Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells

Siegfried Salomon and Holger Puchta

Institut für Pflanzenbiologie und Kulturpflanzenforschung (IPK), Correnstraße 3, D-06466 Gatersleben, Germany

1Corresponding author
e-mail: puchta@ipk-gatersleben.de

To analyze genomic changes resulting from double-strand break (DSB) repair, transgenic tobacco plants were obtained that carried in their genome a restriction site of the rare cutting endonuclease I-SceI within a negative selectable marker gene. After induction of DSB repair via Agrobacterium-mediated transient expression of I-SceI, plant cells were selected that carried a loss-of-function phenotype of the marker. Surprisingly, in addition to deletions, in a number of cases repair was associated with the insertion of unique and repetitive genomic sequences into the break. Thus, DSB repair offers a mechanism for spreading different kinds of sequences into new chromosomal positions. This may have evolutionary consequences particularly for plants, as genomic alterations occurring in meristem cells can be transferred to the next generation. Moreover, transfer DNA (T-DNA), carrying the open reading frame of I-SceI, was found in several cases to be integrated into the transgenic I-SceI site. This indicates that DSB repair also represents a pathway for the integration of T-DNA into the plant genome.

Keywords: end-joining/genome/illegitimate recombination/integration/transformation

Introduction

Double-strand breaks (DSBs) are critical lesions in genomes. A single genomic DSB is able to block cell division in mammalian cells (Huang et al., 1996). Efficient repair of DSBs is therefore important for the survival of all organisms. In principle, DSBs can be repaired via illegitimate or homologous recombination. In higher eukaryotes illegitimate recombination seems to be the main mode of DSB repair (Sargent et al., 1997). Various studies on illegitimate recombination have been performed in mammalian cells (for review see Roth and Wilson, 1988), Xenopus oocytes (e.g. Lehman et al., 1994; Pfeiffer et al., 1994) and yeast cells (e.g. Mezard and Nicolas, 1994). Studies on recombination in somatic plant cells are of particular biological significance since gametes may arise from meristem cells with mutated genomes (Walbot, 1996). Illegitimate recombination in plants has been investigated mainly via integration of transgenes, especially transfer DNAs (T-DNAs) (Matsumoto et al., 1990; Gheyesen et al., 1991; Mayerhofer et al., 1991; Hiei et al., 1994; Ohba et al., 1995; Papp et al., 1996; Iglesias et al., 1997; Takano et al., 1997). Moreover, extrachromosomal illegitimate recombination between plasmids (De Groot et al., 1994; Gorbunova and Levy, 1997), circularization of extrachromosomal T-DNAs (Bakkeren et al., 1989) and deletions at specific positions in the genome (Wessler et al., 1990; Shirley et al., 1992) were characterized at the molecular level. The main conclusion of these studies was that in many cases recombination occurred at short patches of sequences, homologous between both recombination partners—a phenomenon well known in other organisms (Roth and Wilson, 1988). Frequently short ‘filler’ sequences of unknown origin were found to be included in the new junctions. For plants, it was reported that DNA inserted between the ends of an extrachromosomal plasmid consisted mostly of scrambled sequences derived from the same plasmid and in one case from the tobacco genome. However, DSB repair was always accompanied by deletions whether or not filler DNA was incorporated in the newly formed junctions (Gorbunova and Levy, 1997).

Using highly specific restriction endonucleases for induction of breaks at specific loci in eukaryotic genomes (for review see Haber, 1995; Jasin, 1996) it has been possible to characterize DSB-induced homologous recombination in somatic plant cells (Puchta et al., 1993, 1996; Chinarruzzi et al., 1996; Puchta, 1998a). However, the consequences of DSB-induced illegitimate recombination have not yet been analyzed. Therefore, we constructed transgenic tobacco lines which contain within a negative selectable marker (the codA gene, see Stougaard, 1993), an I-SceI site cleavable after Agrobacterium-mediated transient expression of the I-SceI open reading frame (ORF). When DSB repair resulted in genomic alterations associated with the loss of gene function, the cells became selectable by their resistance to 5-fluorocytosine (5-FC). From sequence analysis of these junctions we were able to identify the molecular structure of different genomic changes induced by DSB repair in plants.

Results

Experimental setup

To characterize illegitimate DSB repair, tobacco was transformed via Agrobacterium with the negative selectable marker gene cytosine deaminase (codA; Stougaard, 1993), carrying an I-SceI site between the promoter and coding region (Figure 1). The rationale for using this marker was that cutting the genomic I-SceI site would induce repair processes. Cells in which repair does not result in restoration of the functional gene can be selected by including 5-FC in the cultivation medium. Cells expressing a functional cytosine deaminase convert 5-FC into the highly toxic compound 5-fluorouracil, killing the cells, whereas cells lacking the enzyme grow and can be propagated (Stougaard, 1993). To exclude large deletions or conversions from the homolog resulting in
loss of the entire transgene locus, kanamycin selection was applied simultaneously. The kanamycin-resistance gene resides on the same T-DNA but outside the codA gene (Figure 1). After DNA extraction from resistant calli, recombination junctions were amplified by PCR. The respective PCR products were then cloned and sequenced to elucidate the molecular nature of the event.

**Induction of DSB repair, selection of calli and PCR analysis**

We chose four different tobacco lines each containing a single copy of T-DNA of either the binary vector pBNE3I (line B9) or pCNE3I (lines C15, C19, C25) in their genome for this study. Both binaries carry on their T-DNA, in an identical arrangement, both a codA and a kanamycin-resistance gene (Figure 1). Seedlings of the selfed transgenic mother plants were inoculated via vacuum infiltration with an Agrobacterium strain carrying the binary pCISceI. This binary carries a plant expression cassette for I-SceI on its 2.4 kb long T-DNA (Puchta et al., 1996) but no selectable marker gene. Thus, no selection for the integration of this T-DNA can be applied and the T-DNA may integrate or be lost in the transformed cells after DSB induction. After 3 days of cocultivation the seedlings were transferred to Murashige and Skoog (MS) plates containing 5-FC and kanamycin in the medium. Resistant calli could be detected after several weeks of cultivation. The results of all transformation experiments are summarized in Table I. DNA was extracted from the calli and recombination junctions were amplified by PCR for sequence analysis. PCR fragments could not be detected from all double resistant calli. Since the oligonucleotides used as PCR primers bind within a 2.3 kb region of the T-DNA (Figure 1), deletions involving at least one primer binding site, insertions too large to be amplified by PCR or other major genomic rearrangements might account for this fact. Surprisingly, the I-SceI site was destroyed in only about half of the PCR products. This indicates, that due to the relatively low concentration of 5-FC used for selection, some calli survived in spite of an intact codA gene. Similar results were reported previously when the codA gene was used as negative selectable marker in plants (Schlaman and Hooykaas, 1997). PCR products without I-SceI site were cloned and sequenced for further analysis.

**Classification of recombination junctions**

In total 28 recombination junctions were sequenced. Surprisingly, only 18 of these were reduced in size. In the other 10 cases the size of the locus was increased after DSB repair. The class of junctions with reduced size could be subdivided into those with simple deletions (Table II) and those having deletions associated with insertion of small filler sequences (Table III). The class of junctions

---

**Table I.** Compilation of 5-FC resistant tobacco cells obtained after induction of DSBs by transient expression of I-SceI

<table>
<thead>
<tr>
<th>Transgenic line</th>
<th>Number of seedlings inoculated</th>
<th>Resistant calli</th>
<th>Tested</th>
<th>PCR positive</th>
<th>I-SceI destroyed</th>
<th>Sequenced</th>
</tr>
</thead>
<tbody>
<tr>
<td>B9</td>
<td>644</td>
<td>90</td>
<td>64</td>
<td>36</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>C15</td>
<td>1254</td>
<td>99</td>
<td>63</td>
<td>23</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>C19</td>
<td>229</td>
<td>12</td>
<td>12</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>C25</td>
<td>349</td>
<td>15</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

*a* For a number of 5-FC resistant calli no PCR fragment could be amplified under the conditions used. The most likely explanation is that in these cases the DSB caused a deletion that included at least part of the primer binding sites. Alternatively, insertions too large to be amplified by PCR or other major rearrangements might be responsible for the lack of a PCR fragment.

*b* Due to the relatively low concentration of 5-FC used for selection ‘escapes’ occurred (see Results).

**Table II.** Compilation of DSB repair events associated with simple deletions at the I-SceI site

<table>
<thead>
<tr>
<th>Line</th>
<th>Deletion (bp)</th>
<th>Length of deletion at 35S promoter codA gene</th>
<th>Homology (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C15-13<em>a</em></td>
<td>2</td>
<td>2</td>
<td>--</td>
</tr>
<tr>
<td>C15-531</td>
<td>3</td>
<td>3</td>
<td>--</td>
</tr>
<tr>
<td>B9-39</td>
<td>5</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>C15-23</td>
<td>8</td>
<td>8</td>
<td>--</td>
</tr>
<tr>
<td>C15-671</td>
<td>15</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>B9-48</td>
<td>236</td>
<td>236</td>
<td>--</td>
</tr>
<tr>
<td>C15-67</td>
<td>305</td>
<td>260</td>
<td>45</td>
</tr>
<tr>
<td>C15-44</td>
<td>1055</td>
<td>1047</td>
<td>8</td>
</tr>
<tr>
<td>B9-3</td>
<td>1060</td>
<td>1057</td>
<td>3</td>
</tr>
<tr>
<td>C19-5</td>
<td>1218</td>
<td>390</td>
<td>828</td>
</tr>
</tbody>
</table>

*a* In this case the bottom strand was taken as reference for the nick.

The top strand (see Figure 2) was taken as reference for the nick. The nucleotides of the 3' end of this strand are found next to the 35S promoter in the junction.

---

**Fig. 1.** Schematic map of the T-DNA from the binary plasmid pBNE3I inserted in the tobacco genome to study DSB repair. An I-SceI site was integrated between codA ORF and 35S promoter. The T-DNA used for the transformation also contained, in addition to the codA gene, a kanamycin gene as transformation marker. A4, A10, A1 and S1, S10 and S11 are primers used for the PCR amplification of the recombined junctions. The arrangement of genes on the T-DNA of the binary vector pCNE3I is identical to pBNE3I. RB, right border; LB, left border.
Endonuclease I-SceI cuts within an 18 bp sequence producing 3’-overhangs of four nucleotides (Figure 2). For general compilation of all recombination events the top strand with the nick next to the 35S promoter was taken as reference for the position of the cut in Tables II–V. The 3’-end of the top strand was conserved within most newly formed junctions. There were only three clear exceptions (noted in Tables II–IV) in which the 3’-end of the bottom strand was conserved upon repair.

**Junctions reduced in size**

Simple deletions. In 10 cases, part of the transgene sequence was lost without further change in the locus (Table II). We detected deletions between 2 and 1218 bp. The small deletions of 2–15 bp do not seem to be responsible for the FC-resistant phenotype of the respective cells by themselves. These calli probably escaped the selection together with a number of other calli that maintain the functional I-SceI site (see Table I). In four cases only the end adjacent to the 35S promoter was deleted, whereas in the other six, both ends of the break were involved in the deletion. While in four cases no homology was found between the rejoined sequences (e.g. Figure 2, junctions C15-531 and B9-48), in one case one identical base pair and in five other cases small, partly interrupted homologies from 2 to 7 bp could be detected (e.g. Figure 2, junctions C15-23 and C19-5).

Deletions associated with the insertion of small filler sequences. In eight cases the junctions reduced in size contained extra sequences between the rejoined ends (Table III). With one exception (C15-29) all deletions

<table>
<thead>
<tr>
<th>Line</th>
<th>Deletion (bp)</th>
<th>Length of deletion at</th>
<th>Homology</th>
<th>Length of filler (bp)</th>
<th>Origin of filler sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35S promoter (bp)</td>
<td>codA gene</td>
<td>35S (bp)</td>
<td>codA</td>
<td></td>
</tr>
<tr>
<td>B9-62</td>
<td>88</td>
<td>2</td>
<td>86</td>
<td>nt</td>
<td>nt</td>
</tr>
<tr>
<td>C15-131</td>
<td>117</td>
<td>76</td>
<td>41</td>
<td>nt</td>
<td>nt</td>
</tr>
<tr>
<td>C15-51</td>
<td>262</td>
<td>227</td>
<td>35</td>
<td>nt</td>
<td>nt</td>
</tr>
<tr>
<td>B9-50</td>
<td>395</td>
<td>290</td>
<td>105</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>B9-46</td>
<td>458</td>
<td>439</td>
<td>19</td>
<td>nt</td>
<td>nt</td>
</tr>
<tr>
<td>C15-29</td>
<td>1152</td>
<td></td>
<td>1152</td>
<td>nt</td>
<td>nt</td>
</tr>
<tr>
<td>B9-88</td>
<td>1287</td>
<td>1281</td>
<td>6</td>
<td>nt</td>
<td>nt</td>
</tr>
<tr>
<td>C15-22</td>
<td>1322</td>
<td>1294</td>
<td>30</td>
<td>nt</td>
<td>nt</td>
</tr>
</tbody>
</table>

The top strand (see Figure 2) was taken as reference for the nick.

4 In this case the bottom strand was taken as reference for the nick as nucleotides of the 3’ end of this strand are found next to the 35S promoter in the junction.

b As determined by PCR analysis.

nt: not testable.

Deletions associated with the insertion of T-DNA insertions into the I-SceI site

<table>
<thead>
<tr>
<th>Line</th>
<th>Genomic deletion (bp)</th>
<th>Length of genomic deletion at</th>
<th>Homology</th>
<th>T-DNA length (bp)</th>
<th>Length of T-DNA deletion (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35S promoter (bp)</td>
<td>codA gene</td>
<td>35S</td>
<td>codA</td>
<td></td>
</tr>
<tr>
<td>C25-1</td>
<td>36</td>
<td>5</td>
<td>31</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>C15-47</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>C15-52</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>C15-53</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

The top strand (see Figure 2) was taken as reference for the nick. LB, left border; RB, right border.

with increased size could be subdivided into genomic insertions (Table IV) and T-DNA insertions (Table V).

Table IV. Compilation of DSB repair events associated with the insertion of genomic sequences into the I-SceI site

<table>
<thead>
<tr>
<th>Line</th>
<th>Deletion (bp)</th>
<th>Length of deletion at</th>
<th>Homology</th>
<th>Length of filler (bp)</th>
<th>Origin of filler sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35S promoter (bp)</td>
<td>codA gene</td>
<td>35S (bp)</td>
<td>codA</td>
<td></td>
</tr>
<tr>
<td>B9-86</td>
<td>1</td>
<td>–</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>C15-36</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>nt</td>
<td>7</td>
</tr>
<tr>
<td>B9-44</td>
<td>2</td>
<td>–</td>
<td>2</td>
<td>nt</td>
<td>nt</td>
</tr>
<tr>
<td>C19-1</td>
<td>11</td>
<td>6</td>
<td>5</td>
<td>nt</td>
<td>nt</td>
</tr>
<tr>
<td>B9-82</td>
<td>1</td>
<td>1</td>
<td>–</td>
<td>nt</td>
<td>nt</td>
</tr>
<tr>
<td>B9-1</td>
<td>4</td>
<td>4</td>
<td>–</td>
<td>nt</td>
<td>nt</td>
</tr>
</tbody>
</table>

The top strand (see Figure 2) was taken as reference for the nick.

4 In this case the bottom strand was taken as reference for the nick as nucleotides of the 3’ end of this strand are found next to the 35S promoter in the junction.

nt, not testable.

Table V. Compilation of DSB repair events associated with the insertion of T-DNA into the genomic I-SceI site

<table>
<thead>
<tr>
<th>Line</th>
<th>Genomic deletion (bp)</th>
<th>Length of genomic deletion at</th>
<th>Homology</th>
<th>T-DNA length (bp)</th>
<th>Length of T-DNA deletion (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35S promoter (bp)</td>
<td>codA gene</td>
<td>35S</td>
<td>codA</td>
<td></td>
</tr>
<tr>
<td>C25-1</td>
<td>36</td>
<td>5</td>
<td>31</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>C15-47</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>C15-52</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>C15-53</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>
with wild-type tobacco DNA to test whether inserts were
inserted uninterruptedly into the DSB. The 90 bp insert
B9-86 was practically identical (85/90) to the promoter
region of the unique T92 gene coding for an auxin-binding
protein (DDBJ/EMBL/GenBank accession No. X70903).
Whereas the break-insert junction close to the

codA gene

contained three homologous base pairs, no homology
could be detected at the other end (Figure 3). In case of
the 795 bp insert B9-1 significant homology (311/411) to
the tobacco retrotransposon Tnt1 (Grandbastien
et al., 1989) was detected. Analysis of the inserted sequence
revealed that the ORF coding for the reverse transcriptase
domain of Tnt1 is interrupted by stop codons. Therefore,
the inserted sequence is not itself derived from an active
retrotransposon, although we cannot exclude that the
inserted sequence originates from a cDNA produced by
trans-complementation of an active Tnt1 element. The
sequence is, as expected, repetitive in tobacco (see Figure 4
for Southern blot). Computer analysis indicated that the
200 bp insert C15-36 contains two repeats of a sequence
fully matching the consensus sequence of a tobacco
chromosomal subterminal repeat (Suzuki et al., 1994).
Interestingly, an interrupted 7 bp homology could be
detected between the consensus sequence and the break
site close to the codA gene (Figure 3). The Southern blot
of tobacco DNA restricted with HindIII (Figure 4) revealed
a pattern, in which, in contrast to other repetitive DNAs,
Fig. 3. Junctions of a filler sequence (B9-50; see also Table III) and two genomic insertions (B9-86 and C15-36; see also Table IV). Only top strands are shown. The inserted sequences are marked by dotted boxes. In most cases short patches of homology (marked by black boxes) are found between both recombination partners at the newly formed junctions. With C15-36 only the junction homologous to the core sequences of the tobacco subterminal repeat is shown. Note that in the junction C15-36 the bottom strand was involved in the recombination reaction, as the four nucleotides of the staggered cut site of I-SceI (see Figure 2) are found close to the 35S promoter in the junction.

Fig. 4. Southern blots revealing that repetitive and unique genomic sequences are inserted in DSBs. HindIII-digested tobacco DNA was used for hybridization with the respective inserts as probes. As a control, HindIII-digested mouse DNA was used, which never produced a signal (data not shown). B9-44 (unknown sequence of 330 bp), B9-82 (582 bp partly homologous to 5’/H11032 sequences of a polygalacturonase gene), C15-36 (200 bp homologous to chromosomal subterminal repeat) C19-1 (unknown sequence of 359 bp) and B9-1 (795 bp homologous to Tnt1 retrotransposon). The strongest signals are shifted to the top of the gel, indicating an under-representation of HindIII sites within the chromosomal regions which contain subterminal repeats. A limited region of homology (49/65) was also found between the 582 bp insert B9-82 and 5’ sequences of a polygalacturonase gene (Tebbutt et al., 1994). The insert B9-82 lacks recognition sites for HindIII or EcoRI, yet Southern blots (Figure 4) of tobacco DNA restricted with HindIII or EcoRI (data not shown) and probed with the inserted sequence, revealed two bands without any background. This indicates that two different genomic loci within the allo-tetraploid tobacco genome are homologous to the insert B9-82. No significant homologies to sequences in databases could be detected for the remaining two inserts. From the results of Southern blotting of tobacco DNA we were, however, able to demonstrate that the 359 bp insert C19-1 is a repetitive sequence and has a similar pattern of distribution as the Tnt1 specific insert B9-1 (Figure 4). With the 330 bp insert B9-44, a single strong band was detected in the Southern blot. However, the significant background indicates that there might be sequences related to B9-44 which occur in high copy number in the tobacco genome. Using insert-specific primers and wild-type tobacco DNA we obtained a PCR fragment of the expected size for all inserts, indicating that the inserts were copied in a correct and uninterrupted fashion into the break (data not shown). Our results thus unambiguously demonstrate that different classes of unique and repetitive genomic sequences can be inserted into a DSB in plant cells. Notably, all inserts contained stop codons in all possible reading frames.

T-DNA insertions. In four cases T-DNA molecules used for the transient expression of the I-SceI gene were found to be inserted into the break (Table V), again accompanied by small deletions (6–36 bp) at one or both ends of the break. As often found in transgenic plants the T-DNAs were not integrated completely. Although at the left border from the nicking site 23, 30, 83 or 95 nucleotides and at the right border 3, 11, 13 or 620 nucleotides were missing in the individual cases, most of the genetic information of the T-DNA (1.7 and 3’/H11003 2.3 kb) was integrated into the break. We found in five of eight cases small partly interrupted homologies of 2–9 bp between the T-DNA and the flanking ends of the DSB of transgene locus (Figure 5).

Plants were regenerated from several calli representing all different classes of recombinants. Segregation and PCR analysis of the seedlings obtained demonstrated that the
induced alterations were stably transferred to the next
generation (data not shown).

Discussion

DSB repair in somatic plant cells

Different transgenic tobacco lines which harbored a single
artificial I-SceI site between promoter and ORF of a
cytosine deaminase gene (Stougaard, 1993) were used to
study consequences of illegitimate DSB repair in plants.
Major alterations were detected via loss of function of the
cytosine deaminase, selected with 5-FC in the medium.
However, neither DSB repair by correct ligation nor cells
in which a non-repaired DSB results in cell death can be
analyzed with the assay system. Simultaneous selection
for kanamycin resistance was applied to exclude gene
conversion events from the homolog in hemizygous
plantlets. In contrast to a recent report (Risseeuw
et al., 1997) we found no indication of instability of the
codA gene in our transgene lines. In control experiments in
which the I-SceI gene was not expressed a small number
resistant calli was obtained. These calli had escaped the
selection, as we did not find changes in the transgene
(data not shown).

Sequence analysis of the modified junctions allowed us
to elucidate the processes that occurred during somatic
DSB repair. We could define two main classes, one in
which the size of the locus was reduced due to DSB repair
and the other in which the size was enlarged. Reduction
in size was due to simple deletions or deletions associated
with the insertion of short filler sequences; an increase in
size was due to insertions of genomic sequences or
T-DNAs harboring the I-SceI gene used for transient
expression of the enzyme.

Two different pathways are involved in the
formation of simple deletions

In 10 events the transgene locus was reduced in size
without concomitant insertion of new sequences (Table II).
Thus a simple deletion of sequences occurred before or
during the joining of the two broken ends. We found two
subclasses of such junctions. In one (five events) none or
only 1 bp of homology was involved in the rejoining,
whereas in the other (five events) 2–7 homologous base
pairs present in both recombination partners could be
detected at the junction borders. This is strongly reminis-
cent of data found in other eukaryotic systems (Lehman
et al., 1994; Mason et al., 1996) and for T-DNAs in plants
(Bakkeren et al., 1989; Gheysen et al., 1991; Mayerhofer
et al., 1991), and confirms data recently reported for DSB
repair of plasmid molecules in plants (Gorbunova and
Levy, 1997).

For cells of higher eukaryotes, junctions without homo-
logies were explained by simple ligation, whereas small
patches of homologous nucleotides were considered to be
a prerequisite for the operation of a single-strand annealing
(SSA) mechanism (Lehman et al., 1994; Nicolas et al.,
1995; Mason et al., 1996; Lieber et al., 1997). In the
single-strand annealing reaction two free single-stranded
3’ ends interact with each other via a few complementary
bases (the homologous nucleotides in the newly formed
junction). This interaction is followed by removal of non-
complementary nucleotides at the 3’ ends so that the
consecutive repair synthesis results after ligation in a
double strand. Remarkably, in three out of four one-sided
deletion events no homologous base pairs were detected
between recombination partners at the junctions, whereas
in four out of six two-sided deletion events, small homo-
logies (>2 bp) were present. This may indicate that one-
sided deletions result primarily from a direct ligation
pathway, whereas two-sided deletions are mainly due to a
single-strand annealing-like pathway.

Deletions with small filler sequences

In eight events deletion was accompanied by the insertion
of a short filler sequence (Table III). In five cases filler
sequences were too small (4–19 bp) to draw valid conclu-
sions about their origins. In one case the filler originated
from the deleted region of the 35S promoter and in two
others from the tobacco genome. With one exception in
which the deletion comprised only the codA gene end,
these deletions involved both ends of the break. We
assume in accordance with recent results obtained with
plasmid DNA (Gorbunova and Levy, 1997), that filler
sequences at the genomic break sites might arise via a
modified synthesis-dependent strand annealing (SDSA)
mechanism (Figure 6) originally proposed for homologous DSB repair in Drosophila (Nassif et al., 1994) and recently also for homologous DSB repair in somatic plant cells (Rubin and Levy, 1997; Puchta, 1998a). The mechanism postulates that both 3'-ends of the break act independently. After pairing with a few complementary nucleotides of a donor sequence the ends are able to copy donor sequences via primer elongation into the break (Figure 6). Prerequisites for annealing and priming reactions according to the SDSA model are thus, as for the SSA model, short patches of sequence homologies between the recombination partners. With the junction containing the internal fragment of the 35S promoter, small homologies are indeed present at both ends of the filler (Figure 3). The alternative explanation that free oligonucleotides were captured by the ends of the break seems less probable, especially since such a capture could not be mimicked artificially in mammalian cells (Roth et al., 1991). Moreover, as cells avoid free DNA ends, which would block cell-cycle progression, it seems unreasonable to assume that numerous 'free' nuclear DNA fragments persist for long in the nucleus.

Insertions of genomic sequences into DSBs: a mechanism of genome evolution

For yeast strains deficient in homologous recombination, the integration of retroelement sequences into a DSB induced by the HO endonuclease has been demonstrated recently (Moore and Haber, 1996; Teng et al., 1996). It was suggested that such insertion events might also play a role in genome evolution of higher eukaryotes (Moore and Haber, 1996). We can now demonstrate that DSB repair can indeed be associated with DNA insertions in higher eukaryotes. A broad spectrum of unique and abundant genomic sequences was found inserted into chromosomal breaks of plants. None or only a few (1–11) base pairs were concomitantly lost at the break site. All sequences were integrated in an uninterrupted way in contrast to the insertion of scrambled plasmid sequences associated with large deletions, which were detected recently in an extrachromosomal assay (Gorbunova and Levy, 1997). This is not surprising, as extrachromosomal ‘naked’ plasmid DNA differs in its recombination behavior from plant DNA organized in nuclear chromatin (Puchta and Meyer, 1994). In contrast to what has been proposed for yeast (Teng et al., 1996) we do not favor the involvement of mRNA or cDNA intermediates in the reaction, as none of the inserted sequences contained an uninterrupted ORF over its entire length. The simplest explanation of our data would be an uninterrupted copying mechanism of ectopic chromosomal DNA. A stretch of several nucleotides is homologous between the 3'-end of the I-SceI site and the subterminal repeat (Figure 3, C15-36) as well as one end of the T92 promoter (Figure 3, B9-86). Thus, an annealing and priming reaction seems to be involved in the process and genomic insertions might occur, as already suggested for filler sequences (Gorbunova and Levy, 1997), via an SDSA mechanism (Figure 6). We assume that (under our experimental conditions) in most cases only one 3' end is actively invading a genomic donor sequence (Figure 6C). If the two 3' ends would simultaneously invade different donor sequences, a combination of different genomic sequences would result. We did not detect this type of recombination event. The linkage of the invading DNA strand with the other end of the break can be mediated by a micro-annealing step (Figure 6D) or by a simple ligation to the other end without involvement of homology (Figure 6D'). The latter was most probably the case for the insert originating from the T92 gene.

The assay system applied in this study has certain experimental restrictions as it uses staggered breaks that only partly resemble naturally occurring breaks. Moreover, it does not address the question of the frequency with which DSB repair occurs via conversion from the chromosomal homolog. Nevertheless, our data clearly demonstrate that various kinds of sequences can spread into new positions within the plant genome as a result of DSB repair. As a large fraction of the tobacco genome consists of repeated sequences, it is not surprising to find them inserted into breaks. Current models explain amplification of repetitive genomic sequences by unequal homologous recombination between repetitive sequences in tandem (Flavell, 1985) or by replicative transposition of retroelements (e.g. SanMiguel et al., 1996). However, these models do not explain how transposition-inactive sequences can be transferred into new positions within the plant genome. Through DSB repair nearly any kind of genomic sequence which is available for a copying process might be inserted into new genomic positions. This phenomenon might be
of general importance for higher plants and their genome evolution. [In plants the ‘germline’ is only set apart late in development and alterations occurring in meristem cells can be transferred to the next generation (Walbot, 1996).] Many plants contain large genomes with large amounts of repetitive DNA. The high variability in genome size between closely related species indicates considerable plasticity. It is tempting to speculate that part of the genome size differences might be due to DSB repair. The spread of retrotransposons was postulated as a main force in enlarging plant genomes (Benetzen and Kellogg, 1997). This genomic enlargement, however, seems to be counterbalanced, at least in part, by deletions (Petrov, 1997). DSB repair is a prominent source of deletions. Insertions in turn may partly counterbalance these deletions, as suggested by our results on frequent insertions upon DSB repair. Therefore it will be of significant interest to test the ratio of insertions versus deletions not only for tobacco with a large genome but also for plants with relatively small genomes like that of Arabidopsis.

**T-DNA insertions**

Remarkably, T-DNA was found to be inserted into the DSB in four recombination events (Table IV). In three cases almost the complete sequence of the 2.4 kb long T-DNA was integrated, in one case a fragment of 1.7 kb. In our experimental setup the cauliflower mosaic virus (CaMV) 35S promoter is driving the transient expression of the I-SceI gene. The same promoter is also present adjacent to the I-SceI site in the target locus, representing a sequence homology of 0.5 kb between the incoming T-DNA and the transgene. Using a similar setup we recently were able to show that in about one out of 125–300 transformation events such a construct recombines with the homologous sequence after DSB induction in the target locus (Puchta, 1998a). However, the four insertions analyzed here were not integrated via homologous recombination. Thus even in the vicinity of a homologous sequence T-DNA integration into a genomic break may occur more frequently via illegitimate than homologous recombination. The fact that T-DNA was inserted into the same break in the plant genome in several cases implies that (i) breaks might play a significant role in T-DNA integration and (ii) the number of breaks in a cell might be rate limiting for integration, as genomic breaks seem to be kept to a minimum. Further support for our interpretation can be found in the literature. A recent study on hamster cells reported the integration of plasmid-specific sequences carrying an I-SceI ORF into an I-SceI site in two out of 11 analyzed recombination events, although these experiments could not distinguish between one-sided homologous or illegitimate integration (Pipiras et al., 1998). It was also shown that X-rays, which induce DSBs in the genome can enhance transgene integration in plant cells (Köhler et al., 1989). Using cotransformation it could be demonstrated that different T-DNAs integrate into a single genomic locus (De Block and Debrouwer, 1991; De Neve et al., 1997). This indicates that a surplus of extrachromosomal DNA ends is recognized by the cell as breaks to be ligated, yielding a tandem arrangement of transgenes, which is then linked to free genomic ends.

Various studies on T-DNA integration indicate that there are in principle two different classes of integration patterns (e.g. Matsumoto et al., 1990; Gheysen et al., 1991; Mayerhofer et al., 1991). In about two-thirds to three-quarters of the cases the right border (which is conserved in the recombinant) and the target locus contain no or only one base pair microhomology, whereas longer homologies exist between the left border (combined with truncations) and the pre-insertion site. In these events the two ends behave differently, and it has been postulated that integration occurs via single-strand annealing followed by the right border being linked to genomic sequences directly or indirectly by the VirD2 protein which is covalently bound to the 5’ end of the T-strand (Tinland, 1996). In a second class of events not only the left but also the right border is partly truncated in the recombinants, and homologous nucleotides can be found in most cases also between the right border and the pre-insertion site (e.g. Matsumoto et al., 1990; Mayerhofer et al., 1991). This pattern of integration was found in our study. We therefore suggest that integration events of this class are generally mediated by DSB repair. T-DNA molecules are transferred as single strands to the nucleus, where some of them become double-stranded before integration (Tinland et al., 1994). By interacting with break ends, a single- or double-stranded T-DNA could be copied into the genome via SDSA (Figure 6). Alternatively, a double-stranded T-DNA could also be integrated at the break by two consecutive SSA reactions or by a single-strand annealing reaction and a subsequent simple ligation step.

Our results might also be interesting for practical application in plant biotechnology: as gene targeting is still not a feasible technique for plants (Puchta, 1998b) the establishment of site-specific integration systems is of considerable interest. Sophisticated techniques based either on the Cre–lax system (Albert et al., 1995; Vergunst and Hooykaas, 1998; Vergunst et al., 1998) or on DSB-induced homologous recombination (Puchta et al., 1996) have been established. Inducing DSBs at desired positions in the plant genome may suffice to obtain site-specific integration of transgenes at these sites in numbers which allow easy identification of the appropriate recombinants via PCR using a combination of locus and insert-specific oligonucleotides.

**Materials and methods**

**DNA constructs**

All plasmids were propagated in Escherichia coli strain DH 5α. The plasmid pNE3 (Stougaard, 1993) carrying a codA ORF under the control of plant expression signals was digested with XhoI and religated with a SalI fragment derived from two complementary oligo nucleotides with the recognition site of the restriction endonuclease I-SceI 5’-pCAGGTCACGGTATCCCTTATACCGTCTAGGATCCGTCGACCCG-3’, resulting in the plasmid pNE3I. The binary vector pBin19 (Bevan, 1987) was digested with EcoRI and HindIII, and ligated with the codA-containing EcoRI/HindIII fragment of pNE3I resulting in pBNE3I (see Figure 1). The EcoRI-linearized plasmid pNE3I was cloned into the binary vector pCB200 (Rothstein et al., 1987), resulting in the binary vector pCNE3I. Both binary vectors carry, besides the functional codA gene with a I-SceI site between promoter and ORF, a kanamycin gene as transformation marker in an identical arrangement as depicted for pBNE3I in Figure 1. The binary vectors were transferred into Agrobacterium tumefaciens via electroporation (voltage 2500 V, capacity 25 μF, shunt 201 Ω and 5 ms pulse; EasyJet Plus system, Equibio, UK).

The I-SceI expression vector pCISceI (Puchta et al., 1996) contains a synthetic I-SceI ORF under the control of the cauliflower 35S promoter (Puchta et al., 1993) between T-DNA borders.
Plant transformation and transient expression

In a first series of experiments Nicotiana tabacum L. cv. Petitie Havana line SR1 plants were transformed with the Agrobacterium strains harboring either the binary vector pBIN31 or pCNE31. Vacuum infiltration of tobacco seedlings with Agrobacterium was performed as described (Puchta, 1998c). The seedlings were placed for 3 days on MS plates [0.3 % Gelrite (Merck, USA) without hormones] and transferred afterwards onto MS plates containing hormones [0.5 μg α-naphthalene acetic acid (NAA) and 5 μg 6-benzyl amino purine (BAP) per ml], 100 μg kanamycin sulfate, 200 μg vancomycine (Duchefa, Netherlands) and 500 μg cefotaxime (Duchefa, Netherlands) per ml. For callus induction the infiltrated seedlings were transferred to a growth chamber (24°C; 16 h light/day). Every week the transformed plantlets were transferred onto fresh media. When the calli formed shoots they were transferred to boxes (Magenta, Chicago, USA) containing MS medium without hormones for plant regeneration. After root development the plants were placed in a greenhouse and allowed to grow to maturity. Segregation of the selfed transformatans was tested by germination of the seeds on MS medium supplemented with 300 μg kanamycin sulfate and 500 μg 5-FC per ml.

In a second series of experiments the transgenic lines B9, C15, C19 and C25 were inoculated with an Agrobacterium strain harboring the binary pPC3Sel as described above. Double selection using 100 μg 5-FC (Sigma, USA) and 100 μg kanamycin sulfate per ml medium was applied. After 8 weeks the 5-FC concentration was increased to 200 μg/ml medium. In most cases surviving calli were cut into two pieces after an additional 6 weeks. One piece was used for shoot regeneration and the other was used for DNA isolation.

Plant DNA extraction

DNA extraction from leaf tissues and calli was carried out as described in Fulton et al. (1995) using 1–2 g of plant tissue.

Southern analysis

Southern blotting using the hybridization membrane ‘Hybond N’ (Amersham, UK) was performed as described (Swoboda et al., 1994). The respective DNA fragments were labeled using a random priming labeling kit (MegaFlair DNA labeling system RPN1607, Amersham, UK) and [α-32P]dATP (Amersham, UK). For identification of plants with single copy integration of pCNE31 the codA coding region was used as hybridization probe. The plant DNA was digested with HindIII and HindIII-Scel. To determine the abundance of the inserted genomic sequences in the tobacco genome, specific probes were generated by PCR (see below). Wild-type tobacco and Nicotiana DNA (as negative control) was digested with HindIII or EcoRII (probe B9-44). Hybridizations were performed at 65°C.

PCR and sequence analysis

Genomic DNA was analyzed by PCR using the primers S1 5’-pGGGCTCT-AGAGCTTAACAGCACAGTTGC-3’ and A1 5’-pTGACTCTAGAGCAAGATATATTTCAACA-3’ and A4-5’-pGCCTCTAGAGCCCTCTGTGCTA-3’ in a first amplification and the nested primer pair S1 5’-pGGGCTCTAGAGCAAGATATATTTCAACA-3’ and A4-5’-pGCCTCTAGAGCCCTCTGTGCTA-3’ in a second amplification. Two microliters of the reaction volume from the first amplification served as a template for the second amplification. In the first amplification the PCR reactions were performed by using SAWADY Long PCR Kit (pseqlab, Erlangen, Germany), in the second amplification bioTaq DNA polymerase (biomaster, Köln, Germany) was used. The PCR reactions were performed according to the polymerase manufacturer’s recommendation in a Gene Amp PCR System 2400 (Perkin Elmer, USA). The following program was used for the first amplification: 5 min denaturation at 94°C, 40 cycles of 20 s at 93°C, 35 s at 55°C and 2 min at 68°C; and a final step of 7 min at 68°C. In case no PCR amplification could be detected, primers S10 5’-pGGGGAACCTGGTTCACTGGTAC-3’ and A10 5’-pGGGCTCTAGACATTTTCTGCCCAGCAGGATATA-3’ were used instead of S11 and A4 for the second amplification. The amplification products were cloned into pCR 2.1-Topo vector using the TOPO TA Cloning Kit (Invitrogen, Carlsbad, USA) and propagated in TOP10 One Shot Cells (Invitrogen, Carlsbad, USA) according to the manufacturer’s instructions. When the cloned PCR fragments were too long for a one step sequence analysis, subclonings were performed. Restricted fragments were cloned into pBluescript SK (Stratagene, La Jolla, USA). Sequence analysis was performed with the automatic DNA-sequencer AFL-Express (Pharmacia, Uppsala, Sweden). Standard M13-20Forward, M13Reverse, T3 and T7 primers were used for the sequencing reaction.

For synthesis of hybridization probes clone B9-1 was amplified with primers 5’-pTTGACTCTAGAGCAGTCTCCTTC-3’ and 5’-pATTTTTCTCAATCTCTCTTTCTTCC-3’ and 5’-pATTTTTCTCAATCTCTCTTTCTTCC-3’; clone B9-44 with 5’-pTTTCTCTCTCTTCTTCTTCC-3’ and 5’-pATTGAGGAGGACCTGTGGGTGTA-3’ and 5’-pTTTTCTCTCTTCTTCTTCC-3’; clone C15-36 with 5’-pTTTTATTTCACAACA-3’ and 5’-pATTTTTCTCTCTTCTTCTTCC-3’; clone C19-1 with 5’-pTTTTTCTCTCTTCTTCTTCC-3’ and 5’-pATTGAGGAGGACCTGTGGGTGTA-3’; clone B9-62 with 5’-pCAATTCAATTCAATTCAAC-3’ and 5’-pGGCTATTGACATCTC-3’ and clone B9-46 with 5’-pATGCGAGGATTTTATT-3’ and 5’-pACTCTGATACCTCAGGATG-3’.

DDDB/EMBL/GenBank accession numbers

The sequences of the genomic insertions are deposited under the numbers AF061073 to AF061078.

Acknowledgements

We would like to thank Ingo Schubert, Barbara Hohn, Gernot Presting, Bruno Tinland, Jan Lucht, Frank Hartung and Waltraud Schmidt-Puchta for useful criticism on the manuscript, Jens Stougaard for the plasmid pCNE3I, Peter A. Ley for communication of data prior to publication and Susanne König for sequence analysis. The studies were funded by the grants Pu 137/3-2 and Pu 137/3-3 of the Deutsche Forschungsgemeinschaft.

References


