Plant Molecular Biology **25**: 721–733, 1994.

Non-recombinant background in gene targeting: illegitimate recombination between a *hpt* gene and a defective 5′ deleted *nptII* gene can restore a Km^r phenotype in tobacco

Marcel J.A. de Groot[2,3], Remko Offringa[1], Jürgen Groet[1], Mirjam P. Does[2], Paul J.J. Hooykaas[1,*] and Peter J.M. van den Elzen[2]

[1] *Institute of Molecular Plant Sciences, Clusius Laboratory, Leiden University, Wassenaarseweg 64, 2333 AL Leiden, Netherlands (*author for correspondence);* [2] *MOGEN International nv, Einsteinweg 97, 2333 CB Leiden, Netherlands;* [3] *present address: Calgene Pacific Pty Ltd, 16 Gipps Street, Collingwood, Victoria 3066, Australia*

Received 26 October 1993; accepted in revised form 14 April 1994

Key words: direct gene transfer, gene targeting, homologous recombination, illegitimate recombination, *Nicotiana tabacum*

Abstract

Previously we have demonstrated gene targeting in plants after *Agrobacterium*-mediated transformation. In these initial experiments a transgenic tobacco line 104 containing a T-DNA insertion with a defective neomycin phosphotransferase (*nptII*) gene was transformed with a repair construct containing an otherwise defective *nptII* gene. Homologous recombination between the chromosomally located target and the incoming complementary defective *nptII* construct generated an intact *nptII* gene and led to a kanamycin-resistant (Km^r) phenotype. The gene targeting frequency was 1×10^{-5}. In order to compare direct gene transfer and *Agrobacterium*-mediated transformation with respect to gene targeting we transformed the same transgenic tobacco line 104 via electroporation. A total of 1.35×10^8 protoplasts were transformed with the repair construct. Out of nearly 221000 transformed cells 477 Km^r calli were selected. Screening the Km^r calli via PCR for recombination events revealed that in none of these calli gene targeting had occurred. To establish the origin of the high number of Km^r calli in which gene targeting had not occurred we analysed plants regenerated from 24 Km^r calli via PCR and sequence analysis. This revealed that in 21 out of 24 plants analysed the 5′-deleted *nptII* gene was fused to the hygromycin phosphotransferase (*hpt*) gene that was also present on the repair construct. Sequence analysis of 7 *hpt/nptII* gene fusions showed that they all contained a continuous open reading frame. The absence of significant homology at the fusion site indicated that fusion occurred via a process of illegitimate recombination. Therefore, illegitimate recombination between an introduced defective gene and another gene present on the repair construct or the chromosome has to be taken into account as a standard byproduct in gene targeting experiments.

Introduction

During the past decade a variety of methods for transformation of plants have been developed. Whereas the most common method makes use of *Agrobacterium tumefaciens*-mediated T-DNA transfer [15] a number of other methods such as electroporation involve the introduction of naked

DNA (direct gene transfer) into plants [40]. Irrespective of the transformation method used, foreign DNA appears to integrate into the genome via a process of illegitimate or non-homologous recombination [9, 10, 13, 16, 17, 32, 38]. In contrast to plant and also animal cells, the integration of DNA into lower eucaryotic organisms such as *Saccharomyces cerevisiae* occurs predominantly via homologous recombination [22]. This type of integration, referred to as gene targeting, offers a number of evident advantages over random integration via illegitimate or non-homologous recombination. For instance, it will allow the complete elimination of gene expression by disrupting a specific gene via a targeted insertion, thereby permitting the study of the function of that gene. Furthermore, it will be possible to introduce precise modifications into the genome of an organism or to target the introduced DNA to a locus where its expression is guaranteed. Despite the relatively low frequency gene targeting has become an important tool for modifying the mammalian genome [6, 7]. The first experiments to establish gene targeting in plants have been reported by Paszkowski *et al.* [39] who demonstrated that homologous recombination occurs in plants, albeit with a very low frequency of $0.5 - 4.2 \times 10^{-4}$.

In previous experiments we have demonstrated gene targeting in plants after *Agrobacterium*-mediated transformation [35, 37]. A transgenic tobacco line 104 containing a T-DNA insertion with a defective *nptII* gene was retransformed by *Agrobacterium* with a repair construct containing a defective *nptII* gene with a complementing non-overlapping mutation. Homologous recombination between the chromosomally located target and the incoming repair construct generated an intact *nptII* gene at the target locus and led to a Km^r phenotype. The gene targeting frequency in these experiments was 10^{-5}. In order to compare *Agrobacterium*-mediated transformation and direct gene transfer with respect to their efficiency in gene targeting we also used electroporation to transform the same transgenic tobacco line 104 with the repair construct. A total of 1.35×10^8 protoplasts were transformed and out of nearly 221000 transformants 477 Km^r calli were selected. Screening the Km^r calli via PCR for recombination events revealed that in none of these calli gene targeting had occurred. This means that, apart from the expected restoration via homologous recombination, a functional *nptII* gene can apparently be formed via another mechanism as well. To establish the origin of the relatively high number of Km^r calli in which the *nptII* gene was not restored via homologous recombination we regenerated plants from 24 individual Km^r calli. The results of the detailed PCR and sequence analysis on these plants are presented and discussed.

Materials and methods

Bacterial strains and plasmid constructions

Standard cloning steps were performed according to Sambrook *et al.* [42]. Restriction and modifying enzymes were purchased from Bethesda Research Laboratories and Biolabs and used under the conditions recommended by the supplier. For bacterial cloning *Escherichia coli* K12 strain DH5α [*supE*44 Δ*lac* U169 (∅80*lacZ*ΔM15) *hsdR*17 *recA*1 *endA*1 *gyrA*96 *thi*-1 *relA*1] was used.

The construction of the binary vector pSDM104 which contains the 3′-deleted *nptII* gene was described by Offringa *et al.* [35]. The plasmid pSDM1111 which contains the 5′-deleted *nptII* gene was derived from pSDM1001 [12]. In both pSDM1001 and the positive control construct pSDM7, which was derived from the binary vector pSDM100 [35], we removed the base-pair mutation in the *nptII*-coding region that caused a reduction in the activity of the NPTII protein [49] by exchanging part of the mutant coding region for the wild-type region. In this way pSDM7B was derived from pSDM7. After the base-pair change in pSDM1001, the 137 bp *Sma* I/*Pst* I fragment from the 3′ non-coding region of the *nptII* gene was deleted. Hereafter, this construct was cloned into pLH7 resulting in pSDM1111. The plasmid pLH7 was obtained by

cloning the *hpt* gene that is present in the binary vector pSDM104 into pIC20R [31].

Large-scale isolation and CsCl purification of plasmid DNA was performed as described by Sambrook *et al.* [42]. DNA concentration was determined by measuring the optical density at 260 and 280 nm.

Plant cell transformation

The construction of the plant line 104 (*Nicotiana tabacum* cv. Petit Havana SR1) was described previously by Offringa *et al.* [35]. Mesophyll protoplasts were prepared from leaves of 5–8-week old axenically grown plants of line 104 and electroporated using the Promega X-Cell 450 Electroporation System [12]. The pulse setting was 14.3 ms, 350 V and 1550 μF. For each transformation of 5×10^5 protoplasts 13 μg pSDM1111 plasmid DNA was added. To estimate the transformation frequency of pSDM1111 we performed control electroporations with an equimolar amount of the positive control construct pSDM7B (10 μg). After electroporation the protoplasts were embedded in agarose beads [12]. Seven days after transformation the beads were divided into four equal parts and cultured in 30 ml of A medium [43] containing 50 mg/l kanamycin for selection of transformants. The medium was refreshed weekly by replacing half of the medium with fresh medium; the first time kanamycin was added at 50 mg/l and thereafter the concentration was raised to 100 mg/l. Three weeks after electroporation calli appeared. When 1–2 mm in diameter, the calli were transferred directly to MS medium [34] supplemented with 15 g/l sucrose (MS15), 0.7% Daichin agar (Brunschwig Chemie), 0.1 mg/l NAA, 1 mg/l BAP and 100 mg/l kanamycin for shoot induction. Shoots were rooted on solid MS20 medium containing 100 mg/l kanamycin.

DNA isolation, Southern analysis and PCR

Plant DNA was isolated from young top leaves [33] and purified by CsCl density centrifugation. 10 μg genomic DNA was digested with restriction enzymes and separated on a 0.8% agarose TBE gel. For Southern blots Hybond N^+ membranes were used following the protocol of the supplier (Amersham). DNA probes labelled with α-^{32}P-dCTP were obtained using the random primer labelling kit of Boehringer Mannheim.

The isolation of genomic DNA from callus was according to Lassner *et al.* [28]. Most of the PCR reactions were performed in a Perkin Elmer Cetus DNA Thermal Cycler 480, the other reactions in a Sensa 949 E DNA processor. The reaction mixture contained 1 μg genomic DNA, 25 pmol of each primer, 10 nmol of each dNTP (Sigma), 3 units of *Taq* DNA polymerase and 10 μl of the corresponding 10 × reaction buffer (Promega) in a total volume of 100 μl. The standard PCR protocol was 30 cycles of 1 min 95 °C denaturation, 1 min/55 °C annealing and 2 min/72 °C elongation. The denaturation step of the first cycle lasted 2 min (5 min for the Sensa 949 E DNA processor) and the elongation step of the last cycle was extended to 10 min. The sequence of the primers used for detection of homologous recombination were, 1, 5′-GAACTGACAGAACCGCAACG-3′; 2, 5′-ACCGTAAAGCACGAGGAAGC-3′; 10, 5′-CATGCGATCATAGGCGTCTC-3′. Next to primer 2 the following primers were used in the analysis of non-recombinant Km^r calli: 35S, 5′-GAACTCGCCGTAAAGACTGGCG-3′; AS1, 5′-CCACTGACGTAAGGGATGAC-3′; H3, 5′-AAGCCTGAACTCACCGCGAC-3′; H1, 5′-CCTGACCTATTGCATCTCCC-3′.

Inverse PCR

IPCR was adapted from Does *et al.* [14] and carried out as follows. About 2 μg genomic plant DNA was digested with *Taq* I, ligated under conditions that promoted intramolecular ligation and redigested with *Ban* I. Hereafter, PCR with primers 9 (9P) and 11 (11H) was performed. The sequences of the primers were: 9, 5′-CTATCGTGGCTGGCCACGAC-3′; 9P, 5′-CTATCTG-

CAGCTATCGTGGCTGGCCACGAC-3′; 11, 5′-GCGCTGACAGCCGGAACACG-3′; 11H, 5′-GCATAAGCTTGCGCTGACAGCCGGA-ACACG-3′. Primers 9 and 9P were used at a concentration of 4 μM.

Sequence analysis

In order to clone the IPCR fragments of plant 6B54 and 7F1 amplification was performed using the primers 9P (primer 9 with an extension at the 5′ end that contains a *Pst* I restriction site) and 11H (primer 11 with an extension at the 5′ end that contains a *Hin*d III restriction site). To obtain amplification with primers 9P and 11H the annealing temperature was increased during PCR from 57 °C (10 cycles) to 62 °C (25 cycles). The resulting IPCR fragment was digested with *Pst* I and *Hin*d III and ligated into *Pst* I- and *Hin*d III-digested M13mp18/mp19 vectors. PCR products amplified with primers AS1 and 2 from the plants 1B21, 2B31, 2D2 and 3E21 were cloned into a T-vector [30] that was derived from pIC20H after digestion with *Sma* I. DNA of 3–10 different clones containing one of the IPCR or PCR fragments were mixed to dilute base-pair changes that could have been introduced during the PCR reaction [24]. Sequencing reactions were performed with the Sequenase 2.0 DNA sequencing kit (U.S. Biochemical Corporation) using the M13-40 primer or primer 11. The AS1-2 PCR fragment of plant 2D61 was sequenced directly [48].

Computer analysis

The search for possible topoisomerase I consensus site (5′-(A/T)-(G/C)-(A/T)-T-3′) [8], A + T-rich regions (≥ 5 bp), stretches of alternating purines and pyrimidines (≥ 5 bp), stretches of purines (≥ 5 bp) [26] and short palindromes was done using the MacVector V. 4.0 sequence analysis software (Kodak International Biotechnologies).

Results

Introduction of the repair construct and screening for recombination

For the gene targeting experiments we used plant line 104 (Fig. 1), in which we had previously demonstrated successful gene targeting after *Agrobacterium*-mediated transformation [35, 37]. This plant line harbours an artificial chromosomal target locus consisting of two copies of a T-DNA in an inverted orientation (Fig. 1). Each T-DNA contains a *nptII* gene with a deletion at the 3′ end of the gene and an intact *hpt* gene. The repair plasmid pSDM1111, which is similar to the T-DNA repair construct pSDM101 [35], contains a 5′-deleted *nptII* gene next to the *hpt* gene. Plasmid pSDM1111 does not contain the 1 kb homology that is present between the *hpt* gene and the left T-DNA border sequence of pSDM101. However, the effective homologous region in which homologous recombination has to occur in order to restore an intact *nptII* gene is of exactly the same size in both plasmid and T-DNA constructs. To be able to distinguish a recombination product from the wild-type *nptII* gene that was used as a positive control, a 137 bp marker deletion was introduced in the 3′ non-coding region of the *nptII* gene of repair construct pSDM1111. Before we used the repair construct pSDM1111 for gene targeting we demonstrated that it could restore a complementing 3′-defective *nptII* gene in extrachromosomal homologous recombination (results not shown).

In seven independent experiments we transformed a total number of 1.35×10^8 protoplasts of plant line 104 with pSDM1111, which was digested with *Xho* I and *Kpn* I to create ends that are homologous to the target locus. These experiments yielded about 2.2×10^5 transformants, as calculated from parallel transformations with the positive control construct pSDM7B, and resulted in 477 Km^r calli (Table 1). All the Km^r calli were screened by PCR for the presence of an intact *nptII* gene that had been formed via homologous recombination. The primers used annealed within the regions deleted in the *nptII* gene of either the target or the repair construct. Therefore, only in

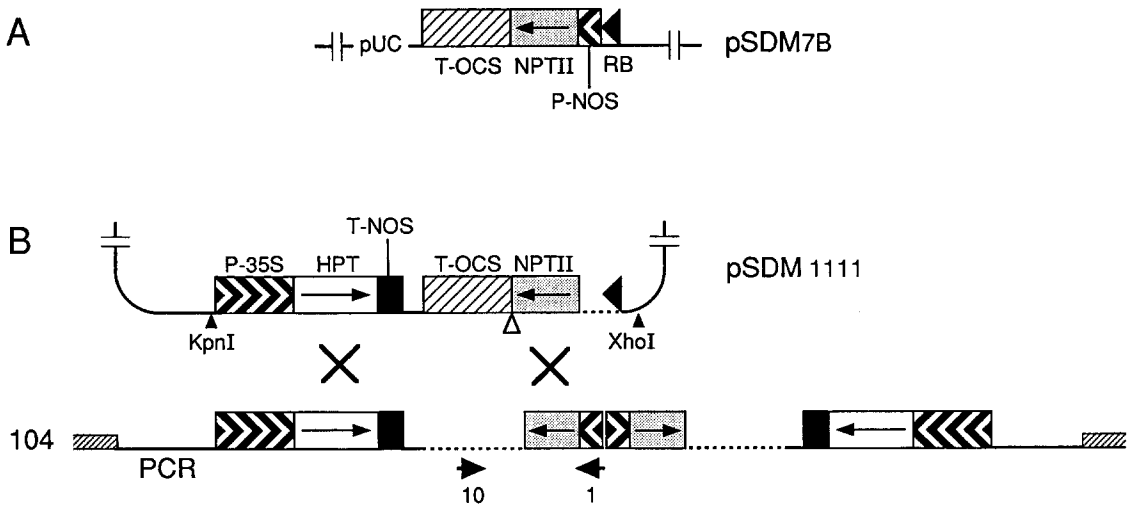

Fig. 1. Strategy for gene targeting of a defective *nptII* gene present at a chromosomal locus in tobacco cells. A. The positive control construct pSDM7B containing the intact *nptII* gene. B. Structure of the chromosomal target for homologous recombination in plant line 104. The target contains two T-DNA inserts derived from the binary vector pSDM104 that are integrated in an inverted orientation. Each T-DNA insert consists of a *hpt* gene and a defective *nptII* gene from which the 3′ part has been deleted. Protoplasts of line 104 were transformed via electroporation with *Xho* I × *Kpn* I-digested pSDM1111 which contains a defective *nptII* gene from which the 5′ part has been deleted. The amount of homology in between the 5′ and 3′ deletions is 613 bp. The *hpt* gene that is present next to the defective *nptII* gene will function as an extra homologous region of 2100 bp. Positions of the primers 1 and 10 that were used in PCR screening are indicated. The fragments amplified with primers 1 and 10 are 1389 bp in size for a wild type and 1252 bp for a recombinant *nptII* gene, respectively. Abbreviations: P-NOS, promoter region of the nopaline synthase (*nos*) gene; NPTII, coding region of the *nptII* gene; T-OCS, transcription terminator of the octopine synthase (*ocs*) gene; P-35S, promoter region of the CaMV 35S transcript; HPT, coding region of the *hpt* gene; T-NOS transcription terminator of the *nos* gene. Symbols: dotted lines, 5′- and 3′- deleted regions of the *nptII* gene; triangle, the 137 bp marker deletion in the 3′-non-coding region of the *nptII* gene; small striped box, flanking chromosomal sequences; thin lines, T-DNA or pUC sequences.

Table 1. Introduction of the repair plasmid pSDM1111 in tobacco protoplasts of line 104.

Experiment[1]	Number of protoplasts	Transformed calli[2]	Km^r calli	Ratio[3]
1	1.6×10^7	13696	10	1:1370
2	2.3×10^7	22448	69	1: 325
3	2.0×10^7	31227	105	1: 297
4	1.5×10^7	31860	22	1:1448
5	1.3×10^7	27248	30	1: 908
6	2.4×10^7	42192	108	1: 391
7	2.4×10^7	52224	133	1: 393
Total		220895	477	1: 463

[1] Numbers 1 to 7 represent independent experiments.
[2] The total number of transformed calli obtained per experiment, as calculated from parallel transformations of 1.0×10^6 protoplasts with the positive control construct pSDM7B.
[3] Ratio between number of Km^r calli obtained and total number of transformed calli.

case an intact *nptII* gene is present a PCR product should be amplified. In that case, primers 1 and 10 will amplify a 1252 bp fragment that is unique to recombinants (Fig. 1). However, it appeared that gene targeting had not occurred in any of the Km^r calli. This means that in our experiments the frequency of homologous recombination was less than 4.5×10^{-6}. Whereas we did not obtain any targeting events when we used electroporation to transform plant line 104, such events were observed at a frequency of 10^{-5} after *Agrobacterium*-mediated transformation [35, 37].

The origin of non-recombinant kanamycin-resistant calli

In order to investigate the origin of the large number of Km^r calli in which gene targeting had not

occurred we regenerated kanamycin-resistant plants from 24 individual calli obtained in experiments 1 to 7 (see Table 1). We expected that certain integration events replaced the deleted *nos* promoter and the region encoding the first twelve N-terminal amino acids of the NPTII enzyme with sequences endogenous to the plant genome, thereby allowing the synthesis of a functional

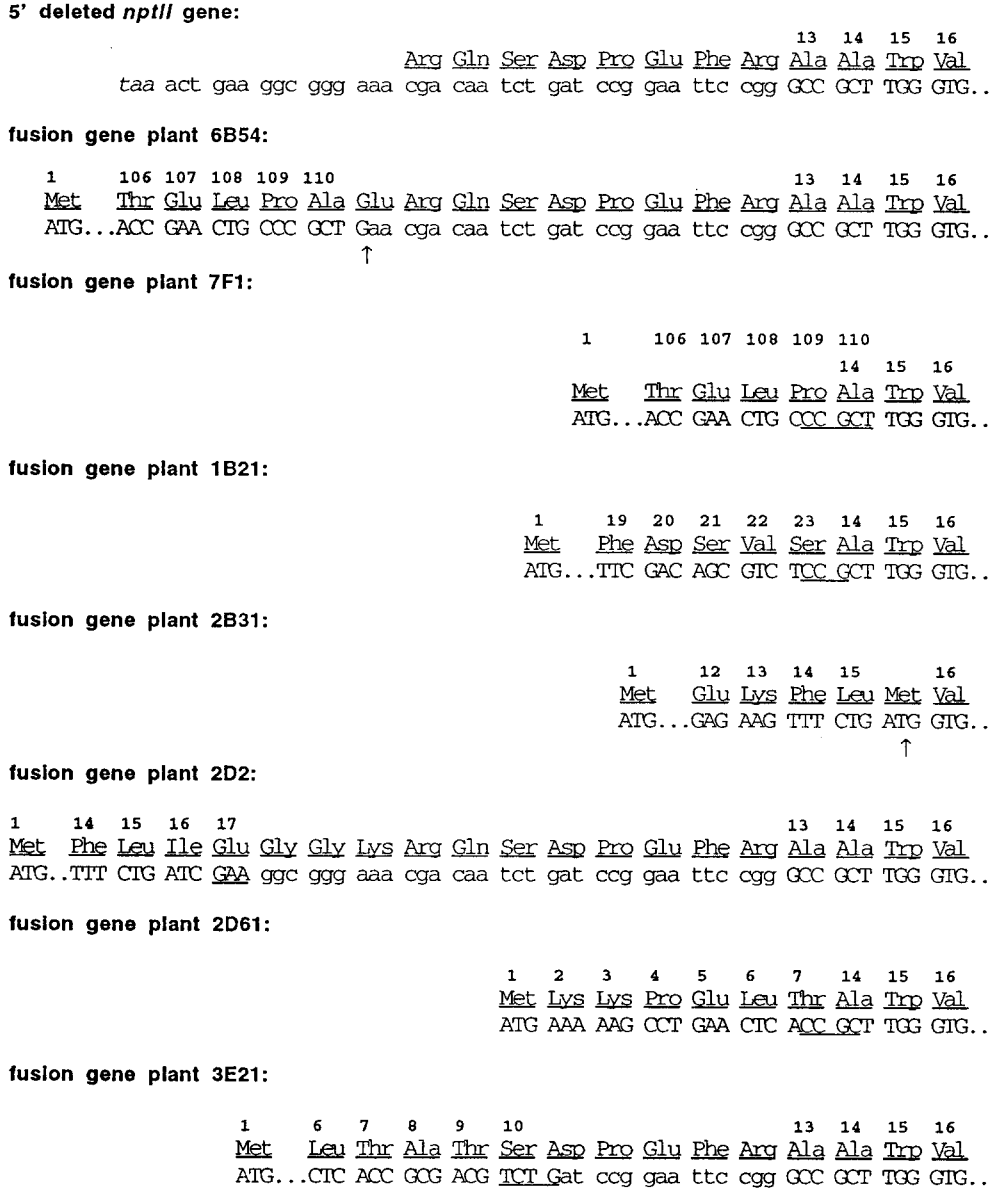

Fig. 2. Outline of the translational fusions between the defective *nptII* gene and the *hpt* gene in non-recombinant Km^r plants. The 5'-deleted *nptII* gene of pSDM1111 is indicated above the *hpt/nptII* gene fusions. The lower-case letters represent the sequence of the *nos* promoter and *Eco* RI linker that are present upstream of the *nptII* sequences. The sequences are divided in triplets which are in frame with the *nptII* gene. An inframe translation termination codon present upstream of the *nptII* gene is depicted in italics. The upper-case letters represent the sequences coding for the HPT and NPTII enzymes. Above the DNA sequences the encoded amino acids and the codon numbers (in bold) are indicated. Regions where the non-homologous recombination occurred are underlined. The exact cross-over point in plant 6B54 and 2B31 is indicated by an arrow.

NPTII enzyme. To test this hypothesis we started screening for plants having only one copy of the repair construct integrated (i.e. one integrated *nptII* gene besides the *nptII* copies at the target locus) in order to be sure that the kanamycin-resistant phenotype indeed resulted from the integration event analysed. Eight out of 24 plants were subjected to Southern analysis (results not shown) and plants 6B54 and 7F1 were found to have a one-copy integration. Subsequently IPCR was used to amplify junc tion fragments between the known *nptII* sequences and flanking sequences. The inverted orientation of the two T-DNA copies at the target locus prevented amplification of a fragment. The IPCR fragments were cloned into M13 vectors and sequenced. In this way we found that the segment flanking the deleted *nptII* gene did not consist of sequences

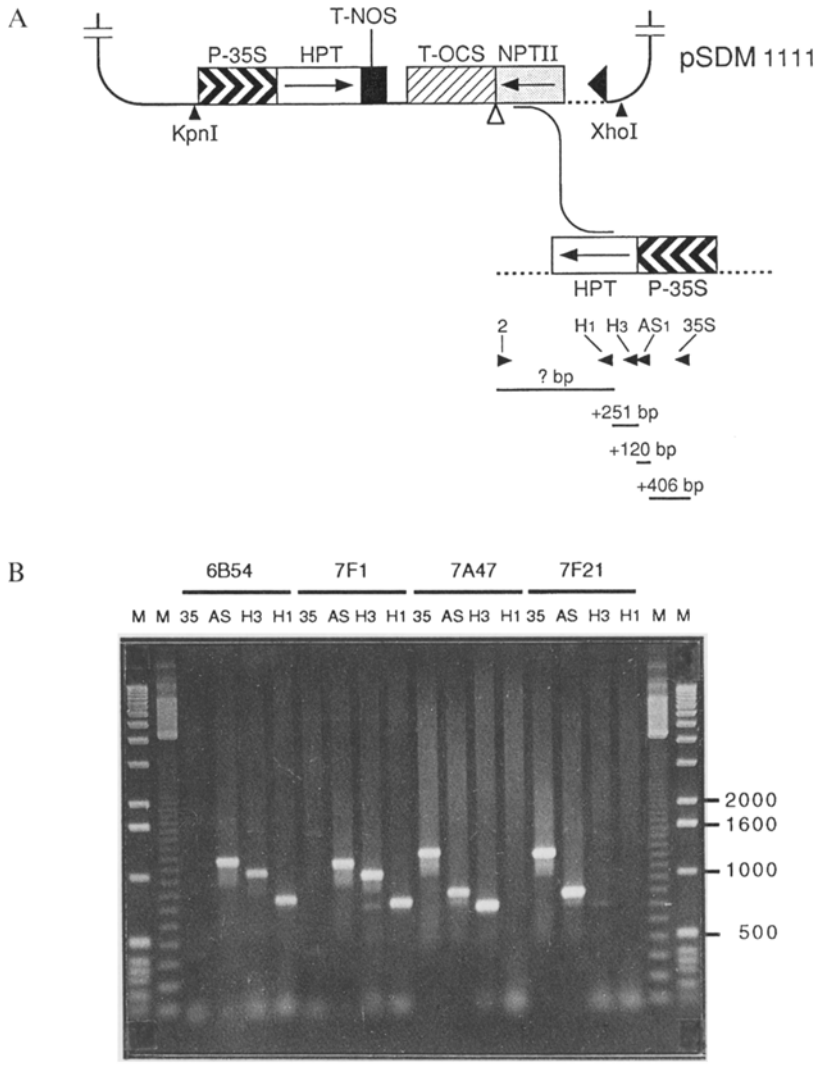

Fig. 3. A. Schematic representation of a fusion between the *nptII* and *hpt* genes and the primer combinations that were used for the PCR analysis. The PCR fragments which are expected are indicated; H1 + 2 should amplify a fragment of unknown size; an amplified H3 + 2 fragment should be 251 bp larger than that of H1 + 2; an amplified AS1 + 2 fragment should be 120 bp larger than that of H3 + 2; an amplified 35S + 2 fragment should be 406 bp larger than that of AS1 + 2. B. PCR analysis on four Km^r plants; 25% of the PCR reaction was ran on a 1.5% TBE agarose gel. Plant names and primer combinations are indicated. M1,2 = 123 bp/1 kb size markers (BRL).

endogenous to the plant genome but instead was completely identical to part of the *hpt* gene. Moreover, the *nptII* and *hpt* sequences were fused in such a way that the *hpt* open reading frame proceeded into the *nptII* open reading frame (Fig. 2). Apparently illegitimate recombination events between the 5′-deleted *nptII* gene and the *hpt* gene can lead to formation of a *hpt/nptII* gene fusion that confers resistance to kanamycin. Whereas theoretically it is possible that the *hpt* sequence was derived from either the *hpt* gene at the target locus or the repair construct, the first possibility seems highly unlikely if one compares the frequency of extrachromosomal homologous recombination and gene targeting [1, 12, 35, 39]. Probably, the illegitimate recombination events occurred extrachromosomally either within one molecule of the repair construct or between two cotransformed repair constructs.

To examine whether other Km^r plants contained similar fusions between the *nptII* and *hpt* genes, we chose a set of primers that anneal at different positions within the *hpt* gene (H1 and H3) or the 35S promoter (35S and AS1; Fig. 3). When a fusion is present in a plant these primers should amplify a PCR fragment in combination with primer 2 which is located in the *nptII* gene (Fig. 3). The results of this PCR analysis (Table 2) demonstrated that *nptII/hpt* fusions are present in 21 out of 24 plants. We did not study the *nptII* genes present in the remaining three plants in more detail, but it is possible that these are fused to endogenous plant genes.

It is evident that a PCR fragment was not always amplified with all four primer sets, but with the exception of plants 3D53 and 2D61 (see later) matching fragments were amplified with primers H3 and 2 or AS1 and 2. This means that all the fusions contain at least the region encoding the first eight amino acids of the HPT enzyme and the minimal 35S promoter [3]. Although the analysis of PCR samples on agarose gels did not allow a precise determination of the fragment size (at best 10–15 bp differences could be detected) it is apparent that in a large number of plants PCR analysis yielded fragments of similar size. To establish whether PCR fragments of similar size corresponded to the presence of specific fusions and whether the fusions contained a continuous *hpt/nptII* open reading frame we determined the sequences of the *hpt/nptII* fusions in plants 1B21, 2B31, 2D2, 2D61 and 3E21 like we did before for plants 6B54 and 7F1 (see above). Indeed we found that in all these fusions the *nptII* and *hpt* genes were fused in such a way that a continuous *nptII/hpt* open reading frame was created (Fig. 2). Nevertheless, as will be described below, all the fusions were different.

Table 2. PCR analysis for the presence of *nptII/hpt* gene fusions in non-recombinant Km^r plants.

Plant line[1]		Amplified fragments with primer combinations			
		35S + 2	AS1 + 2	H3 + 2	H1 + 2
2D61			**831**	–[2]	
7F21	(2)	1260	850	730	
2B31		**1258**	**852**	**732**	
3E21		**1264**	**858**	**738**	
6A53	(≥3)		860	740	
7A13		1270	860	740	
1B21	**(≥5)**		**879**	**759**	
2A51			890	770	
2D3		1300	890	770	
7A45	(≥7)	*800/1000*	890	770	*980*
7A47		1300	890	770	
2D2		1303	897	**777**	*520*
6H32	(2)	*1050*	900	780	
2F6		1370	960	840	
3D53		*920/1080*	*670/800*	840	
6H11			1140	1120	770
7F1	**(1)**		**1137**	**1017**	**776**
6B54	**(1)**		**1170**	**1050**	**799**
7G22		*500/730/1450*	1170/*2000*	1050	800
2D5			1230	1110	860
			1410	1290	1040
6G1			1270	1150	900
2D1		*1000*	*300*		
6E53			*780*	*1600*	
6E62				*720*	

[1] In parenthesis the number of integrated copies as estimated by Southern analysis is indicated. The fusion genes sequenced are depicted in bold. Fusion genes in which PCR fragments are amplified which do not match *nptII/hpt* gene fusions are depicted in italics.

[2] No fragment was amplified because fusion occurred in the region where primer H3 anneals.

Detailed description of the hpt/nptII *gene fusions*

For plant 6B54 the exact fusion site could be established. The fusion gene is predicted to encode a protein in which amino acid 110 (alanine) of the HPT enzyme is linked via a new glutamic acid residue to an arginine encoded by the sequence that is present in front of the deleted *nptII* gene. In plant 7F1 the fusion had occurred in a stretch of 5 nucleotides (5′-CCGCT-3′) that are present in both the *nptII* and the *hpt* gene. The fusion in plant 7F1 encodes a protein in which either amino acid 109 (proline) or 110 (alanine) of the HPT enzyme is fused to amino acid 14 (alanine) or 15 (tryptophan) of the NPTII enzyme. Thus, for the fusion genes of plants 6B54 and 7F1 the fusion points in the *hpt* gene are very close together (1–5 nucleotides) while in the *nptII* gene they are 27–32 nucleotides apart. In plant 1B21 the fusion had occurred in sequence 5′-CCG-3′ that is present in both the *nptII* and the *hpt* gene. The fusion gene of 1B21 encodes a protein in which amino acid 23 (serine) of the HPT enzyme is fused to amino acid 14 (alanine) of the NPTII enzyme. In plant 2B31 the fusion site could be established precisely. At that position the *nptII* and *hpt* genes share no homology. The resulting fusion gene encodes a protein in which amino acid 15 (leucine) of the HPT enzyme is fused via a new methionine to amino acid 16 (valine) of the NPTII enzyme. In plant 2D2 the fusion had occurred in the sequence 5′-GAA-3′ that is present in both the *nptII* and the *hpt* gene. The fusion gene in 2D2 encodes a protein in which amino acid 16 (isoleucine) or 17 (glutamic acid) of the HPT enzyme is fused to either a glutamic acid or glycine encoded by the sequence that is present upstream of the 5′-deleted *nptII* gene. In plant 2D61 the fusion had occurred in the sequence 5′-CCGC-3′ that is present in both the *nptII* and *hpt* gene. Since this fusion occurred in the region of the *hpt* gene where the primer H3 annealed, no fragment was amplified with primer H3 and 2. Nevertheless, the fragment that was amplified with primer AS1 and 2 corresponded with the size calculated from the sequence of the fusion gene. The fusion gene of 2D61 encodes a protein in which amino acid 7 (threonine) of the HPT enzyme is fused to amino acid 14 (alanine) of the HPTII enzyme. Lastly, in plant line 3E21 the fusion had occurred in a sequence, 5′-TCTG-3′, that is present in both the *nptII* and *hpt* genes. The resulting fusion gene encodes a protein in which amino acid 10 (serine) of the HPT enzyme is fused to an aspartic acid residue encoded by the sequence that is present upstream of the 5′-deleted *nptII* gene.

From the results described above it is evident that none of the fusions are identical. Even when the amplified fragments were of a nearly identical size, as was the case with plants 2B31 and 3E21, the fusion sites in the *nptII* and *hpt* genes were still (23–26 nucleotides) apart. As expected, the region in which the fusions occurred in the 5′-deleted *nptII* gene was relatively limited. Most of the variety in fragment length was caused by the large region in which the fusions occurred in the *hpt* gene. In all the sequenced *hpt/nptII* gene fusions the fusion protein encoded contained always at least the valine residue at position 16 and the more C-terminal amino acids of the NPTII enzyme. Thus, although deletion of the 12 N-terminal amino acids of the NPTII enzyme [2] is apparently not enough to completely prevent activity of the enzyme after a fusion, it seems that much larger N-terminal deletions are not tolerated. Site-directed mutagenesis experiments have shown that replacement of valine residue 36 leads to a NPTII enzyme that confers a 20-fold lower level of antibiotic resistance [5]. The fusion genes obtained in our work are predicted to contain at least the first 7 amino acids of the HPT enzyme. Nonetheless, it seems that the deleted part of the NPTII enzyme can be replaced by a variety of amino acid sequences of different composition and length. These are encoded by the *hpt* gene and in a few plants also by the sequence that is present upstream of the 5′ deletion. Apparently, the requirements for restoring a functional NPTII enzyme are not very strict. It could even be that the presence of a translation initiation codon is the only requirement for a functional fusion. Other studies have shown that the NPTII enzyme tolerates large N-terminal fusions [11, 47].

Discussion

In this paper we describe experiments set up to compare electroporation (direct gene transfer) and *Agrobacterium*-mediated transformation with respect to gene targeting in plants. Besides our group [35, 37], three other groups [18, 29, 39] have demonstrated gene targeting using both direct gene transfer and *Agrobacterium*-mediated transformation. However, due to the different experimental designs, it has not been possible to compare gene targeting efficiencies using the two transformation methods. Therefore, we used electroporation to transform protoplasts of plant line 104, which had been used successfully for gene targeting after *Agrobacterium*-mediated transformation [35, 37]. Although the experiments of Offringa *et al.* [35, 37] demonstrated that targeted integration occurred reproducibly in 1 out of 100000 transformed cells, we obtained not a single recombination event from 221000 transformants. However, because of the limited number of targeting events obtained thus far, we are unable to make a statistically justified comparison of the efficiency of both transformation methods with respect to gene targeting. Nevertheless, the results suggest that in the same experimental set-up *Agrobacterium*-mediated transformation is at least as efficient as direct gene transfer in mediating gene targeting in tobacco. Since in tobacco *Agrobacterium*-mediated transformation is clearly more efficient than electroporation, which is a major advantage for gene targeting experiments, we consider *Agrobacterium* transformation as the method of choice for gene targeting in this species. On the other hand, electroporation will probably be very useful in transformation of plant species for which no good *Agrobacterium*-mediated transformation system is available yet. For gene targeting in mammalian cells electroporation is the standard transformation method [7].

In our experiments we unexpectedly obtained a large number of non-recombinant Km^r calli. Considering the experimental set-up of our gene targeting experiments we only expected Km^r calli after homologous recombination between the repair construct and the target locus, since the defective *nptII* gene on the repair construct lacked the *nos* promoter, a translation initiation site and part of the coding region encoding the first twelve amino acids. Furthermore, a stop codon was present upstream of the defective *nptII* gene which had to be deleted before an in-frame fusion with an endogenous gene would be possible. When we started to examine the origin of the non-recombinant Km^r calli we expected that the deleted region would have been replaced with sequences endogenous to the plant genome. Instead we discovered that in 21 out of 24 plants the defective *nptII* gene had been fused to the *hpt* gene that was present on the repair construct to enlarge the region of homology with the target locus. Sequencing showed that the fusion had led to a continuous *hpt/nptII* reading frame. Moreover, the PCR analysis showed that in 95% of the fusions at least the minimal 35S promoter [3] was present. Apparently, illegitimate recombination between the two genes either from one or from two independent repair DNA molecules is relatively efficient as compared to gene targeting. In the experiments where we used *Agrobacterium*-mediated transformation to obtain gene targeting in plant line 104 we had to screen 213 Km^r calli to obtain one gene targeting event [35]. Thus, also in this case we obtained a high background level of Km^r calli. Although we did not study the origin of the non-recombinant Km^r calli in the gene targeting experiments with *Agrobacterium* extensively, PCR analysis of 80 plants (with primers 1, 2, H1 and H2 in one reaction; results not shown) indicated the presence of a *hpt/nptII* fusion in only 6 plants [36]. This suggests that after *Agrobacterium*-mediated transformation the frequency of *hpt/nptII* fusions is about four-fold lower than after electroporation, where primers H1 and 2 amplified a fragment in 7 out of 21 calli. Thus, after *Agrobacterium*-mediated transformation fusions with endogenous genes seem to predominate. This conclusion is in line with the observation that an incoming T-DNA is preferentially integrated into transcriptionally active genes [21, 25]. Moreover, in case of extrachromosomal homologous recombination it has been shown that recombination between T-DNAs is

less efficient than between co-introduced plasmid DNA molecules [4, 12, 35]. Halfter *et al.* [18] have described the restoration of an intact *hpt* gene by gene targeting in *Arabidopsis thaliana* using direct gene transfer in 4 out of 150 selected hygromycin-resistant plants. Apparently, in most of these lines the Hm^r phenotype was due to other events than gene targeting. Therefore, illegitimate recombination between an introduced defective gene and another gene present on the repair construct or the chromosome has to be taken into account as a standard by-product in gene targeting experiments.

From the detailed study of 7 *hpt/nptII* gene fusions we can conclude that these fusions were a result of illegitimate recombination events. Whereas the *nptII* and *hpt* genes shared no homology at the fusion site in plants 2B31 and 6B54, the other five plants shared a small stretch of 3 to 5 homologous nucleotides. This absence of long stretches of homologous DNA is a typical feature of illegitimate recombination in mammalian cells [41]. A remarkable aspect of the small stretches of homology at the fusion site is that 4 out of the 5 share the sequence 5′-CYG-3′ and 3 out of those the sequence 5′-CCG-3′. In the latter three (1B21, 2D61 and 7F1) this sequence is present at exactly the same site in the *nptII* gene. However in the *hpt* gene this sequence is located at different sites. The high frequency with which we find this sequence is surprising since illegitimate recombination junctions in animal cells show a great diversity of sequences and no consensus site has been reported [26, 41, 45].

Two groups demonstrated that the integration of T-DNA in the plant genome is mediated by illegitimate recombination events [16, 32]. The plant/T-DNA junctions show a variety of sequences but they do not include the 5′-CCG-3′ sequence. The significance of this sequence is therefore not clear. Since type I topoisomerases have a possible role in illegitimate recombination ([8] for review) we screened the *nptII* and *hpt* genes for the presence of the consensus recognition sequence. Based on *in vitro* assays with a rat liver and a wheat germ enzyme the following consensus sequence has been derived: 5′-(A or T)-(G or C)-(A or T)-T-3′ [8]. Because this consensus sequence is very degenerate it was no surprise that we found many sites in both the *nptII* and the *hpt* gene. In the *nptII* gene it is partially overlapping the sequence at the fusion site in plant 3E21 and it is present in the immediate vicinity of the fusion site in plants 2D2 and 6B54. However, the fusions in the other plants do not have the consensus sequence present within 10 bp of the fusion sites. In the *hpt* gene the consensus sequence is present at the fusion site in plants 6B54 and 7F1 and within 10 bp of the fusions in the other plants. Nonetheless, while the consensus sequence is only 4 nucleotides the total binding site, which encompasses 20–25 bp, is not well defined [27]. Therefore, it remains questionable whether a consensus site is indeed recognized by topoisomerase I and whether topoisomerase I is involved in the illegitimate recombination junctions presented here. In addition, we searched for other features that are usually implied in illegitimate recombination in mammalian cells such as palindromes, A + T-rich DNA segments [23], runs of contiguous purines and alternating purine/pyrimidine tracks [26]. Some of these features were found within 10 bp of the fusion site but they were not always associated with fusion sites. This is in agreement with the data from illegitimate recombination junctions obtained in mammalian cells [26, 45].

Our findings show that illegitimate recombination events were responsible for the background level of Km^r calli in gene targeting experiments. Whereas design of the targeting vector can have a major influence on the level of background [37], it is equally important to increase the absolute frequency of gene targeting itself because that would render the reported background less significant. Gene targeting experiments in mammalian cells revealed a number of variables influencing the efficiency of gene targeting [6, 7]. One of the most important factors turned out to be the length of homologous DNA. Two groups [19, 46] have tested constructs with different degrees of homology for targeting to the *hprt* locus. Thomas and Capecchi [46] observed that a two-fold increase in homology increased the targeting fre-

quency 20-fold, while Hasty *et al.* [19] observed a 100-fold increase in targeting frequency when the homology was increased 5-fold. More important may be the observation that there seems to be a critical length for efficient gene targeting [19]. The minimal amount of homology needed for obtaining gene targeting was 1.9 kb. However, at least 4.2 kb of homology turned out to be required for efficient gene targeting. A similar observation has been reported by Schulman *et al.* [44] for gene targeting in hybridoma cells. This would mean that the amount of homology used in our experiments (2.6 kb) is not sufficient for efficient gene targeting. The effect of this factor on gene targeting frequencies in plants is currently being investigated by our group.

Acknowledgements

We wish to thank Marry Franke-Van Dijk for performing the PCR analysis on the background Km^r calli obtained after *Agrobacterium*-mediated transformation. We thank John Mason for critical reading of the manuscript. This work was supported by the Netherlands Foundation for Biological Research (BION) with financial aid from the Netherlands Organization for Scientific Research (NWO) as well as by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Technology Foundation (STW).

References

1. Baur M, Potrykus I, Paszkowski J: Intermolecular homologous recombination in plants. Mol Cell Biol 10: 492–500 (1990).
2. Beck E, Ludwig G, Auerswald EA, Reiss B, Schaller H: Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn*5*. Gene 19: 327–336 (1982).
3. Benfey PN, Ren L, Chua NH: The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. EMBO J 8: 2195–2202 (1989).
4. Bilang R, Peterhans A, Bogucki A, Paszkowski J: Single-stranded DNA as a recombination substrate in plants assessed by stable and transient recombination assays. Mol Cell Biol 12: 329–336 (1992).
5. Blázques J, Davies J, Moreno F: Mutations in the *aphA*-2 gene of transposon Tn*5* mapping within the regions highly conserved in aminoglycoside phosphotransferases strongly reduce aminoglycoside resistance. Mol Microbiol 5: 1511–1518 (1991).
6. Bradley A: Modifying the mammalian genome by gene targeting. Curr Opin Biotechnol 2: 823–829 (1991).
7. Capecchi MR: Altering the genome by homologous recombination. Science 244: 1288–1292 (1989).
8. Champoux JJ, Bullock PA: Chapter In: Kucherlapati R, Smith GR (eds) Genetic Recombination, pp. 549–574. American Society for Microbiology, Washington DC (1988).
9. Czernilofsky AP, Hain R, Herrera-Estrella L, Lözz N, Goyvaerts E, Baker BJ, Schell J: Fate of selectable marker DNA integrated into the genome of *N. tabacum* DNA. DNA 5: 101–113 (1986).
10. Czernilofsky AP, Hain R, Baker B, Wirtz U: Studies of the structure and functional organization of foreign DNA integrated into the genome of *Nicotiana tabacum*. DNA 5: 473–482 (1986).
11. Datla RSS, Hammerlindl JK, Pelcher LE, Crosby WL, Selvaraj G: A bifunctional fusion between β-glucuronidase and neomycin phospotransferase: a broad-spectrum marker enzyme for plants. Gene 101: 239–246 (1991).
12. de Groot MJA, Offringa R, Does MP, Hooykaas PJJ, van den Elzen PJM: Mechanisms of intermolecular homologous recombination in plants as studied with single- and double-stranded DNA molecules. Nucl Acids Res 20: 2785–2794 (1992).
13. Deroles SC, Gardner RC: Analysis of the T-DNA structure in a large number of transgenic petunias generated by *Agrobacterium*-mediated transformation. Plant Mol Biol 11: 365–377 (1988).
14. Does MP, Dekker BMM, de Groot MJA, Offringa R: A quick method to estimate the T-DNA copy number in transgenic plants at an early stage after transformation, using inverse PCR. Plant Mol Biol 17: 151–153 (1991).
15. Gasser CS, Fraley RT: Genetically engineering plants for crop improvement. Science 244: 1293–1299 (1989).
16. Gheysen G, Villarroel R, Van Montagu M: Illegitimate recombination in plants: a model for T-DNA integration. Genes Devel 5: 287–297 (1991).
17. Hain R, Stabel P, Czernilofsky AP, Steinbiss HH, Herrera-Estrella L, Schell J: Uptake, integration, expression and genetic transmission of a selectable chimaeric gene by plant protoplasts. Mol Gen Genet 199: 161–168 (1985).
18. Halfter U, Morris PC, Willmitzer L: Gene targeting in *Arabidopsis thaliana*. Mol Gen Genet 231: 186–193 (1992).
19. Hasty P, Rivera-Perez J, Bradley A: The length of homology required for gene targeting in embryonic stem cells. Mol Cell Biol 11: 5586–5591 (1991).

20. Hasty P, Rivera-Perez J, Chang C, Bradley A: Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. Mol Cell Biol 11: 4509–4517 (1991).
21. Herman L, Jacobs A, Van Montagu M, Depicker A: Plant chromosome/marker gene fusion assay to study normal and truncated T-DNA integration events. Mol Gen Genet 244: 248–256 (1990).
22. Hinnen A, Hicks JB, Fink GR: Transformation of yeast. Proc Natl Acad Sci USA 75: 1929–1933 (1978).
23. Hyrien O, Debatisse M, Buttin G, de Saint Vincent BR: A hotspot for novel amplification joints in a mosaic of *Alu*-like repeats and palindromic A + T rich DNA. EMBO J 6: 2401–2408 (1987).
24. Keohavong P, Thilly WG: Fidelity of DNA polymerase in DNA amplification. Proc Natl Acad Sci USA 86: 9253–9257 (1989).
25. Koncz C, Martini N, Mayerhofer R, Koncz-Kalman Z, Körber H, Redei GP, Schell J: High-frequency T-DNA mediated gene tagging in plants. Proc Natl Acad Sci USA 86: 8467–8471 (1989).
26. Konopka AK: Compilation of DNA strand exchange sites for non-homologous recombination in somatic cells. Nucl Acids Res 16: 1739–1758 (1988).
27. Krogh S, Mortensen UH, Westergaard O, Bonven BJ: Eukaryotic topoisomerase I-DNA interaction is stabilized by helix curvature. Nucl Acids Res 19: 1235–1241 (1991).
28. Lassner MW, Peterson P, Yoder JI: Simultaneous amplification of multiple DNA fragments by polymerase chain reaction in the analysis of transgenic plants and their progeny. Plant Mol Biol Rep 7: 116–128 (1989).
29. Lee KY, Lund P, Lowe K, Dunsmuir P: Homologous recombination in plant cells after *Agrobacterium*-mediated transformation. Plant Cell 2: 415–425 (1990).
30. Marchuk D, Drumm M, Saulino A, Collins FS: Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. Nucl Acids Res 19: 1154 (1991).
31. Marsh JL, Erfle M, Wijkes EJ: The pIC plasmid and phage vectors with versatile cloning sites for recombination selection by insertional inactivation. Gene 32: 481–485 (1984).
32. Mayerhofer R, Koncz-Kalman Z, Nawrath C, Bakkeren G, Crameri A, Angelis K, Redei GP, Schell J, Hohn B, Koncz C: T-DNA integration: a mode of illegitimate recombination in plants. EMBO J 10: 697–704 (1991).
33. Mettler IJ: A simple and rapid method for minipreperation of DNA from tissue cultured plant cells. Plant Mol Biol Rep 5: 346–349 (1987).
34. Murashige T, Skoog F: A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol Plant 15: 473–497 (1962).
35. Offringa R, de Groot MJA, Haagsman HJ, Does MP, van den Elzen PJM, Hooykaas PJJ: Extrachromosomal homologous recombination and gene targeting in plant cells after *Agrobacterium* mediated transformation. EMBO J 9: 3077–3084 (1990).
36. Offringa R: Gene targeting in plants using the *Agrobacterium* vector system. Doctoral thesis, Leiden University (1992).
37. Offringa R, Franke-van Dijk MEI, de Groot MJA, van den Elzen PJM, Hooykaas PJJ: Nonreciprocal homologous recombination between *Agrobacterium* transferred DNA and a plant chromosomal locus. Proc Natl Acad Sci USA 90: 7346–7350 (1993).
38. Paszkowski J, Shillito RD, Saul M, Mandak V, Hohn T, Hohn B, Potrykus I: Direct gene transfer to plants. EMBO J 3: 2717–2722 (1984).
39. Paszkowski J, Baur M, Bogucki A, Potrykus I: Gene targeting in plants. EMBO J 7: 4021–4026 (1988).
40. Potrykus I: Gene transfer to plants: Assessment of published approaches and results. Annu Rev Plant Physiol Plant Mol Biol 42: 205–225 (1991).
41. Roth DB, Wilson JH: Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. Mol Cell Biol 6: 4295–4304 (1986).
42. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
43. Saul MW, Shillito RD, Negrutiu I: Direct gene transfer to protoplasts with and without electroporation. In: Gelvin SB, Schilperoort RA (eds) Plant Molecular Biology Manual, pp. A1/1–A1:16. Kluwer Academic Publishers, Dordrecht (1988).
44. Schulman MJ, Nissen L, Collins C: Homologous recombination in hybridoma cells: Dependence on time and fragment length. Mol Cell Biol 10: 4466–4472 (1990).
45. Stary A, Sarasin A: Molecular analysis of DNA junctions produced by illegitimate recombination in human cells. Nucl Acids Res 20: 4269–4274 (1992).
46. Thomas KR, Capecchi MR: Site-directed mutagenesis by gene targeting in mouse embryo derived stem cells. Cell 51: 503–511 (1987).
47. Van den Broeck G, Timko MP, Kaush AP, Cashmore AR, Van Montagu M, Herrera-Estrella L: Targeting of a foreign protein to chloroplasts by fusion to the transit peptide from the small subunit of ribulose-1,5-bisphosphate carboxylase. Nature 313: 358–363 (1985).
48. Winship PR: An improved method for directly sequencing PCR amplified material using dimethyl sulphoxide. Nucl Acids Res 17: 1266 (1989).
49. Yenofsky RL, Fine M, Pellow JW: A mutant neomycin phosphotransferase II gene reduces the resistance of transformants to antibiotic selection pressure. Proc Natl Acad Sci USA 87: 3435–3439 (1990).