Mol Gen Genet (1984) 198:19–24

MGG

The DNA sequence of the transposable element *Ac* of *Zea mays* L.

M. Müller-Neumann, J.I. Yoder, and P. Starlinger
Institut für Genetik, Universität zu Köln, Weyertal 121, D-5000 Köln 41, Federal Republic of Germany

Summary. The sequence of the *Ac* element isolated from the *wx-m7* allele has been determined. The *Ac* element is 4563 bp long. A central portion of roughly 3.1 kb is occupied by three open reading frames, two of which point in one direction and the third in the opposite direction. One of the reading frames potentially encodes a protein with a ten-fold repeat of pro gluN and pro glu dipeptides near its N-terminus. The sequences outside the open reading frames are characterized by the presence of a number of direct and inverted repeats. The *Ac* element may thus have evolved from a simpler progenitor structure. The sequence we have determined for the *Ac* from the *wx-m7* allele differs in a few key positions from that reported for the *Ac* element from the *wx-m9* allele (Pohlman et al. 1984). We have resequenced these positions in both *Ac* elements and find them to be identical. We conclude that the phenotypic differences between the two *waxy* alleles are not caused by structural differences in the *Ac* elements but rather may be attributable to the differences in their insertion sites.

Introduction

The *Ac* element is a member of the *Activator/Dissociation* (*Ac/Ds*) family of maize transposable elements discovered and characterized by McClintock (1949, 1951, 1965). *Ac* is the autonomous element of the family. It exerts the functions necessary for its own transposition as well as for the transposition of the nonautonomous *Ds* elements. In addition, *Ac* elements are capable of causing chromosome breaks at the site of some *Ds* elements, but not at the site of their own integration (McClintock 1949, 1964). If either *Ac* or *Ds* elements are inserted in a gene, a recessive mutation can result. Excision of the element from this site often leads to the restoration of the dominant wild-type phenotype. If the reversion events occur during endosperm development in genes that lead to phenotypes visible on the kernel, the timing and frequency of these events can be inferred from the variegation pattern. Changes in the variegation patterns are occasionally observed and can, in some cases, be found linked to either the *Ac* or to the *Ds* elements (McClintock 1951, 1965). In some cases, *Ac* activity disappears for one or several generations and reappears later. McClintock designated these alterations in *Ds* and *Ac* action as "changes in state" and "changes in phase", respectively (McClintock 1949, 1964, 1978a). The frequency and reversibility of these alterations makes it unlikely that they are caused by point mutations. It would be interesting to know whether they are caused by reversible internal sequence rearrangements of the elements. From the sequence of the *Ds* element inserted in the *sh-m5933* allele of the *Sh* gene encoding sucrose synthase, it was already known that many direct and fewer inverted repeats are present (Döring et al. 1984). In *Escherichia coli*, short direct repeats can be the endpoints of deletions (Farabaugh et al. 1978). Also, inverted repeats can be the breakpoints of reversible inversions, as found in bacteriophage Mu (Kamp et al. 1979) or in the gene encoding flagellar antigens in *Salmonella* (Silverman and Simon 1983). Molecular studies with the *Ac/Ds* family have been reviewed recently (Courage et al. 1984; Döring and Starlinger 1984).

The *Waxy* gene encodes a starch granule-bound UDPG: starch glucosyl transferase, which catalyses amylose synthesis (Nelson and Rines 1962). Two independent insertions of an *Ac* element into the *Waxy* locus (McClintock 1963, 1964) have recently been cloned, *Ac9* and *Ac7* (Fedoroff et al. 1983; Behrens et al. 1984). The *wx-m7* allele carries *Ac* near the 5′ end of the gene, while the *Ac* element in the *wx-m9* allele is inserted in an exon 2.5 kb downstream from this site (Fedoroff et al. 1983; Behrens et al. 1984; A. Gierl and S. Schwarz-Sommer, personal communication). The pattern of somatic reversions from the *waxy* phenotype to the *Waxy* wild-type phenotype is different in the two alleles. In homozygotes carrying three doses of the *Ac* element in the endosperm, reversion events occur earlier in development with the *wx-m7* than with the *wx-m9* allele (O. Nelson, personal communication). In addition, *wx-m7* is not a null mutation. The amylose synthesized in this mutant is distributed in the endosperm in a characteristic pattern (McClintock 1978b).

If the sequences of the *Ac* elements from the *wx-m7* and *wx-m9* alleles should differ, this might explain the phenotypic differences observed. The sequence of the *Ac* element from *wx-m9* has been determined previously (Pohlman et al. 1984). We report here the sequence of the *Ac* element of the *wx-m7* allele and find it virtually identical to *Ac* from *wx-m9*, as we cannot confirm small differences between our sequence and that published by Pohlman et al., 1984. Functional differences found between the two alleles might therefore be sought in the different sites of insertion.

Offprint requests to: M. Müller-Neumann

Materials and methods

Restriction enzymes, T4 DNA ligase, alkaline phosphatase, polynucleotide kinase, DNA polymerase Klenow fragment and other enzymes used were obtained from BRL (Neu-Isenburg), Boehringer (Mannheim) or Biolabs (Schalbach) and used as recommended by the suppliers.

Radioactive chemicals were obtained from Amersham (Braunschweig).

Bacteria, plasmids and plasmid constructions. From the lambda phage carrying a 14.2 maize DNA fragment containing the *Ac* element of the *wx-m7* allele (Behrens et al. 1984), a 6.0 kb *Bgl*II/*Sal*I fragment (pAC7), and its *Eco*RI cleavage fragments, a 3.2 kb *Bgl*II/*Eco*RI fragment (pAC710) and a 2.8 kb *Eco*RI/*Sal*I fragment (pAC720) were ligated into appropriate sites of pBR322. pAC720 was linearized with *Bam*HI and digested with nuclease *Bal*31 for periods of different duration (Legersky et al. 1978). *Cla*I linkers were ligated to the digestion products, circularized after *Cla*I-cleavage, and introduced into HB101 cells (Boyer and Roulland-Dussoix 1969) by transformation. The size of the deletions was determined by restriction analysis and candidates carrying deletions of appropriate size were used for further studies. pAC7 was also prepared from GM33 (dam⁻) cells (Marinus et al. 1973) to achieve digestion with those enzymes that are sensitive to adenine methylation.

DNA sequences were determined by the chemical degradation method fo Maxam and Gilbert (1977), using the gel modification of Ansorge and Baker (1984).

Dot plot analysis (Maizel and Lenk 1981) and the application of Fickett's test code for probability that an ORF encodes a protein (Fickett 1982) were performed on a VAX computer (Devereux et al. 1984).

Standard molecular biology methods. Gel electrophoresis, preparation of plasmids in small or large amounts, transformation etc. were performed as described by Maniatis et al. (1982), sometimes with minor modifications.

Results

The *Ac* element cloned from the *wx-m7* allele is located on a 6 kb *Bgl*II-*Sal*I fragment. We have determined its sequence by the chemical degradation procedure described by Maxam and Gilbert (1977). Most of the sequence was determined from conveniently located restriction sites. In those regions of the sequence where this was not possible, we introduced deletions by *Bal*31 digestion of DNA cleaved at an appropriately located restriction site. Sequences in the central part were determined from both strands and the restriction sites used for labeling DNA fragments were overread in all cases by labeling at a nearby site. A region of approximately 1 kb at both termini is virtually identical to the sequence determined previously for the *Ds* element of the *sh-m5933* allele of the *Shrunken* gene (Döring et al. 1984). In this region, we have determined part of the sequence from one strand only. We observed three differences between the *Ac7* element and the *Ds* element from the *sh-m5933* allele, however only *Ac7* was resequenced in these positions. The 12 differences between the sequence of *Ac7* and that determined by Pohlman et al. (1984) for *Ac9* were extensively examined. This involved determining particular sequences by the dideoxy method, running sequencing gels at 70° C, and end-labeling of DNA fragments at different distances from the site to be checked. We also resequenced these sites (with the exception of positions 3964, 4134, and 4442) in *Ac9* and found them to be identical to *Ac7*.

The sequence of *Ac7* is shown in Fig. 1.

In the sequence, three open reading frames (ORF) were detected, ORF1 (A_{1729}TG to T_{1067}TGA) and ORF2 (A_{3165}TG to T_{1885}TAA) both read to the left, and ORF3 (A_{3739}TG to G_{4191}TGA) reads to the right (Fig. 2).

Sequences possibly qualifying as TATA boxes were found 5′ of ORF1 (T_{1806}ATTTTA), 5′ of ORF2 (T_{3228}TTATTCA) and 5′ of ORF3 (T_{3557}ATTTCCAA). [Note that in Fig. 1 only one strand of the *Ac* sequence is shown, and that ORF1 and ORF2 are encoded by the complementary strand (Fig. 2).]

In searching for polyadenylation sites, we looked for the sequence AATAAA or GATAAA or related sequences (Messing et al. 1983), possibly followed at a distance of 15–20 bp by the sequence CAPyTG (Berget 1984). At the end of ORF1, the sequence A_{1026}ACAAT is found, followed by the sequence C_{996}TCTG.

At the end of ORF2, we observe G_{1784}ATAAA, followed by C_{1765}AATG and by C_{1753}ATG. ORF3 is followed by the sequences G_{4494}ATAAA and G_{4520}ATAA, respectively, neither being followed by the sequence CAPyTG or a close derivative. The resemblance of these sequences to the consensus sequences is not as close as that of A_{2683}ATAAA, followed at a distance of 10 bp by C_{2668}ACTG, again followed 5 bp downstream by A_{2659}AAA in the center of ORF2.

At the beginning of each of the open reading frames we find the translation initiation codon ATG in frame. When we compare these to the consensus sequence A/GNNATGG (Kozak 1981; Messing 1983), we see that the −3 and +4 positions of the met codons are in good agreement for ORF1 (GGAGAATGG) and ORF2 (ATGATATGG), but not for ORF3 (CACAAATGA).

Upon integration into the *waxy* gene, *Ac7* creates an 8 bp direct duplication of host DNA, as can be seen by comparison with the corresponding sequence of the wild-type locus (A. Gierl and S. Schwarz-Sommer, personal communication).

An 11 bp imperfect inverted repeat of the sequence $\frac{T}{C}$AGGGATGAAA is found at the termini of *Ac7*.

Using the dot plot matrix method of Maizel and Lenk (1981), we have inspected the *Ac* sequence for additional direct and inverted repeats (Fig. 3). Clusters of direct repeats are detected in the terminal 1 kb sequences rather than in the center of *Ac7*. Local inverted repeats are scattered throughout the element. In addition, sequences at one end of the element are inverted with respect to sequences at the opposite end, as indicated in Fig. 4.

The base composition of *Ac7* is inhomogeneous. ORF1 and ORF2 have 64% and 61% A+T respectively, and less CpG dinucleotides than expected on a random basis. ORF3 has only 28% A+T and more CpG dinucleotides than expected. The rest of the sequence contains 58% A+T.

Discussion

Ac exerts some functions on *Ds* elements *in trans*. A deletion entirely located in ORF1 (Fedoroff et al. 1983) converts

Fig. 1. Sequence of Ac. Only one strand is shown. ORF1 and ORF2 are encoded by the complementary strand. The translation start and stop signals, TATA boxes and polyadenylation sites (AATAAA or GATAAA) of the three ORF are indicated. In addition the sequence CAPyGT (Berget 1984) is also indicated, if this sequence or a derivative of it is found within the first 30 bp downstream

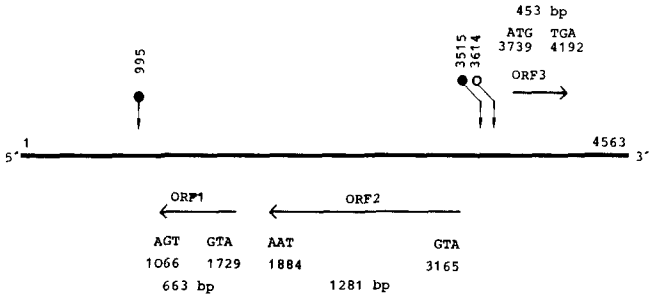

Fig. 2. Block diagram of *Ac*. The open reading frames are indicated by *arrows*. The region between the *solid circles* is deleted in double-*Ds* (Döring et al. 1984). The *open circle* indicates the site of insertion of the central *Ds* element in double-*Ds*

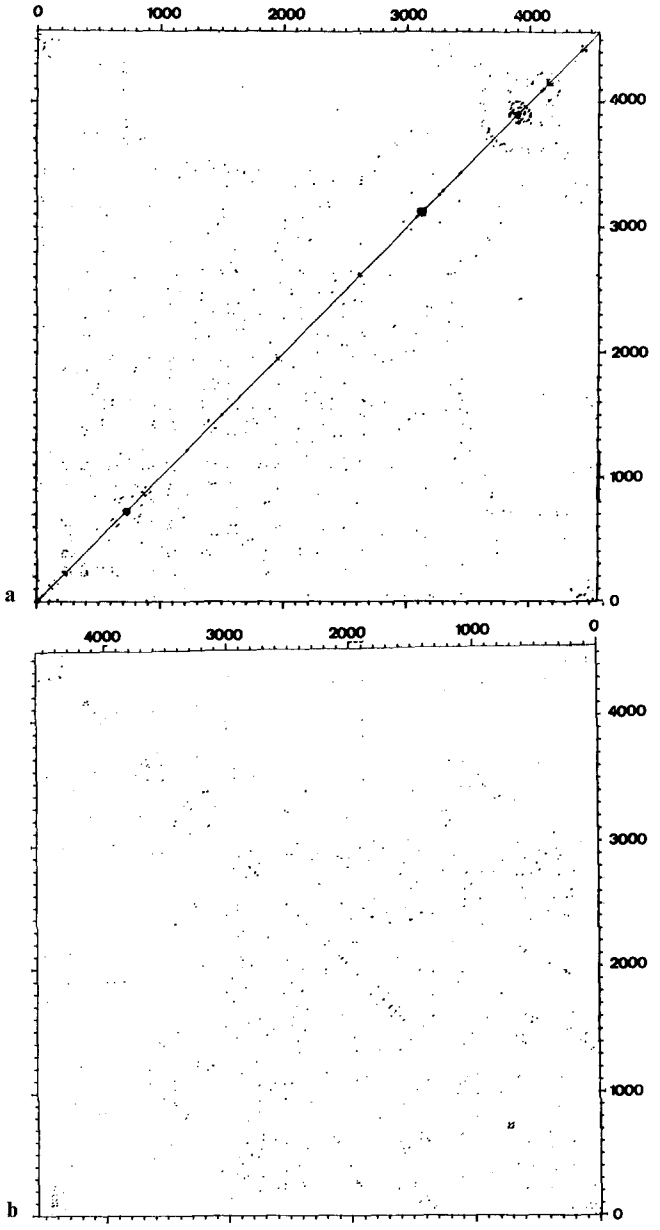

Fig. 3a, b. Dot plots of the sequence of *Ac* against itself (**a**) and against its complement (**b**). Each dot represents a match of a sequence of 20 nucleotides length with no more than 6 mismatches. In (**b**) the diagnoal from upper left to lower right indicates the presence of local inverted repeats. In addition, the accumulation of the dots in the lower left and upper right corners indicates the presence of a few inverted repeats at the two opposite termini of the element

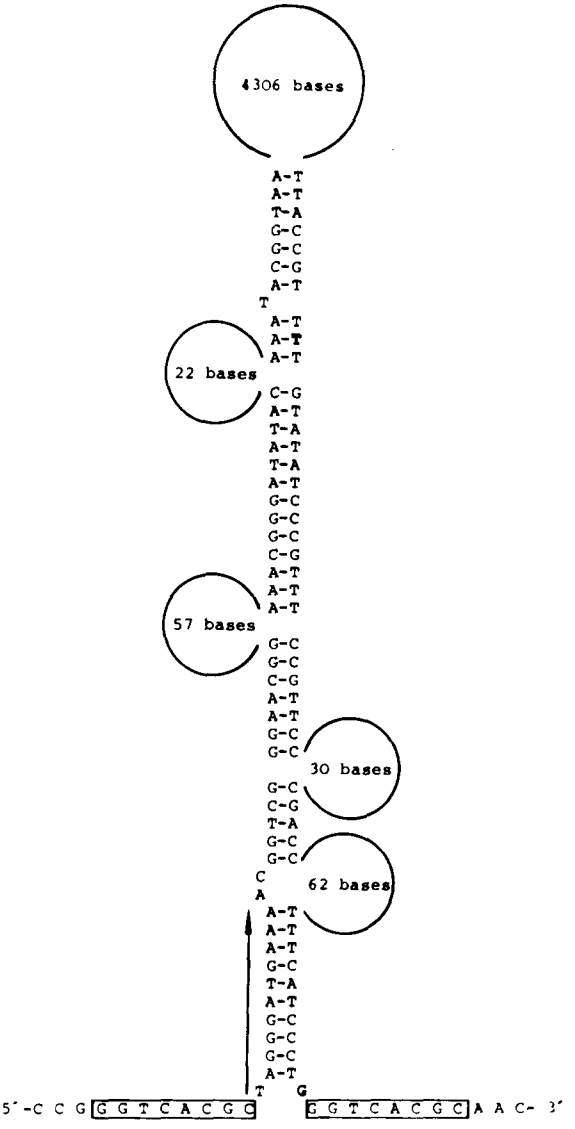

Fig. 4. Secondary structure at the termini of *Ac*. The *arrows* indicate the 11 bp repeats (one mismatch) at the end of *Ac*. The mismatched G is printed in bold face. The 8 bp duplications flanking *Ac* are *boxed in*

Ac9 into *Ds9*, inactivating the transposon function of *Ac*. This implies that *Ac* encodes (a) gene product(s), probably (a) protein(s). It is therefore tempting to analyze the *Ac* sequence for the presence of genes. We begin by asking whether the ORFs potentially encode proteins. By the criterion of Fickett's algorithm (Fickett 1982) ORF1 and ORF2 qualify as potential protein-encoding genes. ORF3 by this criterion is unlikely to encode a protein. We are aware, however, that Fickett's test code has not been evaluated for plant genes.

The above discussion of the ORFs as protein-encoding genes is meaningful only if the genes are unspliced, about which we have no information. If there were no splicing, each protein would have to be encoded by a separate RNA molecule. As shown in Results, the presumptive TATA boxes and polyadenylation sites for the ORFs are not very close to the respective consensus sequences (Messing et al. 1983) while a very good polyadenylation site is located within ORF2. Clearly, RNA and protein studies are needed.

Between the putative transcription start signals and the

first ATG in frame, out of frame met codons are found in both ORF1 and ORF2. These may reduce proper translation starts, as has been found in mammalian cells (Liu et al. 1984). Whether a similar effect is exerted here and whether a low level of correct translation serves a useful function in reducing the level of transposition will be the subject of further studies.

The polypeptide potentially encoded by ORF2 is interesting. At the N terminus, the five amino acids met, ala, ile, val and his are followed by ten repeats of the sequence pro gluN (seven times) or pro glu (3 times). The codons used for this repeated amino acid sequence show third base variation. This suggests a selective value of the amino acid sequence, as this is more uniform than the DNA sequence. A sequence of this kind is not listed in the atlas of protein structure (Dayhoff 1984).

The termini of the *Ds* and *Ac* elements sequenced so far (Döring et al. 1984; Pohlman et al. 1984; Sutton et al. 1984; Merckelbach, personal communication) differ in a characteristic manner. *Ds* elements are terminated by the perfect 11 bp inverted repeat TAGGGATGAAA. In the two *Ac* elements sequenced, one of the 5′ terminal T residues is replaced by a C. Apparently, this base substitution does not abolish transposition capability. Variability of the outermost nucleotides has not yet been observed in transposons of other species.

In addition to the 11 bp imperfect inverted repeat found at the very ends of the element, sequences that extend about 150 bp inside can be arranged in a secondary structure (Fig. 4). This is supported by the observation of a double-stranded structure forming at the ends of *Ac7* in heteroduplex molecules examined by electron microscopy (Behrens et al. 1984). A structure of this kind has also been proposed for the terminal sequence of the transposable elements *Tam1* of *Antirrhinum majus* (Bonas et al. 1984) and the Spm(En) of *Zea mays* (Schwarz-Sommer et al. 1984).

The hexanucleotide CCGTTT or close derivatives of it are found several times near one of the *Ac* termini. Its complement is found several times at the other end. These sequences contribute to the structure shown in Fig. 4 (see also Döring et al. 1984). It is interesting to note that the pentanucleotide CGTTT is part of a dekanucleotide found repeatedly in the FB4 transposon of *Drophila melanogaster*. In both FB4 and in *Ac* this sequence occurs predominantly at the termini, rather than in the center of the element. The center of FB4 is occupied by a large ORF (Potter 1982).

Ac contains other direct and inverted repeats, as shown in Fig. 3. We do not know how they were generated, but it is likely that they are derived from each other by replication errors or recombination events. This has been postulated for the formation of some alleles of the *E. coli* insertion element *IS*2 (Ghosal et al. 1979) and for interfering forms of vesicular stomatitis RNA virus (O'Hara et al. 1984).

A comparison of the DNA sequence of *Ac* and the *Ds* element isolated from the double *Ds* in the *sh-m5933* allele (Döring et al. 1984) shows this *Ds* element to be an internal deletion of *Ac*. The point where the left and the right remaining DNA segments of *Ac* are joined in this *Ds* element is not the same as the point where one copy of this sequence is inserted into the other in double *Ds*, as indicated in Fig. 2. This makes it likely that double *Ds* was created subsequent to the deletion in *Ac*.

The DNA sequences of the *Ac* elements of both the *wx-m7* and the *wx-m9* alleles are identical. The phenotypic difference between the two alleles must therefore have another cause. It will be interesting to see, how the different insertion sites can be responsible for the respective phenotypes, as was suggested for *En (Enhancer)* elements by Peterson (1977).

Acknowledgements. We thank Drs. Pohlman and Messing (Minneapolis, USA), Fedoroff (Baltimore, USA) and Gierl and Schwarz-Sommer (Köln, FRG) for the exchange of unpublished sequence data and J. Riegel for help with DNA sequencing experiments. We further thank Dr. U. Krawinkel for carrying out the dideoxy sequence experiment. This research was supported by Deutsche Forschungsgemeinschaft through SFB 74. John I. Yoder was supported by fellowships of DAAD and Alexander von Humboldt Stiftung.

References

Ansorge W, Baker R (1984) System for DNA sequencing with resolution of up to 600 base pairs. J Biochem Biophys Methods 9:33–47

Behrens U, Fedoroff N, Laird A, Müller-Neumann M, Starlinger P, Yoder J (1984) Cloning of *Zea may* controlling element *Ac* from the *wx-m7* allele. Mol Gen Genet 194:346–347

Berget SM (1984) Are U4 small nuclear ribonucleoproteins involved in polyadenylation? Nature 309:179–182

Bonas U, Sommer H, Saedler H (1984) The 17 kb *Tam1* element of *Antirrhinum majus* induces a 3 bp duplication upon integration into the chalcone synthase gene. The EMBO J 3:1015–1019

Boyer HW, Roulland-Dussoix D (1969) A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J Mol Biol 41:459–472

Courage U, Döring HP, Frommer WB, Kunze R, Laird A, Merckelbach A, Müller-Neumann M, Riegel J, Starlinger P, Tillmann E, Weck E, Werr W, Yoder J (1984) Transposable elements *Ac* and *Ds* at the *Shrunken, Waxy* and *Alcohol Dehydrogenase 1* loci in *Zea mays* L. Cold Spring Harbor Symp Quant Biol, in press

Dayhoff MO (1984) Atlas of protein sequence and structure. NBRF data bank

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the Vax. Nucl Acids Res 12:387–395

Döring H-P, Tillmann E, Starlinger P (1984) DNA sequence of the maize transposable element Dissociation. Nature 307:127–130

Döring H-P, Starlinger P (1984) Barbara McClintock's controlling elements: Now at the DNA level. Cell (in press)

Farabaugh PJ, Schmeissner U, Hofer M, Miller JH (1978) Genetic studies of the *lac* repressor. VII. On the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. J Mol Biol 126:847–863

Fedoroff N, Wessler S, Shure M (1983) Isolation of the transposable maize controlling elements *Ac* and *Ds*. Cell 35:235–242

Fickett JW (1982) Recognition of protein coding regions in DNA sequences. Nucleic Acids Res 10:5303–5318

Ghosal D, Gross J, Saedler H (1979) The DNA sequence of IS2–7 and generation of mini-insertions by replication of IS2 sequences. Cold Spring Harbor Symp Quant Biol 43:1193–1196

Kamp D, Chow LT, Broker TR, Kwoh D, Zipser D, Kahmann R (1979) Site-specific recombination in phage Mu. Cold Spring Harbor Symp Quant Biol 43:1159–1167

Kozak M (1981) Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucl Acids Res 9:5233–5252

Legerski R, Hodnett JL, Gray Jr, HB (1978) Extracellular nucleases of Pseudomonas BAL 31 III. Use of the double-strand deoxyriboexonuclease activity as the basis of a convenient

method for the mapping of fragments of DNA produced by cleavage with restriction enzymes. Nucl Acids Res 5:1445–1464
Liu Ch, Simonsen C-C, Levinson AD (1984) Initiation of translation at internal AUG codons in mammalian cells. Nature 309:5963:82–85
Marinus MG, Morris NR (1973) Isolation of DNA methylase mutants of *E. coli* K12. J Bacteriol 114:1143
Maxam A, Gilbert W (1977) A new method for sequencing DNA. Proc Natl Acad Sci USA 74:560–564
Maizel Jr, JV, Lenk RP (1981) Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc Natl Acad Sci USA 78:7665–7669
Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, New York
McClintock B (1947) Cytogenetic studies of maize and *Neurospora*. Carnegie Inst Washington Year Book 46:146–152
McClintock B (1949) Mutable loci in maize. Carnegie Inst Washington Year Book 48:142–154
McClintock B (1951) Chromosome organization and genic expression. Cold Spring Harbor Symp Quant Biol 16:13–47
McClintock B (1963) Further studies of gene-control systems in maize. Carnegie Inst Washington Year Book 62:486–493
McClintock B (1964) Aspects of gene regulation in maize. Carnegie Inst Washington Year Book 63:592–602
McClintock B (1965) The control of gene action in maize. Brookhaven Symp Biol 18:162–184
McClintock B (1978a) Mechanisms that rapidly reorganize the genome. Stadler Symp 10:25–48
McClintock B (1978b) Development of the maize endosperm as revealed by clones. In: The clonal basis of development. Carnegie Inst, Washington
Messing J, Geraghty D, Heidecker G, Hu N-T, Kridl J, Rubenstein I (1983) Plant gene structure. In: Kosuge T, Meredith CP, Hollaender A (eds) Geneticc engineering of plants. Plenum Press, New York and London, pp 211
Nelson OE, Rines HW (1962) The enzymatic deficiency in the *waxy* mutant of maize. Biochem Biophys Res Comm 9:297–300
O'Hara PJ, Nichol ST, Horodyski FM, Holland JJ (1984) Vesicular stomatitis virus defective interferring particles contain extensive genomic sequence rearrangements and base substitutions. Cell 36:915–924
Peterson PA (1977) The position hypothesis for controlling elements in maize. In: Bukhari AI, Shapiro JA, Adhya SL (eds) DNA insertion elements, plasmids and episomes. Cold Spring Harbor Laboratory, New York, pp 429–435
Pohlman RF, Fedoroff N, Messing J (1984) The nucleotide sequence of the maize controlling element Activator. Cell 37:635–643
Potter SS (1982) DNA sequence of a foldback transposable element in *Drosophila*. Nature (London) 297:201–204
Schwarz-Sommer Zs, Gierl A, Klösgen RB, Wienand U, Peterson PA, Saedler H (1984) The *Spm (En)* transposable element controls the excision of a 2 kb DNA insert at the *wx-m8* locus of *Zea mays*. The EMBO J 3:1021–1028
Silverman M, Simon M (1983) Phase variation and related systems. In: Shapiro JA (ed) Mobile Genetic elements. Academic Press, New York, pp 537–557
Sutton WD, Gerlach WL, Schwartz D, Peacock WJ (1984) Molecular analysis of *Ds* controlling element mutations at the *Adh1* locus of maize. Science 223:1265–1268

Communicated by H. Saedler

Received September 5, 1984