).

In transgenic tobacco plants the 102 N-terminal amino acids of the ORFa protein are not required for the excision of a *Ds* element (Li and Starlinger, 1990). Furthermore, in a transient expression system in *Petunia* protoplasts, the truncated *Ac* ORFa protein catalyses *Ds* excision more efficiently than the full length ORFa protein (Houba-Hérin *et al.*, 1990).

We have begun to identify the functions and their locations on the ORFa polypeptide chain. Here we show that ORFa protein–DNA complexes can be heterogeneous in size. At least 136 amino acids may be removed from the N-terminus of the protein without affecting its DNA binding function. Essential basic amino acids of the DNA binding domain are located between amino acids 190 and 200. This segment has a basic character, but no sequence homology to any other known DNA binding protein domain is found. However, extensive homologies exist between the C-terminal portions of the *Ac* ORFa protein and the *Tam3* open reading frame (Hehl *et al.*, 1991). Several sections in the C-termini of both proteins also have weaker homologies to the *Drosophila* transposon *Hobo*.

Results

Expression of the Ac ORFa protein in Escherichia coli

The *Ac* ORFa gene product is a rare protein in maize. Its abundance in endosperm cells was estimated to be in the range of 1000 molecules per genome equivalent (Fusswinkel *et al.*, 1991). For initial protein–DNA interaction studies ORFa protein was overexpressed in a baculovirus expression system (Hauser *et al.*, 1988). To facilitate biochemical and mutational analyses, we evaluated here the suitability of an *E.coli* expression system.

The initiation codon of ORFa was mutated into an *Nco*I recognition site, and the complete *Ac* ORFa was cloned as an *Nco*I–*Bam*HI fragment behind the T7 gene 10 promoter in plasmid pET-3d (Studier *et al.*, 1990). This construct expresses an unfused, full length *Ac* ORFa protein with a

threonine-to-alanine exchange of the second amino acid. Based on this clone, several deletion mutants of ORFa protein were constructed (for details see Materials and methods).

The ORFa protein derivatives accumulate almost exclusively in insoluble protein bodies. Figure 1 shows a Coomassie-stained SDS−polyacrylamide gel with solubilized inclusion bodies isolated from some of the *E.coli* strains carrying pET-3d derivatives. By Western blotting it was shown that these protein bodies consist predominantly of recombinant protein (data not shown). In addition to the full length translation products, the inclusion bodies contain varying amounts of shorter recombinant proteins, which are either products of degradation, early termination, or late initiation, respectively.

We noted earlier that the *Ac* ORFa protein has, independently of its source (maize, insect cells, *E.coli*), an aberrantly slow electrophoretic mobility in SDS-containing gels (Fusswinkel *et al.*, 1991). Its apparent molecular weight is ~112 kDa, whereas its calculated molecular weight is only 92 kDa. We find that truncated proteins lacking segments from the C-terminus retain this migration anomaly, whereas proteins lacking the 188 N-terminal amino acids have the expected electrophoretic mobilities. We conclude that the structure of the ORFa protein between amino acids 1 and 188 is responsible for the migration anomaly. The 'PQ-region', a 10-fold repeat of the dipeptide proline−glutamine between amino acids 109 and 128 (Figure 6), does not notably contribute to this effect.

The renatured ORFa protein from E.coli has DNA binding activity

Inclusion bodies isolated from *E.coli* expressing the full length ORFa protein were solubilized in guanidinium chloride, renatured by dilution, and employed in electrophoretic mobility shift assays (MSAs). The termini of *Ac* and oligomeric AAACGG hexamers, but neither AAAGGG motifs nor non-specific pUC19 fragments, are complexed by the renatured protein (Figure 2a, lanes 1−4). Methylation of the single cytosine residue in the upper strand (AAA^{m}CGG) abolishes complex formation whereas methylation of the two cytosines in the lower strand (^{m}C^{m}CGTTT) enhances complex formation (data not shown). Accordingly, the renatured ORFa protein preparation has DNA binding properties indistinguishable from those of the insect cell-derived protein (Kunze and Starlinger, 1989). We cannot exclude the possibility that the smaller polypeptides present in some preparations participate in the DNA binding reaction, however, they visibly affect neither formation nor electrophoretic migration of the complexes.

The ORFa protein preparations from inclusion bodies contain a nuclease activity which becomes apparent upon incubation of target DNAs with increasing amounts of protein. As this activity is removed after purification of the ORFa protein from other components of the inclusion bodies by size-exclusion chromatography under denaturing conditions, we conclude that it is not associated with the ORFa protein itself. The DNA binding properties of the ORFa protein (and truncated derivatives; see below) remain unaltered after such purification (see Figure 5).

These results demonstrate that the *Ac* ORFa protein does not require eukaryote- or plant-cell-specific modifications or a complementing factor for binding to its DNA target.

Deletion mapping of the DNA binding domain

Truncated ORFa protein derivatives were renatured and tested in MSAs for complex formation with ORFa protein target sequences. A mobility shift gel with some of them

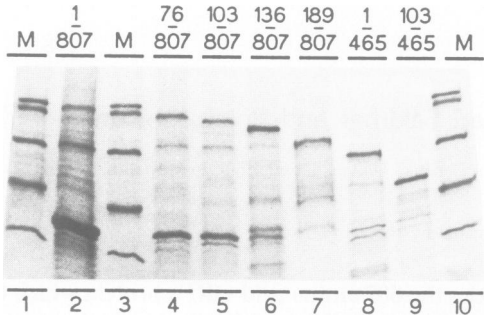

Fig. 1. Coomassie-stained SDS−polyacrylamide gradient gel (4−15%) with solubilized inclusion bodies (~2 μg protein) containing ORFa protein deletion derivatives. Molecular weights of the marker proteins are 116, 97.4, 66, 45 and 29 kDa (lanes 1, 3 and 10). Lane 2, ORFa(1−807); lane 4, ORFa(76−807); lane 5, ORFa(103−807); lane 6, ORFa(136−807); lane 7, ORFa(189−807); lane 8, ORFa(1−465); lane 9, ORFa(103−465).

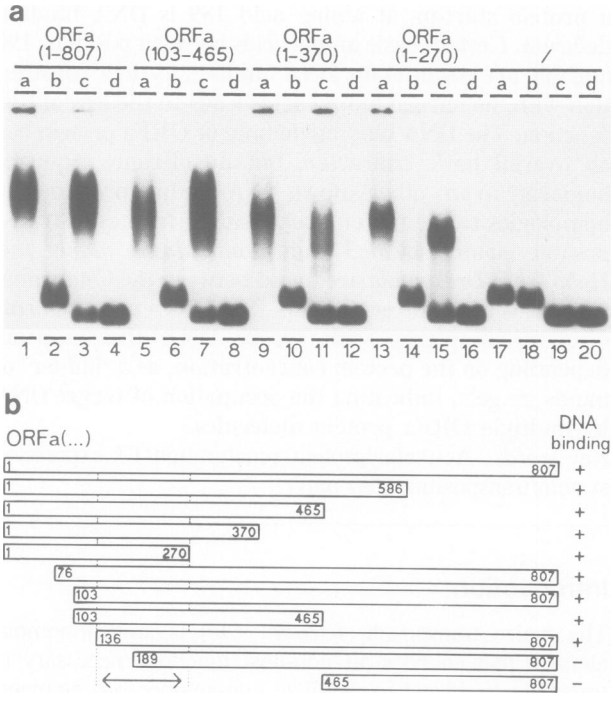

Fig. 2. (**a**) Mobility shift assay with ORFa protein and deletion derivatives. About 0.1 ng (~20 000 c.p.m.) ^{32}P-labelled target DNAs consisting of (a) 185 bp from the 5′-end of *Ac* (lanes 1, 5, 9, 13 and 17); (b) a 242 bp *Hpa*II fragment from pUC19 (lanes 2, 6, 10, 14 and 18); (c) 16 AAACGG hexanucleotides in tandem (lanes 3, 7, 11, 15 and 19); or (d) 14 AAAGGG hexanucleotides in tandem (lanes 4, 8, 12, 16 and 20) were incubated with renatured ORFa protein (lanes 1−4), ORFa(103−465) protein (lanes 5−8), ORFa(1−370) protein (lanes 9−12). ORFa(1−270) protein (lanes 13−16), or no protein (lanes 17−20), respectively. The binding reaction and electrophoresis on an agarose gel were performed under standard conditions (see Materials and methods). (**b**) Summary of the DNA binding tests with ORFa protein deletion derivatives. Formation of specific complexes with the ends of *Ac* or repeated AAACGG motifs in MSAs is indicated by a + to the right of the proteins; a − indicates no binding. The ORFa protein segment between the two dashed lines is sufficient for DNA binding.

is shown in Figure 2a and the experiments are summarized in Figure 2b. At least 135 N-terminal and 537 C-terminal amino acids, respectively, may be removed from the ORFa protein without disrupting the DNA binding domain. However, the deletion of 188 or more amino acids from the N-terminus abolishes DNA binding. All ORFa protein derivatives tested were able to form slowly migrating, diffuse complexes similarly to the wild-type ORFa protein. However, the complexes formed with the truncated derivatives migrated slightly faster. This increase in mobility could be caused by the smaller size of the proteins, by altered aggregation properties, or by a combination thereof, respectively.

The ORFa protein domain sufficient for the recognition of the AAACGG hexameric motifs is located between amino acids 136 and 270. Other protein functions located outside of this region are not required for the protein–DNA interaction.

Basic amino acids are involved in the DNA binding reaction

The *Ac* ORFa amino acid sequence has some conspicuous features near the N-terminus of the protein. The first, ~100 amino acid segment has a rather basic character (17 basic residues versus seven acidic residues). This section is followed by a 10-fold repeat of the dipeptide proline–glutamine or proline–glutamic acid, respectively, between amino acids 109 and 128 (Figure 6). Behind this 'PQ-region' another basic region 48 amino acids in length (amino acids 159–207; 14 basic versus one acidic residues) is situated. In order to investigate whether this basic portion within the DNA binding segment (residues 136–270) is involved in the DNA binding reaction, groups of basic amino acids were deleted or substituted by uncharged amino acids in the ORFa(103–465) protein.

The DNA binding function is neither affected by substitution of lysines 159 and 160, lysines 174 and 176, and histidines 182 and 183 with alanine residues, respectively, nor by deleting amino acids 228–237. The mutant proteins form complexes indistinguishable from those formed with wild-type protein (Figure 3a). An exchange of histidine 191 and arginine 193 with alanine, however, results in a loss of binding to the ends of *Ac* and to the AAACGG motifs. The same result was obtained by substituting histidine 196 and lysine 200 with glutamine. The importance of this segment is confirmed by the lack of DNA binding of mutant ORFa(1–189/197–465), which has amino acids 190–196 deleted (Figure 3a).

These results suggest that the ORFa protein domain responsible for specifically recognizing AAACGG motifs in the ends of *Ac* is located between amino acids 184 and 227, and that the basic residues between positions 190 and 200 are essential for the binding reaction (Figure 3b).

The N-terminus has no DNA binding activity

The *in vivo* analysis of the tobacco protoplast excision assay (Baker *et al.*, 1987; Coupland *et al.*, 1988) of a truncated ORFa protein having 102 amino acids removed from the N-terminus revealed that this derivative is still capable of mobilizing a *Ds* element (Li and Starlinger, 1990). However, its interaction with the target sequences in the ends of *Ac* must be qualitatively altered compared with the wild-type ORFa protein. The wild-type protein can only mobilize with normal frequency a transposable element carrrying at least 250 bp of the *Ac* 5′-end, whereas an element retaining only 186 bp of the *Ac* 5′-end is stable in its presence. In contrast, the ORFa(103–807) protein efficiently mobilizes the transposon with only 186 bp of the *Ac* 5′-end (Li and Starlinger, 1990). Another observation indicating a modulating function of the N-terminus was made by Houba-Hérin *et al.* (1990): in a transient assay in petunia protoplasts the truncated ORFa protein induces excisions of a *Ds* element ~20 times more frequently than the wild-type ORFa protein.

These findings prompted us to test if the 102 N-terminal amino acids of the ORFa protein have themselves the capacity to bind to DNA. The wild-type N-terminus of the ORFa protein was attached to the ATG initiation codon of clone pORFa(103–465/A191/A193). The resulting plasmid expresses the ORFa residues 1–465, but carries the His191→Ala and Arg193→Ala substitutions which lead to an inactivation of the DNA binding domain. In MSAs this protein exhibits no DNA binding activity to any of the DNA fragments tested, containing either 185 bp from the 5′-end of *Ac*, *Ac* nucleotides 148–283 (data not shown), *Ac* nucleotides 180–283 (data not shown), or isolated AAACGG motifs, respectively (Figure 4, lanes 7–8). In contrast, the control construct expressing the ORFa(1–465) protein without amino acid exchanges exhibits normal DNA binding activity (Figure 4, lanes 5–6).

We conclude that the N-terminus does not specifically bind

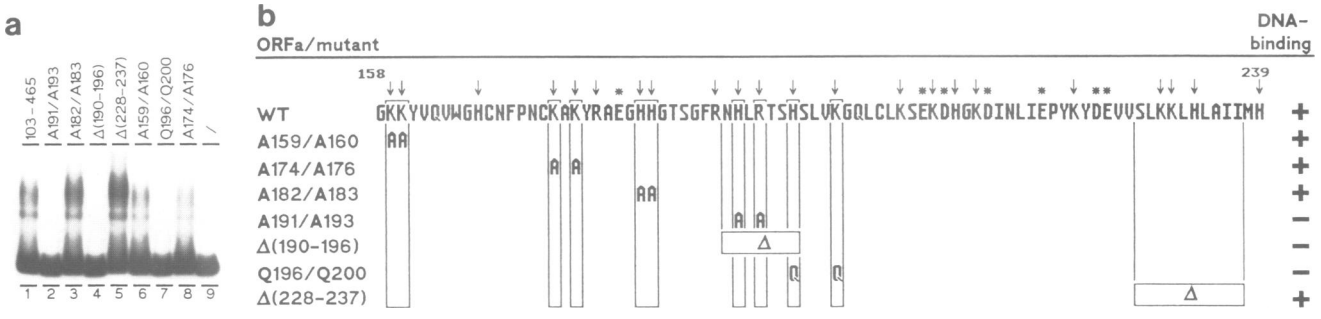

Fig. 3. (**a**) Mobility shift assays with ORFa(103–465) protein and mutants thereof were performed under standard conditions. Target DNA was a ^{32}P-labelled DNA fragment carrying 16 AAACGG hexamers in tandem. Renatured proteins were ORFa(103–465) (lane 1), ORFa(103–465/A191/A193) (lane 2), ORFa(103–465/A182/A183) (lane 3), ORFa(103–189/197–465) (lane 4), ORFa(103–227/238–465) (lane 5), ORFa(103–465/A159/A160) (lane 6), ORFa(103–465/Q196/Q200) (lane 7), ORFa(103–465/A174/A176) (lane 8), no protein (lane 9). (**b**) Effects of mutations introduced into the DNA binding domain of the ORFa(103–465) protein. WT, the wild-type ORFa sequence between amino acids 158 and 239. Basic amino acids are marked by an arrow and acidic residues by an asterisk. The columns show the substituted amino acids of individual mutants. Deletions are depicted by a Δ. The DNA binding properties are indicated to the right of each mutant.

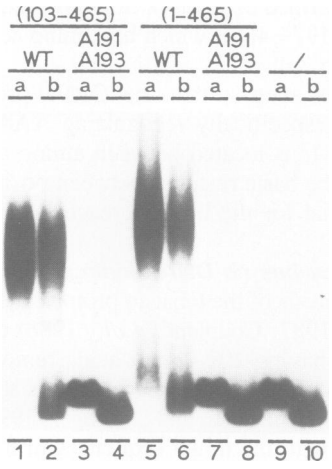

Fig. 4. The N-terminus of ORFa protein has no specific DNA binding function. A mobility shift experiment (standard conditions) is shown. Target DNAs containing either the 185 5′-terminal nucleotides of *Ac* (lanes 1, 3, 5, 7 and 9) or 16 AAACGG hexanucleotides in tandem (lanes 2, 4, 6, 8 and 10) were incubated with renatured ORFa(103−465) protein (lanes 1 and 2), ORFa(103−465/A191/A193) protein (lanes 3 and 4), ORFa(1−465) protein (lanes 5 and 6), ORFa(1−465/A191/A193) protein (lanes 7 and 8), or no protein (lanes 9 and 10), respectively.

to any *Ac* sequences between positions 1 and 283. The observed effects upon deletion of the N-terminus—enhanced transpositional activity and altered requirements for *cis*-acting sequences within the transposable element—are not caused by a specific interaction of the N-terminus with the end of the transposon.

The ORFa protein−DNA complex can be separated into discrete bands

In earlier experiments with ORFa protein-containing nuclear extracts from insect cells it was observed that the ORFa protein forms complexes with a diffuse appearance, which indicates a heterogeneous composition (Kunze and Starlinger, 1989). Under standard reaction conditions the ORFa(1−807) protein and the deletion derivatives used in this study can form similar complexes, too. We examined the complexes formed between the chromatographically purified ORFa(103−465) protein and several target DNAs under varying experimental conditions. Incubation of radiolabelled *Ac* 5′-end or $(AAACGG)_{16}$ fragment with 10 μl protein (the real concentration of active protein is unknown because the renaturation process is incomplete) results in the formation of a slowly migrating, diffuse complex (Figure 5, lanes 4 and 8). A reduction in the amount of protein leads to an increase in mobility of the complex and to its resolution into a 'ladder' of discrete bands that starts with a defined smallest complex band. With smaller target DNAs containing only six or four AAACGG motifs, respectively, a ladder of bands can be obtained too, but not the very slowly migrating complex. Therefore, the large size of the diffuse complexes between the ORFa(103−465) protein and the *Ac* 5′-end or the $(AAACGG)_{16}$ fragment is not caused by a concentration-dependent self-aggregation of protein, but rather by multiple simultaneous protein−DNA interactions.

When using the $(AAACGG)_4$ target, we have repeatedly observed a ladder with only four steps. We are presently investigating if the number of steps is always identical to the number of AAACGG motifs in the target DNA, which would indicate that each motif is bound by a single protein molecule. It may be of interest that the ORFa(103−465) protein apparently has an increased affinity for target DNAs containing a higher number of AAACGG motifs. Although the molar ratios of protein to DNA are nearly constant, 10 μl protein extract are sufficient to complex all of the employed *Ac* 5′-end and $(AAACGG)_{16}$ DNA (lanes 4 and 8), but only a fraction of the $(AAACGG)_4$ DNA (lane 17). Possibly, protein−protein interactions can stabilize the complexes.

Discussion

The migration anomaly of the ORFa protein

The *Ac* ORFa protein synthesized in *E.coli* has very similar properties to the one expressed in insect cells, and to the *Ac* protein identified in maize. It has virtually the same abnormally slow electrophoretic mobility in SDS-containing gels and the same immunological reactivity, it recognizes DNA with the same specificity, and the protein−DNA complexes formed can have a similar diffuse appearance in electrophoretic mobility shift gels as the protein from insect cells. These observations corroborate the assumption that the 112 kDa protein from maize is the full length ORFa protein, and that no eukaryote-specific post-translational protein modification is required for DNA binding.

As bacteria do not faithfully modify eukaryotic proteins, glycosylation and phosphorylation are unlikely to cause the abnormally slow migration in denaturing gels. This conclusion is strengthened by the earlier observation that even an *in vitro* translated fragment of the *Ac* cDNA encoding the 230 N-terminal amino acids of the ORFa protein has an apparent molecular weight of ~39 kDa instead of the calculated 25 kDa (Kunze *et al.*, 1987). Rather, the secondary structure of the N-terminal 188 amino acids of the *Ac* ORFa protein is responsible for the migration anomaly. The 10-fold proline−glutamine/proline−glutamic acid dipeptide repeat between amino acids 109 and 128 (Figure 6) does not play a particular role in this phenomenon.

Reduced electrophoretic mobilities in SDS−polyacrylamide gels were reported for several proteins, among them the DNA binding proteins *Krüppel* (Ollo and Maniatis, 1987), yeast and *Drosophila* HSF (Sorger and Pelham, 1988; Wu *et al.*, 1987) and *Zeste* (Benson and Pirotta, 1987). However, in the cases of *Krüppel* and *Zeste* a high overall or local glutamic acid content is considered to be responsible for this effect. In the *E.coli* TonB protein an N-terminal segment with many proline residues, separated by one amino acid from each other, is supposed to be responsible for its anomalously slow migration (Postle and Good, 1983). This seems not to be the case for the *Ac* protein.

The DNA binding domain of the ORFa protein

Basic amino acids between positions 190 and 200 are required for DNA binding of the ORFa protein. Certain amino acid exchanges within this region abolish DNA binding. We do not yet know if the mutated amino acids are directly involved in the recognition of target DNA, or if another essential protein function is affected, or if both is the case. Hypothetical other functions could be oligomerization or self-aggregation, or interaction with a non-*Ac* protein.

The DNA binding domain of the ORFa protein has no obvious homology to any sequences in the databanks, but

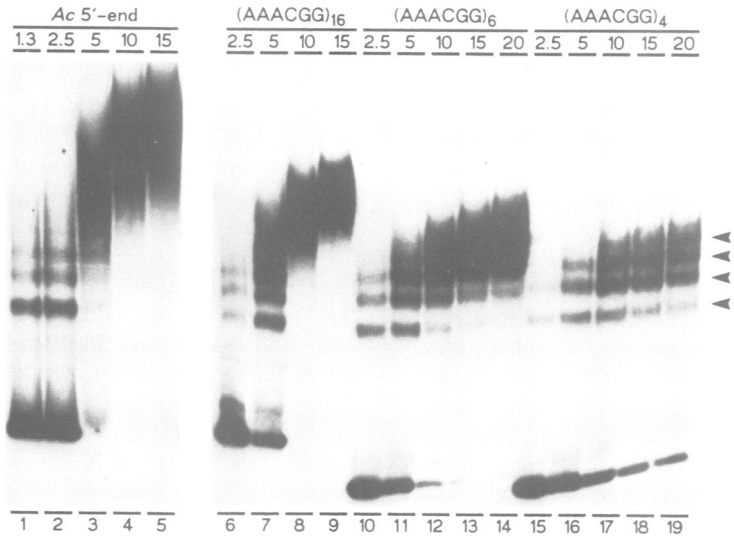

Fig. 5. Composition of the ORFa(103−465) protein−DNA complex. Target DNA (60 000 c.p.m.) containing the 185 5′-terminal nucleotides of *Ac* (lanes 1−5), 16 AAACGG motifs (lanes 6−9), six AAACGG motifs (lanes 10−14) or four AAACGG motifs (lanes 15−19), respectively, were incubated wtih 2 μg poly(dI-dC) and varying amounts of chromatographically purified, renatured ORFa(103−465) protein (indicated in μl on top of each lane) in EMSA buffer. After 10 min incubation at room temperature aliquots of the samples were electrophoresed in a 3.5% polyacrylamide gel, dried and exposed. The arrowheads indicate the four bands seen with the $(AAACGG)_4$ target.

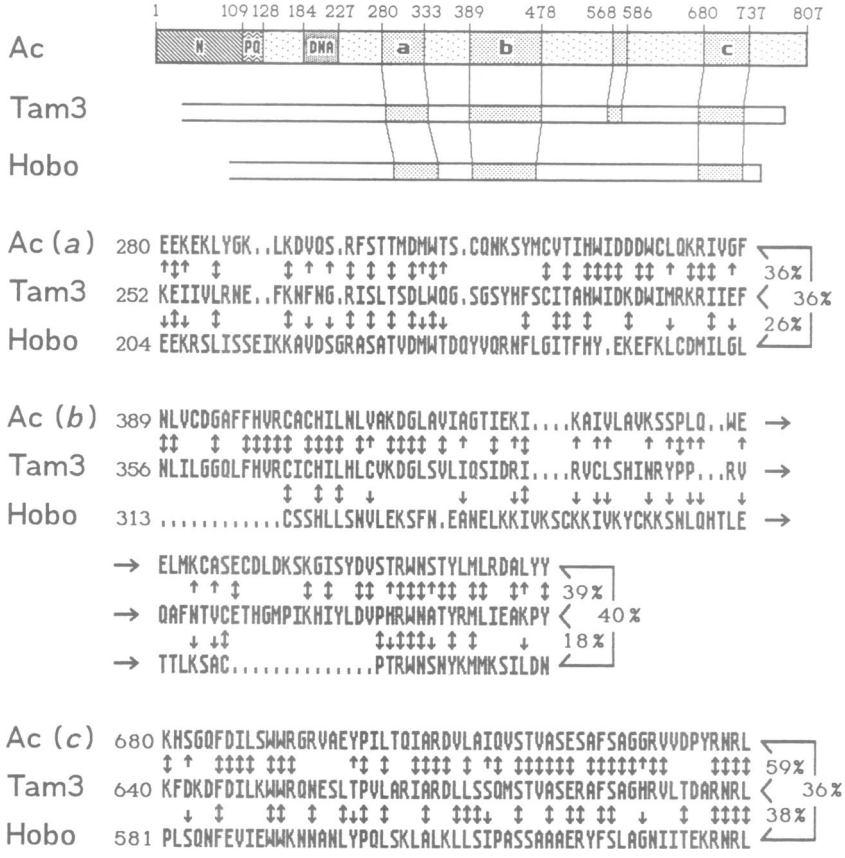

Fig. 6. Diagram of the *Ac* ORFa protein with functional regions and homologies to the open reading frames of *Tam3* and *Hobo*. N depicts the ~100 N-terminal amino acids which may be removed without destroying the transposase function. The sequence of the PQ region is PQPQPQPQPEPQPQPQPEPE. DNA is the DNA binding domain. The amino acid sequences of the segments 'a', 'b' and 'c' of the *Ac* protein (shaded) are shown aligned to the *Tam3* and *Hobo* proteins. Single arrows indicate identical amino acids between the *Ac* protein and the *Hobo* protein. Double arrows indicate identical amino acids between the *Ac* and *Tam3* proteins, and between the *Tam3* and *Hobo* proteins, respectively. The percentage of amino acid identity between the individual sequences is noted to their right.

it shares a basic character with many other DNA binding proteins. Immediately downstream of the essential basic amino acids, between amino acids 214 and 276, a weak homology to a consensus helix−loop−helix structure (Murre *et al.*, 1989) is found. Since the deletion of amino acids 228−237 has no impact on the DNA binding capacity of

the ORFa protein, a functional significance of this homology for DNA binding is rather unlikely.

The structure of the ORFa protein–DNA complex

Our data illustrate that the diffuse appearance of the bands observed earlier (Kunze and Starlinger, 1989) is probably caused by a heterogeneous composition of the ORFa protein–DNA complexes. It must be emphasized that in these earlier experiments with the full length ORFa(1–807) protein, the migration of the complexes was virtually independent of the number of AAACGG motifs on the target DNA and of the protein concentration. Whether these differences are simply due to the smaller size of the ORFa(103–465) protein, thereby facilitating electrophoretic separation, or if particular functions in the protein's C-terminal half were removed is not yet known. The reduced affinity of ORFa(103–465) protein to target DNAs with low numbers of AAACGG motifs, however, reflects the behaviour of the full length protein (Kunze and Starlinger, 1989). It could explain the decline in transposition frequency caused by deletions which remove part of the AAACGG or related motifs from either end of *Ac*, which was described by Coupland *et al.* (1988).

Phage Mu transposase also forms multiple protein–DNA complexes (Kuo *et al.*, 1991). Here, the ladder of bands is caused by monomeric transposase molecules binding to distinct sites on the target DNA. If all binding sites are occupied the size of the complexes does not further increase.

It is assumed that during excision the two ends of the transposable element have to be brought into close proximity. Frey *et al.* (1990) showed that the TNP A protein of the maize *En/Spm* transposable element binds specifically to several recognition sites at either end of the element. They propose that it might serve as a 'glue' for bringing the ends together by non-specific binding of specifically bound TNP A protein to the other transposon end, rather than by protein–protein interaction. The A protein of phage Mu presumably has an equivalent function, but the joining of the ends is mediated by a transient interaction of the protein with an internal operator element which is required for an efficient A protein-dependent synapsis between the ends of Mu (Surette *et al.*, 1987; Craigie and Mizuuchi, 1987; Leung *et al.*, 1989). Protein-mediated DNA loop formation has indeed been observed sometimes when proteins have long range effects on DNA, for instance in the cases of SP1 and the BPV-1 E2 protein (Su *et al.*, 1991; Knight *et al.*, 1991). In both these cases the proteins self-associate to form oligomers. The *Drosophila zeste* protein, which mediates transvection effects, even self-aggregates into complexes of several hundred monomers (Bickel and Pirotta, 1990). Based on the observed DNA binding properties and complex mobility in MSAs, we suggest that the ORFa protein is required for the formation of a complex between the ends of *Ac* by binding to its target sites in either end and joining them through protein–protein interaction.

The N-terminus of the Ac ORFa protein

The 102 N-terminal amino acids of the ORFa protein are required neither for the excision of a *Ds* element (Li and Starlinger, 1990; Houba-Hérin *et al.*, 1990) nor for the DNA binding reaction. Their deletion leads to a 20-fold increased excision frequency of a *Ds* element in the transient assay, indicating an inhibitory role of this segment. To explain the observation that the truncated ORFa protein can mobilize a *Ds* element having only 186 nucleotides of the corresponding *Ac* sequence left at its 5′-end, whereas the full length ORFa protein requires 250 bp of the *Ac* 5′-end, Li and Starlinger (1990) proposed that an interaction of the N-terminus of the ORFa protein with *Ac* DNA between nucleotides 186 and 250 might prevent it from blocking the transposase function of the ORFa protein. We were unable, however, to detect any specific DNA binding capacity of the ORFa N-terminus, although it has been shown earlier that the ORFa protein does in fact bind to *Ac* nucleotides 186–250 (Kunze and Starlinger, 1989). Most likely, the DNA binding domain of the ORFa protein recognizes the five AAACGG or related motifs located within the fragment. We cannot presently exclude the possibility that a DNA binding reaction of the N-terminus may take place only if the protein is simultaneously bound to the AAACGG motifs.

We suggest that the N-terminus of the ORFa protein does not specifically interact with DNA sequences at the 5′-end of *Ac*. Instead, it might reduce the attraction of the ORFa protein to its target sites, for instance by steric hindrance. As a result the truncated ORFa protein could mobilize a transposon having fewer subterminal recognition sites than *Ac*, like the *Ds* element in Li and Starlinger's experiment, that are insufficient for the formation of a functional transposition complex with the full length ORFa protein. In addition, enhanced affinity of the truncated ORFa protein could explain the more efficient mobilization of an *Ac* element that was observed by Houba-Hérin *et al.* (1990).

The 70 kDa ORFa protein derivative in maize

In addition to the 112 kDa ORFa protein in maize endosperm a smaller ORFa derivative with an apparent molecular weight of ~70 kDa occurs that is not recognized by an antiserum directed against the N-terminal 100 amino acids of the ORFa protein (Fusswinkel *et al.*, 1991). If this smaller protein is truncated solely at the N-terminus, we predict that it begins at about amino acid 200 and has lost the AAACGG binding domain. We do not know yet if the 70 kDa protein has an influence on the transposition reaction, for instance if it is involved in the inverse dose effect (McClintock, 1951). A regulatory effect of a truncated transposase is exerted in the case of the *Drosophila* P element, where a C-terminally truncated somatic derivative of the transposase functions as a repressor of transposition (Robertson and Engels, 1989; Misra and Rio, 1990).

The Ac protein has homologies to Tam3 and Hobo

The *Ac* ORFa protein has some conspicuous homologies to the putative protein products of transposable elements *Tam3* from *A.majus* (Hehl *et al.*, 1991) and *Hobo* from *Drosophila melanogaster* (Streck *et al.*, 1986; Calvi *et al.*, 1991). *Ac* and *Tam3* are most similar, as some sections of their proteins have up to 65% identical amino acids (Hehl *et al.*, 1991; Figure 6). As was noted by Calvi *et al.* (1991), the *Hobo* protein seems to be more closely related to *Ac* than to *Tam3*. However, in addition to the similarities at the very C-termini of the three open reading frames (segment 'c' in Figure 6) discussed by Calvi *et al.* (1991) we find sequence similarities between the *Hobo* and *Tam3* open reading frames also in the more centrally located segment 'a' and in part of segment 'b' (Figure 6). All homologies are restricted to the C-terminal two thirds of either protein, whereas the DNA binding

domain of the *Ac* ORFa protein has no homology to any region of the *Tam3* or *Hobo* open reading frames. *Tam3* and *Hobo* have no subterminal clusters of AAACGG or similar motifs. Accordingly, the *Ac* ORFa protein does not bind to any *Tam3* DNA fragment (R.Kunze, unpublished). The homologies between the C-termini of the three proteins suggest a related function, which is not the DNA binding reaction, but possibly a function essential for a transposase. As this putative function could be retained in the 70 kDa ORFa derivative, an influence of this protein on the transposition process is conceivable.

Although complex formation between the ends of the transposon is very likely a prerequisite for the transposition reaction, it possibly is not sufficient to catalyse the steps of excision and reintegration. In prokaryotes, host-encoded proteins participate during the transposition reactions of several mobile elements (for reviews see Howe and Berg, 1989). Kaufman *et al.* (1989) discuss this possibility for the *Drosophila* P-element, whose transposase binds to subterminal motifs and whose TIRs are recognized by a host-encoded protein. Although no host factor binding to *Ac* sequences is known yet, it is conceivable that the *Ac* ORFa protein is active only as part of a multi-component transposition complex. For instance, an as yet unidentified domain of the *Ac* ORFa protein could promote complex formation between the ends of *Ac* as discussed above, whereas another domain might interact with host-encoded proteins. In this model the distribution of *Ac* ORFa protein target sites in the ends of *Ac* determines the appropriate positioning of the transposition complex.

Materials and methods

Plasmids, bacteria and media

We obtained the complete T7 RNA polymerase expression system from F.W.Studier (Rosenberg *et al.*, 1987; Studier *et al.*, 1990). Cloning steps into plasmids pET-3b and pET-3d were performed in *E.coli* strain HMS174. For expression, plasmids were transformed into strain BL21(DE3). Bacteria were grown and induced in media according to the authors' recommendations.

Construction of plasmids

A 1050 bp *Nae*I−*Nar*I fragment and a 1520 bp *Nar*I−*Dra*I fragment from *Ac* cDNA clones pcAcP and pcAcXY (Kunze *et al.*, 1987), respectively, were cloned into the *Sma*I site of pUC19, and a *Bam*HI linker was inserted into the *Sac*I site. Subsequently, a *Nco*I linker (CGCCATGG) was inserted into the *Aha*II site that overlaps with the first ATG codon of the *Ac* cDNA. From this plasmid, pRK19, the *Ac* cDNA was excised as a *Nco*I−*Bam*HI fragment and cloned into the *Nco*I site of the T7 expression vector pET-3d. The resulting plasmid, pORFa, encodes the complete, 807 amino acid *Ac* ORFa, with a Thr → Ala exchange of the second amino acid.

As a common progenitor for the creation of C-terminal deletions of ORFa, plasmid pRL1 was used. In this plasmid the residual 29 *Ac* cDNA-containing nucleotides between the *Bam*HI and *Nco*I sites were deleted. pRL1 was partially digested with *Mae*I and religated with an oligonucleotide (TAGCTAGATCTAGC) that introduces a translation termination codon and a *Bgl*II site into the sequence. From these intermediates the *Nco*I−*Bgl*II fragments were transferred into *Nco*I and *Bam*HI cleaved pET-3d. All plasmids carry the same Thr → Ala amino acid exchange as pORFa. pORFa(1−270) encodes a protein consisting of the N-terminal 270 amino acids of ORFa, followed by a single Ser residue. pORFa(1−370) encodes the 370 N-terminal amino acids. pORFa(1−585) encodes the 585 N-terminal ORFa residues, followed by a serine.

pORFa(1−465) was constructed by exchanging *Asu*II−*Eco*RI fragment from pORFa against the corresponding fragment from pORFa(103−465) (see below).

Several plasmids expressing N-terminal deletions were constructed. pRK19 was partially digested with *Pvu*II and religated with a 12mer *Nco*I linker. The *Nco*I−*Bam*HI fragment from the correct linker-insertion plasmid was inserted into *Nco*I−*Bam*HI-cleaved pET-3d, resulting in plasmid pORFa(76−807). Its reading frame starts with Met−Ala, followed by ORFa amino acids 76−807.

In order to mutate the 10.ATG of the *Ac* cDNA into an *Nco*I site, a cDNA fragment was amplified by PCR between a mismatch-primer (GAGTCAAGATGccATGGCTATTGT) spanning the 10.ATG and a primer (AAACTATGTGATGTTCTCAAGTGA) located downstream of the *Asu*II site in the cDNA (mutations in the *Ac* sequence are indictaed by lower case letters). The amplified fragment was digested with *Nco*I and *Asu*II and exchanged against the longer *Nco*I−*Asu*II fragment in pRK19. pORFa(103−807) was created by inserting the *Nco*I−*Bam*HI fragment from this intermediate into pET-3d.

pORFa(136−807) was made in the same way, except that the mutagenesis primer had the sequence CACCACAGAccAtGGCAAAGAAGT. ORFa amino acid 136 is preceded by a methionine.

pORFa(189−807) was made by linearizing pORFa with *Asu*II, inserting a 10mer *Nco*I linker, and subsequently deleting the *Nco*I fragment. The expressed protein starts with Met−Ala, followed by the indicated ORFa amino acids.

pORFa(465−807) was cloned by filling-in of the *Eco*RI site in the *Ac* cDNA in pORFa, inserting a 10mer *Nco*I linker and deleting the *Nco*I fragment. The protein product of this plasmid contains ORFa amino acids 465−807, preceded by Met−Ala.

pORFa(f11/663−807) expresses ORFa amino acids 663−807 as a C-terminal fusion to 12 amino acids derived from T7 gene 10. It was constructed by cloning the *Bcl*I−*Bam*HI fragment from pRK19 into the *Bam*HI site of pET-3b.

A N- and C-terminally truncated protein is expressed by pORFa(103−465), which is the product of partial cleavage of pORFa(103−807) with *Eco*RI, complete *Bam*HI digestion, filling-in and religation.

Amino acid exchange mutations were introduced into pORFa(103−465) by two-step PCR (Kamman *et al.*, 1989) using pORFa as template, digesting the amplified fragments with *Nco*I and *Asu*II to trim the ends, and ligating them into appropriately opened pORFa(103−465). pORFa(103−465/A159/A160) carries a Lys159→Ala and Lys160→Ala exchange, and was made by amplifications between (i) primers SFo15 (GAGTCAAGATGccATGGCTATTGT) and SFo24 (ACCTGAACGTATgcCgcTCCATCGACCT), and (ii) the product of amplification (i) and primer SFo18 (AAAGTGAGGGCGCAGAGACTTAAC). pORFa(103−465/A174/A176) carries a Lys174→Ala and Lys176→Ala exchange, and was made by amplifications between (i) primers SFo15 and SFo23 (CAGCCCTATACgcAGCCgcGCAATTAGGA), and (ii) the product of (i) and primer SFo18. pORFa(103−465/A182/A183) carries a His182→Ala and His183→Ala mutation. Amplification primers were (i) SFo22 (AGGGCTGAGGGTgcagcTGGAACAAGCGG) and SFo18, and (ii) the product of (i) and SFo15. pORFa(103−465/A191/A193) carries a His191→Ala and Arg193→Ala mutation. Primers were (i) SFo19 (TTTCGAAATgcCTTGgcAACATCACA) and SFo18, and (ii) the product of (i) and SFo15. pORFa(103−465/Q196/Q200) carries a His196→Gln and Lys200→Gln mutation. Primers were (i) SFo20 (ACATCACAaAGTTTAGTTcAAGGTCAG) and SFo18, and (ii) the product of (i) and SFo15.

Internal deletions were introduced by single-step PCR. pORFa(103−189/197−465) carries a deletion of amino acids 190−196. It was made by amplifying between primers SFo25 (ggatttcgaAGTTTAGTTAAAGGTCAGT) and SFo18, and inserting the *Asu*II−*Sca*I digested product into pORFa(103−465) cleaved with the same enzymes. pORFa(103−227/238−465) carries a deletion of amino acids 228−237. A fragment amplified between primers SFo15 and SFo21 (attcatgcaTAACCACTTCATCGTACT) was digested with *Asu*II and *Nsi*I, and subsequently ligated into pORFa(103−465) cleaved with the same enzymes.

pORFa(1−465/A191/A193) was constructed in the same way as pORFa(1−465), except that pORFa(103−465/A191/A193) was taken instead of pORFa(103−465).

Plasmids pORFa(1−465) and pORFa(1−465/A191/A193) were characterized by restriction digests and by analysing the size and immunological reactivity of their expression products. The precursor of pRK19 was completely sequenced. All other plasmids were sequenced across the introduced mutations.

Preparation of inclusion bodies

Fifty ml ZB and 200 μg/ml ampicillin were inoculated with a single colony and grown at 32°C overnight. Prewarmed M9ZB (1 litre), 200 μg/ml ampicillin were inoculated with 2.5 ml overnight culture and shaken at 37°C. After 2 h 2 ml ampicillin (100 mg/ml) were added. At an OD_{600} ~0.6 the culture was induced with IPTG (0.4 mM) and another 2 ml ampicillin (100 mg/ml) were added. The cells were harvested 2−3 h after induction

by centrifugation. One to 3 g cells were resuspended in 15 ml buffer I (10 mM Tris−Cl, 1 mM EDTA, 0.2% Triton X-100, pH 8.3), adjusted to 6 mM $MgCl_2$ and 15 U/ml Benzonase (Merck), and homogenized twice in a French Press under high pressure. Protease inhibitors (0.5 mM PMSF, 1 μg/ml each of aprotinin, leupeptin, pepstatin A and antipain) were added to the lysate, followed by a 15−30 min incubation at 4°C, and 25 min centrifugation at 20 000 *g* (4°C). The pellet was resuspended in 20 ml buffer II (20 mM Tris−Cl, 0.5 M NaCl, 5 mM EDTA, 0.5% Triton X-100, 10% glycerol) with protease inhibitors and centrifuged for another 25 min at 15 000 *g* (4°C). The resulting sediment contains the inclusion bodies. They were suspended in 5 ml buffer II and, after protein concentration determination with the Micro BCA reagent (Pierce), adjusted to ~5 mg/ml.

Chromatographic purification

Recombinant protein was purified from a contaminating nuclease activity by size-exclusion chromatography in a denaturing solvent. Inclusion bodies were dissolved by shaking for at least 1 h at room temperature in denaturation buffer (6 M guanidinium chloride, 0.1 M DTT, 100 mM Tris−Cl, 2 mM EDTA, pH 8.5). After removal of insoluble material by centrifugation, ~1 mg protein was applied to an FPLC-Superose-12 column (Pharmacia) and size fractionated at 0.4 ml/min in denaturation buffer. Eluted fractions (0.25 ml) were analysed by SDS−PAGE. Individual fractions containing ORFa protein (derivatives) were renatured as described below.

Renaturation

Renaturation of the bacterial proteins was performed essentially as described by Jaenicke and Rudolph (1989). Aliquots of the inclusion body suspension were sedimented by centrifugation and dissolved by shaking for at least 1 h at room temperature in denaturation buffer to a final concentration of 2.5 mg/ml. After removal of insoluble components by centrifugation, aliquots were diluted 1:100 or 1:50 with ice-cold renaturation buffer (50 mM Tris−Cl, 50 mM NaCl, 3 mM $MgCl_2$, 5 μM $ZnCl_2$, 0.2% Triton X-100, 5 mM glutathione-reduced, 0.5 mM glutathione-oxidized, 10% glycerol, pH 8.5) and incubated overnight at 4°C.

Mobility shift assays

MSAs were carried out similarly as described by Kunze and Starlinger (1989). Between 0.1 ng and 0.3 ng terminally ^{32}P-labelled target DNA fragment (20−60 000 c.p.m.) were incubated for 10 min at room temperature with varying amounts of renatured protein and 0.5−4 μg poly(dI-dC) in EMSA buffer (15 mM Tris−Cl, 60 mM NaCl, 3 mM $NgCl_2$, 2 μM $ZnCl_2$, 10 mM 2-mercaptoethanol, 0.04% Triton X-100, 4% glycerol, pH 8.3). The samples were electrophoresed on native, low ionic strength 1.2% agarose gels (7.8 mM Tris, 3.3 mM Na-acetate, 1 mM EDTA, pH 8.0) or 3.5% polyacrylamide gels (17.5 mM Tris, 13.5 mM glycine, 1 mM EDTA, pH 8.5) at 7 V/cm. Subsequently, gels were dried onto 3MM paper and autoradiographed.

Acknowledgements

We thank William F.Studier for the T7 expression system, Rainer Rudolph for advice about protein renaturation, the colleagues in Benno Müller-Hill's department for oligonucleotide synthesis, and Brigitte Kisters-Woike for help with computer analyses. We are especially grateful to Peter Starlinger for his constant support, fruitful discussions and suggestions. This work was supported by Deutsche Forschungsgemeinschaft through SFB274.

References

Baker,B., Coupland,G., Fedoroff,N., Starlinger,P. and Schell,J. (1987) *EMBO J.*, **6**, 1547−1555.
Benson,M. and Pirotta,V. (1987) *EMBO J.*, **6**, 1387−1392.
Bickel,S. and Pirotta,V. (1990) *EMBO J.*, **9**, 2959−2967.
Calvi,B.R., Hong,T.J., Findley,S.D. and Gelbart,W.M. (1991) *Cell*, **66**, 465−471.
Coen,E.S. and Carpenter,R. (1988) *EMBO J.*, **7**, 877−883.
Coupland,G., Baker,B., Schell,J. and Starlinger,P. (1988) *EMBO J.*, **7**, 3653−3659.
Craigie,R. and Mizuuchi,K. (1987) *Cell*, **51**, 493−501.
Finnegan,E.J., Taylor,B.H., Dennis,E.S. and Peacock,W.J. (1988) *Mol. Gen. Genet.*, **212**, 505−509.
Frey,M., Reinecke,J., Grant,S., Saedler,H. and Gierl,A. (1990) *EMBO J.*, **9**, 4037−4044.
Fusswinkel,H., Schein,S., Courage,U., Starlinger,P. and Kunze,R. (1991) *Mol. Gen. Genet.*, **225**, 186−192.
Greenblatt,I.M. and Brink,R.A. (1962) *Genetics*, **47**, 489−501.
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.
Hehl,R., Nacken,W.K.F., Krause,A., Saedler,H. and Sommer,H. (1991) *Plant Mol. Biol.*, **16**, 369−371.
Houba-Hérin,N., Becker,D., Post,A., Larondelle,Y. and Starlinger,P. (1990) *Mol. Gen. Genet.*, **224**, 17−23.
Howe,M.M. and Berg,D.E. eds (1989) *Mobile DNA*. American Society for Microbiology, Washington, DC.
Jaenicke,R. and Rudolph,R. (1989) In Creighton,T.E. (ed.), *Protein Structure: a Practical Approach*. IRL Press, Oxford, pp 191−223.
Kamman,K., Laufs,J., Schell,J. and Gronenborn,B. (1989) *Nucleic Acids Res.*, **17**, 5404.
Kaufman,P.D., Doll,R.F. and Rio,D.C. (1989) *Cell*, **59**, 359−371.
Knight,J.D., Li,R. and Botchan,M. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3204−3208.
Kunze,R. and Starlinger,P. (1989) *EMBO J.*, **8**, 3177−3185.
Kunze,R., Stochaj,U., Laufs,J. and Starlinger,P. (1987) *EMBO J.*, **6**, 1555−1563.
Kuo,C.-F., Zou,A., Jayaram,M., Getzoff,E. and Harshey,R. (1991) *EMBO J.*, **10**, 1585−1591.
Leung,P.C., Teplow,D.B. and Harshey,R.M. (1989) *Nature*, **338**, 656−658.
Li,M.-g. and Starlinger,P. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 6044−6048.
McClintock,B. (1951) *Cold Spring Harbor Symp. Quant. Biol.*, **16**, 13−47.
Misra,S. and Rio,D.C. (1990) *Cell*, **62**, 269−284.
Müller-Neumann,M., Yoder,J.I. and Starlinger,P. (1984) *Mol. Gen. Genet.*, **198**, 19−24.
Murre,C., Schonleber McCaw,P. and Baltimore,D. (1989) *Cell*, **56**, 777−783.
Ollo,R. and Maniatis,R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5700−5704.
Pohlman,R.F., Fedoroff,N. and Messing,J. (1984) *Cell*, **37**, 635−643.
Postle,K. and Good,R.F. (1983) *Proc. Natl. Acad. Sci. USA*, **83**, 5235−5239.
Robertson,H.M. and Engels,W.R. (1989) *Genetics*, **123**, 815−824.
Rosenberg,A.H., Lade,B.N., Chui,D.-s., Lin,S.-W., Dunn,J.J. and Studier,F.W. (1987) *Gene*, **56**, 125−135.
Saedler,H. and Nevers,P. (1985) *EMBO J.*, **4**, 585−590.
Sorger,P.K. and Pelhan,H.R.B. (1988) *Cell*, **54**, 855−864.
Streck,R.D., MacGaffey,J.E. and Beckendorf,S.K. (1986) *EMBO J.*, **5**, 3615−3623.
Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60−89.
Su,W., Jackson,S., Tjian,R. and Echols,H. (1991) *Genes Dev.*, **5**, 820−826.
Surette,M.G., Buch,S.J. and Chaconas,G. (1987) *Cell*, **49**, 253−262.
Wu,C., Wilson,S., Walker,B., Dawid,I., Paisley,T., Zimarino,V. and Ueda,H. (1987) *Science*, **238**, 1247−1253.

Received on August 14, 1991; revised on September 23, 1991