<table>
<thead>
<tr>
<th>Title</th>
<th>The Structure and Function of the Hairy-Root-Inducing Plasmid A4 Virulence Loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Hirayama, Takashi</td>
</tr>
<tr>
<td>Citation</td>
<td>Bulletin of the Institute for Chemical Research, Kyoto University (1991), 68(5-6): 321-367</td>
</tr>
<tr>
<td>Issue Date</td>
<td>1991-02-28</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/2433/77354">http://hdl.handle.net/2433/77354</a></td>
</tr>
<tr>
<td>Type</td>
<td>Departmental Bulletin Paper</td>
</tr>
<tr>
<td>Textversion</td>
<td>publisher</td>
</tr>
</tbody>
</table>
The Structure and Function of the Hairy-Root-Inducing Plasmid A4 Virulence Loci

Takashi Hirayama

Received November 13, 1990

KEY WORDS: Agrobacterium/ Ri plasmid/ virulence region/ plasmid evolution

1. INTRODUCTION

Agrobacterium rhizogenes confers hairy root tumors at wound sites on susceptible dicotyledonous plants upon infection (1). One such strain, A4, harbors three plasmids (hairy root inducing plasmids, pRi) of different sizes, pRiA4a (170 kb), pRiA4b (250 kb), and pRiA4c (420 kb). Neither pRiA4a nor pRiA4b is compatible with pRiA4c, which is a co-integrate between them (2,3). pRiA4b is the principal causal agent for hairy root disease, though the coexistence of pRiA4a seems to enhance hairy root formation (2,4). Defined DNA segments of pRiA4b (T-DNA) are transferred from its host bacterium to a plant cell followed by integration of the T-DNA into plant nuclear genomes (5,6,7,8). The T-DNA encodes oncogenic genes such as tms and rol genes. The tms is responsible for the constitutive synthesis of plant phytohormones. The rol genes are essential for the root formation but their actual function is unknown (9,10,11,12). Besides, the T-DNA also encodes genes that direct the synthesis of agropine, one of the unique amino acid derivatives called opines (13). Other A. rhizogenes strains also carry three similar plasmids, but the opines directed by each strain are not always agropine (14).

Another Agrobacterium species, A. tumefaciens, causes proliferation of crown gall tumors but not hairy roots. The causative agent, a tumor-inducing plasmid (pTi), similarly transfers its T-DNA into plant cells (15,16,17). This T-DNA also directs the synthesis of phytohormones and opines such as octopine, nopaline, or agropine, in transformed plant cells (13,18,19). However, sequence similarity between the T-DNA of pRi and pTi is limited at both ends (25 bp border repeats) of the T-DNA, which are indispensable in cis for the transfer, and at genes directing synthesis of phytohormones (2,14,20,21,22). Other similar sequences between the plasmids lie outside the T-DNA, spanning about 35 kb regions (2). These regions are called virulence (vir) loci, which are essential in trans for the T-DNA transfer (23,24,25). Expression of the vir genes is inducible by plant phenolic compounds such as aceto-
syringone, although some vir genes are transcribed even under non-inducing conditions (23,26,27,28,29,30,31).

The author has engaged in studying the pRiA4b vir loci and overall organization of vir genes. The nucleotide sequence of all of the vir genes (except for virB) and their flanking regions have been determined and the function of respective vir genes elucidated. In this review, the results obtained from these studies are summarized and compared with those of the pTi's, mainly octopine-type pTiA6 and nopaline-type pTiC58, and the phylogenetic relationships between pRiA4b and pTi's are discussed. In addition, the author would like to introduce a current model for the plant inducible transcription of the vir genes and that for the T-DNA transfer.

2. OVERALL ORGANIZATION OF pRiA4b vir GENES

Avirulent mutants of pRiA4b have been constructed by the Tn3-HoHol transposon insertion procedure (32), and it has been found that their insertion sites are localized in a region of about 26 kb (33). This region is highly homologous to the pTiA6 and pTiC58 vir loci where nonpathogenic derivatives of pTi carried mutations (3,23,33,34). Furthermore, using various small subfragments as probes, it has been shown that the gene organization in these regions of the three plasmids is quite similar, though some minor distinctions have been noticed (33). The vir loci of these three plasmids, except for pRiA4b virB and pTiC58 virD have been sequenced (cited below individually), and the similarity of their organizations at the nucleotide sequence level confirmed. Organizations of the vir genes of the three plasmids are shown as transcriptional units or operons in Figure 1 (virA, virB, virC, virD, virE, and virG). The similarity of the organizations suggests that all of the vir genes have evolved from a common ancestral set of vir genes.

Among the coding regions, the most critical difference between pRiA4b and pTi's is the lack of the virE gene in pRiA4b, nevertheless a promoter sequence homologous to that of the pTiA6 virE promoter is present (see below). Another struc-
tural difference between pRiA4b and pTiA6 is seen from the 3'-half of virD2 to the whole virD3 region. In contrast to the coding regions, spacer regions are generally less conserved, and the length between some vir genes and their sequences are quite different. The virC and virG operons are close to each other in both pRiA4b and pTiC58 (33,35), while those of pTiA6 are separated by about 2.5 kb, the sequence of which is derived from an IS66 sequence (36). The spacer between virA and virB is 900 bp longer in pRiA4b than in pTiC58, and this 900 bp intervening region of pRiA4b contains a short sequence homologous to IS66 (37). Also in the upstream region of virA in another octopine-type pTi, pTi15955, but not pTiA6 and pRiA4b, significant similarity with the IS66 sequence has been identified (38). Therefore, the overall organization of the vir genes of pRi and pTi is close to one another, but in spacer regions, sequences themselves have considerably diverged among plasmids. Besides, some differences in spacer sequences are caused by rearrangement through the transposition of insertion sequences after phylogenetic separation.

3. CHARACTERISTICS OF THE vir GENES

3.a. The virD locus

The virD locus is an operon composed of four genes, virD1, virD2, virD3, and virD4, which can code for polypeptides of 16.0 kDa, 48.4 kDa, 72.8 kDa, and 73.9 kDa, respectively. Each of these genes is preceded by a potential ribosome binding site sequence (SD sequence; 39). These gene products have been made visible by using an in vitro coupled transcription-translation system of E. coli (unpublished results). All of the transposon-insertions within the virD locus lead to avirulent phenotypes (33). The avirulent phenotype of virD2 mutant comes from the lack of an activity for recombination between the 25 bp border repeats at the ends of T-DNA (33); this activity is believed to be concerned in the earliest event for transfer of the T-DNA from a bacterium to a plant cell (see below). The active site of VirD2 for T-DNA processing is likely to be in its NH2-terminal half, because its COOH-terminal half is quite different from that of pTiA6 (33), nevertheless the two VirD2 proteins are exchangeable with each other (25, 33, 40). The amino acid sequences of these two different COOH-terminal halves, however, display similar hydropathy profiles (33), and thus the COOH-terminal half of VirD2 is likely to contribute to the protein stability (33). The NH2-terminal half of VirD2 being enough for T-DNA processing has been demonstrated with the 3'-truncated virD2 of pTiA6 (41).

VirD2 has biochemically been characterized with pTiA6: VirD2 is a site-specific endonuclease that introduces a nick between the third and fourth residues on the bottom strand (T-strand) of the 25 bp border repeats (42,43); VirD2 appears to act on relaxed covalently closed circular DNA substrates (44); and VirD2 binds the 5'-end of the nicked T-strand DNA, probably covalently (45,46,47).

In contrast with VirD2, the amino acid sequences of both VirD1 and VirD4 are entirely conserved between pRiA4b and pTiA6 (78% and 77% identical residues for VirD1 and VirD4, respectively). VirD1 has a potential type I DNA binding domain, similar to that found in the CloDF13 MobC 16 kDa protein. With pTiA6,
it has been shown that VirD1 has a DNA topoisomerase activity (44). Therefore, VirD1 is likely to be involved in nicking and subsequent reactions at the border repeats of T-DNA processing in conjunction with VirD2. VirD4 is absolutely required for virulence (33,48), but its function is unknown. Its NH2-terminal portion contains a signal-peptide-like sequence at amino acid residues 1 to 27, suggesting that VirD4 is either a membrane protein or a secretory protein.

VirD3 of pRiA4b is quite different from that of pTiA6 both in size and amino acid sequence. One can speculate that these two VirD3 proteins possess different functions for virulence, or that VirD3 is not related to pathogenicity. Transposon-insertion mutations within pRiA4b virD3 become avirulent (33). However, those results have turned out to be due to the defect of VirD4 by a polar effect, because other mutants which carry a short deletion or insertion within virD3 (nonpolar mutants) are pathogenic, at least on Kalanchoe and carrot (49). As to pTiA6 virD5 mutants, different results have been reported, one fully pathogenic and the other pathogenic at a reduced efficiency (23, Y. Machida, personal communication). These discrepant results may have arisen from differences in the mutation sites. Thus, the function of VirD3, if any, seems to be only accessory.

An open reading frame (temporarily termed virD5) has been found at regions downstream from virD4 on both pRiA4b (completely sequenced) and pTiA6 (partially sequenced) (49,50). This frame of pRiA4b can code for 841 amino acid residues, the sequence of which contains three intramolecular repeats, each consisting of about 120 amino acids. The expression of virD5 is independent of a plant factors (44). Since expression of the pRiA4b and pTiA6 genes required for virulence that have been examined so far are inducible by plant factors (28,29,51), VirD5 seems to be dispensable for pathogenicity. It is unknown why such a conserved sequence which is non-essential for virulence is present within the vir loci.

3.b. The virC locus

The virC locus is present upstream from the virD operon, but in the opposite direction. This operon is composed of two genes, virC1 and virC2, which are separated by only 5 bp. Each gene is preceded by a potential ribosome binding site sequence, and one for virC2 is within the virC1 coding sequence. Their gene products, VirCl (25.6 kDa) and VirC2 (22.1 kDa), have been identified by over-production in E. coli cells (unpublished results). Transposon-insertion mutants within virC have attenuated pathogenicity; hairy root formation occurs at a reduced efficiency and/or only on limited host plants. These results on the structure and mutant phenotype of pRiA4b virC (33) are close to those of pTi's (52, 53). Recently, it has become clear that pTiA6 VirCl binds the 24 bp overdrive sequence adjacent to the right border of the T-DNA (54). This overdrive sequence has been identified as an enhancer sequence for crown gall induction (55,56,57). Therefore, VirCl that has been bound to the overdrive sequence may stimulate the VirD2 binding to the right border sequence. The function of VirC2 is not known at all.
3.c. The virE locus

The virE locus has been identified with pTi mutants which give an attenuated virulence or an avirulent phenotype (23). This operon contains two genes in pTiA6 (58) and three genes in pTiC58 (59). The respective genes have partly conserved sequence related to each other. As to pRiA4b, however, the virE coding sequence is completely missing (49, 60), presumably due to deletion during evolution as discussed below. Since crown gall induction requires VirE function, introduction of the virE operon into A. rhizogenes carrying pRiA4b may enhance hairy root induction. The results of such experiments have revealed no difference in pathogenicity between A. rhizogenes with and without the pTiA6 virE operon, indicating that hairy root induction needs no VirE function (49).

Recently, pTiA6 VirE2 has been shown to be a non-sequence specific single-stranded DNA binding protein, and speculated to protect the T-strand from nuclease digestion during T-strand synthesis and subsequent transfer to plant cells (61, 62, 63). Non-requirement of VirE function for hairy root induction with pRiA4b may be due to the presence of a compensating single-stranded DNA binding protein coded for by another portion of pRiA4b.

Although there is no virE coding sequence on pRiA4b, a promoter that acts acetosyringone-dependently is present downstream from virD5. About a 30 bp sequence including this promoter is close to that of the pTiA6 virE promoter region, and particularly, the inverted hexamer repeats (TG(A/T)AA(C/T)) characteristic of acetosyringone-inducible promoters (25, see below) are completely conserved in both plasmids (49). The reason why only the promoter region has been conserved is unknown. This promoter may be used for downstream unidentified gene(s) whose product(s) benefits the host bacteria upon infection to plant cells.

3.d. The virB locus

The virB locus is the largest among the vir operons, and its size is about 9.5 kb. Since sequence data for only a part of the pRiA4b virB operon is available (unpublished results), characteristics of the virB locus can be described on the basis of complete sequence data of the corresponding region of both pTiA6 and pTiC58 (64, 65). The available partial sequences of pRiA4b virB are highly homologous to those of the corresponding regions of pTiA6 and pTiC58, indicating the entire pRiA4b virB sequence is close to those of pTi's. The virB operon contains eleven genes, virB1 to virB11. Several genes are overexpressed and the gene products with expected molecular sizes are identified. Nine out of 11 gene products (26.0 kDa VirB1, 12.3 kDa VirB2, 11.6 kDa VirB3, 23.3 kDa VirB5, 31.8 kDa VirB6, 5.9 kDa VirB7, 26.4 kDa VirB8, 32.1 kDa VirB9, and 40.6 kDa VirB10) contain hydrophobic spanning regions including a signal-peptide-like sequence, suggesting that the majority of VirB proteins associate with the bacterial cell membrane. The remaining two proteins, VirB4 (87.4 kDa) and VirB11 (38.1 kDa) possess a potential ATP-binding site. Also these two proteins, though containing no obvious hydrophobic regions, appear to associate with the cell envelope and the inner membrane, respectively, on
the basis of cell dissection experiments (65). Thus, the majorities of VirB proteins are considered to constitute a pore-like structure in the membrane through which the T-DNA bound to VirD2 passes to plant cells.

3.e. The virG locus

The virG locus contains only one gene. VirG is composed of 241 amino acid residues (27.3 kDa). Translation initiation of virG occurs at an unusual start codon, TTG (66,67). Considerable numbers of such cases have been known (68). This TTG codon as well as most of other TTG start codons are accompanied by not only an SD sequence but also a T-signal sequence, which is complementary to the TΨ loop of initiator tRNA. Since the mutational change of this TTG to ATG does not affect VirG function in Agrobacterium cells (66), this TTG start codon seems no necessity. Nevertheless, this TTG codon together with the translation initiation signals are all conserved among the virG genes of pRiA4b and pTi's (35,69). Therefore, the TTG start codon might contribute to control the intracellular level of VirG at the translational level.

VirG has been overproduced in E. coli cells, purified to homogeneity, and characterized (67). VirG is bound, from the major groove of DNA helix, to the hexamer sequences characteristic of inducible vir promoters. Thus, VirG is likely to act as a positive transcription factor for the vir genes upon plant signals. This interpretation is consistent with the results of pTiA6 virG mutant analysis that mutations within virG abolish all the plant factor-inducible vir expressions (27).

3.f. The virA locus

In the virA locus there is a single gene, which codes for a polypeptide of 829 amino acids (91.2 kDa) (37). VirA contains two hydrophobic domains at amino acid residues 18–39 and 260–280. The former domain is followed by hydrophilic amino acids, just like the signal sequence of proteins secreting across the cytoplasmic membrane (70). The latter domain also precedes a region involving clusters of basic amino acid residues, similar to the 'stop transfer sequences' found in several transmembrane proteins (71). Therefore, these two regions have been assumed to be membrane-spanning, while their inside and outside regions to be periplasmic and cytoplasmic, respectively. This topology has been demonstrated with VirA of pTiB6 close to pTiA6 (72,73). This membrane topology of VirA coincides with an idea offered from pTiA6 virA mutant analysis (27) that VirA has a role both in sensing plant phenolics and activating VirG at the NH₂-terminal and COOH-terminal halves, respectively (74).

The 5'-truncated VirA derivatives but not the intact VirA itself have been overproduced in E. coli and characterized (37, H. Endoh, personal communication). Those mutant VirA proteins (VirA') are autophosphorylated in the presence of ATP, and the phosphorylated VirA' then transfers its phosphoryl group to VirG in vitro. The autophosphorylation site of VirA has been demonstrated with pTiA6 to be His-474 in the middle portion (75), the amino acid sequence around which is highly conserved among the plasmids. It is interesting that the amino acid sequence of the
VirA COOH-terminal end region (residues 715-829) is homologous to that of the VirG NH₂-terminal half (residues 1 to 118), the central Asp residue (Asp-52) of which is thought to receive the phosphoryl group, on the analogy of other positive regulators (37,74). This possible phosphorylation target domain of VirA may contribute to regulation of signal transduction including the phosphotransfer from VirA His-474 to VirG Asp-52.

Autophosphorylation of VirA presumably mimics the sensing reactions of plant signals, and the 5'-truncated VirA' derivatives seems to behave as if it constitutively received a plant signal. Thus, the VirA NH₂-terminal half usually appears to repress autophosphorylation at His-474.

4. PHYLOGENETIC RELATIONSHIPS AMONG pRiA4b, pTiA6, AND pTiC58

Phylogenetic relationships among the three plasmids can be deduced from the similarity of the predicted Vir protein amino acid sequences which are available for all three plasmids (VirCl, VirC2, VirA, and VirG). The identities of amino acid residues for each Vir protein in various combinations are summarized in Table 1. The mean values of identical residues in the four proteins are 92% for between pRiA4b and pTiC58, and 79% for pTiC58 and pTiA6, and for pTiA6 and pRiA4b. These relationships are equally maintained in the individual Vir proteins. Therefore, these four vir genes, probably including other vir genes too, have evolved as a single set, but not separately, from an ancestral set of vir genes. In addition, these values show that pTiC58 is more akin to pRiA4b than to pTiA6, and that pRiA4b and pTiC58 are almost equally distant from pTiA6. Thus, the pTiA6 vir gene set should have initially separated from a common ancestral vir gene set for the three plasmids. However, the replicator structure of pTiC58 should be closer to that of pTiA6 than to that of pRiA4b, because replicators belonging to a single incompatibility group are generally much more similar to each other than those to different incompatibility groups are (76). These seemingly discrepant relationships could be because the functional domains on pRi/pTi (the vir loci, the replicator, the T-DNA etc) have been shuffled during evolution of these three plasmids.

5. TRANSCRIPTIONAL REGULATION OF vir GENES

Expression of vir genes is tightly regulated at the transcriptional level (23,27,28, 29,30). In vegetatively growing bacteria, only virA and virG are significantly ex-
pressed. When bacteria are exposed to plant cells or phenolic compounds such as acetosyringone (26), the expression of all of the vir genes becomes induced at high levels. It is obvious that the VirA and VirG proteins contribute the inducible expression, from the mutational and structural analysis as described above. These two proteins contain conserved domains found in corresponding components, sensor and regulator, of various two-component regulatory systems such as EmvZ-OmpR, PhoR-PhoB, NtrB-NtrC, FixL-FixJ, and so on (35,37,38,66,69,74,77,78). Although there is no evidence for direct interaction between the VirA membrane protein and plant phenolics, it is believed that signaling changes the conformation of the VirA COOH-terminal cytoplasmic domain, leading to autophosphorylation at His-474 (74). Signal-independent autophosphorylation of VirA' carrying only the cytoplasmic domain (75, H. Endoh, personal communication) is interpreted to be that VirA' is structurally unstable and fluctuates among various tertiary structures including one that should be created in response to plant signals. The phosphorylated VirA transfers its phosphoryl group to VirG, presumably on Asp-52 (74, H. Endoh, personal communication, E.W. Nester, personal communication). The resulting phosphorylated VirG is likely to activate expression of the vir genes including virG itself. Therefore, the plant signals both modulate the VirG activity and increase the VirG concentration in the cell.

The promoter structure for vir have been assigned by identifying mRNA start sites for each vir gene (28,30). There are some minor differences between pRiA4b and pTiA6, but the principal inducible promoters are well conserved between the two plasmids. The characteristics of the promoters for pRiA4b vir genes are the following. The virA gene is attended by two promoters. The upstream one (AI promoter) acts constitutively, and the downstream one (AII promoter) is inducible with a low level of constitutive activity. Upstream from virG there are three promoters. The farthest one (GI) is the major inducible promoter, and the other two (GII and GIII promoters) have characteristics similar to the AI and AII promoters, respectively. The virB, virC, and virD genes are each accompanied by one promoter, and all of them work only under inducing conditions. The virE promoter, which is not attended by the virE-coding sequence, has similar characteristics to the AII and GIII promoters. The −35 and −10 region sequences of the constitutive vir promoters resemble the respective consensus sequence of the E. coli promoters (79), while the −35 and −10 regions, particularly the −35 region, of the inducible vir promoters which have no basal activity, display a low degree of sequence similarity to the respective consensus sequences (see Figure 2). Therefore, Agrobacterium RNA polymerase appears to be unable to interact with the inducible vir promoters by itself.

In the upstream regions of inducible promoters, one to four hexamer sequences with the consensus sequences of 5' TG(A/T)AA(C/T) 3' (vir box), phased at interval of integral multiples of 11 bp. Moreover, the helical phase of these vir boxes is nearly opposite to that of the −35 and −10 regions of the promoter, and the farthest upstream vir box is always preceded by an additional vir box in the inverted orientation (see Figure 2; 28). Footprinting experiments have revealed that these
**Structure and Function of the Hairy-root-inducing Plasmid A4 Virulence Loci**

(A) **virA II**

5'-GTTTCATTTGAAACAAACTGAGTCGACGTCTGTGATTTCAAACCCATTTACAAAGCCTACCGTGCGGCCTAAGCGCCACGGGAGTGGGA-3'

**virB**

5'-GAAAACCGTTTTCGCTTCAAATGAAATCGAAAAGAAGAAAACGAAAATCCTAGAGTAACCGACCCTCCCGATAATCGTGAACATCAGAT-3'

**virC**

5'-TAAGGATTATTTCCTCTATAATTGTTACATTTGCAACTATTCTATAACAACAATGAAATATAGTTCAGATAATTATTTTCTTATTCATG-3'

**virD**

5'-ATAGTTGCAAATGTAACAATTATAGAGGAAATAATCCTTATCTGTTCTTGATTCC.AGTTTTTATAGGCGTAGGTTTTCGTCTGCCCCGA-3'

**irE**

5'-CCCCCCCGCAGGCCCGCCACGAATTGCAGTTGAAACACGATATTCGTTCAACGCATTTCGCTGAGGTGCTAGGCTTCGCGTATTCTTGA-3'

**virG I**

5'-GAATGTTACAAAATTACATTTGTAGCAAAGCTCAGCAATCTTTGTCATCAAGCTGAAACATATTGTTTGCATTTTTGTCATGCACGGCT-3'

(B) A C

vir box 5' TG AA

T T

Fig. 2. Nucleotide sequence of the upstream regions of inducible pRiA4b vir genes (28). A) The sequence are aligned with the —10 region of the promoters. The —35 and —10 regions of the promoters are overlined. The vir box sequence are indicated by underlines and those in the inverted orientation are indicated by double underlines. Asterisks indicate the transcription starting sites. B) The consensus sequence of the vir box.

**vir** boxes including the inverted one are actual recognition signals for VirG (67,80, 81). The contact of VirG with DNA occurs from the major groove of the DNA helix, and the tendency of the VirG protein to bind each **vir** box has a gradient upstream to downstream (67). On the basis of these facts, the binding-cade model has been offered (67,82); two VirG molecules initially bind the inverted **vir** boxes, and then cooperative binding of additional VirG molecules occurs toward the downstream regions regardless of the presence or absence of downstream **vir** box sequences, although the presence of phased **vir** boxes probably stimulates the cooperation and stabilizes the protein-DNA complexes. The cooperative binding at the major groove missing the **vir** box seems to occur through loose contact. Since VirG molecules are bound to the DNA helix from one side including the major grooves of the phased **vir** box sequences, the positions thought to interact with RNA polymerase (83) are free from the contact of VirG. The exact mechanisms by which VirG activates the inducible promoters are unknown, but the cooperative binding of VirG appears to facilitate the functional interaction of the RNA polymerase with the inducible **vir** promoter region through guiding RNA polymerase to the promoter by specific interaction of VirG with the enzyme, by a conformational change of the promoter region, or by a combination of these actions. The phosphorylated and non-phosphorylated VirG proteins are likely to bind the same DNA regions of the **vir** boxes, but the binding efficiency with the former is presumably higher than that with the latter, on the analogy of characteristics of another positive regulator PhoB (84).
6. SCHEME FOR T-DNA TRANSFER

For the T-DNA transfer, the contact between bacterial and plant cells is essential, probably through surface components. While the plant surface components involved are not elucidated at all, four different *Agrobacterium* chromosomal regions (chromosomal virulence loci; *chv*) have been shown to be directly involved in attachment of bacteria to plant cells. The linked *chvA* and *chvB* synthesize and excrete β-1,2-glucan (85,86), the *cel* locus synthesizes cellulose fibrils (87), the *exoC* locus affects cyclic glucan and acidic succinoglycan synthesis (88,89), and the *att* locus influences cell surface proteins (90). All of these genes influence the surface composition of bacterial cells, but it is not known exactly how their products enhance attachment to plant cells. Other soil bacteria associating with plants also carry genes that are functionally and structurally close to the *Agrobacterium* chromosomal virulence loci. These chromosomal genes are constitutively expressed, perhaps reflecting an additional role of surface polysaccharides in mediating general bacteria-plant cell interactions. These surface components may be linked to some of the *virB* gene products as described above.

The T-DNA region is defined as the pRi/pTi segment homologous to sequences present in transformed plant cells. The actual sizes of the T-DNA elements vary in different plasmids. pTiC58 T-DNA is one large contiguous segment of roughly 22 kb (91). pTiA6 contains three adjacent T-DNAs (13 kb TL, 1.5 kb TC, and 7.8 kb TR). The TL element harbors phytohormone synthetic genes, TR contains several opine synthetic genes, and the TC does not specify a phenotype in transformed plant cells. pRiA4b contains two separate T-DNAs, and both TL and TR provide oncogenic functions. Each of these T-DNA elements is transferred either independently or entirely (92).

Various T-DNAs integrated into the plant genomes generally end within or proximal to a 25 bp sequence that flanks the T-DNA region of the plasmids as imperfect repeats. These border repeat sequences of pRi/pTi are highly related to each other (20,21,33). A consensus sequence of the 25 bp repeats is 5' TGGCAG-GATATATT(G/C)NPu(G/T)TGTAA(A/T)Py 3'. Only these 25 bp direct repeats are required in cis for its mobilization to the plant cell. Deletion of the right border region abolishes transformation. However, deletion of the left border repeat has little effect on pathogenicity, and none at all when the T-DNA is on a small plasmid vector (93,94). If the orientation of the right border 25 bp sequence is reversed with respect to its natural orientation, the efficiency of T-DNA transfer is greatly attenuated. Thus, the T-DNA is likely to be transferred in a right to left direction, defined by the orientation of the border repeats.

The 25 bp border repeats are surely essential for T-DNA transfer, but their neighboring DNA regions influence the efficiency of transfer. For instance, the DNA sequence flanking the native right or left pTiC58 border are stimulative or repressive, respectively, for the T-DNA transfer (95). As to pTiA6 carrying the four border repeats, a 24 bp DNA sequence (5' TAApTPypNCTGTPuTNTGTT-TGTTTG 3'), termed overdrive, is adjacent to the right copies of the 25 bp border
repeats of the TL and TR T-DNA elements, and is essential for an efficient transfer of each T-DNA element. Overdrive acts like an enhancer; it can stimulate T-DNA transfer when placed in either orientation, on either side, and at variable distances up to 6 kb. As described above, pTiA6 VirCl has recently been shown to bind the overdrive sequence (54). The overdrive-VirCl complex probably brings a conformational change around the right border repeat, facilitating the T-DNA to be processed for transfer. It is strange that pTiA6 overdrive-like sequences can not be found on nopaline-type pTiC58, but only on pRiA4b and other octopine-type pTi's, nevertheless all pTi/pRi have the virCl gene similar to pTiA6 virCl.

Various T-DNA-associated molecules are known to form in Agrobacterium cells upon vir induction: (1) double-stranded circular T-DNA molecules (T-circles) (33,96,97,98); (2) double-stranded breaks at the T-DNA borders, yielding double-stranded linear T-DNAs (99); (3) site specific single-stranded nicks (border nicks) between the third and fourth base of the bottom strand (T-strand) of the 25 bp border repeats (43,100,101); and (4) free, linear, single-stranded copies of the T-strand (100, 102). However, T-circle and double-stranded linear T-DNA molecules are produced in bacteria cells at a low frequency of $10^{-3}$ to $10^{-5}$ (33,97,98,103), an unlikely frequency for an intermediate in an efficient T-DNA transfer system. Furthermore, T-circles produced by recombination or linear double-stranded T-DNAs produced by cleavage at the borders would result in the loss of the T-DNA region from pRi/pTi. It seems unlikely that the T-DNA transfer process would have evolved to be suicidal. Thus, T-circles and double-stranded linear T-DNAs might represent side products during formation of T-strands (eg, through occasional recombination events stimulated by border nicks). In contrast, T-strands are efficiently produced, roughly one copy per bacterium. Also border nicks are detected at the corresponding frequency (95,101).

As described in an earlier section, mutations located within the first half of the virD operon, encoding VirD1 and VirD2, block the production of both border nicks and T-strands. If relaxed covalently closed circular molecules are used as substrate in vitro, nicking reaction occurs with VirD2 alone. Thus, VirD2 acts as a site specific endonuclease which recognizes and cleaves the lower strand of the 25 bp T-DNA border repeat sequences, and VirD1 converts closed circular DNA substrates to the relaxed form by its helicase activity (44). Border nicks seem to serve as sites for the initiation and termination of T-strand production, and T-strands could be generated either by displacement from the T-DNA region on the plasmid by a helicase activity and/or by replacement strand synthesis using the top strand of the T-DNA as template (Figure 3). These properties of T-strands explain earlier genetic studies for polar T-DNA transfer. Several activities other than the border endonuclease and helicase, including polymerase(s) and repair enzyme(s), are expected to be essential components for the production of T-strand molecules. Since there are no other vir or chromosomal virulence mutants deficient in T-strand production alone, some function necessary to complete the generation of the T-strand molecules might be essential bacterial functions encoded by the Agrobacterium chromosome. Alternatively, VirD1/VirD2 might possess additional activities. Indeed, VirD2 has an ac-
tivity of covalently binding to the 5' end of the T-strand, possibly working as a pilot into plant cells (46,47). Because the T-strand is a linear single-stranded DNA, single-stranded DNA binding proteins might facilitate its synthesis and processing. Such a function is specified by virE2 of both pTiA6 and pTiC58. VirE2 is the most abundant protein produced in vir-induced cells, suggesting a structural rather than an enzymatic role of VirE2 in the T-DNA transfer process. VirE2 binds tightly and cooperatively to DNA in a non-sequence specific manner (61,62,63). Thus, several possible roles can be considered for VirE2, including packaging of the T-strand and protecting the T-strand from nuclease during its transit. However, pTi virE2 mutants are avirulent only on some plant hosts, and pRiA4b naturally carries no virE2 gene, indicating that VirE2 acts only supplementarily or its function can be replaced by a similar protein such as the host single-stranded DNA binding protein.

The T-strand may be packaged into a viral-like particle with VirE2. However, since T-strands are not produced abundantly and T-DNA transfer requires close physical contact between bacteria and the plant cell, the process is not analogous to viral infection. Assuming that Agrobacterium uses evolutionarily conserved mechanisms...
Structure and Function of the Hairy-root-inducing Plasmid A4 Virulence Loci

for T-DNA transfer, the characteristics of T-DNA transfer described above are most similar to DNA transfer between bacterial cells via conjugation. Border nicks are analogous to nicks at the origin of conjugal DNA transfer, the T-strand is comparable to the linear donor single-stranded DNA molecule during bacterial conjugation, and the process of the T-strand synthesis might be similar to replacement strand synthesis intermediates of conjugal donor DNA according to the rolling circle model. Supporting evidence for T-DNA transfer being bacterial conjugation modified to plant cells is that the transfer origin from a conjugal E. coli plasmid (RSF1010) can substitute for the T-DNA borders in directing DNA transfer to plant cells from Agrobacterium (104). This hybrid transfer system requires an intact vir region and the RSF1010 mobilization loci. The RSF1010 oriT and its cognate mobilization proteins are likely to generate a conjugative DNA transfer intermediate which is then transferred to plant cells using the Agrobacterium vir specific transfer machinery.

In contrast to characteristics of T-DNA processing in bacteria, the processing between the entry of T-strands into plant cells and their integration in the plant nuclear genomes are little understood. Integration of single copy of T-DNA is frequent, but multiple copies of the T-DNA are more frequently incorporated into genomes of various dicotyledonous plant species (91,105,106). Some are inserted as a unit of multiple copies, and others disperse to various different chromosomal positions. Since no multimeric T-DNA forms have been observed in vir-induced bacterial cell, it is more likely that they arise in the plant cell or during transfer to the plant cell. Potentially T-DNA tandem repeats are the result of replication, repair, and ligation of the T-DNA during and/or before insertion into plant DNA. Nucleotide sequence analysis of the T-DNA integration sites has revealed that T-DNA insertion is not dependent on a specific target DNA sequence (107). In addition, a variety of complex rearrangements has been found in the vicinity of integration sites (107). On the T-DNA side, the right junction points in transformed cell DNAs fall within or a few bases from the right 25 bp border repeat sequence, and the corresponding left junction points are distributed more widely, over 100 bp internal to and including the left 25 bp border repeat sequence (108). Thus, T-DNA insertions occur more precisely on its right side than its left side, suggesting that T-DNA integration, like the generation of the transferable T-strand copy, is directed by the T-DNA right border. The variety of integrated T-DNAs as well as target DNA rearrangements suggest that T-DNA is integrated through a multistep process involving several different types of recombination, replication, and repair reactions most likely mediated by host plant encoded enzymes. However, the high efficiency of Agrobacterium transformation suggests that additional factors may play a role. Polar T-DNA transfer as well as targeting of the T-DNA to the plant cell nucleus may be mediated by protein(s) linked to the right end of the T-DNA. The efficient T-DNA integration may also be in part due to the single-stranded form of the transferred T-DNA, for general recombination involves invasion of target sequences by single-stranded donor DNA.
ACKNOWLEDGEMENT

The author greatly thanks Drs. M. Takanami and A. Oka for thoughtful discussion and for help in preparing the manuscript. The author is grateful to Dr. T. Aoyama and Mr. H. Endoh for help in experimental work.

REFERENCES

(23) S.E. Stachel and E.W. Nester, EMBO J., 5, 1445 (1986).
(35) B.S. Powell, G.K. Powell, R.O. Morris, P.M. Rogowsky, and C.I. Kado, Mol. Microbiol., 1,
Structure and Function of the Hairy-root-inducing Plasmid A4 Virulence Loci


(73) L.S. Melchers, T.J.G. Regenburg-Tuink, R.B. Burrett, N.J.A. Sedee, R.A. Schilperoort, and
T. Hirayama

(100) S.E. Stachel, B. Timmerman, and P. Zambrski, EMBO J., 6, 857 (1987).
The Papers Published by the Staff Member of the Institute from July 1989 to June 1990

Nuclear Chemistry


Measurement of the Mass of the Electron Neutrino Using Electron Capture in


(338)
Physical Chemistry


Epitaxial Growth Mechanism and Structure of Organic Thin Films,


**Analytical and Inorganic Chemistry**

Steric Effects of Polymethylene Chain Length on Liquid-Liquid Extraction of Copper(II) with Bis(4-Acylpyrazol–5-One) Derivatives, S. Miyazaki, H. Mukai,


Preparation and Spectroscopic Studies of Alkoxide-derived V2O5-GeO2 Solids


Photoelectrochemical Behavior of Iron Oxide Thin Film Electrodes Prepared by


Mechanical Properties of Bone after Implantation of Apatite-Wollastonite Con-


Ferromagnetic Resonance of Fe/Mg Multilayered Films with Artificial Super-


Organic Chemistry


Addition-Elimination Strategy for Asymmetric Induction: a Chiral Sulfoxide


**Polymer Chemistry**


Direct Observation of Polymer Crystals by High-Resolution TEM, M. Tsuji, Kaigai Kobunshikenkyu, 36, 20, (1990), in Japanese.


TEM Study on Oriented Thin Films of Poly(aryl-ether-tether-ketone) [PEEK]

**Biochemistry**

Sequence Analysis by Computer-Introduction to IDEAS, K. Nakai and M.
Kanehisa, In “Information analysis of nucleic acid and proteins” (H Uchida, ed.), Tokyo

Peptide Sequence Library for Protein Expert System, Y. Seto, Y. Ikeuchi and

Utilizing Protein Primary Structure Databases, M. Kanehisa, In “Proteins II”

Computer Analysis of Functional Sites in Proteins and Nucleic Acids, M.

Human Genome Project, M. Kehehisa, Kagaku, 45, 600, (1990), in Japanese.

Fragment Peptide Library for Classification and Functional Prediction of Proteins,

Conformation of Membrane Fusion-Active 20-Residue Peptides with or without
Lipid Bilayers. Implication of a-Helix Formation for Membrane Fusion, S.

Membrane Fusion and Lysis by the Amphiphilic Peptides Induced by Charge
Neutralization: a Model Study of Virus Membrane Fusion Mechanism, M.
Murata, S. Takahashi, T. Kanaseki and S. Ohnishi, Biophysics of the Cell Surface,

Role of Cysteine Residues in Tryptophanase for Monovalent Cation-Induced
Activation, M. Tokushige, N. Tsujimoto, T. Oda, T. Honda, N. Yumoto, S. Ito,

Dynamic Structure and Function of Biological Macromolecules, Y. Hiragi, K.

Temperature-Jump Solution Small-Angle Scattering and Structural Kinetics,
Y. Hiragi, Y. Sano, H. Inoue, T. Ueki, K. Kajiwara and H. Nakatani, Horyako,

Structural Characteristics of Cation K-Carrageenate in Sol and Gel States as
Observed by Small-Angle X-Ray Scattering, K. Kajiwara, H. Urakawa, M.


Purification and Characterization of Thermostable Aspartate Aminotransferase


Reconstitution of Nucleosomes in vitro with a Plasmid Carrying the Long Termi-


Structural Features of Oncogenes, Y. Kawarabayasi, J. Kato and T. Takeya, Protein and nucleic acids; structure and biological information, 29, (1989).