The EMBO Journal vol.10 no.6 pp.1571 – 1578, 1991

Genomic homologous recombination *in planta*

Susannah Gal, Barbara Pisan, Thomas Hohn, Nigel Grimsley[1] and Barbara Hohn

Friedrich Miescher-Institut, PO Box 2543, CH-4002 Basel, Switzerland
[1]Present address: INRA CNRS, Laboratoire de Biologie Moléculaire, Castanet-Tolosan, France

Communicated by B.Hohn

A system for monitoring intrachromosomal homologous recombination in whole plants is described. A multimer of cauliflower mosaic virus (CaMV) sequences, arranged such that CaMV could only be produced by recombination, was integrated into *Brassica napus* nuclear DNA. This set-up allowed scoring of recombination events by the appearance of viral symptoms. The repeated homologous regions were derived from two different strains of CaMV so that different recombinant viruses (i.e. different recombination events) could be distinguished. In most of the transgenic plants, a single major virus species was detected. About half of the transgenic plants contained viruses of the same type, suggesting a hotspot for recombination. The remainder of the plants contained viruses with cross-over sites distributed throughout the rest of the homologous sequence. Sequence analysis of two recombinant molecules suggests that mismatch repair is linked to the recombination process.

Key words: agroinfection/*Brassica napus*/cauliflower mosaic virus/mismatch repair/transgenic plants/

Introduction

The genomes of organisms vary in a population on an evolutionary time scale but also within a single individual. On an evolutionary scale, the alterations of genomes are the means by which organisms evolve and new species are created (for review, see Flavell, 1982). On an individual level, many processes are involved which alter the genomic DNA of certain cells in a single organism, but rather little is known about the controlling factors associated with these changes. One type of genomic change which is fairly amenable to study is homologous recombination. Homologous recombination is well studied in prokaryotes (Smith, 1989) and simple eukaryotes (Holliday, 1964; Petes and Hill, 1988). Studies in animal (Bollag *et al.*, 1989) and plant cell culture (Peterhans and Paszkowski, 1990) have been performed as well, but to a very limited extent in whole organisms. In animal systems, a few specific examples of homologous recombination in somatic cells of a whole organism exist. These include the dilute coat-colour mutation in mice (Seperack *et al.*, 1988), the retinoblastoma gene in humans (Knudson, 1985), immunoglobulin rearrangement in lymphoid cells of mammals (Engler and Storb, 1988), and in *Drosophila*, P element (Engels, 1983) and rDNA recombination (Hawley and Marcus, 1989). Analysis of homologous recombination in whole plants, however, is hampered by the lack of well characterized and easily scorable markers.

Plant genomes contain a large amount of repeated DNA sequences (as much as 75% of the nuclear DNA; Flavell, 1980; Thompson and Murray, 1981) which exists as non-coding elements and as multigenic, highly homologous gene families. These sequences could provide targets for homologous recombination resulting in continuous alteration of the genome and the potential for lethal deletions. Thus mechanisms must have evolved to suppress or control this variation and to maintain a relatively stable genome. These control mechanisms may be under the influence of genetic and physiological signals and may be subject to manipulation by appropriate alterations of these signals. A marker for recombination in a whole plant would therefore be a useful tool for studying the mechanisms used to control genomic homologous recombination.

We describe an engineered marker for homologous recombination which can be scored on intact plants. The approach uses cauliflower mosaic virus (CaMV), an 8 kb double-stranded DNA virus (for review, see Gronenborn, 1987), as the marker. The replication cycle of CaMV involves reverse transcription of an RNA intermediate (Bonneville *et al.*, 1988). Extrachromosomal recombination has been seen between CaMV strains present in turnips (Howell *et al.*, 1981; Lébeurier *et al.*, 1982), often involving the region near the 35S promoter (Choe *et al.*, 1985; Dixon *et al.*, 1986; Geldreich *et al.*, 1986; Grimsley *et al.*, 1986b; Stratford and Covey, 1989; Vaden and Melcher, 1990). This is the promoter for the genomic RNA, the 5′ end of which is the site of a template switch during reverse transcription to form the first DNA strand (Bonneville *et al.*, 1988). We have introduced into *Brassica napus* plants a CaMV DNA construction that would only produce viable virus after recombination. These plants produced viruses which were analysed for the frequency and sequence specificity of the recombination.

This system offers several distinct advantages over others which have been described previously (Burk and Menser, 1964; Peterhans *et al.*, 1990), including visual scoring for recombination (as viral infection produces chlorotic plaques and vein clearing on leaves), high sensitivity (theoretically a single event of recombination would produce infectious virus), and the ability to distinguish different recombination events (the duplicated homologous regions provided are from two viral strains with known sequence differences). This last attribute of our system allowed us to analyse different cross-over events and to propose structures for the recombination intermediate.

Results

Production and analysis of transgenic plants

We constructed a vector with CaMV sequences such that only after recombination could a viable virus be produced. This vector was introduced into *B.napus* plants using *Agrobacterium*-mediated transformation, using a mixture of *A.tumefaciens* and *A.rhizogenes*, as described in Materials and methods. The construction integrated into plants is shown in Figure 1A (a simple structure of a single T-DNA is shown while those actually found in plants are often tandem repeats, P.Swoboda and S.Gal, unpublished). Reverse transcription of a transcript from the integrated 35S promoter (labelled '35S' transcript in Figure 1A) will not produce virus since the complete viral sequence is not coded for on this transcript. However, if homologous recombination occurs between the indicated homologous regions (Figure 1B), viable virus could be produced. The homologous regions in this construction are derived from two different strains of CaMV and have multiple base pair differences, allowing us to distinguish between the different possible recombinant viruses (Figure 1C). The transgenic plants expressed nopaline synthetase, and Southern blot analysis confirmed the presence of CaMV sequences in the arrangement shown in Figure 1A (data not shown).

The progeny of the transgenic plants produced virus. These plants did not show virus at early stages in development but only at later times during growth. (A more detailed description of the timing of the appearance of virus in the transgenic plants will be published elsewhere.) By outcrossing, we have obtained transgenic plants which have little of the dark green, wrinkled leaf phenotype typical of *A.rhizogenes* transformed *B.napus* plants so that we were able to visualize the viral symptoms of vein clearing and chlorotic spots on leaves (Figure 2). Thus a marker for intrachromosomal homologous recombination in plants was established which did not require a destructive assay for scoring.

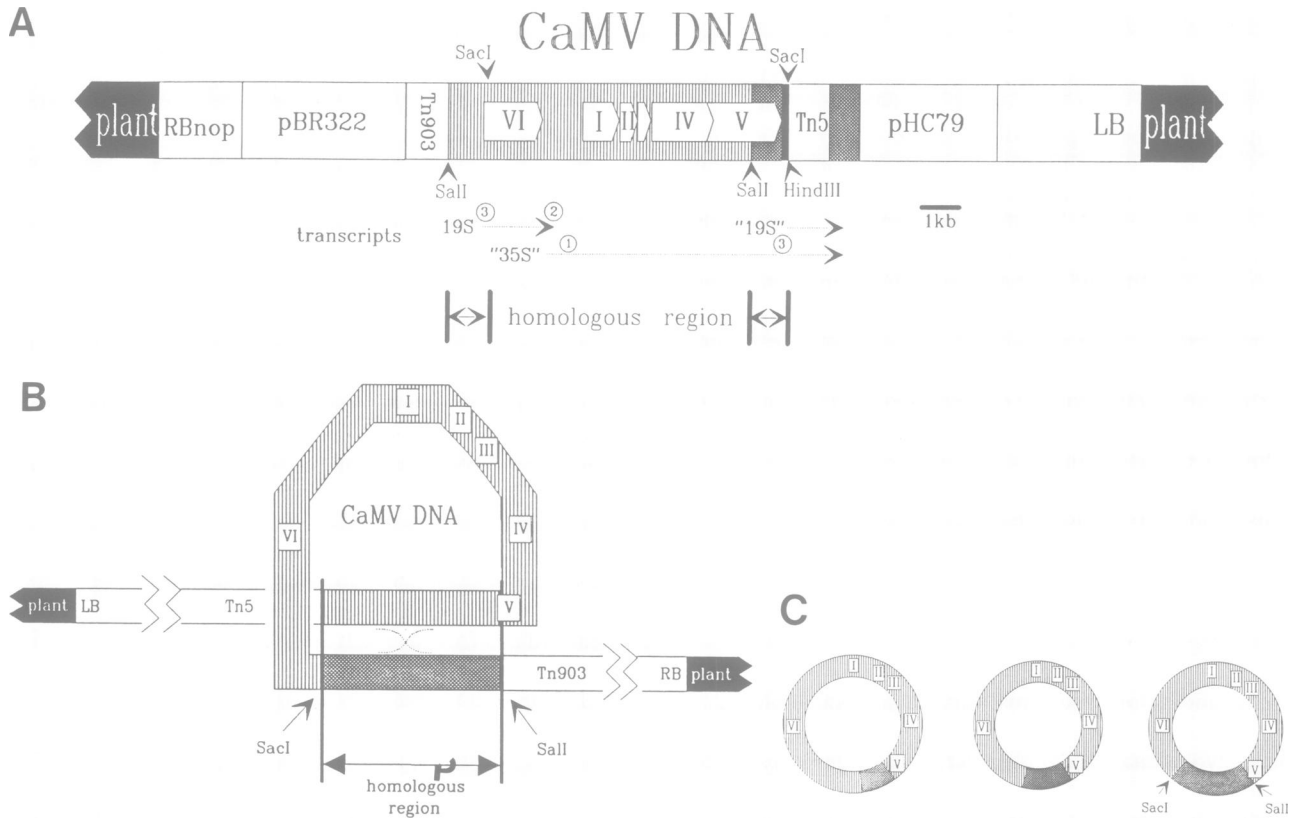

Fig. 1. A. Construction for monitoring homologous recombination in plants. As described in Materials and methods, the plasmid pEAP21 was integrated into the T-DNA of *Agrobacterium tumefaciens* and the structure of the transforming DNA once integrated in the plant genome is shown. RB and LB indicate the right and left borders, respectively, of the T-DNA, *nop* is the gene for nopaline synthetase. The CaMV sequences are indicated by ▥ blocks (D/H strain), ▩ blocks (4184 strain), and ■ blocks (S strain) with the open reading frames for some of the CaMV genes shown as open arrows, the larger ones labelled with roman numerals. Tn*903* is kanamycin resistance expressed in bacteria and Tn*5* is kanamycin resistance expressed in plants because it is under the control of the CaMV 19S promoter. The transcripts expected in plants expressed from the viral promoters are indicated by ·····>. Two 19S transcripts are produced, one coding for the natural gene VI protein and the other, labelled '19S' coding for the Tn*5* gene. The transcript from the CaMV 35S promoter is labelled '35S' since the transcript encodes Tn*5* instead of the natural gene VI protein. The ~1 kb homologous regions provided are located between the *Sal*I and *Sac*I (*Hin*dIII) sites and are indicated. The *Sal*I and *Sac*I sites are indicated only within the CaMV part of the construction, the *Hin*dIII site indicated is not unique. Illegal template switching using 19S and 35S transcripts would produce viable virus in our system if the reverse transcriptase initiated at the normal primer binding site (1), moved to the 5′ end of the '35S' transcript and switched to the 19S transcript (site 2), utilized the 19S transcript as template and then at the 5′ end of this transcript switched back to the '35S' transcript (sites marked 3). **B.** A possible structure for an intramolecular homologous recombination intermediate to form viable viral DNA from the construction shown in A. Cross-over (shown by the crossing dotted lines) anywhere along the homologous regions would produce viable virus. For simplicity, the homologous region representing the 4184/S CaMV strain is shown only with dots. The length of the homologous region is not strictly to scale. The orientation of the right and left T-DNA borders have been reversed with respect to Figure 1A. **C.** Structures of possible recombinant viruses after crossing-over in different parts of the homologous regions.

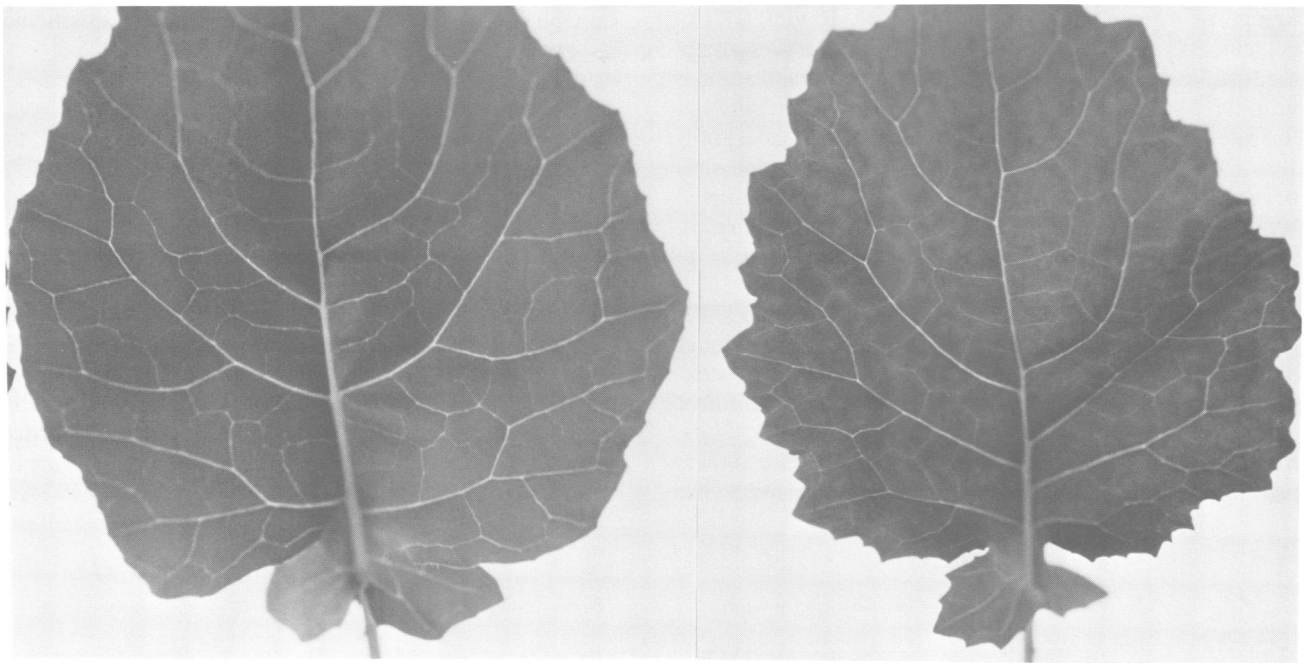

Fig. 2. Leaves of transgenic *Brassica napus* plants transformed with the construction shown in Figure 1A. The leaf on the left does not contain replicating virus and the leaf shown on the right contains replicating CaMV and shows typical viral symptoms of vein clearing and chlorotic spots.

Mapping of recombinant viruses

The viruses produced in the original transgenic plants and the resulting progeny plants could have a number of different sequences depending on the site of recombination used for excision of viral DNA from its proviral state (Figure 1). Within the 1033 bp homologous region, there are 34 bp differences between the D/H viral strain and the region containing sequences from CaMV strains 4184 and S. These base pair changes, all but one of which are included in the coding region for the gene V protein, are conservative at the protein level. Only two base pair changes actually cause an amino acid change and in both cases arginine is substituted by lysine. Thus we did not expect a difference in the viability of different recombinant viruses. These base pair differences are spread fairly randomly throughout the homologous region (see Figure 4) and in some cases produce restriction enzyme site differences which have been exploited to map the recombination or cross-over points. Virus DNA was isolated directly, either from the transgenic plant or from turnips inoculated with an extract from a transgenic plant, restricted by these distinguishing enzymes, and analysed by Southern blots. In some cases, polymerase chain reaction (PCR) products derived from these DNA samples were analysed on ethidium bromide-stained gels. No difference in the resulting pattern was seen when viral DNA samples from the transgenic plant, a turnip inoculated with the transgenic plant extract, or PCR products produced from these DNA samples were analysed. Over 40 different plants were analysed from three subsequent generations from a single transgenic plant which contained multiple T-DNA insertion loci. The majority of the plants (37 out of 41) contained a single major viral DNA species, as judged by the detection of single band patterns after digestion, using three distinguishing restriction enzymes (data not shown).

The restriction enzyme digestion patterns of the various viral DNAs found in different plants allowed us to divide them into four classes, based on the regions of the

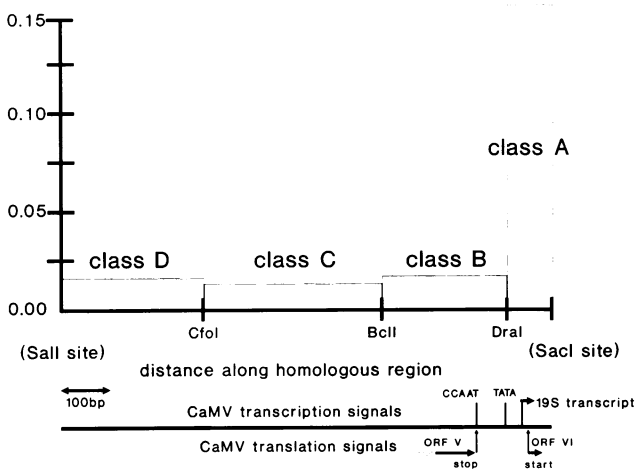

Fig. 3. Distribution of recombination sites in the recombinant viruses. Recombinant viruses from 37 individual transgenic plants were analysed by restriction enzyme digestion and from this analysis, separated into four classes, A–D, based on the site within the homologous region used for the cross-over. The *x* axis represents the homologous region between the *SalI* and *SacI* sites showing the postions of the distinguishing restriction enzyme sites *CfoI*, *BclI* and *DraI* which separate the recombination points into four classes. The *y* axis represents the frequency of the appearance of a virus in a class and is the number of plants containing a virus which recombined in this region divided by the length of the block in base pairs. In the lower part of the figure is a representation of the genetic information of this region of CaMV.

recombination events. As shown in Figure 3, viruses arising from recombination events distributed throughout the homologous region were observed, but $>50\%$ of the plants (21 out of 37) contained viral DNA with a restriction enzyme digestion pattern identical to the 4184/S type of sequence (class A), suggesting a hotspot for recombination.

The remaining 16 plants analysed fell into three classes

in approximately equal numbers. Class B viral DNAs contained all 4184/S derived restriction enzyme sites except the final D/H strain specific *Dra*I site, suggesting that these viral DNAs had been produced by recombination between this site and the *Bcl*I site. Class C DNAs contained 4184/S sites 5′ of and including the *Cfo*I site and D/H type sites beyond that, suggesting that the recombination occurred between the *Cfo*I and the *Bcl*I sites. The final class of viruses, D, contained DNA that originated by recombination 5′ to the *Cfo*I site.

No correlation was found between the type of viral DNA produced by one plant and those produced by its progeny; plants with identical genomic inserts (siblings from back-crossed plants) were found to produce viruses with different cross-over points (Southern blot data not shown). Some transgenic plants produced the virus early while others produced virus later in development (P.Swoboda and S.Gal, unpublished data), and there seems to be no particular recombinant produced at these different times.

Sequence analysis of recombinants

In order to confirm the predicted positions of the recombination sites and to analyse them more precisely, viral DNA in this region was cloned and sequenced. From each of the four classes described above, fragments from three independent plants were cloned and the entire 1019 bp region sequenced (Figure 4). We have indicated the simplest possible pattern for the recombination or crossing-over sites, defined as a region between distinguishing bases which is bordered by bases from the two different sequences. (The use of the words recombination or crossing-over site is not meant to imply a mechanism.) Many more such sites may exist but are silent because of the lack of sequence differences. For each virus isolate, 4–6 clones were

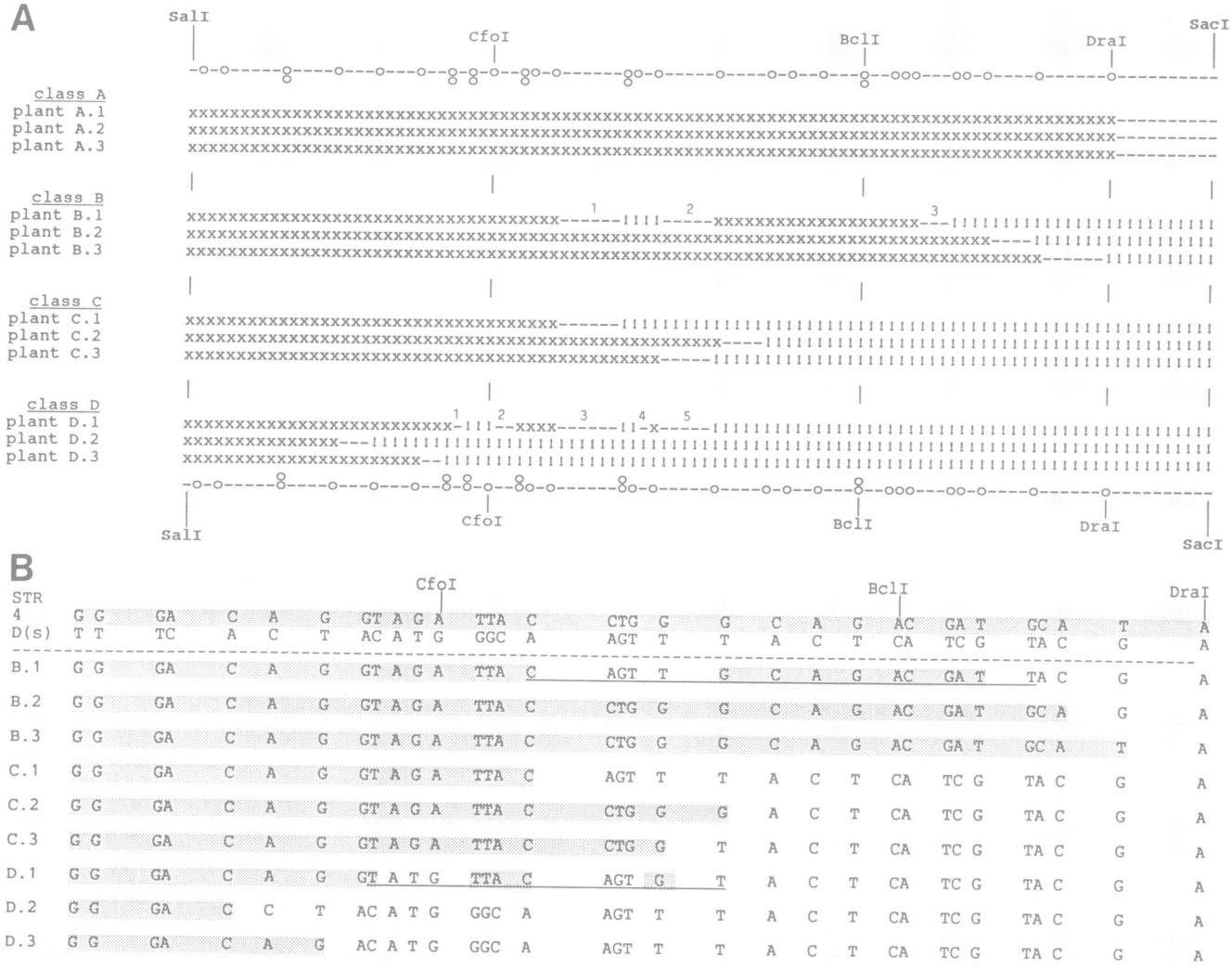

Fig. 4. Sequences of recombinant viruses from some transgenic plants. Sequences as symbols (A) or bases (B). **A.** Top and bottom lines show a comparison of the homologous sequences with the positions of the distinguishing restriction enzyme sites shown. Each character represents 10 bp, a '–' representing sequence identity between the two CaMV strains and each 'o' representing one base different in that 10 bp. The middle lines represent the sequences of the 12 recombinant viral DNAs from 12 different transgenic plants, classified as in Figure 3. The '×' indicates 4184/S CaMV strain; '!' indicates D/H CaMV; '–' indicates a cross-over region. Two plants (B.1 and D.1) showed multiple crossing-over sites which are marked with small numbers. **B.** The individual bases which differ between the two strains are indicated on the top lines: the D/H strain (D) as sense strand (s) with respect to transcriptional information; the 4184/S strain (4) in the complementary sense strand (shaded). The positions of the three distinguishing restriction enzyme sites are marked. The distance between the different base changes are approximated by the number of spaces between the bases. Below the line are the sequences for the nine viral DNAs from classes B, C and D labelled as in part A. Bases are shaded for the 4184/S sequence and not for the D/H sequence. For the two viral DNAs which contain multiple recombination points (B.1 and D.1), the length of the possible heteroduplex formed during the recombination event is underlined.

completely sequenced in this region; no differences in the sequences from the different clones of a single viral isolate were found. In some cases, direct sequencing of PCR products was performed (Gal and Hohn, 1990) and again, only one type of viral DNA was detected.

All class A viruses showed 4184/S sequences only (Figure 4A) confirming the restriction enzyme mapping, and indicating that the recombination occurred beyond the *Dra*I site. Because there are no other base pair differences between the two CaMV sequences used 3′ to this site, the cross-over point cannot be mapped more precisely. In the three other classes of virus, the site(s) of recombination was mapped to be within the expected regions of the homologous sequence. However, in 2 cases, multiple recombination events were detected (plants B.1 and D.1, Figure 4A). In the case of plant B.1, three recombination sites are found in the recombinant viral DNA, marked with small numbers in panel A of the Figure. In the case of plant D.1, 5 crossing-over points would explain the sequence of the recombinant DNA.

Discussion

A sensitive marker for intrachromosomal recombination

We have created plants which contain a partial dimer of CaMV that requires homologous recombination to form viable viral DNA. We believe that the construction is a marker for genomic recombination in plants for several reasons. First, the progeny plants of the original transformants produce virus. As virus is not detected in young plants from infected parents, the virus does not pass through seed. These plants later produce virus, following recombination, which is only possible if the entire genetic information for the virus were incorporated into the plant genome. Second, the types of recombinant viruses found in plants that have identical T-DNA insertions (siblings) vary between plants. This indicates that each type of virus is produced by an independent recombination event. If the constuction is integrated in its original form and there are no constraints or sequence preferences for the recombination event, we would expect all possible combinations of the homologous information to be present in the resultant viral sequences. In general, this is what we found. No particular type of virus has been observed in different lines of transgenic plants. This would argue against any pre-rearrangement of the transgene in a germ line cell, alleviating the need for recombination, which would then cause a plant in a subsequent generation to contain a virus identical to that in the parent and in its siblings. Also, the time of appearance of the virus in the transgenic plants is random (P.Swoboda and S.Gal, unpublished) which is what one might expect if the rearrangement were an infrequent event but once having occurred, virus could spread rapidly in the transgenic plant. The method we have devised is a sensitive indicator for recombination, one which enables scoring of intact plants and which lacks the selection pressure inherent in other recombination systems (Wirtz *et al.*, 1987; Peterhans *et al.*, 1990).

There is a recombinational hotspot

In general, all regions of the provided homology have been found to be used for recombination. However, a large proportion of the viruses in the transgenic plants are of one particular type or cross-over point, in the 3′-most part of the homologous stretch downstream of the *Dra*I site, and thus we refer to it as a hotspot for recombination. The recombinational hotspot may reflect a true event at the nucleic acid level and not selection of a particular virus. As noted earlier (Results), the base pair differences between the two viral sequences are conservative, and are therefore not likely to introduce a bias for or against particular recombinant viruses.

There are a number of possible explanations for the observation of a hotspot. The first is that this region contains the longest stretch of perfect homology (122 bp) between the two viral strains used to construct the duplicated regions. It has been shown for other systems, that the length of perfect homology affects the efficiency of the recombination process (bacteria, Shen and Huang, 1986; Watt *et al.*, 1985; mammalian cells, Waldman and Liskay, 1988).

Another possible explanation for the recombination hotspot is that in this region a part of the promoter for the 19S transcript of the virus is located (see Figure 3). Transcriptional enhancement of recombination has been observed in other systems, notably in yeast (Thomas and Rothstein, 1989; Stewart and Roeder, 1989) and mammalian tissue culture cells (Alt *et al.*, 1986; Blackwell *et al.*, 1986; Nickoloff and Reynolds, 1990). However, the possible involvement of the promoter to enhance the recombination in this region is not reflected by an increased number of recombinant viruses in the region just 5′ of the *Dra*I site (class B in Figure 3) nor in the positions of the crossing-over sites in the sequenced viral isolates of this class (Figure 4).

A third possible explanation for the recombination hotspot is that recombination occurred at the RNA level. During normal virus replication, reverse transcription of the 35S transcript involves a template switch at the 5′ end of the RNA after initiation at the primer binding site (marked 1 in Figure 1A) (Bonneville *et al.*, 1988). It is believed that the RNA terminal repeat serves to promote this switching which takes place primarily intramolecularly, but which can also occur between different 35S RNA molecules. Since the 19S RNA has an identical 3′ end, one can imagine that an 'illegal' template switch using this molecule would be possible (see legend to Figure 1A). The 19S transcript has been found in CaMV infected *B.napus* plants in large excess relative to a 35S transcript (Covey *et al.*, 1990). This proposed replication scheme would produce viable virus that would contain the 4184/S strain of viral DNA in the homologous region. At the sequence level, that is what we observe. 'Illegal' template switching has been proposed as a mechanism for the producuction of recombinant viruses after inoculation of turnips with mutant viruses (Stratford and Covey, 1989; Vaden and Melcher, 1990) and in agroinfection (Grimsley *et al.*, 1986b).

As noted earlier, the recombination site in the class A viruses cannot be mapped more precisely, for instance to the 5′ end of the 19S transcript, because there are no sequence differences in the strains in the region between the *Dra*I site and the *Sac*I site. Thus we cannot distinguish between these possiblities at the present time.

By what mechanism does the recombination event occur?

As stated above, the hotspot for the recombinant viruses could be explained by RNA recombination and template

switching. But for the other ~50% of the plants, DNA recombination is the most likely explanation. The form of the construction would not allow a viral promoter to produce an RNA moiety which contained the D/H strain of the homologous region unless a read-through product from the bacterial Tn903 gene was produced (see Figure 1A). We searched for such a product by Northern blots using RNA from transgenic plants with a probe for the Tn903 region and found no specific hybridization signal (data not shown). Therefore, any recombinant viral DNA containing D/H viral strain sequences most likely arose from DNA recombination. The existence of two examples of recombinant viral DNAs showing multiple recombination points is consistent with resolution of a DNA heteroduplex intermediate and not with multiple RNA template switching (see below). The reciprocal product of recombination, namely that left in the plant chromosome, would have to be analysed to distinguish between the possibility of gene conversion and reciprocal homologous recombination.

Multiple recombination events

Recombination events outside of the hotspot region were detected at several different sites. Of the nine viral DNAs analysed, two show multiple cross-over sites between the two viral sequences. Recombination events can only be detected if they occur in a region which shows differences between the two viral strains used. Thus, other crossing-over events may occur in the viral recombination targets but in regions where the sequences do not differ and therefore are 'silent' sites. The number of recombination points may therefore represent an underestimate. Multiple crossing-over events have been mapped at the restriction enzyme level in agroinfection experiments using CaMV (Grimsley *et al.*, 1986b) and a recently reported engineered animal retrovirus system (Hu and Temin, 1990).

There are two possible explanations for the observation of multiple crossing-over events. One is that there are multiple recombination events and that each resolves independently to form the recombinant viral DNA. This possibility is unlikely for the following reasons. The probability of two different recombination events occurring twice in the same plant would be a multiple of the probability of one single event. Extrachromosomal recombination between two different replicating viruses or post-escape gene conversion events, as the source of viral DNAs containing multiple cross-over sites, are unlikely possibilities because these plants contain only a single type of virus. However, 'pre-escape' microconversion events between the duplicated regions, as have been observed in yeast (Wheeler *et al.*, 1990) and chicken cells (Buerstedde *et al.*, 1990), could account for multiple crossing-over sites if that converted region were then used for recombination.

A second, more likely explanation for the multiple crossing-over points would be mismatch repair during the heteroduplex resolution. In most models of recombination, heteroduplex formation between interacting homologous sequences is a crucial intermediate in generating a cross-over (Holliday, 1964). A region of heteroduplex covering as few as 100 bp would be sufficient to generate multiple mismatches in the homologous region of 1 kb provided by our system. The resolution of these mismatched bases could produce the small regions representing the reciprocal parent strand. In other systems this has been proposed to explain structures found after formation of mismatches *in vitro* (Abastado *et al.*, 1984; Brown and Jiricny, 1988) as well as resolved products of recombination *in vivo* (Borts *et al.*, 1990; Wheeler *et al.*, 1990). When the recombinant viral DNAs are considered as a product of mismatch repair, two features are apparent. Firstly, the sites for the multiple cross-overs in the two recombinant viral DNAs, B.1 and D.1, are

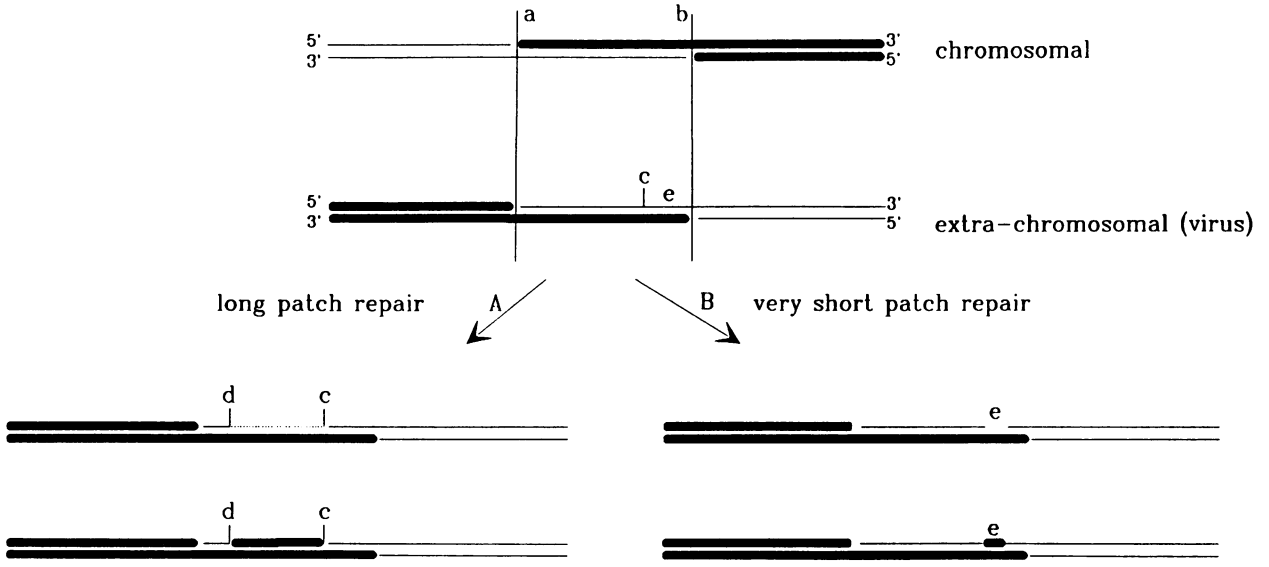

Fig. 5. A model for the generation of multiple recombination points (similar to Radman, 1988). The top two sets of lines represent the double-stranded products after reciprocal recombination to produce a chromosomal product and an extrachromosomal product which will produce the viral DNA. The length of heteroduplex between a and b could have been produced by strand invasion at one point and branch migration of the Holliday structure to the other. The extrachromosomal product can resolve the heteroduplex intermediate in two ways (A and B). One would involve nicking at c and 3′→5′ exonuclease activity to remove the upper strand to d, which would then be repaired using the lower strand as a template to form a patch of converted sequence. The other pathway would involve single base mismatch repair at e to remove the upper strand base and replace it with the corresponding lower strand base. The viral DNA of plant D.1 could be explained as having arisen by a combination of pathways A and B. These two mechanisms are similar to the pathways of long patch repair (A) and very short patch repair (B) proposed for prokaryotes (Radman, 1988).

in regions of multiple differences between the two strains. (These multiple differences are in fact required to detect multiple cross-overs.) Sites of multiple differences may be better targets for the repair enzymes since they are more likely to be recognized by them. These sites may have a more distorted structure due to the mismatches which could result in strand separation, initiating repair. However, not every region which contains a similar density of base pair differences was used for multiple crossing-over in the viruses analysed (for example in the region of the *BclI* site, Figure 4). Second, in the viral DNA from plant D.1, there is a single base cross-over site (between 4 and 5, Figure 4A). One possible heteroduplex between the two genomic sequences would produce a G-T mismatch (Figure 4B), which, if repaired to G-C, would favour the 4184 CaMV strain in that site. This is what we observe in the resultant virus. The repair of G-T mismatches primarily to G-C is known to be the preference for repair in human and bacterial systems (Modrich, 1987; Brown and Jiricny, 1988). Unfortunately, nothing is known about mismatch repair in plants for comparison.

We thus propose that, in at least the two plants in which multiple cross-over sites were observed, extended regions of heteroduplex were formed during the recombination event (at least 300 bp and 200 bp long for plants B.1 and D.1, respectively; underlined in Figure 4B). This heteroduplex was then resolved by small patch and single base mismatch repair to form the resultant recombinant viruses (Figure 5). Final resolution of the heteroduplex involving excision and repair of the lower strand (not shown for simplicity) would precede viral replication and spread to explain the existence of only one viral DNA type per plant. This model is similar to repair mechanisms in prokaryotes (Radman, 1988).

This is the first report of a system to monitor intra-chromosomal homologous recombination in a whole higher organism. The plants described provide a means of studying the various factors which control recombination. Future work will focus on plant development and environmental changes which may affect recombination in plants.

Materials and methods

Construction of recombination substrate

The 1.2 kb fragment carrying bacterial kanamycin resistance from Tn903 was cloned into pUC8. Next the 1.2 kb fragment was isolated after cutting this plasmid completely with *SalI* and partially with *Hin*dIII and ligating it to the large fragment resulting from cutting pHC79 (Hohn and Collins, 1980) with *SalI* and *Hin*dIII. This vector was then cut with *Bst*EII and *SalI* and the large fragment isolated and ligated to a similarly cut pCaMV6Km (Paszkowski *et al.*, 1986) to produce the plasmid pEAP1. Into the unique *SalI* site of pEAP1, one genome of CaMV strain D/H (Balazs *et al.*, 1982) was cloned to result in the vector pEAP21. This vector was recombined into *Agrobacterium tumefaciens* C58 strain 3850 (Zambryski *et al.*, 1983) by standard procedures. Figure 1A shows the described construction after the T-DNA has integrated into the plant genome.

The two homologous regions of CaMV DNA in this construction are contained between the *SalI* site (position 4814 relative to *Bst*EII at postion 126 in the D/H strain of CaMV) and the *Hin*dIII site (position 5828) as shown in Figure 1A. The sequence differences between the two provided regions are 34 bases over the 1033 bp. Sequences of the strains used have three bases different from that published in the available gene bank in the homologous region (S, Franck *et al.*, 1980; D/H, Balazs *et al.*, 1982, 1984 and Howarth *et al.*, 1981). The laboratory CaMV strain D/H has a C→G transversion at position 5208, and a T→C transition at position 5251 such that a second *BclI* site was created. The other homologous region is a combination of two CaMV strains with the first part from 4184 representing the sequence from *SalI* to *Eco*RV (positions 4413–5290 relative to *Bst*EII at 127) and the other part from S contained between the *Eco*RV and the *Hin*dIII sites (See Figure 1). The laboratory strain 4184 has one change of an A→G at position 4562.

Containment conditions

Plant transformation and agroinfection-experiments using *Agrobacterium* containing complete viral sequences were conducted in BL3 containment facilities.

Plant growth and transformation

Brassica napus var. Brutor plants were maintained under greenhouse conditions prior to transformation. Transgenic plants were kept in isolated chambers with growth conditions of 16 h light, 8 h dark at 24°C. Inoculation of turnips with transgenic plant extracts was performed as described previously (Grimsley *et al.*, 1986a).

The method of plant transformation was essentially that of Guerche *et al.* (1987) using co-inoculation of the *A.tumefaciens* strain containing the CaMV construction described above and a wild-type *A.rhizogenes* A4 strain on plant petiole pieces (Pisan, 1990). The resultant hairy roots were cultured and analysed for the expression of the transgene nopaline synthetase as described (Petit *et al.*, 1986; Aerts *et al.*, 1979). Transformed roots were regenerated and plants with sufficient root formation transferred to soil. The plants had a reduced seed set and wrinkled, dark green leaves typical of *A.rhizogenes*-transformed *B.napus* plants. Several independent transformed roots have been isolated, one of which was followed in this study. Probing of genomic DNA from the original transformant on a Southern blot with a T-DNA border specific probe showed multiple bands (data not shown) suggesting that this plant contained multiple insertions of the T-DNA at different loci. This has been confirmed by segregation analysis of both nopaline synthetase activity and viral DNA on Southern blots of selfed and back-crossed progeny plants. We have continued to back-cross these plants to obtain plants with single integration sites within the plant chromosome. The analysis of these plants will be published elsewhere (P.Swoboda, and S.Gal, unpublished data).

DNA, RNA isolation, Southern and Northern blotting

Total DNA isolations from plants were performed by two methods (Burr and Burr, 1981; Saghai-Maroof *et al.*, 1984). Isolation of CaMV viral particles was performed by the method of Gardner and Shepherd (1980). Southern blots were performed using Zeta probe nylon membrane (Bio-Rad Laboratories, Richmond, CA), with alkaline blotting and hybridization in the presence of non-fat dried milk as described (Sambrook *et al.*, 1989). RNA was isolated in the presence of guanidinium chloride (Sambrook *et al.*, 1989), separated by electrophoresis in formaldehyde gels (Fourney *et al.*, 1988), blotted to nitrocellulose membranes and hybridized as described (Shen *et al.*, 1986). Radioactive probes for Southern and Northern blots were made using the random primer labelling kit from Boehringer-Mannheim (Mannheim, Germany) and [α-^{32}P]dATP (Amersham, Little Chalfont, UK).

PCR amplification, cloning and sequencing of recombinant viruses

Material containing ~0.5 ng virus (0.1 fmoles) was amplified using primers starting 10 bp outside of the *SalI* and *Sac*I (position 5822) sites. Primer 1, located upstream of the *SalI* site had the sequence 5′-TGGACGAAGC-TTTCCGTGTGTTCAGAAAGT and primer 2, located downstream of the *Sac*I site had the sequence 5′-AGCTCTTGCTAAGCTTA-TTTTTGCTCTTAC. These primers would only amplify the sequence in a replicating virus and not from the integrated pre-recombination construction. The primers were added at a concentration of 1 μM with 200 μM of each nucleotide (Pharmacia, Uppsala, Sweden) in a buffer containing 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris–HCl, pH 8.4 at room temperature, and 0.01% BSA in a total volume of 100 μl. After heating the sample for 5 min at 95°C, 2 units of *Taq* polymerase (Perkin Elmer Cetus Corporation, Norwalk, CT) was added and 30 cycles were performed in the DNA Thermal Cycler Version 2.1 (Perkin Elmer Cetus Corporation) as follows: 1 min 94°C, 2 min 58°C, 4 min 72°C. The final cycle was followed by an additional 7 min at 72°C, and maintained at 4°C until use. Proteins were removed from the PCR products by phenol–chloroform extraction and buffer, nucleotides and primers were removed by two runs through a Centricon-30 column (Amicon, Danvers, MA). PCR products were analysed directly by cutting with restriction enzymes, separated on a 1% agarose gel and visualized with ethidium bromide.

For cloning and sequence analysis, viral DNA or PCR samples were digested with *Sac*I and *SalI* (Biofinex, Praroman, Switzerland) and the 1019 bp fragment cloned into standard vectors after separation in low melting point agarose (Sigma, St Louis, MO). This fragment lacks 26 bp of the homologous sequence (that between the *Sac*I and final *Hin*dIII sites), but

there are no sequence differences between the strains in this region. Sequencing was performed on double-stranded plasmid DNA preparations from an alkaline lysis procedure (Sambrook *et al.*, 1989) using Sequenase® (US Biochemicals, Cleveland, OH) and gel-purified viral DNA specific primers as well as the universal and reverse primers located on the vector.

Acknowledgements

The authors would like to thank Peter Swoboda, Joe Jiricny and Tom Brown for helpful discussions, Holger Puchta and Jean-Marc Bonneville for critical reading of the manuscript, Markus Briker for assistance with the plants, and Georges Pelletier for help with the plant transformation procedure. S.Gal was the recipient of a National Science Foundation (USA) postdoctoral fellowship.

References

Abastado,J.P., Cami,B., Dinh,T.H., Igolen,J. and Kourilsky,P. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 5792–5796.

Aerts,M., Jacobs,M., Hernalsteens,J.P. and Van Montagu,M. (1979) *Plant Sci. Lett.*, **17**, 43–50.

Alt,R.W., Blackwell,T.K., Depinho,R.A., Reth,M.G. and Yancopoulos,G.D. (1986) *Immunol. Rev.*, **89**, 5–30.

Balazs,E., Guilley,H., Jonard,G. and Richards,K. (1982) *Gene*, **19**, 239–249.

Blackwell,T.K., Moore,M.W., Yancopoulos,G.D., Suh,H., Lutzker,S., Selsing,E. and Alt,F.W. (1986) *Nature*, **324**, 585–589.

Bollag,R.J., Waldman,A.S. and Liskay,R.M. (1989) *Annu. Rev. Genet.*, **23**, 199–225.

Bonneville,J.M., Hohn,T. and Pfeiffer,P. (1988) In Domingo,E., Holland,J.J. and Ahlquist,P. (eds) *RNA Genetics. Vol. II. Retroviruses, Viroids and RNA Recombination*. CRC Press, Boca Raton, FL, pp. 23–42.

Borts,R.H., Leung,W.-Y., Kramer,W., Kramer,B., Williamson,M., Fogel,S. and Haber,J.E. (1990) *Genetics*, **124**, 573–584.

Brown,T.C. and Jiricny,J. (1988) *Cell*, **54**, 705–711.

Buerstedde,J-M., Reynaud,C-A., Humphries,E.H., Olson,W., Ewert,D.L. and Weill,J-C. (1990) *EMBO J.*, **9**, 921–927.

Burk,L.G. and Menser,H.A. (1964) *Tobacco Sci.*, **18**, 101–104.

Burr,B. and Burr,F.A. (1981) *Genetics*, **98**, 143–156.

Choe,I.S., Melcher,U., Richards,K., Lébeurier,G. and Essenberg,R.C. (1985) *Plant Mol. Biol.*, **5**, 281–289.

Covey,S.N., Turner,D.S., Lucy,A.P. and Saunders,K. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1633–1637.

Dixon,L., Nyffenegger,T., Delley,G., Martinex-Izquierdo,J. and Hohn,T. (1986) *Virology*, **150**, 463–468.

Engels,W.R. (1983) *Annu. Rev. Genet.*, **17**, 315–344.

Engler,P. and Storb,U. (1988) In Kucherlapati,R. and Smith,G.R. (eds) *Genetic Recombination*. American Society for Microbiology, Washington, DC, pp. 667–700.

Flavell,R.B. (1980) *Annu. Rev. Plant Physiol.*, **31**, 569–596.

Flavell,R.B. (1982). In Dover,G.A. and Flavell,R.B. (eds) *Genome Evolution*, Academic Press, London, pp. 301–323.

Fourney,R.M., Miyakoshi,J., Day,R.S.III and Paterson,M.C. (1988) *BRL Focus*, **10**, 5–7.

Franck,A., Guilley,H., Jonard,G., Richards,K. and Hirth,L. (1980) *Cell*, **21**, 285–294.

Gal,S. and Hohn,B. (1990) *Nucleic Acids Res.*, **18**, 1076.

Gardner,R.C. and Shepherd,R.J. (1980) *Virology*, **106**, 159–161.

Geldreich,A., Lébeurier,G. and Hirth,L. (1986) *Gene*, **48**, 277–286.

Grimsley,N., Hohn,B., Hohn,T. and Walden,R. (1986a) *Proc. Natl. Acad. Sci. USA*, **83**, 3282–3286.

Grimsley,N., Hohn,T. and Hohn,B. (1986b) *EMBO J.*, **5**, 641–646.

Gronenborn,B. (1987) In Hohn,T. and Schell,J. (eds) *Plant DNA Infectious Agents. Plant Gene Research*, Springer-Verlag, New York, pp. 1–21.

Guerche,P., Jouanin,L., Tepfer,D. and Péllétier,G. (1987) *Mol. Gen. Genet.*, **206**, 382–386.

Hawley,R.S. and Marcus,C.H. (1989) *Annu. Rev. Genet.*, **23**, 87–120.

Hohn,B. and Collins,J. (1980) *Gene*, **11**, 291–298.

Holliday,R. (1964) *Genet. Res.*, **5**, 282–304.

Howarth,A.J., Gardner,R.C., Messing,J. and Shepherd,R.J. (1981) *Virology*, **112**, 678–685.

Howell,S.H., Walker,L.L. and Walden,R.M. (1981) *Nature*, **293**, 483–486.

Hu,W.-S. and Temin,H.M. (1990) *Science*, **250**, 1227–1233.

Knudson,A.G.Jr (1985) *Cancer Res.*, **45**, 1437–1443.

Lébeurier,G., Hirth,L., Hohn,B. and Hohn,T. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2932–2936.

Modrich,P. (1987) *Annu. Rev. Biochem.*, **56**, 435–466.

Nickoloff,J.A. and Reynolds,R.J. (1990) *Mol. Cell. Biol.*, **10**, 4837–4845.

Paszkowski,J., Pisan,B., Shillito,R.D., Hohn,T., Hohn,B. and Potrykus,I. (1986) *Plant Mol. Biol.*, **6**, 303–312.

Peterhans,A. and Paszkowski,J. (1990) In Bennett,A.B. and O'Neill,S.D. (eds) *Horticultural Biotechnology*, Wiley-Liss, Inc., New York, pp. 155–170.

Peterhans,A., Schlüpmann,H., Basse,C. and Paszkowski,J. (1990) *EMBO J.*, **9**, 3437–3445.

Petes,T.C. and Hill,C.W. (1988) *Annu. Rev. Genet.*, **22**, 147–168.

Petit,A., Berkaloff,A. and Tempé,J. (1986) *Mol. Gen. Genet.*, **202**, 388–393.

Pisan,B. (1990) Ph.D. thesis, University of Basel, Basel, Switzerland.

Radman,M. (1988) In Kucherlapati,R. and Smith,G.R. (eds), *Genetic Recombination*, American Society for Microbiology, Washington, DC, pp. 169–190.

Saghai-Maroof,M.A., Soliman,K.M., Jorgensen,R.A. and Allard,R.W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 8014–8018.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Seperack,P.K., Strobel,M.C., Corrow,D.J., Jenkins,N.A. and Copeland,N.G. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 189–192.

Shen,D-W., Fojo,A., Roninson,I.B., Chin,J.E., Soffir,R., Pastan,I. and Gottesman,M.M. (1986) *Mol. Cell. Biol.*, **6**, 4039–4044.

Shen,P. and Huang,H.V. (1986) *Genetics*, **112**, 441–457.

Smith,G.R. (1989) *Cell*, **58**, 807–809.

Stewart,S.E. and Roeder,G.S. (1989) *Mol. Cell. Biol.*, **9**, 3464–3472.

Stratford,R. and Covey,S.N. (1989) *Virology*, **172**, 451–459.

Thomas,B.J. and Rothstein,R. (1989) *Cell*, **56**, 619–630.

Thompson,W.F. and Murray,M.G. (1981) In Strumpf,P.K. and Cona,E.E. (eds), *Biochemistry in Plants*, Academic Press, New York, pp. 10–81.

Vaden,V.R. and Melcher,U. (1990) *Virology*, **177**, 717–726.

Waldman,A.S. and Liskay,R.M. (1988) *Mol. Cell. Biol.*, **8**, 5350–5357.

Watt,V.M., Ingles,C.J., Urdea,M.S. and Rutter,W.J. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4768–4772.

Wheeler,C.J., Maloney,D., Fogel,S. and Goodenow,R.S. (1990) *Nature*, **347**, 192–194.

White,J.H., DiMartino,J.F. anderson,R.W., Lusnak,K., Hilbert,D. and Fogel,S. (1988) *Mol. Cell. Biol.*, **8**, 1253–1258.

Wirtz,U., Schell,J. and Czernilofsky,A.P. (1987) *DNA*, **6**, 245–253.

Zambryski,P., Joos,H., Genetello,C., Leemans,J., Van Montagu,M. and Schell,J. (1983) *EMBO J.*, **2**, 2143–2150.

Received on January 16, 1991; revised on February 27, 1991