The EMBO Journal vol.3 no.5 pp.1015 – 1019, 1984

The 17-kb Tam1 element of *Antirrhinum majus* induces a 3-bp duplication upon integration into the chalcone synthase gene

Ulla Bonas, Hans Sommer and Heinz Saedler*

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

*To whom reprint requests should be sent

Communicated by H.Saedler

The DNA sequence of the termini and the flanking regions of the 17-kb transposable element Tam1 was determined. Tam1 is integrated in the chalcone synthase gene of the *niv-53* mutant of *Antirrhinum majus*. The element has a 13-bp perfect inverted repeat at its termini and appears to induce a 3-bp duplication of the target site upon integration. The DNA sequence of a *niv*$^+$ revertant was analyzed and found to differ from the wild-type sequence by an additional 2 bp that seem to derive from the target site duplication. Stretches of homologous sequences have been found between the ends of Tam1, within each terminus of the element, and between the termini and target site sequences. Structural similarities between the ends of Tam1 and the Spm-I8 element of *Zea mays* reflect a possible horizontal spread of a common progenitor.

Key words: Antirrhinum majus/chalcone synthase gene/DNA sequence/Tam1/transposable element

Introduction

Flowers of the snapdragon *Antirrhinum majus,* line T53, carrying the unstable mutation *nivea-recurrens (niv-rec)* exhibit a characteristic variegated phenotype consisting of numerous streaks or flakes of dark red against a white background. First described by Kuckuck (1936) and further characterized by Harrison and Carpenter (1973), this mutation has recently been analyzed at a molecular level in our laboratory. After establishing that the *nivea* locus encodes the enzyme chalcone synthase, which plays a key role in anthocyanin synthesis (Heller and Hahlbrock, 1980; Spribille and Forkmann, 1982; Reif, unpublished data), we subsequently found that the *niv-rec* mutant, renamed *niv-53*, contains a 17-kb insert (Bonas *et al.*, 1984). The insert, called Tam1 (Transposon antirrhinum majus 1), apparently inactivates the chalcone synthase *(chs)* gene, producing colorless tissue. However, as verified by Southern hybridization analysis (Bonas *et al.*, 1984), the element is frequently excised in individual somatic and germinal cells, restoring pigmentation in these cells and their clonal progeny.

Here we describe the sequence at the target site of Tam1 integration in a wild-type strain and in a *niv*$^+$ revertant of *niv-53* as well as the sequences of the Tam1 termini.

Results

Cloning of the DNA fragments containing the chs gene-Tam1 junctions

The 17-kb element Tam1, which is integrated in the chalcone synthase *(chs)* gene of *A. majus*, was shown to be responsible for the *niv-rec* mutation in line T53 (Bonas *et al.*, 1984). Previously we described the cloning in λ1059 of two overlapping fragments (14 kb and 15 kb long) containing the entire Tam1 element. Since one of these fragments could not be removed from the vector by *Bam*HI restriction digests, a second cloning program was initiated. λ gt WES was chosen as a cloning vector for the 4.4-kb and 4.7-kb *Eco*RI fragments containing the junctions between the *chs* gene and the Tam1 element (Bonas *et al.*, 1984). Two clones, Am53-R, containing a 4.4-kb fragment, and Am53-L, harboring a 4.7-kb *Eco*RI insert, were analyzed further.

A comparison of the restriction maps of Am53-R and Am53-L with the wild-type *chs* fragment present in clone Am3 is shown in Figure 1. Each clone contains one region that is homologous to the *chs* gene and a second region that shows no homology to any part of the wild-type sequence. The pattern of restriction sites indicates that the right-hand junction between Tam1 and the *chs* gene is present in Am53-R while Am53-L contains the left-hand one.

Heteroduplex studies have shown that Tam1 is integrated ~3.45 – 3.85 kb from the left (5′) *Eco*RI site and 2.25 – 1.85 kb from the right end of the Am3 clone (Bonas *et al.*, 1984). For sequence analysis of the border regions between Am3 and Tam1, appropriate fragments were isolated. The parts of Am53-R and Am53-L that have been sequenced are illustrated in Figure 1. Both strands of the 520-bp *Taq*I/*Hin*fI fragment from Am53-R shown in Figure 1 were sequenced completely with the help of various overlapping subfragments. One strand of the adjoining *Hin*fI/*Ava*II fragment of Am53-R was also sequenced as well as the overlapping region between these two major fragments. The sequence of both strands of the 170-bp *Dde*I fragment of Am53-L shown in Figure 1 was determined.

DNA sequence of the Tam1 termini

To define the point of Tam1 integration, the wild-type sequence of the Am3 fragment was compared with the corresponding sequences within the clones Am53-R and Am53-L. Figure 2 shows the sequences of the junction fragments and the relevant portion of the *chs* gene. Tam1 appears to be integrated in the promoter region of the *Antirrhinum chs* gene. This conclusion is based on the following reasoning. Antibodies against the *chs* enzyme from parsley cross-react with the *Antirrhinum* enzyme (Wienand *et al.*, 1982). No amino acid sequence data is available for the *Antirrhinum* enzyme, but a cDNA sequence encoding the N-terminal end of the parsley enzyme is known (Reimold *et al.*, 1983). A nucleotide sequence comparable with this parsley sequence is present in Am3. On the basis of this sequence, the ATG underlined in Figure 2 is thought to be the translational initiation codon. Further upstream is a sequence resembling a TATA box. Tam1 is inserted 17 bp upstream from the putative TATA box and 127 bp upstream from the ATG codon. This result is in agreement with the heteroduplex studies reported previously (Bonas *et al.*, 1984). The clones Am53-R and AM53-L each contain the sequence ATA at the junctions between the *chs* gene and the Tam1 element. However, this triplet occurs only once in the wild-type Am3 sequence. Therefore, the Tam1

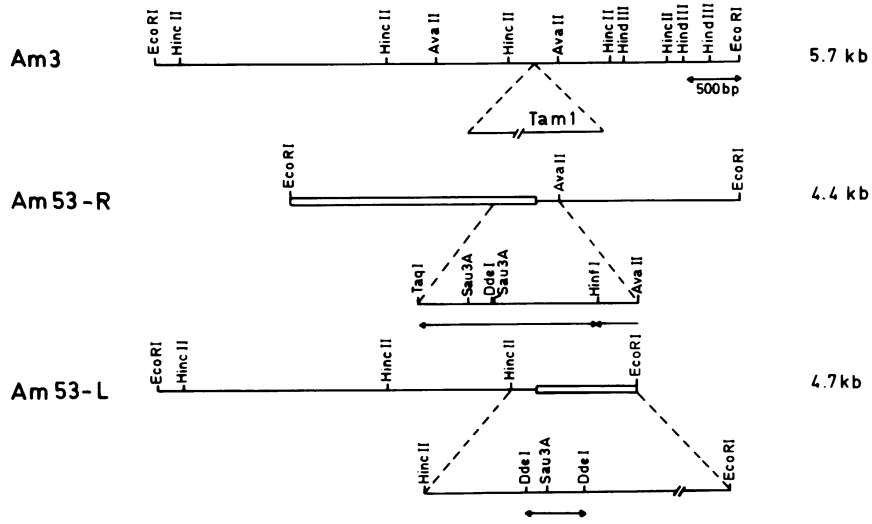

Fig. 1. Comparison of the restriction maps of Am53-R and Am53-L with Am3. The 4.4-kb and 4.7-kb *Eco*RI fragments containing the junctions between the *chs* gene and the Tam1 element were isolated from λ gt WES recombinant clones. Clone Am3, containing the 5′ region of the *chs* gene was used as a probe. The homology between either Am53-R or Am53-L and Am3 is indicated by the common restriction sites drawn on the single line. The double line represents Tam1 sequences. The position of Tam1 integration in the *chs* gene is indicated by the triangle below the Am3 restriction map. For sequence analysis, various fragments of Am53-R and Am53-L were subcloned. A double-headed arrow beneath the enlarged map indicates that both strands of the fragment were sequenced, while a single-headed arrow means that only one strand was sequenced.

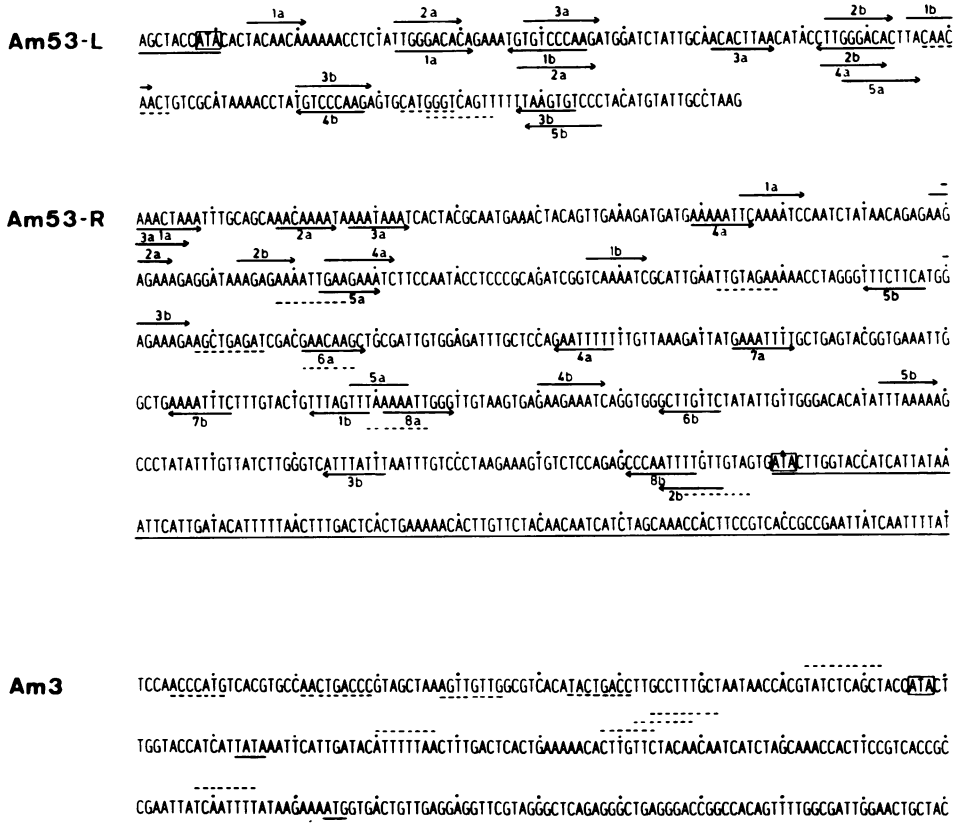

Fig. 2. Sequence of the Tam1 ends and flanking sequences. The sequence of Am53-L begins with the *Dde*I site of the 170-bp *Dde*I fragment (see Figure 1). It contains the junction between the left end of Tam1 and the *chs* gene. The Am53-R sequence includes the right-hand (3′) junction of Tam1 and starts within the *Hinf*I/*Ava*II fragment shown in Figure 1. From Am3 only the relevant portion is shown. In Am53-R and Am53-L, sequences homologous to Am3 are underlined with a continuous line. The 3-bp boxed-in sequence at the left and the right termini of Tam1 is found only once in the Am3 sequence. In the Am3 sequence, the presumed TATA box 17-bp downstream of the integration site and a possible ATG initiation codon are underlined. Regions of homology within Am53-L and Am53-R are designated by arrows. Arrows drawn below the sequences refer to inverted repeats. Those above the sequences indicate direct repeats within each end. Only repeats 8 bp or larger are considered. The repeat symmetries are paired by number 1a/1b, etc. Inverted repeats present in Am3 and Am53-L are designated by dotted lines below the Am3 sequence; those present in Am3 and the right end of Tam1 are indicated by dotted lines above the Am3 sequence.

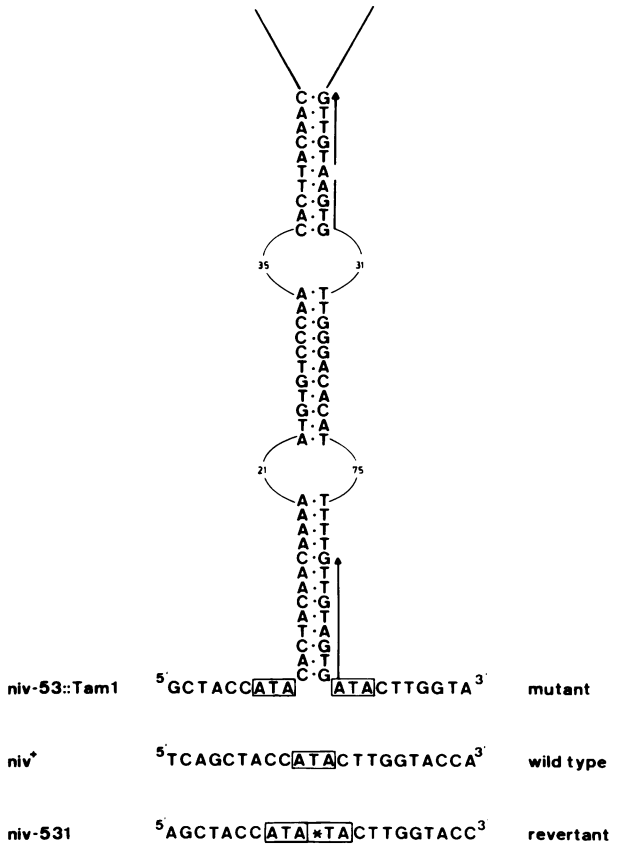

Fig. 3. Sequence homology between the ends of Tam1: a possible secondary structure. The 5′ and 3′ termini of Tam1 are drawn in a stem-and-loop structure by matching stretches of sequence homology. Three sections of inverted repeat homology can be seen: a 13-bp IR at the very end, followed by an 11-bp IR and a 10-bp IR. As indicated by arrows, the innermost 10-bp IR is identical to the terminal 13-bp IR except for the TA base pair as illustrated by the interruption in the arrow. The small numbers included in the loops refer to the number of nucleotides present between the paired regions. The boxed-in 3-bp sequence flanking the element is present only once in the wild-type sequence. In the revertant *niv-531* the duplication appears to have been retained except for a single A/T base pair indicated by (*).

element appears to generate a 3-bp duplication upon integration.

Comparison of the sequences of the ends of the Tam1 element reveals that there is extensive homology between them (Figure 3). Three patches of inverse sequence homology exist between the ends, a 13-bp inverted repeat (IR) at the very end, an 11-bp IR in the middle and a 10-bp IR in the inner part of the Tam1 termini. Remarkably enough, the 10-bp IR is identical to the first 9 bp of the terminal 13-bp IR except for a single base pair. An additional TA base pair is present in the 10-bp IR at the position indicated in Figure 3.

DNA sequence of the niv⁺ revertant

A *niv⁺* revertant, in which Tam1 has been excised thus restoring activity of the *nivea* locus, was cloned into vector EMBL4 (see Materials and methods). Sequence analysis of the *niv-531* revertant revealed that the sequence of the site at which Tam1 was integrated is different from the wild-type sequence (Figure 3). Two additional nucleotides (TA) are present at the former site of integration. It appears that the total Tam1 element was excised together with one of the immediately flanking A residues in the reversion process. The nucleotide sequence of the *chs* gene surrounding the target site in the revertant was shown to be identical to the wild-type sequence.

Discussion

Tam1 appears to be a transposable element

The data presented here show that the 17-kb insert Tam1 has structural features in common with other transposable elements. It contains inverted repeats at its termini and generates a duplication of the target site upon integration (Figure 3).

Excision of Tam1

To gain information about the mechanism of Tam1 excision, the *chs* sequence of an *Antirrhinum niv⁺* revertant was compared with that of a wild-type line. In the revertant two additional base pairs are present at the former integration site (Figure 3). These can be interpreted to be remnants of the duplication generated during the integration process. Pigment synthesis is obviously not affected by these additional two base pairs while in the case of the 17-kb insert in the *niv-rec* mutant, gene expression is inhibited. It should be noted that the wild-type strain sequenced (a commercial variety), is not the progenitor of the *niv-rec* line T53 so that some uncertainty about the exact target site sequence remains.

A second case in which the duplication seems to have been retained in revertants of transposable element-induced mutations has also been reported. In *Zea mays*, Ds induces an 8-bp duplication at the *adh1* locus (Sachs *et al.*, 1983; Döring *et al.*, 1984). In four revertants analyzed, the 8-bp duplication of the target site is retained but in an altered form. At the junctions between the two direct repeats, inversions and deletions of 1 or 2 bp were found (Sachs *et al.*, 1983). This mode of excision may be characteristic of plant transposable elements. In *Drosophila*, for example, excision of both a P element (O'Hare and Rubin, 1983) and a foldback element (Collins and Rubin, 1983) results in precise restoration of the original sequence.

Since excision must involve cleavage at the ends of Tam1 and rejoining of the *chs* sequences at the target site, it is possible that the ends of the element are somehow brought together in this process. As shown in Figure 3, an elaborate stem and loop structure can be formed by pairing stretches of sequence homology between the ends of Tam1. One could envisage that a protein (transposase) specifically recognizes the terminal 13 bp of Tam1 and brings or holds the ends together. After proper cleavage at these ends the element would be excised, followed by religation of the target site sequences.

Since the 10-bp IR in the Tam1 terminus is almost identical to the endmost 13-bp IR (Figure 3), it too may be a substrate for transposase activity, although less effectively due to the additional TA base pair. Occasional cleavage of Tam1 at the 10-bp IR would generate a rudimentary 210-bp element at the locus. Whether or not this rudimentary element would have any effect on gene expression is a moot point. In analogy to the additional base pairs in the *niv-53* revertant, it may not interfere with gene expression at all. On the other hand, it might inhibit expression partially or even completely. At any rate, since it would be too small to code for any products of its own, it should be incapable of autonomous excision and therefore produce a non-variegated phenotype. However, since the element would still contain intact termini, excision ought to occur in response to signals from an active Tam1 ele-

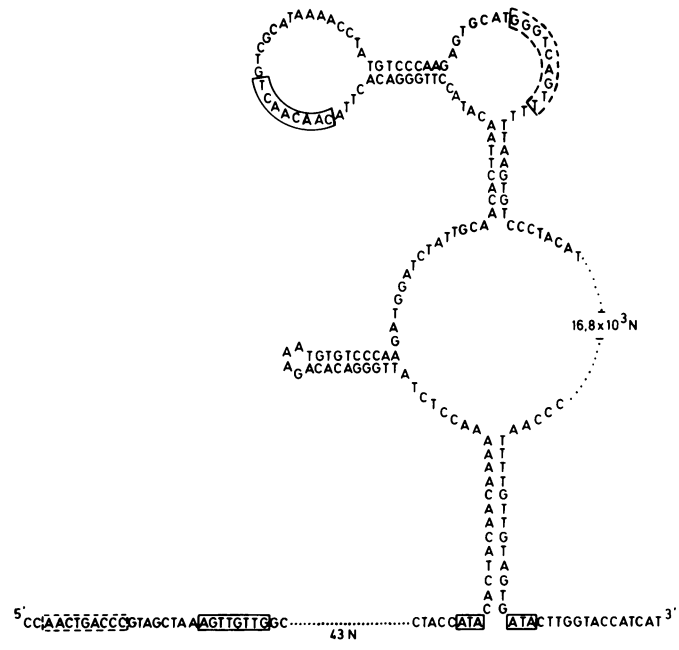

Fig. 4. The left terminus of Tam1. One of several possible secondary structures which can be formed within the left terminus of the element using stretches of inverted repeat homology is shown (see also Figure 2). It extends from the terminal 13-bp IR stem formed between the two termini of Tam1. The single-stranded loops exhibit homology to the target sequence as illustrated by the boxed-in sequences which are inverted repeats of sequences to the left of the integration site of Tam1 in the promoter region of the *chs* gene. A similar structure can be drawn for the right end of Tam1. N: nucleotides.

ment elsewhere in the genome. This would then lead to a variegated phenotype and a two-component system similar to the Ac/Ds (McClintock, 1951) and En/I (Peterson, 1953) systems described in *Z. mays*. We are currently attempting to isolate a system of this kind by genetic means.

Possible significance of sequence homology for integration and/or excision of Tam1

Stretches of extensive sequence homology can be found within each Tam1 terminus (Figure 2; direct and inverted repeats larger than 8 bp are indicated). By pairing these sectors, several possible secondary structures can be formed. Considering ony the left end of Tam1, a structure reminiscent of a tRNA can be drawn, as illustrated in Figure 4. The single-stranded loops exposed in this structure contain extensive homology to sequences in the *chs* gene in inverse orientation. One could imagine that these sequences interact with the homologous sequences of the target site during integration and/or excision of Tam1. The two boxed-in sequences in Figure 4 are found respectively 53 bp and 70 bp upstream of the integration site of Tam1 in the *chs* gene. Similar patches of sequence homology are found between the right end of Tam1 and neighbouring *chs* gene sequences. Interactions of the kind proposed would lead to pronounced site specificity for Tam1 integration. This can be elucidated by analyzing other Tam1 integration events.

Speculations on the origin of new Tam1 derivatives

Considering the structure shown in Figure 4, it can be speculated that recombination occurs within the paired sectors leading to new Tam1 derivatives. Recombination between direct repeats would lead to deletion of part of the terminus while recombination between inverted repeats would result in inversion within the terminus. Recombinational events of this kind could be expected to affect the element's influence on gene expression or its ability to excise. For this purpose we are investigating *niv-rec* plants that exhibit an altered phenotype compared to the original high-flaking *niv-rec* phenotype.

Comparison of Tam1 and other transposable elements in plants

The 13-bp perfect inverted repeats of Tam1 (Figure 3) are identical to the first 13 bp of the 14-bp IR of Tam2 (Upadhyaya *et al.*, in preparation), a 5-kb insertion sequence isolated from a stable white *nivea* mutant of *A. majus*. Comparison of the sequences of Tam1 and Tam2 reveals extensive homology between the first 200 bp of their left termini while the major part of the elements are non-homologous. However, a stem and loop structure similar to that of Tam1 can be formed by pairing regions of homology between the ends. Therefore a common function such as a transposase may operate on both elements.

There are structural similarities between the *Antirrhinum* elements Tam1 and Tam2 and transposable elements from other plant species. The terminal 5 bp of a 3.4-kb insertion in the lectin gene *(Le1)* of soybean are identical to those of Tam1 and Tam2 (Vodkin *et al.*, 1983). Furthermore, the termini of Tam1 and Tam2 are homologous to 12 bp out of 13 bp at the very ends of the receptor component (Spm-I8) of the Spm controlling element system present in the *waxy^{m-8}* mutant of maize (Schwarz-Sommer *et al.*, 1984). All these elements contain the sequence CACTA at the very ends of their termini. Moreover, they all produce a 3-bp duplication of the target site upon integration. The structural similarities at the ends may reflect a related mechanism of transposition and/or horizontal spread of a common progenitor of these transposons during evolution.

Materials and methods

Plant material

A. majus line T53, genotype *niv-53::Tam1*, is a homozygous *nivea-recurrens (niv-rec)* mutant that exhibits a highly variegated phenotype (Harrison and Carpenter, 1973). The *niv*$^+$ revertant line 101, genotype *niv-531*, provided by B.J.Harrison (John Innes Institute, Norwich, UK), was isolated from line T53 and inbred for at least three generations. All plants were grown at 20–25°C in the greenhouse.

Isolation of plant DNA and cloning procedures

Total DNA was extracted from frozen plant material, as described previously (Bonas *et al.*, 1984).

For construction of genomic T53 clones in λ gt WES, 200 μg DNA was completely digested with *Eco*RI and separated on an 0.8% agarose gel. The DNA of the 4.4-kb and 4.7-kb regions were electroeluted onto DEAE-cellulose paper (Dretzen *et al.*, 1981). Preparation of the λ gt WES vector, ligation and *in vitro* packaging were performed as described (Wienand *et al.*, 1982). Recombinant phages were screened by plaque hybridization (Benton and Davis, 1977) to ^{32}P-labeled Am3, a 5.7-kb *Eco*RI fragment carrying a major part of the *chs* gene (Wienand *et al.*, 1982). 8 x 10^4 recombinants from the cloning of the 4.4-kb region were obtained yielding ~200 positive signals. From the recombinant phages carrying inserts in the size range of 4.7 kb, 26 positively hybridizing plaques were isolated. Clones Am53-R and Am53-L were investigated further.

DNA from the *niv*$^+$ revertant line 101 was partially digested with *Mbo*I (100 μg DNA; 0.37 U *Mbo*I/μg DNA in 1 h) and size fractionated on a sucrose gradient according to Karn *et al.* (1980). λ EMBL4 vector arms were prepared by digestion with *Bam*HI + *Sal*I as described by Frischauf *et al.* (1983). Purification of the vector arms was carried out by PEG precipitation (A.Gierl, unpublished). 2.5 μg plant DNA (fragments >14 kb) was ligated with 3.5 μg EMBL4 arms, packaged *in vitro* and plated onto Q364, a P2 lysogen which allows selection for recombinants (Karn *et al.*, 1980). 5 x 10^5 recombinant phage plaques were hybridized to ^{32}P-labeled Am3 DNA. Eight positive clones with ~14-kb inserts were isolated. Six of these clones contain-

ed a 5.7-kb *Eco*RI fragment homologous to the 5.7-kb *Eco*RI fragment from a wild-type strain. Clone Am101-2 was selected for further subcloning and sequence analysis.

Subcloning into plasmid vectors

The 4.7-kb *Eco*RI insert from the λgt WES clone Am53-L was subcloned into pACYC184. From clone Am53-R, a 3.1-kb *Eco*RI/*Hind*III fragment was cloned into pBR328. HB101 was used as a host for transformation. For sequence analysis appropriate fragments from both clones and from Am101-2 were cloned into the *Sma*I site of pUC9 and transformed into MC1022.

Preparation of phage and plasmid DNAs

Purification of bacteriophage λ and extraction of phage DNA was carried out as described by Yamamoto *et al.* (1970). Plasmid DNA was isolated as described by Holmes and Quigley (1981) and purified twice on a CsCl gradient.

Sequence analysis

DNA fragments were digested with appropriate enzymes, separated on 2% agarose gels and purified by phenolization. Both strands of 3'- and 5'-labeled fragments were sequenced using the technique described by Maxam and Gilbert (1980). 3 kb of the Am3 fragment have been sequenced (Sommer, unpublished data). Only the relevant portion is presented here for comparison.

Acknowledgements

We thank Dr.P.Nevers for very valuable and helpful discussions and critical reading of the manuscript. We also thank I.Will and T.Akmandor for technical assistance.

References

Benton,W.D. and Davis,R.W. (1977) *Science (Wash.),* **196**, 180-182.
Bonas,U., Sommer,H., Harrison,B.J. and Saedler,H. (1984) *Mol. Gen. Genet.,* in press.
Collins,M. and Rubin,G.M. (1983) *Nature,* **303**, 259-260.
Döring,H.P., Tillmann,E. and Starlinger,P. (1984) *Nature,* **307**, 127-130.
Dretzen,G., Bellard,M., Sassone-Corsi,P. and Chambon,P. (1981) *Anal. Biochem.,* **112**, 295-298.
Frischauf,A.M., Lehrach,H., Poutska,A. and Murray,N. (1983) *J. Mol. Biol.,* **170**, 827-842.
Harrison,B.J. and Carpenter,R. (1973) *Heredity,* **31**, 309-323.
Heller,W. and Hahlbrock,K. (1980) *Arch. Biochem. Biophys.,* **200**, 617-619.
Holmes,D.S. and Quigley,M. (1981) *Anal. Biochem.,* **114**, 193-197.
Karn,J., Brenner,S., Barnett,L. and Cesareni,G. (1980) *Proc. Natl. Acad. Sci. USA,* **77**, 5172-5176.
Kuckuck,H. (1936) *Z. Induct. Abstamm. Vererbgsl.,* **71**, 429-440.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.,* **65**, 499-560.
McClintock,B. (1951) *Cold Spring Harbor Symp. Quant. Biol.,* **16**, 13-47.
O'Hare,K. and Rubin,G.M. (1983) *Cell,* **34**, 25-35.
Peterson,P.A. (1953) *Genetics,* **38**, 682-683.
Reimold,U., Kröger,M., Kreuzaler,F. and Hahlbrock,K. (1983) *EMBO J.,* **2**, 1801-1805.
Sachs,M.M., Peacock,W.J., Dennis,E.S. and Gerlach,W.L. (1983) *Maydica,* **28**, 289-302.
Schwarz-Sommer,Zs., Gierl,A., Klösgen,R.B., Wienand,U., Peterson,P.A. and Saedler,H. (1984) *EMBO J.,* **3**,
Spribille,R. and Forkmann,G. (1982) *Phytochemistry,* **21**, 2231-2234.
Vodkin,L.O., Rhodes,P.R. and Goldberg,R.B. (1983) *Cell,* **34**, 1023-1031.
Wienand,U., Sommer,H., Schwarz,Zs., Shepherd,N., Saedler,H., Kreuzaler, F., Ragg,H., Fautz,E., Hahlbrock,K., Harrison,B. and Peterson,P.A. (1982) *Mol. Gen. Genet.,* **187**, 195-201.
Yamamoto,K.R., Alberts,B.M., Benzinger,R., Lawhorne,K. and Treiber,G. (1970) *Virology,* **40**, 734-744.

Received on 26 January 1984; revised on 13 February 1984