9. Homology recognition during t-DNA integration into the plant genome

CSABA KONCZ, KINGA NÉMETH, GEORGE P. RÉDEI and JEFF SCHELL

Contents

Introduction

In bacteria, it is common that chromosomal gene expression is modulated by plasmid or phage genes present at a high copy number. Some extrachromosomally encoded functions are particularly important for genetic exchange between bacteria. Using homologous, site-specific or illegitimate recombination, plasmids or phages can induce DNA transfer to related or unrelated species by conjugation or transduction, respectively. Conjugational transfer of DNA is one of the major forms of horizontal transfer of genetic information between lower prokaryotic and eukaryotic organisms (Davies 1990; Heinemann 1991; Smith 1991).

Agrobacterium: A Relict with Ancient DNA Transfer Mechanism?

Agrobacterium is a soil microorganism that lives in association with the plant rhizosphere. Agrobacteria are hosts for diverse plasmids and phages. Ti and Ri plasmids unique to *Agrobacterium* encode several functions that enable the bacterium to live in close association with plants. These functions are potentially pathogenic because, with their help, the bacterium can alter signal perception, metabolic and developmental pathways of plant cells. Whereas modification of bacterial functions is achieved by plasmid-encoded regula-

tory proteins, genetic colonization of plants requires the transfer of *Agrobacterium* plasmid genes into the plant genome. This interkingdom DNA transfer is meaningful because Ti and Ri plasmid genes transferred to plants are recognized by the plant transcription system (for review, see Binns & Thomashow 1988; Hooykaas 1989; Kado 1991; Zambryski 1988; Zambryski et al. 1989). The capability of interkingdom DNA transfer is not unique to Ti and Ri plasmids of *Agrobacterium*, but can also be demonstrated between other organisms (e.g. between *E. coli* and yeast), when means for detecting the transfer, maintenance and expression of plasmid genes in foreign hosts are available (Buchanan-Wollaston et al. 1987; Heinemann & Sprague 1989). It is remarkable that, among extrachromosomal DNAs persisting in known bacterial species, only *Agrobacterium* plasmids harbor genes with eukaryotic transcription signals. Whether these genes represent a selective conservation of an ancient genetic exchange system, a recent retrotransfer of genetic information from plants to agrobacteria, or a unique evolution of DNA sequences that provide selective advantage in plant-bacterial interaction is an intriguing question (Zambryski et al. 1989; Otten et al. 1992).

Ti and Ri Plasmid Functions for Genetic Engineering of Plants

Regulation of agrobacterial functions resulting in plant pathogenicity is achieved by a limited number of genes located partly in the virulence (*vir*) region of Ti and Ri plasmids. A chief function affecting bacterial signal perception is defined by the *vir*A gene that encodes a membrane-associated chemoreceptor. VirA specifically activates a transcription factor, the product of gene *vir*G, by phosphorylation. VirA belongs to a large family of bacterial chemosensors including PhoB, OmpR and NtrC, that coordinately regulate bacterial responses to environmental stimuli (Mekalanos 1992; Stock et al. 1989; Winans 1991). Certain environmental changes, such as pH, directly affect the conformation, and thereby the activity of the transmembrane VirA receptor (Jin et al. 1990a; Turk et al. 1991). In addition, a binding affinity to different carrier proteins involved in chemical signalling enables the VirA chemoreceptor to modulate the expression of virulence genes in response to various chemicals, such as sugars or amino acids. Whereas transport proteins recognizing these compounds are accessory modulators of VirA activity, carrier proteins binding plant phenolics are specific regulators of the VirA chemoreceptor (Ankenbauer et al. 1991; Cangelosi et al. 1990; Chang and Winans 1992; Huang et al. 1990, Lee et al. 1992; Shimoda et al. 1990; Pazour et al. 1991). Phenolics show species-dependent and developmental stage dependent patterns of synthesis and perform different metabolic, structural and defense functions in plants (Dixon & Lamb 1990). Activation or repression of VirA activity is dependent on the type of phenolic compound (Fortin et al. 1992; Spencer & Towers 1988).

Perception of phenolic signals results in VirA-mediated activation of VirG, a LysR-type transcription activator. VirG binds to *vir*-box DNA sequences representing conserved *cis*-regulatory elements found in the promoters of Ti and Ri plasmid virulence genes/operons *vir*B, C, D, E, F and H (Close et al. 1987; Han et al. 1992; Jin et al. 1990b, 1990c; Mantis & Winans 1992; Pazour et al. 1992; Powell & Kado 1990; Rogowsky et al. 1987). Function and regulation of these *vir* genes have extensively been reviewed (Binns & Thomashow 1988; Hooykaas 1989; Kado 1991; Zambryski 1988).

VirG is similar to a wide family of chromosomally encoded transcription factors that regulate metabolic pathways of bacterial cells (Henikoff et al. 1988; Winans et al. 1986). By partial complementarity between LysR-type transcription activators, metabolic pathways affect the expression of virulence genes and conversely, VirG can modulate the activity of different chromosomal and plasmid genes (Aoyama et al. 1991; Winans 1990).

The start of plant colonization is marked by significant changes in bacterial functions. Agrobacteria become immobilized on plant cell walls by the synthesis of cellulose fibrils, and initiate the copying and single-stranded transfer of Ti or Ri plasmid DNA segments into plant cells (for review see Binns & Thomashow 1988; Kado 1991; Zambryski 1988). The boundaries of the transferred DNA, the T-DNA, are defined by specific 25-bp direct repeats. These repeats are homologous to the *oriT* region of conjugative bacterial plasmids and to vegetative replication origins of single-stranded DNA phages (Cook & Farrand 1992; Waters & Guiney 1993; Waters et al. 1991). The 25-bp repeats and neighboring DNA sequences interact with a protein complex (i.e. T-DNA relaxation complex), known subunits of which are encoded by the *vir* C and D operons (De Vos & Zambryski 1989; Mozo & Hooykaas 1992; Shurvinton & Ream 1991; Toro et al. 1989; Zambryski 1992). Functionally essential components of the T-DNA relaxation complex are the VirD1 topoisomerase and VirD2 endonuclease that produce a nick at the third base-pair of 25-bp T-DNA border repeats (Albright et al. 1987; Dürrenberger et al. 1989; Yanofsky et al. 1986; Wang et al. 1987). A nick at the right 25-bp border probably provides a free 3′-end for priming and determines the polarity of strand-replacement DNA synthesis which is terminated by a nick in the left T-DNA border (Albright et al. 1987; Stachel et al. 1986, 1987). Copying the T-DNA with right-border to left-border polarity requires a number of chromosomally encoded proteins (e.g. DNA polymerase) and is facilitated by a single-stranded DNA binding protein encoded by a Ti plasmid virulence gene, *vir*E2. It is remarkable that unlike other single-stranded DNA-binding proteins, VirE2 can strongly bind and completely cover the T-strand, the single-stranded DNA product of this segmental conjugative DNA synthesis (Christie et al. 1988; Citovsky et al. 1989; Gietl et al. 1987; Sen et al. 1989).

In addition to VirE2, the VirD2 protein of the relaxation complex remains bound to the T-strand, probably by covalent interaction with the 5′-end of single-stranded DNA (De Vos & Zambryski 1989; Dürrenberger et al. 1989; Herrera-Estrella et al. 1988; Howard et al. 1989). Functional analysis of the

*vir*B operon indicates that VirD2 and VirE2 may play an important role in interactions with VirB proteins that constitute membrane pores mediating active transfer of the T-strand into plant cells (for review see Ward et al. 1991; Zambryski 1992). Interestingly, during conjugative transfer of the T-strand into plants, the conjugation of whole Ti or Ri plasmids is suppressed. This is explained by a "cross-talk" between regulation of virulence genes and other genes controlling bacterial conjugation of Ti and Ri plasmids, as well as by the absence of compounds required for induction of Ti and Ri plasmid conjugation. In case of Ti plasmids, the synthesis of such compounds (e.g. octopine and agrocinopines) is encoded by T-DNA genes which are silent in *Agrobacterium*, but expressed in plants (for review see Beck von Bodman et al. 1992; Gelvin & Habeck 1990; Habeeb et al. 1991; Lintig et al. 1991; Steck & Kado 1990; Zambryski et al. 1989).

Transfer and Expression of T-DNA in Plants

Little is known about how the T-strand-VirD2-E2 DNA-protein complex is transported through plant membranes. The recent finding of functionally active nuclear localization signals (NLS) in both VirD2 and VirE2 proteins suggests that they can interact with NLS-binding proteins of nuclear pores and mediate uptake of the T-strand into plant cell nuclei (Citovsky et al. 1992; Herrera-Estrella et al. 1990; Howard et al. 1992; Newmeyer 1993; Shurvinton et al. 1992; Tinland et al. 1992; Zambryski 1992). Integration of the T-DNA predictably requires an interaction between T-strand associated and chromatin-bound proteins, as well as with enzymes involved in DNA synthesis, recombination and repair. DNA replication and repair in eukaryotes are regulated by cell cycle control and signalling pathways of different growth factors (Murray 1992). Therefore, it is intriguing that genes located in the T-DNAs of Ti and Ri plasmids encode proteins involved in the synthesis and modification of major plant growth factors, auxin and cytokinin (for review see Zambryski et al. 1989). It is not known whether T-DNA genes are transcribed before conversion of the T-strand to double-stranded form during integration into plant nuclear DNA. If T-DNA genes are expressed only after integration, the induction of DNA synthesis should be independent of T-DNA encoded functions. In fact, the transfer and integration of the T-DNA is not affected by inactivation of T-DNA genes (Joos et al. 1983; Zambryski et al. 1983). Thus, any DNA fragment carrying 25-bp T-DNA border repeats can be transformed with the help of virulence gene functions from *Agrobacterium* into plants (Hoekema et al. 1983). Nonetheless, T-DNA genes regulating cell division by plant growth factors are required for the proliferation of transformed cells. This function of T-DNA genes can be replaced by exogeneously provided growth factors in vitro when artificial T-DNAs are used for plant transformation (for review see Zambryski et al. 1989).

In planta transformation experiments with a T-DNA vector carrying an intron-containing *uidA* gene revealed expression of the β-glucuronidase reporter enzyme in differentiated tissues, suggesting that T-DNA transfer can occur into non-dividing cells (Vancanneyt et al. 1990). Thus, the induction of DNA synthesis during T-DNA integration is either not necessarily linked to the onset of cell division or T-DNA genes can be expressed before integration into plant chromosomes. Genetic analysis of virulence properties of different *Agrobacterium* strains indicate that bacterial synthesis of plant growth factors (e.g. gene *tzs* encoded cytokinin production) increases the efficiency of T-DNA transformation in various plant hosts (Alt-Moerbe et al. 1988). It is conceivable, therefore, that bacterial production of plant growth factors may trigger DNA synthesis during T-DNA transfer. Alternatively, the T-DNA transfer itself may incite a plant cell response, activating DNA synthesis, repair or recombination. The success of meristem and seed transformation experiments shows that T-DNA transfer can also occur into dividing cells which may give rise to cell lineages producing reproductive cells recovered in transgenic seed progeny (Feldmann 1992; Feldmann & Marks 1987).

Genomic Targets of T-DNA Integration

Although T-DNA genes are typical eukaryotic genes, no significant homology has been found between T-DNA coding regions and plant nuclear genes. Therefore, it appears unlikely that homologous recombination is involved in the T-DNA integration process. Cytological and DNA hybridization studies of T-DNA inserts in various plants have shown that insertions occur at random chromosomal positions and, usually, conservatively maintain T-DNA sequence elements located between the left- and right-border repeats (Ambros et al. 1986; Chyi et al. 1986; Wallroth et al. 1986). Nevertheless, T-DNAs different from intact single-copy insertions are frequently found. Complex inserts may consist of direct or inverted repeats of two or more T-DNAs integrated into the same genomic loci, and it is not unusual that some of these repeats are truncated at their ends (De Block & Debrouwer 1991; Gheysen et al. 1990; Jorgensen et al. 1987; Kwok et al. 1985; Simpson et al. 1982). In the few cases available, nucleotide sequence data indicate that plant DNA sequences flanking T-DNA insertions can suffer deletions, rearrangements and may be duplicated to yield perfect or imperfect, direct or inverted repeats (Gheysen et al. 1987; Matsumoto et al. 1990).

To gain more insight into the integration mechanism at both chromosomal and DNA sequence levels, two approaches were developed. In a genetic approach, promoterless reporter genes, such as kanamycin phosphotransferase (*aph*(3')II), β-glucuronidase (*uidA*) and luciferase (*luxAB* and *luxF*) genes, were inserted in close proximity to T-DNA borders. Integration of the T-DNA into transcribed plant genomic loci could thus be detected by selecting or screening for active plant gene-reporter gene fusions in diverse tissues of T-DNA transformed plants (André et al. 1986; Fobert et al. 1991; Herman et

al. 1990; Kertbundit et al. 1991; Koncz et al. 1989; Topping et al. 1991; Walden et al. 1990). A screening for *aph*(3′)II expression in three plant species showed that T-DNA insertions induce transcriptional gene fusions, on average, in 30% of transgenic plants. The frequency of gene fusions normalized for the copy number of T-DNA insertions was found to be similar in both *Nicotiana* and *Arabidopsis*. These plant species differ strikingly in the complexity and organization of their nuclear genome. Therefore, the data suggests that transcriptionally active chromatin is a major target for T-DNA integration (Koncz et al. 1989). Transformation with "enhancer search" T-DNAs carrying a "core" TATA-box promoter-driven reporter gene at their borders resulted in similar data and showed that T-DNA insertions frequently occur in the vicinity of transcription regulatory DNA sequences (Goldsbrough & Bevan 1991; Topping et al. 1991). Moreover, T-DNA inserts carrying multiple copies of upstream regulatory elements of the Cauliflower Mosaic Virus (CaMV) 35S RNA promoter at their ends were successfully used for activation of the expression of regulatory genes flanking T-DNA integration sites in tobacco (Hayashi et al. 1992). These data are consistent with the hypothesis that DNA synthesis, recombination and repair are coordinately regulated with transcription in plants and other eukaryotes (for review see Andrews & Herskowitz 1990; Heintz et al. 1992; Merill et al. 1992).

T-DNA Integration Involves Illegitimate Recombination

A second approach to explore the mechanism of T-DNA integration aims at the characterization of T-DNA insertion sites in plant chromosomes. Nucleotide sequence analysis of target sites before and after T-DNA integration is expected to unravel the mode of interaction between T-strand and target plant DNA sequences, and to suggest a possible role for proteins participating in the integration process. So far, 16 target sites and corresponding T-DNA-plant DNA insert junctions were isolated from *Nicotiana* and *Arabidopsis* (Gheysen et al. 1991; Matsumoto et al. 1990; Mayerhofer et al. 1991). Based on nucleotide sequence comparison between target sites and insert junctions, T-DNA insert structures can be classified as a) simple unique inserts or b) complex inserts with one or more copies of T-DNA, showing rearrangements of T-DNA or plant DNA sequences at the insert junctions.

Analysis of border sequences of simple unique inserts showed that junction formation with plant DNA may lead to short deletions at both ends of the T-strand. These deletions usually remove only a few nucleotides from the T-strand termini, including the processed *nic*-sites. Terminal sequences of truncated T-DNA ends were found to show homology with short segments of target plant DNA that mark the breakpoints of a deletion formed in the target DNA during T-DNA integration. Truncated ends of the T-DNA are precisely joined to breakpoints of target site deletions, indicating that T-strand termini can anneal by partial homology to free ends of target DNA during integration. Such free DNA

ends may derive from a single nick in the target that is extended to a gap following T-strand invasion. The position of base-paired segments between T-strand and target DNAs seems to define the position of nicks in the target DNA. These nicks correspond to junction-points between T-DNA and plant DNA, as well as to breakpoints of the target deletion formed during integration. Thus, once stabilization of the T-strand occurs by pairing, processing and ligation of interacting DNA ends, nicking in the second strand of the target DNA mediate repair synthesis of the second strand of the T-DNA. Examples of T-DNA inserts carrying truncated ends and terminal homologies with plant DNA targets are shown in Figure 1, together with a simplified model displaying putative events of the integration process.

It is important to note that the temporal order of recombination events shown by this single-stranded gap repair model is tentative. Nicking of both target DNA strands may occur before interaction with the T-DNA. Thus, termini of the T-strand may interact with a double-stranded break in the target. Growing strands (i.e. leading and lagging strands, or Okazaki-fragments) of replicating forks can provide alternative insertional targets for T-DNA integration. Therefore, the single-stranded gap repair model is convertible to a double-strand break repair model that equivalently explains T-DNA integration events (see Mayerhofer et al. 1991).

The mechanism by which terminally truncated T-DNAs integrate into plant chromosomomes is probably identical to the process of illegitimate recombination that mediates the integration of double-stranded DNAs into nuclear and replicating viral genomes of animal cells (Anderson et al. 1984; Marvo et al. 1983; Subramani & Berg 1983). As outlined by general models of homologous and non-homologous (illegitimate) recombination, the same recombination mechanisms are equally effective with single- and double-stranded DNAs (Bollag et al. 1989; Cox & Lehman 1987; Dressler & Potter 1982; Lin et al. 1984; Roth et al. 1985, 1986; Seidman 1987; Wilson et al. 1982). Observation of high-frequency transformation (and recombination) with single-stranded DNAs in mammals and plants suggests that integration of the T-DNA probably uses those mechanisms that mediate the integration of any DNA into chromosomes of plants or mammals (Bilang et al. 1992; Lin et al. 1987; Furner et al. 1989; Rauth et al. 1986; Rodenburg et al. 1989).

A basic difference between homologous and nonconservative recombination processes is that for stabilization of initial synapsis between non-homologous recombinational substrates, an initial pairing and processing of single-stranded DNA ends is absolutely required (Anderson & Eliason 1986; Brenner et al. 1985; Runitz & Subramani 1984). Pairing of 5 to 7 bp between DNA ends was found to be sufficient to trigger illegitimate recombination in animal cells. A recombination assay based on self-ligation of T-DNAs carrying a CaMV replicon resulted in similar data. As observed for truncated ends of stable chromosomal T-DNA inserts, few complementary base-pairs at the joining ends of linear T-DNAs is sufficient for formation of illegitimate junctions yielding circular DNA molecules in plants (Bakkeren et al.

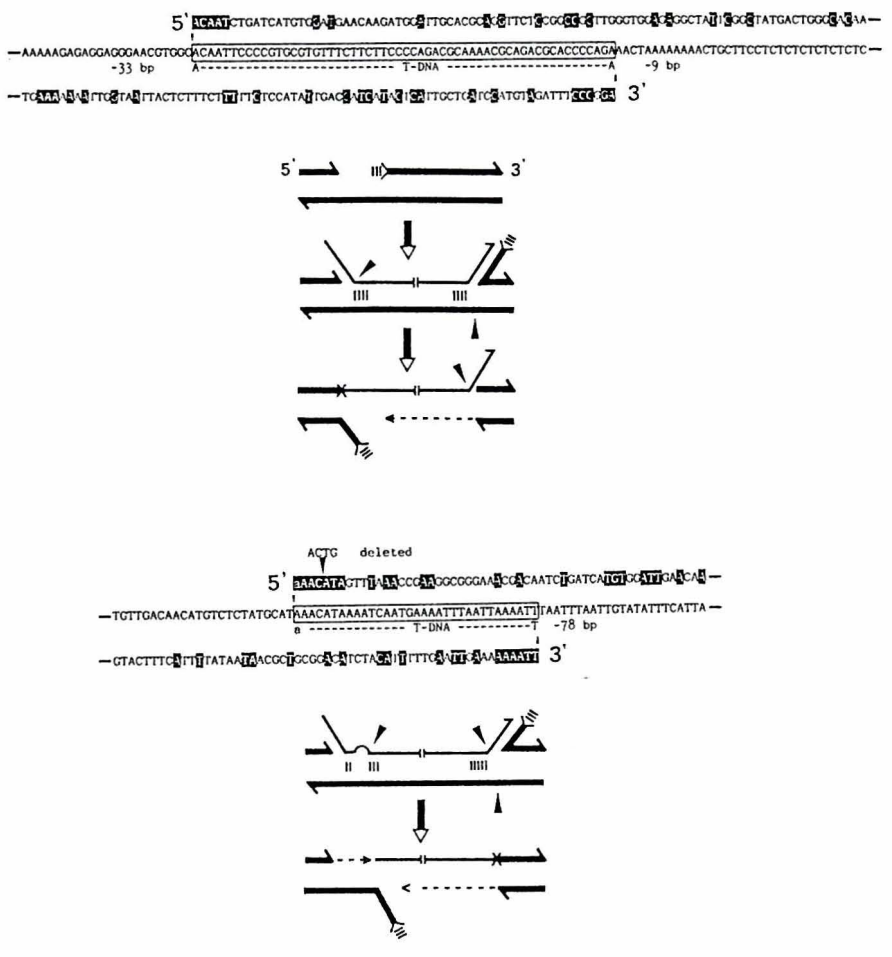

Figure 1. Simple unique T-DNA inserts with terminal deletions. **Upper section**, *sequence alignment*: 5′ (above) and 3′ (below) sequences of the T-strand found at T-DNA-plant DNA insert junctions are aligned with the sequence of corresponding plant target DNA (middle). At the 5′-end an ACAAT sequence and at the 3′-end a CCCgGA sequence (printed in black) of the T-strand show homology with the target DNA. These sequences were found to be precisely joined to breakpoints (marked by small vertical black arrows) of a deletion (boxed sequences) formed by the integration process in the target DNA. During integration 33 bp (–33 bp, marked to the left below the target site sequence) were deleted from the 5′-end of the T-strand, whereas the 3′-end suffered a deletion of 9 bp (–9 bp, marked to the right below the target DNA sequence). Other nucleotides matching between DNA sequences of target site and T-strand termini are also printed in black. Note a near-equal distribution of homology with the target DNA in both T-strand termini. A schematic model below the DNA sequence alignment shows the most likely version of integration events. According to this scheme, the T-strand (marked by an interrupted thin black line) invades a nick in the target DNA (marked by thick black bars) that is being extended to a gap during the integration. (Flush ends of plant DNA, (bars), and of the T-strand (lines) mark 5′ DNA ends, whereas half arrow-head at their other ends label 3′ DNA ends.)

1989). Interactions between the DNA ends during illegitimate recombination are not substantially different from the mode of interactions between DNA strands during homologous recombination, because homologous recombination ultimately leads to double-stranded breaks and repair at the newly joined ends (Craig 1988; Sancar & Sancar 1988; Sugawara & Haber 1992; Szostak et al. 1983; Takahashi et al. 1992; Wake et al. 1985). Therefore, integration of the T-DNA and other single-stranded DNAs may well be described by any scheme which is compatible with the double-strand break model of general recombination.

Possible Role of VirD2 and VirE2 Proteins in Illegitimate Recombination of the T-DNA

In the majority of simple T-DNA inserts, homologous segments with the target DNA and deletions are found only at the 3′-end of the T-strand. Plant DNA junctions with the 5′-end of the T-strand usually contain three nucleotides from the 25-bp T-DNA border repeat. These 5′-end junctions are thus precisely formed at the position of the *nic*-site as shown by examples in Figure 2.

It is well documented that the last nucleotide of the *nic*-site is tightly associated with the VirD2 protein. Nuclear localization signals of VirD2 probably aid the entry of the T-strand to plant cell nuclei (for review see Herrera-Estrella et al. 1990; Howard et al. 1992; Shurvinton et al. 1992; Tinland et al. 1992; Zambryski 1992). VirD2 is a site-specific, single-stranded DNA endonuclease that in *Agrobacterium* interacts with the VirD1 subunit of the border processing relaxation complex. Frequent preservation of the *nic*-site in 5′-end T-strand junctions suggests that VirD2 may actively participate in the integration process. VirD2 may lead the T-strand to nicks in the target, possibly via interaction with chromatin-bound proteins, such as topoisomerases. Based

The ends of the T-strand invade the gap and anneal with target DNA sequences by finding partial homology (marked by vertical lines between the T-strand and target plant DNA). At the position of base-paired segments between T-strand and target DNAs, nicks are formed (marked by large black triangles) that remove the overhangs from both T-strand and plant DNA termini. Alternatively, the ends may be removed by exonucleolitic activities, as indicated by an arrow representing a 5′ to 3′ exonuclease activity at the 5′-end of displaced target DNA strand. A nick in the second (lower) strand of the target, formed at the position of homologous segment between the 3′-end of the T-strand and the target DNA, provides a free 3′-end that is used as primer for DNA repair synthesis of the second strand of the T-DNA (marked by a scattered line with arrow head). **Lower section**, *sequence alignment*: At the 5′-end of the T-strand a deletion of 4 bp was found. The alignment indicates that the 5′-end probably formed a heteroduplex with an AAACATA target sequence which was removed during integration. The 3′-end of the T-strand annealed with a target segment at an AAAATT sequence that is localized 78 bp upstream of the left 25-bp border sequence. Integration scheme: The process is identical with the one described above, except for putative heteroduplex repair at the 5′-end. Due to annealing of an internal T-strand sequence with the target, a segment of 78 bp is removed from the 3′-end of the T-strand.

on analogies between DNA gyrases, topoisomerases and VirD1-VirD2 subunits of the T-DNA relaxation complex, it is tempting to speculate that covalent binding to DNA results in an activated VirD2 protein that can even mediate the ligation of the T-strand to free 3′-ends within plant DNA targets. In fact,

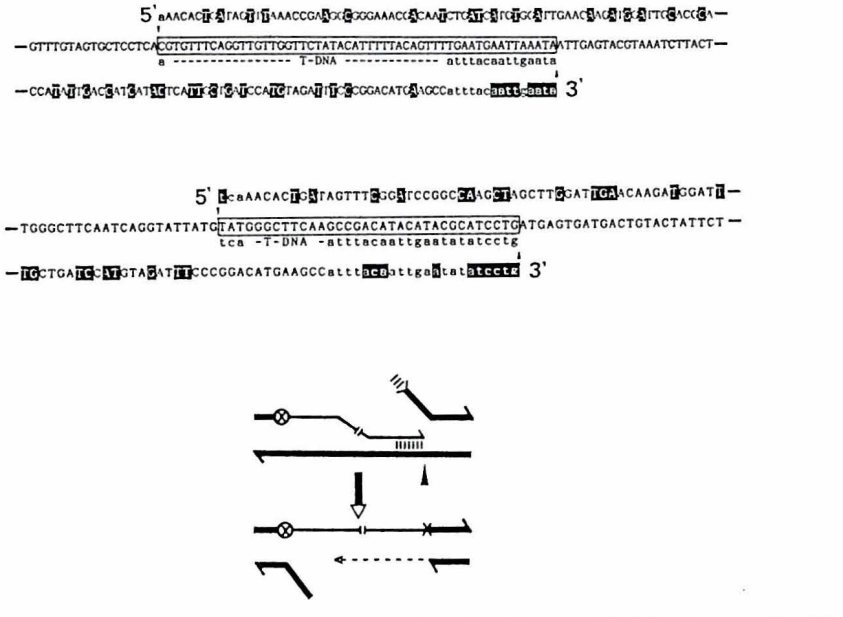

Figure 2. Simple unique T-DNA inserts retaining the 5′ *nic*-site. Symbols are listed in Figure 1 legend. Nucleotide sequences representing portions of 25-bp T-DNA border repeats retained in the T-strand after processing are printed in lower case below the target DNA sequence in the middle. **Upper** *sequence alignment*: A plant DNA junction with the 5′-end of the T-strand contains a single base-pair from the 25-bp repeat that represents a secondary *nic*-site used occasionally during T-strand processing in *Agrobacterium*. The 5′-end does not show terminal homology with the target. The 3′-end of the T-strand carries a DNA segment homologous with an AATTgAATA sequence in the target DNA. This segment of homology is located precisely at the right breakpoint of a deletion in the target. The other breakpoint of target deletion was found to be identical with the position of junction-point between the 5′-end of the T-strand and plant DNA. **Lower** *sequence alignment*: The 5′-end of the T-strand is joined to plant DNA at the canonical *nic*-site that retains three nucleotides (TCA) from the 25-bp T-DNA border repeat. The 3′-end of the T-strand carries a terminal homology to an ATCCTG segment of target DNA that is located exactly at the breakpoint of target site deletion. *Integration scheme*: As described in the text, the integration process is identical to that shown in Figure 1, except a VirD2-mediated ligation event at the 5′-end of the T-strand (marked by a circle with X). In this simplified scheme the T-strand invades a nick in the target. The 5′-end of the T-strand is ligated to the 3′-end within the nick. The 3′-end of the T-strand invades the nick and slides along the target by displacing a target DNA strand. Exonucleolitic processing of displaced target DNA strand extends the nick to a gap. Movement of the 3′-end of the T-strand is terminated by base-pairing with complementary DNA sequences in the target DNA. At the position of base-paired segment between T-strand and target DNA sequences, a nick is introduced into the second strand of target DNA. This nick provides a primer for synthesis of the second strand of T-DNA.

in vitro experiments with purified VirD1 and VirD2 proteins indicate that VirD2 is active as a nicking-closing (endonuclease-ligase) enzyme on single-stranded DNA substrates (Vogel & Das 1992; F. Jasper personal communication).

VirD2-enhanced ligation of the T-strand to 3′-ends of nicks in the target may greatly enhance the illegitimate recombination process described above. Interaction of the 3′-end of T-strand with the target may be facilitated at the same time by the VirE2 protein that is probably co-transferred with the T-strand into plant cell nuclei.

The invasive 3′-end of the T-strand appears to function analogously to a recombinational "breathing" strand. The movement of the invading T-strand in the target may be triggered by a RecA-like protein that is able to "slide over" longer mismatches (Bianchi & Radding 1983; Dressler & Potter 1982; Smith 1987; Pang et al. 1992). The T-strand donated VirE2 protein may induce unwinding the target by binding to the displaced target DNA strand. The resulting structure, carrying a free 3′-end of the T-strand and an open loop/gap in the target, is similar to a replication fork. The essential function of single-stranded DNA binding proteins in DNA replication (Chase & Williams 1986) suggests that VirE2 can even function as a subunit of a DNA polymerase complex binding to T-strand-formed replication forks.

Further events of the integration are greatly dependent on the fate of the 3′-end of the T-strand. Movement of the 3′-end of the T-strand in the target DNA loop/gap is stopped by finding short, complementary sequences within the target. If the displaced target DNA strand and the unpaired 3′-end segment of the T-strand are exonucleolitically processed and ligated, the integration process will lead to a heteroduplex with a looped out T-strand. Heteroduplex resolution is expected to yield a nick in the second target DNA strand at the position of the base-paired segment between the T-strand and target DNAs. This leads to a relaxed gap structure with a free 3′ target DNA end that may serve as primer for DNA repair synthesis of the second strand of the T-DNA. From this model, it is possible to predict that the integration process ultimately results in a deletion of target DNA sequences. In fact, such deletions were identified and localized between nick sites corresponding to junction-points between T-DNA and plant DNA (Figure 2). The integration mechanism of simple unique inserts with intact 5′ *nic*-sites thus likely involves well-known events of general recombination, DNA repair and replication, except the step of VirD2-mediated ligation at the 5′-end of integrating T-strand (see references above).

Complicated Cases of T-DNA Integration: Filler Dnas and Repeats

The integration mechanism of complex T-DNA inserts is probably not different from that of simple unique inserts, but may differ in the timing of events occurring at the joining DNA ends. Complex inserts may carry single-copy T-DNA, the ends of which are either intact or deleted. Junction-points between

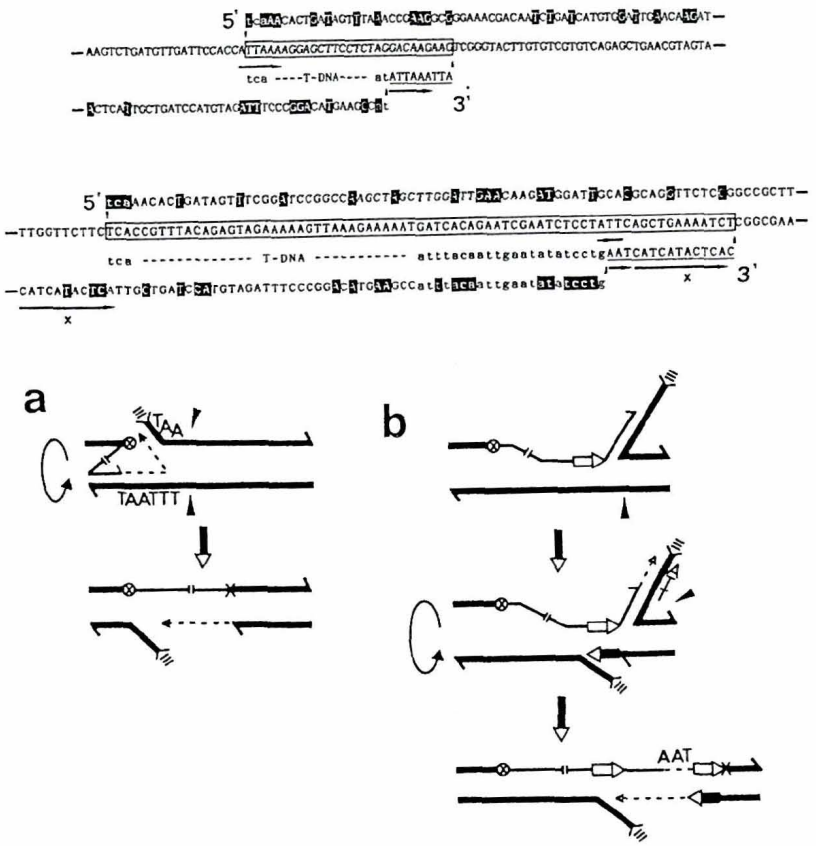

Figure 3. Complex T-DNA inserts: formation of filler DNAs by DNA polymerase template switch and slipping at the junctions. Symbols are listed in Figure 1. **Upper** *sequence alignment:* The 5′-end of the T-strand is joined to plant DNA via the canonical *nic*-site (TCA nucleotides). The 3′-end of the T-strand shows scattered homology with the target. Between the 3′ terminus of the T-strand and the corresponding breakpoint of target deletion a filler DNA sequence, ATTAAATTA, was identified. A part of this filler DNA sequence was found complementary to a TAATTT sequence of target DNA, whereas the reverse complement of remaining nucleotides was identified at the junction site with the 5′-end of the T-strand. Simulation of a template switch at the position of junction between plant DNA and the 5′-end of the T-strand adds exactly the filler DNA sequence to the 3′-end of the T-strand, as primer. An integration scheme drawn in *section a)* shows putative events following this template switch. These include: ligation of the 3′-end of the T-strand with the processed 5′-end of the displaced target DNA strand, nicking the second target DNA strand at the position of template switch, and synthesis of the second strand of the T-DNA using this nick as primer. **Lower** *sequence alignment*: The 5′-end of the T-strand carries a canonical *nic*-site and is precisely joined to the breakpoint of target deletion. The 3′-end of the T-strand is separated by a filler DNA from the corresponding breakpoint of target DNA deletion. As above, this filler DNA consists of two segments. One segment is a direct repeat of a T-strand sequence located upstream of the 3′ terminus (labelled by X). The other segment is represented by nucleotides AAT, the reverse complement of which is located

the left or right T-DNA ends and plant DNA in complex inserts differ from the breakpoints of target deletions produced by the integration. Gaps between the T-DNA ends and breakpoints of target deletions are filled with DNA of apparently unknown origin. A careful analysis of these "filler" DNA sequences, however, always reveals homologies to DNA segments located within the T-DNA or target DNA at various distances upstream or downstream of insert junctions. Mechanisms generating such filler DNA repeats are known from the analysis of recombinational junctions in *E. coli*, yeast, plant and animal cells (Albertini et al. 1982; Bakkeren et al. 1989; Brenner et al. 1984, 1985; Lin et al. 1984, 1987; Roth et al. 1989; Whoriskey et al. 1987; Wilson et al. 1982). A common feature of these mechanisms is the action of DNA polymerase and repair enzymes on the DNA ends (provided by integrating DNA and nicks, gaps, breaks or replication loops in the target) before the junctions are formed. These events can be classified according to whether

- modification of DNA ends of the target or
- of the integrating DNA occurs before integration, and
- whether the ends of target and integrating DNAs interact with each other during these modifications.

An independent class of rearrangements takes place after integration and usually involves recombination between repeats that generate further deletions, inversions, translocations and amplifications (for review see Lichtenstein et al. this volume, chapter 00; Albertini et al. 1982; Das et al. 1990; Edlund & Normark 1981; Gloor et al. 1991; Liskay et al. 1984; McArthur et al. 1991; Selker 1990; Tovar & Lichtenstein 1992; Whoriskey et al. 1987).

Pre-insertional modifications may occur in the target DNA before T-strand integration and may be caused by the repair of nicks, gaps or breaks in the target DNA, or by "mistakes" during DNA replication. The majority of aberrant junctions arise, however, by interaction of the 3′-end of the T-strand with the target before the end-ligation reaction. Thus, filler DNAs are frequently formed when DNA polymerase uses the 3′-end of the T-strand as primer and initiates a "run off" DNA synthesis by copying the displaced target DNA strand. This template switch copying results in the formation of short inverted repeat of a target sequence at the 3′-end of the T-strand (see Figure 3a).

The template switch reaction leads to a blocked structure with a short double-stranded "pin" (see Figure 3b) that is usually resolved by nicking the target DNA strand at the position of base-paired segment between T-strand and

in the target DNA exactly at the position where the 3′-end of the T-strand can be positioned by sequence alignement. At this position, the 3′-end shows terminal homology to an ATaTCCTg sequence of the target DNA. As depicted in *section b)*, this arrangement suggests that after base-pairing with the target, the 3′-end of the T-strand primed an abortive template switch by copying the displaced target DNA strand. This resulted in AAT nucleotides of filler DNA. At the same time, nicking the second strand of target DNA at the position of template switch provided a 3′ DNA end for copying a segment from the T-strand. Relaxation of the replication fork to a double-strand break led to repairing the staggered DNA ends which filled the gap X providing the filler DNA, and resulted in the synthesis of the second strand of the T-DNA.

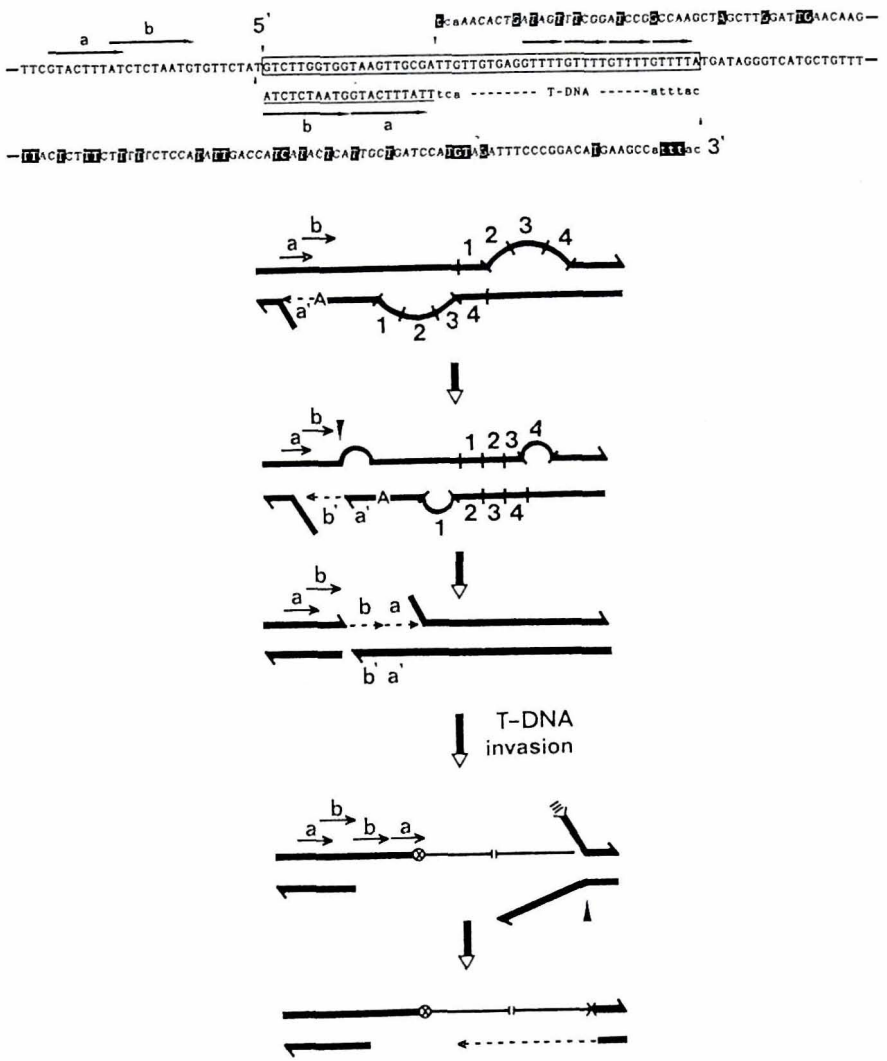

Figure 4. A complex T-DNA insert indicates a pre-insertional target site rearrangement. Symbols are described in Figure 1. *Sequence alignment:* The 5′-end of the T-strand carries a canonical *nic*-site which is separated by a filler DNA from the corresponding breakpoint of target deletion. This filler DNA contains two segments which represent overlapping target DNA sequences, "a" and "b", located upstream of the junction with the 5′-end of the T-strand. The 3′-end of the T-strand carries only three base-pairs terminal homology with the target. Note, however, that the target contains a GTTTT motif, repeated four times, which may weakly interact with GATTT and CATTT motifs in the T-strand. The length of "b" filler DNA repeat is 10 bp. "b" sequence in the target overlaps by a single base-pair with sequence "a" of 9 bp, and is located downstream of "a". In contrast, "b" is located upstream of "a" in the filler DNA, without a single base-pair overlap. A possible explanation for these sequence arrangements is provided in the scheme below. Reverse repetition of "a" and "b" sequences in the filler DNA indicates

target DNAs. DNA synthesis primed by this nick results in initiation of DNA synthesis on the T-strand as template, which may be interrupted by relaxation of the double-stranded break structure, yielding DNA polymerase slipping. These sequential events generate filler DNAs in the junctions that may contain short, direct or inverted repeats of T-strand and/or plant DNA sequences flanking the insertion sites. A unique case shown in Figure 4 indicates that similar events in a double-strand break or replication fork structure may modify the ends of target plant DNA before integration of the T-strand. Several other examples for formation of aberrant T-DNA insert junctions have been described by Gheysen et al. (1991).

Future Prospects: Studies of Repeats and Post-Insertional Rearrangements

An important conclusion drawn from the analysis of simple and complex T-DNA integration events is that the ends of invading T-strand are usually located sterically close to each other. This explains why, in spite of the large size of T-DNA inserts (6 to 40 kb), the integration mechanism usually results, if at all (Németh unpublished), in small target site deletions. Based on the mechanisms of DNA replication and recombination referred to above, it can be deduced that copying of the T-strand may lead to generation of direct or inverted T-DNA repeats, if ligation of the 5′-end of the T-strand did not occur early enough during the integration process. However, little information is available so far about DNA sequence junctions between T-DNA repeats, and it is also unclear whether such repeats can be formed in *Agrobacterium* during T-DNA transfer. Further study of possible VirD2-mediated single-stranded

a two-step copying event. Disappearance of the single base-pair overlap between "a" and "b" suggests that staggered DNA ends, with a single-base overhang, primed DNA copying in a double-strand break (marked by small black arrows at the 5′ breakpoint in the sequence alignment). A false alignment between GTTTT repeats of target DNA would place a nick in the lower strand exactly in the position of sequence "a". (Invasion of the target by the 3′-end of the T-DNA may provide a reason for false annealing, i.e. looping the target may display gradual movement of the 3′-end). Copying "a" is interrupted by a correction of annealing between GTTTT repeats (or by further movement of the T-strand in the target). Thus, after elongation on "a", the free 3′ DNA end is shifted and starts copying "b". Once "b" is copied, a heteroduplex is generated and resolved by nicking (labelled by a black vertical triangle in the second step of the scheme). This leads to a double-strand break, the ends of which are repaired. T-strand invasion occurs at this time-point by ligation of the 5′-end. At the same time the 3′-end of the T-strand reaches its final position by overlapping with the last GTTTT repeat, where a nick is formed. Finally, the 3′-end of the T-strand is ligated with the recessed displaced strand of the target and the second strand of the T-DNA is synthesized by DNA repair. Although apparently much phantasy is required to simulate such an integration event, analogous examples found in *E. coli* and mammals indicate that such complex events are common during DNA integration. As in our case, these examples suggest that integration events may interfere with DNA repair or replication by causing mistakes that may lead to generation of multiple repeats in filler DNAs (Wilson et al. 1982; Albertini et al. 1982; Brenner et al. 1984; Lin et al. 1984).

ligation events is also required for a precise description of the T-DNA multimerization process.

T-DNA is a powerful tool for induction of insertional gene mutations in plants. Recent analysis of genetic linkage between gene mutations and T-DNA inserts in *Arabidopsis* revealed, however, that a large proportion of gene mutations induced by T-DNA transformation are unlinked to T-DNA inserts (Feldmann 1992; Koncz et al. 1992 unpublished). The relative abundance of such unlinked mutations has prompted discussions about general mutagenenicity of the T-DNA transformation process, emphasizing the role of somaclonal variations in plants (a term frequently used for description of still obscure molecular events). The data described above indicate that although the basic model of T-DNA integration is relatively well established, the available results can hardly explain complicated illegitimate events, such as the induction of unlinked mutations by T-DNA insertions.

However, the T-DNA integration model implies several predictions that may help in approaching these problems. Thus, it is possible that the ends of integrating T-strand are able to interact with different chromosomal loci located sterically close in the nucleus. Such interactions may result in chromosome aberrations, such as breakage, deletion or translocation. Generation of filler DNAs at T-DNA border junctions may result in copying of repeats which are potential targets for recombination. Recombination between repeats was demonstrated to generate deletions, inversions, translocations or DNA amplification in bacterial, yeast and mammalian systems (see references above). Interaction of the T-strand with both leading and lagging DNA strands of bi-directional replication forks may also lead to unequal exchange, breakage or inversion events between sister chromatids.

When transferred into the plant, the T-strand carries a bacterial pattern of methylation. Because T-DNA integration involves DNA replication and repair, demethylation of neighboring plant DNA sequences, possibly repeats, may also be induced. T-DNA insertions can thus be considered as false- or undermethylated islands in plant chromatin. Several lines of evidence indicate that foreign-DNA is a target of corrections by methylation that down-regulates the expression of T-DNA genes (Hobbs et al. 1990; Goring et al. 1991; Jorgensen 1990; Kilby et al. 1992; Matzke & Matzke 1991; Meyer et al. 1992; Scheid et al. 1991; Renckens et al. 1992). At the same time, many examples show that alteration of the methylation state of repeated DNA sequences may cause fatal consequences in replication, repair and recombination at multiple genomic loci in several organisms. Interestingly, in all systems studied so far, duplication of undermethylated DNA sequences is required for induction of a correction mechanism, referred to as RIP (rearrangement induced premeiotically, or repeat-induced point mutations) (Kricker et al. 1992; Leonhardt et al. 1992; Meehan et al. 1992; Selker 1990). Would RIP target T-DNA inserts? If so, how is this mechanism connected to the regulation of recombination and DNA replication? Answering these questions is an intriguing task for future studies of T-DNA integration.

Reference

Albertini AM, Hofer M, Calos MP & Miller JH (1982) On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell 29: 319–328.

Albright LM, Yanofsky MF, Leroux B, Ma D & Nester EW (1987) Processing of the T-DNA of *Agrobacterium tumefaciens* generates border nicks and linear, single-stranded T-DNA. J Bacteriol 169: 1046–1055.

Alt-Moerbe J, Neddermann P, von Lintig J, Weiler EW & Schröder J (1988) Temperature-sensitive step in Ti plasmid *vir*-region induction and correlation with cytokinin secretion by *Agrobacteria*. Mol Gen Genet 213: 1–8.

Ambros PF, Matzke AJM & Matzke MA (1986) Localization of *Agrobacterium rhizogenes* T-DNA in plant chromosomes by *in situ* hybridization. EMBO J 5: 2073–2077.

Anderson RA & Eliason SL (1986) Recombination of homologous DNA fragments transfected into mammalian cells occurs predominantly by terminal pairing. Mol Cell Biol 6: 3246–3252.

Anderson RA, Kato S & Camerini-Oterno RD (1984) A pattern of partially homologous recombination in mouse L cells. Proc Natl Acad Sci USA 81: 206–210.

André D, Colau D, Schell J, Van Montagu M & Hernalsteens J-P (1986) Gene tagging in plants by a T-DNA insertion that generates APH(3′)II plant gene fusions. Mol Gen Genet 204: 512–518.

Andrews BJ & Herskowitz I (1990) Regulation of cell cycle-dependent gene expression in yeast. J Biol Chem 265: 14057–14060.

Ankenbauer RA, Best EA, Palanca CA & Nester EW (1991) Mutants of the *Agrobacterium tumefaciens virA* gene exhibiting acetosyringone-independent expression of the *vir* regulon. Mol Plant Microb Interact 4: 400–406.

Aoyama T, Takanami M, Makino K, Oka A (1991) Cross-talk between the virulence and phosphate regulons of *Agrobacterium tumefaciens* caused by an unusual interaction of the transcriptional activator with a regulatory DNA element. Mol Gen Genet 227: 385–390.

Bakkeren G, Koukolikova-Nicola Z, Grimsley N & Hohn B (1989) Recovery of *Agrobacterium tumefaciens* T-DNA molecules from whole plants early after transfer. Cell 57: 847–857.

Beck von Bodman S, Hayman TG & Farrand SK (1992) Opine catabolism and conjugal transfer of the nopaline Ti plasmid pTiC58 are coordinately regulated by a single repressor. Proc Natl Acad Sci USA 89: 643–647.

Bianchi ME & Radding CM (1983) Insertions, deletions and mismatches in heteroduplex DNA made by RecA protein. Cell 35: 511–520.

Bilang R, Peterhans A, Bogucki A & Paszkowski J (1992) Single-stranded DNA as recombination substrate in plants as assessed by stable and transient recombination assays. Mol Cell Biol 12: 329–336.

Binns AN & Thomashow MF (1988) Cell biology of *Agrobacterium* infection and transformation of plants. Annu Rev Microbiol 42: 575–606.

Bollag RJ, Waldmann AS & Liskay MR (1989) Homologous recombination in mammalian cells. Annu Rev Genet 23: 199–225.

Brenner DA, Kato S, Anderson RA, Smigocki AC & Camerini-Otero RD (1984) The recombination and integration of DNAs introduced into mouse cells. Cold Spring Harbor Quant Biol XLIX: 151–160.

Brenner DA, Smigocki AC & Camerini-Otero D (1985) Effect of insertions, deletions, and double-stranded breaks on homologous recombination in mouse L cells. Mol Cell Biol 5: 684–691.

Buchanan-Wollaston V, Passiatore JE, Cannon F (1987) The *mob* and *oriT* mobilization functions of a bacterial plasmid promote its transfer to plants. Nature 328: 172–175.

Cangelosi GA, Ankenbauer RG & Nester EW (1990) Sugars induce the *Agrobacterium* virulence genes through a periplasmic binding protein and a transmembrane signal protein. Proc Natl Acad Sci USA 87: 6708–6712.

Chang CH & Winans SC (1992) Functional roles assigned to the periplasmic, linker, and receiver domains of the *Agrobacterium tumefaciens* VirA protein. J Bacteriol 174: 7033–7039.

Chase JW & Williams KR (1986) Single-stranded DNA binding proteins required for DNA replication. Annu Rev Biochem 55: 103–136.

Christie PJ, Ward JE, Winans SC & Nester EW (1988) The *Agrobacterium tumefaciens virE2* gene product is a single-stranded-DNA-binding protein that associates with T-DNA. J Bacteriol 170: 2659–2667.

Chyi YS, Jorgensen RA, Goldstein D, Tanksley SD & Loaiza-Figueroa F (1986) Localization and stability of *Agrobacterium*-mediated T-DNA insertions in the *Lycopersion* genome. Mol Gen Genet 204: 64–69.

Citovsky V, Wong ML & Zambryski P (1989) Cooperative interaction of *Agrobacterium* VirE2 protein with single-stranded DNA: Implications for the T-DNA transfer process. Proc Natl Acad Sci USA 86: 1193–1197.

Citovsky V, Zupan J, Warnick D & Zambryski P (1992) Nuclear localization of *Agrobacterium* VirE2 protein in plant cells. Science 256: 1802–1805.

Close TJ, Rogowsky PM, Kado CI, Winans SC, Yanofsky MF & Nester EW (1987) Dual control of *Agrobacterium tumefaciens* Ti plasmid virulence genes. J Bacteriol 169: 5113–5118.

Cook DM & Farrand SK (1992) The *oriT* region of the *Agrobacterium tumefaciens* Ti plasmid pTiC58 shares DNA sequence identity with the transfer origins of RSF1010 and RK2/RP4 and with T-region borders. J Bacteriol 174: 6238–6246.

Cox MM & Lehman IR (1987) Enzymes of general recombination. Annu Rev Biochem 56: 229–262.

Craig NL (1988) The mechanism of conservative site-specific recombination. Annu Rev Genet 22: 77–105.

Das OP, Levi-Minzi S, Koury M, Benner M & Messing J (1990) A somatic gene rearrangement contributing to genetic diversity in maize. Proc Natl Acad Sci USA 87: 7809–7813.

Davies J (1990) Interspecific gene transfer: where next? TIBTECH 8: 198–203.

De Block M & Debrouwer D (1991) Two T-DNAs co-transformed into *Brassica napus* by a double *Agrobacterium tumefaciens* infection are mainly integrated at the same locus. Theor Appl Genet 82: 257–263.

De Vos G & Zambryski P (1989) Expression of *Agrobacterium* nopaline-specific VirD1, VirD2 and VirC1 proteins and their requirement for T-strand production in *E. coli*. Mol Plant Microb Interact 2: 43–52.

Dixon RA & Lamb CJ (1990) Molecular communication in interactions between plants and microbial pathogens. Annu Rev Plant Physiol Plant Mol Biol 41: 339–367.

Dressler D & Potter H (1982) Molecular mechanism in genetic recombination. Annu Rev Biochem 51: 727–761.

Dürrenberger F, Crameri A, Hohn B & Koukolikova-Nicola Z (1989) Covalently bound VirD2 protein of *Agrobacterium tumefaciens* protects the T-DNA from exonucleolitic degradation. Proc Natl Acad Sci USA 86: 9154–9158.

Edlund T & Normark S (1981) Recombination between short DNA homologies causes tandem duplication. Nature 292: 269–271.

Feldmann KA (1992) T-DNA insertion mutagenesis in *Arabidopsis*: seed infection/transformation. In: Koncz C, Chua N-H & Schell J (eds) Methods in *Arabidopsis* Research (pp 274–289). World Scientific, Singapore.

Feldmann KA & Marks DM (1987) *Agrobacterium*-mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue culture approach. Mol Gen Genet 208: 1–9.

Fobert PR, Miki BL & Iyer VN (1991) Detection of gene regulatory signals in plants revealed by T-DNA-mediated fusions. Plant Mol Biol 17: 837–851.

Fortin C, Nester EW & Dion P (1992) Growth inhibition and loss of virulence in cultures of *Agrobacterium tumefaciens* treated with acetosyringone. J Bacteriol 174: 5676–5685.

Furner IJ, Higgins ES & Berrington AW (1989) Single-stranded DNA transforms plant protoplasts. Mol Gen Genet 220: 65–68.

Gelvin SB & Habeck LL (1990) *vir* genes influence conjugational transfer of the Ti plasmid of *Agrobacterium tumefaciens*. J Bacteriol 172: 1600–1608.

Gheysen G, Van Montagu M & 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.

Gheysen G, Herman L, Breyne P, Van Montagu M & Depicker A (1990) Cloning and sequence analysis of truncated T-DNA inserts from *Nicotiana tabacum*. Gene 94: 155–163.

Gheysen G, Villarroel R & Van Montagu M (1991) Illegitimate recombination in plants: a model for T-DNA integration. Genes & Devel 5: 287–297.

Gietl C, Koukolikova-Nicola Z & 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.

Gloor GB, Nassif NA, Johnston-Schiltz DM, Preston CR & Engels WR (1991) Targeted gene replacement in *Drosophila* via P element-induced gap repair. Science 253: 1110–1117.

Goldsbrough A & Bevan M (1991) New patterns of gene activity in plants detected using an *Agrobacterium* vector. Plant Mol Biol 16: 263–269.

Goring DR, Thompson L & Rothstein S (1991) Transformation of a partial nopaline synthase gene into tobacco suppresses the expression of a resident wild type gene. Proc Natl Acad Sci USA 88: 1770–1774.

Habeeb L, Wang L & Winans SC (1991) Transcription of the octopine catabolism operon of the *Agrobacterium* tumor-inducing plasmid pTiA6 is activated by a LysR-type regulatory protein. Mol Plant Microb Interact 4: 379–385.

Han DC, Chen CY, Chen YF & Winans SC (1992) Altered-function mutations of the transcriptional regulatory gene *virG* of *Agrobacterium tumefaciens*. J Bacteriol 174: 7040–7043.

Hayashi H, Czaja I, Lubenow H, Schell J & Walden R (1992) Activation of a plant gene by T-DNA tagging: auxin-independent growth *in vitro*. Science 258: 1350–1353.

Heinemann JA (1991) Genetics of gene transfer between species. Trends Genet 7: 181–185.

Heinemann JA & Sprague GF (1989) Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. Nature 340: 205–209.

Heintz NH, Dailey L, Held P & Heintz N (1992) Eukaryotic replication origins as promoters of bidirectional DNA synthesis. Trends Genet 8: 376–381.

Henikoff S, Haughn GW, Calvo JM & Wallace JC (1988) A large family of bacterial activator proteins. Proc Natl Acad Sci USA 85: 6602–6606.

Herman L, Jacobs A, Van Montagu M & Depicker A (1990) Plant chromosome/marker gene fusion assay for study of normal and truncated T-DNA integration events. Mol Gen Genet 224: 248–256.

Herrera-Estrella A, Chen Z, Van Montagu M & 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.

Herrera-Estrella A, Van Montagu M & Wang K (1990) A bacterial peptide acting as a plant nuclear targeting signal: the amino-terminal portion of *Agrobacterium* VirD2 protein directs a β-galactosidase fusion protein into tobacco nuclei. Proc Natl Acad Sci USA 87: 9534–9537.

Hobbs SLA, Kpodar P & DeLong MOC (1990) The effect of T-DNA copy number, position and methylation on reporter gene expression in tobacco transformants. Plant Mol Biol 15: 851–864.

Hoekema A, Hirsch PR, Hooykaas PJJ & Schilperoort RA (1984) A binary plant vector strategy based on separation of *Vir*- and T-region of the *Agrobacterium tumefaciens* Ti plasmid. Nature 303: 17–180.

Hooykaas PJJ (1989) Transformation of plant cells via *Agrobacterium*. Plant Mol Biol 13: 327–336.

Howard EA, Winsor BA, De Vos G & 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.

Howard EA, Zupan JR, Citovsky V & Zambryski P (1992) The VirD2 protein of *A. tumefaciens* contains a C-terminal bipartite nuclear localization signal: implications for nuclear uptake of DNA in plant cells. Cell 68: 109–118.

Huang MLW, Cangelosi GA, Halperin W & Nester EW (1990) A chromosomal *Agrobacterium tumefaciens* gene required for effective plant signal transduction. J Bacteriol 172: 1814–1822.

Jin S, Roitsch T, Ankenbauer RG, Gordon MP & Nester EW (1990a) The VirA protein of *Agrobacterium tumefaciens* is autophosphorylated and is essential for *vir* gene regulation. J Bacteriol 172: 525–530.

Jin S, Roitsch T, Christie PJ & Nester EW (1990b) The regulatory VirG protein specifically binds to a *cis*-acting regulatory sequence in transcriptional activation of *Agrobacterium tumefaciens* virulence genes. J Bacteriol 172: 531–537.

Jin S, Prusti RK, Roitsch T, Ankenbauer RG & Nester EW (1990c) Phosphorylation of the VirG protein of *Agrobacterium tumefaciens* by the autophosphorylated VirA protein: essential role in biological activity of VirG. J Bacteriol 172: 4945–4950.

Joos H, Inzé D, Caplan A, Sormann M, Van Montagu M & Schell J (1983) Genetic analysis of T-DNA transcripts in nopaline crown galls. Cell 32: 1057–1067.

Jorgensen R (1990) Altered gene expression in plants due to *trans*-interactions between homologous genes. Trends Biotech 8: 340–344.

Jorgensen R, Snyder C & Jones JDG (1987) T-DNA is organized predominantly in inverted repeat structures in plants transformed with *Agrobacterium tumefaciens* C58. Mol Gen Genet 207: 471–477.

Kado CI (1991) Molecular mechanism of crown gall tumorigenesis. Crit Rev Plant Sci 10: 1–32.

Kertbundit S, De Greve H, Deboeck F, Van Montagu M & Hernalsteens J-P (1991) *In vivo* random β-glucuronidase gene fusions in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 88: 5212–5216.

Kilby NJ, Leyser HMO & Furner IJ (1992) Promoter methylation and progressive transgene inactivation in *Arabidopsis*. Plant Mol Biol 20: 103–112.

Koncz C, Martini N, Mayerhofer R, Koncz-Kálmán Z, Körber H, Rédei GP & Schell J (1989) High-frequency T-DNA-mediated gene tagging in plants. Proc Natl Acad Sci USA 86: 8467–8471.

Koncz C, Németh K, Rédei GP & Schell J (1992) T-DNA insertional mutagenesis in *Arabidopsis*. Plant Mol Biol 20: 963–976.

Kricker MC, Drake JW & Radman M (1992) Duplication-targeted DNA methylation and mutagenesis in the evolution of eukaryotic chromosomes. Proc Natl Acad Sci USA 89: 1075–1079.

Kwok WW, Nester EW & Gordon MP (1985) Unusual plasmid DNA organization in an octopine crown gall tumor. Nucleic Acids Res 13: 459–471.

Lee K, Dudley MW, Hess KM, Lynn DG, Joeger RD & Binns AN (1992) Mechanism of activation of *Agrobacterium* virulence genes: identification of phenol-binding proteins. Proc Natl Acad Sci USA 89: 8666–8670.

Leonhardt H, Page AW, Weier HU & Bestor TH (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71: 865–873.

Lin F-L, Sperle KM & Sternberg NL (1987) Extrachromosomal recombination in mammalian cells as studied with single- and double-stranded DNA substrates. Mol Cell Biol 7: 129–140.

Lin F-W, Sperle K & Sternberg N (1984) Model for homologous recombination during transfer of DNA into mouse L cells: role for DNA ends in the recombination process. Mol Cell Biol 4: 1020–1034.

Lintig J, Zanker H & Schröder J (1991) Positive regulators of octopine-inducible promoters in the nopaline and octopine catabolism regions of Ti plasmids. Mol Plant Microb Interact 4: 370–378.

Liskay RM, Stachelek JL & Letsou A (1984) Homologous recombination between repeated chromosomal sequences in mouse cells. Cold Spring Harbor Quant Biol XLIX: 183–189.

Mantis NJ & Winans SC (1992) The *Agrobacterium tumefaciens vir* gene transcriptional activator *virG* is transcriptionally induced by acid pH and other stress stimuli. J Bacteriol 174: 1189–1196.

Marvo SL, King SR & Jaskunas SR (1983) Role of short regions of homology in intermolecular illegitimate recombination events. Proc Natl Acad Sci USA 80: 2451–2456.

Matsumoto S, Ito Y, Hosoi T, Takahashi Y & 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 224: 309–316.

Matzke MA & Matzke AJM (1991) Differential inactivation and methylation of a transgene in plants by two suppressor loci containing homologous sequences. Plant Mol Biol 16: 821–830.

Mayerhofer R, Koncz-Kalman Z, Nawrath C, Bakkeren G, Crameri A, Angelis K, Rédei GP, Schell J, Hohn B & Koncz C (1991) T-DNA integration: a mode of illegitimate recombination. EMBO J 10: 697–704.

McArthur JG, Beitel LK, Chamberlain JW & Stanners CP (1991) Elements which stimulate gene amplification in mammalian cells: role of recombinogenic sequences/structures and transcriptional activation. Nucleic Acids Res 19: 2477–2484.

Meehan R, Lewis J, Cross S, Nan X, Jeppesen P & Bird A (1992) Transcriptional repression by methylation of CpG. J Cell Sci Suppl 16: 9–14.

Mekalanos JJ (1992) Environmental signals controlling expression of virulence determinants in bacteria. J Bacteriol 174: 1–7.

Merill GF, Morgan BA, Lowndes NF & Johnston LH (1992) DNA synthesis control in yeast: An evolutionary conserved mechanism for regulating DNA synthesis genes? Bioessays 14: 823–830.

Meyer P, Linn F, Heidmann I, Meyer AH, Niedenhof I & Saedler H (1992) Endogeneous and environmental factors influence 35S promoter methylation of a maize A1 gene construct in transgenic petunia and its colour phenotype. Mol Gen Genet 231: 345–352.

Mozo T & Hooykaas PJJ (1992) Factors affecting the rate of T-DNA transfer from *Agrobacterium tumefaciens* to *Nicotiana glauca* plant cells. Plant Mol Biol 19: 1019–1030.

Murray AW (1992) Creative blocks: cell-cycle checkpoints and feedback controls. Nature 359: 599–604.

Newmeyer DD (1993) The nuclear pore complex and nucleocytoplasmic transport. Curr Opinion Cell Biol 5: 395–407.

Otten L, Canaday J, Gérard JC, Fournier P, Crouzet P & Paulus F (1992) Evolution of agrobacteria and their Ti plasmids - a review. Mol Plant Microb Interact 5: 279–287.

Pang Q, Hays JB & Rajagopal I (1992) A plant cDNA that partially complements *Escherichia coli recA* mutations predicts a polypeptide not strongly homologous to RecA proteins. Proc Natl Acad Sci USA 89: 8073–8077.

Pazour GJ, Ta CN & Das A (1991) Mutants of *Agrobacterium tumefaciens* with elevated *vir* gene expression. Proc Natl Acad Sci USA 88: 6941–6945.

Pazour GJ, Ta C & Das A (1992) Constitutive mutations of *Agrobacterium tumefaciens* transcriptional activator *virG*. J Bacteriol 174: 4169–4174.

Powell BS & Kado CI (1990) Specific binding of VirG to the *vir* box requires a C-terminal domain and exhibits a minimum concentration threshold. Mol Microbiol 4: 2159–2166.

Rauth S, Song K-Y, Ayares D, Wallace L, Moore PD & Kucherlapati R (1986) Transfection and homologous recombination involving single-stranded DNA substrates in mammalian cells and nuclear extracts. Proc Natl Acad Sci USA 83: 5587–5591.

Renckens S, De Greve H, Van Montagu M & Hernalsteens JP (1992) *Petunia* plants escape from negative selection against a transgene by silencing the foreign DNA via methylation. Mol Gen Genet 233: 53–64.

Rodenburg CW, De Groot MJA, Schilperoort RA & Hoyykaas PJJ (1989) Single-stranded DNA used as an efficient new vechicle for plant protoplast transformation. Plant Mol Biol 13: 711–719.

Rogowsky PM, Close TJ, Chimera JA, Shaw JJ & Kado CI (1987) Regulation of the *vir* genes of *Agrobacterium tumefaciens* plasmid pTiC58. J Bacteriol 169: 5101–5112.

Roth DB & Wilson JH (1986) Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. Mol Cell Biol 6: 4295–4304.

Roth DB, Porter TN & Wilson JH (1985) Mechanism of nonhomologous recombination in mammalian cells. Mol Cell Biol 5: 2599–2607.

Roth DB, Chang WB & Wilson JH (1989) Comparison of filler DNA at immune, nonimm-

mune, and oncogenic rearrangements suggests multiple mechanisms of formation. Mol Cell Biol 9: 3049–3057.
Runitz J & Subramani S (1984) The minimum amount of homology required for homologous recombination in mammalian cells. Mol Cell Biol 4: 2253–2258.
Sancar A & Sancar GB (1988) DNA repair enzymes. Annu Rev Biochem 57: 29–67.
Scheid OM, Paszkowski J & Potrykus I (1991) Reversible inactivation of a transgene in *Arabidopsis thaliana*. Mol Gen Genet 228: 104–112.
Seidman MM (1987) Intermolecular homologous recombination between transfected sequences in mammalian cells is primarily nonconservative. Mol Cell Biol 7: 3561–3565.
Selker E (1990) Premeiotic instability of repeated sequences in *Neurospora crassa*. Annu Rev Genet 24: 579–613.
Sen P, Pazour GJ, Anderson D & Das A (1989) Cooperative binding of *Agrobacterium tumefaciens* VirE2 protein to single-stranded DNA. J Bacteriol 171: 2573–2580.
Shimoda N, Toyoda-Yamamoto A, Nagamine J, Usami S, Katayama M, Sakagami Y & Machida Y (1990) Control of expression of *Agrobacterium vir* genes by synergistic actions of phenolic signal molecules and monosaccharides. Proc Natl Acad Sci USA 87: 6684–6688.
Shurvinton CE & Ream W (1991) Stimulation of *Agrobacterium tumefaciens* T-DNA transfer by overdrive depends on a flanking sequence but not on helical position with respect to the border repeat. J Bacteriol 173: 5558–5563.
Shurvinton CE, Hodges L & Ream W (1992) A nuclear localization signal and the C-terminal omega sequence in the *Agrobacterium tumefaciens* VirD2 endonuclease are important for tumor formation. Proc Natl Acad Sci USA 89: 11837–11841.
Simpson RB, O'Hara PJ, Kwok W, Montoya AL, Lichtenstein C, Gordon MP & Nester EW (1982) DNA from the A6S/2 crown gall tumors contain scrambled Ti-plasmid sequences near its junction with the plant DNA. Cell 29: 1005–1014.
Smith GR (1987) Mechanism and control of homologous recombination in *Escherichia coli*. Annu Rev Genet 21: 179–201.
Smith GR (1991) Conjugational recombination in *E. coli*: myths and mechanisms. Cell 64: 19–27.
Spencer PA & Towers GHN (1988) Specificity of signal compounds detected by *Agrobacterium tumefaciens*. Phytochem 27: 2781–2785.
Stachel SE, Timmerman B & Zambryski P (1986) Generation of single-stranded T-DNA molecules during initial stages of T-DNA transfer from *Agrobacterium tumefaciens* to plant cell. Nature 322: 706–712.
Stachel SE, Timmerman B & Zambryski P (1987) Activation of *Agrobacterium tumefaciens vir* gene expression generates multiple single-stranded T-strand molecules from the pTiA6 T-region: requirement for 5′ *vir*D gene products. EMBO J 6: 857–863.
Steck TR & Kado CI (1990) Virulence genes promote conjugative transfer of the Ti plasmid between *Agrobacterium* strains. J Bacteriol 172: 2191–2193.
Stock JB, Ninfa AJ, Stock AM (1989) Protein phosphorylation and regulation of adaptive responses in bacteria. Microbiol Rev 53: 450–490.
Subramani S & Berg P (1983) Homologous and nonhomologous recombination in monkey cells. Mol Cell Biol 3: 1040–1052.
Sugawara N & Haber JE (1992) Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. Mol Cell Biol 12: 563–575.
Szostak JW, Orr-Weaver TL & Rothstein RJ (1983) The double-strand break model for recombination. Cell 33: 25–35.
Takahashi NK, Yamamoto K, Kitamura Y, Luo SQ, Yoshikura H & Kobayashi I (1992) Nonconservative recombination in *Escherichia coli*. Proc Natl Acad Sci USA 89: 5912–5916.
Tinland B, Koukolikova-Nicola Z, Hall MN & Hohn B (1992) The T-DNA-linked VirD2 protein contains two distinct functional nuclear localization signals. Proc Natl Acad Sci USA 89: 7442–7446.
Topping JF, Wei W & Lindsey K (1991) Functional tagging of regulatory elements in the plant genome. Development 112: 1009–1019.
Toro N, Datta A, Carmi OA, Young C, Prusti RK & Nester EW (1989) The *Agrobacterium tumefaciens virC1* gene product binds to overdrive, a T-DNA transfer enhancer. J Bacteriol 171: 6845–6849.
Tovar J & Lichtenstein C (1992) Somatic and meiotic chromosomal recombination between inverted duplications in transgenic tobacco plants. Plant Cell 4: 319–332.
Turk SCHJ, Melchers LS, den Dulk-Ras H, Regensburg-Tuink AJG & Hooykaas PJJ (1991) Environmental conditions differentially affect *vir* gene induction in different *Agrobacterium* strains. Role of the VirA sensor protein. Plant Mol Biol 16: 1051–1059.
Vancanneyt G, Schmidt R, O'Connor-Snachez A, Willmitzer L & Rocha-Sosa M (1990) Construction of an intron-containing marker gene: splicing of the intron in transgenic plants and its use in monitoring early events in *Agrobacterium*-mediated plant transformation. Mol Gen Genet 220: 245–250.
Vogel AM & Das A (1992) Mutational analysis of *Agrobacterium tumefaciens virD2*: Tyrosine 29 is essential for endonuclease activity. J Bacteriol 174: 303–308.
Wake CT, Vernaleone F & Wilson JH (1985) Topological requirements for homologous recombination among DNA molecules transfected into mammalian cells. Mol Cell Biol 5: 2080–2089.
Walden R, Koncz C & Schell J (1990) The use of gene vectors in plant molecular biology. Methods Mol Cell Biol 4: 175–194.
Wallroth M, Gerats AGM, Rogers SG, Fraley RT & Horsch RB (1986) Chromosomal localization of foreign genes in *Petunia hybrida*. Mol Gen Genet 202: 6–15.
Wang K, Stachel SE, Timmerman B, Van Montagu M & Zambryski P (1987) Site-specific nick in the T-DNA border sequence is a result of *Agrobacterium vir* gene expression. Science 235: 587–591.
Ward JE, Dale EM & Binns AN (1991) Activity of the *Agrobacterium* T-DNA transfer machinery is affected by *virB* gene products. Proc Natl Acad Sci USA 88: 9350–9354.
Waters V & Guiney DG (1993) Processes at the nick region link conjugation, T-DNA transfer and rolling circle replication. Mol Microbiol 9: 1123–1130.
Waters VL, Hirata KH, Pansegrau W, Lanka E & Guiney DG (1991) Sequence identity in the nick regions of IncP plasmid transfer origins and T-DNA borders of *Agrobacterium* Ti plasmids. Proc Natl Acad Sci USA 88: 1456–1460.
Whoriskey SK, Nghiem WH, Leong PM, Masson JM & Miller JH (1987) Genetic rearrangements and gene amplification in *Escherichia coli*: DNA sequences at the junctures of amplified gene fusions. Genes & Devel 1: 227–237.
Wilson JH, Berget PB & Pipas JM (1982) Somatic cells efficiently join unrelated DNA segments end-to-end. Mol Cell Biol 2: 1258–1269.
Winans SC (1990) Transcriptional induction of an *Agrobacterium* regulatory gene at tandem promoters by plant-released phenolic compounds, phosphate starvation, and acidic growth media. J Bacteriol 172: 2433–2438.
Winans S (1991) Two way signalling in *Agrobacterium*-plant interactions. Microbiol Rev 56: 12–31.
Winans SC, Ebert PR, Stachel SE, Gordon MP & Nester EW (1986) A gene essential for *Agrobacterium* virulence is homologous to a family of positive regulatory loci. Proc Natl Acad Sci USA 83: 8278–8282.
Yanofsky MF, Porter SG, Young C, Albright LM, Gordon MP & Nester EW (1986) The *virD* operon of *Agrobacterium tumefaciens* encodes a site-specific endonuclease. Cell 47: 471–477.
Zambryski P (1988) Basic processes underlying *Agrobacterium*-mediated DNA transfer to plant cells. Annu Rev Genet 22: 1–30.
Zambryski P (1992) Chronicles from the *Agrobacterium*-plant cell DNA transfer story. Annu Rev Plant Physiol Plant Mol Biol 43: 465–490.
Zambryski P, Joos H, Genetello C, Leemans J, Van Montagu M and Schell J (1983) Ti plasmid vectors for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. EMBO J 2: 2143–2150.
Zambryski P, Tempé J & Schell J (1989) Transfer and function of T-DNA genes from *Agrobacterium* Ti and Ri plasmids in plants. Cell 56: 193–201.