4650–4657 *Nucleic Acids Research, 1997, Vol. 25, No. 22* *© 1997 Oxford University Press*

Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions

Vera Gorbunova and Avraham A. Levy*

Department of Plant Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

Received July 7, 1997; Revised and Accepted September 30, 1997

ABSTRACT

Double strand DNA breaks in plants are primarily repaired via non-homologous end joining. However, little is known about the molecular events underlying this process. We have studied non-homologous end joining of linearized plasmid DNA with different termini configurations following transformation into tobacco cells. A variety of sequences were found at novel end junctions. Joining with no sequence alterations was rare. In most cases, deletions were found at both ends, and rejoining usually occurred at short repeats. A distinct feature of plant junctions was the presence of relatively large, up to 1.2 kb long, insertions (filler DNA), in ~30% of the analyzed clones. The filler DNA originated either from internal regions of the plasmid or from tobacco genomic DNA. Some insertions had a complex structure consisting of several reshuffled plasmid-related regions. These data suggest that double strand break repair in plants involves extensive end degradation, DNA synthesis following invasion of ectopic templates and multiple template switches. Such a mechanism is reminiscent of the synthesis-dependent recombination in bacteriophage T4. It can also explain the frequent 'DNA scrambling' associated with illegitimate recombination in plants.

INTRODUCTION

Living organisms have evolved different repair mechanisms to eliminate highly lethal double strand breaks (DSBs) from chromosomal DNA. Repair of DSB is achieved either by homologous recombination or by illegitimate recombination via non-homologous DNA end joining. Non-homologous end joining is involved in the formation of a wide range of chromosomal rearrangements, including DNA integration, deletions, insertions, VDJ recombination and transposition. In higher eukaryotes, DSB repair and DNA integration occur more frequently via illegitimate recombination than via homologous recombination. This is in contrast to prokaryotes and lower eukaryotes where homologous recombination is the preferred pathway.

In vertebrates, inspection of junctions produced during naturally occurring chromosomal rearrangements showed that joining tends to preserve the original sequences (for a review, see 1). Changes, mainly restricted to the immediate vicinity of breakpoints, consist of a few base substitutions, nucleotide losses and occasional nucleotide additions. Furthermore, in all species examined, joining frequently occurs at short repeats (2–8). These features, often found in naturally occurring junctions, were confirmed in studies employing artificial linear DNA substrates with defined non-homologous terminal configurations. Such substrates were introduced *in vivo* to cultured cells, or *in vitro* to cell extracts (5,6,9–17). All these systems were shown to efficiently join DNA ends via a non-homologous joining process. It has been proposed that some alignment proteins which bind DNA termini facilitate end joining by stabilizing energetically weak single base matches (16). Recently, several proteins involved in non-homologous end joining have been characterized: Ku70, Ku 80, Rad50 (18–21) and yeast homolog of mammalian DNA ligase IV (22).

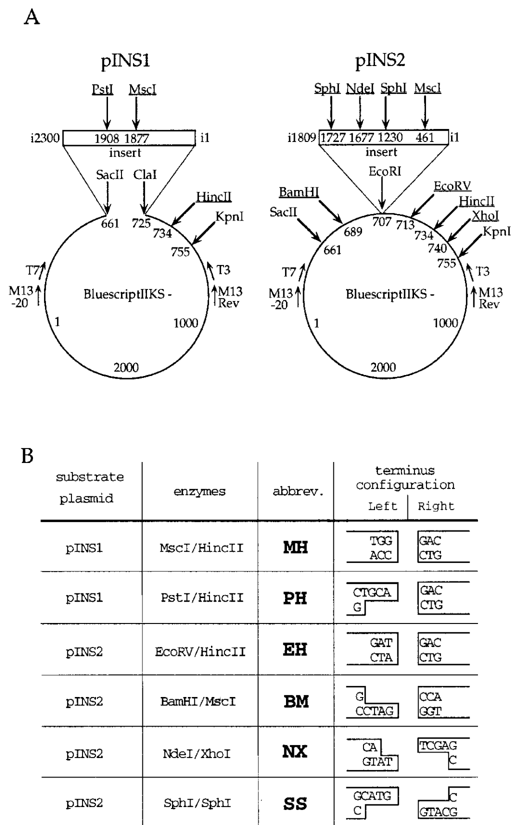
Extra nucleotides of filler DNA generated by terminal transferase activity are commonly found at VDJ joints in mammalian lymphoid cells (23–25). Filler DNA of up to 40 bp is also found in ~10% of the junctions in non-lymphoid cells that do not contain terminal transferase activity (5,6,26,27). Several possibilities have been proposed for the origin of filler DNA, such as patching with oligonucleotides and misincorporation of single nucleotides at the ends of DNA molecules (6,26,28). Filler DNA was also observed in *Drosophila*, following P element-induced gap repair (29). It was proposed to be generated through DNA synthesis primed by a broken end invading a non-homologous template. In yeast strains deficient in homologous recombination, DSB are repaired via non-homologous end joining and ~1% of the novel junctions contain filler DNA whose origin was, in all cases, from the cDNA of the yeast Ty1 retrotransposon (30,31).

In plants, DSB repair via homologous recombination received much attention (32–35) even though it is less efficient than non-homologous end joining. The only previous direct assays of non-homologous end joining in plants were through the analysis of footprints left upon transposable elements excision (36–39). These footprints usually consist of minor mutations, i.e. a few base pair deletions, additions or inversions. However, repair of the transposon-induced DSB might not be characteristic of non-homologous end joining, since transposase or other transposition-specific proteins are likely to play a role in protection and repair of the broken site. Information gathered from the analysis of naturally occurring junctions in plants contrasts with the data obtained from transposable elements. In

*To whom correspondence should be addressed. Tel: +972 8 9342734; Fax: +972 8 9344181; Email: lplevy@weizmann.weizmann.ac.il

maize, sequencing of five spontaneous deletions at the *Waxy* locus revealed that in all cases filler DNA (1–131 bp) was present at the novel joints (40). Another spontaneous deletion in maize, the *bz-R* allele, also contains filler DNA (41). Sequencing of one fast-neutron-induced and one X-ray-induced deletion in *Arabidopsis* revealed complex rearrangements, such as inversions and insertions at the deletion endpoints (3). Another example of non-homologous end joining in plants is T-DNA integration. In most cases this type of illegitimate recombination also shows deletions as well as the presence of filler DNA at the junction between T-DNA borders and the genomic insertion site (42). The available data on non-homologous end joining in plants, although limited, suggest that rearrangements at novel joints are more complex than in other organisms. If this is confirmed, non-homologous end joining adds to the panoply of processes conferring high plasticity to the plant genome (43), making it a potentially important factor in plant genome evolution.

In this study we demonstrate that plant cells are capable of joining the ends of linearized plasmids introduced by direct transformation. In the majority of events analyzed, end joining was associated with deletions ranging from a few base pairs up to 1 kb. Joining frequently occurred at short repeats. Only in a small proportion of the junctions, blunt or compatible cohesive ends were joined directly without any degradation. In a few cases, 5′ protruding single strands (PSSs) were filled-in. Interestingly, ~30% of junctions contained insertions of filler DNA ranging in size from 2 bp to 1.2 kb. Filler DNA was a patchwork of sequences homologous to internal regions of the plasmid, or to tobacco genomic sequences. Our data suggest that filler DNA was generated by repair synthesis involving copying from ectopic sites located in the same plasmid molecule or in the plant chromosome. These data demonstrating extensive rearrangements at sites of end joining, contrast with similar works in vertebrates, and make DSB repair via non-homologous end joining an important driving force in plant genome evolution.



substrate plasmid	enzymes	abbrev.	terminus configuration Left	Right
pINS1	MscI/HincII	**MH**	TGG / ACC	GAC / CTG
pINS1	PstI/HincII	**PH**	CTGCA / G	GAC / CTG
pINS2	EcoRV/HincII	**EH**	GAT / CTA	GAC / CTG
pINS2	BamHI/MscI	**BM**	G / CCTAG	CCA / GGT
pINS2	NdeI/XhoI	**NX**	CA / GTAT	TCGAG / C
pINS2	SphI/SphI	**SS**	GCATG / C	C / GTACG

Figure 1. (**A**) Structure of plasmids used for the analysis of end joining. pINS1 and pINS2 are bluescriptIIKS derivatives carrying inserts of known sequence within the polylinker. Linear DNA substrates derived from these plasmids were generated by duplicate restriction digests within the insert or the polylinker. Restriction sites used for linearization are underlined. Numbers within the plasmid circle correspond to the position in the bluescriptIIKS sequence as described in GenBank. Numbers within the insert box refer to the position within the insert with i1 corresponding to the first nucleotide of the insert. Small arrows represent the PCR primers used to amplify junction sites. (**B**) Terminus configurations of linearized substrates prepared from the indicated plasmids.

MATERIALS AND METHODS

Plant material

In all experiments, wild type *Nicotiana tabacum* var. *Samsun* plants were used.

Plasmids

Plasmids pINS1 and pINS2 (Fig. 1A) are bluescriptIIKS (Stratagene) derivatives carrying inserts of a known sequence within the polylinker; the inserts were chosen for the convenience of restriction sites. pINS1 contains a 2300 bp portion of a kanamycin resistance gene cloned into *Sac*II and *Cla*I sites, and pINS2 contains a 1809 bp cDNA of unknown function cloned into an *Eco*RI site.

Substrate preparation

Linear DNA substrates carrying different terminus configurations were prepared from pINS1 and pINS2 by duplicate restriction cuts (Fig. 1). Two different substrates with blunt/blunt (*Msc*I–*Hin*cII) and 3′PSS/blunt (*Pst*I–*Hin*cII) terminus configurations were derived from pINS1. Four different substrates with blunt/blunt (*Eco*RV–*Hin*cII), 5′PSS/blunt (*Bam*HI–*Msc*I), 5′PSS/5′PSS (*Nde*I–*Xho*I) and 3′PSS/3′PSS (*Sph*I–*Sph*I) were derived from pINS2. Restriction enzymes were used according to recommendations of the suppliers.

Protoplast transformation

Protoplasts were prepared from sterile tobacco leaves (44). Polyethylene glycol-mediated transformation was performed as described (45) without the addition of carrier DNA. In each experiment, five million protoplasts were transformed with 100 μg of linearized plasmid. After transformation, protoplasts were resuspended in 0.5 M sucrose-containing VKM medium (44) at a density of 10^5 cells/ml, and kept overnight in the dark at 25 °C.

Plant DNA extraction

Total DNA was isolated from young leaves of tobacco as described (46). To extract DNA from protoplasts the same protocol was used with two modifications: the volume of all buffers was scaled down in order to perform the procedure in an Eppendorf tube, and the isopropanol precipitation step was omitted.

PCR and sequence analysis of end joining products

Total DNA extracted from 2.5 million protoplasts was amplified with nested PCR with standard M13-20/M13Rev primers in the first round and T3/T7 primers in the second round. An aliquot of 2 µl of the reaction from the first round was used as a template in the second round. All PCR amplifications were performed with a cycle of 1 min denaturation at 94°C, 1 min of annealing at 55°C and 1 min of extension at 72°C, repeated 30 times. PCR products were digested with *Sac*II+*Kpn*I, cloned into the corresponding sites of the bluescriptIIKS, and sequenced with M13-20 and/or M13Rev primers. Sequencing was done by the Sanger's method using an Applied Biosystems 373A DNA sequencer.

Southern analysis

For Southern analysis, 5 µg of genomic DNA was digested with *Xho*I+*Xba*I, *Spe*I, *Bgl*II and *Eco*RV, fractionated on 0.8% agarose gel, and transferred to a nitrocellulose membrane purchased from MSI. Hybridization was performed according to manufacturer's instructions. A 150 bp *Sac*II–*Kpn*I fragment from the pINS1-PH8 plasmid, containing the filler DNA, was radiolabeled by the random priming method (47) and used as a probe.

RESULTS

Experimental system

To study non-homologous end joining in plant cells, we transformed tobacco protoplasts with linear DNA fragments derived from plasmids digested by six different combinations of duplicate restriction cuts (Fig. 1B). Plasmids used in the experiments are shown in Figure 1A. Following transformation, protoplasts were incubated overnight and total DNA was extracted and amplified by PCR with primers T3+T7 and M13-20+M13rev (Fig. 1A). Since the plasmids were digested between the primer annealing sites, no PCR products could be obtained unless the plasmids were circularized or two plasmids joined. PCR products of various sizes were obtained (data not shown), indicating end joining of linear molecules. To exclude the possibility that PCR products arose as a result of PCR artifacts, linearized plasmids were mixed with tobacco genomic DNA and amplified with the same set of primers as the DNA extracted from protoplasts. No PCR products were obtained (data not shown).

Joining can occur between compatible ends (head to head) or between different ends (head to tail). To eliminate head to head junctions, PCR products were digested with *Sac*II and *Kpn*I (Fig. 1A) and cloned into the corresponding sites of bluescript. Such cloning procedure also eliminates potential non-specific PCR products.

Sequence analysis of novel junctions

The PCR-amplified novel junctions did not have preferential size. Out of 52 clones, 34 different sequences were found. These sequences are presented in Figure 2, and the characteristics of the junctions are summarized in Table 1. To exclude the possibility that junctions were generated in *Escherichia coli* as the result of modifying the cloning vector, only junctions that retained some of the insert sequences from pINS1 and pINS2 (Fig. 1) were considered.

Table 1. Number of base pairs deleted, inserted, or matched at sites of end joining

Junction[a]	Base pairs deleted[b]		Match[c]	Insert
	Left	Right	base pairs	base pairs
MH1	0	0	0	0
MH2	88	0	1	0
MH3	90	8	0 (13/16)	0
MH4	85	7	5 (13/16)	0
MH5	77	13	7 (13/16)	0
MH6	0	22	–	2
MH7	0	14	2 (7/10)	0
MH8	3	0	0	0
MH9	263	7	–	277
MH10	0	12	–	200
BM1	23	284	–	205
BM2	0	406	–	36
NX1	9	7	2 (4/5)	0
NX2	1	0	1	0
NX3	2	7	1	0
NX4	1	10	2	0
NX5	0	3	1	0
PH1	45	14	7 (11/13)	0
PH2	54	9	3 (11/13)	0
PH3	53	8	4 (11/13)	0
PH4	135	13	2 (5/6)	0
PH5	163	16	8	0
PH6	157	0	0	0
PH7	273	15	–	116
PH8	385	18	–	180
SS1	0	0	4	0
SS2	9	460	2	0
SS3	14	609	0	0
SS4	17	471	5	0
SS5	2	130	2	0
SS6	4	1058	–	170
SS7	12	1117	–	1229
SS8	5	1190	–	1056
EH1	0	0	0	0

[a]Junctions sequences are shown in Figure 2.
[b]The exact number of deleted bases at each end cannot be determined when joining occured at regions of microhomology, in these cases the number of bases distal to microhomology region is given.
[c]The number of perfectly matched nucleotides at the junction is reported; numbers in parenthesis show longer imperfect matches.

```
pINS1   MH          MscI ↓   ↓ HincII
·····················TGG     GAC··········
·····················ACC     CTG··········

MH1·····················TGG/GAC··········

··········CAG·85bp·TGG       GAC··········
MH2·····················CAGAC··········

·GCGCAGGGGCGCCCGGT·74bp·TGG    GACCTCGAGGGGGGCCCGGTA
MH3··············GCGC/GGGGGGCgCCGGTA··········
MH4··············GCGCAGGGGGGCCCGGTA··········
MH5··············GCGCAGGGGCGCCCGGTA··········

·····················TGG     GAC·19bp/C··········
MH6·····················TGG/ga/C··········

··········GGCTATCGTGGCTGG      GACCTCGAGGGGGG······
MH7··············GGCTATCGTGGCTGGG··········

··················GC/TGG     GAC··········
MH8·····················GC/GAC··········

··········TGGA/260bp·TGG     GACCTCG/AGGGGGGG·····
MH9················TGGA/ins 277bp/AGGGGGGG········

·····················TGG     GACCTCGAGGGG/GGG·····
MH10··················TGG/ins 200bp/GGG········
```

```
pINS2   BM          BamHI ↓   ↓ MscI
·····················G       CCA··········
·····················CCTAG   GGT··········

··········CGG/22bp·G         CCA·281bp/TCA········
BM1················CGG/ins 205bp/TCA··········

··············TAGTG          CCA·403bp/TCT········
··············ATCACCTAG      GGT·403bp/AGA········
BM2·············TAGTG/GATC/ins 36bp/TCT··········
```

```
pINS2   NX          NdeI ↓   ↓ XhoI
·····················CA      TCGAG··········
·····················GTAT    C··········

··········AAGGTGCATAGCA        TCGAGGGGGGGGCC······
NX1·················AAGGGGCCC··········

··········AAGGTGCATAGCA        TCGAGGGGGGGGCCC······
··········TTCCACGTATCGTAT      CCCCCCCGGG······
NX2············AAGGTGCATAGCATCGAGGGGGGGGCCC··········
···············TTCCACGTATCGTAGCTCCCCCCCGGG··········

··········AAGGTGCATAGCA        TCGAGGGGGGGGCCC······
NX3··············AAGGTGCATAGGGGCCC··········

··········AAGGTGCATAGCA        TCGAGGGGGGGGCCC······
NX4················AAGGTGCATAGCCC··········

··········AAGGTGCATAGCA        TCGAGGGGGGGGCCC······
··········TTCCACGTATCGTAT      CCCCCCCGGG······
NX5·············AAGGTGCATAGCATA/GGGGCCC··········
················TTCCACGTATCGTAT/CCCCGGG··········
```

```
pINS1   PH          PstI ↓   ↓ HincII
·····················CTGCA   GAC··········
·····················G       CTG··········

AGGGGCGCCCGGT/40bp·CTGCA      GACCTCGAGGGGGGGCCCGGTA
PH1··············AGGGGCGCCCGGTA··········
PH2··············AGGGGGGCCCGGTA··········
PH3··············AGGGGGGGCCCGGTA··········

····TTGGGTGG·130bp·CTGCA      GACCTCGAGGGGGGGCCC····
PH4·················TTGGGTGGCCC··········

··ATCCGGTACC·158bp·CTGCA      GAC·13bp·CCGGTACCC····
PH5·················ATCCGGTACCC··········

·········GGC/152bp·CTGCA      GAC··········
PH6····················GGC/GAC··········

·········TAC/268bp·CTGCA      GACCTCGAGGGGGGG/CCC···
PH7··············TAC/ins 116nt/CCC··········

·········GTT/380bp·CTGCA      GACCTCGAGGGGGGGCCC/GGT
PH8··············GTT/ins 180nt/GGT··········
```

```
pINS2   SS          SphI ↓   ↓ SphI
·····················GCATG   C··········
·····················C       GTACG··········

·····················GCATG   C··········
·····················C       GTACG··········
SS1·····················GCATGC··········

··········AATTGTATGCATG        C·459bp·TTAA······
SS2·····················AATTAA··········

········CA/GAATTGTATGCATG      C·608bp/TCC······
SS3·····················CA/TCC··        ····

··GTTCAAAACAGAATTGTATGCATG     C·470bp·TCAAATC······
SS4··················GTTCAAATC··········

·····················GCATG     C·125bp·CAGT······
·····················C         GTACG·125bp·GTCA······
SS5·····················GCAGT··········
························CGTCA··········

················ATG/CATG       C·1057bp/ACT······
SS6··············ATG/ins 170bp/ACT··········

·········AGA/ATTGTATGCATG      C·1116bp/CAT······
SS7··············AGA/ins 1229bp/CAT··········

················TAT/GCATG      C·1189bp/AAT······
SS8··············TAT/ins 1056bp/AAT··········
```

```
pINS2   EH          EcoRV ↓   ↓ HincII
·····················GAT     GAC··········
·····················CTA     CTG··········

EH1················GAT/GAC··········
```

Figure 2. Novel DNA sequences at the end junctions. The shaded box shows the sequence of both strands of the substrate termini, the restriction enzymes used for linearization and the plasmid used (pINS1 or pINS2 described in Fig. 1A). End joining products are designated by two capital letters, each representing the restriction enzyme (P, *Pst*I; H, *Hinc*II; M, *Msc*I; B, *Bam*HI; S, *Sph*I) used for substrate linearization. In some of the boxes, the line(s) above end joining products show sequences of the substrate, in the terminal and subterminal regions, that are related to the sequence at the novel junction. Similarity regions at the substate termini are underlined. Identical sequences at the junctions which cannot be attributed to either substrate end are double underlined, additional matches in the adjacent regions are single underlined. In the junctions showing no homology between the joined ends the break between parental sequences is indicated by a slash. Lower case letters designate bases of unknown origin. When the large portion of the subsrate end was deleted, instead of the full sequence the number of deleted base pairs is given, similarly, the sequence of large insertions (ins) is not shown but, instead, the size of the insertion is indicated. Complementary strands are shown when end joining is thought to involve filling in (NX2, NX5 and SS5) or annealing of restriction overhangs (SS1).

The majority of junctions were associated with deletions in at least one of the termini. The size of the deletions varied from a few base pairs to 1.1 kb (Table 1). Probably, the only limiting factor for deletion size was the loss of the primer binding site. Even when substrates with blunt or compatible cohesive termini were used, perfect joining, i.e., no sequence alterations, was observed only in three out of 19 clones, whereas the rest of the clones contained deletions.

A characteristic feature of junctional breakpoints was the presence of short repeats of 1–16 bp (Fig. 2 and Table 1). Short repeats were observed in 18 out of 24 junctions that did not carry filler DNA (Table 1). Three (MH1, EH1 and NX5 in Fig. 2) out of six junctions with no repeats were apparently generated by blunt end ligation. These data suggest that the mechanism for end joining involves the recognition of patch homologies by overlap formation between fortuitously matching base pairs of interacting DNA termini. In cases where end joining occurred at sites of relatively long (12–16 bp) imperfect repeats (junctions MH3-5 and PH1-3 in Fig. 2), novel junctions usually retained one of the 'parental' sequences at the site of a repeat, except for junction MH3 where the resulting sequence differed from either 'parent'.

In three out of 23 junctions, where substrates with cohesive termini were used, complementary nucleotides were filled in. Such a small proportion of 'fill-in' junctions is possibly explained by a very strong tendency to end degradation.

Another outstanding feature of junctions generated in tobacco cells was the presence of filler DNA. Filler DNA was observed in 10 out of 34 junctions and its length varied from 2 bp to 1.2 kb. Filler DNA was frequently associated with deletions in one or both termini (Table 1).

Origin of filler DNA at junctional breakpoints

Filler DNA in nine junctions was long enough to enable further analysis. DNA sequences were compared to the GenBank database. The structure of the filler DNA and its junctions with plasmid termini is summarized in Figure 3. Junctions PH7, BM1, BM2, MH9, MH10 and SS6 contained each a stretch of filler DNA which is identical to a certain region of bluescript (backbone of the pINS1 and pINS2 plasmids), located quite far from the linearized termini. In junctions PH7, BM1 and SS6, 1–3 nucleotides of unknown origin were found at the borders between the plasmid and the filler DNA. In PH8 junction, the filler DNA was not homologous to any GenBank sequence. Finally, filler DNA in junctions SS7 and SS8 (Fig. 3B) had a complicated structure, consisting of multiple reshuffled sequences homologous to bluescript, and of regions that did not have homology in GenBank. Interestingly, bluescript homologous regions were found in both direct and inverse orientations relative to pINS2.

Junctional breakpoints between the plasmid and filler DNA contained microhomologies of 1–11 bp at least at one insert termini. In the filler DNA of junctions SS7 and SS8 joining between bluescriptIIKS homologous segments also occurred at short repeats.

To study the origin of filler DNA in the PH8 junction, tobacco genomic DNA was digested with restriction enzymes that do not cut within the filler DNA, and was hybridized to a probe corresponding to the whole 180 bp filler DNA segment. The resultant Southern blot is shown in Figure 4. Strong hybridization signal and multiple bands were obtained, indicating that the filler DNA was derived from multicopy tobacco genomic sequences.

DISCUSSION

Sequence analysis of novel joints formed upon non-homologous end joining of linear fragments introduced into plant cells shows the following features: (i) formation of large deletions; (ii) frequent occurrence of short repeats at the deletion breakpoints; (iii) insertions of filler DNA at the novel joints; (iv) filler DNA of simple or complex structure; and (v) the presence of short repeats at the sites of switching from the substrate to the filler DNA sequence. These rearrangements are more complex than those typically found with non-homologous end joining in other non-plant species.

We discuss below how the assay described in this work can be used to study non-homologous end joining in plants, and the underlying mechanism of non-homologous end joining. We also compare non-homologous end joining in plants and other species.

Illegitimate recombination in plants

The extensive rearrangements described here, namely, large deletions and simple or complex filler DNA insertions; either a general feature of illegitimate recombination and DSB repair in plants or they might be limited to the the experimental conditions of our end joining assay. Although non-homologous end joining has not been systematically studied in plants, insight into its mechanism can be obtained through previous analysis of spontaneous (40) or induced (3) deletions, or through the analysis of T-DNA insertions into the plant genome (42). In all cases deletions and filler DNA with varying degree of complexity were observed. In addition, recent work in our laboratory (48) also shows that abortive DSB repair in plants is associated with the occurrence of deletions and filler DNA. Therefore, we conclude that the non-homologous end joining system described here is a good assay for DSB repair mechanisms occurring in genomic DNA. Taken together, our data and previous reports suggest that illegitimate recombination in plants occurs through an error prone mechanism. Such a mechanism has important evolutionary implications. Plants differ from other higher eukaryotes by their immobile life style which requires high plasticity in body structure, physiology and genome. Consequently, plants might have been selected to tolerate polyploidy and extensive DNA rearrangements (43). The error prone DSB repair mechanism described here may contribute to the plasticity of the plant genome: insertion of filler DNA from templates *in cis* or *in trans* can cause gene amplifications and increase the amount of repetitive DNA in plants.

Deletions are often large and are associated with short repeats

Only three out of 34 novel joints consisted of perfect joining, and only in three joints was one of the cohesive termini filled in (Table 1). The majority of end-joining events involved deletions ranging from a few base pairs to 1.2 kb with an average of 250 bp (Table 1). These deletions are larger than those reported in similar assays in other organisms. In *Xenopus*, the best characterized *in vitro* end-joining system, deletions are rare events (15). Larger deletions (up to 150 bp) were found *in vivo*, in injected *Xenopus* eggs (49). In *Schizosaccharomyces pombe*, nucleotides are frequently lost *in vivo* from PSS tails, but degradation usually does not affect the duplex regions (10). In mammalian *in vivo* systems, a somewhat more extensive terminal degradation was observed, although it

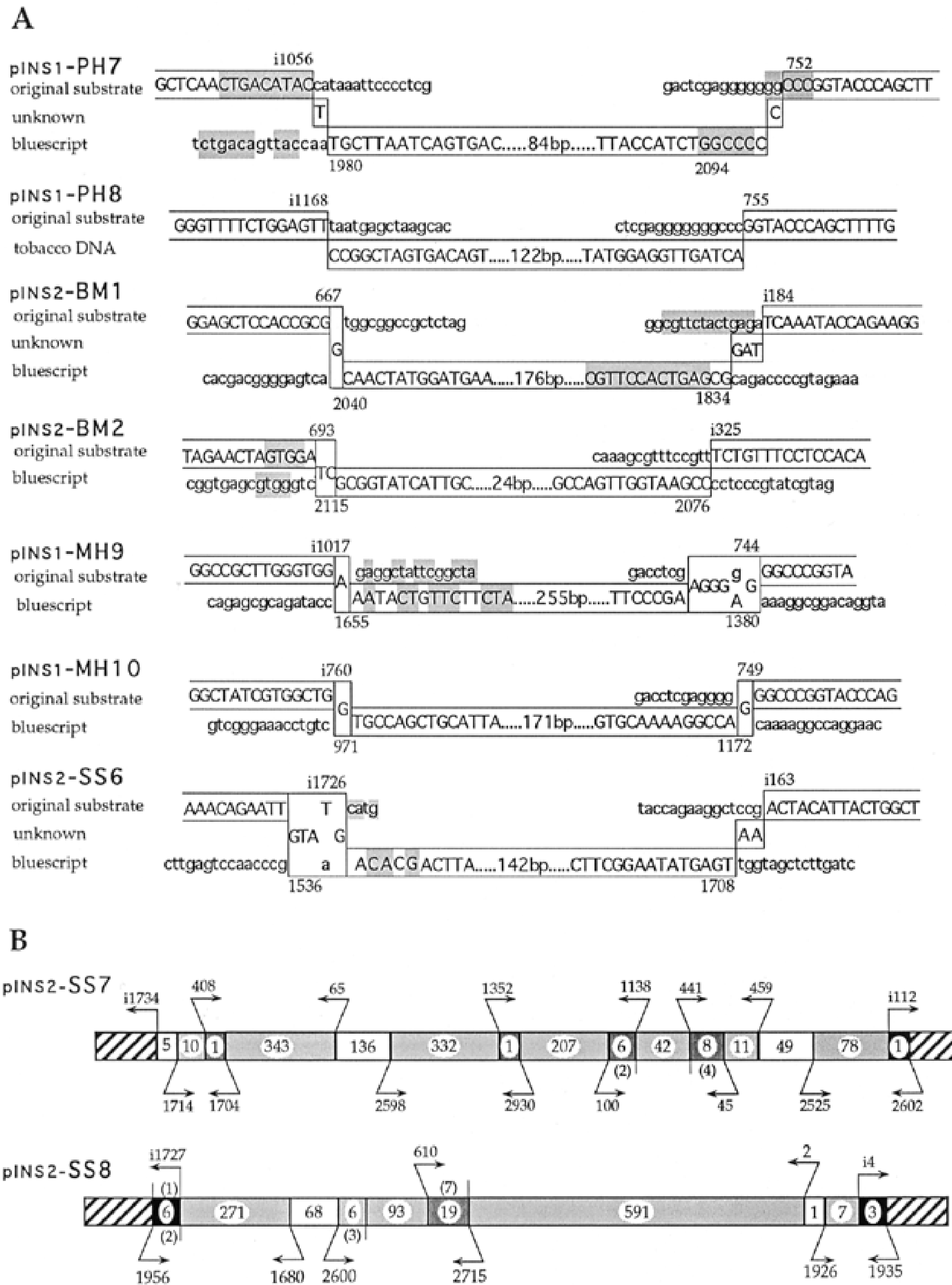


Figure 3. Structure of novel end junctions carrying filler DNA. Junctions are designated as in Figure 2 and Table 1. Substrate ends generated by restriction enzymes are shown in the same orientation as in Figure 2. (**A**) Junctions carrying 'simple' insertions, i.e. filler DNA corresponding to a single fragment of known homology. Sequences of the novel junctions are shown in capital, boxed letters. Small letters indicate the regions flanking the recombination sites in the original ends (upper line) or in the segment of Bluescript homologous to the filler DNA (bottom line). Numbers correspond to the positions in the bluescriptIIKS sequence; those with the letter i denote the position in the insert of the corresponding plasmid (see Fig. 1). Regions of homology that might have played a role in junction formation are shaded. (**B**) Junctions carrying filler DNA of a complex structure. Numbers on the arrows refer to positions in the bluescriptIIKS sequence; numbers preceded by the letter i refer to positions in the insert of pINS2. Hatched boxes represent the termini of the substrate plasmid; shaded boxes represent segments of the filler DNA which are 100% homologous to the bluescriptIIKS; filled boxes denote the regions where homologous segments overlap, with the number of mismatches given in brackets on the same side where the position of the corresponding segment is marked by the arrows; white boxes represent regions which are non-homologous to the bluescriptIIKS. Numbers inside the boxes correspond to the number of base pairs within the segment.

rarely exceeded 15 bp from the original termini (5,6). Overall, deletions reported here in transfected plant cells are larger (up to 1200 bp) than those typically reported in other species. Moreover, the upper limit to deletion size in our assay was the position of the primers (deletions spanning the primer annealing site could not be amplified). Such differences may reflect a stronger exonuclease activity or a lesser protection of DNA ends in plant cells compared to other species. Interestingly, in yeast increased deletion

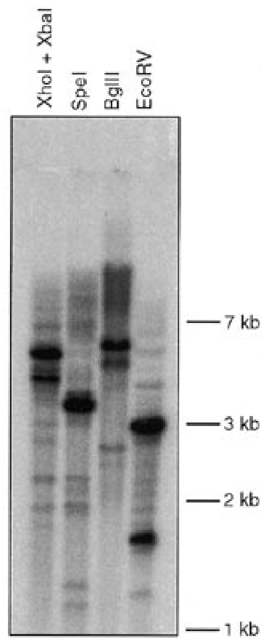


Figure 4. Genomic Southern blot probed with the filler DNA from the junction PH8. Genomic DNA was extracted from leaves of wild type tobacco, digested with the enzymes indicated and probed with the *Kpn*I/*Sac*II fragment of plasmid PH8 carrying the filler DNA.

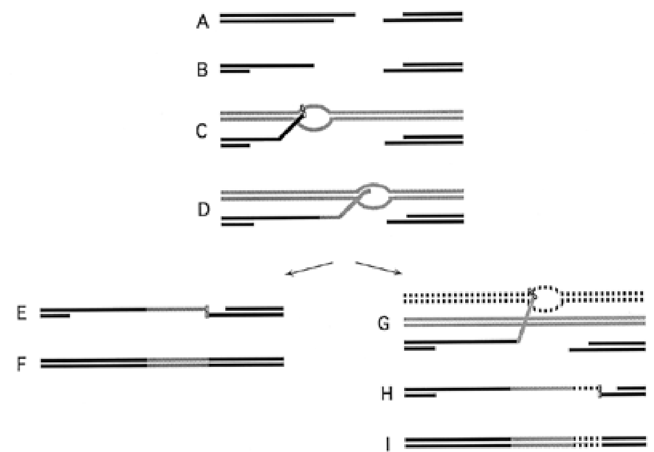


Figure 5. Model for filler DNA formation. (**A**) DNA ends before joining. (**B**) DNA ends are degraded and protruding 3′ single strands are formed. (**C**) A protruding single strand invades an ectopic site that shares a short region of homology (small boxes). (**D**) DNA synthesis proceeds according to the SDSA mechanism (57), with the newly synthesized strand being immediately displaced from the template. The newly synthesized DNA corresponds to filler DNA. When synthesis is aborted, the free single-stranded tails can anneal at a region of microhomology (**E**) and DNA synthesis and nick ligation complete the repair process (**F**), resulting in a simple filler DNA formation. Alternatively, the free single standed tails can reinvade a new template (**G**) leading to the formation of complex filler DNA structure (**H, I**).

size at the junctions formed by non-homologous end joining was found in cells defective for Ku homologs (18,19). The error prone DSB repair reported here for plants might be related to the expression of the Ku-like proteins.

In most deletions analyzed, end joining ocurred at short repeats ranging in size from 1 to 16 bp (Table 1). This finding is similar to previous reports regarding non-homologous end joining in other species (2–8). The frequent occurrence of junctional breakpoints at short repeats seems therefore to be a highly conserved feature of the DSB repair process. Various mechanisms were proposed to account for the generation of such junctions. Short repeats could be exposed by single strand degradation of duplex DNA followed by subsequent annealing of the single strands at regions of microhomologies (4,50). Alternatively, end joining could occur via single strand invasion in duplex DNA, followed by annealing at regions of microhomologies and nucleolytic trimming of the displaced single strand(s). Once complementary nucleotides anneal, a transient joint can be stabilized, either by ligation or by DNA synthesis. Lehman *et al*. (49) showed that end joining in *Xenopus* oocyte nuclear extracts does not occur in the absence of dNTPs or in the presence of aphidicolin (an inhibitor of DNA synthesis). They proposed that 3′ tails might prime DNA synthesis at sites of microhomology, and that duplex elongation might stabilize the newly formed joint. The role of DNA synthesis in the end joining process is further discussed below.

Models for filler DNA formation

Our finding that filler DNA was homologous to sequences of the substrate plasmid or of the tobacco genome indicates that it was not produced by untemplated repair synthesis. It could in principle originate from the intercalation and ligation of DNA fragments present in the cell. However, filler DNA such as presented in Figure 3B is highly unlikely to be formed by a multiligation process. In addition, previously reported filler DNA found at the sites of chromosomal rearrangements was in many cases derived from sequences near the junctional breakpoints (40,41,51–56), making intercalation of ‘floating’ DNA fragments an unlikely process. A more plausible mechanism for filler DNA insertion is through templated DNA synthesis induced upon DSB repair. Repair synthesis via synthesis-dependent strand annealing (SDSA) (57) or a similar mechanism can best explain the observed complex filler DNA insertions. Such a repair mechanism was recently proposed to explain the occurrence of filler DNA associated with illegitimate integration of transfected DNA in mammalian genome (27).

According to the SDSA DSB repair model, protruding 3′ ends formed at the break site invade a homologous template. DNA synthesis proceeds independently from each 3′ end, with newly synthesized DNA being immediately displaced from the template. This process eventually produces complementary single-stranded ‘tails’. Then the two tails anneal and the remaining single stranded regions are filled-in thus completing the repair process.

Filler DNA can be produced (Fig. 5) when, instead of an homologous sequence, an unrelated template is invaded and synthesis is primed by a short region of homology (Fig. 5C). Unrelated template can be invaded in the absence of an homologous template (like in our extrachromosomal system) or when homologous sequence is not readily available as for example during chromatin condensation, chromatids separation etc. After a portion of the template has been copied (Fig. 5D), synthesis can be aborted and newly synthesized DNA is displaced from its template and joins to the second end (Fig. 5E). Such a

process would result in a simple filler DNA insert (Fig. 5F). Alternatively, a new template can be reinvaded (Fig. 5G) generating filler DNA of a complex structure (Fig. 5H–I), as found in this work (Fig. 3B). Filler DNA can be captured *in cis* from the same molecule, or *in trans* from another molecule or from genomic DNA. From our data and from previous literature (40,41,51–56) capturing *in cis* seems to be more efficient. SDSA is a potentially error prone mechanism, since unstable interaction between the newly synthesized DNA and the template can facilitate abortion of DNA synthesis and subsequent template switch. It might therefore explain filler DNA and complex rearrangements accompanying DSB repair in plants. SDSA model was also proposed by Nassif *et al.* (29) to explain complex conversion events observed upon P element excision. Similarly, SDSA could explain DNA rearrangements associated with abortive DSB repair in plants (48).

For the bluescript-derived filler DNA, we could establish the regions flanking the invasion sites. This enabled us to identify the putative microhomology regions involved in strand invasion and template switching. Interestingly, we found that in several cases (Fig. 3) very short regions of homology were sufficient to prime DNA synthesis from the invading strand, in contrast to *in vitro* systems like PCR, where considerable homology is required for priming. This suggests that *in vivo*, annealing and priming of DNA synthesis might be facilitated by putative alignment factors. Furthermore, these data provide direct molecular evidence, in addition to previous biochemical evidence (49), for the importance of DNA synthesis during end joining, and for the role played by regions of microhomology in priming such synthesis.

ACKNOWLEDGEMENTS

We thank E.Rubin for critical reading of the manuscript, Y.Avivi for editing the manuscript and D.Berlin for technical help. This work was supported by the Forchheimer foundation and by Dr Chutick's doctoral fellowship to V.G.

REFERENCES

1. Roth,D.B. and Wilson,J.H. (1988) In Kucherlapati,R. and Smith,G.R. (eds), *Illegitimate Recombination in Mammalian Cells*. American Society for Microbiology, Washington, DC, pp. 621–653.
2. Stringer,J.R. (1982) *Nature* **269**, 363–366.
3. Shirley,B.W., Hanley,S. and Goodman,H.M. (1992) *Plant Cell*. **4**, 333–347.
4. Thaker,J., Chalk,J., Ganesh,A. and North,P. (1992) *Nucleic Acids Res*. **20**, 6183–6188.
5. Roth,D.B., Porter,T.N. and Wilson,J.H. (1985) *Mol. Cell. Biol*. **5**, 2599–2607.
6. Roth,D.B. and Wilson,J.H. (1986) *Mol. Cell. Biol*. **6**, 4295–4304.
7. Ruley,H.E. and Fried,M. (1983) *Nature* **304**, 181–184.
8. Marvo,S.L., King,S.R. and Jaskunas,S.R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1391–1395.
9. Goedecke,W., Vielmetter,W. and Pfeiffer,P. (1992) *Mol. Cell. Biol*. **12**, 811–816.
10. Goedecke,W., Pfeiffer,P. and Vielmetter,W. (1994) *Nucleic Acids Res*. **22**, 2094–2101.
11. Lehman,C.W., Clemens,M., Worthylake,D.K., Trautman,J.K. and Carroll,D. (1993) *Mol. Cell. Biol*. **13**, 6897–6906.
12. Munz,P.L. and Young,C.S.H. (1991) *Virology* **183**, 160–169.
13. North,P., Ganesh,A. and Thacker,J. (1990) *Nucleic Acids Res*. **18**, 6205–6210.
14. Pfeiffer,P., Thode,S., Hancke,J. and Vielmetter,W. (1994) *Mol. Cell. Biol*. **14**, 888–895.
15. Pfeiffer,P. and Vielmetter,W. (1988) *Nucleic Acids Res*. **16**, 907–924.
16. Thode,S., Schafer,A., Pfeiffer,P. and Vielmetter,W. (1990) *Cell* **60**, 921–928.
17. Winegar,R.A., Lutze,L.H., Rufer,J.T. and Morgan,W.F. (1992) *Mutagenesis* **7**, 439–445.
18. Boulton,S.J. and Jackson,S.P. (1996) *EMBO J.* **15**, 5093–5103.
19. Boulton,S.J. and Jackson,S.P. (1996) *Nucleic Acids Res*. **24**, 4639–4648.
20. Milne,G.T., Jin,S., Shannon,K.B. and Weaver,D.T. (1996) *Mol. Cell. Biol*. **16**, 4189–4198.
21. Weaver,D.T. (1995) *Trends Genet*. **11**, 388–392.
22. Teo,S.H. and Jackson,S.P. (1997) *EMBO J*. **16**, 4788–4795.
23. Landau,N.R., Schatz,D.G., Rosa,M. and Baltimore,D. (1987) *Mol. Cell. Biol*. **7**, 3237–3243.
24. Concannon,P., Pickering,L.A., Kung,P. and Hood,L. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6598–6602.
25. Alt,F.W. and Baltimore,D. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4118–4122.
26. Roth,D.B., Chang,X.-B. and Wilson,J.H. (1989) *Mol. Cell. Biol*. **9**, 3049–3057.
27. Marrihew,R.V., Marburger,K., Pennington,S.L., Roth,D.B. and Wilson,J.H. (1996) *Mol. Cell. Biol*. **16**, 10–18.
28. Anderson,R.A., Kato,S. and Camerini-Otero,R.D. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 206–210.
29. Nassif,N., Penney,J., Pal,S., Engels,W.R. and Gloor,G.B. (1994) *Mol. Cell. Biol*. **14**, 1613–1625.
30. Moore,J.K. and Haber,J.E. (1996) *Nature* **383**, 644–646.
31. Teng,S.-C., Kim,B. and Gabriel,A. (1996) *Nature* **383**, 641–644.
32. Puchta,H., Dujon,B. and Hohn,B. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 5055–5060.
33. Puchta,H., Dujon,B. and Hohn,B. (1993) *Nucleic Acids Res*. **21**, 5034–5040.
34. Shalev,G. and Levy,A.A. (1997) *Genetics*, **146**, 1143–1151.
35. Chiurazzi,M., Ray,A., Viret,J.F., Perera,R., Wang,X.H., Lioyd,A.M. and Signer,E.R. (1996) *Plant Cell* **8**, 2057–2066.
36. Coen,E.S, Robbins,T.P., Almeida,J., Hudson,A. and Carpenter,R. (1989) in Berg,D.E and Howe,M.M. (eds), *Consequences and Mechanism of Transposition in Antirrhinum majus*. American Society for Microbiology, Washington, DC. pp. 413–436.
37. Baran,G., Echt,C., Bureau,T. and Wessler,S. (1992) *Genetics* **130**, 377–384.
38. Scott,L., LaFoe,D. and Weil,C. (1996) *Genetics* **142**, 237–246.
39. Saedler,H. and Nevers,P. (1985) *EMBO J*. **4**, 585–590.
40. Wessler,S., Tarpley,A., Purugganan,M., Spell,M. and Okagaki,R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8731–8735.
41. Ralston,E.J., English,J.J. and Dooner,H.K. (1988) *Genetics* **119**, 185–197.
42. Gheysen,G., Villarroel,R. and Van Montagu,M. (1991) *Genes Dev*. **5**, 287–297.
43. Walbot,V. and Cullis,C.A. (1985) *Ann. Rev. Plant Physiol*. **36**, 367–396.
44. Aviv,D. and Galun,E. (1985) *Heredity* **76**, 135–136.
45. Vardi,A., Bleichman,S. and Aviv,D. (1990) *Plant Sci*. **69**, 199–206.
46. Dellaporta,S.L., Wood,J. and Hicks,J.B. (1983) *Plant Mol. Biol. Rep*. **1**, 19–21.
47. Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem*. **132**, 6–13.
48. Rubin,E. and Levy,A. (1997) *Mol. Cell. Biol*. in press.
49. Lehman,C.W., Trautman,J.K. and Caroll,D. (1994) *Nucleic Acids Res*. **22**, 434–442.
50. Lehman,C.W. and Carroll,D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10840–10844.
51. Anderson,O.D., Litts,J.C., Gautier,M.F. and Greene,F.C. (1984) *Nucleic Acids Res*. **12**, 8129–8144.
52. Goldberg,S.Z., Kuebbing,D., Traubert,D., Schafer,M.P., Lewis,S.E., Popp,R.A. and Anderson,F.W. (1986) *J. Biol. Chem*. **261**, 12368–12374.
53. Doseff,A., Martienssen,R. and Sundaresan,V. (1991) *Nucleic Acids Res*. **19**, 579–584.
54. Nicholls,R.D., Fischel-Ghodsian,N. and Higgs,D.R. (1987) *Cell* **49**, 369–378.
55. Pulak,R.A. and Anderson,P. (1988) *Mol. Cell. Biol*. **8**, 3748–3754.
56. Spitz,R.A. and Orkin,S.H. (1982) *Nucleic Acids Res*. **10**, 8025–8029.
57. Formosa,T. and Alberts,B.M. (1986) *Cell* **47**, 793–806.