Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of *Agrobacterium*

Annette C. Vergunst*, Miranda C. M. van Lier, Amke den Dulk-Ras, Thomas A. Grosse Stüve, Anette Ouwehand, and Paul J. J. Hooykaas†

Clusius Laboratory, Institute of Biology, Leiden University, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands

Edited by Stanley Falkow, Stanford University, Stanford, CA, and approved December 2, 2004 (received for review August 24, 2004)

Several human pathogens and the plant pathogen *Agrobacterium tumefaciens* use a type IV secretion system for translocation of effector proteins into host cells. How effector proteins are selected for transport is unknown, but a C-terminal transport signal is present in the proteins translocated by the *A. tumefaciens* VirB/D4 type IV secretion system. We characterized this signal in the virulence protein VirF by alanine scanning and further site-directed mutagenesis. The Cre recombinase was used as a reporter to measure the translocation efficiency of Cre–Vir fusions from *A. tumefaciens* to *Arabidopsis*. The data unambiguously showed that positive charge is an essential characteristic of the C-terminal transport signal. We increased the sensitivity of this translocation assay by modifying the Cre-induced readout in host cells from kanamycin resistance to GFP expression. This improvement allowed us to detect translocation of the VirD2 relaxase protein in the absence of transferred DNA, showing that attachment to the transferred DNA is not essential for transport by the VirB/D4 system. We also found another translocated effector protein, namely the VirD5 protein encoded by the tumor-inducing plasmid. According to secondary structure predictions, the C termini of all VirB/D4-translocated proteins identified so far are unstructured; however, they contain a characteristic hydropathic profile. Based on sequence alignments and mutational analysis of VirF, we conclude that the C-terminal transport signal for recruitment and translocation of effector proteins by the *A. tumefaciens* VirB/D4 system is hydrophilic and has a net positive charge with a consensus motif of R-X(7)-R-X-R-X-R-X-X(n)>.

translocation signal | type IV secretion | Cre recombinase reporter assay for translocation | VirF protein | effector protein

Certain bacterial pathogens use specialized secretion systems that span the bacterial envelope to inject effector proteins directly into eukaryotic host cells. One such system, the type IV secretion system (T4SS), is used by *Agrobacterium tumefaciens* for the induction of the plant tumor crown gall in plants and by pathogens such as *Brucella* spp., *Bartonella* spp., *Helicobacter pylori*, and *Legionella pneumophila* to provoke disease in humans and animals (1–4). This versatile family of T4SSs not only transports effector proteins but also includes a large group that is involved in conjugative DNA transfer within and between bacterial species, as well as in interkingdom transfer to plants, yeasts, and fungi (5–8). The translocated substrates have been identified for some bacteria (1, 9, 10), but, in most cases, it is still not known how the bacteria subvert host cells and cause disease.

A. tumefaciens causes crown gall disease on plants by transferring a nucleoprotein complex and several effector proteins by means of its *virB/D4*-encoded T4SS into host cells (11, 12). The VirD2 protein initiates conjugative DNA processing of a region of the tumor-inducing plasmid, the T region, resulting in release of a single-stranded transferred DNA (T-DNA) molecule. The VirD2 relaxase remains covalently associated to the 5′ end of the T-DNA and is thought to act as a pilot to mediate transfer of the complex through the T4SS and into the host cell nucleus. Subsequent expression of the genes located on this T-DNA disturbs the plant's hormonal balance, causing uncontrolled cell division and development of a tumor. The ssDNA-binding protein VirE2 is independently translocated into host cells (13, 14) and is thought to protect the T strand against nucleases (15). In the plant cell, VirD2 and VirE2 together are thought to ensure nuclear targeting of the complex by virtue of their nuclear localization signals (16, 17). Translocation of VirF and VirE3 is necessary for full virulence on some host plants. Although the precise role of these effector proteins in the infection process has not yet been elucidated, VirF is somehow involved in proteolytic degradation of target proteins by the proteasome (18). The VirB/D4 secretion system is composed of the 11 VirB subunits and the inner membrane protein VirD4. The architecture of the secretion complex (19) and the pathway through which the effector substrates pass are becoming clearer. Cascales and Christie (14) recently demonstrated, in an elegant study using a T-DNA immunoprecipitation assay, that the coupling protein VirD4 is the first component of the T4SS to interact with the T-DNA/VirD2 transfer intermediate. The VirE2 protein interacts at the cell poles of the bacterium with VirD4 (20), strongly supporting the model that VirD4 is the cytoplasmic component of the T4SS that sorts not only nucleoprotein complexes but also the effector proteins for translocation and that T4SS are actually committed protein translocation systems (21, 22). Translocation of the effector proteins VirF, VirE2, and VirE3 is mediated by a C-terminal transport signal (13, 20, 23, 24). Sequence comparison suggested that an RPR motif might play a role in recognition by the secretion apparatus; however, it is not known how the effector proteins and the nucleoprotein complex are recognized by the T4SS.

Here, we report a detailed analysis of the VirF translocation signal by using the previously developed Cre recombinase reporter assay for translocation (CRAfT) (13, 23), in which the site-specific recombinase Cre is used as a reporter to detect translocation of Vir proteins into host cells. We also provide evidence for translocation of another effector protein, VirD5, as well as for the relaxases VirD2, and MobA from IncQ plasmid RSF1010 in the absence of their cognate DNA substrate by using an optimized GFP reporter plant line. A consensus motif for the transport signal of translocated proteins of the *A. tumefaciens* T4SS is proposed.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: CRAfT, Cre recombinase reporter assay for translocation; T4SS, type IV secretion system; T-DNA, transferred DNA.

*Present address: Institut National de la Santé et de la Recherche Médicale Unite 431, Unité de Formation et de Recherche Médecine CS83021, Avenue Kennedy, 30908 Nimes Cedex 02, France.

†To whom correspondence should be addressed. E-mail: p.j.j.hooykaas@biology.leidenuniv.nl.

Materials and Methods

Bacterial Strains and Growth Conditions. Bacteria were grown in LC medium (10 g/liter bacto-tryptone/5 g/liter yeast extract/8 g/liter NaCl, pH 7) with appropriate antibiotics. *A. tumefaciens* strain LBA1100 (25), containing octopine type pTiB6 with a complete virulence (*vir*) region but lacking T-DNA and transfer functions, was used for most transport experiments with *A. thaliana* root explants. Selection conditions of the strains and methods for introduction of plasmids are described in detail elsewhere (26). In some experiments, LBA2587, a derivative of LBA1100 with a precise deletion of *virD4* (27), was used.

Plasmid Constructions. To create site-directed mutants of Cre::virFΔ42N (pSDM3155) (13, 26), we used the protocol described by Sawano and Miyawaki (28). The Cre::VirFΔ42N fusion protein was chosen for mutagenesis studies because it was the most efficiently transferred Cre::Vir protein analyzed (13). Further details of the cloning procedures, creation of targeted Cre fusions, sequences of oligonucleotides, plasmid, and DNA sources can be found in *Supporting Materials and Methods* and Tables 2 and 3 which are published as supporting information on the PNAS web site. DNA sequences, encoding the C-terminal 19 aa of VirF, 10 aa of VirF, 50 aa of VirD1, VirD2, VirD3, VirD5, Ysa, Atu6154, and MobA, as well as full-length VirD2, were translationally fused to *cre* in plasmid pSDM3197 (26). All *cre* control and *cre-vir* genes used in this study are expressed from the *A. tumefaciens virF* promoter sequence, and the chimeric proteins contain an N-terminally located simian virus 40 nuclear localization signal sequence to ensure nuclear targeting after Vir-mediated translocation into host cells. All plasmids were introduced into *A. tumefaciens* by electroporation (29). Expression was verified by Western blot analysis as described earlier (23).

Plant Material. Two different transgenic *Arabidopsis thaliana* ecotype C24 lines were used for detection of translocation of chimeric Cre-Vir proteins from *A. tumefaciens*. In line 3043 (13) (Fig. 2, which is published as supporting information on the PNAS web site) Cre-mediated excision results in expression of a neomycin phosphotransferase (*npt II*) gene, leading to kanamycin resistance. In this study, we isolated a second reporter plant line that has GFP expression as a readout to visualize Cre activity. A single-copy transgenic line was selected after *A. tumefaciens*-mediated transformation of root explants with pCB1 (30) and selection for phosphinothricin resistance (31) (Fig. 2). Southern blot analysis confirmed the presence of a single T-DNA insert, and the offspring of homozygous CB1 plants was used in further experiments.

Protein Translocation Experiments. Transformation of *A. thaliana* roots is described elsewhere (23, 31). Briefly, seedlings from *A. thaliana* 3043 or CB1 were grown for 10 days. Roots were collected and precultured for 3 days, followed by a 3-day cocultivation period with *A. tumefaciens*. The 3043 root explants were transferred to fresh shoot induction medium containing 50 mg/liter kanamycin and 100 mg/liter timentin (23) to select for Cre-mediated induction of *npt II* expression. Two Petri dishes, each containing at least 200 root explants (accurately determined after distribution of the explants) were used per strain. The number of kanamycin-resistant calli per root explant was determined 2 weeks after cocultivation. The transfer efficiency of Cre fusion proteins is expressed as percent of the control, Cre::VirFΔ42N, that contains the C-terminal 160 aa of WT VirF fused to Cre, and which was included in each experiment to allow comparison. A single-sample Student *t* test with a hypothetical mean value of 100% was performed for each mutant to determine whether the average translocation efficiency of the mutant differed significantly from WT ($P < 0.05$). The GFP marker in CB1 roots allowed assaying for translocation directly after cocultivation by fluorescence microscopy (Leica MZ FLIII microscope and a Sony 3CCD color video camera).

Results

Delimitation of the C-Terminal Translocation Signal of VirF. Previously, it was established that the *A. tumefaciens* effector proteins VirE2, VirE3, and VirF possess a C-terminal transport signal (13, 23, 24) in which the only apparent common element was an RPR tripeptide. By using the CRAfT assay and *A. thaliana* reporter line 3043, in which Cre-mediated excision of a *lox*-flanked DNA sequence by translocated Cre-Vir fusion proteins from *Agrobacterium* is detected as kanamycin resistance (Fig. 2), the translocation signal was reduced to the C-terminal 37 aa of VirF (13). Here, we extended the analysis of N-terminal deletions of VirF and found that the C-terminal 19 aa were still sufficient to transfer Cre into 3043 host cells with an efficiency of 25.5 ± 5.4% (SE); $n = 6$] of Cre::VirFΔ42N (see *Materials and Methods* for definition of transfer efficiency). We could not, however, detect transfer of a Cre fusion with the C-terminal 10 aa to the kanamycin reporter line (see below for transfer by using a more sensitive GFP reporter line). Importantly, there was no transfer of Cre::VirF19C from *virD4* mutant LBA2587 (27), showing that translocation mediated by this short sequence also still depends on the presence of an intact T4SS.

Truncation of the C terminus was found to have severe effects on transfer efficiency. Removal of the last residue of VirF resulted in an extreme decrease in transfer (by 96%; Table 1). Transfer of a mutant lacking the last 3 aa was not detectable. These data suggest that the last amino acid is critical, or that the position of other residues from the C terminus is critical for efficient signal recognition (see below).

Alanine Scanning and Site-Directed Mutagenesis Reveal the Importance of Positively Charged Residues in the C-Terminal Transport Signal of VirF. Chou–Fasman secondary structure analysis of VirE2, VirE3, and VirF predicts unstructured C termini. To gain more insight into the nature of the C-terminal translocation signal of VirF, we performed alanine scanning, replacing 15 residues in the C terminus of the chimeric protein Cre::VirFΔ42N independently by Ala by a primer mutagenesis approach (28). The data of a large number of independent cocultivation experiments (summarized in Table 1) unambiguously showed that mutation of the Arg residues R187, R195, R197, and R200 into Ala significantly decreased the translocation efficiency of VirF (by 85%, 84%, 60%, and 80%, respectively); these Arg residues are therefore important for efficient recognition by the transfer apparatus. We tested whether R173 and R177 located further from the C terminus were equally important, but mutation of neither of these amino acids resulted in a significant decrease in transfer. Although a single-sample Student *t* test indicated that translocation of the P196A mutant protein differed significantly from WT, the reduction by only 22% indicates that Pro is not a critical residue. Mutation of any of the other residues did not result in a significant change in translocation efficiency as compared with WT.

The data suggest that the four Arg residues in the C-terminal 20 aa of VirF are important for recognition by the VirB/D4 T4SS, although R197 appears to be less important than the other three. Conversion of the last 2 amino acid positions to Ala did not lead to a decrease in transfer, whereas truncation by removal of the C-terminal residue (Leu) dramatically reduced transfer. This finding suggests that the presence of an Arg residue at the −3 position (R200) from the C terminus may play an important role in efficient signal recognition. Given that positively charged Arg residues are an important characteristic of the signal, we determined whether Arg could

Table 1. Mutational analysis of the C-terminal transport signal of VirF

Mutation*	Sequence of amino acids 173–202	$n^{\dagger}$	Efficiency[‡]	*P*
WT VirF	RPIARSIKTAHDDARAELMSADRPRSTRGL	10	100	
Ala mutants				
R173A	**A**PIARSIKTAHDDARAELMSADRPRSTRGL	3	69 ± 21	0.2725
R177A	RPIA**A**SIKTAHDDARAELMSADRPRSTRGL	3	67 ± 25	0.3207
D185A	RPIARSIKTAHD**A**ARAELMSADRPRSTRGL	4	78 ± 9	0.10994
R187A	RPIARSIKTAHDDA**A**AELMSADRPRSTRGL	4	15 ± 3	<0.0001
E189A	RPIARSIKTAHDDARA**A**LMSADRPRSTRGL	3	100 ± 28	0.9895
L190A	RPIARSIKTAHDDARAE**A**MSADRPRSTRGL	2	68 ± 8	0.1524
M191A	RPIARSIKTAHDDARAEL**A**SADRPRSTRGL	3	103 ± 31	0.9271
S192A	RPIARSIKTAHDDARAELM**A**ADRPRSTRGL	2	69 ± 14	0.2722
D194A	RPIARSIKTAHDDARAELMSA**A**RPRSTRGL	3	80 ± 16	0.3559
R195A	RPIARSIKTAHDDARAELMSAD**A**PRSTRGL	7	16 ± 3	<0.0001
P196A	RPIARSIKTAHDDARAELMSADR**A**RSTRGL	3	78 ± 0.1	<0.0001
R197A	RPIARSIKTAHDDARAELMSADRP**A**STRGL	6	40 ± 10	0.0016
S198A	RPIARSIKTAHDDARAELMSADRPR**A**TRGL	3	124 ± 23	0.4101
T199A	RPIARSIKTAHDDARAELMSADRPRS**A**RGL	2	74 ± 3	0.071
R200A	RPIARSIKTAHDDARAELMSADRPRST**A**GL	3	20 ± 11	0.0195
G201A	RPIARSIKTAHDDARAELMSADRPRSTR**A**L	2	80 ± 18	0.4696
L202A	RPIARSIKTAHDDARAELMSADRPRSTRG**A**	2	93 ± 13	0.6748
Arg mutants				
R187D	RPIARSIKTAHDDA**D**AELMSADRPRSTRGL	3	10 ± 2	0.0006
R187K	RPIARSIKTAHDDA**K**AELMSADRPRSTRGL	2	57 ± 13	0.1812
R195D	RPIARSIKTAHDDARAELMSAD**D**PRSTRGL	3	4 ± 2	0.0003
R195K	RPIARSIKTAHDDARAELMSAD**K**PRSTRGL	3	72 ± 24	0.3746
R197D	RPIARSIKTAHDDARAELMSADRP**D**STRGL	4	23 ± 8	0.0025
R197K	RPIARSIKTAHDDARAELMSADRP**K**STRGL	2	65 ± 9	0.1583
R200D	RPIARSIKTAHDDARAELMSADRPRST**D**GL	3	0.7 ± 0.4	<0.0001
R200K	RPIARSIKTAHDDARAELMSADRPRST**K**GL	2	76 ± 13	0.3137
Double Arg mutants				
R195/197A	RPIARSIKTAHDDARAELMSAD**A**P**A**STRGL	3	0.5 ± 0.3	<0.0001
R195/197D	RPIARSIKTAHDDARAELMSAD**D**P**D**STRGL	3	6 ± 4	0.0022
R195/200D	RPIARSIKTAHDDARAELMSAD**D**PRST**D**GL	3	0 ± 0	<0.0001
Other mutants				
P196I	RPIARSIKTAHDDARAELMSADR**I**RSTRGL	3	65 ± 4	0.0121
R195/197N	RPIARSIKTAHDDARAELMSAD**N**P**N**STRGL	3	8 ± 4	0.0022
L190/M191G	RPIARSIKTAHDDARAE**GG**SADRPRSTRGL	3	52 ± 15	0.0875
P196/S198I	RPIARSIKTAHDDARAELMSADR**I**R**I**TRGL	3	35 ± 9	0.0187
C-terminal truncations				
R200 Stop	RPIARSIKTAHDDARAELMSADRPRST*	3	0 ± 0	<0.0001
G201 Stop	RPIARSIKTAHDDARAELMSADRPRSTR*	3	5 ± 3	0.001
L202 Stop	RPIARSIKTAHDDARAELMSADRPRSTRG*	2	4 ± 0.4	0.003

Transport experiments are performed as described in *Materials and Methods.*

*Mutations are indicated by a one-letter code and position followed by the amino acid into which it was modified. Stop, introduction of premature stop codon.

[†]Total number of independent transformation experiments.

[‡]Transfer efficiency is expressed as percentage of kanamycin-resistant calli per root explant of the WT control (mean ± SE). A single-sample two-tailed Student *t* test was performed to determine the probability that transfer of the mutant differs significantly from WT ($P < 0.05$). The transfer efficiency of WT VirF in 10 independent experiments was 0.54, 0.83, 0.76, 0.52, 0.28, 1.08, 0.94, 0.88, 1.14, and 0.81 kanamycin-resistant calli per root explant.

be replaced by another positively charged amino acid, Lys. As shown in Table 1, replacement of the Arg residues at positions R187, R195, R197, and R200 by Lys did not reduce the transfer efficiency significantly, clearly indicating that positive charge rather than the precise structure of the Arg residue itself is important for signal recognition. In agreement with this idea, replacement of an Arg residue by the acidic Asp residue resulted in an even stronger decrease in transfer (especially for R200) than mutation to Ala. Simultaneous mutation of R195 and R200 into Asp resulted in a complete loss of transfer (even when tested by using the sensitive GFP reporter line described below). These findings highlight the strong correlation between a positively charged C terminus and transfer ability.

Development of a Super-Sensitive Reporter Plant Line for CRAfT. To develop a more sensitive and faster assay to detect protein translocation, we isolated a reporter line of *A. thaliana* containing a single-copy insertion of pCB1 (30). CB1 (Fig. 2) allows the use of GFP fluorescence as a readout for Cre-Vir transfer from *A. tumefaciens*. In contrast to reporter line 3043, root explants can be assayed for protein translocation immediately after the 3-day cocultivation period with *A. tumefaciens.* Fig. 1*A* gives an impression of the efficiency with which Cre activity can be visualized. Transformation of CB1 roots with a T-DNA construct containing a *cre* gene behind a strong plant promoter based on the mannopine synthase sequence [pBigMac-*cre* (32)] resulted in high numbers of host cells expressing GFP 3 days

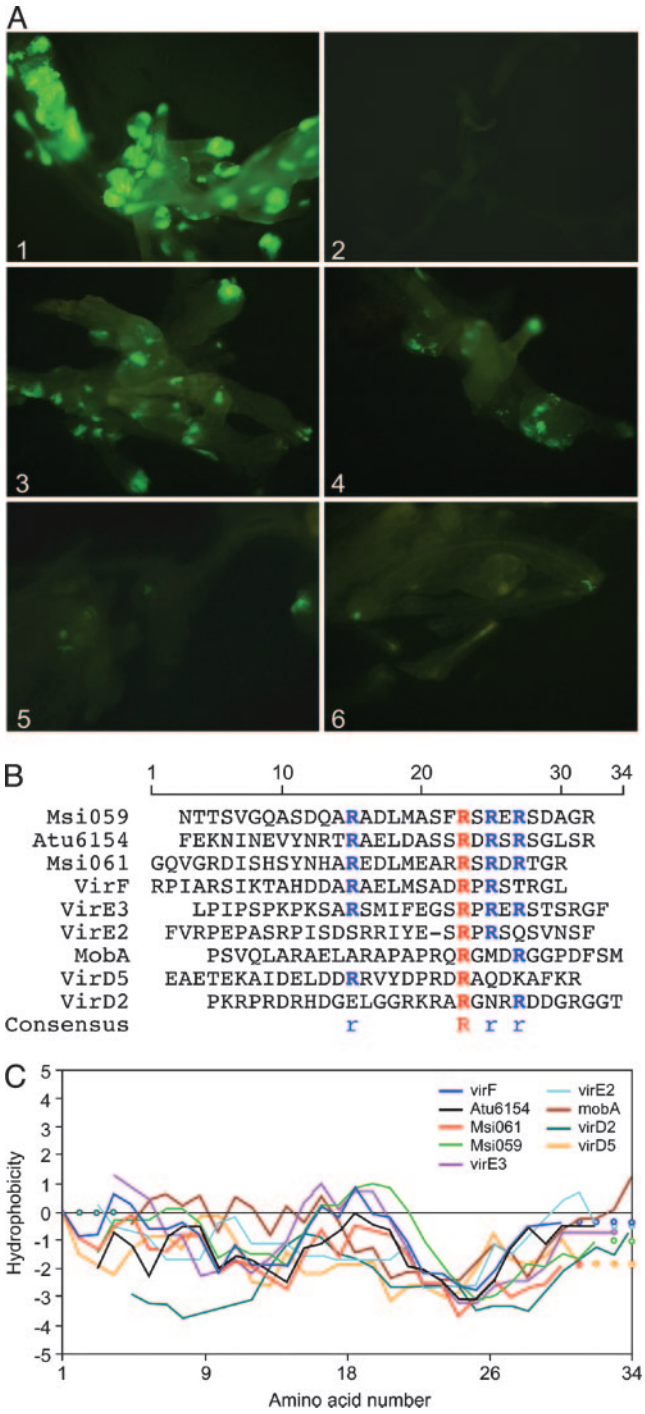

Fig. 1. Visualization of protein translocation into host cells and alignment of transport signals. (*A*) Root explants of *A. thaliana* GFP reporter line CB1 3 days after cocultivation with *A. tumefaciens* (*A1*) containing a T-DNA vector with the *cre* gene behind a strong plant promoter or expressing Cre or Cre–Vir fusions from a bacterial promoter (*A 2–6*). (*A2*). Cre alone. (*A3*) Cre::VirF37C. (*A4*) Cre::VirF19C. (*A5*) Cre::VirF10C. (*A6*) Cre::VirD2. (*B*) Protein alignment [MultAlin, which can be accessed at http://prodes.toulouse.inra.fr/multalin/multalin.html (39)] of the C-terminal 30 aa of *A. tumefaciens* VirF, VirE3, VirE2, VirD5, and VirD2 from pTiB6, and Atu6154 from pTiC58, *M. loti* proteins Msi059 and Msi061 and RSF1010 MobA. (*C*) Kyte–Doolittle hydropathy plot of the alignment in *B*.

after cocultivation. Cocultivation with the negative control strain lacking T-DNA but expressing Cre alone from the bacterial *virF* promoter [LBA1100 (pSDM3197) (26)] rarely (<1 in 5,000 explants) resulted in a GFP fluorescing host cell. In contrast, protein translocation of Cre:VirFΔ42N and Cre::VirF37C resulted in high numbers of GFP fluorescing cells per root explant (see Fig. 1*A*). Transport mediated by the C-terminal 19 aa of VirF was reduced by ≈75% compared with transfer by the VirFΔ42N protein, in agreement with results described above using the kanamycin reporter line 3043. A chimeric protein consisting of the C-terminal 10 aa of VirF fused to Cre was able to induce GFP expression in CB1 host cells, although we had been unable to detect translocation of this chimeric protein into 3043 cells. This finding highlights that this regeneration-independent reporter system can detect remarkably low levels of translocation, and shows that a minimal signal for recognition by the VirB/D4 transport system must be present in this short 10-aa sequence.

VirB/D4-Dependent Translocation of Relaxase Proteins VirD2 and MobA in the Absence of a DNA Substrate. According to a current model, T4SSs are dedicated protein translocation machines that have evolved to transfer also nucleoprotein complexes (4, 21). The relaxase protein that is covalently bound to the 5′ end of the ssDNA molecule is thought to possess the T4SS transport signal and thus act as a pilot for the complex. Therefore, we analyzed translocation of the VirD2 and the MobA relaxase protein of IncQ plasmid RSF1010, which can be mobilized by the VirB/D4 T4SS (25), in the absence of bound DNA. We detected translocation of Cre::VirD2 in one experiment by using reporter plant line 3043 with kanamycin resistance as a readout (nine kanamycin-resistant calli in 480 root explants, 0.8% of Cre::VirFΔ42N), but were unable to reproduce this result. Furthermore, we did not find evidence for translocation of Cre fused to the 50 C-terminal aa of VirD2. However, we were able to reproducibly show translocation of a Cre::VirD2 fusion protein by using the GFP reporter system (Fig. 1*A*) in ≈20–30 cells per 500 explants. During the course of our experiments, Dot/Icm-dependent translocation of the MobA protein from a *Legionella* donor to an *Escherichia coli* recipient was reported (10). Here, we obtained evidence for translocation of Cre fused to the C-terminal 48 aa of MobA from *A. tumefaciens* strain LBA1100 to plant cells, both with reporter line 3043 [2 ± 1.1% (SE); n = 3] of Cre::VirFΔ42N), as well as with line CB1 (data not shown). Translocation of both Cre::MobA48C and Cre::VirD2 was VirB/D4-dependent because transfer was not detected from *virD4* mutant LBA2587. LBA1100 contains neither a T-DNA nor an *IncQ* plasmid, indicating that VirD2 and MobA translocation can occur in the absence of DNA transfer. The presence of an RKRAR and an RQR sequence respectively in the C termini of VirD2 and MobA, and a net positive charge of +4 and +2 in the last 20 aa, respectively, are in agreement with the requirements of the C-terminal T4SS signal as revealed by our analysis of VirF (see above).

VirD5 Is a Transported Effector Protein. Based on the presence of the RxR tripeptide in MobA and VirD2, respectively, and the finding that P196 was not critical for translocation of VirF, we searched the *vir* region of the Ti plasmid for genes encoding proteins with a positively charged C terminus containing an RxR motif, and selected VirD1 (RRR), VirD3 (RLR), VirD5 (RDR), Atu6154 (RDRSR), and Ysa (RDR) (8) as candidates. Protein fusions of the C-terminal 50 aa of these proteins to Cre were expressed in LBA1100, as confirmed by Western blot analysis (data not shown). Root explants of both *A. thaliana* 3043 and CB1 were cocultivated with LBA1100 expressing the respective fusion proteins, but we could not detect translocation of VirD1, VirD3, and Ysa. VirD5 transport, however, was detected to both reporter plant lines [3043: 26.6 ± 0.4% (SE), n = 2 of Cre::VirFΔ42N], which is consistent with earlier observations (33) that VirD5 contains eukaryotic DNA binding motifs and

nuclear localization signal sequences, and may, therefore, play a role in the host cell during the infection process.

In addition, we detected transport of the fusion with the Atu6154 protein. This protein, which is encoded by the nopaline pTiC58 plasmid, is closely related to the VirF protein of the octopine Ti plasmid, but has a different function, because it cannot complement for the absence of VirF (21).

The C Termini of VirB/D4-Translocated Proteins Reveal a Consensus Arg Motif and Show Similarity in Hydropathic Profile. By using CRAfT, the VirD2, VirD5, VirE2, VirE3, VirF, Atu6154, and MobA proteins have been shown to be translocated by the *A. tumefaciens* T4SS directly into host cells. A recent study (27) showed that *Mesorhizobium loti* contains a VirB/D4 T4SS that is implicated in nodulation processes on several leguminous host plants. The CRAfT assay was used to show that two *M. loti* proteins, Msi059 and Msi061, both of which are involved in nodulation, can be translocated by the *A. tumefaciens* T4SS (27). An alignment of the 30 C-terminal aa of these nine translocated proteins (Fig. 1*B*) highlights the importance of Arg residues, and suggests a consensus sequence of R-X(7)-R-X-R-X-R. The incidence of Arg residues among the C-terminal 20 residues is higher than expected for the Arg composition of *A. tumefaciens* proteins (6.64%), and the net charge varies between +1 and +4.

Besides the resemblance in the presence of Arg residues in the aligned proteins, the Kyte–Doolittle hydropathy profiles of the C termini also show similarity (Fig. 1*C*). To analyze the importance of this characteristic profile, we constructed three mutants in which the profile of the VirF C terminus was modified, without changing the net charge; P196I, L190/M191G, and P196/S198I (see Table 1 and Fig. 3, which is published as supporting information on the PNAS web site). Although not significant for L190/M191G, transfer of P196I and P196/S198I mutant proteins was significantly reduced by 35% and 65%, respectively, compared with WT. This result indicates that the presence of the characteristic profile in the C terminus helps to ensure efficient transfer, but that it is not essential for signal recognition.

The hydropathic profiles of the C termini of the other mutant proteins used in this study were not changed compared with WT VirF, except for R195/197A. To find out whether modification of charge and/or profile in this mutant protein caused the dramatic decrease in transfer efficiency, we created the R195/197N mutant protein that has a similar hydropathic profile (see Fig. 3), but a net charge of −2 compared with WT. This mutant protein was strongly reduced in transfer, indicating that charge is more critical than the characteristic hydropathic profile for efficient transfer. Regardless of changing the characteristic profile, the C termini of the profile mutants are still hydrophilic. The lack of detectable translocation of VirD1 and VirD3, selected as candidate effectors based on the C-terminal RxR motif, may be due to the hydrophobic character at their C terminus. The striking resemblance in hydropathic profile between the translocated proteins suggests that the *A. tumefaciens* C-terminal transport signal not only has an Arg-rich consensus sequence but that it is also hydrophilic.

Discussion

A central question in T4SS biology is how this translocation system can transfer both effector proteins and DNA molecules from donor to recipient. In this respect, the *A. tumefaciens* VirB/D4 T4SS is an appealing focus for study because it transfers both DNA molecules into host cells, and, independent of these DNA molecules, the effector proteins VirE2, VirE3, and VirF (13, 26).

Most bacterial protein secretion systems recognize their substrates through a signal in the N terminus. There is a clear consensus in the cleavable N-terminal signal peptide sequences recognized by the Sec-dependent and twin Arg translocation (TAT) systems (34, 35). So far, no consensus has been found to define the N-terminal signal that is present in the effector proteins recognized by the type III secretion system. In contrast, the transport signal of the effectors VirE2, VirE3, and VirF of the *A. tumefaciens* T4SS is located in the C terminus (13, 23, 24).

In this study, we identified VirD5 and Atu6154 encoded by the nopaline type Ti plasmid as effector proteins of the *Agrobacterium* transfer system, and we showed that both proteins also carry a C-terminal transport signal. Preliminary data (A.C.V., T.A.G.S., A.O., and P.J.J.H., unpublished work) show that VirD5 is targeted to the plant cell nucleus and that VirD5 is not essential for tumor formation, but may, like VirF and VirE3, play a role in optimizing the transfer process and thus enlarge host range. Atu6154 is related to the octopine Ti plasmid VirF protein but cannot complement a *virF* mutation (21). Possibly, the nopaline pTi-encoded protein has adopted another function during infection and will be an interesting subject for further studies.

To define the C-terminal translocation signal of the *A. tumefaciens* T4SS, we used a deletion and mutagenesis approach for the effector protein VirF. Alanine scanning provided evidence that Arg residues in the C-terminal part of the protein at positions 187, 195, 197, and 200, but not 173 and 177, are important for translocation. Substitution of any of the other residues within the C-terminal 20 aa with Ala did not result in a significant decrease of transfer. Subsequent site-directed mutagenesis revealed that these important Arg residues could be replaced with Lys, but not Asp, without significantly affecting translocation efficiency. These data show an apparent correlation of positive charge with transport function, and we propose that the VirF signal is likely to interact with a complementary charged domain in the coupling protein VirD4.

Interestingly, in the accompanying article, Nagai *et al.* (36) performed a detailed analysis of the Dot/Icm-translocated *Legionella pneumophila* RalF protein, and showed that a Leu residue at the −3 position in the C terminus is critical for transfer. Similarly, in the *A. tumefaciens* VirF protein, the Arg residue at the −3 position is very important for transfer. In contrast, Simone *et al.* (24) reported that the C-terminal 5 aa of VirE2 are dispensable for transfer. Close inspection of the sequence shows that these 5 aa do not contain a positively charged residue, and that an Arg residue becomes located at the −3 position after removal of the C-terminal 5 aa of VirE2. Replacement of the C-terminal residue L202 in VirF by Ala does not affect translocation, showing that this is not a critical residue. However, removal of L202, resulting in a protein that is truncated by one amino acid led to an almost complete loss of transfer. These findings suggest that critical residues for transfer require at least two additional residues at the C terminus. Based on our findings for the VirF signal and an alignment (Fig. 1 *B* and *C*) of the C termini of the so-far-identified proteins, we extended the RPR sequence predicted to be part of the transport signal based on the effector proteins VirE2, VirE3, and VirF (13) to a consensus R-X(7)-R-X-R-X-R-X-X(n)> sequence that is hydrophilic and has a net positive charge. The maximum value of n has to be determined experimentally because our data do not allow us to draw conclusions about the maximum allowed distance of the motif from the C terminus.

Accumulating evidence suggests that it is the coupling protein that recruits both the protein substrates and the nucleoprotein complex to the T4SS (4, 14, 20, 22, 37, 38), the latter by virtue of the relaxase protein that is covalently bound to the DNA. Our studies show that Cre::VirD2 chimeric proteins are translocated into host cells in the absence of T-DNA, strongly suggesting that the relaxase component VirD2 indeed provides the transport signal for transfer of the nucleoprotein complex. Recently, it was shown that the IncQ plasmid RSF1010 MobA relaxase is translocated from *Legionella* to *E. coli* by the Dot/Icm T4SS (10),

suggesting that MobA, similar to VirD2, carries the transport signal. Here, we show that the MobA protein can also be translocated by the *A. tumefaciens* T4SS into plant cells. Indeed, the MobA protein contains the C-terminal consensus sequence and hydropathic profile present in the *A. tumefaciens* effector proteins. Moreover, as the DNA binding and relaxase functions are not present in these C-terminal 48 aa of MobA, we can now definitely conclude that translocation of relaxase proteins such as MobA and VirD2 from donor to (prokaryotic or eukaryotic) recipient can occur, irrespective of whether they have a covalently bound DNA molecule attached.

Chou–Fasman secondary structure analysis predicts an unstructured C terminus for the translocated A. *tumefaciens* proteins, suggesting a mobile and open structure. This finding is in line with the results described by Nagai *et al.* (36) in the accompanying manuscript that indicate that the translocation signal for recruitment of the RalF effector protein by the *L. pneumophila* Dot/Icm system is disordered, and is thus probably flexible. Besides, a long α-helix upstream of the sequence that contains the transport signal may be required to project the signal for optimal interaction with the coupling protein and the T4SS (36). The minimal information required for recognition of the VirF signal must be present in the C-terminal 10 aa of VirF because those amino acids were able to translocate Cre into host cells; however, the efficiency of translocation was reduced dramatically compared with a 19-aa sequence. Besides the absence of R187, a lack of structural information may be the reason for the inefficient translocation. Similarly, the Cre::VirD2–50C fusion may lack such features that resulted in inefficient and undetectable transport. Small differences in other properties, such as the spacing of the Arg residues, the preference for Arg residues (even though Lys can replace these Arg residues in VirF) or the characteristic hydrophilic profile may influence the efficiency of translocation, and thereby create an organized translocation of the different effector proteins during infection. The fact that not all proteins with the features of the *A. tumefaciens* T4SS transport signal are exported suggests that either other so far unrecognized features may be concealed in the T4SS signal or that additional properties in those proteins may be incompatible with translocation. Interestingly, we were unable to detect translocation of a Cre::GFP::VirF fusion protein from *A. tumefaciens* by using the sensitive GFP reporter line (data not shown). This result is in line with an earlier suggestion that GFP may block translocation at a step after recruitment to VirD4 (20), and suggests that the T4SS may only transport unfolded proteins.

The finding that two *M. loti* proteins that are involved in nodulation can be translocated by the *A. tumefaciens* T4SS system (27) is in line with the close evolutionary relatedness between those species. In contrast, we were unable to detect translocation by the *A. tumefaciens* VirB/D4 system of the *L. pneumophila* RalF protein (A.C.V. and A.d.D.-R., unpublished data). Together, this result indicates that although there are common features in C-terminal transport signals of different T4SS, additional characteristics determine specificity for the cognate T4SS and the VirD4 coupling factor. An intriguing question then is how promiscuous plasmids, such as the IncQ plasmids evolved to be able to hitchhike on different T4SSs. One possibility is that the MobA protein of the incQ plasmid has combined minimal information needed for secretion by different T4SS. Otherwise, it may contain multiple C-terminal signals. Further detailed analysis will have to show which is the case.

Further studies to the interaction of effector molecules with components of the T4SS and the coupling protein will undoubtedly provide detailed insight into the molecular mechanism of protein translocation and give direction to the development of novel antimicrobials against pathogens that use a T4SS for pathogenesis.

We thank Dr. R. Heidstra and Dr. B. Scheres (Utrecht University, Utrecht, The Netherlands) for kindly providing pCB1, E. Jurado Jácome for performing initial VirD2 experiments, H. Wevers for technical assistance, P. Hock for drawing the figures (Leiden University), Dr. C. Roy for sharing unpublished data (Yale University School of Medicine, New Haven, CT), and Dr. H. Shuman (Columbia University Medical Center, New York) for his advice on *Legionella* constructions. We also thank Dr. J. Louwerse and Dr. F. Garcia Rodriguez for helpful discussion and Dr. D. O'Callaghan for helpful comments on the manuscript. This work was supported by European Union-Framework 5, Quality of Life and Management of Living Resources Grant QLRT-2000-01200.

1. Cascales, E. & Christie, P. J. (2003) *Nat. Rev. Microbiol.* **1,** 137–149.
2. Christie, P. J. & Vogel, J. P. (2000) *Trends Microbiol.* **8,** 354–360.
3. Christie, P. J. (2001) *Mol. Microbiol.* **40,** 294–305.
4. Llosa, M. & O'Callaghan, D. (2004) *Mol. Microbiol.* **53,** 1–8.
5. Bundock, P., den Dulk-Ras, A., Beijersbergen, A. & Hooykaas, P. J. J. (1995) *EMBO J.* **14,** 3206–3214.
6. de Groot, M. J., Bundock, P., Hooykaas, P. J. & Beijersbergen, A. G. (1998) *Nat. Biotech.* **16,** 839–842.
7. Waters, V. L. (2001) *Nat. Genet.* **29,** 375–376.
8. Zhu, J., Oger, P. M., Schrammeijer, B., Hooykaas, P. J. J., Farrand, S. K. & Winans, S. C. (2000) *J. Bacteriol.* **182,** 3885–3895.
9. Nagai, H. & Roy, C. R. (2003) *Cell. Microbiol.* **5,** 373–383.
10. Luo, Z.-Q. & Isberg, R. R. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 841–846.
11. Gelvin, S. B. (2000) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51,** 223–256.
12. Lai, A. & Kado, C. I. (2000) *Trends Microbiol.* **8,** 361–369.
13. Vergunst, A. C., Schrammeijer, B., den Dulk Ras, A., de Vlaam, C. M. T., Regensburg-Tuïnk, A. J. G. & Hooykaas, P. J. J. (2000) *Science* **290,** 979–982.
14. Cascales, E. & Christie, P. J. (2004) *Science* **304,** 1170–1173.
15. Abu-Arish, A., Frenkiel-Krispin, D., Fricke, T., Tzfira, T., Citovsky, V., Wolf, S. G. & Elbaum, M. (2004) *J. Biol. Chem.* **279,** 25359–25363.
16. Zupan, J. R., Citovsky, V. & Zambryski, P. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 2392–2397.
17. Ward, D. V., Zupan, J. R. & Zambryski, P. C. (2002) *Trends Plant Sci.* **7,** 1–3.
18. Schrammeijer, B., Risseeuw, E., Pansegrau, W., Regensburg-Tuïnk, A. J. G., Crosby, W. L. & Hooykaas, P. J. J. (2001) *Curr. Biol.* **11,** 258–262.
19. Yeo, H. J. & Waksman, G. (2004) *J. Bacteriol.* **186,** 1919–1926.
20. Atmakuri, K., Ding, Z. & Christie, P. J. (2003) *Mol. Microbiol.* **49,** 1699–1713.
21. Regensburg-Tuïnk, A. J. G. & Hooykaas, P. J. J. (1993) *Nature* **363,** 69–70.
22. Llosa, M., Gomis, R. F., Coll, M. & de la Cruz, F. (2002) *Mol. Microbiol.* **45,** 1–8.
23. Vergunst, A. C., Lier, M. van, den Dulk Ras, A. & Hooykaas, P. J. J. (2003) *Plant Physiol.* **133,** 978–988.
24. Simone, M., McCullen, C. A., Stahl, L. E. & Binns, A. N. (2001) *Mol. Microbiol.* **41,** 1283–1293.
25. Beijersbergen, A., den Dulk Ras, A., Schilperoort, R. A. & Hooykaas, P. J. J. (1992) *Science* **256,** 1324–1327.
26. Schrammeijer, B., den Dulk Ras, A., Vergunst, A. C., Jurado Jácome, E. & Hooykaas, P. J. J. (2003) *Nucleic Acids Res.* **31,** 860–868.
27. Hubber, A., Vergunst, A. C., Sullivan, J. T., Hooykaas, P. J. J. & Ronson, C. W. (2004) *Mol. Microbiol.* **54,** 561–574.
28. Sawano, A. & Miyawaki, A. (2000) *Nucleic Acids Res.* **28,** E78.
29. den Dulk Ras, A. & Hooykaas, P. J. (1995) *Methods Mol. Biol.* **55,** 63–72.
30. Heidstra, R., Welch, D. & Scheres, B. (2004) *Genes Dev.* **18,** 1964–1969.
31. Vergunst, A. C., De Waal, E. C. & Hooykaas, P. J. J. (1998) *Arabidopsis Protocols* (Humana, Totowa, NJ).
32. Vergunst, A. C. & Hooykaas, P. J. J. (1998) *Plant Mol. Biol.* **38,** 393–406.
33. Schrammeijer, B., Beijersbergen, A., Idler, K. B., Melchers, L. S., Thompson, D. V. & Hooykaas, P. J. J. (2000) *J. Exp. Bot.* **51,** 1167–1169.
34. Dilks, K., Rose, R. W., Hartmann, E. & Pohlschröder, M. (2003) *J. Bacteriol.* **185,** 1478–1483.
35. Cao, T. B. & Saier, M. H., Jr. (2003) *Biochim. Biophys. Acta* **1609,** 115–125.
36. Nagai, H., Cambronne, E. D., Kagan, J. C., Amor, J. C., Kahn, R. A. & Roy, C. R. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 826–831.
37. Escudero, J., den Dulk Ras, A., Regensburg-Tuïnk, A. J. G & Hooykaas, P. J. J. (2003) *Mol. Microbiol.* **47,** 891–901.
38. Szpirer, C. Y., Faelen, M. & Couturier, M. (2000) *Mol. Microbiol.* **37,** 1283–1292.
39. Corpet, F. (1988) *Nucleic Acids Res.* **16,** 10881–10890.