VirD2 of Agrobacterium tumefaciens : functional domains and biotechnological applications

Kregten, M. van

VirD2 of *Agrobacterium tumefaciens*: functional domains and biotechnological applications

Maartje van Kregten

VirD2 of *Agrobacterium tumefaciens*: functional domains and biotechnological applications

PROEFSCHRIFT

ter verkrijging van

de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus prof.mr. P.F. van der Heijden

volgens besluit van het College van Promoties

ter verdedigen op donderdag 19 mei 2011

klokke 15.00 uur

door

Maartje van Kregten

geboren te Den Haag

in 1982

Samenstelling promotiecommissie:

Promotor:	Prof. dr. P.J.J. Hooykaas
Co-promotor:	Dr. ir. E.J. van der Zaal
Overige leden:	Dr. ir. A.C. Vergunst (INSERM-Université Montpellier 1)
	Prof. dr. J. Memelink
	Prof. dr. H.P. Spaink
	Prof. dr. C.A.M.J.J. van den Hondel

Printed and bound by Wöhrmann Print Service, Zutphen, the Netherlands

ISBN: 978-90-8570-747-9

Contents

Chapter 1

Introduction

1.1 General introduction

Agrobacterium tumefaciens is a pathogen that genetically transforms its hosts, dicotyledonous plants, causing crown gall disease. Plant wounds not only form the entry points for the bacteria, but also release phenolic compounds like acetosyringone that trigger the virulence genes of the bacterium. Recognition of these compounds by *Agrobacterium* triggers a series of events finally leading to genetic transformation of plant cells. Crucial for this process is the presence of a tumor-inducing plasmid, the Ti-plasmid. During transformation by *Agrobacterium*, a strand of DNA derived from the Ti-plasmid is transported to the recipient cell. This T-(transferred) strand can integrate into the recipient cell genome, after which the genes encoded on it become expressed. The expression of genes of plant hormone synthesizing enzymes that are present on the T-DNA leads to the uncontrolled tumorous growth of plants cells, thus forming the crown gall. Depending on the *Agrobacterium* strain causing the infection, T-DNA encoded genes enable crown gall cells to synthesize specific opines, amino acid derivatives which *Agrobacterium* can use as its sole source of carbon and nitrogen (Zhu *et al.*, 2000). Well-known opines include compounds such as nopaline, octopine, and mannopine. By having food produced by another organism, *Agrobacterium* creates an ecological niche for itself. Although the normal targets of *Agrobacterium*-mediated transformation (AMT) are dicotyledonous plants, transformation of a range of other species has been made possible in a laboratory setting. Hosts include, but are not limited to, monocotyledonous plants (Hooykaas – van Slogteren *et al.*, 1991), yeast (Bundock *et al.*, 1995) and other fungi (de Groot *et al.*, 1998).

Insights into the mechanism by which *Agrobacterium* processes and transfers its T-DNA in the form of a single stranded T-strand to the recipient cells have led to its development as a very important tool in plant biotechnology. The T-DNA is delimited by two imperfect repeating sequences in the Ti plasmid, the Right Border and Left Border (RB and LB) (Yadav *et al.*, 1982). None of the genes located between the repeats is necessary for T-DNA transfer, but transfer is mediated by *vir* genes located in a region of the Ti plasmid that is adjacent to the T-region. Any DNA sequence naturally present between these borders will be translocated to the recipient. In the natural situation, the genes encoding the Vir proteins are located within the *vir* region of the Ti-plasmid, usually adjacent to the T-DNA. While the *vir* region is required for processing and transfer of a T-strand, the *vir* genes

themselves are located outside of the RB and LB sequences and are thus not transferred to recipient cells. In genetic terms, *vir* genes act *in trans* upon the formation and transfer of a T-strand. The invention of the binary vector system, in which the *vir* genes are located on a different plasmid than the T-DNA, opened the way for easy transformation of a range of plant and other species, as any gene and any marker can be transferred in a relatively easy way to the recipient (Hoekema *et al.*, 1983). Genetic transformation of plants via AMT has thus become the most widely used tool in fundamental as well as applied plant research; not only for gaining insight into plant gene functioning, but also for breeding purposes, to generate crops that are, for example, drought- or disease-resistant. In the decades following these discoveries, much effort has been dedicated to unraveling the mechanism behind AMT. More recent findings that *Agrobacterium* is also able to transfer proteins to eukaryotic cells has further inspired the research community, making *Agrobacterium* and especially its *vir* region a paradigm for all kinds of transfer events from bacterial to eukaryotic cells (Vergunst *et* al., 2000, Juhas *et al.*, 2008, Pitzschke and Hirst, 2010).

The molecular mechanism enabling DNA and protein transfer by *Agrobacterium* resembles both bacterial plasmid conjugation and pathogenesis of a variety of bacterial species in its utilization of a Type IV Secretion System (T4SS). This large protein complex serves as a passage between *Agrobacterium* and the recipient. The T-DNA and Vir proteins enter the recipient cell via a T4SS encoded by genes within the *vir* region of the Ti plasmid. Such T4SSs are common in bacteria and quite conserved. The genes and operons encoding them show colinearity, i.e. their order in the genome is similar (reviewed in Schröder and Lanka, 2005). In *Agrobacterium* a second T4SS distinct from the T4SS used in pathogenesis is used to conjugate the pTi between *Agrobacterium* cells. Analysis of its origins shows that the different components have probably been acquired from a different transmissible plasmid, rather than from a duplication of the *vir* region (Alt-Morbe *et al.*, 1996). The complex multi-component structure of the T4SS requires a large investment of energy in order to be assembled. It stands to reason that they are not constitutively expressed, but are expressed when the chance of a successful transfer event is greatest. In *Agrobacterium*, the VirB/D4 T4SS becomes expressed when *Agrobacterium* detects substances from a wounded plant cell in the vicinity (see 1.2), while the T4SS involved in the conjugation of the pTi becomes expressed when opines are present (White and Winans, 2007).

The exact mechanism of T-strand and concomitant protein translocation by *Agrobacterium* has yet to be elucidated. Since similar translocation events are found in a variety of pathogenic bacteria, the biology of T4SSs becomes very relevant for health care.

For instance, several pathogens that target humans, but also other mammals, use their T4SS to translocate virulence factors called effector proteins to their hosts. Well-known pathogens utilizing this mechanism include *Helicobacter pylori* (Backert and Selbach, 2008), *Bartonella* species (Dehio, 2008) and *Legionella pneumophilia* (Swanson and Hammer 2000). In addition, plasmid exchange between bacteria via T4SSs (plasmid conjugation) leads to a major challenge in modern medicine by causing the rapid spreading of antibiotic resistance genes (Hawkey and Jones, 2009). Apart from conjugative plasmids, which encode a T4SS, there are also mobilizable plasmids that do not encode a T4SS, but only have the genes required for DNA processing; they use the T4SS encoded by other plasmids to "hitchhike" into a recipient cell (Francia *et al.*, 2004). T4SSs are still being discovered in ever more distant bacterial species, indicating that they are highly important for the exchange of genetic information between bacteria and for manipulation of their environment.

1.2 The mechanism of *Agrobacterium*-mediated transformation

As briefly mentioned above, when a plant cell is wounded, it releases phenolic compounds such as acetosyringone. These compounds are detected by *Agrobacterium* via the VirA protein, which is a receptor kinase spanning the inner membrane (Figure 1) (Lee *et al.*, 1995). Upon detection of phenolic compounds, VirA phosphorylates VirG, which then binds to the *vir* boxes within the regulatory region of other *vir* genes, thus stimulating their expression (Lee *et al.*, 1996). The newly expressed Vir proteins perform several tasks in the cell. The T-strand is prepared by the nicking reaction of VirD1 and VirD2, and this process is enhanced by binding of VirC1 to the overdrive sequence, adjacent to the RB (see 1.3 for details). VirD2 stays covalently attached to the T-strand. This protein-DNA complex is referred to as the T-complex. VirD2 interacts with VirC1 and the 3 VirD2-Binding Proteins (VBP1-3) to enhance recruitment to the T4SS (see 1.4 and Chapter 3) (Atmakuri *et al.*, 2007, Guo *et al.*, 2007a and b).

Proteins VirB1-11, encoded by the *virB* operon, and VirD4, form the acetosyringone-inducible T4SS of *Agrobacterium* (see 1.5 for details), via which the T-complex and several vir effector proteins are translocated to the cytosol of the recipient cell. In the recipient cell, the T-complex is imported into the nucleus, where it is released from its interactions (see 1.6 for details). Finally, the T-strand can integrate into the genome of the recipient cell and the genes on the T-strand can become expressed (see 1.7 for details).

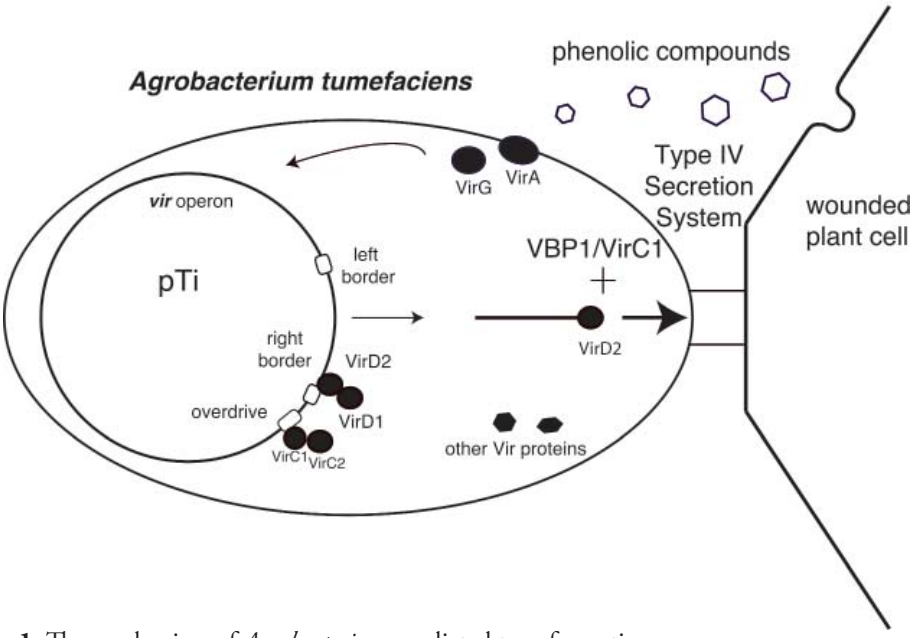

Figure 1: The mechanism of *Agrobacterium*-mediated transformation

1.3 Generation of the T-strand

For the release of the T-strand, a single strand break (nick) is generated at the border sequences that flank the T-DNA. This border nicking is performed by VirD2, a protein belonging to a large family of relaxases which are essential for conjugation and mobilization of plasmids. Relaxases can be placed in several different families and the two families discussed in this thesis can be classified according to their catalytic sites, which contain one or two tyrosine residues. (Frost *et al.*, 2005, Garcillan-Barcia *et al.*, 2009). Examples of relaxases that contain one catalytically active tyrosine residue are MobA of plasmid RSF1010, TraI of plasmid RP4, and VirD2, while examples of relaxases that have two such residues are TrwC of plasmid R388, and TraI of the F plasmid. (Francia *et al.*, 2004). Usually, relaxases contain other well-characterized functional domains C-terminal from their relaxase domain, such as primase or helicase domains (Garcillan-Barcia *et al.*, 2009). VirD2 differs in that respect as it contains a domain of unknown function at its C-terminus (Shurvinton *et al.*, 1992).

At the site of initiation (the RB in the T-region of *Agrobacterium*, *oriT* in conjugative and mobilizable plasmids), the complex of proteins required for nicking of the borders/ *oriT*, called the relaxosome, is assembled. In *Agrobacterium*, the relaxosome consists of the

cofactor VirD1 and the relaxase VirD2, but other systems require different co-factors besides the relaxase (see below). The role of VirD1 in processing of the T-strand is much less clear than that of VirD2. From *in vitro* experiments it is known that, in addition to VirD2, VirD1 is required to nick double-stranded DNA, while VirD2 alone is sufficient to nick single-stranded DNA (Scheiffele *et al.*, 1995).

VirC1, the first of the two proteins encoded by the *virC* operon, binds to the overdrive sequence, a sequence near the RB, but outside of the T-DNA (Toro *et al.*, 1989). This sequence acts as an enhancer of the nicking reaction and of T-strand formation (Peralta *et al.*, 1986, Van Haaren *et al.*, 1987, Toro *et al.*, 1989). VirC2 is also a DNA-binding protein, with a ribbon-helix-ribbon DNA binding domain (Lu *et al.*, 2009). VirC1 and VirC2 are not essential for T-strand processing, but they increase the amount of T-strands produced (Atmakuri *et al.*, 2007).

The only domain of VirD2 important for T-strand processing is the relaxase domain. However, it is not uncommon for relaxases to contain a helicase domain C-terminally from the relaxase domain, which also plays a role in T-strand processing (e.g. relaxase/helicase TrwC of plasmid R338 and relaxase/helicase MobA of plasmid RSF1010 (Garcillan-Barcia *et al.*, 2009)). The presence of the helicase domain can facilitate the separation of the DNA strands prior to making a single strand break or, alternatively, aid in releasing the T-strand once this break has been made. Cofactors with other auxiliary functions are also known; plasmid RP4 has in its relaxosome the cofactors TraJ, to recruit its cognate relaxase/helicase TraI, and TraH and TraK to stabilize the interactions between the relaxase and the DNA. The F plasmid relaxosome requires besides relaxase TraI, the TraY protein to bend the DNA into a conformation that can be more easily processed by the relaxase, and Integration Host Factor to aid in assembly of the relaxosome (reviewed in Byrd and Matson, 1997).

Border cleavage occurs by attack of the catalytically active tyrosine residue (Tyr-29) of VirD2 on the phosphodiesterbond of the *nic* position within the RB. Other relaxases also have a conserved tyrosine residue at a position quite close to the N-terminus performing this action on the *nic* site of their cognate *oriT* (classification of the relaxase family is based on this feature; see Garcillan-Barcia *et al.*, 2009). This reaction creates a covalent bond between the relaxase protein and the DNA (Byrd and Matson, 1997). Relaxase-mediated cleavage of single stranded DNA requires Mg^{2+} ions and water and is isoenergetic; therefore, the reaction is reversible. Ironically, the formation of the covalent protein-DNA bond itself does not actually relax the DNA in the sense that tension due to DNA supercoiling is not decreased; another factor is required to release the tension (Byrd and Matson, 1997).

In all systems but the T-strand transferring mechanism of *Agrobacterium*, a single relaxase-mediated cleavage is enough to create the protein-DNA intermediate. However, to release a perfect pTi-derived T-strand, a second cleavage needs to be performed at the LB. In that manner, only the region in between the RB and LB is transferred as a T-strand to recipient cells, rather than the whole T-DNA containing plasmid. Models for the initiation and termination of T-strand processing are presented in 1.4. Although T-strand transfer is the norm, the whole T-DNA containing plasmid can be transferred when the LB is deleted (Caplan *et al.*, 1985, Miranda *et al.*, 1992). Also when both borders are present, sometimes border skipping is observed, resulting in transfer of the entire plasmid (Van der Graaff *et al.*, 1996, Kononov *et al.*, 1997).

It is nowadays generally believed that uncoupling of the relaxase only occurs after transfer of the relaxase moiety of the T-complex to the recipient cell, although actual experimental evidence is scarce. For the relaxase TrwC from plasmid R388, nicking activity after translocation has been demonstrated in the recipient cell (Draper *et al.*, 2005). In chapter 2, experimental evidence of the translocation of VirD2 as part of the T-complex further confirms this model. These data, demonstrating protein transfer, fit well into the emerging picture of the T4SS as a protein secretion machine, rather than a DNA-transferring machine (Vergunst *et al.*, 2000, Christie and Vogel, 2000, Juhas *et al.*, 2008, Schröder and Lanka, 2005).

1.4 Models for the initiation and termination of T-strand processing

Initiation of T-strand processing has been well described: the catalytic Tyr residue of the relaxase in question creates a covalent bond between the relaxase and the 5' end of the DNA at the *nic* site within oriT or at the RB. This reaction is reversible. (Pansegrau and Lanka, 1996, reviewed in Byrd and Matson, 1997). In the case of plasmids, this one reaction and subsequent processing is all that is required, while for *Agrobacterium*, processing starts at the RB and needs to be terminated at the LB. The exact mechanism of LB cleavage is not yet clarified, although it is clear that the main product of processing in *Agrobacterium* is the T-strand. In the case of *Agrobacterium*, transfer of the entire plasmid can happen when the LB is skipped, which can result in the integration of vector backbone into the genome of the recipient cell (Ramanathan and Veluthambi, 1995, Van der Graaff *et al.*, 1996, Kononov *et al.*, 1997). DNA sequences from the plasmid backbone have been found in 75% of a

population of AMT-generated tobacco plants have (Kononov *et al.*, 1997). Most of these backbone sequences were linked to (part of) the T-DNA, via the LB or RB. Ramanathan and Veluthambi (1995) observed T-DNA integrations originating from the LB of a Ti plasmid, indicating that processing can also accidentally start from the LB. This is also supported by earlier descriptions of processed T-strand intermediates originating from both borders, with protein bound to the 5' end (Dürrenberger *et al.*, 1989, Steck *et al.*, 1989).

After initiation of T-strand processing, the T-strand needs to be separated from the plasmid. There are two possibilities for this process. One model states that, in analogy to rolling circle replication, which occurs in bacteriophages and some plasmids, the T-strand is released by displacement synthesis, mediated by DNA polymerase starting from the 3' OH end of the RB/oriT (reviewed in Lanka and Wilkins, 1995, reviewed in Byrd and Matson, 1997, reviewed in Khan, 2005). Another possibility is that the T-strand is separated from its complementary strand by helicase activity (reviewed in Lanka and Wilkins, 1995, reviewed in Byrd and Matson, 1997). In both cases, the single stranded T-strand must be relatively stable, allowing for its translocation to the recipient cell (reviewed in Byrd and Matson, 1997). Before the relaxase-T-strand bond is broken, the relaxase needs to be translocated, so that release of the T-strand takes place within the recipient cell. This implies that the donor and recipient cell should be in close contact during the processing of the T-strand. In that manner, T-strand processing should be coupled to T-strand translocation, with the relaxase protein at the 5' end piloting the T-strand into the recipient cell as current models suggest (reviewed in Chen *et al.*, 2005 Christie, 2004, Llosa *et al.*, 2002). Evidence has been found for the translocation of TrwC of plasmid R388 (Draper *et al.*, 2005). Additional evidence for the translocation of VirD2, piloting the T-strand, is presented in Chapter 2, while Vergunst and co-workers have shown that low-level translocation of VirD2 and MobA (of plasmid RSF1010) is possible in the absence of a T-strand.

While the formation of the T-strand-relaxase complex takes place in the donor cell, breakage of the relaxase-T-strand bond would then occur in the recipient cell. Efficient bacterial conjugation therefore requires cell-cell contact between donor and recipient (reviewed in Alvarez-Martinez and Christie, 2009). Only for *Agrobacterium*, VirD2-T-strand complexes can be precipitated from *Agrobacterium* cells that are not in contact with recipient cells (but induced to express *vir* genes by administration of acetosyringone) (Cascales and Christie, 2004). After the T-strand is translocated to the recipient cell, a combination of transferred and resident enzymatic activities will determine the final outcome of the DNA transfer. In the case of conjugation, a double-standed circular plasmid will be formed. For

Agrobacterium-mediated DNA transfer, the T-strand finally integrates into the genome of the recipient. In either case, the relaxase is eventually released from the T-strand. Evidence suggests that VirD2 is not responsible for ligating the T-strand to the host genomic DNA (Zieniemowicz *et al*., 2000). Therefore, VirD2 is probably released from the T-strand prior to integration, but the exact mechanism behind this process remains obscure.

1.5 The Type IV Secretion System

After the T-strand is released, it is recruited to the Type IV Secretion System. The T4SS is a multiprotein complex which spans the inner membrane, the periplasm and the outer membrane. In *Agrobacterium,* it extends into the pilus, which protrudes outside of the cell. As already mentioned in section 1.1, the T4SS of *Agrobacterium* that is induced by plant phenolic compounds and which is responsible for the translocation of Vir proteins and the T-strand to the recipient plant cells is composed of the products of the *virB* operon, VirB1-11, in combination with the coupling protein VirD4. *Agrobacterium* also has a second T4SS, encoded by its *tra* operon, which is used for the conjugative transfer of the pTi between *Agrobacterium* cells (Alt-Mörbe *et al*., 1996). For many years the T4SS, such as that encoded by the F plasmid of *E. coli*, was viewed as a channel that must have evolved for translocation of DNA (Frost *et al*., 1994). The coupling protein could serve to "push" DNA through the secretion channel (Llosa *et al*., 2002).

In a seminal paper describing the translocation of the virulence proteins VirE2 and VirF protein of *Agrobacterium* independently of the T-strand, it was proposed that T4SSs might in principle be regarded as protein secretion channels, through which DNA is only transported when bound to the relaxase (Vergunst *et al*., 2000). Evidence for translocation of the relaxases VirD2 and MobA in the absence of T-DNA transfer was later found, as well as translocation of VirE2, VirE3, VirD5 and VirF (Vergunst *et al*, 2005, reviewed in Christie, 2004, Ding *et al*., 2003, Lawley *et al*., 2003).

More recently, T4SS have been identified in pathogenic microorganisms that do not transfer DNA at all, but exclusively mediate the translocation of virulence factors into host cells. Such T4SS are essential for virulence.

T4SS have thus two main functions: conjugation and effector protein translocation. For conjugation to occur, the relaxase-processed DNA intermediate needs to be formed and recognized by the T4SS which then provides it with a passageway into the recipient

cells (Christie, 2004, Llosa *et al.*, 2002). When the T4SS is only used for the translocation of effector proteins, the requirements for the T4SS are approximately the same as for conjugation: a coupling protein to mediate interaction between the effector protein and the T4SS secretion channel and facilitation of transfer (in)to the recipient cell.

A third class of T4SS was more recently discovered in some bacteria, which are involved in the uptake from or release of DNA into the environment. In *Helicobacter pylori*, a DNA uptake system employs a T4SS lacking a coupling protein (Hofreuter *et al.*, 2001). The mechanism of DNA uptake is not yet clear, although it has recently been suggested that the *Helicobacter* homolog of VirB4 energizes DNA transport (Stingl *et al.*, 2010)

In *Neisseria gonorrhea*, DNA is released into the extracellular space via a T4SS (Hamilton *et al.*, 2005). The very complex structure of the T4SS has been the subject of many reviews (e.g. Alvarez-Martinez and Christie, 2009, Gomez-Ruth and Coll, 2006, Schroder and Lanka, 2005). Interestingly, employing the DNA moiety as a detectable target, the order in which the T-strand encounters the structural proteins of the T4SS of *Agrobacterium* has been elucidated by immunoprecipitation of cross-linked Vir protein containing complexes (Cascales and Christie, 2004). Determination of the order of contacts of just the secreted proteins within the T4SS is technically extremely challenging and has not yet been reported.

What is currently known about translocated proteins is that they contain a translocation signal. This T4SS secretion signal has been identified in the C-terminal end of translocated Vir proteins, containing several arginine residues, making the signal positively charged. The relaxase of mobilizable plasmid RSF1010, MobA, also contains such a signal at its C-terminus. The processed relaxase-T-DNA complex of RSF1010 is known to "hitchhike" through the T4SS of *Agrobacterium* (Bravo-Angel *et al.*, 1999). The translocation signal of VirF can even be used to aid the translocation of unrelated proteins that normally cannot pass through the T4SS, such as Cre (Vergunst *et al.*, 2000 and 2005).

1.6 Travel into the nucleus

The T-complex produced by *Agrobacterium*, consisting of VirD2 covalently bound to the T-strand, is delivered into the cytosol of the recipient cell. There it becomes coated by VirE2, which is delivered separately from the T-complex to the recipient cell (Otten *et* al., 1984, Figure 2). It has been postulated that the function of this coating could be to protect the T-strand from degradation by recipient cell nucleases (Citovsky *et al.*, 1989). However, this

has never been shown *in vivo*, only *in vitro* using λ exonuclease (Dürrenberger *et al.*, 1989). The T-complex is then imported into the recipient cell nucleus, due to the presence of a nuclear localization signal (NLS) in VirD2.

In *Arabidopsis thaliana*, nuclear import is mediated by the importin AtKapα, which interacts with VirD2 (Ballas and Citovsky, 1997). Several other interactors have been described for VirD2 in *Arabidopsis*: several cyclophilins (Deng *et al.*, 1998) and the TATA-box Binding Protein (TBP) (Bako *et al.*, 2003). Interaction of VirD2 with cyclophilins is discussed in more detail in Chapter 2. Although the interaction between VirD2 and TBP suggests that a T-complex, when arriving within the nucleus, might by targeted to sites where TBP is present, this is probably not the case. It has been shown that integration sites of T-strands are distributed randomly throughout the genome (Kim *et al.*, 2007). While VirD2 can mediate border fusion *in vitro* (Pansegrau *et al.*, 1996), VirD2 does not appear to ligate the T-strand to the plant genome (Ziemienowicz *et al.*, 2000).

Another factor aiding in nuclear import is VirE2, which interacts with VirE2 Interacting Protein 1 (VIP1). VIP1 is a transcription factor in the *Arabidopsis* defense pathway, and is phosphorylated by MPK3 when plants perceive the presence of *Agrobacterium* or other pathogens. VIP1 then travels to the nucleus, may bind VirE2 and may take the T-complex along. In this way, the plant's own defense mechanism may help bring the T-complex into the nucleus (Djamei *et al.*, 2007). *In vitro*, VIP1 also binds to plant nucleosomes, which may link the T-DNA to the chromatin via a VIP1-VirE2-T-DNA complex (Lacroix *et al.*, 2008).

Before or during integration into the recipient's genome, the T-strand needs to be freed from its binding partners. How VirD2 is released from the T-strand is currently unknown. The F-box protein VirF may play a role in the decoating of VirE2, with the help of the VirE2-interacting protein VIP1. VIP1 brings VirE2 into contact with VirF, which, interacts with the tobacco proteolysis machinery. Through this interaction, both VIP1 and VirE2 are destabilized. Thus this mechanism may help to uncoat the T-DNA (Tzfira *et al.*, 2004). Interestingly, VirF is a host range factor: requirement for this factor is not absolute, but it is needed to induce larger tumors on certain plant species. For instance, VirF is not needed for the transformation of *Saccharomyces cerevisiae* (Bundock *et al.*, 1995). Also, *Arabidopsis* was found to contain an F-box protein that can functionally replace VirF (Zaltsman *et al.*, 2010).

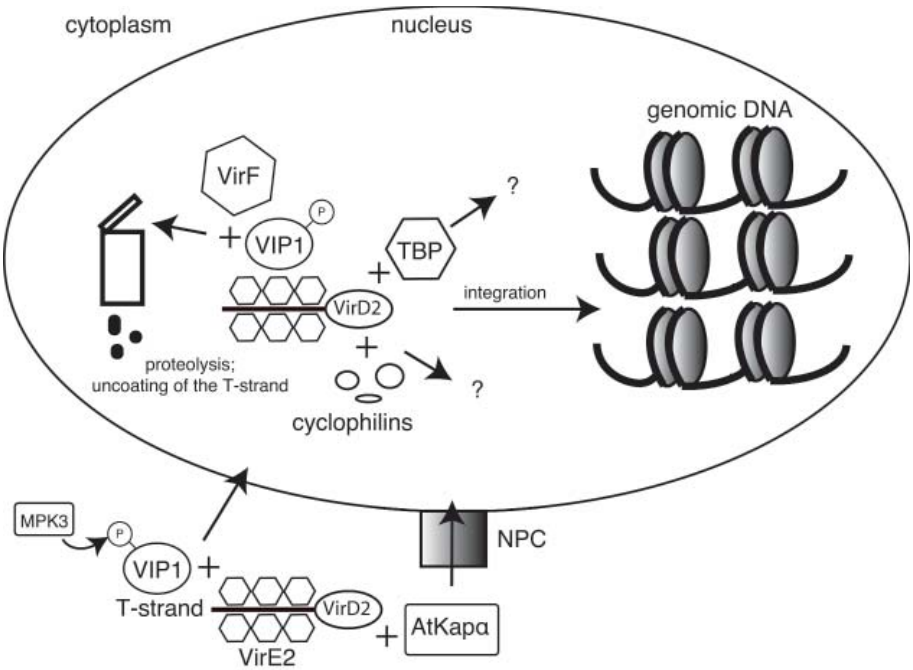

Figure 2: Travel of the VirD2-T-strand complex into the nucleus

1.7 T-DNA integration

In the case of AMT of higher plants, T-DNA insertions have been found to be randomly dispersed over the genome of the recipient (Kim *et al.*, 2007). Although transferred as a single strand, evidence has been presented that indicates that T-DNA can become double-stranded before integration (Tzfira *et al.*, 2003). However, other integration events that were analyzed in detail can only be explained by assuming that the T-DNA was single-stranded at the time of integration (Tzfira *et al.*, 2004).

The main mechanism via which any exogenous DNA integrates in higher eukaryotes is non-homologous recombination (NHR). This consists of two pathways: non-homologous end-joining (NHEJ) and microhomology-mediated recombination (MMEJ). NHEJ is the main pathway of T-DNA integration, and often results in small deletions or insertions of both the T-DNA and the genomic locus where it is inserted. During NHEJ, both ends of the break are processed and ligated. When integrated via NHEJ, the T-DNA is also subject to such processing (reviewed in Bleuyard *et al.*, 2006). In yeast, it was found that enzymes from the NHEJ pathway were essential for the integration of T-DNA lacking homology to the

yeast genome (Van Attikum *et al.*, 2001). When integration via NHEJ occurs, often random 'filler' DNA is added to the site of integration. Filler DNA may introduce microhomology, enabling integration via MMEJ. Filler DNA is also associated with preservation of the T-DNA ends (Windels *et al.*, 2003).

In contrast to the error-prone NHR events, homologous recombination (HR)-mediated events can lead to very precise site-specific alteration of the genome by exchanging the genetic locus or gene of interest, or a part thereof, for a modified version that is introduced into the cell, a process also called gene targeting (GT). Homologous recombination can only occur when the incoming DNA has homology to the genome of the recipient. The break is processed to produce a single stranded overhang, which invades the homologous region in another stretch of DNA. This region is used to bridge the gap and serve as a template for repair. Eventually, this creates a Holliday structure of four strands of DNA. When this structure has been resolved, the remaining nick in the DNA is repaired (Bleuyard *et al.*, 2006).

When integration events occur via NHEJ, such an exchange does not take place; the gene of interest will integrate elsewhere and thus co-exist with the original locus. Moreover, sequences at random positions in the genome will be disturbed. Obviously, HR is the desired pathway for accomplishing site-directed mutagenesis. The field of GT is of high interest for both plant and medical science; fundamentally, to enable the study of the function of genes in their context, and applied, for the development of gene therapy. Much effort is devoted to develop an efficient way to achieve GT in a variety of species.

The chance of integration at a desired position is increased by the presence of a DSB at this site in the genome. This was first established in mammalian cells (Jasin *et al.*, 1985). In plants, this was shown by Puchta *et al.*, who showed in protoplasts that expression of homing endonuclease I-SceI (see below) will increase the chance of homologous recombination (HR) of a substrate carrying an I-SceI recognition site (Puchta *et al.*, 1993). Integration of the T-DNA into the DSB site via NHR also happens (Salomon and Puchta, 1998).

1.8 Tools for targeted integration: zinc fingers nucleases and homing endonucleases

As mentioned above, it has been known for a rather long time that the presence of DSBs in the genome is stimulatory to GT in higher eukaryotes (Jasin *et al.*, 1985). To create DSBs in a genome, different DNA binding proteins have been designed and discovered. The two

most used enzymes for the creation of a DSB are zinc finger nucleases (ZFNs) and homing endonucleases (HEs).

Zinc fingers (ZFs) are DNA-binding motifs; a specific sequence of 30 amino acids can be predicted to bind a certain three basepair sequence (triplet). The "zinc finger code" has been described for many of the GNN, ANN, and CNN triplets (Segal *et al.*, 1999, Dreier *et al.*, 2001, Dreier *et al.*, 2005). For TNN triplets, the available code is still limited. By combining different ZFs, polydactyl ZF domains (PZFs) can be made, which can target a longer, specific sequence of DNA. These PZFs can in principle be combined with other functional groups to create novel functional proteins, e.g. with activator or repressor domains, to create artificial transcription factors, or with a nuclease domain, to create a ZFN.

To create a ZF nuclease, which has DNA cleaving activity, a nuclease domain thus needs to be fused to the PZF. The nuclease domain of choice is the nuclease domain of the restriction enzyme *Fok*I. In *Fok*I, the DNA binding and cleaving activities are located in separate domains of the protein. Therefore, the *Fok*I nuclease domain can be made available for fusion to ZFs, creating a novel restriction enzyme (the zinc finger nuclease, ZFN) with a custom-made recognition sequence, which was first achieved by Kim *et al.*, in 1996. The ZFNs have been eagerly used by groups from a variety of research fields to create genomic lesions in a large number of organisms (reviewed in Carroll, 2008) to study DNA repair mechanisms or explore the possibilities for therapeutic genetic manipulation. With ZFN technology on the rise, hopes are high for the development of novel therapeutic gene targeting strategies for human disease and crop improvement in plants, as well as fundamental scientific questions in the area of DNA repair.

Apart from zinc fingers, homing endonucleases (HEs) can also be used for the introduction of highly specific DSBs in a genome (e.g. Chiurazzi *et al.*, 1996, Pacher *et al.*, 2007). HEs have been discovered in different organisms. Some are encoded by open reading frames present in introns of otherwise non-related genes, while others, designated inteins, are intervening sequences which are spliced out post-translationally (Stoddard, 2005). They function as selfish genetic elements; there is no obvious benefit for the host organism. The HE copies the homing endonuclease ORF into an allele that does not contain it, or to unrelated sites that happen to contain the recognition site. The HE recognizes a long (14-40 bp) DNA sequence, which is present in the homologous allele that does not contain the coding sequence for the HE. Sequence recognition is somewhat relaxed, indicating that the homing endonuclease will not be lost when small mutations in the homologous allele will

occur (Jurica and Stoddard, 1999).

Five families of homing endonucleases have been identified: the LAGLIDADG, HNH, His-Cys box, and GIY-YIG families, and a largely uncharacterized family in the introns of cyanobacteria (Jurica and Stoddard, 1999). The LAGLIDADG class is most extensively characterized, and consists of ORFs in introns mainly found in the mitochondria and chloroplasts of single-celled eukaryotes. The well-studied homing endonucleases I-SceI, I-CreI and HO belong to this group. Both I-SceI and HO contain two LAGLIDADG motifs, indicating that they function as monomers, while homing endonucleases that contain only one LAGLIDADG motif function act as homodimers. I-CreI is an example of the latter class (Stoddard, 2006, Chevalier *et al.*, 2004).

Homing endonucleases of the LAGLIDADG family create DSBs with 4 nucleotide 3'-OH overhangs. The ends are processed and the homologous allele is used as a template for HR (Jurica and Stoddard, 1999). The homing endonuclease is only involved in the initial creation of the DSB and not in the subsequent repair.

I-SceI was the first homing endonuclease to be identified. It was discovered that the yeast mitochondrial marker ω was inherited in a non-Mendelian way, with ω^- converting into ω^+ at a very high rate (reviewed in Dujon, 1989). Only later it was found that ω was an optional intron in the *LSU* mitochondrial gene and the non-Mendelian inheritance pattern was caused by the activity of I-SceI, encoded by an ORF in the optional intron, causing the optional intron to be copied into the other allele (Dujon, 1980).

The yeast mating-type switching enzyme HO is an example of a domesticated homing endonuclease; it is encoded by a free-standing gene, and contains only remnants of the sequences required for intein splicing. Exactly how HO became domesticated is unknown. HO is responsible for changing the mating type of a yeast cell from **a** to α or vice versa. Inactive loci for **a** or α flank the locus for the active mating type (MAT locus). HO creates a DSB whereupon either one of the inactive loci are copied into the active locus (Pietrokovski, 1994, Strathern *et al*, 1982, Kostriken *et al.*, 1983). This ensures that there is a heterogeneous population, capable of mating to create diploid cells.

Although the application of ZFNs for site-directed mutagenesis can be accompanied with problems (of a mostly technical nature, rather than fundamental) (Durai et al., 2005), recent progress in this field is very impressive (reviewed in Wu *et al.*, 2009). Therefore it can be expected that problems such as to how to design the optimal PZF domains for a particular genetic locus without any undesired effects at other loci can in principle be solved. Nevertheless, alternatives to PZF-based technology will be most welcome. The development

of HEs with modified recognition sequences is now coming of age, which will enable the application of homing endonucleases at a broader scale in the future (Pâques and Duchateau, 2007, Galetto *et al*., 2009).

1.9 Outline of this thesis

The focus of the research described in this thesis is on VirD2. This protein is at the center of AMT as it is not only essential for T-strand processing, but also accompanies the T-strand to its final destination in the recipient genome. At the start of the research leading to this thesis, it was already known that only the N-terminal half of VirD2 is required for T-strand processing. The function of the C-terminal half was basically unknown; only the very C-terminal end had been implied as important for translocation and/or integration of the T-DNA. Chapter 2 describes experiments in which we defined a minimal version of VirD2, stripped of all but the relaxase domain. When supplied with a T4SS translocation signal derived from the virulence protein VirF, this minimal VirD2 restores virulence of a VirD2 mutant *Agrobacterium* strain. This indicates that the relaxase domain is the only domain absolutely required for virulence. The other domains of VirD2, the domain of unknown function (DUF) and the C-terminus, were dispensable and could be replaced by the VirF translocation signal. DUF and the C-terminus are possibly involved in the localization of the T-complex to the T4SS, a function that apparently can be taken over by the very small T4SS translocation signal of VirF (F). It also gives support for the prevailing model of the T4SS as a protein – rather than a DNA - translocation machine. In Chapter 3 we describe studies to elucidate the exact role of DUF in localization. We determined, by tagging GFP to VirD2 truncations, that the N-terminal 60 amino acids of DUF, DUF-60, are crucial for polar localization of VirD2. In VirD2-204, fluorescent signal is spread throughout the cell, while in VirD2-264 polar localization is restored to near wild type levels. The translocation signal of VirF, which consists of only 37 amino acids, also acts by restoring polar localization to VirD2 truncations. The phenotype of polar localization correlates with the ability of the corresponding *Agrobacterium* strain to transform plants.

Protein translocation from *Agrobacterium* to any recipient is of interest both for fundamental research and for biotechnological applications. By creating fusions to Cre and demonstrating Cre recombinase activity in the recipient, Vergunst *et al.* proved that several Vir proteins were translocated. Moreover, it was shown that these Vir proteins contained

C-terminal translocation signals, as these protein parts could lead to translocation of Cre fusion proteins to recipient cells via the T4SS of *Agrobacterium* (Vergunst *et al*., 2000 and 2005). Our finding that a T4SS translocation-deficient VirD2 mutant could be made T4SS translocation-competent again by addition of the translocation signal of VirF (designated F) led us to develop VirD2-fusion proteins that can actually be translocated to plants. While N-terminal fusion of protein domains to VirD2 easily lead to impairment of the relaxase function, C-terminal fusions are likely to block the native T4SS translocation signal of VirD2. However, by placing the fusion protein in between VirD2 and F, the relaxase function remained intact and translocation ability was restored. Detection of VirD2-mediated translocation events is easily done by assessing the expression of marker genes encoded on the T-DNA. Furthermore, VirD2 is covalently attached to a T-strand, which results in a possibility to combine the delivery of DSB-inducing protein domains with the concomitant delivery of a DNA molecule that could potentially be used as repair template.

In Chapter 4, we describe the translocation efficiency of several VirD2 fusion proteins. The proteins studied are of great interest for the development of site-directed mutagenesis methods, as they contained zinc fingers or the homing endonucleases I-SceI and HO. All tested protein domains could be translocated, albeit with very different efficiencies. Especially the I-SceI domain was found to be translocated at a relatively high level in a translational fusion with VirD2 and F. Recovery of I-SceI induced damage after *in planta* expression showed that fusion with VirD2 and F did not abrogate the nuclease activity of the fusion protein. We can therefore conclude that protein domain delivery into plant cells via *Agrobacterium*-mediated transformation using VirD2 fusion proteins is feasible.

In Chapter 5, we show that the recombinant relaxase/effector proteins VirD2-I-SceI-F and VirD2-204-I-SceI-F retain cleavage activity on an I-SceI site after delivery to plant cells by *Agrobacterium*-mediated transformation. Several I-SceI-induced mutations were recovered from *Arabidopsis* transformants treated with such recombinant relaxase effector protein, but no GT events were discovered.

The results presented highlight the biotechnological possibilities of functional protein domain delivery via VirD2, and present a lead for further investigation into the development of site-directed mutagenesis strategies in plants via *Agrobacterium*-mediated transformation.

1.10 References

Alt-Morbe, J., Stryker, J. L., Fuqua, C., Li, P. L., Farrand, S. K., and Winans, S. C. 1996. The conjugal transfer system of *Agrobacterium tumefaciens* octopine-type Ti plasmids is closely related to the transfer system of an IncP plasmid and distantly related to Ti plasmid vir genes. J. Bacteriol. 178(14): 4248-57.

Alvarex-Martinez, C.E., and Christie, P.J. 2009. Biological diversity of prokaryotic Type IV Secretion Systems, Microbiol. and Mol. Biol. Rev. 73(4): 775-808.

Atmakuri, K., Cascales, E., Burton, O. T., Banta, L., and Christie, P. J. 2007. *Agrobacterium* ParA/MinD-like VirC1 spatially coordinates early conjugative DNA transfer reactions. EMBO (Eur. Mol. Biol. Organ.) J. 26(10): 2540-2551.

Backert, S., and Selbach, M. 2008. Role of type IV secretion in *Helicobacter pylori* pathogenesis. Cell. Microbiol. 10(8): 1573-1581.

Bako, L., Umeda, M., Tiburcio, A.F., Schell, J., and Koncz, C. 2003. The VirD2 protein of *Agrobacterium*-transferred DNA interacts with the TATA box binding–protein and a nuclear protein kinase in plants. Proc. Natl. Acad. Sci. U.S.A. 100(17): 10108-10113.

Ballas, N, and Citovsky, V. 1997. Nuclear localization binding protein from *Arabidopsis* mediates nuclear import of *Agrobacterium* VirD2 protein. Proc. Acad. Natl. Sci. U.S.A. 94(20): 10723-10728.

Bleuyard, J.Y., Gallego, M.E., and White, C.I. 2006. Recent advances in understanding of double-strand break repair machinery of plants. DNA repair. 5: 1-12.

Bravo-Angel, A. M., Gloeckler, V., Hohn, B., and Tinland, B. 1999. Bacterial conjugation protein MobA mediates integration of complex DNA structures into plant cells. J. Bacteriol. 181(18): 5758-5765.

Bundock, P., den Dulk-Ras, A., Beijersbergen, A., and Hooykaas, P. J. 1995. Trans-kingdom T-DNA transfer from *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*. EMBO (Eur. Mol. Biol. Organ.) J. 14(13): 3206-14.

Byrd, D.R., and Matson, S.W. 1997. Nicking by transfesterification: the reaction catalyzed by a relaxase. Mol. Microbiol. 25(6): 1011-1022.

Caplan, A.B., Van Montagu, M., and Schell, J. 1985. Genetic analysis of integration mediated by single T-DNA borders. J. Bacteriol. 161(2): 655-664.

Carroll, D. 1008. Zinc finger nucleases as gene therapy agents. Gene Ther. 15(22): 1463-1468.

Cascales, E., and Christie, P.J. 2004. Definition of a bacterial type IV secretion pathway for a DNA substrate. Science. 304(5674): 1170-1173.

Chen, I., Christie, P.J., and Dubnau, D. 2005. The ins and outs of DNA transfer in bacteria. Science. 310(5753):1456-1460.

Chevalier, B, Sussman, D., Otis, C., Noel, A. J., Turmel, M., Lemieux, C., Stephens, K., Monnat,

R. J., Jr., and Stoddard, B. L. 2004. Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. Biochemistry. 43(44): 14015-14026.

Chiurazzi, M, Ray, A., Viret, J.F., Perera, R., Wang, X.H., Lloyd, A.M., and Signer, E.R. 1996. Enhancement of somatic introchromosomal homologous recombination in Arabidopsis by the HO endonuclease. The Plant Cell. 8: 2057-2066.

Christie, P.J. 2004. Type IV secretion: the *Agrobacterium* VirB/D4 and related conjugation systems. Biochem. Biophys. Acta. 1693: 219-234.

Christie, P.J., and Vogel, J.P. 2000. Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. Trends in Microbiol. 8(8): 354-360.

Christie, P.J., Ward, J.E., Winans, S.C, and Nester, E.W. 1998. The *Agrobacterium tumefaciens virE2* gene product is a single-stranded-DNA-binding protein that associates with T-DNA. J. Bacteriol. 170(6): 2659-2667.

Citovsky, V., Wong, M. L., and Zambryski, P. 1989. Cooperative interaction of *Agrobacterium* VirE2 protein with single-stranded DNA: implications for the T-DNA transfer process. Proc. Natl. Acad. Sci. U.S.A. 86: 1193-1197.

Dehio, C. 2008. Infection-associated type IV secretion systems of *Bartonella* and their diverse roles in host cell interaction. Cell. Microbiol. 10(8):1591-1598.

Deng, Chen, L., Peng, W. T., Liang, X., Sekiguchi, S., Gordon, M. P., Comai, L., and Nester, E. W. 1999. VirE1 is a specific molecular chaperone for the exported single-stranded-DNA-binding protein VirE2 in *Agrobacterium*. Mol. Microbiol. 31(6): 1795-1807.

Ding, Z., Atmakuri, K., and Christie, P.J. 2003. The ins and outs of bacterial type IV secretion substrates. Trends Microbiol. 11(11): 527-535.

Djamei, A., Pitzschke, A., Nakagami, H., Rajh, I and Hirt, H. 2007. Trojan horse strategy in *Agrobacterium* transformation: abusing MAPK defense signaling. Science. 318(5849): 453-456.

Draper, O., Cesar, C.E., Machon, C., de la Cruz, F., and Llosa, M. 2005. Site-specific recombase and integrase activities of a conjugal relaxase in recipient cells. Proc. Natl. Acad. Sci. U.S.A 102(45): 16385-16390.

Dreier, B, Beerli, R.R, Segal, D.J., Flippin, J.D, and Barbas, C.F. 2001. Development of zinc finger domains for recognition of the 5'-ANN-5' family of DNA sequences and their use in the construction of artificial transcription factors. J. Biol. Chem. 276: 29466-29478.

Dreier, B, Fuller, R.P., Segal, D.J., Lund, C.V., Blancefort, P., Huber, A., Koksch, B., and Barbas, C.F. 2005. Development of zinc finger domains for recognition of the 5'-CNN-5' family of DNA sequences and their use in the construction of artificial transcription factors. J. Biol. Chem. 280: 35588-35597.

Dujon, B. 1980. Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci. Cell 20(1): 185-97

Dujon, B. 1989. Group I introns as mobile genetic elements: facts and mechanistic speculations – a review. Gene. 82: 91-114.

Durai, S., Mani, M., Kandavelou, K., Wu, J., Porteus, M.H., and Chandrasegaran, S. 2005. Zinc finger nucleases: custom-designed scissors for genome engineering of plant and mammalian cells. Nucleic Acids Res. 33(18): 5978-5990.

Dürrenberger, F, Crameri, A., Hohn, B., and Koukolikova-Nicola., Z. 1989. Covalently bound VirD2 from *Agrobacterium tumefaciens* protects the T-DNA from exonucleolytic degradation. Proc. Natl. Acad. Sci. U.S.A 86(23): 9154-9158.

Francia M.V., Varsaki, A., Garcillán-Bárcia, M.P., Lahorre, A., Drainas, C., and de la Cruz, F. 2004. A classification scheme for mobilization regions of bacterial plasmids. FEMS Microbiol. Rev. 28(1): 79-100.

Francis, K.E., and Spiker, S. 2005. Identification of *Arabidopsis thaliana* transformants without selection reveals a high occurrence of silenced T-DNA integrations. The Plant Journal. 41: 464-477.

Frost, L.S., Ippen-Ihler, K., and Skurray, R.A. 1994. Analysis of the sequence and gene products of the transfer region of the F sex factor. Microbiol. Rev. 58(2): 162-210.

Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. 2005. Mobile genetic elements: the agents of open source evolution. Nat. Rev. Microbiol. 3(9):722-732.

Galetto, R., Duchateau, P., and Pâques, F. 2009.. Targeted approaches for gene therapy and the emergence of engineered meganucleases. Expert Opin. Biol. Ther 9(10): 1289-1303.

Garcillan-Barcia, M. P.,Jurado, P., González-Pérez, B., Moncalián, G., Fernández, L.A., and de la Cruz, F. 2007. Conjugative transfer can be inhibited by blocking relaxase activity within recipient cells by intrabodies. Mol. Microbiol. 63(2): 404-416.

Garcillan-Barcia, M. P., Francia, M.V., and De la Cruz, F. 2009. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev 33(3): 657-87.

Gomis-Ruth, F.X., and Coll, M. 2006. Cut and move: protein machinery for DNA processing in bacterial conjugation. Curr. Opin. Struct. Biol. 16(6): 744-752.

de Groot, M. J., Bundock, P., Hooykaas, P. J., and Beijersbergen, A. G. 1998. *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. Nat Biotechnol. 16(9): 839-42.

van der Graaff, E., den Dulk-Ras, Am, and Hooykaas, P.J.J. 1996. Deviating T-DNA transfer from *Agrobacterium tumefaciens* to plants. Plant Mol. Biol. 31: 677-681.

Guo, M, Hou, Q., Hew, C. L., and Pan, S. Q. 2007a. *Agrobacterium* VirD2-binding protein is involved in tumorigenesis and redundantly encoded in conjugative transfer gene clusters. Mol. Plant Microb. Interact. 20(10): 1201-1212.

Guo, M. Jin, S., Sun, D., Hew, C. L., and Pan, S. Q. 2007b. Recruitment of conjugative DNA transfer substrate to *Agrobacterium* type IV secretion apparatus. Proc. Natl. Acad. Sci. U.S.A. 104(50): 20019-20024.

Hamilton, H.L., Dominguez, N.M., Schwartz, K.J., Hackett, K.T., and Dillard, J.P. 2005. *Neisseria gonorrhoeae* secretes chromosomal DNA via a novel Type IV Secretion System. Mol. Microbiol. 55(6): 1704-1721.

Hawkey P.M., and Jones A.M. The changing epidemiology of resistance. 2009. J Antimicrob.

Chemother. (64) Suppl 1:i3-10.
Hoekema, A., Hirsch, P.R., Hooykaas, P.J.J., and Schilperoort, R.A. 1983. A binary plant vector strategy based on separation of vir- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. Nature. 303: 179–180.
Hofreuter, D, Odenbreit, S., and Haas, R. 2001. Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. Mol. Microbiol. 41(2): 379-391.
Hooykaas-van Slogteren, G. M., Hooykaas, P. J., and Schilperoort, R. A. 1992. Expression of Ti plasmid genes in monocotyledonous plants infected with *Agrobacterium tumefaciens*. Biotechnology 24: 382-383.
Jasin, M. 1985. High frequency of homologous recombination in mammalian cells between endogenous and introduced SV40 genomes. Cell. 43(3): 695-703.
Juhas, M., Crook, D.W., and Wood, D.W. 2008. Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. Cell. Microbiol. 10(12): 2377-2386.
Jurica, M.S., and Stoddard, B.L. Homing endonucleases: structure, function and evolution. 1999. CMLS Cell. Mol. Life Sci. 1304-1326.
Khan, S.A. 2005. Plasmid rolling-circle replication: highlights of two decades of research. Plasmid. 53: 126-136.
Kim, S.I., Veena, and Gelvin, S.B. 2007. Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the *Arabidopsis* genome generated under non-selective conditions. The Plant J. 51: 779-791.
Kim, Y.G., Cha, J., and Chandrasegaran, S. 1996. Hybrid restriction enzymes: zinc finger fusions to *Fok*I cleavage domain. Proc. Natl. Acad. Sci. U.S.A. 93: 1156-1160.
Kononov, M.E., Bassuner, B., and Gelvin, S.B. 1997. Integration of T-DNA binary vector 'backbone' sequences into the tobacco genome: evidence for multiple complex patterns of integration. The Plant J. 11(5): 945-957.
Kostriken, R., Strathern, J.N., Klar, A.J., , Hicks, J.B., and Heffron, F. 1983. A site-specific endonuclease essential for mating-type switching in *Saccharomyces cerevisiae*. Cell. 35(1): 167-74.
Lacroix, B., Loyter, A., and Citovsky, V. 2008. Association of the *Agrobacterium* T-DNA-protein complex with plant nucleosomes. Proc. Natl. Acad. Sci. U.S.A. 104(40): 15429-15434.
Lacroix, B., and Citovsky, V. 2009. *Agrobacterium* aiming for the host chromatin. Comm. and Integrative Biol. 2(1): 42-45.
Lanka, E., and Wilkins, B. M. DNA processing reactions in bacterial conjugation. 1995. Annu. Rev. Biochem. 64: 141-169.
Lawley, T.D., Klimke, W.A., Gubbins, M.J. and Frost, L.S. 2003. F factor conjugation is a true Type IV secretion system. FEMS Microbiol. Lett. 224(1): 1-15.
Lee, Y. W., Jin, S., Sim, W. S., and Nester, E. W. 1996. The sensing of plant signal molecules by *Agrobacterium*: genetic evidence for direct recognition of phenolic inducers by the VirA protein. Gene. 179(1): 83-8.

Lee, Y. W., Jin, S., Sim, W. S., and Nester, E. W. 1995. Genetic evidence for direct sensing of phenolic compounds by the VirA protein of *Agrobacterium tumefaciens*. Proc. Natl. Acad. Sci. U.S.A. 92(26): 12245-12249.

Llosa, M., Gomis-Ruth, F.X., Coll, M., and de la Cruz, F. 2002. Bacterial conjugation: a two-step mechanism for DNA transport. Mol. Microbiol. 45(1): 1-8.

Lu, J., dan Dulk-Ras, A., Hooykaas, P. J. J., and Glover, M. J. *Agrobacterium tumefaciens* VirC2 enhances T-DNA transfer and virulence through its C-terminal ribbon-helix-helix DNA-binding fold. 2009. Proc. Natl. Acad. Sci. U.S.A. 106(24): 9643-9648.

Miranda, A., Janssen, G., Hodges, L., Peralta, E.G., and Ream, W. 1992. *Agrobacterium tumefaciens* transfers extremely long T-DNAs by a unidirectional mechanism. J. Bacteriol. 174(7): 2288-2297.

Otten, L., Piotrowiak, G., Hooykaas, P., Dubois, M., Szegedi, E., and Schell., J. 1984. Restoration of virulence of *vir* region mutants of *Agrobacterium tumefaciens* strain B6S3 by coinfection with normal and mutants *Agrobacterium* strains. Mol Gen. Genet. 195: 159-163.

Pâques, F. and Duchateau, P. 2007. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr. Gene Ther. 7: 49-66.

Pacher, M., Schmidt-Puchta, W. and Puchta, H. 2007. Two unlinked double-strand breaks can induce reciprocal exchanges in plant genomes via homologous recombination and non-homologous end joining. Genetics. 174: 21-29.

Pansegrau, W., and Lanka, E. 1996. Mechanisms of initiation and termination reactions in conjugative DNA processing. J. Biol. Chem. 271(22): 13068-13076.

Peralta, E.G., Helmiss, R., and Ream, W. 1986. Overdrive, a T-DNA transmission enhancer on the *A. tumefaciens* tumour-inducing plasmid. The EMBO (Eur. Mol. Biol. Organ.) J. 5(6): 1137-1142.

Pietrokovski, S. 1994. Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. Protein Sci. 3: 2340-2350.

Pitzschke, A, and Hirt, H. 2010. New insights into an old story: *Agrobacterium*-induced tumour formation in plants by plant transformation. EMBO (Eur. Mol. Biol. Organ.) J. e-pub ahead of print.

Podevin, N., De Buck, S., De Wilde, C., and Depicker, A. 2006. Insights into recognition of the T-DNA border repeats as termination sites for T-strand synthesis by *Agrobacterium tumefaciens*. Transgenic Res. 15: 557-571.

Puchta, H. 2002. Gene replacement by homologous recombination in plants. Plant Mol. Biol. 48: 173-182.

Puchta, H., Dujon, B., and Hohn, B. 1993. Homologous recombination in plant cells is enhanced by *in vivo* induction of double strand breaks into DNA by a site-specific endonuclease. Nucleic Acids Res. 21(22): 5034-5040.

Ramanathan, V., and Veluthambi, K. 1995. Transfer of non-T-DNA portions of the *Agrobacterium tumefaciens* Ti plasmid pTiA6 from the left terminus of T_L-DNA. Plant Mol. Biol. 28: 1149-1154.

Scheiffele, P, Pansegrau, W., and Lanka, E. 1995. Initiation of *Agrobacterium tumefaciens* T-DNA processing. Purified proteins VirD1 and VirD2 catalyze site- and strand-specific cleavage of

superhelical T-border DNA *in vitro*. J. Biol. Chem. 270(3): 1269-1276.
Schroder, G. and Lanka, E. 2005. The mating pair formation of conjugative plasmids – A versatile secretion machinery for transfer of proteins and DNA. Plasmid. 54(1): 1-25.
Segal, D.J., Breier, B, Beerli, R.R., and Barbas, C.F. 1999. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-CNN-3' DNA target sequences. Proc. Natl. Acad. Sci. 96: 2758-2763.
Shurvinton, C.E., Hodges, L., and 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(24):11837-11841.
Stoddard, B. L. 2005. Homing endonuclease structure and function. Q. Rev. Biophys. 38(1): 49-95.
Strathern, J. N., Klar, A. J., Abraham, J. A., Ivy, J. M., Nasmyth, K. A., and McGill, C. 1982. Homothallic switching of yeast mating type cassettes is initiated by a double-stranded cut in the MAT locus. Cell. 31(1): 183-192.
Stingl, K., Müller, S., Scheidgen-Kleyboldt, G., Clausen, M., and Maier, B. 2010. Composite system mediates two-step DNA uptake into *Helicobacter pylori*. Proc. Natl. Acad. Sci. U.S.A. 107(3): 1184-1189.
Steck, T., Close, T.J., and Kado, C.I. 1989. High levels of double-stranded transferred DNA (T-DNA) processing from a nopaline Ti plasmid. Proc. Natl. Acad. Sci. U.S.A. 86: 2133-2137.
Swanson, M.S., and Hammer, B.K. 2000. *Legionella pneumophilia* pathogenesis: a fateful journey from amoebae to macrophages. Annu. Rev. Microbiol. 54:567-613.
Toro, N., Datta, A., Yanovsky, M., and Nester, E. 1988. Role of the overdrive sequence in T-DNA border cleavage in *Agrobacterium*. Proc. Natl. Acad. Sci. U.S.A. 85(22): 8558-8562.
Toro, N., Datta, A., Carmi, O.A., Young, C., Prusti, R.K., and Nester, E.W. 1989. The *Agrobacterium tumefaciens virC1* gene product binds to overdrive, a T-DNA transfer enhancer. J. Bacteriol. 171(12): 6845-6849.
Tzfira, T., Frankman, L.R., Vaidya, M., and Citovsky, V. 2003. Site-specific integration of *Agrobacterium tumefaciens* T-DNA via double-stranded intermediates. Plant Physiol. 133: 1011-1023.
Tzfira, T., Li., J., Lacroix, B., and Citovsky, V. 2004. *Agrobacterium* T-DNA integration: molecules and models. TRENDS in genetics. 20(8): 375-383.
Tzfira, T., Valdya, M., and Citovsky, V. 2004. Involvement of targeted proteolysis in plant genetic transformation by *Agrobacterium*. Nature. 431: 87-92.
Yadav, N. S., Vanderleyden, J., Bennett, D. R., Barnes, W. M., and Chilton, M. D. 1982. Short direct repeats flank the T-DNA on a nopaline Ti plasmid. Proc. Natl. Acad. Sci. U.S.A. 79(20): 6322-6326.
Van Attikum, H., Bundock, P., and Hooykaas, P.J.J. 2001. Non-homologous end-joining proteins are required for *Agrobacterium* T-DNA integration. EMBO (Eur. Mol. Biol. Organ.) J. 20(22): 6550-6558.
Van der Graaff, E., Den Dulk-Ras., A., and Hooykaas, P.J.J. 1996. Deviating T-DNA transfer from *Agrobacterium tumefaciens* to plants. Plant Mol. Biol. 31(3): 677-681.

Van Haaren, M.J.J., Sedee, N.J.A., Schilperoort, R.A., and Hooykaas, P.J.J. 1987. Overdrive is a T-region transfer enhancer which stimulates T-strand production in *Agrobacterium tumefaciens*. Nucleic Acids Res. 15(21): 8983-8997.

Vergunst, A. C., Schrammeijer, B., den Dulk-Ras, A., de Vlaam, C. M., Regensburg-Tuïnk, T. J., and Hooykaas, P. J. 2000. VirB/D4-dependent protein translocation from *Agrobacterium tumefaciens* into plant cells. Science. 290(5493): 979-982.

Vergunst, A. C., van Lier, M. C., den Dulk-Ras, A., Stuve, T. A., Ouwehand, A., and Hooykaas, P. J. 2005. Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of *Agrobacterium*. Proc. Natl. Acad. Sci. U.S.A. 102(3): 832-837.

White, C.E., and Winans, S.C. 2007. Cell-cell communication in the plant pathogen *Agrobacterium tumefaciens*. Phil. Trans. R. Soc. Lond. B. Biol. Sci. 362(1483): 1135-1148.

Windels, P., De Buck, S., Van Bockstaele, E., De Loose, M., and Depicker, A. 2003. T-DNA integration in *Arabidopsis* chromosomes. Presence and origin of filler DNA sequences. Plant Physiol. 133: 2061-2068.

Zaltsman, A., Krichevsky, A., Loyter, A., and Citovsky, V. 2010. *Agrobacterium* induces expression of a host F-box protein required for tumorigenicity. Cell Host & Microbe. 7(3): 197-209.

Zhu, J., P. M. Oger P. M., Schrammeijer, B. Hooykaas, P. J. Farrand, S. K. and Winans, S. C. 2000. The bases of crown gall tumorigenesis. J Bacteriol 182(14): 3885-3895.

Ziemienowicz, A., Tinland, B., Bryant, J, Gloeckler, V., and Hohn, B. 2000. Plant enzymes but not *Agrobacterium tumefaciens* VirD2 mediate T-DNA ligation *in vitro*. Mol. Cell Biol. 20(17): 6317-6322.

Chapter 2

Agrobacterium-Mediated T-DNA Transfer and Integration by Minimal VirD2 Consisting of the Relaxase Domain and a Type IV Secretion System Translocation Signal

Maartje van Kregten, Beatrice I. Lindhout,
Paul J. J. Hooykaas, and Bert J. van der Zaal

Modified from Van Kregten *et al.*, 2009,
Mol. Plant-Microbe Interact. 22(11):1356-1365.

Abstract

The VirD2 protein of *Agrobacterium tumefaciens* is essential for processing and transport of the T-DNA. It has at least three functional domains: a relaxase domain at the N-terminus, a bipartite nuclear localization signal (NLS), and a sequence called ω at the C terminus. We confirm here that deletions of the C-terminal part of VirD2 led to lack of transfer of T-DNA but, for the first time, we report that virulence is restored when these truncations are supplemented at the C terminus by a short translocation signal from the VirF protein. The lack of virulence of C-terminal deletions suggests that the C-terminal part contains all or part of the translocation signal of VirD2. Using a novel series of mutant VirD2 proteins, the C-terminal half of VirD2 was further investigated. We demonstrate that the C-terminal 40 amino acids of VirD2, which include the NLS and ω, contain all or part of the translocation domain necessary for transport of VirD2 into plant cells, while another element is present in the middle of the protein. The finding that a type IV secretion system transport signal at the C-terminus of VirD2 is necessary for virulence provides evidence for the role of VirD2 as the protein driving translocation of the T-strand from *Agrobacterium tumefaciens* into plant cells.

Introduction

Throughout recent decades, *Agrobacterium tumefaciens* has been widely used as a tool for the genetic engineering of a variety of organisms, which include dicotyledonous plants, monocotyledonous plants (Hooykaas-van Slogteren *et al.,,* 1984), yeast (Bundock *et al.,,* 1995), and other fungi (de Groot *et al.,* 1998).

During transformation by *A. tumefaciens*, a single-stranded (ss)DNA segment is transferred from the bacterium to the recipient cell, where it is integrated into the genome. This transferred segment or T-strand, also known as T-DNA, is derived from a particular region (T-region) of a tumor-inducing (Ti)-plasmid, where it is flanked by two imperfect direct repeats, the left (LB) and right (RB) border sequences. Among the Ti plasmid-encoded virulence (Vir) proteins that are required for T-strand formation and translocation to the recipient cell (Christie and Vogel 2000, Zhu *et al.,* 2000), VirD2 is absolutely

essential. Together with VirD1, VirD2 forms a relaxosome at the RB of the T-region, which initiates the formation of the T-strand. VirD1 and VirD2 are both required for recognition of the border repeat sequences flanking the T-strand but it is the relaxase domain of the VirD2 protein which actually cleaves the LB and RB sequence on one of the DNA strands, enabling the release of the T-strand (Jayaswal *et al.,* 1987, Lessl and Lanka 1994, Scheiffele *et al.,* 1995). VirD2 remains covalently attached to the 5′ end of the T-strand through an N-terminal tyrosine residue (Tyr29) (Dürrenberger *et al.,* 1989, Scheiffele *et al.,* 1995, Vogel and Das 1992).

After its formation within *A. tumefaciens*, the VirD2-Tstrand complex is transported to the cytoplasm of the recipient eukaryote, along with but independently from other virulence proteins such as VirE2, VirF (Vergunst *et al.,* 2000), VirD5 (Vergunst *et al.,* 2005), and VirE3 (Schrammeijer *et al.,* 2003). These transport or translocation events are mediated by the type IV secretion system (T4SS) of *A. tumefaciens*, a product of the *virB* operon, associated with the coupling protein VirD4 (Christie *et al.,* 2005). It is a matter of debate how the complex is translocated into recipient cells, whether translocation is driven by the relaxase or by the T-DNA. To support the favored model of relaxase-driven translocation of the T-DNA (Cascales and Christie 2003, Chen *et al.,* 2005), evidence has been found of low levels of translocation of relaxases MobA (plasmid RSF1010) and VirD2 without the presence of T-DNA (Vergunst *et al.,* 2005), indicating that transit through the T4SS is possible for unbound relaxases. However, the fact that translocation of these proteins is only detected at a very low level could very well mean that the presence of a covalently bound T-DNA moiety is the normal prerequisite for their transfer. In that manner, recognition and transfer of T-DNA by the T4SS rather than the relaxase would be the predominant reason for the transport of relaxases. This issue is still not resolved. A possible mechanism for T-DNA-driven translocation has been suggested when it was discovered that TrwB, the coupling protein of the T4SS of plasmid R388, possesses DNA-dependent ATPase activity (Tato *et al.,* 2005). Another model suggests that ssDNA binding VirE2 proteins in the recipient cell form a structure that can pull the T-strand into the recipient cell (Grange *et al.,* 2008).

It is hypothesized that VirE2 binds to the T-strand in the cytoplasm of the recipient cell and aids in protection of the T-strand from attack by nucleases (Citovsky *et al.,* 1989). VirD2 is thought to do the same for the 5′ end (Dürrenberger *et al.,* 1989). Although VirD2 and VirE2 both contain one or more nuclear localization signals (NLS), nuclear targeting of the T-strand complex has been primarily attributed to an NLS of VirD2. Although the

N-terminal part of VirD2 proteins contains a monopartite NLS composed of amino acids 32 through 35 (KRAR) (Rossi *et al.,* 1993), the very C-terminal end of VirD2, which is conserved between different VirD2 proteins, has a bipartite NLS. Both of them are functional in a modular fashion, able to guide reporter proteins into the nucleus, but the C-terminal bipartite NLS has been shown to be essential for nuclear uptake of the T-complex (Howard *et al.,* 1992, Relić *et al.,* 1998, Tinland *et al.,* 1995). Recently, it was shown that VirE2 is able to interact with importin α when bound to ssDNA, suggesting a role for both VirD2 and VirE2 in nuclear uptake of the T-complex (Bhattacharjee *et al.,* 2008).

The VirD2 protein encoded by the Ti-plasmid of the widely used octopine type strains of *Agrobacterium* is a 47.5-kDa protein of 424 amino acids (Fig. 1). Alignment of VirD2 proteins from different *Agrobacterium* strains has shown that the protein consists of two distinct parts. Although the N-terminal relaxase domains are virtually identical, the C-terminal halves of VirD2 proteins are rather variable (Howard *et al.,* 1992, Shurvinton *et al.,* 1992). Despite this variation, several conserved sequences are present at the C-terminal end of VirD2. Howard and associates (1992) aligned VirD2 proteins from different *Agrobacterium* strains and found that the bipartite NLS is conserved at the C-terminal end. Shurvinton and associates (1992) also aligned the C-terminal end of different VirD2 proteins and identified the bipartite NLS and the ω sequence as conserved between the different VirD2 proteins.

Based on an alignment of the C terminus of translocated proteins (including VirD2, VirF, VirE2, VirE3, and the relaxase MobA of plasmid RSF1010), a putative C-terminal translocation signal containing several conserved arginine residues was annotated (Vergunst *et al.,* 2000) (Fig. 1). In VirD2, this signal partly overlaps with the bipartite NLS also present at the C-terminal end.

The very C-terminal amino acid sequence, designated the omega (ω) sequence (DGRGG) (Fig. 1), has been shown to be important for the function of VirD2. Insertion or deletion of amino acids of ω was shown to have detrimental effects on the final number of T-strand integration events in recipient plant cell genomes (Bravo-Angel *et al.,* 1998, Shurvinton *et al.,* 1992). However, the precise role of VirD2 regarding T-strand integration events has been the object of debate. Mysore and associates (1998) reported ω to be important for integration of the T-DNA into the host genome. However, Tinland and associates (1995) reported that this might at least partially be a consequence of VirD2-mediated protection of the 5′ end of the T-strand against nucleolytic attack, thus not necessarily reflecting a reduced efficiency of the integration step itself. Furthermore, Bravo-Angel and associates (1998)

published that ω was important for T-DNA translocation to the recipient cell, whereas the efficiency and the pattern of integrations were unaffected by ω mutations.

VirD2 contains a large domain of unknown function (DUF) between the relaxase domain and the very C-terminal end (Fig. 1). This DUF consists of amino acids residues which are poorly conserved between different VirD2 proteins (Howard *et al.,* 1992). Although substantial in size, no enzymatic functions have thus far been attributed to the VirD2 DUF. However, it would be surprising if rapidly evolving bacteria would maintain such a large domain when it had no function. If it has no enzymatic activity, it might be required for recruiting interaction partners in the recipient cell. Indeed, several *Arabidopsis thaliana* cyclophilins were shown to interact with sequences within the DUF (Deng *et al.,* 1998).

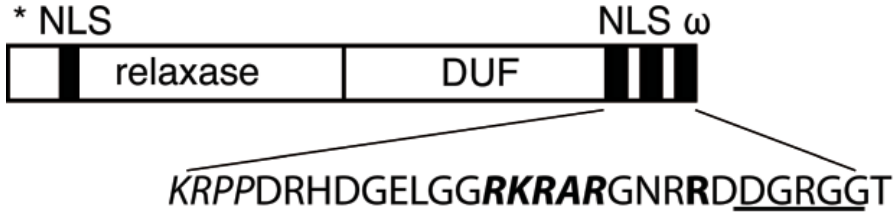

Figure 1: Schematic representation of the VirD2 coding region. Within the N-terminal 228 amino acids, the relaxase domain is encoded, which is essential for T-strand processing (Scheiffele *et al.,* 1996; Steck *et al.,* 1990). Tyrosine residue 29 (indicated by *) binds covalently to the 5′ end of the T-strand (Vogel *et al.,* 1992). The N-terminal nuclear localization signal (NLS) (Rossi *et al.,* 1993) is indicated by a black box. The domain of unknown function (DUF) of VirD2 consists mainly of amino acid residues which are poorly conserved between VirD2 proteins (Howard *et al.,* 1992). At the very C-terminal end, which consists of approximately 40 amino acids that are conserved between different VirD2 proteins, the bipartite NLS (in italics) has been shown to function in nuclear uptake of the T-complex (Relić *et al.,* 1998; Tinland *et al.,* 1995). The ω domain (in italics and underlined), another element of the conserved very C-terminal end, has been described as being important for T-DNA transfer (Bravo-Angel *et al.,* 1998) or for integration of T-DNA within the plant genome (Mysore *et al.,* 1998; Shurvinton *et al.,* 1992). The C-terminal 40 amino acids of VirD2 contain a consensus sequence implicated in translocation through the type IV secretion system of *Agrobacterium tumefaciens* (indicated in bold) (Vergunst *et al.,* 2005). Figure not drawn to scale.

In this study, we used a novel series of mutants of VirD2 to investigate the requirements for translocation and integration of T-strands into recipient plant cells. Evidence is presented that part of the DUF and the C-terminal 40 amino acids are required for translocation of the VirD2-T-strand complex but that the interactions of plant proteins

with the DUF described thus far do not influence the transformation efficiency. A minimal version of VirD2, consisting of only the relaxase domain, still meets the requirements for translocation when supplemented with a short translocation signal derived from VirF. The *virF* gene is present only in some Ti plasmids. Apart from its ability to enhance the virulence of octopine strains of *Agrobacterium* on some host plants, the native VirF protein is not at all involved in T-strand formation or its transfer to host cells (Regensburg-Tuïnk and Hooykaas 1993). Altogether, the data reported here provide compelling evidence that, once a T-strand has been made within *A. tumefaciens*, the translocation of the VirD2-T-strand complex is protein driven rather than DNA driven. Moreover, once inside a recipient plant cell, the DUF is not required for transformation of the recipient cell. The broader significance of these results is discussed.

Results

Providing VirD2 with FLAG and SV40 NLS sequences and the VirF protein translocation signal.

In our study, we investigated in detail which sequences within the C-terminal DUF of the VirD2 protein are essential for virulence of *A. tumefaciens* and for translocation of T-strands to plant cells. For this purpose, we made a set of VirD2 C-terminal deletion constructs. For convenience of protein immunodetection, we added a FLAG tag at the N terminus. To ensure nuclear targeting when the endogenous C-terminal bipartite NLS of VirD2 was deleted, an SV40 NLS was added downstream of the FLAG tag, N-terminal of the VirD2 moieties (Fig. 2). In order to compensate for the possible loss of a T4SS secretion signal, which was postulated to be present at the very C-terminal end (Vergunst *et al.,* 2005), we also made a series of constructs to which we added the C-terminal 37 amino acids of VirF (F). This particular F sequence was shown to enable translocation of proteins through the T4SS of *A. tumefaciens* using Cre-fusion proteins to detect protein translocation directly (Vergunst *et al.,* 2005).

Prior to testing VirD2 deletion constructs, we first tested the transfer of T-DNA via the tagged VirD2 derivatives that formed the starting point of the experiment. Hence, the activities of FLAG-NLS-VirD2 and FLAG-NLS-VirD2-F were established in *Arabidopsis* root transformation assays, using *Agrobacterium* sp. strain LBA2585 (Δ*virD2* and ΔT-DNA).

The wide-host-range plasmid pCAMBIA2301, which contains a T-DNA sequence with the plant-selectable *nptII* gene and a β-glucuronidase (*GUS)::intron* reporter gene between the T-DNA border repeats, was used as a T-strand donor. In such a root transformation assay, GUS staining directly after the 3 days of cocultivation can be taken as a measure for the total amount of T-strand molecules that have been transferred from the bacterial cell to the nuclei of plant cells; thus, those that have already been integrated into the genome as well as those that have not yet been incorporated and might be only transiently present (Tinland *et al.,* 1995).

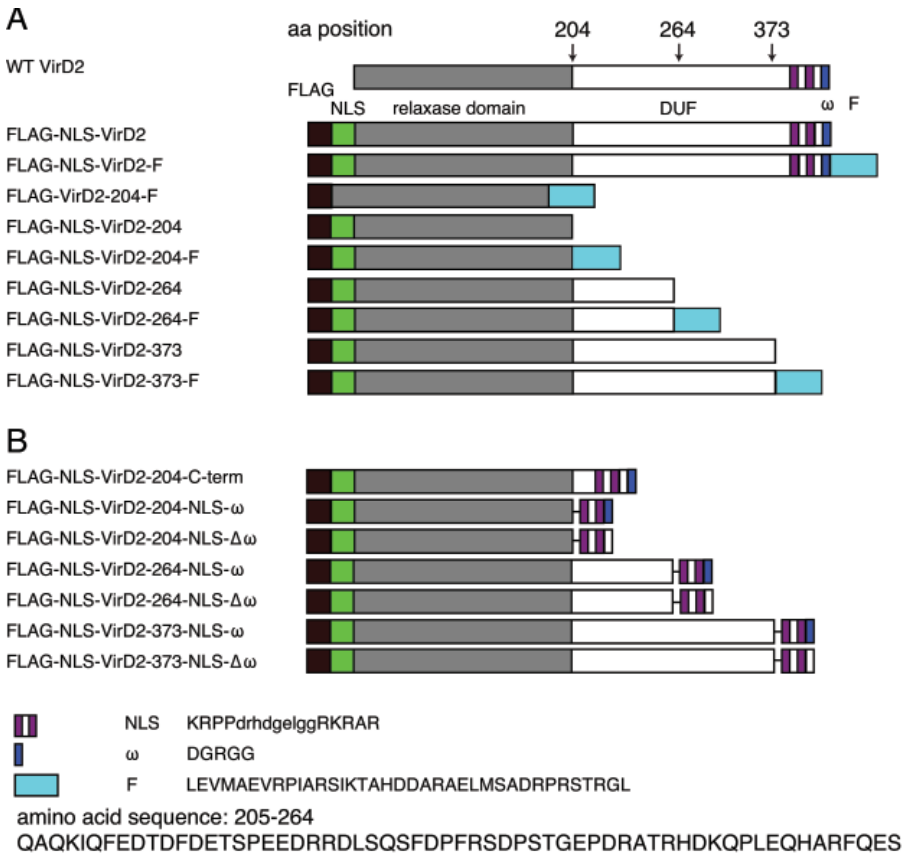

Fig. 2. A, Schematic overview of the protein structure of VirD2 truncations. **B,** Schematic overview of the protein structure of VirD2 truncations complemented with the different elements of the C-terminal end. Pictures not drawn to scale.

The root transformation assays (Table 1) clearly demonstrated that fusing a FLAG-NLS sequence at the N terminus of VirD2 did not abolish its activity, although it reduced the transformation efficiency by approximately 60% when compared with wild-type (WT) VirD2 protein. Therefore, the VirD2 protein tolerates the N-terminal modifications while retaining most of its T-strand transferring capacity. The C-terminal addition of F did not hamper VirD2 function at all. In fact, FLAG-NLS-VirD2-F is transferred at the efficiency of WT VirD2. However, C-terminal addition of F to WT VirD2, without FLAG and NLS, resulted in a twofold reduction of transformation rate (data not shown), indicating that WT levels of transformation cannot be boosted further by the F sequenc

Table 1. Root transformation assays

Construct[a]	**Transient transformation**[b]
WT VirD2	1
FLAG-NLS-VirD2	0.39 ± 0.13
FLAG-NLS-VirD2-F	1.07 ± 0.21
FLAG-VirD2-F	0.89 ± 0.20

[a] WT = wild type, NLS = nuclear localization signal.

[b] For every experiment, at least 100 explants were scored for β-glucuronidase-positive spots. Data were normalized against data for WT VirD2. n = 3.

Deletion of the DUF of FLAG-NLS-VirD2 still allows proficient transient and stable T-DNA transfer to *A. thaliana* root explants when F is present.

We investigated the effects of deletions in the C terminus of FLAG-NLS-VirD2 while leaving the N-terminal relaxase domain intact, because relaxase activity is obviously essential for T-DNA processing (Lessl *et al.,* 1994, Scheiffele *et al.,* 1995). Although the smallest version of VirD2 that was still able to process DNA correctly was previously reported to consist of 228 amino acids (Scheiffele *et al.,* 1995, Steck *et al.,* 1990), we decided to further minimize the relaxase domain of VirD2 to 204 amino acids, based upon sequence similarities between different relaxases.

The first series of deletion mutants (Fig. 2) comprised the following proteins: FLAG-NLS-VirD2-204, consisting of the N-terminal 204 amino acids of VirD2, which is the relaxase domain; FLAG-NLS-VirD2-264, consisting of the N-terminal 264 amino acids of VirD2, which is the relaxase domain and part of the DUF; FLAG-NLS-VirD2-373, consisting of VirD2, from which only the very C-terminal 51 amino acids have been

truncated, which includes the 40 amino acids in which the translocation signal is most likely present.

The expression of the various chimerical proteins in acetosyringone- induced *Agrobacterium* strains was determined by Western blotting using an anti-FLAG antibody (Fig. 3A). In all samples, a clear signal was present at a somewhat higher-than expected position but agreeing well with predicted molecular weights. Fusion with F adds 4 kDa to the protein. Degradation seemed to be minimal, suggesting that the different FLAG -NLS-VirD2 truncations were expressed and were stable in *A. tumefaciens* Proteins containing a C-terminal F translocation signal seemed to be produced in higher amounts than those without F.

In order to evaluate the plant-cell-transforming ability of the different *Agrobacterium* strains, it was decided to use root explants of *A. thaliana*. In this assay, even low amounts of transformation events are easily detected. GUS staining directly after cocultivation was taken as a measure for transient transformation, and the number of successful integrations was measured by the number of transgenic kanamycin-resistant calli that developed after several weeks on selective medium. All data were normalized against WT VirD2 (VirD2 expressed under the control of the *virD* promoter).

Translocation of T-strands by the FLAG-NLS-VirD2 C-terminal truncations was completely abolished (Table 2). Not even FLAG-NLS-VirD2-373, which is only deleted for the C-terminal 51 amino acids, was able to induce any GUS activity or callus formation, indicating that the very C-terminal end provides an essential function in translocation of the VirD2-Tcomplex. However, by providing a C-terminal VirF (F) translocation sequence to the FLAG-NLS-VirD2 truncations, translocation of T-strands was to a large extent restored. This effect was not enhanced by the presence of longer parts of the DUF (Table 2). The F sequence apparently was sufficient as a dominant modular domain enabling the passage of hybrid proteins and their covalently bound T-strands into recipient plant cells. No obvious difference between the ratio of transient transformation and stable integration was observed for any of the constructs, indicating that the DUF and the C-terminal end of VirD2 have no special role in T-DNA integration efficiency. To determine the necessity of the addition of the SV40 NLS, we created FLAG-VirD2-204-F, in which no extra NLS is present. Surprisingly, it performed at an even higher level than FLAG-NLS-VirD2-204-F, indicating that either the small NLS present in the relaxase domain (Rossi *et al.,* 1993) is sufficient for nuclear localization or that VirD2 is guided to the nucleus by another protein or by the VirF protein translocation signal. However, no previously undescribed putative NLS sequences

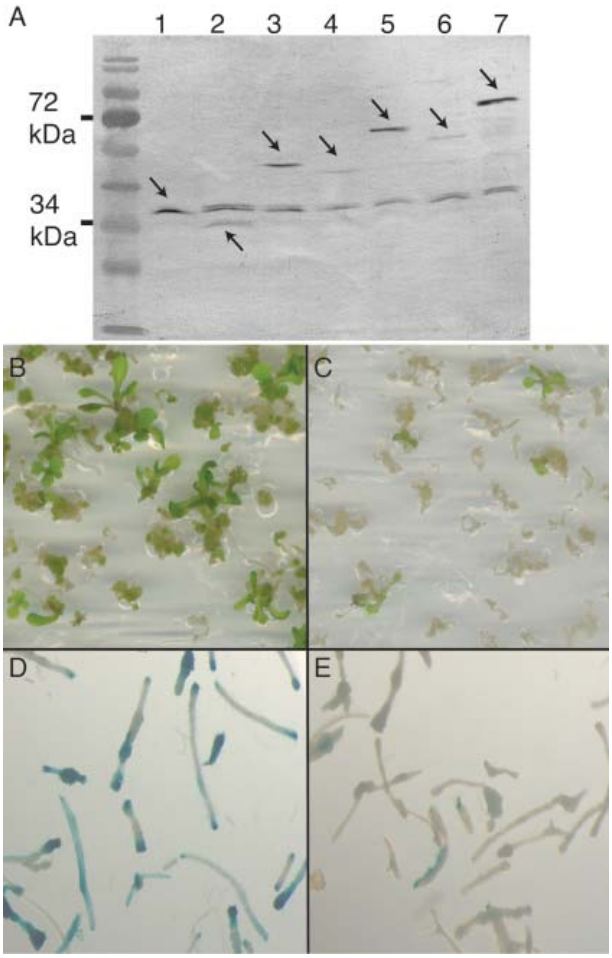

Fig. 3. A, Expression of VirD2 truncation mutants. Western blot showing expression of FLAG-NLS-VirD2-(F) and truncations thereof. Arrows indicate position of the bands corresponding to the FLAG-NLS-VirD2 truncations. Left to right: FLAG-NLS-VirD2-204-F, FLAG-NLS-VirD2-204, FLAG-NLS-VirD2-264-F, FLAG-NLS-VirD2-264, FLAG-NLS-VirD2-373-F, FLAG-NLS-VirD2-373, and FLAG-NLS-VirD2-F. In all lanes, a nonspecific band is also present. **B** and **C,** Callus formation after 3 weeks in *Arabidopsis thaliana* root explants transformed with *Agrobacterium tumefaciens* LBA2585 carrying pCAMBIA2301 and expressing **B,** wild-type (WT) VirD2 and **C,** FLAG-NLS-VirD2-264-NLS-Δω. **D** and **E,** β-Glucuronidase-positive spots in *Arabidopsis thaliana* root explants transformed with *Agrobacterium tumefaciens* LBA2585 carrying pCAMBIA2301 and expressing **D,** WT VirD2 and **E,** FLAG-NLSVirD2-264-NLS-Δω.

could be identified in VirF (PredictNLS progam.)

Domains of VirD2 required for translocation

Considering the experiments described above, the complete DUF as well as the remaining C-terminal sequences of VirD2 can, in fact, be replaced with a short VirF-derived protein translocation sequence which, by itself, is unlikely to perform any functions other than mediating passage through the T4SS. At the same time, because none of the C-terminal VirD2 deletion constructs were competent for translocation without F, part of the native VirD2 C terminus must be required for translocation. It has been reported that the native VirD2 protein contains a T4SS translocation sequence of its own and a candidate sequence at the C-terminal end has been postulated (Vergunst *et al.,* 2005). In order to define the sequences of VirD2 that are sufficient for translocation, we created internal deletion mutations in the DUF. Between the relaxase domain and the re-added C-terminal sequences, a linker sequence encoding the amino acids GGGS was provided to ensure flexibility between the two linked sequences (Wriggers *et al.,* 2005). We thus created FLAG-NLS-VirD2-204-C-term, consisting of the relaxase domain and the C-terminal sequence downstream from amino acid 374 (51 amino acids). This sequence includes the putative translocation signal, including the bipartite NLS and the ω domain (Fig. 2). To address the question of which C-terminal sequences are required for translocation, we also created FLAG-NLS-VirD2-204-NLS-ω, consisting of the relaxase domain, the C-terminal bipartite NLS, and ω (Fig. 2). In this construct, a 4-amino-acid sequence (GNRR) is deleted between the bipartite NLS and ω. We also created FLAG-NLS-VirD2- 204-C-term-Δω, consisting of the C-terminal sequence from amino acids 396 through 418**.** This sequence includes the bipartite NLS and the arginine-rich region but has ω precisely deleted. Using these VirD2 truncations, we tested for the effects of a full DUF deletion on competence of the truncated VirD2 protein for T-DNA translocation and virulence of the corresponding *Agrobacterium tumefaciens* strain. As was already evident from the earlier series of experiments (Table 2), a FLAG-NLS-VirD2-204 version of VirD2 without any further C-terminal additions was completely unable to deliver T-strands into plant cells. By adding any of the short VirD2-derived C-terminal sequences mentioned above, it was reproducibly found that the resulting FLAG-NLS-VirD2-204 derivatives were all able to deliver T-strands to *Arabidopsis* root explants, albeit only at a 1 to 5% level of the WT VirD2 protein (Table 3). Hence, for

restoration of appreciable VirD2 activity with VirD2-derived C-terminal sequences, the presence of at least part of the DUF seemed to be required. Therefore, FLAG-NLS-VirD2-264-NLS-ω and FLAG-NLS-VirD2-264- C-term-Δω were created, as well as the full DUF-containing versions FLAG-NLS-VirD2-373-NLS-ω and FLAG-NLSVirD2- 373-C-term-Δω. Remarkably, all these proteins could restore the virulence of VirD2 to the level conferred by FLAG-NLS- VirD2.

Altogether, our results show that the very C-terminal sequence of VirD2 does not contain a translocation signal that is as efficient as F in supporting T-DNA translocation. Furthermore, in our work, we did not find a reduction in T-DNA translocation when ω was deleted or a small internal deletion of 4 amino acids was present between NLS and ω. However, when VirD2-derived C-terminal sequences were used instead of the F, the presence of a small internal part of the DUF corresponding to amino acids 205 to 264 of WT VirD2 restored the virulence of the corresponding *A. tumefaciens* strains to the same level as the ones that possessed F as a C-terminal sequence. Addition of the larger part of DUF (amino acids 205 to 373) did not lead to any further enhanced T-strand transfer, thus providing compelling evidence that the sequence of DUF between amino acids 205 and 264 confers an important function in T-DNA translocation within the native VirD2 protein.

Table 2: C-terminal truncations

Construct[a]	**Transient transformation**[b]	**Stable integration**[c]
WT VirD2	1	1
FLAG-NLS-VirD2-F	1.07 ± 0.21	0.92 ± 0.17
FLAG-VirD2-204-F	0.82 ± 0.13	nd
FLAG-NLS-VirD2-204	0	0
FLAG-NLS-VirD2-204-F	0.57 ± 0.19	0.51 ± 0.21
FLAG-NLS-VirD2-264	0	0
FLAG-NLS-VirD2-264-F	0.61 ± 0.21	0.49 ± 0.23
FLAG-NLS-VirD2-373	0	0
FLAG-NLS-VirD2-373-F	0.65 ± 0.21	0.61 ± 0.22

[a]WT = wild type, NLS = nuclear localization signal. [b]For every experiment, at least 100 explants were scored for β-glucuronidase-positive spots. Data were normalized against data for WT VirD2. b Amount of calli per explants was scored 3 and 4 weeks after cocultivation. Data were normalized against WT VirD2. For every experiment, at least 100 explants were scored for callus formation. nd = not determined. $n = 3$.

Table 3. Root transformation assays

Construct[a]	Transient transformation[b]	Stable integration[c]
WT VirD2	1	1
FLAG-NLS-VirD2	0.39 ± 0.13	n.d.
FLAG-NLS-VirD2-204	0	0
FLAG-NLS-VirD2-204-C-term	0.05 ± 0.01	0.04 ± 0.01
FLAG-NLS-VirD2-204-C-term-Δω	0.05 ± 0.01	0.01 ± 0.01
FLAG-NLS-VirD2-204-NLS-ω	0.08 ± 0.01	n.d.
FLAG-NLS-VirD2-264-NLS-ω	0.33 ± 0.19	0.25 ± 0.15
FLAG-NLS-VirD2-264-C-term-Δω	0.41 ± 0.13	0.22 ± 0.10
FLAG-NLS-VirD2-373-NLS-ω	0.37 ± 0.01	0.24 ± 0.04
FLAG-NLS-VirD2-373-C-term-Δω	0.40 ± 0.18	n.d.

a WT = wild type, NLS = nuclear localization signal. b For every experiment, at least 100 explants were scored for β-glucuronidase- positive spots. Data were normalized against data for WT VirD2. c Amount of calli per explants was scored 3 and 4 weeks after cocultivation. Data were normalized against WT VirD2. For every experiment, at least 100 explants were scored for callus formation. n.d. = not determined

Discussion

By constructing a novel series of VirD2 mutants which was used for *Arabidopsis* and *Nicotiana glauca* transformations, we have shown that the large C-terminal half of VirD2, mostly consisting of a large DUF, can, in fact, be deleted without any serious consequences for the normal VirD2 functions, provided that a short 37-amino-acid translocation signal (F) of the VirF protein is present instead. These data shed a new light upon the truly essential domains of VirD2 in members of genus *Agrobacterium* itself and for the translocation of VirD2 via the T4SS of *A. tumefaciens*, a paradigm for conjugative DNA transfer and effector protein translocation.

Although multiple protein substrates have now been identified for the T4SS of *A. tumefaciens* (e.g., VirE2, VirE3, and VirF) (Citovsky *et al.,* 1989, Christie 2004, García-Rodríguez *et al.,* 2006), definitive proof that T-strand transfer is guided by the VirD2 protein is, in fact, lacking. For TrwC, the relaxase and helicase of plasmid R388, translocation independent of DNA transfer to the recipient has been shown (Draper *et al.,* 2005), although its efficacy remained unknown. Regarding VirD2, Cre recombinase reporter assays for protein translocation (CRAfT) assays indicated that VirD2 can mediate

Cre transfer to plant cells in the absence of T-DNA, albeit at a very low level (Vergunst *et al.,* 2005). For our experiments, CRAfT assays based upon the recombinogenic activity of Cre within plant cells were not feasible. Fusion of VirD2 to the N or C terminus of Cre resulted in near complete loss of recombination events in planta (*unpublished data*). We have taken these data as indications that either the Cre protein does not tolerate a VirD2 moiety or such fusions can no longer be translocated. Using T-strand transfer as a reporter system, as was done in our study, we present an alternative to CRAfT which, moreover, could give insights regarding the actual function of VirD2 in T-strand formation and translocation. Our data clearly demonstrated that properties of the VirD2 relaxase protein are the main determinants behind T-DNA translocation from *A. tumefaciens* through the T4SS into plant cells. Once delivered to the T4SS, DNA-dependent ATPase activity of coupling proteins can subsequently generate the energy required for DNA transfer (Llosa *et al.,* 2002, Tato *et al.,* 2005), or perhaps VirE2 can aid the translocation of the T-DNA from the cytoplasm of the recipient cell (Grange *et al.,* 2008).

The conclusion that T-strand translocation first of all depends on translocation of the VirD2 relaxase and not, by any means, upon the transferred DNA itself was particularly evident for the following reason: all VirD2 truncations based upon FLAG-NLS-tagged VirD2, including a complete truncation of the DUF, performed at 50 to 60% of WT VirD2 in a root transformation assay, provided that they were equipped with a C-terminal translocation signal of VirF (F). In none of the cases was any obvious difference between transient transformation (GUS staining directly after cocultivation) and stable integration (formation of kanamycin resistant calli) observed. Removal of F abolished T-strand translocation completely, strongly suggesting that C-terminal sequences of VirD2 are, indeed, required for translocation, thus corroborating data from other types of VirD2 mutants (Bravo-Angel *et al.,* 1998, Mysore *et al.,* 1998, Shurvinton *et al.,* 1992).

Our finding that DUF deletion mutants could still function in translocation of T-strands confirmed and extended earlier data that part of the DUF is not essential for VirD2 function (Koukolíková-Nicola *et al.,* 1993). In that study, it was shown that 111 to 142 amino acids of the C-terminal end of VirD2 could be deleted, provided that a 38-amino-acid VirD2 C-terminal sequence or the larger part thereof was fused to the N-terminal 275 amino acids. Although these data suggested that the C-terminal end of VirD2 could function as a rather independent moiety that was able to complement a severe C-terminal truncation, our data demonstrate that the situation is more complex. When the DUF extends to amino acid 373 or 264 of VirD2, it is, indeed, possible to complement VirD2 function with a short

VirD2-derived C-terminal sequence (Table 3). However, further truncation of the DUF down to amino acid residue 204 yielded an N-terminal half of VirD2 that was refractory to complementation with VirD2-derived C-terminal sequences but the protein regained T-strand delivering capacity when supplemented with F. Our data provide the first evidence that the amino acid sequence flanked by residues 204 and 264 has a biological function, in contrast to the larger C-terminal part of the DUF. This amino acid sequence does not show similarity to any proteins other than VirD2 and contains no known domains, as determined by BLAST, Pfam, and Scan- Prosite searches. As will be discussed below, the 204-to-264 sequence is most likely important for translocation processes within *A. tumefaciens* On the other hand, our data strongly indicate that the 37-amino-acid VirF translocation signal is extremely potent, possibly combining recruitment to the T4SS within *A. tumefaciens* cells as well as passage through the T4SS to eukaryotic cells.

As far as reported interactions of VirD2 domains with host-encoded proteins are concerned, our data, in combination with the earlier VirD2 truncations (Koukolíková-Nicola *et al.,* 1993), enabled a reevaluation regarding their biological relevance. Some plant cyclophilins - proteins involved in protein folding, protein degradation, and signal transduction (Wang and Heitman 2005) - have been reported to interact with VirD2 (Deng *et al.,* 1998). Roc1 was found to interact with amino acids 174 to 337 of VirD2. The full interaction domain with Roc1 is only present in our longest VirD2 truncation, VirD2-373. CypA was reported to interact with amino acids 274 to 337 of VirD2. Only FLAG-NLS-VirD2-373-F contains this interacting domain. Because, in our experiments, plant cell transformation via these truncated VirD2 proteins was easily obtained, these cyclophilins can be of only limited importance for VirD2 function *in planta*. Of course, it cannot be excluded that these interactions are necessary for correct folding of WT VirD2 but they apparently are not required for folding of the deletion constructs. Moreover, a study in our lab showed that deletion of all cyclophilins in a yeast strain did not inhibit transformation by *A. tumefaciens* (H. Van Attikum and P. J. J. Hooykaas, *unpublished results*). A biologically relevant interaction with the TATA box-binding protein, which binds to the N-terminal half of VirD2 (Bakó *et al.,* 2003), is still possible in our mutants.

Based upon our series of constructs, the 51-amino-acid sequence forming the very C terminus of VirD2 does not contain any elements sufficient for mediating efficient T-strand transfer to *Arabidopsis* root cells. Obviously, when transfer of T-strands to host cells is inhibited, their integration into the host genome will not be observed. Mostly for this reason, the involvement of VirD2-derived sequences during T-DNA integration into

the host genome has, thus far, been controversial (Bravo-Angel *et al.,* 1998, Mysore *et al.,* 1998, Tinland *et al.,* 1995, Ziemienowicz *et al.,* 2000). Considering our data, where the presence of a very different heterologous C-terminal sequence derived from VirF allowed for quite efficient transient and stable transformation events, even by severely truncated VirD2 proteins, it can be concluded that the C-terminal part of VirD2 is not of any special importance for T-DNA integration. Altogether, it is much more likely that the region C-terminal of amino acid 204 functions as a complex translocation signal, having at least two components: one for interaction with the T4SS and the other, we hypothesize, to enable recruitment of the T-complex to the T4SS.

For VirD2, several interactions responsible for recruitment of VirD2 and the T-strand to the T4SS within *A. tumefaciens* cells have recently been reported: with the VirD2-binding proteins (VBP) 1 to 3 (Guo *et al.,* 2007a and b) and with VirC1 (Atmakuri *et al.,* 2007). For the VBP, no additional function is known, but VirC1 is also responsible for enabling the generation of a high number of T-strands (Atmakuri *et al.,* 2007). Our experiments clearly demonstrated that the amino acid sequence corresponding to residues 204 to 264 of the VirD2 protein was able to strongly enhance the ability of VirD2 mutant proteins regarding T-strand transfer (Table 3). This effect was not further enhanced by a more extensive area of the DUF containing amino acids 204 to 373. Therefore, we hypothesize that the amino acids 204 to 264 provide an important interaction site of VirD2 with one or more *A. tumefaciens* proteins. Although binding of VBP1-3 was roughly mapped to the N-terminal half of VirD2 (Guo *et al.,* 2007a), it probably still includes amino acid residues of the 204-to-264 region. The VirD2 region responsible for the interaction with VirC1 is currently unknown. The presence of two signals in VirD2, one for translocation and one for recruitment to the T4SS, is interesting. We hypothesize that this could provide the bacterium with an extra measure of control, preventing the translocation of VirD2 without a T-strand attached. Of the DUF of VirD2, amino acids 265 to 373 seem to be dispensable in both *A. tumefaciens* and the two plant species used in this study. However, we cannot exclude that this particular sequence represents a host range factor in other plant species.

Acknowledgments

We thank C. N. Fontaine for technical assistance and A. Vergunst for advice in the early stage of the project.

Materials & methods

Construction of VirD2 mutants

Cloning steps were performed according to standard techniques in *Escherichia coli* DH5α. Both *E. coli* and *A. tumefaciens* were cultured in LC (tryptone at 10 g/liter, yeast extract at 5 g/liter, and NaCl at 8/liter) with the appropriate antibiotics. Primer and DNA oligomer sequences are listed in Table 4.

Construction of pSKN

Bluescript plasmid pSK+ was digested with *Sac*I and *Kpn*I, removing the existing polylinker sequences. Four primers (A to D) were used to construct a novel polylinker: A, *Sac*I-*Xho*I F; B, *Sac*I-*Xho*I R; C, *Xho*I-*Kpn*I F; and D, *Xho*I-*Kpn*I R. *Not*I sites flanking the former *Sac*I and *Kpn*I sites can be used to isolate *Not*I fragments for ligation into pBFF or pBFFstop (discussed below) with conservation of reading frame in the correct orientation. Next to the *Not*I site and just upstream of polylinker I, an 11-codon sequence encodes the SV40 nuclear localization signal (Fig. 4, sequence boxed) (Dunn *et al.,* 1988).

Construction of VirD2 truncations and modifications in pSKN

For construction of pSKN-VirD2N, the complete open reading frame (ORF) of VirD2 without the stop codon was obtained by polymerase chain reaction (PCR) on DNA isolated from *Agrobacterium* LBA1115 (Hood *et al.,* 1993) using primers D2NheF and D2XhoR. The purified 1.3-kb DNA fragment was digested with *Nhe*I and *Xho*I and cloned in similarly digested pSKN, creating pSKN-VirD2. The VirD2 truncation mutants were created using this construct. By digestion with *Nde*I and *Xho*I, a 693-bp C-terminal fragment was removed from the VirD2 coding region of pSKN-VirD2 and the desired mutated VirD2 proteins were reconstituted by supplying different *Nde*I/*Xho*I fragments, with these sites or appropriate overhangs provided via primer sequences. For the creation of VirD2-204, the *Nde*I/*Xho*I gap in pSKN-VirD2 was filled in with a double-stranded oligo with *Nde*I/*Xho*I-compatible overhangs using annealed VirD2-204FW and VirD2-204 RV. VirD2-264 was

made by filling of the *Nde*I/*Xho*I gap with a *Nde*I/*Xho*I-digested PCR fragment, made with primers D2FW and D2-264RV. VirD2-373 was made similarly by filling the gap with the *Nde*I/*Xho*I-digested PCR-product from primers VirD2 FW and VirD2 373 RV.

To create versions of VirD2-204, VirD2-264, and VirD2-373 that were complemented with C-terminal sequences from VirD2, pSKNVirD2 truncations were digested with *Xho*I and *Eco*RI. The same sites were present in all pairs of oligo sequences used. Between the end of the VirD2 truncations and the C-terminal sequences, a sequence encoding the small amino acid linker GGGGS was inserted to secure flexibility of the resulting structures (Wriggers *et al.,* 2005). In this set of constructs, no translated polylinker sequences were present, unlike in pBFF constructs (discussed below). Klenow extension of annealed oligos was used to create these constructs, essentially as described (Uhlman, 1988), followed by *Xho*I/*Eco*RI digestion of the double- stranded DNA fragments: pSKNVirD2-204/264/373 C-term-Δω, LINLSTSFW and LINLSTSRV; and pSKNVirD2-204/264/ 373 NLS-ω, LINLSOFW and LINLSORV.

Construction of *A. tumefaciens* expression vectors pBFF and pBFFstop

For expression of the different VirD2 proteins in *A. tumefaciens*, the vectors pBFF and pBFFstop were constructed. The wide-host-range, non-mobilizable plasmid pBFF was based on plasmid pRL662 (Vergunst *et al.,* 2000) and constructed as follows: an undesirable *Xho*I site in the plasmid was removed by Klenow fill-in to give pBBR6ΔXho. This vector was digested with *Not*I and the 5′ overhangs were filled in with T4 DNA polymerase in the presence of dNTPs. After heat inactivation of the enzyme, vector DNA was precipitated and digested with *Xma*I. A 1,300-bp *Xma*I/*Sca*I fragment containing the promoter and the complete ORF of the VirF gene sequence of the octopine type *A. tumefaciens* (Ti plasmid pTiB6) was cloned into the vector to give pBBR6ΔXhoVirF. Subsequently, most of the VirF coding sequence was removed by digestion with *Bsp*HI, which cuts in the VirF ATG translational start codon, and *Xho*I, which cuts just upstream of the 37-amino-acid VirF C-terminal T4SS transfer domain (Vergunst *et al.,* 2000). A double-stranded FLAG oligomer consisting of annealed FLAG NotFW and FLAG-NotRV was digested with *Nco*I and *Xho*I and cloned into the *Bsp*H1/*Xho*I-digested vector to give the final plasmid pBFF, which thus combines a unique *Not*I cloning site between an N-terminal FLAG-tag and F (Fig. 3). In pBFF, the truncated VirD2 protein was expressed as FLAG-NLSVirD2 truncation F, including 30 amino acids between the VirD2 truncation and F as a result of translation of

polylinker sequences from pSKN. In pBFFstop, the stop codon was in the polylinker of pBFF, resulting in proteins that end on 31 random amino acids encoded by the polylinker. The presence of this random amino acid sequence did not affect transformation efficiency (data not shown). All FLAG-NLS-VirD2 constructs were expressed under the control of the *virF* promoter, which performed equally well as the *virD* promoter (data not shown). To create pBFFstop, pBFF was cut with *Not*I and a double-stranded *Not*I-compatible DNA fragment harboring several stop codons was introduced at this site, restoring the single *Not*I site that was originally present but prohibiting any in-frame read-through into the VirF C-terminal sequence. The oligos annealed to generate this fragment were FW stop and RV stop.

To create pBRRVirD2, which expresses VirD2 protein under control of the *virF* promoter, without any N- or C-terminal fusions, pBBR6ΔXhoVirF was digested with *Bsp*HI, treated with Klenow to fill in recessed ends, and subsequently digested with *Xho*I. A PCR product was generated with the primers D2StuIFW and D2XhoIRV. The amplicon was digested with *Stu*I and *Xho*I and inserted into the vector. pBFF-VirD2stop, which expresses an N-terminally FLAG and SV40 NLS-tagged VirD2 protein with a wild-type C-terminal sequence, thus without any VirF sequences or amino acids arbitrarily encoded by polylinker sequences, was created by digesting pBFFstopVirD2 and pBRRVirD2 with *Drd*I and *Rsr*II. The 4,413-bp fragment of pBRRVirD2 was then ligated to the 1,240-bp *Drd*I/*Rsr*II-fragment of pBFFVirD2, restoring the VirD2 ORF.

pSDM3149 (A. Vergunst, A. Den Dulk-Ras, and P. J. J. Hooykaas, *unpublished*) was based on plasmid pRL662 (Vergunst *et al.,* 2000). Essentially, it is a wide-host-range, non-mobilizable plasmid containing the *virD* promoter and *virD2* gene as amplified from Ti plasmid pTiA6 (Rossi *et al.,* 1993). For root transformation experiments, the plasmids were transferred to *A. tumefaciens* LBA2585 (GV3101, pPM2260, Δ*virD2*, and ΔT-DNA) (Bravo-Angel *et al.,* 1998) by electroporation. Vector pCAMBIA2301 was used as T-stand donor (T-DNA contains GUS::intron and the kanamycin resistance marker). pCAMBIA2301 was transferred to LBA2585 by triparental mating (Ditta *et al.,* 1980).

Root transformation assay in *Arabidopsis thaliana* ecotype C-24

Root transformation experiments were performed as described earlier (Vergunst *et al.,* 2000). In short, seedlings of *Arabidopsis thaliana* ecotype C-24 were grown for 10 days in liquid medium, after which roots were harvested and pre-incubated for 3 days on

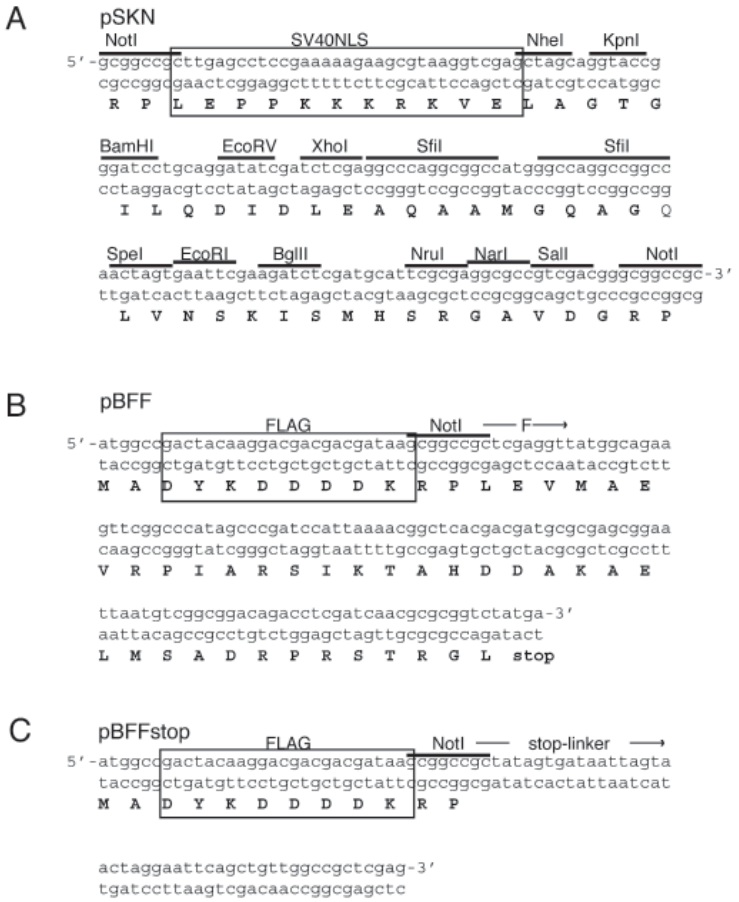

Figure 4: A, pSKN; **B,** pBFF; and **C,** pBFFstop vectors. Amino acids encoded by the top strand are indicated below the double-stranded sequence.

medium containing 2,4-dichlorophenoxyacetic acid at 0.5 ml/liter and kinetin at 0.05 mg/liter, followed by cocultivation with *Agrobacterium tumefaciens* LBA2585 containing the appropriate VirD2 donor vector and pCAMBIA2301 as T-DNA donor for 2 days on solid medium containing 2,4-dichlorophenoxyacetic acid at 0.5 ml/liter, kinetin at 0.05 mg/liter, and 20 μM acetosyringone. After this period, roots were either stained for GUS activity or placed on medium containing N6-(isoprofenyl)adenine at 5 mg/liter and indole 3-acetic acid at 0.15 mg/liter and, for selection, kanamycin at 50 mg/liter and timentin at 100 mg/liter to kill remaining *A. tumefaciens*.

Kanamycin-resistant calli were counted either after 2 and 3 weeks (WT VirD2) or after 3 and 4 weeks (others) for several independent experiments (more details can be found in the figure captions). The highest numbers of the pairs of calli counts were used

for further calculations regarding the potency of different mutants to lead to transfer and integration of T-strands. In that manner, obvious false differences were excluded because these were predominantly caused by fusion of closely neighboring calli on root segments in positive controls, which should lead to an underestimation of callus formation. Data were normalized against values for WT VirD2 (from *A. tumefaciens* LBA2585 containing pSDM3149 and pCAMBIA2301).

For the determination of transient transformation events, at least 100 root explants per construct per experiment were stained for a few hours or overnight in staining solution containing X-gluc after 3 days of cocultivation with *Agrobacterium* strains.

Western blot assay

Agrobacterium cultures were induced with 3 μM acetosyringone for at least 4 h before being harvested. A crude protein extract was made by boiling *Agrobacterium* cells in Laemmli buffer with β-mercapto-ethanol for 10 min. Insoluble proteins and debris were spun down and the supernatant was used for Western blotting. Proteins were separated on 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis gels, transferred to nitrocellulose membrane, and treated with FLAG antibody conjugated to alkaline phosphatase (cat no. A9469; Sigma- Aldrich). Detection of alkaline phosphatase activity was performed using nitroblue tetrazolium–5-bromo-4-chloro-3-indolyl phosphate solution (Sigma-Aldrich) according to instructions from the manufacturer.

Table 4. Primer and DNA oligomer sequences

Primer	**Sequence**[a]
A (*Sac*I-*Xho*I F)	GCGGCCGCTTGAGCCTCCGAAAAAGAGCGTAAGGTCGAGCTAGCGGTACCGGGATCCTGCAGGATATCGATCTCGAGGCCC
B (*Sac*I-*Xho*I R)	GGGCCTCGAGATCGATATCCTGCAGGATCCCGGTACCTGCTAGCTCGACCTTACGCTTCTTTTTCGGAGGCTCAAGCGGCCGCCAGCT
C (*Xho*I-*Kpn*I F)	TCGAGGCCCAGGCGGCCATGGGCCAGGCCGGCCAACTAGTGAATTCGAAGATCTCGATGCATTCGCGAGGCGCCGTCGACGGGCGGCCGCGTAC
D (*Xho*I-*Kpn*I R)	GCGGCCGCCCGTCGACGGCGCCTCGCGAATGCATCGAGATCTTCGAATTCACTAGTTGGCCGGCCTGGCCCATGGCCGCCTGGGCC
D2NheF	GCGAGCTAGCGATGCCCGATCGCGCTCAAG
D2XhoR	GCGACTCGAGGGTCCCCCCGCGCCCATC
D2FW	CCTGGATGCGACTTCGCGAGC
D2-264RV	CATACTCGAGGGACTCCTGGAAACGGGCG
VirD2-373RV	CATACTCGAGGCGAGTGTCTTGCTCACCGAC
LiNLSTSFW	GATCCTCGAGGGAGGTGGTGGAAGTAAGCGTCCGCGTGACCGTCACGATGGGAATTGGGTG
LINLSTSRV	CTAGGAATTCCTAGCGACGATTACCTCTTGCACGTTTGCGTCCACCCAATTCTCCATCGTGA
LINLSOFW	GGATCCTCGAGGGAGGTGGTGGAAGTAAGCGTCCGCGTGACCGTCACGATGGAGAATTGGGTGGAC
LINLSORV	CCTAGGAATTCCTAGGTCCCCCCGCGCCCATCTCTTGCACGTTTGCGTCCACCCAATTCTCCATCG
VirD2NotI	GGATCGCGGCCGCTTGAGCTAGCGATGCCC
VirD2RV	CCATGGCGGCCGCCCGTCGACGGCG
FLAG-NotFW	GACGCTCTGTTTCTCTCACCACAGCCATGGCCGACTACAAGGACGACGACGATAAGCGGCCGCTCGAGACTAGTGAGCT
FLAG-NotRV	ACTAGTCTCGAGCGGCCGCTTATCGTCGTCGTCCTTGTAGTCGGCCATGGCTGTGGTGAGAGAAACAGAGCGTCAGCT
FWstop	GGCCGCTATAGTGATAATTAGTAACTAGGAATTCAGCTGTT
RVstop	GGCCAACAGCTGTATTCCTAGTTACTAATTATCACTATAGC
D2StuIFW	GAAGGCCTGATCGCGCTCAAGTAATCATTCGC
D2XhoIRV	GACTCGAGCTAGGTCCCCCCGCGCCCAT

[a]Initial VirD2 truncations were created in the high copy vector pSKN (a bluescript derivative) and subcloned into destination vectors pBFF and pBFFstop as*Not*I fragments.

References

Atmakuri, K., Cascales, E., Burton, O. T., Banta, L. M., and Christie, P. J. 2007. *Agrobacterium* ParA/MinD-like VirC1 spatially coordinates early conjugative DNA transfer reactions. EMBO (Eur. Mol. Biol. Organ.) J. 26:2540-2551.

Bakó, L., Umeda, M., Tiburcio, A.F., Schell, J., and Koncz, C. 2003. The VirD2 pilot protein of *Agrobacterium*-transferred DNA interacts with the TATA box-binding protein and a nuclear protein kinase in plants. Proc. Natl. Acad. Sci. U.S.A. 100:10108-10113.

Bhattacharjee, S., Lee, L. Y., Oltmanns, H., Cao, H., Veena, Cuperus, J., and Gelvin, S. B. 2008. MPa-4, an *Arabidopsis* importin α isoform, is preferentially involved in *Agrobacterium*-mediated plant transformation. Plant Cell 20:2661-2680.

Bravo-Angel, A. M., Hohn, B., and Tinland, B. 1998. The omega sequence of VirD2 is important but not essential for efficient transfer of T-DNA by *Agrobacterium tumefaciens*. Mol. Plant-Microbe Interact. 11:57-63.

Bundock, P., den Dulk-Ras, A., Beijersbergen, A., and Hooykaas, P. J. 1995. Trans-kingdom T-DNA transfer form *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*. EMBO (Eur. Mol. Biol. Organ.) J. 14:3206-3214.

Cascales, E., and Christie, P. J. 2003. The versatile bacterial type IV secretion systems. Nat. Rev. Microbiol. 1:137-149.

Chen, I., Christie, P. J., and Dubnau, D. 2005. The ins and outs of DNA transfer in bacteria. Science 310:1456-1460.

Christie, P. J. 2004. Type IV secretion: the *Agrobacterium* VirB/D4 and related conjugal systems. Biochem. Biophys. Acta 1694:219-234.

Christie, P. J., and Vogel, J. P. 2000. Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. Trends Microbiol. 8:354-360.

Christie, P. J., Atmakuri, K., Krishnamoorthy, V., Jakubowski, S., and Cascales, E. 2005. Biogenesis, architecture and function of bacterial type IV secretion systems. Annu. Rev. Microbiol. 59:451-485.

Citovsky, V., Wong, M. L., and Zambryski, P. 1989. Cooperative interaction of *Agrobacterium* VirE2 protein with single-stranded DNA: Implications for the T-DNA transfer process. Proc. Natl. Acad. Sci. U.S.A. 86:1193-1197.

Deng, W., Chen, L., Wood, D. W., Metcalfe, T., Liang, X., Gordon, M. P., Comai., L., and Nester, E. W. 1998. *Agrobacterium* VirD2 protein interacts with plant host cyclophilins. Proc. Natl. Acad. Sci. U.S.A. 95:7040-7045.

Ditta, G., Stanfield, S., Corbin, D., and Helinski, D. R. 1980. Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of *Rhizobium meliloti*. Proc. Natl. Acad. Sci. U.S.A. 77:7347-7351.

Draper, O., César, C. E., Machón, C., de la Cruz, F., and Llosa, M. 2005. Site-specific recombinase and integration activities of a conjugative relaxase in recipient cells. Proc. Natl. Acad. Sci. U.S.A. 102:16385-16390.

Dunn, J. J., Krippl, B., Bernstein, K. E., Westphal, H., and Studier, F. W. 1988. Targeting bacteriophage T7 RNA polymerase to the mammalian cell nucleus. Gene 68:259-266.

Dürrenberger, F., Crameri, A., Hohn, B., and Koukolíková-Nicola, Z. 1989. Covalently bound VirD2 protein of *Agrobacterium tumefaciens* protects the T-DNA from exonucleolytic degradation. Proc. Natl. Acad. Sci. U.S.A. 86:9154-9158.

García-Rodríguez, F. M., Schrammeijer, B., and Hooijkaas, P. J. 2006. The *Agrobacterium* VirE3 effector protein: a potential plant transcriptional activator. Nucleic Acids Res. 34:6469-6504.

Grange, W., Duckely, M., Husale, S., Jacob, S., Engel, A., and Hegner, M. 2008. VirE2: a unique ssDNA-compacting molecular machine. PLoS Biol. 6:e44.

De Groot, M. J., Bundock, P., Hooykaas, P. J., and Beijersbergen, A. G. 1998. *Agrobacterium*-mediated transformation of filamentous fungi. Nat. Biotechnol. 16:839-842.

Guo, M., Hou, Q., Hew, C. L., and Pan, S. Q. 2007a. *Agrobacterium* VirD2-binding protein is involved in tumorigenesis and redundantly encoded in conjugative transfer gene clusters. Mol. Plant-Microbe Interact. 10:1201-1212.

Guo, M., Jin, S., Sun, D., Hew, C. L., and Pan, S. Q. 2007b. Recruitments of conjugative DNA transfer substrate to *Agrobacterium* type IV secretion apparatus. Proc. Natl. Acad. Sci. U.S.A. 104:20019-20024.

Hood, E. E., Gelvin, S. B., Melchers, and L. S., Hoekema, A. 1993. New *Agrobacterium* helper plasmids for gene transfer to plants. Transgen. Res. 2:208-218.

Hooykaas-van Slogteren, G. M., Hooykaas, P. J., and Schilperoort, R. A. 1992. Expression of Ti plasmid genes in monocotyledonous plants infected with *Agrobacterium tumefaciens*. Biotechnology 24:382-383.

Howard, E. A., Zupan, J. R., Citovsky, V., and Zambryski, P. C. 1992. The VirD2 protein is *A. tumefaciens* contains a C-terminal bipartite nuclear localisation signal: implications for nuclear uptake of DNA in plant cells. Cell 68:109-118.

Jayaswal, R. K., Veluthambi, K., Gelvin, S. B., and Slightom., J. L. 1987. Double-stranded cleavage of T-DNA and generation of single-stranded T-DNA molecules in *Escherichia coli* by a *virD*-encoded border-specific endonuclease from *Agrobacterium tumefaciens*. J. Bacteriol. 169:5035-5045.

Koukolíková-Nicola, Z., Raineri, D., Stephens, K., Ramos, C., Tinland, B., Nester, E. W., and Hohn, B. 1993. Genetic analysis of the *virD* operon of *Agrobacterium tumefaciens*: a search for functions involved in transport of T-DNA into the plant cell nucleus and in T-DNA integration. J. Bacteriol. 175:723-731.

Lessl, M., and Lanka, E. 1994. Common mechanisms in bacterial conjugation and Ti-mediated T-DNA transfer to plant cells. Cell 77:321-324.

Llosa, M., Gomez-Rüth, F. X., Coll, M., and de la Cruz, F. 2002. Bacterial conjugation: a two-step

mechanism for DNA transport. Mol. Microbiol. 45:1-8.
Mysore, K. S., Bassuner, B., Deng, X. B., Darbinian, N. S., Motchoulski, A., Ream, W., and Gelvin, S. B. 1998. Role of the *Agrobacterium tumefaciens* VirD2 protein in T-DNA transfer and integration. Mol. Plant-Microbe Interact. 11:668-683.
Regensburg-Tuïnk, A. J., and Hooykaas, P. J. 1993. Transgenic *N. glauca* plants expressing bacterial virulence gene virF are converted into hosts for nopaline strains of *A. tumefaciens*. Nature 363:69-71.
Relić, B., Andjelković, M., Rossi, L., Nagamine, Y., and Hohn, B. 1998. Interaction of the DNA modifying proteins VirD1 and VirD2 of *Agrobacterium tumefaciens*: analysis by subcellular localization in mammalian cells. Proc. Natl. Acad. Sci. U.S.A. 95:9105-9110.
Rossi, L., Hohn, B., and Tinland, B. 1993. The VirD2 protein of Agrobacterium tumefaciens carries nuclear localization signals important for transfer of T-DNA to plants. Mol. Gen. Genet. 239:345-353.
Scheiffele, P., Pansegrau, W., and Lanka, E. 1995. Initiation of *Agrobacterium tumefaciens* T-DNA processing. Purified proteins VirD1 and VirD2 catalyze site- and trans-specific cleavage of superhelical T-border DNA *in vitro*. J. Biol. Chem. 270:1269-1276.
Schrammeijer, B., den Dulk-Ras, A., Vergunst, A. C., Jurado Jacome, E., and Hooykaas, P. J. 2003. Analysis of Vir protein translocation from *Agrobacterium tumefaciens* using *Saccharomyces cerevisiae* as a model: evidence for transport of a novel effector protein VirE3. Nucleic Acids Res. 31:860-868
Shurvinton, C. E., Hodges, L., and 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. U.S.A. 89:11837-11841.
Steck, T. R., Lin, T. S., and Kado, C. I. 1990. VirD2 gene product form the nopaline plasmid pTiC58 has at least two activities required for virulence. Nucleic Acids Res. 18:6953-6958.
Tato, I., Zunzunegui, S., de la Cruz, F., and Cabezon, E. 2005. TrwB, the coupling protein involved in DNA transport during bacterial conjugation, is a DNA-dependent ATPase. Proc. Natl. Acad. Sci. U.S.A. 102:8156-8161.
Tinland, B., Schoumacher, F., Gloeckler, V., Bravo-Angel, A. M., and Hohn, B. 1995. The *Agrobacterium tumefaciens* virulence D2 protein is responsible for precise integration of T-DNA into the plant genome. EMBO (Eur. Mol. Biol. Organ.) J. 14:3585-3595.
Uhlmann, E. 1988. An alternative approach in gene synthesis: use of long selfpriming oligodeoxynucleatides for the construction of doublestranded DNA. Gene 71:29-40.
Vergunst, A. C., van Lier, M. C. M., den Dulk-Ras, A., Stüve, T. A., Ouwehand, A., and Hooykaas, P. J. 2005. Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of *Agrobacterium*. Proc. Natl. Acad. Sci. U.S.A. 102:832-837.
Vergunst, A. C., Schrammeijer, B., den Dulk-Ras, A., de Vlaam, C. M., Regensburg-Tuïnk, T. J., and Hooykaas, P. J. 2000. VirB/D4-dependent protein translocation from *Agrobacterium* into plant cells. Science 290:979-982.
Vogel, A. M., and Das, A. 1992. Mutational analysis of *Agrobacterium tumefaciens virD2*: tyrosine 29

is essential for endonuclease activity. J. Bacteriol. 174:303-308.
Wang, P., and Heitman, J. 2005. The cyclophylins. Genome Biol. 6:226. Wriggers, W., Chakravanarty, S., and Jennings, P. A. 2005. Control of protein functional dynamics by peptide linkers. Biopolymers 80:736-746.
Zhu, J., Oger, P. M., Schrammeijer, B., Hooykaas, P. J., Farrand, S. K., and Winans, S. C. 2000. The bases of crown gall tumorigenesis. J. Bacteriol. 182:3885-3895.
Ziemienowicz, A., Tinland, B., Bryant, J., Gloeckler, V., and Hohn, B. 2000. Plant enzymes but not *Agrobacterium* VirD2 mediate T-DNA ligation in vitro. Mol. Cell Biol. 20:6317-6311.

Chapter 3

The DUF domain of VirD2 determines recruitment of VirD2 to the Type 4 Secretion System within *Agrobacterium tumefaciens*

Maartje van Kregten, Frederik Spanhoff, Paul J.J. Hooykaas and Bert J. van der Zaal

Abstract

VirD2 is the *Agrobacterium tumefaciens* protein responsible for T-strand processing. Besides its relaxase domain, it contains a large domain of unknown function. In a previous study we have shown that this DUF domain can be replaced by the very short C-terminal Type 4 Secretion System translocation sequence of the unrelated protein VirF. We have confirmed this result here in *Nicotiana glauca* and *Kalanchoë daigramontiana*. Also, we have studied the subcellular localization of GFP-tagged VirD2 constructs. In the absence of DUF, the VirD2 protein is randomly distributed in the bacterium, but that full length VirD2 shows a localization at the cell poles in more than 60% of the cells. Polar localization is determined for the most part by the N-terminal 60 amino acids of DUF. Surprisingly, the short C-terminus of VirF is able to substitute for DUF in targeting truncations of VirD2 to the cell poles. Polar localization of VirD2 truncations correlates with the virulence of the corresponding *Agrobacterium* strain.

Introduction

Agrobacterium tumefaciens causes crown gall disease in dicotyledonous plants by transferring some of its own DNA (the transferred or T-DNA) into these plants, where it integrates into the plant's genome. The T-DNA, situated on the large tumor-inducing plasmid (pTi) contains genes causing the plant cells to divide rapidly, forming a tumorous overgrowth, and to produce opines, which *Agrobacterium* can use as carbon source (recently reviewed in Pitzschke and Hirt, 2010). In the laboratory, it was shown that *Agrobacterium* can mediat DNA transfer also to monocotyledonous plants (Hooykaas – van Slogteren *et al.*, 1991), yeast and other fungi (Bundock *et al.*, 1995, de Groot *et al.*, 1998).

The T-DNA is delimited by a 24 basepair direct repeat: the left and right border (LB and RB) sequences (Yadav *et al.*, 1982). After it was discovered that any DNA sequence present between the LB and RB would be transferred to the plant recipient, the binary vector system was developed, in which the LB and RB are present on an extra plasmid, instead of on the large (~200kb) pTi, which is difficult to manipulate (Hoekema *et al.*, 1983). Since then, *Agrobacterium*-mediated transformation (AMT) has become one of the

preferred methods of plant and fungal transformation.

AMT starts upon the perception of phenolic compounds, released by wounded plant cells, by *Agrobacterium*. This triggers a signaling cascade which results in expression of bacterial virulence (*vir*) genes (reviewed in Zhu *et al.*, 2000, Pitzschke and Hirt, 2010). For T-DNA processing, the VirD1 and VirD2 proteins are essential. VirD2 is a relaxase, able to nick the border sequences to release a single-stranded DNA molecule called the T-strand. The VirD1 protein is supposed to aid by local unwinding of the DNA (Jayaswal *et al.*, 1987, Lessl and Lanka, 1994, Scheiffele *et al.*, 1995). After processing, VirD2 remains covalently attached to the 5' end of the T-strand to form the T-complex (Dürrenberger *et al.*, 1989, Scheiffele *et al.*, 1995, Vogel and Das, 1992, reviewed in Lanka and Wilkins, 1995, and Gelvin, 2000). VirD2 is, as such, responsible for the translocation of the T-complex to the recipient cell (Van Kregten *et al.*, 2009, Llosa *et al.*, 2002). Translocation occurs via the Type IV Secretion System (T4SS) formed by the VirB1-11 proteins and the coupling factor VirD4 (reviewed in e.g. Christie, 2004, Llosa *et al.*, 2002).

Apart from the T-complex, other Vir proteins like VirE3 (Schrammeijer *et al.*, 2003), VirE2 and VirF (Vergunst *et al.*, 2000), and VirD5 (Vergunst *et al.*, 2005) are translocated to the recipient species during AMT. All these translocated proteins probably interact directly or indirectly with the VirD4 protein, which is the first to act in the process of translocation (Cascales and Christie, 2004). Since the VirD4 protein (Kumar and Das, 2002) and the VirB proteins forming the T4SS proper (Judd *et al.*, 2005) are found at one or both poles of the rod-shaped *Agrobacterium* cells, polar localization of transport substrates might be expected as well. Recently, using deconvolution microscopy, components and substrates of the T4SS were found to be located around the perimeter of the cell in a helical fashion, rather than at the poles (Aguilar *et al.*, 2010). For VirE2, which was shown to interact with VirD4 in an immunoprecipitation assay, polar localization has indeed been demonstrated, but to our knowledge this has not been demonstrated for other translocated Vir proteins. Although attempts have been made, an interaction between VirD2 and VirD4 has not yet been found, suggesting that their interaction is weak or that another binding partner is involved (Atmakuri *et al.*, 2003).

In a previous study (Van Kregten et al., 2009), we have focused on the functional properties of the large domain of unknown function (DUF) which, except for some short terminal sequences, forms the larger part of C-terminal half of VirD2 (Fig. 1). The DUF is relatively unstructured and is poorly conserved between VirD2 proteins from related *Agrobacterium* species (Howard *et al.*, 1992). As we demonstrated, the N-terminal half of

VirD2, consisting of the relaxase domain, can successfully be used for AMT of *Arabidopsis thaliana* root explants when supplemented with the C-terminal T4SS translocation signal of VirF (VirD2-204-F). This means that the DUF domain and the C-terminal sequences, in total about half of the protein, can be replaced by a very short (37 amino acids) sequence from a protein that is not involved in T-DNA processing. However, when only short C-terminal sequences of VirD2 were fused to the relaxase domain VirD2-204, such as combinations of the nuclear localization signal (NLS) (Rossi *et al.*, 1993), the arginine-rich putative T4SS translocation signal (Vergunst *et al.*, 2005), and the omega (ω) sequence (Bravo-Angel *et al.*, 1998, Mysore *et al.*, 1998, Shurvinton *et al.*, 1992), AMT remained severely hampered. Remarkably, AMT was restored when a stretch of amino acids corresponding to residues 205 to 264 of VirD2 was fused in between the relaxase domain and the short C-terminal VirD2 sequence, providing compelling evidence that these 60 amino acids play an important role in AMT within the native VirD2 protein (Van Kregten et al., 2009).

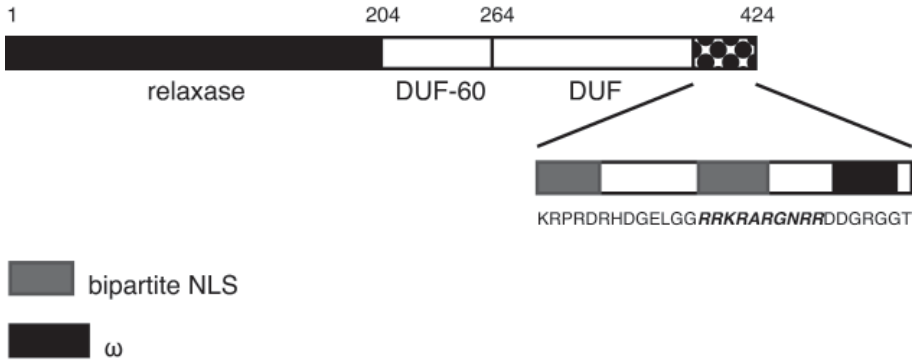

Figure 1: Schematic overview of the domains of VirD2. The N-terminal half of VirD2 consists of the relaxase domain (Scheiffele *et al.*, 1996, Steck *et al.*, 1990). The C-terminal half encompasses the domain of unknown function (DUF), of which DUF-60 is functionally important (Van Kregten *et al.* 2009). The C-terminal domain consists of a bipartite NLS, the arginine-rich sequence important in translocation, and the ω sequence (Howard *et al.*, 1992, Vergunst *et al.*, 2005, Shunvinton *et al.*, 1992).

In the present study, we wanted to investigate whether the importance of the 205-264 region of the DUF domain (DUF-60) for AMT was also evident when using other plant species than *Arabidopsis thaliana*. In addition to that, we wanted to gain more insight in the precise function of the DUF-60 domain. The VirD2 region downstream of amino acid residue 204 is not required for the relaxase reaction (Scheiffele *et al.*, 1995, Steck *et al.*, 1990, Van Kregten et al., 2009), but it may be involved in contacting elements of the T4SS

or in recruitment to this system. However, no interaction between VirD2 and VirD4, the coupling protein of the T4SS, has been found (Atmakuri *et al.*, 2007), but through its DUF-60 domain may still interact with another component of the T4SS and thus be involved in polar localization of the T-complex.

Our current results provide evidence that the DUF-60 function is not restricted to AMT of *Arabidopsis*, but can also be observed in AMT of *Nicotiana glauca* and *Kalanchoë daigramontiana*. By using GFP-tagged VirD2 constructs, we found that DUF-60 is crucial for polar localization of GFP-tagged VirD2. The short T4SS translocation signal of VirF can substitute for DUF-60 in inducing polar localization of VirD2. The relevance of these findings is discussed.

Results

DUF-60 can be replaced by the C-terminal end of VirF in plant transformation

The DUF domain of VirD2 is dispensable for transformation of *Arabidopsis* when the T4SS translocation signal of VirF is fused to C-terminally truncated versions of VirD2, such as VirD2-204 (Van Kregten *et al.*, 2009). Fusion of the native C-terminal end of VirD2 to VirD2-204 resulted in only a very low transformation efficiency, but addition of DUF-60 (60 amino acids located adjacent to amino acids 204) in between VirD2-204 and the VirD2-derived C-terminal peptide restored transformation efficiency to normal levels (Van Kregten *et al.*, 2009). Apparently, the C-terminal translocation signal of VirD2 needs assistance from DUF-60 to fulfill its function.

To investigate if the effect of DUF-60 in other plant species is similar to the effect observed in *Arabidopsis*, several VirD2 truncations were made, all N-terminally tagged with the FLAG tag and the SV40 NLS (omitted in names of constructs for brevity). The FLAG tag was used for easy detection of protein expression and the SV40 NLS to ensure nuclear localization. A schematic overview of the constructs used in this study is shown in Fig. 2. We created VirD2-204, in which the C-terminal half is completely deleted, and VirD2-264, in which DUF-60 is included, but the rest of the C-terminus is deleted. These constructs were also supplemented with elements of the C-terminal end of VirD2, to create VirD2-CtermΔω and VirD2-264-CtermΔω. This C-terminal end lacks the 5 C-terminal amino acids, but is sufficient to enable translocation (Van Kregten *et al.*, 2009). The constructs

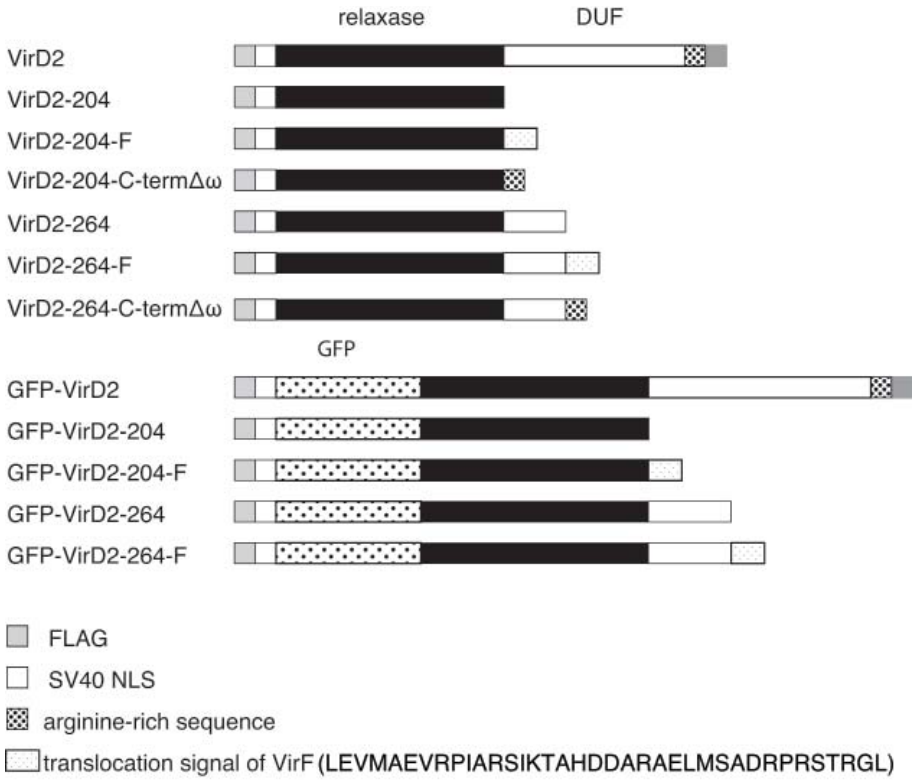

Figure 2: A schematic overview of the proteins used in this study.

were also supplemented with the T4SS translocation signal of VirF, to create VirD2-204-F and VirD2-264-F.

We inoculated *Nicotiana glauca, Kalanchoë daigramontiana, Solanum lycopersicum* and *Helianthus annuus* with *Agrobacterium* strain LBA2569 (Δ*virD2*, ΔT-DNA, Vergunst, Den Dulk-Ras and Hooykaas, unpublished) expressing the relevant VirD2 truncations or LBA1010 as positive control (wild type strain, Koekman *et al.*, 1982). Tumor formation was photographed after three (*N. glauca*) or six (*K. daigramontiana*) weeks. *H. annuus* and *S. lycopersum* only formed tumors when inoculated with the wild type control *Agrobacterium* strain (data not shown).

Agrobacterium strains that expressed the VirD2 truncations VirD2-204 or VirD2-264 induced no tumor formation in *N. glauca* or *K. daigramontiana*. Tumors were induced by *Agrobacterium* strains expressing VirD2-204-F and VirD2-264-F, the truncations that were complemented with the C-terminal 37 amino acids of VirF. An *Agrobacterium* strain expressing VirD2-264-CtermΔω caused smaller tumors than strains expressing VirD2-

204-F and VirD2-264-F. An *Agrobacterium* strain expressing VirD2-204-CtermΔω caused smaller tumors than an *Agrobacterium* strain expressing VirD2-264-CtermΔω (Fig. 3, Table 1). Apparently, as previously found for *Arabidopsis* root transformations (Van Kregten *et al.*, 2009), the C-terminal sequence of VirF can also replace the C-terminal half of VirD2 in tumor formation on other plant species. Hence, just as for *Arabidopsis*, the DUF domain of VirD2 is not required for transformation of *N. glauca* or *K. daigramontiana*. By combining our previous results with the newly gathered data, it seems that DUF does not play a role as host range factor in the tested plant species.

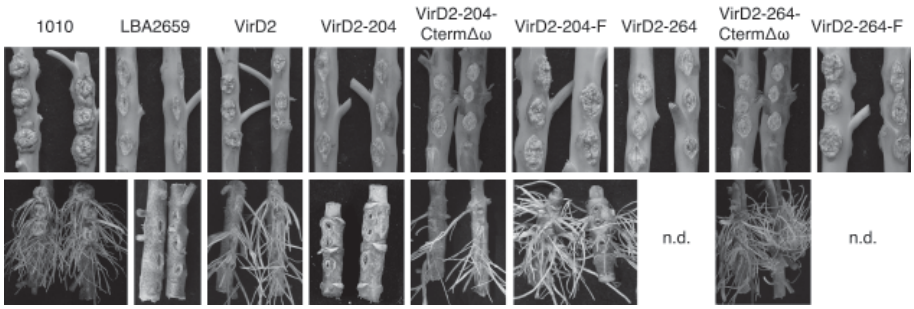

Figure 3: Tumor assay in *Nicotiana glauca* (top) and *Kalachoë tubiflora* (bottom). LBA1010 is a wild type *Agrobacterium* strain, LBA2569 is LBA1010Δ*virD2*, and is used for the expression of VirD2 mutants. All VirD2 constructs are N-terminally tagged with FLAG and NLS. Left to right: LBA1010, LBA2569, LBA2569 expressing VirD2, LBA2569 expressing VirD2-204, LBA2569 expressing VirD2-204-CtermΔω, LBA2569 expressing VirD2-204-F, LBA2569 expressing VirD2-264, LBA2569 expressing VirD2-264-CtermΔω, LBA2569 expressing VirD2-264-F Tumors were photographed after 4 weeks (*N. glauca*) and 6 weeks (*K. daigramontiana*). n.d.: not determined.

DUF-60 and the C-terminus of VirF direct GFP-VirD2 to the cell poles

Since our results pointed towards a function of DUF-60 as part of the T4SS translocation signal within *Agrobacterium* cells, we went on to investigate if DUF-60 may function as a factor influencing recruitment of VirD2 to the T4SS. We thus fused GFP to the VirD2 truncations in order to visualize the subcellular localization of these proteins in *virD2* deleted *Agrobacterium* strain LBA2569. GFP-VirD2, GFP-VirD2-204, and GFP-VirD2-264 were expressed under control of the *virF* promoter. As control, free GFP was expressed in *Agrobacterium* strain LBA1100 (ΔT-DNA, Beijersbergen *et al.*, 1992), from plasmid pSDM1761 (Bloemberg *et al.*, 2000). After induction of the *Agrobacterium* cells

Table 1

construct[a]	plant species	
	N. glauca	*K. daigramontiana*
LBA1010	+++	+++
LBA2569	-	-
VirD2	+	+
VirD2-204	-	-
VirD2-204-CtermΔω	+/-	+/-
VirD2-204-F	++	++
VirD2-264	-	n.d.[b]
VirD2-264- CtermΔω	+	+
VirD2-264-F	++	n.d.

[a]LBA1010 is a wild type *Agrobacterium* strain, LBA2569 is LBA1010Δ*virD2*. All VirD2 constructs are N-terminally tagged with FLAG and NLS. Tumor formation was scored after 3 (*N. glauca*) or 6 (*K. daigramontiana*) weeks.

[b]n.d.: not determined

with acetosyringone, localization of the fluorescent signal was determined by confocal microscopy.

As shown by the results in Table 2 and Figure 4, a control strain expressing free GFP showed fluorescent signal distributed throughout the entire cell. Cells expressing GFP-VirD2 showed polar localization in 63% of cells. Cells expressing GFP-VirD2-204, from which DUF is deleted, showed a dramatic drop in polar localization, to only 8%. In cells expressing GFP-VirD2-264, polar localization is restored to 50% of cells. Apparently, DUF-60 is very important for polar localization of VirD2.

The addition of the C-terminus of VirF restores to a large extent the ability of VirD2 truncations to mediate transformation of *Arabidopsis* root cells (Van Kregten *et al.*, 2009). To investigate if the C-terminal T4SS translocation signal of VirF, consisting of only 37 amino acid residues, is also sufficient to restore the normal pattern of polar localization of VirD2 truncations in induced *Agrobacterium* cells, we expressed GFP-VirD2-204-F in *Agrobacterium* strain LBA2569 and determined its subcellular localization. As shown in Table 2 and illustrated in Fig. 4, the addition of the C-terminal 37 amino acids of VirF to GFP-VirD2-204 restored the amount of cell showing polar localization to 41%. When both DUF-60 and the C-terminus of VirF are present, the percentage of polar localization of GFP-VirD2-264-F is restored to 62%, the level also found for GFP-VirD2. However, cells expressing GFP-VirD2-264-F do not show neat localization to the cell poles, but had

a somewhat spotted phenotype. Apart from fluorescent signal at the cell poles, they had additional spots in other places in the cell (e.g. the middle cell in the picture of GFP-VirD2-264-F in Fig. 4).

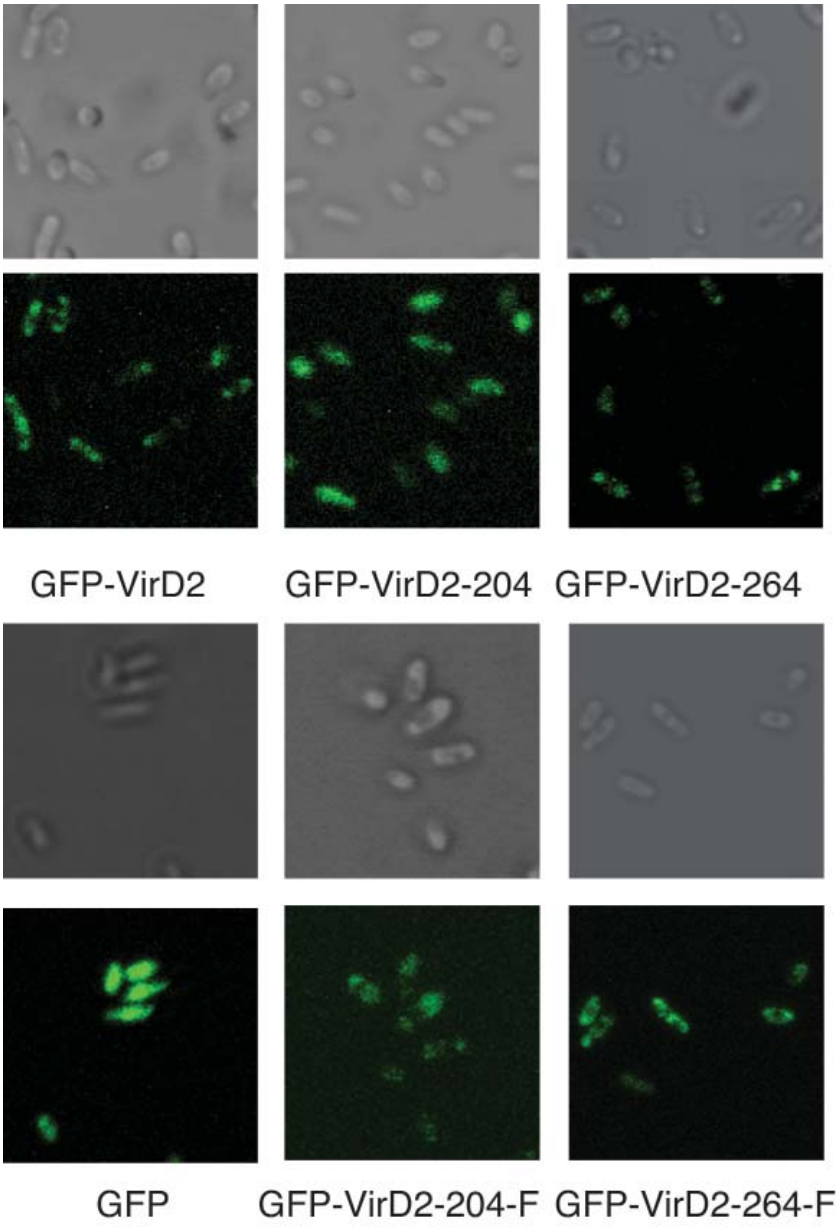

Figure 4: Localization of GFP-tagged VirD2 truncations in *Agrobacterium*. Upper pictures: visual field, lower pictures: fluorescence. The proteins produced in *Agrobacterium* strain LBA2569 (Δ*virD2*) are indicated below the images.

In summary, the data regarding the localization of VirD2 established that DUF-60 is crucial for determining polar localization of VirD2 within *Agrobacterium* cells. When DUF-60 is lacking, almost all polar localization disappears. The short C-terminal sequence of VirF can compensate for the lack of DUF-60, restoring the polar localization of VirD2. These results are discussed below.

Table 2: subcellular localization of VirD2 truncations in *Agrobacterium*

Agrobacterium strain expressing	percentage of cells showing:	
	polar localization*	signal in entire cell
GFP	0	100
GFP-VirD2-204	8	92
GFP-VirD2-264	50	50
GFP-VirD2	62	38
GFP-VirD2-204-F	41	59
GFP-VirD2-264-F	62**	38

* For every strain, at least 100 cells were scored.
** mostly with "spotted" phenotype

Discussion

Agrobacterium's VirD2 protein is remarkably multifunctional. It is responsible for the processing of the T-DNA (Jayaswal *et al.*, 1987, Lessl and Lanka, 1994, Scheiffele *et al.*, 1995). It is also responsible for the translocation of itself and the T-strand, to which it is covalently bound, through the T4SS of *Agrobacterium* into the cytosol of the recipient cell (Van Kregten *et al.*, 2009, Llosa *et al*,. 2002). Furthermore, it plays a role in nuclear import of the complex (reviewed in Gelvin, 2010).

In this study, we have further investigated the function of the large DUF domain in VirD2, making up almost half of its size (Fig. 1) and thus far a part of VirD2 to which a clear function has yet to be assigned. We focused on DUF-60, the N-terminal 60 amino acids of DUF, since a possible function of DUF seems to be contained within this region (Van Kregten *et al.*, 2009).

Our results concerning tumor formation in *N. glauca* and *K. daigramontiana* (Fig. 3 and Table 1) correlate with previous results obtained in *Arabidopsis* (Van Kregten *et al*, 2009). In all plant species tested so far, the requirements for successful transformation

are either DUF-60 and the C-terminal end of VirD2, or the C-terminus of VirF that are attached to the relaxase domain. The fact that the C-terminus of VirF can substitute for DUF-60 and the C-terminal end of VirD2 in very different plant species argues against a subtle function of DUF-60 in recipient plant cells. In case such a subtle function would exist, for instance via interaction with a particular host protein, different plant species would be expected to react differently upon the presence of DUF-60. For the VirF protein, such a host dependent attribution to virulence has been described (Hooykaas *et al.*, 1984, Tzfira *et al.*, 2004). The data presented here and previously strongly suggest that this is not the case for DUF-60. Obviously, just testing three different plant species would not suffice for drawing a definitive conclusion that DUF-60 does not exert any function in some plant species, had we not discovered that the function of DUF-60 lies within *Agrobacterium* cells.

If DUF-60 is important for the transformation process, but not required for a function in the host plant species, it is logical to assume it fulfills a function within *Agrobacterium* cells. Since the relaxase reaction does not require DUF, we focused our efforts on the subcellular localization of VirD2. A likely function of DUF-60 is recruitment of the T-complex to the T4SS. Since the T4SS is located at the poles of *Agrobacterium* cells, DUF-60-dependent polar localization would provide evidence for this hypothesis. Indeed, we found that deletion of DUF resulted in an almost complete loss of polar localization: in cells expressing GFP-VirD2-204, polar localization dropped to 8% of the cells, wherein GFP-VirD2 localized to the cell poles in 62% of cells (Table 2, Fig. 4). Our results on GFP-VirD2 are comparable to results found by other groups who found polar localization in 55% (Guo *et al.*, 2007b), and between 50 to 70% polar localization (Atmakuri *et al.*, 2007) of *Agrobacterium* cells expressing GFP-VirD2. In a recently published article, T4SS components and substrates were shown to be positioned in a helical array around the circumference of the cell, rather than at the cell poles (Aguikar *et al.*, 2010). It would be interesting to apply deconvolution microscopy, the technique used by Aguilar and co-workers, to the strains used in this study, especially for further investigation of GFP-VirD2-264, since the 'spotted' phenotype observed in strains expressing this construct is reminiscent of such helical arrays.

Polar localization was predominantly determined by DUF-60, as shown by construct GFP-VirD2-264, which displayed polar localization in 50 % of cells. We cannot exclude that a larger domain than DUF-60 mediates polar localization; however, a construct expressing GFP-DUF-60, which could theoretically demonstrate a modular DUF-60 potential for polar localization, proved to be toxic for *E. coli* cells, and could therefore not be prepared.

For all VirD2 truncations tested, the extent of polar localization of VirD2 correlated

with virulence of the host bacterium: constructs displaying polar localization and containing a T4SS translocation signal can be translocated and can cause transformation, whereas those not locating to the poles do not mediate transformation (Tables 1 and 2, Fig. 3 and 4, and Van Kregten *et al.*, 2009). This strongly indicates that polar localization is required for translocation via the T4SS of *Agrobacterium*.

Apparently, the translocation of VirD2 is a two-step process: recruitment of VirD2 to the cell poles and subsequent interactions with the T4SS resulting in translocation. For polar recruitment, DUF-60 is required, while for translocation to a recipient cell, sequences of the C-terminal end of VirD2 are required in addition. As can be inferred from our data, the short and continuous stretch of 37 C-terminal amino acids of VirF suffices for both polar localization and translocation. To our knowledge, possible interaction partners of VirF within *Agrobacterium* cells have never been described, so the mechanism of polar localization of VirF is currently unknown. In addition to that, not much is known about the recruitment to the T4SS of other translocated Vir proteins either. Only for VirE2 it was found that this protein interacts directly with the coupling protein VirD4 at the cell poles. If VirD4 is deleted, polar localization of VirE2 is lost as well (Atmakuri *et al.*, 2003). No data is available on domain requirements of VirE2 for interaction with VirD4. In a related system, the requirements for translocation of the Bep proteins of *Bartonella henselae* seem comparable to those of VirD2. An T4SS system largely similar to that of *Agrobacterium* exists in *Bartonella henselae*: the requirements for translocation of the effector proteins BepA-G are the presence of the Bep intracellular delivery (BID) domain and a C-terminal sequence rich in charged amino acids, which serves as T4SS translocation signal. Together, these two domain serve probably as a bipartite translocation signal (Schulein *et al.*, 2005).

To explain differences in requirements for translocation between VirD2 and VirF, we hypothesize that the presence of an extra domain required for translocation of VirD2 must serve an extra purpose as well. We propose that DUF-60 serves as an interaction domain. Several interactions between VirD2 and *Agrobacterium* proteins have been described. VirD2 interacts with VirC1 and the VirD2-Binding Proteins 1-3 (VBP1-3) (Atmakuri *et al.*, 2007, Guo *et al.*, 2007a). For all these factors, loss of interaction with VirD2 has been reported to result in loss of polar localization of VirD2 (Atmakuri *et al.*, 2007, Guo *et al.*, 2007a, Guo *et al.*, 2007b). Both VirC1 and VBP interactions seem equally required for polar localization. We hypothesize that DUF-60 represents a binding site for one or more of these interactants. The site of interaction between VirD2 and VirC1 has not been mapped yet, but VBP1 was shown to interact with the N-terminal half of VirD2, upstream of a *Bam*HI restriction site

within the VirD2 encoding region of the octopine-type VirD2 (Guo *et al.*, 2007a). This includes DUF-60. Of VirC1, it was shown that it not only interacts with VirD2, but also binds to the overdrive sequence on the Ti plasmid (Toro *et al.*, 1989). By these interactions, VirC1 could coordinate the T-strand processing reaction, ensuring that VirD2 can only be translocated after having processed a T-DNA by linking the T-strand processing reaction to polar localization.

We propose that, possibly with some extension into the relaxase domain, DUF-60 is the T4SS recruitment domain of VirD2. In that manner, DUF-60 mediates the polar localization of VirD2 due to its interaction with VirC1 and/or the VBP proteins. We propose that VirD2 has two domains for translocation; DUF-60, controlling recruitment of the protein to the T4SS, and a translocation signal, controlling the actual translocation process.

Acknowledgements

The authors would like to thank Gerda Lamers and Philippe Sakalis for invaluable help with microscopy.

Materials & methods

Bacterial strains and plasmids

Cloning steps were perfomed in *E. coli* strain DH10B. *Agrobacterium tumefaciens* wild type strain LBA1010 (Koekman *et al.*, 1982) and Δ*virD2* strain LBA2569 (C58, pTiB6, Δ*virD2*; Vergunst, A., Den Dulk-Ras, A. and Hooykaas, P.J.J., unpublished) were used for tumor formation assays. LBA2569 and LBA1100 (ΔT-DNA, Beijersbergen *et al.*, 1992) were used for microscopy studies.

Constructs used for tumor assays were described previously (Van Kregten *et al.*, 2009). For microscopy analysis, GFP was amplified from pEGFP (a gift from dr. De Boer, Add2X Biosciences) using primers GFP fw (GAATGCTCGAGGTGAGCAAGGGCGAGGAGCTG) and GFP rv (CTTACCTCGAGCTTGTACAGCTCGTCCATGCC). The product was digested with *Xho*I and ligated to similarly digested pSKN VirD2C (Van der Zaal, unpublished), to yield pSKN-GFP-VirD2. pSKN-GFP-VirD2 was digested with *Xmn*I and *Nde*I and the fragment of interest was ligated to similarly digested pSKN-VirD2-204 and pSKN-

VirD2-264 (described in Van Kregten *et al*., 2009) to yield pSKN-GFP-VirD2-204 and pSKN-GFP-VirD2-264.

All pSKN plasmids were digested with *Not*I, and fragments containing the genes-of-interest of the plasmids were ligated to similarly digested pBFF or pBFFstop (an *Agrobacterium* expression vector, allowing expression of the gene of interest under control of the acetosyringone-inducible *virF* promoter; described in Van Kregten *et al*., 2009), to yield pBFF-GFP-VirD2, pBFF-GFP-VirD2-204, pBFF-GFP-VirD2-264, and pBFFstop-GFP-VirD2, pBFFstop-GFP-VirD2-264, pBFFstop-GFP-VirD2-264. A schematic drawing of the proteins can be found in Fig. 2.

Microscopy

All constructs were electroporated into *Agrobacterium* strain LBA2569. Fluorescence was examined in cells from fresh overnight cultures induced with 200 μM acetosyringone. Aliquots of overnight cultures were resuspended in 0.9% NaCl solution, and visualized using a Zeiss Imager confocal microscope.

Tumor assay

Three- to four-week-old seedling of *Nicotiana glauca, Kalanchoë daigramontiana*, *Solanum lycopersum* and *Helianthus annus* were inoculated with 20 μL of an *Agrobacterium tumefaciens* suspension in physiological salt solution, at an OD of 0.8, from a fresh overnight culture. Tumors were photographed after four weeks (*N. glauca*), or six weeks (other species).

References

Aguilar, J., Zupan, J., Cameron, T.A. and Zambryski, P.C. 2010. *Agrobacterium* Type IV Secretion System and its substrates form helical arrays around the circumference of virulence-induced cells. *Proc. Natl. Acad. Sci. USA.* 107(8):3758-3763.

Atmakuri, K., Cascales, E., Burton, O.T., Banta, L.M., and Christie, P.J. 2007. *Agrobacterium* ParA/MinD-linked VirC1 spatially coordinates early conjugative DNA transfer reactions. EMBO (Eur. Mol. Biol. Organ.) J. 26(10):2540-2551.

Atmakuri, K., Ding, Z., and Christie, P.J. 2003. VirE2, a type IV secretion substrate, interacts with the VirD4 transfer protein at cell poles of *Agrobacterium tumefaciens*. Mol. Microbiol. 49(6):1699-1713.

Beijersbergen, A., den Dulk-Ras, A., Schilperoort, R.A., and Hooykaas, P.J.J. 1992. Conjugative transfer by the virulence system of *Agrobacterium tumefaciens*. Science. 256(5061):1324-1327.

Bloemberg, G.V., Wijfjes, A.H., Lamers, G.E., Stuurman, N., and Lugtenberg, B.J. 2000. Simultaneous imaging of *Pseudomonas fluorescens* WCS365 populations expressing three different autofluorescent proteins in the rhizosphere: new perspectives for studying microbial communities. Mol Plant-Microbe Interact. 13(11):1170-1176.

Bravo-Angel, A.M., Hohn, B., and Tinland, B. 1998. The omega sequence of VirD2 is important but not essential for efficient transfer of T-DNA by *Agrobacterium tumefaciens*. Mol. Plant-Microb. Interact. 11(1):57-63.

Bundock, P., den Dulk-Ras, A., Beijersbergen, A., and Hooykaas, P. J.J. 1995. Trans-kingdom T-DNA transfer from *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*. EMBO (Eur. Mol. Biol. Organ.) J. 14(13): 3206-3214.

Cascales, E., and Christie, P.J. 2004. Definition of a bacterial type IV secretion pathway for a DNA substrate. Science. 304(5674):1170-1173.

Christie, P.J. 2004. Type IV secretion: the *Agrobacterium* VirB/D4 and related conjugation systems. Biochem. Biophys. Acta. 1693(1-3):219-234.

De Groot, M. J., Bundock, P., Hooykaas, P. J.J., and Beijersbergen, A. G. 1998. *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. Nat Biotechnol. 16(9):839-842.

Dürrenberger, F, Crameri, A., Hohn, B., and Koukolikova-Nicola., Z. 1989. Covalently bound VirD2 of *Agrobacterium tumefaciens* protects the T-DNA from exonucleolytic degradation. Proc. Natl. Acad. Sci. U.S.A. 86(23):9154-9158.

Gelvin, S.B. 2000. *Agrobacterium* and plant genes involved in T-DNA transfer and integration. Annu. Rev. Plant Physiol. Plant Mol. Biol. 51:223-256.

Gelvin, S.B. 2010. Finding a way to the nucleus. Curr. Opin. Microbiol. 13(1):53-58.

Guo, M., Hou, Q., Hew, C.L., and Pan, S.Q. 2007a. *Agrobacterium* VirD2-binding protein is involved in tumorigenesis and redundantly encoded in conjugative transfer gene clusters. Mol Plant-Microbe Int. 20(10):1201-1212.

Guo, M., Jin, S., Sun, D., Hew, C.L., and Pan, S.Q. 2007b. Recruitment of conjugative DNA transfer substrate to *Agrobacterium* type IV secretion apparatus. Proc. Natl. Acad. Sci. U.S.A. 104(50):20019-20024.

Hooykaas, P.J.J., Hofker, M., Den Dulk-Ras, H., and Schilperoort, R.A. 1984. A comparison of virulence determinants in an octopine Ti plasmid, a nopaline Ti plasmid, and an Ri plasmid by complementation analysis of *Agrobacterium tumefaciens* mutants. Plasmid. 11(3):195-205.

Hooykaas-van Slogteren, G. M., Hooykaas, P. J., and Schilperoort, R. A. 1992. Expression of Ti plasmid genes in monocotyledonous plants infected with *Agrobacterium tumefaciens*. Biotechnology. 24:382-383.

Howard, E. A., Zupan, J. R., Citovsky, V., and Zambryski, P. C. 1992. The VirD2 protein of *A. tumefaciens* contains a C-terminal bipartite nuclear localisation signal: implications for nuclear uptake

of DNA in plant cells. Cell. 68(1):109-118.
Jayaswal, R. K., Veluthambi, K., Gelvin, S. B., and Slightom., J. L. 1987. Double-stranded cleavage of T-DNA and generation of single-stranded T-DNA molecules in *Escherichia coli* by a *virD*-encoded border-specific endonuclease from *Agrobacterium tumefaciens*. J. Bacteriol. 169(11):5035-5045.
Judd, P.K., Kumar, R.B., and Das. A. 2005. Spatial location and requirements for the assembly of the *Agrobacterium tumefaciens* type IV secretion apparatus. Proc. Natl. Acad. Sci. U.S.A. 102(32):11489-11503.
Koekman, B. P., Hooykaas, P. J. J. and. Schilperoort R. A. 1982. Localization of the replication control region on the physical map of the octopine Ti plasmid. Plasmid. 7(2):119-132.
Kumar, R.B., and Das, A. 2002. Polar location and functional domains of the *Agrobacterium tumefaciens* DNA transfer protein VirD4. Mol. Microbiol. 43(6):1523-1532.
Lanka, E., and Wilkins, B.M. 1995. DNA processing reactions in bacterial conjugation. Annu. Rev. Biochem. 64:141-169.
Lessl, M., and Lanka, E. 1994. Common mechanisms in bacterial conjugation and Ti-mediated T-DNA transfer to plant cells. Cell. 77(3):321-324.
Llosa, M., Gomis-Ruth, F.X., Coll, M., and de la Cruz, F. 2002. Bacterial conjugation: a two-step mechanism for DNA transport. Mol. Microbiol. 45(1):1-8.
Mysore, K. S., Bassuner, B., Deng, X. B., Darbinian, N. S., Motchoulski, A., Ream, W., and Gelvin, S. B. 1998. Role of the *Agrobacterium tumefaciens* VirD2 protein in T-DNA transfer and integration. Mol. Plant-Microbe Interact. 11(7):668-683.
Pitzschke, A, and Hirt, H. 2010. New insights into an old story: *Agrobacterium*-induced tumour formation in plants by plant transformation. EMBO (Eur. Mol. Biol. Organ.) J. 29(6):1021-1032.
Rossi, L., Hohn, B., and Tinland, B. 1993. The VirD2 protein of *Agrobacterium tumefaciens* carries nuclear localization signals important for transfer of T-DNA to plants. Mol. Gen. Genet. 239(3):345-353.
Scheiffele, P., Pansegrau, W., and Lanka, E. 1995. Initiation of *Agrobacterium tumefaciens* T-DNA processing. Purified proteins VirD1 and VirD2 catalyze site- and trans-specific cleavage of superhelical T-border DNA *in vitro*. J. Biol. Chem. 270(3):1269-1276.
Schrammeijer, B., den Dulk-Ras, A., Vergunst, A.C., Juralda Jácome, E., and Hooykaas, P.J.J. 2003. Analysis of Vir protein translocation from *Agrobacterium tumefaciens* using *Saccharomyces cerevisiae* as a model: evidence for transport of a novel effector protein VirE3. Nucleic Acids Res. 31(3):860-868.
Schulein, R., Guye, P., Rhomberg, T.A., Schmid, M.C., Schröder, G., Vergunst, A.C., Carena, I., and Dehio, C. 2005. A bipartite signal mediates the transfer of type IV secretion substrates of *Bartonella henselae* into human cells. Proc. Natl. Acad. Sci. U.S.A. 102(3):856-861.
Shurvinton, C. E., Hodges, L., and 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. U.S.A. 89(24):11837-11841.
Steck, T.R., Lin, T.S., and Kado, C.I. 1990. VirD2 gene product from the nopaline plasmid pTiC58

has at least two activities required for virulence. Nucleic Acids Res. 18(23):6953-6958.
Tzfira, T., Vaidya, M., and Citovsky, V. 2004. Involvement of targeted proteolysis in plant genetic transformation by *Agrobacterium*. Nature. 431(7004):87-92.
Van Kregten, M, Lindhout, B.I., Hooykaas, P.J.J, and van der Zaal, B.J. 2009. *Agrobacterium*-mediated T-DNA transfer and integration by minimal VirD2 consisting of the relaxase domain and a Type IV Secretion System translocation signal. Mol. Plant-Microbe Interact. 22(11): 1356-1365.
Vergunst, A.C., van Lier, M.C., den Dulk-Ras, A., Stuve, T.A., Ouwehand, A., and Hooykaas, P.J.J. 2005. Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of *Agrobacterium*. Proc. Natl. Acad. Sci. U.S.A. 102(3):832-837.
Vergunst, A.C., Schrammeijer, B., den Dulk-Ras, A., de Vlaam, C.M., Regensburg-Tuïnk, T.J., and Hooykaas, P.J.J. 2000. VirB/D4-dependent protein translocation from *Agrobacterium tumefaciens* into plant cells. Science. 290(5493):979-982.
Vogel, A.M., and Das, A. 1992. Mutational analysis of *Agrobacterium tumefaciens virD2*: tyrosine 29 is essential for endonuclease activity. J. Bacteriol. 174(1):303-308.
Yadav, N.S., Vanderleyden, J., Bennett, D.R., Barnes, W.M., and Chilton, M.D. 1982. Short direct repeats flank the T-DNA on a nopaline Ti plasmid. Proc. Natl. Acad. Sci. U.S.A. 79(20):6322-6326.
Zhu, J., Oger, P.M., Schrammeijer, B., Hooykaas, P.J.J., Farrand, S.K., and Winans, S.C. 2000. The bases of crown gall tumorigenesis. J. Bacteriol. 182(14):3885-3895.

Chapter 4

Translocation of novel recombinant effector proteins from *Agrobacterium tumefaciens* to *Arabidopsis thaliana*

Maartje van Kregten, Paulo de Boer, Paul J.J. Hooykaas and Bert J. van der Zaal

Abstract

In this study, novel effector proteins for use in plant transformation were created by fusion of several biologically interesting protein domains to the relaxase VirD2 of *Agrobacterium tumefaciens*. All effector proteins, when equipped with the C-terminus of VirF, were translocated through the Type IV Secretion System of *Agrobacterium* to *Arabidopsis thaliana* roots, albeit with different efficiencies. When artificial DNA-binding zinc fingers were added, the translocation efficiency dropped from 63% of wild type VirD2 for one zinc finger to 1% for six zinc fingers. Addition of a *Fok*I endonuclease domain did not decrease translocation efficiency any further. Compared to a HO nuclease moiety, an I-SceI homing endonuclease moiety was transferred very efficiently. Moreover, expression of the VirD2-I-SceI-F fusion protein in *Arabidopsis* plants demonstrated that this protein still possessed site specific nuclease activity. Fusing proteins-of-interest to VirD2 proved to be a very sensitive way to determine if they can be translocated by simply monitoring T-DNA encoded *GUS* gene activity. Moreover, the fact that zinc finger nucleases as well as homing endonucleases can be translocated to recipient eukaryotic cells, employing a protein moiety which is also able to transfer a single stranded DNA molecule, offers intriguing possibilities for gene targeting purposes.

Introduction

Targeted integration of transgenes at a predetermined locus in a complex genome, also known as gene targeting (GT), would be the ideal way to create organisms with precisely the trait altered that is of interest for study or for biotechnological applications. Unfortunately, in higher eukaryotes and especially in plants, foreign DNA predominantly integrates at random positions in the genome. This may cause the transgene to be expressed at different levels, the expression levels of endogenous genes to be disturbed, and the presence of extra copies of the transgene - or parts thereof - at various positions in the recipient genome. Both for the study of fundamental processes of plant biology and the generation of genetically altered crops, a method for reliable gene targeting would be more than welcome.

The problem regarding gene targeting in plants lies in the pathway employed for

the integration of foreign DNA into the plant genome. In most cases, DNA integration happens via non-homologous end-joining (NHEJ), in which the incoming DNA is ligated to the plant genome in an error-prone way. However, for biotechnological applications, the pathway of homologous recombination (HR), in which DNA is integrated via recombination between incoming and target DNA, would be the pathway of choice since the locus of integration can then be predetermined by inserting homology with the desired integration site. Unfortunately, in plants NHEJ is much more prevalent than HR. The earliest attempts at GT in several plant species resulted in very low efficiencies, estimated at 10^{-6} to 10^{-4} (Paszkowski *et al.*, 1988, Offringa *et al.*, 1990, Halfter *et al.*, 1992, Hrouda and Paszkowski, 1994, Risseeuw *et al.*, 1995, Miao and Lam, 1995). Since then, it has been demonstrated in plants that the presence of a double-stranded break (DSB) in the genome near the integration site will enhance the possibility of integration via HR (Puchta *et al.*, 1996). Since this discovery, methods have been developed to induce DSBs in the genome to optimize the levels of HR, most of them using zinc finger nucleases but also homing endonucleases.

Zinc finger nucleases can be regarded as artificial restriction enzymes consisting of two protein domains: one consisting of artificial zinc finger (ZF) moieties and an endonuclease (N) domain. The ZF domain represents a DNA binding module, which can be made by linking individual Cis_2-His_2 type ZF moieties with known DNA-binding characteristics, thereby forming a polydactyl ZF (PZF) structure binding to a DNA sequence of choice. With a single ZF essentially interacting with three consecutive basepairs of DNA, several ZFs need to be combined to generate the specificity needed to target a specific site in a complex genome. The N domain of ZFNs is thus far represented by the non-specific nuclease domain of the restriction enzyme *Fok*I, which is active after dimerization (Bitinaite *et al*, 1998, Kim *et* al., 1996, reviewed in Carroll, 2008). The development of ZFN technology has already resulted in many successful gene targeting events in a variety of eukaryotic organisms (Bibikova *et al.*, 2002, Wright *et al.*, 2005, Doyon *et al.*, 2008, De Pater *et al.*, 2009, reviewed in Weinthal *et al.*, 2010).

Another method of creating DSBs in the genome is by using natural homing endonucleases (HEs) that have very long, and therefore highly specific, asymmetric recognition sites. As their name implies, HEs are involved in the process of intron homing. HEs are unusual in the sense that they are encoded by elements inside another open reading frame, either in introns or as inteins, intervening protein sequences that are spliced out post-translationally. Their function is solely to copy their coding sequence into other loci, making

HEs a selfish genetic element for which no benefit to the host has been found (Jurica and Stoddard, 1999, Stoddard, 2005). The advantage of HEs is that their recognition sites are often so long that they are mostly absent in a given genome. Insertion of a HE recognition site, such as for I-SceI (Puchta *et al.*, 1996), thus allows for controlled experiments using a single DSB in a genome. Although not yet as flexible as the ZFN system, recent reports that HEs with modified recognition sites can be generated will undoubtedly lead to the use of HEs as important tools in molecular genetics (reviewed in Pâques and Duchateau, 2007, Galetto *et al.*, 2009).

Obviously, for getting protein tools into cells, a delivery system will be required. By using the Type 4 Secretion System (T4SS) of the natural genetic engineer *Agrobacterium tumefaciens*, it should be possible to precisely deliver DNA as well as protein substrates to a wide variety of cells. The T4SS is nowadays considered to be a passageway between the *Agrobacterium* cell and the recipient cell through which both protein and single stranded DNA are translocated (reviewed in Christie, 2004, Ding *et al.*, 2003, Lawley *et al.*, 2003). In nature, *Agrobacterium* transfers part of its tumor-inducing plasmid – the T- or transferred strand, flanked by the right (RB) and left border (LB) repeats – to cells of susceptible dicotyledonous plant hosts, causing the crown gall tumor disease (reviewed in Zhu *et al.*, 2000, Pitzschke and Hirt, 2010). In a laboratory setting, the host range of *Agrobacterium* has been extended to, amongst others, monocotyledonous plants and fungi (resp. Hooykaas-van Slogteren *et al.*, 1984, Bundock *et al.*, 1995, De Groot *et al.*, 1992), thus making *Agrobacterium*-mediated transformation (AMT) a widely used tool in biotechnology.

For genetic transformation of target cells, the protein VirD2 is essential. It is a strand processing relaxase which participates in liberating the T-strand by nicking the RB and LB sequences. During transfer of the single stranded T-strand via the T4SS, the VirD2 protein remains covalently bound at the 5' end (Dürrenberger *et al.*, 1989, Scheiffele *et al.*, 1995, Vogel and Das, 1992).

Besides DNA-bound VirD2, several other virulence (Vir) proteins can pass through the T4SS. This has been shown for VirE2 (Citovsky *et al.*, 1992), VirF (Vergunst *et al.*, 2000), VirD5 (Vergunst *et al.*, 2005), and VirE3 (Schrammeier *et al.*, 2003). These proteins share a C-terminal stretch of positively charged amino acids, which are likely to function as a T4SS translocation signal (Vergunst *et al.*, 2005). In the seminal study regarding Vir protein translocation, the recombinase protein Cre – which is unrelated to any of the Vir proteins - was shown to be able to translocate to *Arabidopsis* root cells when fused to the C-terminus of VirF (Vergunst *et al.*, 2000). In the same study, the CRAfT (Cre Recombinase reporter

Assay for Translocation) assay was developed, allowing detection of Cre protein transfer in an indirect manner by monitoring Cre-mediated recombination events in a special *Arabidopsis* target line.

In order to develop a system for the translocation of proteins via the T4SS of *Agrobacterium*, the most straightforward solution would thus be to equip proteins of interest with a C-terminal sequence derived from the C-terminus of VirF. Applying this strategy enabled us recently to present evidence for the translocation of severely truncated versions of the relaxase VirD2. Of these truncations, we showed that they are still functional in T-strand processing, and lack only the translocation function necessary to translocate the VirD2-T-strand to the recipient cell. When the VirD2 truncations are fused to the C-terminus of VirF, translocation is restored, thus providing definitive proof that T-strand transfer is protein-driven (Van Kregten *et al*., 2009).

Our finding that the VirD2 protein can be manipulated rather extensively while maintaining the natural biological function of interest, prompted us to investigate whether a VirD2 framework might be a useful vehicle for accommodation of other protein domains while preserving all biological functions of interest. Detection of VirD2 translocation is easy; genes on the T-strand, which is covalently bound to VirD2, will make detection simple and does not require a special reporter line of the target organism. In that manner, starting from the natural VirD2 function, it would be possible to construct novel protein tools for molecular genetics. Concerning the goal of GT in higher eukaryotes, in particular plants, such protein tools should combine (1) an ability to transfer T-strand molecules to the large range of organisms amendable for AMT, with (2) additional functions that will facilitate site-specific integration of co-transferred DNA into the recipients genome.

As a first step towards VirD2-mediated genome engineering (VirGEN) we investigated whether functional VirD2 derivatives can be combined with biotechnologically interesting protein domains. We thus demonstrate that ZF domains as well as complete ZFNs can in principle be translocated, together with incoming T-strands, although the efficiency declines severely with more complicated PZF domains. Two HEs, I-SceI and HO can also be translocated in fusion to VirD2; in particular the I-SceI moiety was were translocated very efficiently, at the high level of 71% of wild type VirD2. The implications of these findings are discussed.

Results

Translocation of VirD2 fusion proteins

To create novel fusion proteins with multiple functions, both in T-strand production and translocation from *Agrobacterium* to the recipient cell, and as effector protein in the recipient cell, we made fusion proteins containing VirD2 or VirD2-204, the relaxase domain of VirD2, of which we have shown that it is proficient in T-strand translocation (Van Kregten *et al.*, 2009). To easily detect fusion protein expression, we added the FLAG tag N-terminally, and to ensure nuclear localization, the NLS of SV40 was added just downstream of the FLAG tag. To enable translocation, the C-terminal 37 amino acid residues of VirF constituting a T4SS translocation signal (Vergunst *et al.*, 2000) were added at the C-terminus. While addition of the short N-terminal FLAG tag and NLS sequence did not seriously hamper VirD2 activity (Van Kregten *et al.*, 2009), preliminary experiments indicated that much longer N-terminal protein sequences could be inhibitory, possibly for the reason that the catalytic Tyr residue at position 29 that nicks the DNA in T-strand processing is normally present at the N-terminus (Vogel and Das, 1992). Moreover, it can easily be envisaged that N-terminal fusion of protein domains to VirD2 might lead to biased conclusions regarding their effects upon the natural VirD2 activity; in case that the hybrid protein should fall apart, a C-terminal fragment comprised of a VirD2 domain directly followed by a C-terminal VirF sequence is likely to possess relaxase as well as translocation activity. With structural or enzymatic protein modules inserted in between the VirD2 proper and the F sequence, transfer of the complete proteins is close to guaranteed since the VirD2 module has then become dependent upon the C-terminus of VirF. VirD2 was thus fused to (P)ZF moieties comprised of 1 to 6 ZFs, to 1 to 3 (P)ZF moieties combined with the nuclease domain of *Fok*I, creating ZFN-type fusion domains, and to the HEs I-SceI and HO. Of the latter two, a fusion to just the relaxase domain of VirD2 (VirD2-204) was also made. A schematic drawing of the novel recombinant proteins can be found in Fig. 1. Expression of constructs in *Agrobacterium* was confirmed by Western blot, using an anti-FLAG antibody, except for VirD2-HO-F (Fig. 2).

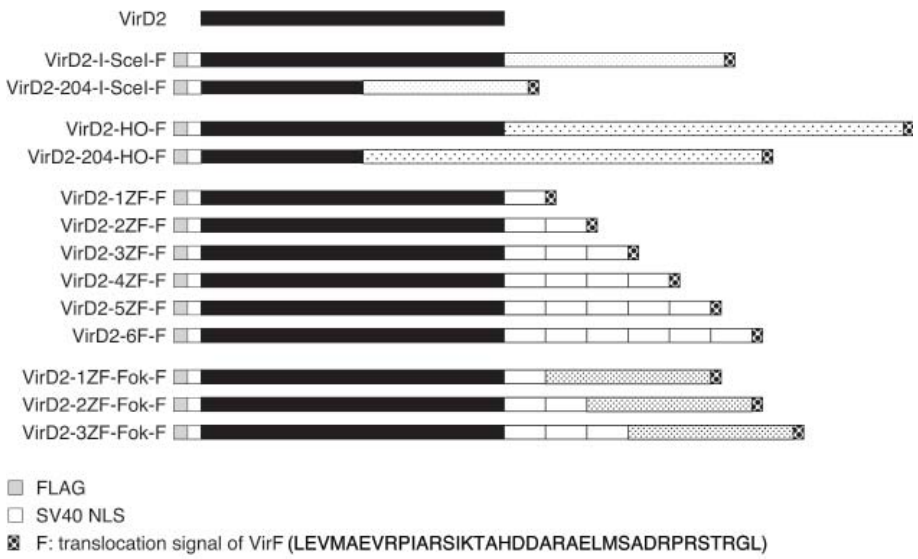

Figure 1: Schematic drawing of the novel recombinant effector proteins. Drawing is not to scale.

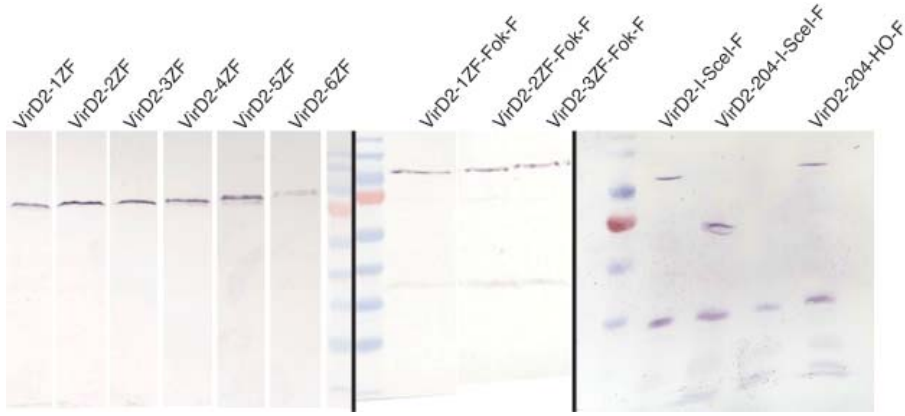

Figure 2: Western blot indicating expression of the novel recombinant effector proteins, expressed in *Agrobacterium* strain LBA2585 (Δ*virD2*, ΔT-DNA). The red band in the marker indicates 70 kDa (Fermentas).

To investigate the ability of the novel hybrid proteins to pass through the T4SS, we cocultivated *Agrobacterium* strain LBA2585 (Δ*virD2*, ΔT-DNA; Bravo-Angel *et al*., 1998) containing the expression vector for the relevant construct as well as the wide-host range vector pCAMBIA2301, the donor for a T-strand sequence with *in planta* expressed *GUS* and *NPTII* (kanamycin) marker genes, with *Arabidopsis* C-24 root explants. The number of GUS positive spots was determined after three days of cocultivation and taken as an

indication of transfer of T-strand molecules to the root cells, either transiently present or integrated into the genome, as described previously (Tinland *et al.*, 1995, Van Kregten *et al.*, 2009). All data were normalized to the level obtained with an LBA2585 strain harboring an expression plasmid providing wild-type (WT) VirD2, thus without any extra tags or protein domains. For all constructs, at least three independent experiments were performed.

As can be seen in Table 1, the efficiency of translocation differs widely between the different constructs. For VirD2 fusions with an increasing number of ZFs, a gradual decline in translocation efficiency was observed. While a construct consisting of VirD2 and 1 ZF still translocated at the relatively high level of 63% of wild type VirD2, addition of a second ZF and third ZF lowered the efficiency to 48% and 30%, respectively. Further increasing the number of ZFs about halved the transfer efficiency with every extra ZF added, until only 1% of transfer efficiency was observed when six ZFs were present.

To determine if the translocation of an entire ZFN could be a possibility, we created fusion proteins consisting of VirD2, 1 to 3 ZFs, the nuclease domain of *Fok*I, and the C-terminus of VirF. Remarkably, these constructs translocated at the same level as their counterparts that did not contain a *Fok*I nuclease domain (Table 1). Hence, the *Fok*I protein domain in itself does not at all hamper translocation of a VirD2 fusion via the T4SS.

Translocation of recombinant VirD2-ZFN-F proteins is thus possible, but for biotechnological applications the levels of translocation might be insufficient, especially for those proteins with 5 or 6 ZFs which should possess superior DNA-binding characteristics (Neuteboom *et al.*, 2006). Therefore, we set out to investigate translocation of the other group of DSB-inducing proteins that are widely used *in planta*, the homing endonucleases. We created fusion proteins of the monomeric HEs I-SceI, the use of which in plant research has already been documented (e.g. Puchta *et al.*, 1993, Puchta *et al.*, 1996, Orel *et al.*, 2003), and HO (Chiurazzi *et al.* 1996). Recombinant proteins consisting of VirD2, I-SceI and the C-terminus of VirF translocated at 71% of the level of wild type VirD2 (Table 1). A similar HO fusion was transferred to a lesser extent, only at 29% of wild type VirD2. To investigate whether the level of translocation could be further increased when the size of the fusion protein was diminished, we also made recombinant proteins using only the N-terminal relaxase half of VirD2, up to amino acid residue 204. In our previous study (Van Kregten *et al.*, 2009), VirD2-204 fused to the C-terminus of VirF was shown to be able to direct T-strands to recipient cells with only a modest loss in efficiency compared to the two times larger protein possessing the complete VirD2 domain. As can be seen in Table 1, we found that the translocation efficiencies of VirD-204-based fusions with I-SceI and HO

were lower than those of the similar constructs containing full length VirD2; for the I-SceI fusion, translocation dropped from 71% to 51% of the level of wild type VirD2. We thus did not detect a net benefit regarding protein and T-strand translocation by reducing the bulkiness of the fusion proteins by deleting the dispensable C-terminal half of VirD2.

To determine whether stable integration of the T-strand into the genome of *Arabidopsis* cells was still possible using novel recombinant proteins, stable integration was assessed by determining the number of transgenic kanamycin-resistant calli developing on root segments cocultivated with *Agrobacterium* strains expressing VirD2-I-SceI-F, VirD2-204-I-SceI-F, and VirD2-HO-F (Table 2). When comparing the stable integration results with the previously obtained data regarding the amount of GUS-positive spots after three days of cocultivation, no significant differences were found (see Table 1 and 2). Hence, most supposedly transient events finally led to stable presence of one or more T-DNAs within the recipient genome, demonstrating that the addition of a large protein domain to VirD2 does not hamper integration of the T-DNA.

All the biologically interesting proteins tested in this study could be made translocation-competent by the addition of VirD2 and the C-terminus of VirF. However, thus far the translocation of in particular the I-SceI domain was most compatible with the T4SS of *Agrobacterium*. For further proof of principle that novel recombinant effect proteins can be designed which combine DNA translocating potential with DSB-inducing capacity, it is essential to demonstrate that the fusion proteins retain their nuclease activity in the recipient cells. Therefore, we tested the *in planta* activity of the constructs containing I-SceI and HO.

Activity of the nuclease moiety

The activity of the VirD2 moiety of the fusion protein is warranted by its translocation and the subsequent detection by means of expression of the genes on the T-strand; if the VirD2 domain would be non-functional, this could not occur. For obtaining evidence that fusion of enzyme domains within a VirD2 framework does not affect the enzymatic properties of interest, in this case nuclease activity on chromatin-embedded DNA *in planta*, further tests are required. The activity of the specific nuclease moieties in the set of constructs mentioned above can be assessed in *Arabidopsis* lines that have a specific target site inserted into their genome. Expression of an active, site specific nuclease in such plant lines should induce a DSB in the DNA at its cognate recognition site. To a certain extent, erroneous repair will

Table 1: Transient transformation by novel recombinant effector proteins

construct	transient transformation[a]
wild type VirD2	1
FLAG-NLS-VirD2-1ZF-F	0.63 +/- 0.04
FLAG-NLS-VirD2-2ZF-F	0.48 +/- 0.09
FLAG-NLS-VirD2-3ZF-F	0.30 +/- 0.08
FLAG-NLS-VirD2-4ZF-F	0.12 +/- 0.04
FLAG-NLS-VirD2-5ZF-F	0.06 +- 0.03
FLAG-NLS-VirD2-6ZF-F	0.01 +/- 0.01
FLAG-NLS-VirD2-1ZF-Fok-F	0.59 +/- 0.07
FLAG-NLS-VirD2-2ZF-Fok-F	0.48 +/- 0.05
FLAG-NLS-VirD2-3ZF-Fok-F	0.34 +/- 0.10
FLAG-NLS-VirD2-I-SceI-F	0.71 +/- 0.09
FLAG-NLS-VirD2-204-I-SceI-F	0.51 +/- 0.11
FLAG-NLS-VirD2-HO-F	0.29 +/- 0.13
FLAG-NLS-VirD2-204-HO-F	0.17 +/- 0.08

[a] n=3 or greater, at least 100 explants were scored for GUS-positive spots per experiment.

Table 2: Stable integration of T-DNA by novel recombinant effector proteins

construct	stable transformation[a]
wild type VirD2	1
VirD2-I-SceI-F	0.63 +/- 0.07
VirD2-204-I-SceI-F	0.52 +/- 0.14
VirD2-HO-F	0.28 +/- 0.14

[a]n=3 or greater, at least 100 explants were scored for formation of kanamycin-resistant calli per experiment.

lead to mutations at this restriction site, rendering it insensitive to further cleavage by the nuclease moiety. As mentioned above, we focused on I-SceI fusion proteins as their transfer through the T4SS was very efficient. Furthermore, detection of erroneously repaired I-SceI target sites is very much facilitated by commercially available I-SceI enzyme; the use of HO is less attractive in this respect as such an enzyme source is lacking. For ZFN target sites, any commercially available enzyme can in principle be used for digestion of a cognate 6 bp target site chosen to be present in between the recognition half sites for the ZFN (Lloyd *et al.*, 2005). In this study, however, we refrained from further testing ZFNs for a combination of two reasons. Firstly, as demonstrated above, ZFNs with complex PFZ

domains consisting of five or six ZFs apparently are much less suitable substrates for T4SS transfer. Secondly, although three fingered PZF domains can be transferred via VirD2, the ZFNs based upon such simple PZF domains that were produced in our lab did not lead to site directed mutations in a chromosomal yeast reporter system (De Pater *et al*, 2009).

To assess the activity of the nuclease moiety of the VirD2-I-SceI-F, VirD2-204-I-SceI-F, VirD2-HO-F or VirD2-204-HO-F proteins upon chromatin-embedded DNA in plant cells, we produced the relevant VirD2 fusion protein under control of the *RPS5a* promoter, which is primarily active in dividing cells, in plant lines that contained a single copy of a locus with a recognition site for either I-SceI or HO. A similar approach was followed previously for complex ZFNs (De Pater et al., 2009).

At least 10 primary transformed plants were grown and their seeds were sown on selective medium. From the second generation T2 seedlings, DNA was extracted and the part of the locus containing the I-SceI or HO site was PCR-amplified. To screen for mutated recognition sites, the PCR product was digested with I-SceI (New England Biolabs), or *Hind*III (Fermentas). *Hind*III is the restriction enzyme cutting closest to the HO recognition site, since HO is not commercially available (see Fig. 3). In PCR products derived from plant lines expressing VirD2-I-SceI-F and VirD2-204-I-SceI-F, enzyme-resistant DNA was readily found and sequenced (Fig 3). The slower migrating DNA bands, each corresponding to a different transgenic plant line expressing VirD2-I-SceI-F or VirD2-204-I-SceI-F, in which the I-SceI recognition site is destroyed, vary in intensity (Fig. 3 and data not shown). This indicates that the fraction of damaged I-SceI recognition sites within DNA samples derived from different plant lines is variable, possibly due to different expression levels of the proteins. Since VirD2-I-SceI-F and VirD2-204-I-SceI-F are expressed under the control of the *RPS5a* promoter, the protein is mainly expressed in dividing cells. Initial damage occurring in dividing cells will thus be inherited in clonal sections, the size of each section depending on the number of daughter cells generated after the mutagenic event. On average, under the given circumstances, about one third of the total PCR-product displayed I-SceI-resistance, indicating that about one third of all available sites in T2 seedlings have been erroneously repaired. Of course, it is impossible to determine the amount of breaks repaired correctly, but it may very well be that only a minority of breaks will lead to a detectable footprint.

Sequence analysis damage induced by VirD2-I-SceI-F or VirD2-204-I-SceI-F showed that most events had led to small deletions or insertions (Fig. 3). Of course, very large deletions or insertions will be missed by PCR screening; very large deletions will

remove the binding site of a primer, and very large insertions will cause the formation of products that will not be amplified in the PCR reaction.

A similar strategy for finding evidence of mutations induced by expression of VirD2-HO-F and VirD2-204-HO-F was unsuccessful. This might indicate that HO is inactive when fused to VirD2 or, alternatively, it might reflect a technical difficulty when repair of HO-induced breaks in dividing plant cells would produce mainly small (6 bp or less) deletions that do not affect the neighboring *Hind*III-site (Fig. 3). Nevertheless, the fact that VirD2-(204)-I-SceI-F fusion proteins which can process and transfer T-DNA to a recipient plant cell do also posses nuclease activity when expressed *in planta* strongly indicates novel VirD2 based protein tools are feasible.

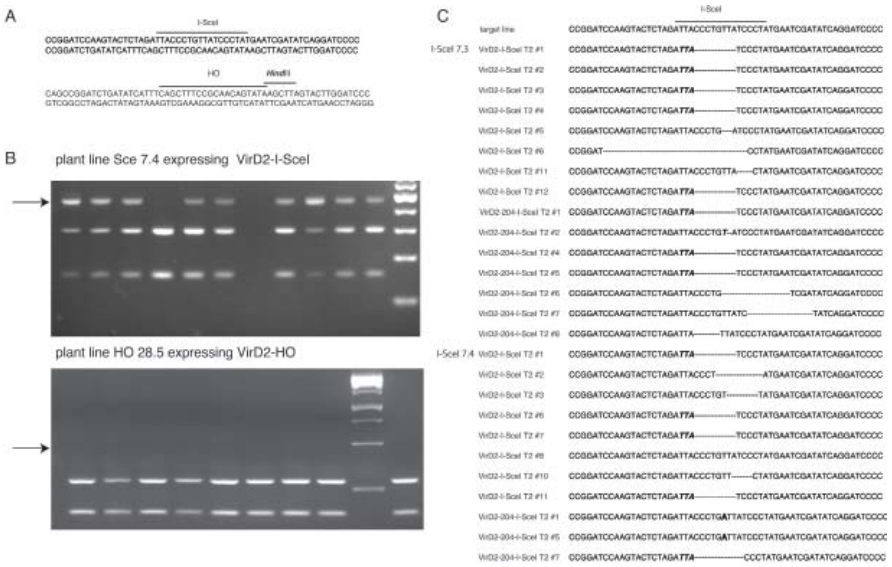

Figure 3: evidence for nuclease activity of VirD2-fusion proteins, **A**: recognition sites of I-SceI and HO, **B**: examples of digested PCR product, with the position of resistant bands indicated by an arrow, **C**: sequences of resistant bands. **Figure 3**: evidence for nuclease activity of VirD2-fusion proteins, **A**: recognition sites of I-SceI and HO, **B**: examples of digested PCR product, with the position of resistant bands indicated by an arrow, **C**: sequences of resistant bands.

Discussion

In this study, we report that the creation of a novel class of recombinant effector proteins for use in *Agrobacterium*-mediated transformation is feasible. In their most desirable form, these proteins posses T-strand processing activity in *Agrobacterium*, can be translocated to a plant cell recipient together with the T-strand, which can contain DNA sequences of interest for delivery to the recipient cells, and, when present in the plant cell, they can induce the formation of a site specific DSB in the recipients genome

To create these proteins, we made fusions of three protein domains: 1) the *Agrobacterium* relaxase protein VirD2, which is responsible for T-strand processing within *Agrobacterium* cells, 2) a biologically interesting effector moiety which can exert a function in a recipient cell, and 3) the T4SS translocation signal of the C-terminus of the *Agrobacterium* VirF protein (see Fig. 1 for a schematic drawing of the proteins). The effector domains used in the present work share a common feature: they are in principle able to induce a site-specific DSB in a complex genome, an event that can greatly GT (reviewed in Pâques and Duchateau, 2007, Carroll, 2008). Since the main tools for the induction of DSBs are ZFNs and HEs, we have included members of both classes in our study.

Regarding the possibility to include a complete ZFN moiety in between the N-terminal VirD2 domain and the C-terminal T4SS translocation signal, the situation is rather complex. Our data demonstrate that the *Fok*I nuclease domain passes the T4SS without any difficulty and, in addition to that, its fusion to VirD2 also leaves the essential VirD2 functions regarding T-strand processing and transfer unaffected. However, complete ZFN transfer seems to be hampered by the required ZF domains; while addition of one ZF still allows for about 68% of WT VirD2 translocation, the addition of more ZFs results in a steady decrease, with a PZF domain consisting of six ZFs being transferred at only 1% of WT VirD2 translocation (see Table 1). The experiments described in the present study do by themselves not yet provide definitive proof that the actual passage through the T4SS is the limiting factor for transfer of complex PZF domains since an inhibitory action of these domains upon the VirD2 functions associated with T-strand processing cannot be excluded. However, using the CRAfT assay, previous results indicated that detectable Cre activity in recipient plant cells was already severely diminished when only a single artificial ZF moiety was fused in between Cre and the C-terminus of VirF. Addition of more complex PZF

domains, consisting of three or more ZFs, nearly nullified detectable Cre activity in target cells (Lindhout and Van der Zaal, unpublished data). Since these Cre fusion proteins had nothing to do with T-strand processing, it thus appears that Cis_2-His_2 type ZF moieties indeed hamper translocation of proteins via the T4SS of *Agrobacterium*. Nevertheless, in contrast to Cre fusions, a VirD2 framework still allows for considerable transfer of three fingered PZF domains at about 30% of WT VirD2 levels, with or without a *Fok*I nuclease domain (Table 1). This might still hold promise for VirD2-mediated ZFN transfer.

In the present study, we did not pursue further investigation of ZFN transfer since three fingered ZFNs currently present in our lab do not suffice to generate footprints in yeast cells (De Pater *et al.*, 2009, Van der Zaal *et al.*, unpublished data). Obviously, it will be very interesting to explore the biotechnological applications of VirD2-ZFN fusions with more potent three-fingered ZFNs, possibly produced using the OPEN system (Maeder *et al.*, 2008), instead of modular assembly as was done in our lab. However, it should be realized that the commonly used ZFNs function as a dimer and that naturally occurring recognition sites are very rarely palindromic, requiring a different PZF domain for each half site. Hence, when the proteins constituting a ZFN are administered in small quantities, as it is most likely the case when using T4SS transfer during AMT, the correspondingly reduced chance of two ZFN subunits encountering each other is likely to diminish ZFN activity. Creating a ZFN that functions as a monomer might therefore be highly advantageous. The recent description of a monomeric ZFN in which the *Fok*I domain already consists of a covalently bound dimer which is `sandwiched` between two PZF domains (Mori *et al.*, 2009) is therefore very promising, although an extra PZF domain for sequence specification might very well again complicate T4SS-mediated transfer.

While the overall picture regarding the option that VirD2-ZFN-F fusions can be used for VirD2-mediated genome engineering (VirGEN) is not yet clear, the possibilities for VirD2-HE-F fusions are very promising. The homing endonuclease HO, in fusion to VirD2 and the C-terminus of VirF, translocates at 29% of the efficiency of wild type VirD2. Furthermore, the HO moiety is not inhibitory to integration of the T-strand; stable integration is approximately equal to transient transformation (see Tables 1 and 2). Translocation of HO is remarkable, considering the fact that HO contains 5 putative zinc fingers at its C-terminus (Russell *et al.*, 1986). Apparently, these ZFs are in a conformation that is less inhibitory to translocation than the ZFs stitched together in a PZF domain of similar size. Hence, simply attributing failure of translocation to increased bulkiness of fusions proteins provides insufficient reason for translocation difficulties with complex

PZFs. Individual assessment of the translocation ability of proteins remains is required until the general requirements for protein translocation are more clear. A great step toward possibly defining these requirements has been taken with the elucidation of the structure of the inner and outer membrane complex of the T4SS of plasmid pKM101 (Fronzes *et al.*, 2009, Chandran *et al.*, 2009).

Of all novel protein tools tested, VirD2-I-SceI-F translocates at the highest level of 70% of WT VirD2. Reducing the bulkiness of the total fusion protein, by using only the N-terminal half of the VirD2 protein which constitutes the active relaxase moiety, did not present any advantage over fusion to full length VirD2. In fact, the translocation rate of VirD2-204-I-SceI-F dropped to 51% of wild type. This might indicate that, although not required for relaxase activity, part of the deleted C-terminal half of VirD2 can positively contribute to transfer of fusion proteins, as was suggested previously (Van Kregten *et al.*, 2009). Also for HO fusions, the use of the VirD2-204 domain was not advantageous for transfer efficiency (see Table 1). As can be seen in Table 2, using VirD2 or VirD2-204 as a framework for the fusion protein does not make much difference in terms of integration efficiency. Just as in earlier work (Tinland *et al.*, 1995, Bravo-Angel *et al.*, 1998), an active role for VirD2 in integration of the T-DNA remain unlikely.

Whatever the future may hold for VirD2-mediated transfer of ZFN domains, recombinant effector proteins that involve HEs are highly promising to develop as DSB-inducing agents for GT in plants, provided that proof of nuclease activity of these fusion proteins can be obtained. As mentioned above, we did not succeed in finding footprints indicative of DSB-formation after expressing HO-containing fusion proteins *in planta*. This might reflect technical difficulties, as mutations other than larger than 6 bp deletions or the gain of large insertions in a HO site remained unnoticed with the procedures used. Otherwise, HO might be a difficult enzyme for use in plants. To our knowledge, HO has been used as DSB-creating agent in plants once and then under the control of the strong, constitutive *35S* promoter (Chiurazzi *et al.*, 1996). Since we expressed VirD2-(204)-HO-F under the control of the relatively weak *RPS5a* promoter, protein levels might have remained too low. Furthermore, it has been shown that HO is a rather unstable protein (Kaplun *et al.*, 2000), another possible complication for successful application in our experimental set-up. In contrast to HO fusion proteins, footprints confirming the activity of VirD2-(204)-I-SceI-F were readily found, mostly as small deletions or insertions (Fig. 3). This is consistent with findings in experiments performed in mammalian cells (Rouet *et al.*, 1994, Liang *et al.*, 1998).

Altogether, we have shown that several novel biologically interesting proteins that can in principle induce site specific DSBs can be translocated through the T4SS of *Agrobacterium* by fusion to VirD2 and F, although to a different extent. For I-SceI-containing fusion proteins, the T-strand processing and translocation ability of VirD2 remained largely intact and the nuclease activity was detectable when fusions proteins were expressed in plants. With this proof of principle, VirGEN could be realized, holding great promise for applications in GT. Although naturally occurring HE type nucleases like I-SceI have a fixed recognition site, different groups are working on strategies to accommodate the required flexibility within the HE scaffold in order to be able to generate proteins that can recognize chosen sequences of interest (Chen *et al.*, 2009, reviewed in Pâques and Duchateau, 2007, Galetto *et al.*, 2009). When the initial successes result in a generic strategy for selecting-site specific DNA-binding domains, VirGEN as we envisage it will easily come of age. By linking the T-strand and the protein of interest, it should become possible to deliver a T-strand to a specific site in a complex genome which is being cleaved in the same time frame, thus increasing the chances of GT and HR when the T-strand has homology to the locus of interest.

Materials and methods

Cloning of expression vectors for VirD2 fusion proteins

Cloning was performed using standard techniques in *Escherichia coli* strain DH5α. Both *E. coli* and *Agrobacterium* were cultured in LC medium containing the appropriate antibiotics. Primer sequences are listed in Table 3.

I-SceI was amplified from pUC18-I-SceI, a kind gift of Dr. H. Puchta, University of Karlsruhe, Germany, using primers I-SceI fw and I-SceI rv. The product was digested with *Spe*I and *Bgl*II and ligated to similarly digested pSKN-VirD2 or pSKN VirD2-204 (described in Van Kregten *et* al., 2009) to yield respectively pSKN-VirD2-I-SceI, or pSKN-VirD2-204-I-SceI.

HO was amplified from *Saccharomyces cerevisiae* using primers HO start and HO end. The product was digested with *Eco*RI and *Spe*I and ligated to similarly digested pSKN-VirD2 or pSKN-VirD2-204 to yield respectively pSKN-VirD2-HO and pSKN-VirD2-204-HO. The plasmids pSKN-VirD2 and pSKN-VirD2-204 are high copy *E. coli* cloning

vectors, derived from pSKN, in which the gene of interest becomes fused to the SV40 NLS sequence (for details, see Van Kregten *et al*, 2009).

The fragment containing the gene of interest of all pSKN-based vectors was released by digestion with *Not*I and ligated to *Not*I-digested *Agrobacterium* expression vector pBFF. As described previously, the pBFF plasmid allows for *virF* promoter-driven expression of N-terminal FLAG-tagged and SV40 NLS-containing fusion proteins, equipped with the 37 C-terminal amino acid translocation signal of the VirF protein (Van Kregten *et al.*, 2009). A schematic drawing of the proteins expressed by these constructs can be found in Fig. 1.

For expression of the same proteins in *Arabidopsis* plants under control of the *RPS5a* promoter, primarily active in embryos and meristems, *Not*I fragments were cloned into the previously described pGPTV-derived wide host range T-DNA vector pRF (Lindhout *et al.*, 2006, De Pater *et al.*, 2009) to create pRF-VirD2-I-SceI, pRF-VirD2-204-I-SceI, pRF-VirD2-HO and pRF-VirD2-204-HO. A schematic drawing of these proteins can be found in Fig. 1.

The zinc finger domains used were obtained from constructs in pSKN-SgrAI (Neuteboom *et al.*, 2006) by digestion with *Sfi*I and ligated into similarly digested pSKN-VirD2. The fragment containing the gene of interest was released by digestion with *Not*I and ligated to *Not*I-digested *Agrobacterium* expression vector pBFF.

The nuclease domain of *Fok*I was amplified from pSDM3835 (De Pater *et al.*, 2009) using primers Fok fw and Fok rv. The PCR product was cloned into pJET1.2 (Fermentas), excised by digestion with *Spe*I and ligated to relevant similarly digested pBFF VirD2-ZF vectors. A schematic drawing of the proteins expressed by these constructs can be found in Fig. 1.

Cloning of T-DNA vectors

To create binary vectors with an HO or I-SceI restriction site on T-DNA, pSDM3832 (De Pater *et al.*, 2009) was digested with *Bam*HI. Double-stranded oligo's were created from Scesite fw and Scesite rv or HOsite fw and HOsite rv, and ligated to *Bam*HI-digested pSDM3832, creating pSDM3832-I-SceI and pSDM3832-HO.

Root transformation

Root transformation experiments were performed as described earlier (Vergunst et al.,

2000). In short, seedlings of *Arabidopsis thaliana* ecotype C-24 were grown for 10 days in liquid medium, after which roots were harvested and pre-incubated for 3 days on medium containing 2,4-dichlorophenoxyacetic acid at 0.5 mg/liter and kinetin at 0.05 mg/liter. Subsequently, root segments were cocultivated for two days on the same medium but supplemented with 20 μM acetosyringone (Sigma-Aldrich) with *Agrobacterium tumefaciens* strain LBA2585, which has a deletion for *virD2* and the T-DNA (Bravo-Angel *et al.*, 1998), supplied *in trans* with the appropriate pBFF-derived VirD2 fusion protein expression vector and pCAMBIA2301 (www.cambia.org) as T-DNA donor. After this period, root segments were stained in staining solution containing X-gluc. For at least 100 root segments per construct per experiment, the amount of spots were counted. Data were normalized against data for T-DNA transfer by WT VirD2, as described previously (Van Kregten *et al.*, 2009). For the determination of stable integration, root segments were placed on medium containing N6-(isoprofenyl)adenine at 5 mg/liter and indole 3-acetic acid at 0.15 mg/liter. For selection of transformed plant cells, kanamycin was present at 50 mg/liter. Timentin (100 mg/liter) was used to kill remaining *Agrobacterium* cells. After two to three weeks, kanamycin-resistant calli were counted. Data were normalized against data for kanamycin-resistant callus formation by wild type VirD2, as described before (Van Kregten *et al.*, 2009).

Footprinting

Arabidopsis thaliana ecotype Col-0 lines containing an HO or I-SceI recognition site were created using floral dip (Clough and Bent, 1998) with *Agrobacterium* strain AglI (Lazo *et al.*, 1991) harboring pSDM3832-HO or pSDM3832-I-SceI as T-DNA donor. Homozygous phosphinotricin-resistant lines containing a single copy of the site of interest, namely Sce 7.3 and 7.4, and HO 28.5, were selected and were subsequently transformed using AglI containing pRF-VirD2-I-SceI, pRF-VirD2-204-I-SceI, pRF-VirD2-HO or pRF VirD2-204-HO. For every construct, at least 10 kanamycin-resistant transformants were allowed to grow and set seed. From their seedlings, DNA was extracted and footprinting was determined. The locus containing the restriction site was amplified with primers SP250 and SP251, as reported previously (De Pater *et al.*, 2009). For loci containing an I-SceI site, the PCR product was digested with I-SceI (NEB). For loci containing an HO site, the PCR product was digested with *Hind*III (Fermentas), because the enzyme HO is not commercially available. Enzyme-resistant bands were isolated from gel, cloned into pJET1.2 (Fermentas) and sequenced.

Western blot

Agrobacterium cultures were induced with 200 μM acetosyringone (Sigma-Aldrich) for at least 4 h before being harvested. A crude protein extract was made by boiling *Agrobacterium* cells in Laemmli buffer with β-mercapto-ethanol for 10 min. Insoluble proteins and debris were spun down and the supernatant was used for Western blotting. Proteins were separated on 10% sodium dodecyl sulfate polyacrylamide gel electrophoresis gels, transferred to nitrocellulose membrane (Immobilon-P, Millipore) and treated with FLAG antibody conjugated to alkaline phosphatase (cat no. A9469; Sigma-Aldrich). Detection of alkaline phosphatase activity was performed using nitroblue tetrazolium–5-bromo-4-chloro-3-indolyl phosphate solution (Sigma-Aldrich) according to instructions from the manufacturer.

Table 3: primer sequences

I-SceI fw:	GGATACTAGTGATGAAAAACATCAAAAAAAACCAGGTAATG
I-SceI rv:	TGATAGAGGAGGCTTTGAAAGGACTTGTAGATAGG
HO start:	GGATACTAGTTATGCTTTCTGAAAACACGACTATTCTGAT
HO end:	GGATCGAATTCTTGCAGATGCGCGCACCT
Fok fw:	TTCAACTAGTCAAAAGTGAACTGG
Fok rv:	AAACTAGTTGAAAGTTTATCTCGCCGTTATT
Scesite fw:	GATCTGATATCGAATTCATAGGGATAACAGGGTAATCTAGAGTACTTG
Scesite rv:	GATCCAAGTACTCTAGATTACCCTGTTATCCCTATGAATTCGATATCA
HOsite fw:	GATCTGATATCATTTCAGCTTTCCGCAACAGTATAAGCTTAGTACTTG
HOsite rv:	GATCCAAGTACTAAGCTTATACTGTTGCGGAAAGCTGAAATGATATCA
SP250:	CTCTGCCGTCTCTCTATTCG
SP251:	CTTGAAGAAGTCGTGCTGCTT

References

Bravo-Angel, A.M., Hohn, B., and Tinland, B. 1998. The omega sequence of VirD2 is important but not essential for efficient transfer of T-DNA by *Agrobacterium tumefaciens*. Mol. Plant-Microbe

Interact. 11:57-63.
Bibikova, H., Golic, M., Golic, K.G., and Carroll, D. 2002. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. Genetics. 161(3):1169-1175.
Bitinaite, J., Wah, D. A., Aggarwal, A. K., and Schildkraut, I. 1998. *Fok*I dimerization is required for DNA cleavage. Proc. Natl. Acad. Sci. 95(18):10570-10575.
Bundock, P., den Dulk-Ras, A., Beijersbergen, A., and Hooykaas, P. J. J. 1995. Trans-kingdom T-DNA transfer form *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*. EMBO (Eur. Mol. Biol. Organ.) J. 14:3206-3214.
Carroll, D. 2008. Zinc finger nucleases as gene therapy agents. Gene Ther. 15(22):1463-1468.
Chen, Z., Wen, F., Sun, N., and Zhao, H. 2009. Directed evolution of homing endonuclease I-SceI with altered sequence specificity. Protein Eng. Des. Sel. 22(4):249-256.
Chandran, V. Fronzes, R., Duquerroy, S., Cronin, N., Navaza, N., and Waksman, G. 2009. Structure of the outer membrane of a type IV secretion system. Nature. 462:24-31.
Chiurazzi, M, Ray, A., Viret, J. F., Perera, R., Wang, X. H., Lloyd, and A. M., and Signer, E. R. 1996. Enhancement of somatic intrachromosomal homologous recombination in Arabidopsis by the HO endonuclease. Plant Cell. 8(11):2057-2066.
Christie, P.J. 2004. Type IV secretion: the *Agrobacterium* VirB/D4 and related conjugation systems. Biochem. Biophys. Acta. 1693:219-234.
Citovsky, V., Zupan, J., Warnick, D., and Zambryski, P. 1992. Nuclear localization of *Agrobacterium tumefaciens* protein in plant cells. Science. 256:1802-1805.
Clough, S.J. and Bent, A.F. 1998. Floral dip: a simplified method for *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana*. Plant J. 16(6):735-743.
De Groot, M. J., Bundock, P., Hooykaas, P. J., and Beijersbergen, A. G. 1998. *Agrobacterium*-mediated transformation of filamentous fungi. Nat. Biotechnol. 16:839-842.
De Pater, S., Neuteboom, L.W., Pinas, J.E., Hooykaas, P.J.J., and van der Zaal, B.J. 2009. ZFN-induced mutagenesis and gene targeting in *Arabidopsis* through *Agrobacterium*-mediated floral dip transformation. Plant Biotech. J. 7:821-835.
Ding, Z., Atmakuri, K., and Christie, P.J. 2003. The ins and outs of bacterial type IV secretion substrates. Trends Microbiol. 11(11):527-535.
Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J., Gregory, P.D., Urnov, F.D., and Amacher, S.L. 2008. Heritable targeted disruption in zebrafish using designed zinc-finger nucleases. Nat. Biotechnol. 26(6) :702-708.
Dürrenberger, F, Crameri, A., Hohn, B., and Koukolikova-Nicola., Z. 1989. Covalently bound VirD2 from *Agrobacterium tumefaciens* protects the T-DNA from exonucleolytic degradation. Proc. Natl. Acad. Sci. 86(23):9154-9158.
Fronzes, R., Schäfer, E., Wang, L., Saibil, H.R., Orlova, E.V., and Waksman, G. 2009. Structure of a Type IV Secretion System Core Complex. Science. 323:266-268.
Galetto, R., Duchateau, P., and Pâques, F. 2009. Targeted approaches for gene therapy and the

emergence of engineered meganucleases. Expert Opin. Biol. Ther. 9(10):1289-1303.
Halfter, U., Morris, P.C., and Willmitzer, L. 1992. Gene targeting in *Arabidopsis thaliana*. Mol. Gen. Genet. 231:186-193.
Hrouda, M., and Paszkowski, J., 1994. High fidelity extrachromosomal recombination and gene targeting in plants. Mol. Gen. Genet. 243:106-111.
Jurica, M.S., and Stoddard, B.L. Homing endonucleases: structure, function and evolution. 1999. CMLS Cell. Mol. Life Sci. 1304-1326.
Hooykaas-van Slogteren, G. M., Hooykaas, P. J. J., and Schilperoort, R. A. 1992. Expression of Ti plasmid genes in monocotyledonous plants infected with *Agrobacterium tumefaciens*. Biotechnology 24:382-383.
Kaplun, L, Ivantsiv, I., Kornitzer, D. and Raveh, D. 2000. Functions of the DNA damage response pathway target Ho endonuclease of yeast for degradation via the ubiquitin-26S proteosome system. Proc. Natl. Acad. Sci. USA 97:10077-10082.
Kim, Y.G., Cha, J., and Chandrasagaran, S. 1996. Hybrid restriction enzymes: zinc finger fusions to *Fok*I cleavage domain. Proc. Natl. Acad. Sci. 93:1156-1160.
Lawley, T.D., Klimke, W.A., Gubbins, M.J. and Frost, L.S. 2003. F factor conjugation is a true Type IV secretion system. FEMS Microbiol. Lett. 224(1):1-15.
Lazo, G.R., Stein, P.A., and Ludwig, R.A. 1991. A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. Biotechnology. 9:963-967.
Liang, F., Han, M., Romanienko, P.J., and Jasin, M. 1998. Homology-directed repair is a major double-strand repair pathway in mammalian cells. Proc. Natl. Acad. Sci. 95(9):5172-5177.
Lloyd, A., Plaisier, C.L., Carroll, D., and Drews, G.N. 2005. Targeted mutagenesis using zinc finger-nucleases in *Arabidopsis*. Proc. Natl. Acad. Sci. 102(6):2232-2237.
Lindhout, B.I., Pinas, J.E., Hooykaas, P.J.J., and van der Zaal, B.J. 2006. Employing libraries of artificial zinc finger transcription factors to screen for homologous recombination mutants in *Arabidopsis*. Plant J. 48(3):475-483.
Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichtinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A., Unger-Wallace, E., Sander, J.D., Müller-Lerch, F., Fu, F., Pearlberg, J., Go¨ bel, C., Dassie, J.P., Pruett-Miller, S.M., Porteus, M.H., Sgroi, D.C., Lafrate, A.J., Dobbs, D., McCray Jr, P.B., Cathomen, T., Voytas, D.F. and Joung, J.K. 2008. Rapid "open-source" engineering of customized zincfinger nucleases for highly efficient gene modification. Mol. Cell. 31:294–301
Miao, Z.H., and Lam, E. 1995. Targeted disruption of the TGA3 locus in *Arabidopsis thaliana*. Plant J. 7:359-365.
Mori, T., Kagatsume, I., Shinomiya, K., Aoyama, Y., and Sera, T. 2009. Sandwiched zinc-finger nucleases harboring a single-chain *Fok*I dimer as a DNA-cleavage domain. Biochem. Biophys. Res. Commun. 390:694-697.
Neuteboom, L.W., Lindhout, B.I., Saman, I.L., Hooykaas, P.J.J., and van der Zaal, B.J. 2006. Effects

of different zinc finger transcription factors on genomic targets. Biochem. Piophys. Res. Commun. 339:263-270.

Offringa, R., de Groot, M.J., Haagsman, H.J., Does, M.P., van den Elzen, P.J., and Hooykaas, P.J.J. 1990. Extrachromosomal homologous recombination and gene targeting in plant cells after *Agrobacterium*-mediated transformation. EMBO J. 9:3077-3084.

Orel, N, Kyryk, A., and Puchta, H. 2003. Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome. Plant J. 35(5):604-612.

Paszkowski, J., Baur, M., Bogucki, A., and Potrykus, I. 1988. Gene targeting in plants. EMBO J. 7:4021-4026.

Pâques, F. and Duchateau, P. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. 2007. Curr. Gene Ther. 7:49-66.

Pitzschke, A, and Hirt, H. 2010. New insights into an old story: *Agrobacterium*-induced tumour formation in plants by plant transformation. EMBO J. 29(6):1021-1032.

Puchta, H., Dujon, B., and Hohn, B. 1993. Homologous recombination in plant cells is enhanced by *in vivo* induction of double strand breaks into DNA by a site-specific endonuclease. Nucleic Acids Res. 21(22):5034-5040.

Puchta, H., Dujon, B., and Hohn, B. 1996. Two different but relate mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc. Natl. Acad. Sci. 93:5505-5060.

Risseeuw, E., Offringa, R., Franke-van Dijk, M.E.I., and Hooykaas, P.J.J. 1995. Targeted recombination in plants using *Agrobacterium* coincides with additional rearrangements at the target locus. Plant J. 7:109-119.

Rouet, P., Smith, F., and Jasin, M. 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell. Biol. 14(12):8096-8106.

Russell, D.W., Jensen, R., Zoller, M.J., Burke, J., Errede, B., Smith, M., and Herskowitz, I. 1986. Structure of the *Saccharomyces cerevisiae HO* gene and analysis of its upstream regulatory region. Mol. Cell. Biol. 6(12):4281-4294.

Scheiffele, P., Pansegrau, W., and Lanka, E. 1995. Initiation of *Agrobacterium tumefaciens* T-DNA processing. Purified proteins VirD1 and VirD2 catalyze site- and trans-specific cleavage of superhelical T-border DNA *in vitro*. J. Biol. Chem. 270:1269-1276.

Schrammeier, B., den Dulk-Ras, A., Vergunst, A. C., Jurado Jacome, E., and Hooykaas, P.J.J. 2003. Analysis of Vir protein translocation from Agrobacterium tumefaciens using Saccharomyces cerevisiae as a model: evidence for transport of a novel effector protein VirE3. Nucleic Acids Res. 31(3):860-868.

Shukla, V.K., Doyon, Y., Miller, J.C., DeKelver, R.C., Moehle, E.A., Worden, S.E., Mitchell, J.C., Arnold, N.L., Gopalan, S., Meng, X., Choi, V.M., Rock, J.M., Wu., Y.Y., Katibah, G.E., Zhifang, G., McCaskill, D., Simpson, M.A., Blakeslee, B., Greenwalt, S.A., Butler, H.L., Hinkley, S.J., Zhang, L.,

Rebar, E.J., Gregory, P.D., and Urnov, F.D. 2009. Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. Nature. 459(7254):437-441.

Stoddard, B. L., 2005. Homing endonuclease structure and function. Q. Rev. Biophys. 38(1). 49-95.

Tinland, B., Schoumacher, F., Gloeckler, V., Bravo-Angel, A. M., and Hohn, B. 1995. The *Agrobacterium tumefaciens* virulence D2 protein is responsible for precise integration of T-DNA into the plant genome. EMBO (Eur. Mol. Biol. Organ.) J. 14:3585-3595.

Townsend, J.A., Wright, D.A., Winfrey, R.J., Lu, F., Maeder, M.L., Joung, J.K., and Voytas, D.F. 2009. High-frequency modification of plant genes using engineered zinc-finger nucleases. Nature. 459:442-446.

Van Kregten, M, Lindhout, B.I., Hooykaas, P.J.J., and Van der Zaal, B.J. 2009. *Agrobacterium*-mediated T-DNA transfer and integration by minimal VirD2 consisting of the relaxase domain and a Type IV Secretion System translocation signal. Mol. Plant-Microbe Interact. 22(11): 1356-1365.

Vergunst, A.C., Schrammeier, B., den Dulk-Ras, A., de Vlaam, C.M., Regensburg-Tuïnk, T.J., and Hooykaas, P.J.J. 2000. VirB/D4-dependent protein translocation from *Agrobacterium* into plant cells. 2000. Science. 290(5493):979-982.

Vergunst, A.C., van Lier, M. C., den Dulk-Ras, A., Stuve, T. A., Ouwehand, A., and Hooykaas, P.J.J. 2005. Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of Agrobacterium. 2005. Proc. Natl. Acad. Sci. 102(3): 832-837.

Vogel, A. M., and Das, A. 1992. Mutational analysis of *Agrobacterium tumefaciens virD2*: tyrosine 29 is essential for endonuclease activity. J. Bacteriol. 174:303-308.

Weinthal, D., Tovkach, A., Zeevi, V., and Tzfira, T. 2010. Genome editing in plant cells by zinc finger nucleases. Trends Plant Sci. 15(6):308-321.

Wright, D.A., Townsend, J.A., Winfrey, R.J. Jr., Irwin, P.A., Rajagopal, J., Lonosky, P.M., Hall., B.D., Jondle, M.D., and Voytas, D.F. 2005. High-frequency homologous recombination in plants mediated by zinc-finger nucleases. Plant J. 44(4):693-705.

Zhu, J, Oger, P.M., Schrammeijer, B., Hooykaas, P.J.J., Farrand, S.K., and Winans, S.C. 2000. The bases of crown gall tumorigenesis. J. Bacteriol. 182(14):3885-3895.

Chapter 5

Agrobacterium-mediated delivery of a meganuclease into target plant cells

Maartje van Kregten, Paulo de Boer, Johan E. Pinas, Paul J.J. Hooykaas and Bert J. van der Zaal

Abstract

Novel recombinant effector proteins were tested for their ability to be used as tools for the induction of targeted double-strand breaks in the *Arabidopsis* genome. These proteins consisted of the VirD2 relaxase of *Agrobacterium tumefaciens* fused to the homing endonuclease I-SceI and the C-terminal translocation signal of the *Agrobacterium* protein VirF. In this study, we present data showing that the novel proteins are indeed capable of inducing double-strand breaks in the genome of an *Arabidopsis* line containing an I-SceI recognition site, after their passage through the T4SS. As such DSBs would stimulate homologous recombination, we searched for targeted integration events of the accompanying homologous T-DNA. However, insertion of the T-DNA at the I-SceI site via homologous recombination could not be demonstrated among about a thousand transgenic plantlets screened, indicating that such frequency is still below one in a thousand.

Introduction

Agrobacterium tumefaciens is the preferred vector for the genetic modification of plants. *Agrobacterium*-mediated transformation (AMT) can be achieved with numerous plant species, sometimes in a very simple manner, like floral dip, which involves dipping *Arabidopsis thaliana* flowers in an *Agrobacterium* suspension. This will generate transformed seeds (Clough and Bent, 1998). During AMT, *Agrobacterium* transports DNA (T- or transferred strands) and several different virulence proteins into plant cells. The virulence protein VirD2 is responsible for the liberating of the T-strand and remains covalently attached to the T-strand at the 5' end (Dürrenberger *et al*. 1989, Scheiffele *et al*., 1995). After translocation into the plant cell, the T-DNA can integrate into the plant genome.

Genetic modification, resulting in incorporation of introduced DNA molecules into the genome, can nowadays be achieved in many plant species. However, it is not yet possible to efficiently steer these processes in such a manner that integration of introduced DNA molecules takes place at the desired position in the genome, leaving the rest of the genome unaltered. Such events are called gene targeting (GT) events.

GT can be achieved if the introduced DNA integrates via homologous recombination

(HR). For this to happen, the introduced DNA should contain homology to the desired integration site. Unfortunately, integration usually takes place via one of several pathways of non-homologous recombination (NHR). Integration via NHR results in mutation of the genomic locus, and the integrated genes may suffer from position effects and like gene silencing. To create a strategy to induce GT in plants, it is therefore important to find a way to boost the amount of integrations via HR. Delivery of T-DNA via *Agrobacterium* does not result in efficient gene targeting (Offringa *et al.*, 1990). It has been demonstrated that introducing double-strand breaks (DSBs) in the genome induces integration via HR: introducing a single DSB in the genome of tobacco caused an increase of two orders of magnitude in the amount of integrations via HR in the vicinity of the DSB (Puchta *et al.*, 1996). Thus far in GT strategies, which employ endonucleases, both the GT DNA template and the genes encoding the DSB-inducing enzymes were co-introduced. These genes for the endonuclease may be integrated into the genome, which is undesirable, as this again may cause mutation and these genes may have to be removed later on. This could be avoided by administering the mRNA or the DSB-inducing enzyme itself, which would disappear from the cells in due time, after their action.

To generate a single DSB in the genome of a higher eukaryote, an enzyme with a long recognition sequence is required. Suitable DSB-inducing enzymes are zinc finger nucleases (ZFNs) and homing endonucleases (HEs). ZFNs consist of several zinc finger DNA binding domains, coupled to a nuclease domain derived from the restriction enzyme *Fok*I. By using different combinations of zinc finger DNA binding domains, each binding to a triplet of base pairs, it is possible to create ZFNs that bind to (almost) any unique site in the genome (reviewed in Durai *et al.*, 2005). HEs are endonucleases that have a very high specificity due to their extremely long (17-20 bp) recognition sites (Stoddard, 2005). Since the specificity of HEs cannot yet be fully modulated, as yet a cognate target site must first be introduced into the plant. In certain species of fungi, a method in which the DSB-inducing agent is administered as an enzyme is already in use. In these cases, restriction enzymes are applied to enhance genomic integration of linear DNA. This method is called restriction enzyme-mediated integration (REMI). Since it involves the transformation of cells with a mixture of linearized plasmid DNA and a restriction enzyme that is capable of generating the compatible cohesive ends in the genome (Kuspa, 2006), its concept is close to the ideal situation where a truly site-specific enzyme is introduced.

We have chosen to create fusion proteins consisting of the *Agrobacterium* VirD2 protein and the HE I-SceI. We have used the T-strand as GT template, thereby creating

a single protein-DNA complex potentially capable of inducing GT. There is no need to synthesize and purify this complex; it is produced within *Agrobacterium* and transferred into the plant by AMT, making this set-up technically very easy.

We have already shown that fusion proteins consisting of VirD2, different HEs or ZFNs, followed by the T4SS translocation signal of VirF can be translocated via AMT into *Arabidopsis* (Chapter 4). Of the fusion proteins that translocated most efficiently, VirD2-I-SceI-F and VirD2-204-I-SceI-F, we have also shown that they can still cleave an I-SceI recognition site when expressed *in planta*, indicating that I-SceI tolerated fusions at both termini (Chapter 4). However, passage through the Type 4 Secretion System (T4SS), which functions as the translocation channel between *Agrobacterium* and the plant, may disrupt the structure of VirD2-I-SceI-F. Apart from that, the fusion protein in this set-up is covalently bound to the T-strand, which was of course not the case when the fusion protein was expressed *in planta*. Furthermore, it remains unknown how many protein molecules are translocated during AMT. Therefore, it is crucial first find out whether activity of the I-SceI moiety of the fusion protein can be detected after AMT.

In this study, we indeed found evidence for post-AMT DSB-inducing activity of the fusion proteins, in the form of induced mutation of the target site. These results show that the novel recombinant proteins VirD2-I-SceI-F and VirD2-204-I-SceI-F are bi-functional; their VirD2 moiety functions in T-strand production and translocation, and their I-SceI moiety functions in the induction of DSBs. Even at the (probably) low level at which the proteins are translocated by *Agrobacterium* into the recipient cell, they can still induce detectable DSBs. These results indicate that it should be possible to develop VirD2-mediated genome engineering (VIRgen) into a tool for GT.

Results

Agrobacterium-mediated delivery of the homing endonuclease I-SceI into plant cells

In our study, we investigated whether it is possible to use *Agrobacterium tumefaciens* to co-deliver an active homing endonuclease protein together with a T-DNA into *Arabidopsis thaliana*. Previous results have shown that a fusion protein consisting of VirD2, I-SceI, and the C-terminal T4SS translocation signal of VirF can be translocated at a relatively high frequency from *Agrobacterium* into *Arabidopsis* root cells. The same construct based on the shorter VirD2-204, consisting of the N-terminal 204 amino acid residues of VirD2

can also be translocated (Chapter 4). VirD2-I-SceI-F and VirD2-204-I-SceI-F have at their N-terminus a FLAG-tag and an SV40 NLS. The FLAG tag was used to confirm expression in *Agrobacterium*, and the NLS was added to ensure nuclear localization of the construct (Chapter 4).

To investigate whether active I-SceI could be delivered into plant cells as fusion proteins and are translocated in sufficient amount to actually bring about DSBs in the recipient, we transformed an *Arabidopsis* target line (Sce 7.3) using the floral dip method. Target line Sce 7.3 is homozygous for a single copy of a locus containing an I-SceI recognition site positioned between the Rps5a promoter and the coding sequence of a *GFP::GUS* fusion protein (Fig 1).

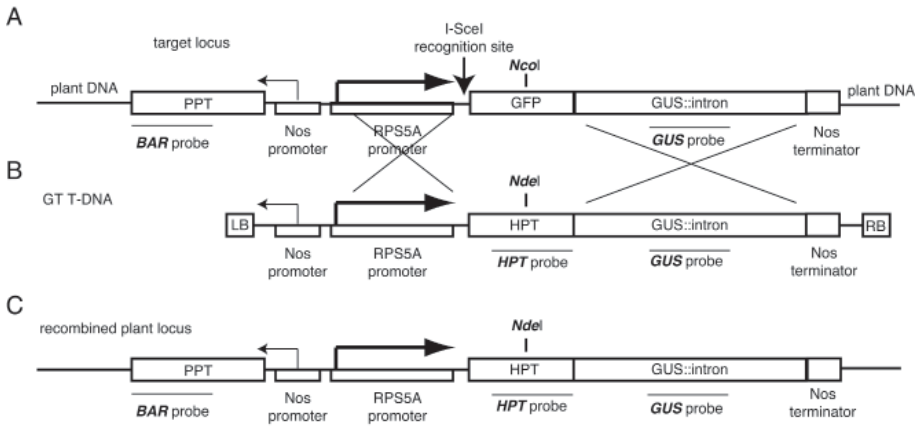

Figure 1: A: The stably integrated targeting locus, with an I-SceI recognition site between the Rps5a promoter and *GFP*. B: the gene targeting T-DNA. LB: left border, RB: right border. Homology to the targeting locus is indicated. C: the locus after a successful GT-event. Note the exchange of *GFP* for *HPT.* The positions of relevant restriction sites and probes are indicated.

For floral dip, we used *Agrobacterium virD2* deletion mutant LBA2585 (Bravo-Angel *et al.*, 1998), containing an expression plasmid for either VirD2-I-SceI-F or VirD2-204-I-SceI-F, or for wild type VirD2, and the binary vector pSDM3834 as T-DNA donor. The T-DNA contained the Rps5a promoter and the *GUS* coding sequence, but lacks the I-SceI recognition site and has the *HPT* coding sequence, instead of the *GFP* coding sequence (De Pater *et al.*, 2009). See Fig. 1 for details.

Transformants were selected on plates containing hygromycin, and DNA was isolated pooled leaf material, each pool representing 10 newly transformed, hygromycin resistant plants. A total of 104 VirD2 pools, 108 VirD2-I-SceI-F pools and 33 VirD2-204-

I-SceI-F pools were collected, corresponding to 1040, 1080, and 330 individual transgenic plants, respectively.

To detect whether transformation had been accompanied by I-SceI nuclease activity, we analyzed the DNA pools for the presence of damaged I-SceI recognition sites. To this end, a small product containing the recognition site of I-SceI in the target locus was amplified by PCR and subsequently digested with I-SceI (New England Biolabs). In none of the 86 VirD2 pools that gave a PCR product, I-SceI-resistant DNA was detected. However, among 95 VirD2-I-SceI-F pools which gave a PCR product, three pools contained I-SceI-resistant DNA. For the VirD2-204-I-SceI-F pools, 28 pools gave a PCR product, two of which contained I-SceI resistant DNA. After cloning and sequencing, mutations in the I-SceI site were found, indicative of misrepaired DSBs (Fig. 2).

The relative intensity of the I-SceI-resistant DNA, compared to the total amount of DNA that was loaded per sample, suggested that approximately one plant per pool of 10 plants contained a footprint (Fig. 2). We therefore investigated the individual plants of a pool to identify the individual plant(s) containing the footprint. Altogether, we identified a single plant containing a footprint in pools VirD2-I-SceI-F 48, 92 and 103 and in pools VirD2-204-I-SceI-F 21 and 33. Gel analysis of I-SceI-digested DNA demonstrated that approximately half of the PCR product was digested (Fig. 2), indicating that the primary transformed plant was close to being heterozygous for the footprint.

After sequencing, it became evident that the different plants studied had different mutations at the I-SceI target site: one point-mutation (C to A), one insertion of a single base (a T), a deletion of 8 basepairs, and a deletion of 16 basepairs combined with the insertion of 10 basepairs, (Fig. 2). These results thus showed that after translocation, the VirD2-(204)-I-SceI-F proteins were able to find and cut the I-SceI target site.

As the VirD2-I-SceI-F and VirD2-204-I-SceI-F fusion proteins retained nuclease activity after translocation, a concomitantly delivered T-DNA might be captured at the target locus by homologous recombination. Therefore we set out to screen the same pools of plants mentioned above for the presence of gene targeting events.

Screening for a GT event

As described above, we found that the I-SceI moiety of VirD2-I-SceI-F and VirD2-204-I-SceI-F can indeed make a DSB at I-SceI recognition sites in the genome after AMT. The T-DNA was accompanied by these VirD2 fusion proteins and contained extensive

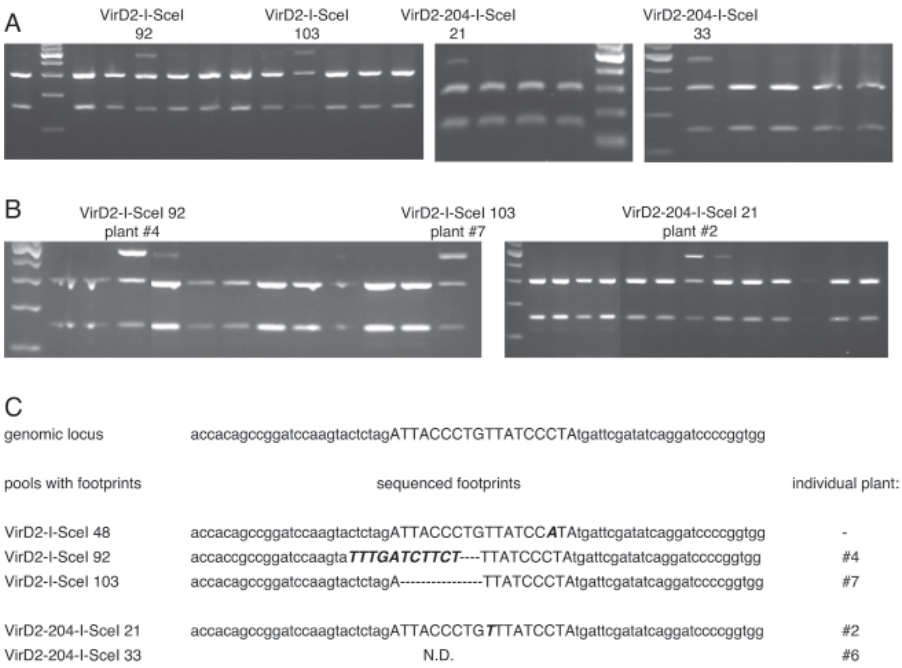

Figure 2: A: pools VirD2-I-SceI 92 and 103, and pools VirD2-204-I-SceI 21 and 33 are representative examples of pools that show an I-SceI-resistant band after digestion, B: individual plants of pools VirD2-I-SceI 92 (plant #4) and 103 (plant #7) and of pool VirD2-204-I-SceI 21 (plant #2) show an I-SceI-resistant band after digestion, C: sequence of genomic locus, with the I-SceI recognition site in capitals, sequences of footprints found individual plants, N.D., not determined.

homology to the pre-inserted locus with the I-SceI recognition site. Since formation of a DSB by VirD2-I-SceI-F or VirD2-204-I-SceI-F would stimulate the targeted integration of the T-DNA via HR, we searched for GT events within our pools of transformants.

The T-DNA differs from the integrated locus by the absence of the *PPT* gene, and the presence of *HPT* instead of *GFP*, allowing for selection of transformants on hygromycin (Fig. 1). Random integration would also confer resistance to hygromycin. Therefore, all pools were screened by PCR for the presence of GT events, as described before (De Pater *et al.*, 2009).

Direct screening for PCR products indicative of GT events, using a primer in *PPT* and a primer in *HPT* was subject to high background signals, probably due to recombination of intermediate products in the PCR reaction sharing extensive sequence overlap (De Pater and Van der Zaal, unpublished observations). Therefore, we pre-screened the pools by first amplifying the entire locus surrounding the target site, by using a primer

in *PPT* and a genomic primer downstream of the integrated target locus. In this procedure, the PCR product is enriched for the targeting site, while off-target random integrations are not amplified. The product of the first PCR was diluted and used as template for the second reaction. In this reaction, nested primers BAR fw in the *PPT* gene and SP284 (De Pater *et al.*, 2009) in the *HPT* gene were used to amplify GT-specific PCR products. As control for the quality of the template, a PCR reaction with nested primers BAR fw and SP251 (De Pater *et al.*, 2009), specific for *GFP* was performed, which always results in a product.

In the 104 VirD2 control pools, no evidence for gene targeting events was found, nor in the 33 VirD2-204-I-SceI-F pools. However, among the 108 VirD2-I-SceI-F pools, pool 80 was found to contain a *HPT*-specific band. Of the 10 plants in pool 80, 9 had survived and of these plants, samples were taken for analysis. In plant #2 the *HPT*-specific band was recovered (Fig. 3A and B). Material of the progeny of plant VirD2-I-SceI-F 80-2 and of the progeny of three random plants from this pool and two random plants from pool 105 were harvested for Southern blot analysis. Unfortunately, Southern blot analysis indicated that GT had not taken place in plant VirD2-I-SceI-F 80-2: after hybridization with a *PPT* probe no difference to the original target line was observed (Fig 3C). The DNA was digested with *Nco*I, which cuts in *GFP* and therefore affects the original target locus and not a GT event, which should contain *HPT* instead of *GFP*. Random integrations will not be detected by hybridization with the *PPT* probe. The other side of the locus was tested with a *GUS* probe. Using this probe, also no differences between the parental line Sce 7.3 and plant #2 of pool VirD2-I-SceI 80 were seen (Fig. 3D), thereby ruling out both a GT event as well as an ectopic integration event of a recombined locus.

To detect the amount for random integrations of the T-DNA, a Southern blot was prepared in which the same samples were digested with *Nde*I and probed using a *HPT* probe, which detects random integrations and GT events. On the blot two to four bands were seen, indicating of a limited number of integrations (Fig. 3E).

In summary, novel recombinant effector proteins consisting of the *Agrobacterium* relaxase protein VirD2 and the monomeric homing endonuclease I-SceI, combined with a T4SS translocation signal derived from the *Agrobacterium* protein VirF, are functional after AMT. They function in the generation of detectable DSBs, indicating that passage through the T4SS does not necessarily (or to a limited extent) interfere with their activity. No GT events were recovered from the plants analyzed in this study. These results are discussed below.

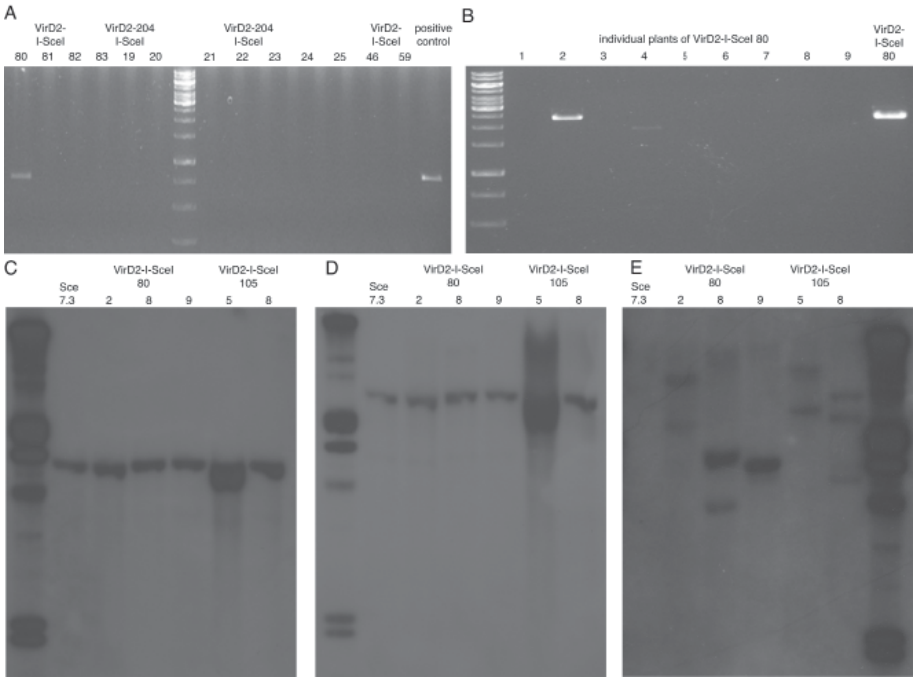

Figure 3: A: PCR with primers in *PPT* and *HPT* on DNA of different pools, indentifying VirD2-I-SceI 80 as a candidate for GT, and (B) on individual plants of pools VirD2-I-SceI 80, identifying plant #2 as the candidate for GT in pool 80. C: Southern blot analysis on individual plants #2, 8 and 9 of pool VirD2-I-SceI 80 and individual plants #5 and 8 of pool VirD2-I-SceI 105. Left lane: DIG III marker (Roche Applied Science) Samples were digested with *NcoI* and a probe in *PPT* was used. D: as C, but using a probe in *GUS*. E: as C, but using a probe in *HPT*. Samples were digested with *NdeI*.

Discussion

We have developed a method for simultaneous delivery via AMT of a DSB-inducing agent, I-SceI, and a GT template in the form of the T-strand. This DSB-inducing agent and the T-stand are covalently bound via VirD2, which is the *Agrobacterium* protein responsible for T-strand processing, and to the T4SS translocation signal of VirF. This creates a protein-DNA complex which is capable of inducing DSBs and delivery of the GT template for integration via HR at the same time.

Previously, we have demonstrated T-strand translocation via VirD2-I-SceI-F and VirD2-204-I-SceI-F fusion proteins. We have demonstrated that the fusion proteins still

possessed I-SceI activity when produced *in planta*.

It has been suggested that VirD2 is (partially) unfolded in order to be able to translocate through the T4SS (Atmakuri *et al.*, 2004, Christie, 2004). To our knowledge, this has not been tested experimentally. However, the crystal structure of the T4SS has recently been solved, and its representation shows that there is a narrow passage which a substrate likely has to pass (Fronzes *et al.*, 2009). Therefore, activity after *in planta* expression, indicating tolerance of I-SceI for fusions with protein domains at both its ends, may not be representative for its activity after AMT. Furthermore, T-strand formation will result in the covalent attachment of the T-strand to the VirD2-I-SceI-F fusion protein, which can also influence the I-SceI moiety. Apart from that, it is also expected that AMT will not bring high amounts of VirD2-T-strand complexes into the recipient cell. Therefore, it is crucial to test that the I-SceI moiety of VirD2-I-SceI-F retains its activity after translocation through the T4SS.

To determine the activity of VirD2-I-SceI-F and VirD2-204-I-SceI-F after AMT, we screened pools of transformants for damaged genomic I-SceI recognition sites. Evidence for the post-AMT activity of the novel recombinant proteins was found in 2 of 280 plants transformed by an *Agrobacterium* strain expressing VirD2-204-I-SceI, while control experiments using wild type VirD2 did not result in any footprints (Fig. 2). Activity of the I-SceI moiety of VirD2-I-SceI-F leads to detectable footprints in 3 out of 95 VirD2-I-SceI-F pools of 10 plants. We recovered the individual plants containing the footprint and they seemed to be heterozygous for the footprint (Fig. 2). This result can be explained by the fact that they were transformed by floral dip. The target tissues of floral dip are the female reproductive tissue and cells of the embryo (Desfeux *et al.*, 2000). An event at such an early stage of development leads to either a heterozygote or a chimeric plant with a substantial amount of tissue containing the footprint.

The nature of the I-SceI-induced damage in the present study consists of small deletions and insertions. This is in line with our earlier findings (Chapter 4) and with experiments performed in mammalian cells (Rouet *et al.*, 1994, Liang *et al.*, 1998). Apparently, just as in other systems, erroneous repair by NHEJ results in changes at the DSB site. These data indicate that the function of I-SceI is not, or only to a limited extent, disrupted by passage through the T4SS and the presence of the T-strand, although we cannot rule out that some VirD2 molecules are translocated without being bound to a T-strand.

Considering our data on the post-AMT activity of VirD2-I-SceI-F and VirD2-204-I-SceI-F, we can conclude that they retain their ability to create DSBs after AMT, and

detectable footprints can be recovered at a frequency of 1 transformant in 320 transformants generated by an *Agrobacterium* strain expressing VirD2-I-SceI-F and 1 transformant in 140 transformants generated by an *Agrobacterium* strain expressing VirD2-204-I-SceI-F, while in transformants generated by an *Agrobacterium* strain expressing wild type VirD2, no footprints could be discovered at all. The true DSB-inducing activity of I-SceI after AMT will be higher than the amount of damaged I-SceI recognition sites that can be recovered. This is because perfect repair is likely to occur frequently, since NHEJ is not necessarily error prone. Moreover, just as all HEs, I-SceI is known to display some flexibility in its recognition site (Jurica and Stoddard, 1999). Therefore, some mutated recognition sites cannot be detected, since they will still be cleaved by I-SceI.

To determine if the use of the chimeric VirD2 nuclease fusions, VirD2-(204)-I-SceI-F, can be instrumental for inducing HR at the cognate recognition site, we screened all pools for GT events. The T-strand translocated by VirD2-(204)-I-SceI-F contains regions of extensive homology to the target locus (Fig. 1). Successful GT events will result in the replacement of *GFP* by *HPT*, a strategy used in a previous study from our lab (De Pater *et al.*, 2009). Although a candidate plant was identified, Southern blot analysis showed that no actual GT event had taken place (Fig. 3). Further analysis showed that two to four random integrations per plant can be observed (Fig 3E). It has previously been shown that four to six transformations are normal for AMT via floral dip (De Buck *et al.*, 2009). Even though the amount of plants tested is low, the data indicate that fusion of VirD2 to I-SceI still results in a normal number of T-DNA integrations.

The efficiency of GT in this setup is of course dependent on the efficiency with which the I-SceI site is cleaved and on the time that the DSB is present. It is difficult to estimate the amount of protein that is translocated to a recipient cell. It is known that an *Agrobacterium* cell accumulates about 50 T-strands in 24 hours after induction of the *vir* genes (Atmakuri *et al.*, 2007). However, it is unknown if these T-strands are all translocated, and to how many recipient cells. Furthermore, some VirD2 molecules may also be translocated without being bound to a T-strand; translocation of unbound VirD2 has been demonstrated, although at a very low level (Vergunst *et al.*, 2005).

As indicated above, apart from the amount of protein delivered, another important factor is how efficiently the cell deals with DSBs, thus how long a DSB exists. It may be that DNA repair in the cell types targeted by floral dip is so efficient that virtually all of the lesions created by VirD2-I-SceI-F are swiftly repaired, without leaving a footprint (De Pater *et al.*, 2009).

To further chart the possibilities of VirD2-mediated genome engineering (VirGEN), more footprinting and GT events need to be generated, by simply screening more plants. Testing more target lines would be wise, since the efficiency of HR may depend on the genomic locus involved (D'Halluin *et al.*, 2008). VirGEN could also be improved by using a method for high throughput screening of GT events without having to use PCR, e.g. the cruciferin system in which GT events yield fluorescent seeds (Shaked *et al.*, 2005). Using another method of AMT to target somatic cells, rather than cells very early in development as are targeted in floral dip, will probably not yield higher frequencies of GT. It has been shown that early in development, cell are more likely to repair DSBs via HR than NHEJ. The older the cell, the less likely repair via HR becomes (Boyko *et al.*, 2006).

When it would truly come of age, VirGEN has as a major advantage in the fact that the nuclease is administered transiently. In addition to that, with *Agrobacterium* producing the nuclease moiety of interest, it is not necessary to purify proteins prior to their delivery to cells of interest. Interestingly, since I-SceI mutants with different recognition sites have been developed (Dojon *et* al., 2006, Chen *et al.*, 2009, Joshi *et* al., 2010, reviewed in Galetto *et al.*, 2009), VirGEN using I-SceI as the nuclease has potential also for newly developed target sites, This raises expectations of one day being able to target endogenous loci of choice with engineered HEs, in combination with *Agrobacterium*-mediated DNA and protein delivery.

Materials and methods

Cloning

Cloning was performed using standard techniques in *E. coli* strain DH5α. Both *E. coli* and *A. tumefaciens* were cultured in LC medium containing the appropriate antibiotics. The cloning of the constructs used in this study is described in Chapter 4.

Generation of plant line Sce 7.3

The generation of target *Arabidopsis* Col-0 line Sce 7.3 is described in Chapter 4. The position of the inserted locus was determined by TAIL-PCR using primers NOS1, 2, and 3 (De Pater *et al.*, 2009), and the degenerate primers AD2 (Liu *et al.*, 1995), using RedTaq polymerase (Sigma-Aldrich). Primer sequences are listed in Table 1. The TAIL-PCR product was excised

from gel, cloned into the pGEMT-easy vector (Promega) and sequenced. Through TAIL-PCR analysis (Liu *et al.*, 1995), we determined that the insert was located in chromosome 2, in the MATE efflux gene (AT2G38330.1). The remainders of the right border sequence are deleted, as well as the adjacent 5 bp.

Generation of pools

Agrobacterium strain LBA2585 (Bravo-Angel *et al*, 1998), containing gene targeting construct pSDM3834 (described in De Pater *et al.*, 2009) and the relevant VirD2 expression construct, was used to transform Sce 7.3 by floral dip (Clough and Bent, 1998). Seeds were sown on selection medium containing hygromycin (15 µg/mL) and transformants were rescued. Pools containing leaf material of 10 transformants were made and genomic DNA was extracted as described (De Pater *et al.*, 2009).

Analysis of footprints

Using primers SP250a and SP251 (De Pater *et al.*, 2009), a fragment containing the I-SceI site was amplified from pooled DNA material, using Phusion polymerase (Finnzymes). The product was digested overnight with I-SceI (New England Biolabs) and analyzed on a 2% agarose gel. Undigested PCR product was excised from gel, cloned into pJET1.2 (Fermentas) and sequenced. Of any pool containing a footprint, the individual plants were analyzed for the presence of the footprint.

Analysis of gene targeting

Two consecutive PCR reactions, using Phusion polymerase (Finnymes), were performed to determine the presence of GT events in pools of transformants. The first PCR was performed to amplify the entire genomic locus, using 1 µL genomic DNA, with primers SP283 (De Pater *et al.*, 2009) and RV2. In this PCR, initial denaturation of the template was performed at 98°C for 3 minutes. Then, 35 cycles of 98°C for 20 seconds, 60°C for 20 seconds and 72°C for 3 minutes were performed.

The second PCR was performed on 1 µL of a 1000-fold dilution of the product of the first PCR. Primers were BAR fw and SP251 (De Pater *et al.*, 2009) for the detection of *GFP*-specific PCR fragments, and BAR fw and SP284 (De Pater *et al.*, 2009) for the

detection of *HPT*-specific PCR fragments. Initial denaturation was performed at 98°C for 3 minutes. Then, 30 cycles of 98°C for 20 seconds, 60°C for 20 seconds and 72°C for 1 minute were performed. The PCR products were analyzed on 1% agarose gel.

Southern blot

Plant tissue was disrupted to a powder under liquid N_2 in a TissueLyser (Retch). DNA was isolated using a CTAB procedure (Murray and Thompson, 1980), and 5 µg of DNA was digested for Southern blot analysis, using either *Nde*I or *Nco*I (Fermentas), and separated on 0.7% agarose gel. It was then blotted onto Hybond-N (Amersham) and hybridized with DIG-labeled probes according to the manufacturer's instructions, supplemented with 50 µg/ml herring sperm DNA and a DIG-labeled probe for either *HPT*, *GUS*, or *PPT*. Detection was performed using the DIG wash and block buffer set and CDP-star, according to the manufacturer's instructions (Roche Diagnostics).

Probes were labeled in a PCR-reaction using DIG-labeling mix (Roche Diagnostics, Mannheim, Germany). Primers were MC141 and MC142 (De Pater *et* al., 2009) for *HPT*, BAR1 and BAR2 for *PPT*, and GUS and GUS3 for *GUS*.

Table 1: Primers

NOS1	GATTGAATCCTGTTGCCGGTCTT	(De Pater *et al*, 2009)
NOS2	GCATGACGTTATTTATGAGATGG	(De Pater *et al*, 2009)
NOS3	CGCAAACTAGGATAAATTATCGC	(De Pater *et al*., 2009)
AD2	NGTCGASWGANAWGAA	(Liu *et al*., 1995)
SP250a	CTCTGCCGTCTCTCTATTCG	
SP251	CTTGAAGAAGTCGTGCTGCTT	(De Pater *et al*., 2009)
SP284	CACGAGATTCTTCGCCCTCC	(De Pater *et al*., 2009)
BAR FW	GTCGAGATCTGGATTGAGAGTG	
RV2	GTCGCTGAGAAGAAGTGGAG	
BAR1	AACCCACGTCATGCCAGTTCC	
BAR2	CGGCGGTCTGCACCATCGTC	
MC141	CGATTCCGGAAGTGCTTGAC	(De Pater *et al*., 2009)
MC142	GGTCGGCATCTACTCTATTC	(De Pater *et al*., 2009)
GUS	AGACTGTAACCACGCGTCTG	
GUS3	GCCTAAAGAGAGGTTAAAGCC	

References

Atmakuri, K., Cascales, E., and Christie, P.J. 2004. Energetic compounds VirD4, VirB11, and VirB4 mediate early DNA transfer reactions required for bacterial Type 4 Secretion. Mol. Microbiol. 54(5):1199-1211.

Bibikova, H., Golic, M., Golic, K.G., and Carroll, D. 2002. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. Genetics. 161(3):1169-1175.

Boubakour-Azzouz, I., and Richetti, M. 2008. Low joining efficiency and non-conservative repair of two distant double-strand breaks in mouse embryonic stem cells. DNA Repair (Amst.). 7(2):149-161.

Boyko, A., Zemp, F., Filkowski, J, and Kovalchuk, I. 2006. Double-strand break repair in plants is developmentally regulated. Plant Physiol. 141:488-497.

Bravo-Angel, A.M., Hohn, B., and Tinland, B. 1998. The omega sequence of VirD2 is important but not essential for efficient transfer of T-DNA by *Agrobacterium tumefaciens*. Mol. Plant-Microbe Interact. 11:57-63.

Carroll, D. 1008. Zinc finger nucleases as gene therapy agents. Gene Ther. 15(22):1463-1468.

Chen, Z, Wen, F., Sun, N., and Zhao, H. 2009. Directed evolution of homing endonuclease I-SceI with altered sequence specificity. Protein Eng. Des. Sel. 22(4):249-256.

Chilton, M.D., and Que, Q. 2003. Targeted integration of T-DNA into the tobacco genome at double-stranded breaks: new insights on the mechanism of T-DNA integration. Plant Physiol. 133(3):956-965.

Christie, P.J. 2004. Type IV secretion: the *Agrobacterium* VirB/D4 and related conjugation systems. Biochem. Biophys. Acta. 1693:219-234.

Clough, S.J., and Bent, A.F. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J. 16(6):735-743.

De Buck, S., Podevin, N., Nolf, J., Jacobs, A., and Depicker, A. 2009. The T-DNA integration pattern in *Arabidopsis* transformants is highly determined by the transformed target cell. The Plant J. 60:134-145.

De Pater, S., Neuteboom, L.W., Pinas, J.E., Hooykaas, P.J., and van der Zaal, B.J. 2009. ZFN-induced mutagenesis and gene-targeting in *Arabidopsis* through *Agrobacterium*-mediated floral dip transformation. Plant Biotech. J. 7:821-835.

Desfeux, C., Clough, S.J., and Bent, A.F. 2000. Female reproductive tissues are the primary target of *Agrobacterium*-mediated transformation by the *Arabidopsis* floral dip method. Plant Physiol. 123:895-904.

D'Halluin, K., Vanderstraeten, C., Stals, E., Cornelissen, M., and Ruiter, R. 2008. Homologous recombination: a basis for targeted genome optimization in crop species such as maize. Plant Biotech.

J. 6(1):93-102.
Ding, Z., Atmakuri, A., and Christie, P.J. 2003. The ins and outs of bacterial Type IV Secretion System substrates. Trend Microbiol. 11(11):527-535.
Doyon, J.B., Pattanayak, V., Meyer, C.B., and Liu, D.R. 2006. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. J. Am. Chem. Soc. 128(7):2477-2484.
Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J., Gregory, P.D., Urnov, F.D., and Amacher, S.L. 2008. Heritable tagreted disruption in zebrafish using designed zinc-finger nucleases. Nat. Biotechnol. 26(6) :702-708.
Dujon, B. 1989. Group I introns as mobile genetic elements: facts and mechanistic speculations – a review. Gene. 82:91-114.
Durai, S., Manu, M., Kandavelou, K., Wu, J., Porteus, M.H., and Chandrasegaran, S. 2005. Zinc finger nucleases : custom-designed scissors for genome engineering. Nucleic Acids Res. 33(18):5978-5990.
Dürrenberger, F, Crameri, A., Hohn, B., and Koukolikova-Nicola., Z. 1989. Covalently bound VirD2 from *Agrobacterium tumefaciens* protects the T-DNA from exonucleolytic degradation. Proc. Natl. Acad. Sci. U.S.A. 86(23):9154-9158.
Gong, W.J., and Golic, K.G. 2003. Ends-out, or replacement, gene targeting in *Drosophila*. Proc. Natl. Acad. Sci. U.S.A. 100(5):2556-2561.
Hanin, M., Volrath, S., Bogucki, A., Ward, E., and Paszkowski, J. 2001. Gene targeting in *Arabidopsis*. Plant J., 28:671-677.
Halfter, U., Morris, P.C., and Willmitzer, L. 1992. Gene targeting in *Arabidopsis thaliana*. Mol. Gen. Genet. 231:186-193.
Hrouda, M., and Paszkowski, J. 1994. High fidelity extrachromosomal recombination and gene targeting in plants. Mol. Gen. Genet. 243:106-111.
Joshi, R., Ho, K.K., Tenney, K., Chen, J.H., Golden, B.L., and Gimble, F.S. 2010. Evolution of I-SceI homing endonuclease with increased DNA recognition site specificity. J. Mol. Biol. e-published before print.
Jurica, M.S., and Stoddard., B.L. 1999. Homing endonucleases: structure, function and evolution. Cell. Mol. Life Sci. 55:1304-1326.
Kim, Y.G., Cha, J., and Chandrasegaran, S. 1996. Hybrid restriction enzymes: zinc finger fusions to *Fok*I cleavage domain. Proc. Natl. Acad. Sci. U.S.A. 93(3):1156-1160.
Kuspa, A. Restriction enzyme-mediated integration (REMI) mutagenesis. 2006. Methods Mol. Biol. 346:201-209.
Lazo, G.R., Stein, P.A., and Ludwig, R.A. 1991. A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. Biotechnology. 9:963-967.
Lawley, T.D., Klimke, W.A., Gubbins, M.J. and Frost, L.S. 2003. F factor conjugation is a true Type IV secretion system. FEMS Microbiol. Lett. 224(1):1-15.
Lee, K.Y., Lund, P., Lowe, K., and Dunsmuir, P. 1990. Homologous recombination in plant cells

after *Agrobacterium*-mediated transformation. Plant Cell. 2:415-425.
Liang, F., Han, M., Romanienko, P.J., and Jasin, M. 1998. Homology-directed repair is a major double-strand repair pathway in mammalian cells. Proc. Natl. Acad. Sci. U.S.A. 95(9):5172-5177.
Liu, Y.-G., Mitsukawa, N., Oosumi, T., and Whitter, R.F. 1995. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal interlaced PCR. Plant J. 8:457-463.
Lloyd, A., Plaisier, C.L., Carroll, D., and Drews, G.N. 2005. Targeted mutagenesis using zinc finger-nucleases in *Arabidopsis*. Proc. Natl. Acad. Sci. U.S.A. 102(6):2232-2237.
Matzke, M.A., Mette, M.F., and Matzke, A.J.M. 2000. Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates. Plant Mol. Biol. 43:401-415.
Miao, Z.H., and Lam, E. 1995. Targeted disruption of the TGA3 locus in *Arabidopsis thaliana*. Plant J. 7:359-365.
Murray, M.G., and Thompson, W.F. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. 8(19):4321-4325.
Mysore, K.S., Bassuner, B., Deng, X.B., Darbinian, N.S., Motchoulsky, A., Ream, W., and Gelvin, S.B. 1998. Role of the *Agrobacterium tumefaciens* VirD2 protein in T-DNA transfer and integration. Mol. Plant.-Microb. Interact. 11:688-683.
Ninomiya, Y., Suzuki, K., Ichii, C., and Inoue, H. 2004. Highly efficient gene replacements in *Neurospora* strain deficient for nonhomologous end-joining. Proc. Natl. Acad. Sci. U.S.A. 101(33):12248-12253.
Offringa, R., de Groot, M.J., Haagsman, H.J., Does, M.P., van den Elzen, P.J., and Hooykaas, P.J.J. 1990. Extrachromosomal homologous recombination and gene targeting in plant cells after *Agrobacterium*-mediated transformation. EMBO (Eur. Mol. Biol. Organ.) J. 9:3077-3084.
Orel, N., Kyryk, A., and Puchta, H. 2003. Different pathways of homologous integration are used for the repair of double-strand breaks within tandemly arranged sequenced in the plant genome. Plant J. 35(5):604-612.
Rouet, P., Smith, F., and Jasin, M. 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell. Biol. 14(12):8096-8106.
Paszkowski, J., Baur, M., Bogucki, A., and Potrykus, I. 1988. Gene targeting in plants. EMBO (Eur. Mol. Biol. Organ.) J. 7:4021-4026.
Paques, F., and Duchateau, P. 2007. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr Gene Ther. 7(1):49-66.
Pitzschke, A, and Hirt, H. 2010. New insights into an old story: *Agrobacterium*-induced tumour formation in plants by plant transformation. EMBO (Eur. Mol. Biol. Organ.) J. 29(6):1021-1032.
Puchta, H., Dujon, B., and Hohn, B. 1996. Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc. Natl. Acad. Sci. U.S.A. 93:5505-5060.
Reiss, B., Klemm, M., Kosak, H., and Schell., J.1996. RecA protein stimulates homologous

recombination in plants. Proc. Natl. Acad. Sci. U.S.A. 93(7):3094-3098.
Risseeuw, E., Offringa, R., Franke-van Dijk, M.E.I., and Hooykaas, P.J.J. 1995. Targeted recombination in plants using *Agrobacterium* coincides with additional rearrangements at the target locus. Plant J. 7:109-119.
Scheiffele, P., Pansegrau, W., and Lanka, E. 1995. Initiation of *Agrobacterium tumefaciens* T-DNA processing. Purified proteins VirD1 and VirD2 catalyze site- and trans-specific cleavage of superhelical T-border DNA *in vitro*. J. Biol. Chem. 270:1269-1276.
Shaked, H., Melamed-Bessudo, C., and Levy, A.A. 2005. High-frequency gene targeting in *Arabidopsis* plants expressing the yeast *RAD54* gene. Proc. Natl. Acad. Sci. U.S.A. 102(34):12265-12269.
Shukla, V.K., Doyon, Y., Miller, J.C., DeKelver, R.C., Moehle, E.A., Worden, S.E., Mitchell, J.C., Arnold, N.L., Gopalan, S., Meng, X., Choi, V.M., Rock, J.M., Wu., Y.Y., Katibah, G.E., Zhifang, G., McCaskill, D., Simpson, M.A., Blakeslee, B., Greenwalt, S.A., Butler, H.L., Hinkley, S.J., Zhang, L., Rebar, E.J., Gregory, P.D., and Urnov, F.D. 2009. Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. Nature. 459(7254):437-441.
Siegl, T., Petzke, L., Welle, E., and Luzhetskyy, A. 2010. I-SceI endonuclease: a new tool for DNA repair studies and genetic manipulations in streptomycetes. Appl. Microbiol. Biotechnol. 87(4):1525-1532.
Stoddard, B. L., 2005. Homing endonuclease structure and function. Q. Rev. Biophys. 38(1). 49-95.
Tanaka, S, Ishii, C., Hatakeyama, C., and Inoue, H. 2010. High efficient gene targeting on the *AGAMOUS* gene in an *Arabidopsis AtLIG4* mutant. Biochem. Biophys. Res. Comm. 396:289-293.
Tinland, B., Schoumacher, F. Gloekler, V., Brova-Angel, A.M., and Hohn, B. 1995. The *Agrobacterium tumefaciens* virulence D2 protein is responsible for precise integration of T-DNA into the plant genome. EMBO (Eur. Mol. Biol. Organ.) J. 14:3585-3595.
Van Kregten, M, Lindhout, B.I., Hooykaas, P.J., and van der Zaal, B.J. 2009. *Agrobacterium*-mediated T-DNA transfer and integration by minimal VirD2 consisting of the relaxase domain and a Type IV Secretion System translocation signal. Mol. Plant-Microbe Interact. 22(11): 1356-1365.
Vergunst, A.C., Schrammeier, B., den Dulk-Ras, A., de Vlaam, C.M., Regensburg-Tuïnk, T.J., and Hooykaas, P.J. 2000. VirB/D4-dependent protein translocation from *Agrobacterium* into plant cells. 2000. Science. 290(5493):979-982.
Vergunst, A.C., van Lier, M. C., den Dulk-Ras, A., Stuve, T. A., Ouwehand, A., Hooykaas, P. J. 2005. Positive charge is an important feature of the C-terminal transport signal of the VirB/D4-translocated proteins of Agrobacterium. Proc. Natl. Acad. Sci. U.S.A. 102(3): 832-837.
Vogel, A. M., and Das, A. 1992. Mutational analysis of *Agrobacterium tumefaciens virD2*: tyrosine 29 is essential for endonuclease activity. J. Bacteriol. 174:303-308.
Weinthal, D., Tovkach, A., Zeevi, V., and Tzfira, T. 2010. Genome editing in plant cells by zinc finger nucleases. Trends Plant Sci. 15(6):308-321.
Wright, D.A., Townsend, J.A., Winfrey, R.J. Jr., Irwin, P.A., Rajagopal, J., Lonosky, P.M., Hall., B.D., Jondle, M.D., and Voytas, D.F. 2005. High-frequency homologous recombination in plants

mediated by zinc-finger nucleases. Plant J. 44(4):693-705.
Ziemienowicz, A., Tinland, B., Bryant, J., Gloeckler, V., and Hohn, B. 2000. Plant enzymes but not *Agrobacterium* VirD2 mediate T-DNA integration *in vitro*. Mol. Cell. Biol. 20:6317-6311.

Summary

Summary

Agrobacterium tumefaciens is a Gram-negative bacterium that naturally occurs in soil. In dicotyledonous plants, it causes the crown gall disease. *Agrobacterium* causes this disease by transferring a strand of DNA (the transferred or T-strand) of its tumor-inducing plasmid (pTi) into the plant cells. Once inside the plant cell, the T-strand enters the nucleus where it integrates into the genome of the plant. The genes encoded on the T-strand cause rapid plant cell division, resulting in the formation of a tumor, and the production of nutrients that can be used by *Agrobacterium*.

The T-strand region of the Ti plasmid is delimited by the two border sequences: the Left and Right Border (LB and RB). Any DNA sequence present between the LB and RB will be transferred to a plant cell. This allows for the generation of transgenic plants in which DNA sequences of choice have been integrated. In the laboratory, the natural host range of *Agrobacterium* has been expanded to include, amongst others, monocotyledonous plants and fungi, making *Agrobacterium* an invaluable tool for biotechnology.

The research described in this thesis was aimed at enlarging the understanding of the *Agrobacterium* protein VirD2. Using its relaxase domain to nick the border sequences delimiting the T-region, VirD2 is responsible for processing of the T-strand, to which it remains covalently attached. In addition, the translocation of the T-strand into the recipient cell is directed by VirD2. We aimed to elucidate the requirements of different domains of VirD2 for translocation. Furthermore, we also developed VirD2 into a tool for site-directed mutagenesis in plants. We used the model plant *Arabidopsis thaliana* (thale cress), since rapid transformation techniques are available and the plant has a relatively short life cycle.

Chapter 1 describes in detail the mechanism of *Agrobacterium*-mediated transformation (AMT) and its biotechnological applications. Mechanistically, AMT is related to plasmid conjugation and pathogenesis of certain bacteria in that the translocation of DNA and proteins occurs via a Type 4 Secretion System (T4SS). The T4SS is a large protein structure, adapted to translocate protein (and protein-bound DNA) to recipient cells. T-strands preferentially integrate into double-strand breaks (DSBs) in the genome of the recipient cell. To artificially induce DSBs in eukaryotic cells, two classes of proteins are commonly used: the zinc finger nucleases and the homing endonucleases, both with advantages and drawbacks.

The research described in **Chapter 2** was aimed at defining the essential domains of VirD2. To this end, we created VirD2 truncations, consisting of the relaxase domain – responsible of the processing of the T-strand – while a large domain of unknown function (DUF) was not at all or only partially present. For its translocation to plant cells, a translocation signal was required at the C-terminal end of VirD2. When appropriate, we replaced it by the translocation signal of the unrelated protein VirF. In that manner, we found that DUF can be deleted completely when the translocation signal of VirF is present. When the endogenous C-terminal sequences of VirD2 are present, only 60 amino acid residues of DUF (DUF-60) are required for translocation. The fact that the translocation signal of an unrelated protein can restore T-strand transformation provides further evidence that T-strand translocation is protein-driven, rather than driven by the T-strand, as suggested by earlier models regarding the function of the T4SS of *Agrobacterium*. In this study, we have provided compelling evidence supporting a change of opinion regarding the T4SS function.

In **Chapter 3**, the role of DUF-60 is further investigated. The results of Chapter 2 could be explained by assuming DUF-60 is required for an interaction, either with a factor within plants or a factor within *Agrobacterium*. The C-terminal end of VirF can apparently substitute for this function, while the C-terminal end of VirD2 alone is insufficient and required support from DUF-60. We investigated this by inoculating *Agrobacterium* strains expressing relevant VirD2 mutants onto two different plant species, *Nicotiana glauca* and *Kalanchoë tubiflora*. These experiments showed that DUF-60 does not fulfill a role in these plant species. Instead, using GFP-tagged VirD2 mutants, we discovered that DUF fulfills an important function with *Agrobacterium* cells. We found that DUF is of crucial importance for the localization of VirD2 to the cell poles of *Agrobacterium*, and polar localization of VirD2 via DUF-60 therefore seems essential for the virulence of an *Agrobacterium* strain. Remarkably, the C-terminal end of VirF can restore polar localization of VirD2-204, and can therefore substitute for DUF-60.

It is known that induction of DSBs in the genome induces DNA repair via homologous recombination, which is the desired pathway for the correct integration of foreign DNA, such as the T-strand. The technology developed thus far employs zinc finger nucleases (ZFNs) or homing endonucleases (HEs) for induction of site-specific DSBs in complex genomes. VirD2 is potentially a very interesting candidate protein for mediating site-directed mutagenesis in plants. It has the unique advantage that it is covalently bound to the T-strand, which can be used as repair template for the incorporation of desired mutations into the plant genome. In **Chapter 4** experiments concerning novel types of

DSB-inducing fusion proteins based on VirD2 are described. Fusions were made consisting of the VirD2 protein, or a large part thereof, combined with 1 up to 6 zinc finger (ZF) domains, as well as complete ZFNs. Furthermore, we made fusions of VirD2 and two different homing endonucleases, I-SceI and HO. To ensure passage through the T4SS, we added the translocation signal of VirF (F). It was then tested if these proteins could still be translocated from *Agrobacterium* to *Arabidopsis* root cells. Remarkably, all fusion proteins could be translocated. However, the translocation efficiency varied widely between the different fusion proteins. An increasing number of ZFs caused a steep decline in translocation efficiency of VirD2-based fusion proteins. While fusion of VirD2 to HO also severely decreased translocation efficiency, similar types of fusions to I-SceI still translocated at high efficiency and were shown to display I-SceI nuclease activity when expressed *in planta*.

In **Chapter 5** we determined that the VirD2-I-SceI-F and VirD2-204-I-SceI-F fusion proteins remained active after AMT. Since passage through the T4SS may involve (partial) unfolding of the protein, this was not self-evident. Activity was determined by screening for damaged I-SceI recognition sites in a plant line in which a targeting locus was pre-inserted. In 3 out of 950 plants, a damaged I-SceI recognition site was discovered. For the development of strategies for gene targeting, the process in which a particular gene is altered without disturbing the rest of the genome, the VirD2-I-SceI-F fusion protein holds promise. This is because the VirD2-I-SceI fusion protein is administered transiently and is covalently bound to the T-strand, ensuring the T-strand is in the vicinity of the DSB when it is made. Also, current advances in modification of HE recognition specificity will result in the creation of HEs suitable to create DSBs in endogenous loci.

In conclusion, the data presented in this thesis reveal previously unknown functions of VirD2. The requirements of VirD2 domains for T-complex translocation have been elucidated and a function has been assigned to the previously uncharacterized DUF domain. Following up on these results, we have created novel recombinant effector proteins that combine the T-strand processing and translocation properties of VirD2 and the (potential) DSB-inducing activity of a foreign protein. The VirD2-I-SceI-F fusion protein turned out to be translocated at a high efficiency, and showed DSB-inducing activity *in planta* after *Agrobacterium*-mediated transformation. These results show that VirD2 is a promising candidate for further development into a tool for the induction of transient effects in organisms susceptible to *Agrobacterium*-mediated transformation.

Nederlandse samenvatting

Samenvatting

Agrobacterium tumefaciens is een Gram-negatieve bacterie, die van nature voorkomt in de grond. In dicotyle planten veroorzaakt hij een ziekte die zich uit in de vorming van kroongallen. *Agrobacterium* doet dit door een enkelstrengs kopie van een stukje DNA (de T-streng) van zijn tumor-inducerende plasmide (pTi) over te dragen naar de plantencel. Vanuit het cytosol gaat het VirD2-T-streng complex naar de celkern, waar de T-streng in het genoom van de plant kan integreren. De genen op het in het genoom geïntegreerde DNA zorgen voor celdeling, waardoor de kroongallen ontstaan, en waarin voedingsstoffen worden geproduceerd voor *Agrobacterium*.

De T-gebied wordt geflankeerd door indentieke sequenties in het Ti plasmide: de linker en rechter border (LB en RB). Elke DNA-sequentie die tussen LB en RB ligt, zal door *Agrobacterium* worden overgedragen naar de plant. Hierdoor is het mogelijk om planten te genereren waarin elke gewenste DNA-sequentie in het genoom is opgenomen. In het laboratorium is het mogelijk om *Agrobacterium* naast zijn natuurlijke gastheren ook onder andere monocotyle planten en schimmels, waaronder ook bakkersgist, te laten transformeren. *Agrobacterium* heeft zich hiermee onmisbaar gemaakt in de biotechnologie.

Het onderzoek waarvan de resultaten zijn beschreven in dit proefschrift was gericht op het eiwit VirD2 dat een belangrijke functie vervult in *Agrobacterium*. Het relaxase domein van VirD2 is verantwoordelijk voor de productie van de T-streng. VirD2 blijft covalent gebonden aan de T-streng na de productie hiervan. Verder is VirD2 ook verantwoordelijk voor de translocatie van de T-streng naar de plantencel. Wij hebben de verschillende domeinen waaruit VirD2 bestaat onderzocht, waardoor duidelijk is geworden welke domeinen absoluut nodig zijn voor de translocatie van de T-streng. Ook hebben we VirD2 ontwikkeld tot werktuig voor het maken van gerichte mutaties in het plantengenoom. Hierbij hebben we *Arabidopsis thaliana* (de zandraket) gebruikt als modelorganisme, omdat voor deze plant snelle en makkelijke transformatietechnieken beschikbaar zijn en de plant een korte levenscyclus heeft.

In **Hoofdstuk 1** wordt het mechanisme van *Agrobacterium*-gemedieerde transformatie (AMT) in detail beschreven, evenals de biotechnologische toepassingen hiervan. AMT lijkt qua mechanisme erg op conjugatie van plasmiden en pathogenese van verschillende bacteriën; alle maken gebruik van een Type 4 Secretiesysteem (T4SS) om DNA

en/of eiwitten over te dragen naar andere cellen. Het T4SS is een groot complex bestaande uit verschillende eiwitten en is geëvolueerd om eiwit (en DNA dat gebonden is aan eiwit) te transporteren naar andere cellen. T-strengen integreren preferentieel in dubbelstrengs breuken (DSBs) in het genoom van de ontvangende cel. Om artificiële DSBs te induceren worden voornamelijk twee klassen eiwitten gebruikt: de zinc finger nucleases en de homing endonucleases. Beide klassen hebben hun eigen voor- en nadelen.

Het onderzoek beschreven in **Hoofdstuk 2** was gericht op het vinden van de essentiële domeinen van VirD2. Hiertoe hebben we truncaties van VirD2 gemaakt, bestaande uit het relaxase domein – verantwoordelijk voor de productie van de T-streng- terwijl het domein met onbekende functie (DUF) niet of slechts gedeeltelijk aanwezig was. Translocatie wordt mogelijk gemaakt door een translocatiesignaal aan het C-terminale einde van het eiwit. Wanneer van toepassing hebben we dit signaal vervangen door het translocatiesignaal van het eiwit VirF, dat niet verwant is aan VirD2.

We hebben gevonden dat DUF volledig kan worden vervangen door het translocatiesignaal van VirF. Wanneer het eigen C-terminale einde van VirD2 aanwezig is, is de helft van DUF (60 aminozuurresiduen; DUF-60) noodzakelijk voor translocatie. De vinding dat het translocatiesignaal van een eiwit dat niet verwant is aan VirD2, wel de translocatie van de T-streng kan herstellen, is bewijs dat de translocatie van VirD2 en de T-streng mogelijk wordt gemaakt door VirD2, en niet door de T-streng. Dit was eerder gepostuleerd, maar onze experimenten hebben hier weer nieuw experimenteel bewijs aan toegevoegd.

In **Hoofdstuk 3** is de rol van DUF-60 verder uitgezocht. De resultaten beschreven in hoofdstuk 2 kunnen worden verklaard door aan te nemen dat DUF een factor is die belangrijk is voor interacties in de gastheer, of in *Agrobacterium* zelf. Het C-terminale einde van VirF kan kennelijk dienen als vervanging voor deze functie, terwijl het C-terminale einde van VirD2 zelf onvoldoende is en ondersteuning van DUF-60 nodig heeft. We hebben dit onderzocht door twee soorten planten, *Nicotiana glauca* en *Kalanchoë tubiflora*, te infecteren met *Agrobacterium*-stammen die verschillende VirD2-mutanten tot expressie brengen. Uit deze experimenten konden wij concluderen dat DUF-60 niet van belang is voor het infecteren van deze soorten. In plaats hiervan hebben we ontdekt dat DUF-60 een cruciale rol speelt in *Agrobacterium*. Door VirD2 te fuseren met GFP (een eiwit dat groene fluorescentie vertoont), hebben we ontdekt dat DUF-60 cruciaal is voor de lokalisatie van VirD2 aan de polen van de cel, en de mate van deze zogeheten polaire lokalisatie bleek te correleren met de virulentie van een *Agrobacterium*-stam. Opmerkelijk genoeg is het

C-terminale einde van VirF voldoende om polaire lokalisatie van VirD2-204 te herstellen. Het C-terminale einde van VirF kan kennelijk de functie van DUF-60 overnemen.

Het is bekend dat inductie van DSBs in het genoom, DNA reparatie via homologe recombinatie bevordert. Via dit mechanisme kan vreemd DNA, zoals de T-streng, foutloos geïntegreerd worden in het plantengenoom. In biotechnologische strategieën gericht op de specifieke inductie van enkele DSBs in complexe genomen, worden DSBs gewoonlijk geïnduceerd door zinc finger nucleases (ZFNs) en homing endonucleases (HEs). VirD2 is in potentie een interessante kandidaat als vehikel voor gerichte mutagenese in planten. Het heeft als uniek voordeel dat het covalent gebonden is aan de T-streng, die gebruikt kan worden om de gewenste homologe DNA-sequenties in de cel te brengen In **Hoofdstuk 4** zijn experimenten beschreven met fusie-eiwitten op basis van VirD2. Fusie-eiwitten zijn gemaakt van VirD2 en 1 tot en met 6 zinc finger (ZF) domeinen en complete zinc finger nucleases (ZFN). Ook zijn fusie-eiwitten gemaakt met de homing endonucleases I-SceI en HO. Om translocatie door het T4SS te garanderen, hebben we het translocatiesignaal van VirF (F) toegevoegd. Daarna is getest of deze eiwitten nog steeds getransloceerd kunnen worden van *Agrobacterium* naar wortelcellen van *Arabidopsis*. Alle geteste eiwitten konden inderdaad getransloceerd worden. We hebben gevonden dat de translocatie-efficiëntie daalt met een toenemende hoeveelheid ZF-domeinen. Fusie van VirD2 aan HO zorgt er ook voor dat het fusie-eiwit met verminderde efficiëntie transloceert. Een fusie van VirD2 en I-SceI wordt met hoge efficiëntie getransloceerd en bleek I-SceI-specifieke nuclease-activiteit te vertonen na expressie *in planta*.

In **Hoofdstuk 5** presenteren wij data waaruit blijkt dat de fusie-eiwitten VirD2-I-SceI-F en VirD2-204-I-SceI-F hun activiteit behouden na AMT. Omdat translocatie door het T4SS kan inhouden dat het eiwit (gedeeltelijk) ontvouwen wordt, was dit niet vanzelfsprekend. De activiteit van I-SceI in deze fusie-eiwitten is bepaald door te screenen voor beschadigde I-SceI herkenningssites in een plantenlijn waarin een herkenningssite was ingebracht. In 3 van de 950 planten is een beschadigde I-SceI-herkenningssite ontdekt. De hier beschreven fusie-eiwitten zijn veelbelovend voor het ontwikkelen van strategieën voor gene targeting, het proces waarin een specifiek gen veranderd wordt zonder dat er een effect is op de rest van het genoom. Deze eiwitten zijn veelbelovend, omdat de fusie-eiwitten slechts tijdelijk aanwezig zijn en bovendien covalent gebonden zijn aan de T-streng, waardoor de T-streng automatisch in de buurt is van een DSB wanneer die wordt gemaakt. Onderzoek gericht op het aanpassen van de herkenningssequentie van HEs zal in de toekomst varianten van HEs opleveren die endogene loci in een complex genoom zullen herkennen.

In dit proefschrift presenteer ik data over voorheen onbekende functies van VirD2 van *Agrobacterium tumefaciens*. Het is duidelijk geworden welke domeinen van VirD2 precies nodig zijn voor translocatie van T-strengen en we hebben een functie kunnen toekennen aan DUF. Vervolgens hebben we nieuwe recombinante relaxase/ effector-eiwitten geproduceerd die zowel T-strengen kunnen produceren als deze kunnen transloceren via het T4SS, en bovendien (potentieel) een DSB kunnen veroorzaken. Vooral het VirD2-I-Sce-F fusie-eiwit bleek met hoge efficiëntie getransloceerd te worden en bleek in staat te zijn om DSBs ter veroorzaken *in planta* na *Agrobacterium*-gemedieerde transformatie. Deze resultaten bewijzen dat VirD2 een interessante kandidaat is voor verdere ontwikkeling tot biotechnologisch hulpmiddel voor de inductie van tijdelijke effecten in organismen die getransformeerd kunnen worden via *Agrobacterium tumefaciens*.

Curriculum Vitae

Curriculum Vitae

Maartje van Kregten werd geboren op 10 mei 1982 in Den Haag. In 2000 behaalde zij het VWO-diploma aan de Dalton Scholengemeenschap Den Haag. In hetzelfde jaar startte zij haar studie bio-farmaceutische wetenschappen aan de Universiteit van Leiden. Tijdens haar studie liep zij stages bij de vakgroep Toxicologie van het Leiden Amsterdam Centre for Drug Research, bij de vakgroep Moleculaire Genetica van het Leiden Institute for Chemistry en bij de Neuroscience Unit aan de universiteit van Helsinki. Ze studeerde af in 2006. Van 2006 tot 2010 werkte ze aan haar promotieonderzoek bij de vakgroep Moleculaire Ontwikkelingsgenetica van het Instituut Biologie Leiden, waarvan de resultaten beschreven staan in dit proefschrift. Van augustus tot december 2010 heeft ze dit werk voortgezet bij Add2X Biosciences.