THE T-DNA ON ITS WAY FROM *AGROBACTERIUM TUMEFACIENS* TO THE PLANT

B. Hohn[1], Z. Koukolíková-Nicola[1], F. Dürrenberger[1,2], G. Bakkeren[1,3] and C. Koncz[4]

[1]Friedrich Miescher-Institut, P.O. Box 2543, CH-4002 Basel, Switzerland; [2]Present Address: University of Geneva, Department of Molecular Biology, CH-1211 Geneva; [3]Present Address: Biotechnology Laboratory, Vancouver, B.C., Canada V6T 1W5; [4]MPI für Züchtungsforschung, D-5000 Köln 30, FRG.

ABSTRACT

Agrobacterium tumefaciens can, under suitable conditions, transfer the T-segment of its tumor inducing plasmid into plants. There it is usually detected integrated in nuclear DNA and expressing its natural or chimaeric genes. The virulence D2 protein, responsible for the endonucleolytic processing of T-DNA, was found covalently attached to the right end of free T-DNA molecules. Unintegrated T-DNA was isolated from plants using a plant viral replicon as T-DNA. Analysis of resulting viruses revealed that the right end of T-DNA was much better preserved than the left one, implying protection by a (virulence?) protein. Comparison of T-DNA integrated in plant DNA with respective preinsertion sites showed that integration invariably was accompanied by a small deletion at the insertion site. Short homologies between target DNA and segments at or close to T-DNA ends seem to have been instrumental in the integration process which may have been aided by virulence protein(s) at the right junction.

INTRODUCTION

Microorganisms of the genus *Agrobacterium* have developed a sophisticated mechanism of genetically and stably transforming their hosts. As a consequence the plant, under the government of the transferred DNA, deviates part of its metabolic resources to build products which only the inciting bacterium can catabolize.

The bacterium only transfers a specific segment of its large Ti (Tumor inducing) plasmid, the T-DNA (Tranfer DNA), into plant cells. Analysis of the transfer process has revealed sequences and functions that are required in *cis* and in *trans* to the T-DNA, respectively. The T-DNA is delimited by two almost perfect direct repeats of 25 bp called border sequences. The T-DNA transfer is mediated by products of the virulence region, located on the Ti plasmid. The bacterium does not carry the T-DNA in a constitutively transferable form but has to be induced by wounded plant cells to render its T-DNA

H. Hennecke and D. P. S. Verma (eds.),
Advances in Molecular Genetics of Plant-Microbe Interactions, Vol. 1, 19–27.

movable. The virulence region, also localized on the Ti plasmid, but separated from the T-DNA, is responsible for this DNA transfer. Signal molecules synthesized in wounded plant tissue induce the expression of virulence genes, the induction being mediated by the action of the VirA and VirG proteins [1-3]. Two virulence loci are directly involved in processing the T-DNA. Proteins D1 and the N-terminal part of D2 code for a DNA topoisomerase and borderspecific endonuclease, respectively [4; 5-8]. As a result a specific nick is introduced in the lower strand of the border repeat sequences. Besides these nicked molecules single stranded T-DNA molecules of lower strand polarity, the so called T-strands, are found as well as doublestranded linear T-DNA versions and a small amount of covalently closed circular molecules [1-3]. The virulence *E* locus codes only for auxiliary functions as in its absence virulence is not abolished but diminished. Virulence E2 protein is a sequence unspecific single stranded DNA binding protein [9-12].

In this communication we will describe T-DNA molecules detected in induced bacterial cells, speculate about the transfer process, describe molecules related to unintegrated T-DNA in plants and finally discuss comparisons of T-DNA inserts into plant DNA with respective preinsertion sequences.

THE T-DNA IN THE BACTERIUM

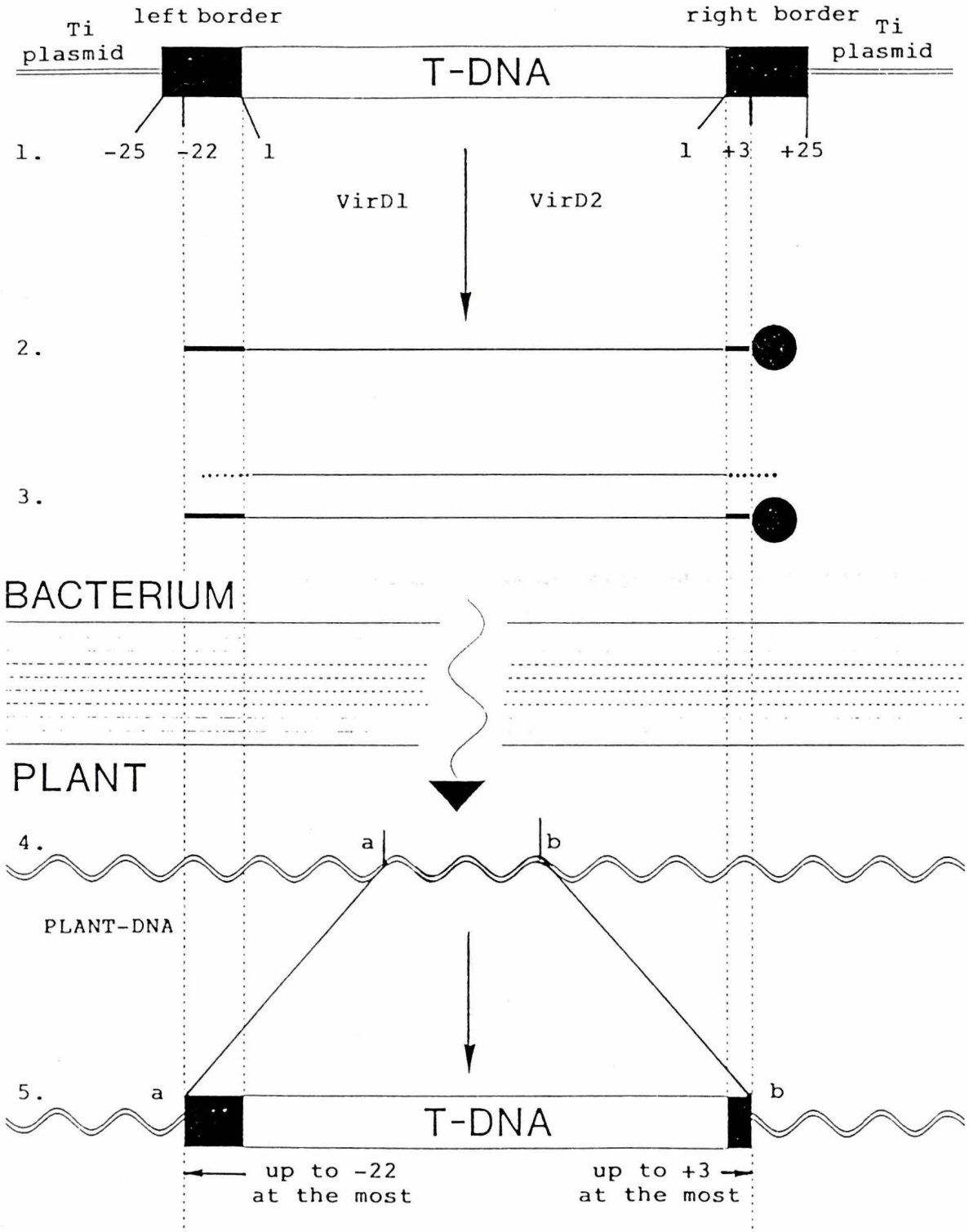
Upon induction of the virulence region specific nicks are introduced in the lower strand of border sequences (Fig.). Single stranded molecules of lower strand polarity, composed exclusively of T-DNA, are formed as a result, possibly by displacement by a newly synthesized bottomstrand [3] see Fig. structure 2. Other T-DNA molecules found in induced *Agrobacterium* cells consist of double stranded linear (structure 3) and a small amount of covalently closed circular molecules. Processing at the border thus explains the absolute requirement of border sequences for transformation, as found in genetic experiments. The endonucleolytic activity resides in the combination of the proteins VirD1 and the N-terminal half of virD2, as experiments in *Agrobacterium* and in *Escherichia coli* [4] have shown. Thereby virD1 exhibits topoisomerase activity whereas in VirD2 the sequence specificity seems to reside.

The VirD2 protein (filled circles in the Figure), after having cleaved the T-DNA between the third and fourth basepair of each border sequence, remains attached to the 5' end of the lower T-DNA strand [13-17]. This protein-DNA association was found to be resistant to SDS, mercaptoethanol, mild alkali, piperidine, and hydroxylamine, indicating that it involves a covalent linkage [16]. The attached protein rendered the right end of doublestranded T-DNA resistant to 5'→ 3' exonucleolytic attack, at least *in vitro* [16]. Thus, this bond may prevent T-DNA degradation also *in vivo* and inhibit religation of T-DNA to the Ti plasmid. The major role(s) of attached VirD2 protein, however, may be in later stages such as T-DNA transfer and/or integration.

T-DNA TRANSFER

In what form the DNA travels to the plant is not known, although an involvement of single

THE T-DNA ON ITS WAY FROM *AGROBACTERIUM TUMEFACIENS* TO THE PLANT



stranded T-DNA at some stage is very likely. It may be complemented to a double stranded form in the bacterium, in the plant, or only upon integration. Adaptation to preexisting bacterial mechanisms seems to be a possible route for the evolution of this unique interkingdom DNA transfer. Several principles may have been combined: a) perception and transduction of environmental stimuli, followed by activation of genes responding to the stimuli, c) induction of proviruses and d) conjugation. Accordingly, T-DNA may be mobilized like conjugating plasmid DNA or like a virus. The need for attachment is not an argument against a virus theory as only closely juxtaposed cells may represent the correct trigger for a virus-transmission. Tumor-inducing virus-particles have been looked for, but so far not found. The single stranded nature of a large part of processed T-DNA molecules may suggest their direct involvement in the transfer, along a relatively safe "mating bridge". Complete safety may not be guaranteed, however, since protecting virulence E2 protein molecules may have to be recruited. The extracellular complementation of *virE* mutants [18,11] can be taken as an argument in favour of involvement of this protein in the actual transfer.

Mobilization functions and origin of transfer of the mobilisable plasmid RSF 1010, in conjuction with Ti plasmid virulence genes, could also be demonstrated to transfer genes to plants [19]. This process may actually mimic an intermediate in evolution, between conjugation between bacteria and "modern" T-DNA transfer from bacterium to plant.

THE T-DNA IN THE PLANT

Once inside the plant-nucleus, by whatever route, T-DNA may not immediately integrate. Experiments to explore the fate of T-DNA inside the plant by chemical methods are difficult to interpret. A different approach to analyze T-DNA molecules in the plant, presumably before/independently of their insertion in the nuclear DNA, has been taken with the use of agroinfection [20]. In these experiments T-DNA was constructed to consist exclusively of a plant viral genome. Once inside a plant, the T-DNA's life as a plant virus can begin as soon as it is circularized. This genetic rescue of T-DNA molecules is plant specific, as opposed to any chemical rescue which would severely suffer from contamination by *Agrobacterium* specific material. Analysis of circle joints, conveniently cloned and amplified by the virus, revealed the following: sequences at the right border were relatively well conserved, with 40% of all rescued molecules including the third base of the border corresponding to the nicking site in the bacterium. This contrasts with a more ragged representation of sequences at the left border. However, even here the rule was followed that not a single base left of the left nick site is ever transferred. A left border mutant (deletion of a base in the conserved core sequence) diminished the efficiency of agroinfection altogether. In addition, it allowed transmission of sequences to the plant which actually were located beyond the nick site, outside of the T-DNA.

These results imply that the right T-DNA end is protected from degradation during transfer. Since the VirD2 protein is found covalently attached to the 5' end of the right T-DNA terminus in *Agrobacterium* (see above), these results suggest the conservation of this

bond in the plant.

As discussed earlier [16], several functions can be imagined for the VirD2 protein covalently attached to T-DNA, possibly in association with other virulence proteins and/or plant proteins: protection of T-DNA during transfer, targeting to the plant nucleus, priming of DNA replication, and integration into the host genome.

T-DNA INTEGRATION

T-DNA integrates into nuclear DNA. *In situ* hybridisation and mapping using genetic or RFLP markers have established that no preferential target for T-DNA integration could be observed [reviewed in 3]. T-DNA insertion units consist of one to several copies of T-DNA. Independent locations of several inserts are also found frequently and have been shown to segregate in subsequent generations. Tandem arrangements of T-DNA have been found to be composed of direct and inverted repeats. Both of these have been found in single T-DNA arrays [3].

Several junctions of T-DNA element to T-DNA element and of T-DNA to plant DNA have been mapped and some of the joints have been sequenced [reviewed in 1,3]. Of the right border repeat maximally three bases have been found retained whereas 24 bases have been found preserved from the left border. Analysis of T-DNA molecules rescued as replicating units allowed a similar conclusion, based on many independently isolated viruses [20]. Since they did not have to go through an integration step, the similarity of the T-DNA ends in the two systems allows the conclusion that integration *per se* did not lead to major distortion of T-DNA ends.

Only two studies have been undertaken until recently, in which the junction sequences of integrated T-DNA was compared to the sequence of unoccupied target DNA. One study revealed a 158 bp duplication of target sequences, now bordering inserted DNA [21]. Some minor rearrangements of target sequences, a small deletion as well as the insertion of small stretches of DNA of unknown origin seem to have accompanied the integration event. The other study comparing pre- and postinsertion sequences revealed that a rather truncated T-DNA had integrated, possibly using small regions of homology at the junctions and again with concommitant creation of a small target deletion [22].

In a more extended analysis Mayerhofer et al [23] analysed the integration pattern of T-DNA into the genome of *Arabidopsis thaliana*. Analysis of seven transformants revealed:

1) no sequence homologies between target sequences and no special features of the target sequences. However, preferential integration into potentially transcribed areas of the analysed and other transformants has been noted, as transcriptional and translational fusions of a T-DNA gene with adjacent cellular sequences were found at similar frequencies in transformants of *Arabidopsis thaliana* and *Nicotiana tabacum* [24]. Since these two species have a markedly differing density of transcribed sequences these results cannot be interpreted as random insertions. Also in mammalian systems DNA in the process of being transcribed may more easily be accessible to invading (Adeno- or Retrovial [25,26]) DNA or, more specifically (but possibly related), may contain more nicks that can serve as entry

points for integrating DNA.

2) small homologies between the right T-DNA end and the target sequence (at b, see Figure). In cases where the right T-DNA end was "complete" (i.e. including up to three bp of border sequence), homologies were less apparent. This implies a special role of the right T-DNA end in accomplishing or helping assimilation of T-DNA into plant DNA and suggests involvement of special proteins such as virulence proteins, possibly in combination with others, in the integration process.

3) small homologies between the left T-DNA end and the target sequence (at a, see Figure).

4) Small rearrangements (direct and inverse duplications) of T-DNA or insert sequences in three cases at one T-DNA/plant DNA junction each.

5) small (29 to 73 bp) deletions (between a and b, Figure) of target DNA.

As suggested by these results, T-DNA (in single- or double stranded form, but in any case with an exposed left 3' end) actively or passively searches for entry into chromosomal DNA. A nick may be detected (or produced?) by the virulence protein armed right T-DNA end; ligation at this point may be followed or preceded by homology search in nearby target sequences. T-DNA may help in exposing target homologies or may exploit DNA replication dependent on temporary single strandedness of target DNA. This homology search must be followed by annealing, mismatch repair, repair synthesis, exo- and endonucleolytic "adaptations" of the recombining partners, and ligation.

Assimilation of right and left end must be independent of each other, as suggested by the independent entry points (a and b, see Figure). The close proximity of these, however, implies a transient link between the T-DNA ends. Such a link could be established by host encoded end to end joining proteins such as have been proposed for *Xenopus laevis* eggs [27], or by (a) virulence protein(s) attached to the right T-DNA end but having affinity to the left end, or a combination of the two. Alternatively, or in addition, the involvement of a 3' protruding upper T-DNA strand in the adherence of T-DNA ends could be invoked. Such T-DNA molecules (see Figure, structure 3) have indeed been detected in induced cells of *Agrobacterium*, but their significance in T-DNA transfer, if any, has remained unknown [Z. Koukolíková-Nicola et al,; 1, and unpublished].

Analysis of tandem T-DNA insertions is not as complete. Such arrangements of T-DNA were found to be composed of direct and inverted repeats. Interestingly, the breakpoints of each pair of T-DNA elements involved in an inverted repeat were, as far as analysed, identical [28]. This may most easily be explained by replication of a double stranded T-DNA molecule, with template strand switching of the replication machinery at the end of a particular element. Since such structures have never been detected in the bacterium they most likely arose in the plant. This would imply that the T-DNA contains a replication origin. A double stranded T-DNA molecule, the suggested template, either arrived from the bacterium as such or is a single stranded molecule converted to the double stranded form. Whether replication of T-DNA before integration is a prerequisite in general is not known. A different kind of mechanism has to be invoked in explaining the head to head link found in a plant transformed with two different T-DNA molecules originating

from one *Agrobacterium* strain [29]: in this case ligation of two double stranded T-DNA elements must have occurred, since 5' ends of single stranded DNA molecules cannot ligate to each other. Thus it seems that several mechanisms may operate within plants to replicate, assemble and integrate T-DNA.

SUMMARY

The T-DNA of *Agrobacterium tumefaciens* is localized like a prophage on the Ti-plasmid. Upon induction of the virulence genes by a plant signal it excises and assembles into a kind of transfer complex. Then it is expelled into a plant cell *via* a bacterium-plant bridge, in a process mimicking certain aspects of bacterial conjugation. In the target cell the T-DNA "plasmid" or "virus" integrates, exploiting bacterial and plant functions.

These are certainly simplifications. Undoubtedly, a more detailed understanding of the mechanisms used in this unique example of interkingdom DNA transfer will be required.

REFERENCES

[1] Koukolíková-Nicola, Z., Albright, L. and Hohn, B. (1987) 'The mechanism of T-DNA transfer from *Agrobacterium tumefaciens* to the plant cell', in Th. Hohn and J. Schell (eds.), Plant DNA Infectious Agents, New York Springer Verlag, pp. 109-148.

[2] Ream, W. (1989) '*Agrobacterium tumefaciens* and interkingdom genetic exchange', Annu. Rev. Phytopathol. 27, 583-618.

[3] Zambryski, P., Tempé, J. and Schell, J. (1989) 'Transfer and function of T-DNA genes from *Agrobacterium* Ti and Ri plasmids in plants', Cell 56, 193-201.

[4] Ghai, J. and Das, A. (1989) 'The *virD* operon of *Agrobacterium tumefaciens* Ti plasmid encodes a DNA-relaxing enzyme', Proc. Natl. Acad. Sci. USA 86, 3109-3113.

[5] Yanofsky, M.F., Porter, S.G., Young, C., Albright, L.M., Gordon, M.P. and Nester, E.W. (1986) 'The *virD* operon of *Agrobacterium tumefaciens* encodes a site-specific endonuclease', Cell 47, 471-477.

[6] Alt-Moerbe, J., Rak, B. and Schröder, J. (1986) 'A 3,6-kbp segment from the *vir* region of Ti plasmids contains genes responsible for border sequence-directed production of T region circles in *E. coli*', EMBO J. 5, 1129-1135.

[7] Yamamoto, A., Iwahashi, M., Yanofsky, M.F., Nester, E.W., Takebe, I. and Machida, Y. (1987) 'The promoter proximal region in the *vir*D locus of *Agrobacterium tumefaciens* is necessary for the plant-inducible circularization of T-DNA', Mol. Gen. Genet. 206, 174-177.

[8] Jayaswal, R.K., Veluthambi, K., Gelvin, S.B. and Slightom, J.L. (1987) 'Double-stranded cleavage of T-DNA and generation of single stranded T-DNA molecules in *Escherichia coli* by a *virD*-encoded border-specific endonuclease from *Agrobacterium tumefaciens*', J. Bacteriol. 169, 5035-5045.

[9] Gietl, C., Koukolíková-Nicola, Z. and Hohn, B. (1987) 'Mobilization of T-DNA from

Agrobacterium to plant cells involves a protein that binds single stranded DNA', Proc. Natl. Acad. Sci. USA 84, 9006-9010.

[10] Das, A. (1988) '*Agrobacterium tumefaciens virE* operon encodes a single stranded DNA-binding protein', Proc. Natl. Acad. Sci. USA 85, 2909-2913.

[11] Christie, P.J., Ward, J.E., Winands, S.C. and Nester, E.W. (1988) 'The *Agrobacterium tumefaciens virE2* gene product is a single stranded-DNA-binding protein that associates with T-DNA', J. Bacteriol. 170, 2659-2667.

[12] Citovsky, V., De Vos, G. and Zambryski, P. (1988) 'Single stranded DNA binding protein encoded by the *virE* locus of *Agrobacterium tumefaciens*', Science 240, 501-504.

[13] Herrera-Estrella, A., Chen, Z., Van Montagu, M. and Wang, K. (1988) 'VirD proteins of *Agrobacterium tumefaciens* are required for the formation of a covalent DNA - protein complex at the 5' terminus of T-strand molecules', EMBO J. 7, 4055-4062.

[14] Ward, E.R. and Barnes, W.M. (1988) 'VirD protein of *Agrobacterium tumefaciens* very tightly linked to the 5' end of T-strand DNA', Science 242, 927-930.

[15] Young, C. and Nester, E.W. (1988) 'Association of the *virD2* protein with the 5' end of T strands in *Agrobacterium tumefaciens*', J. Bacteriol. 170, 3367-3374.

[16] Dürrenberger, F., Crameri, A., Hohn, B. and Koukolíková-Nicola, Z. (1989) 'Covalently bound *virD2* protein of *Agrobacterium tumefaciens* protects the T-DNA from exonucleolytic degradation', Proc. Natl. Acad. Sci. USA 86, 9154-9158.

[17] Howard, E.A., Winsor, B.A., De Vos, G. and Zambryski, P. (1989) 'Activation of the T-DNA transfer process in *Agrobacterium* results in the generation of a T-strand-protein complex: Tight association of VirD2 with the 5' ends of T-strands', Proc. Natl. Acad. Sci. USA 86, 4017-4021.

[18] Otten, L., DeGreve, H., Leemans, J., Hain, R., Hooykaas, P.J.J. and Schell, J. (1984) 'Restoration of virulence of *vir* region mutants of *Agrobacterium tumefaciens* strain B653 by coinfection with normal and mutant *Agrobacterium* strains', Mol. Gen. Genet. 175, 159-163.

[19] Buchanan-Wollaston, V., Passiatore, J.E. and Cannon, F. (1987) 'The *mob* and *oriT* mobilization functions of a bacterial plasmid promote its transfer to plants', Nature 328, 172-174.

[20] Bakkeren, G., Koukolíková-Nicola, Z., Grimsley, N. and Hohn, B. (1989) 'Recovery of *Agrobacterium tumefaciens* T-DNA molecules from whole plants early after transfer', Cell 57, 847-857.

[21] Gheysen, G., Van Montagu, M. and Zambryski, P. (1987) 'Integration of *Agrobacterium tumefaciens* transfer DNA (T-DNA) involves rearrangements of target plant DNA sequences', Proc. Natl. Acad. Sci. USA 84, 6169-6173

[22] Matsumoto, S., Ito, Y., Hosoi, T., Takahashi, Y. and Machida, Y. (1990) 'Integration of *Agrobacterium* T-DNA into a tobacco chromosome: Possible involvement of DNA homology between T-DNA and plant DNA', Mol. Gen. Genet., in press.

[23] Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Bakkeren, G., Crameri, A., Angelis, K., Redei, G.P., Schell, J., Hohn, B. and Koncz, C. (1990) 'T-DNA integration: a mode of illegitimate recombination in plants', submitted.

[24] Koncz, C., Martini, N., Mayerhofer, R., Koncz-Kalman, Z., Körber, H., Redei, G.P. and Schell, J. (1989) 'High frequency T-DNA-mediated gene tagging in plants', Proc. Natl. Acad. Sci. USA 86, 8467-8471.

[25] Schulz, M., Freisem-Rabien, U., Jessberger, R. and Doerfler, W. (1987) 'Transcriptional activities of mammalian genomes at sites of recombination with foreign DNA', J. Virol. 61, 344-353.

[26] Scherdin, U., Rhodes, K. and Breindl, M. (1990) 'Transcriptionally active genome regions are preferred targets for retrovirus integration', J. Virol. 64, 907-912.

[27] Thode, S., Schäfer, A., Pfeiffer, P. and Vielmetter, W. (1990) 'A novel pathway of DNA end-to-end joining', Cell 60, 921-928.

[28] Jorgensen, R., Snyder, C. and Jones, J.D.G. (1987) 'T-DNA is organized predominantly in inverted repeat structures in plants transformed with *Agrobacterium tumefaciens* C58 derivatives', Mol. Gen. Genet. 207, 471-477.

[29] Jouanin, L., Bouchez, D., Drong, R.F., Tepfer, D. and Slightom, J.L. (1989) 'Analysis of TR-DNA/plant junctions in the genome of a *convolvulus arvensis* clone transformed by *Agrobacterium rhizogenes* strain A4', Plant Mol. Biol. 12, 75-85.