VirD2 of Agrobacterium tumefaciens : functional domains and biotechnological applications

Kregten, M. van

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).

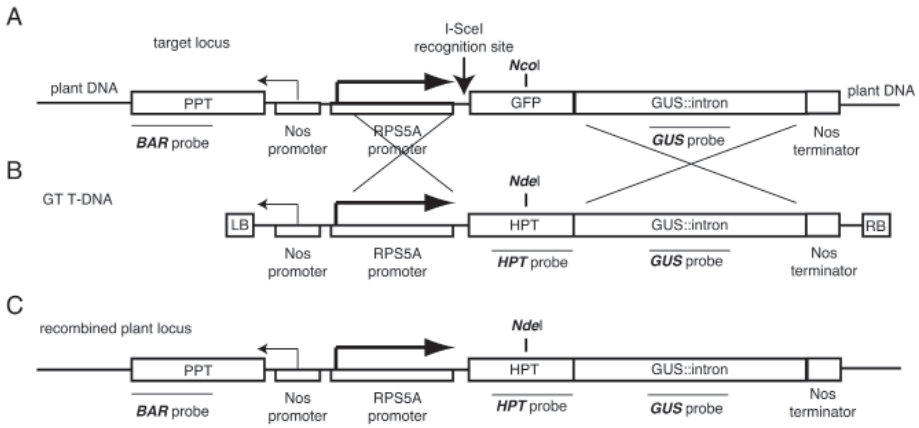

Figure 1: A: The stably integrated targeting locus, with an I-SceI recognition site between the Rps5a promoter and *GFP*. B: the gene targeting T-DNA. LB: left border, RB: right border. Homology to the targeting locus is indicated. C: the locus after a successful GT-event. Note the exchange of *GFP*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

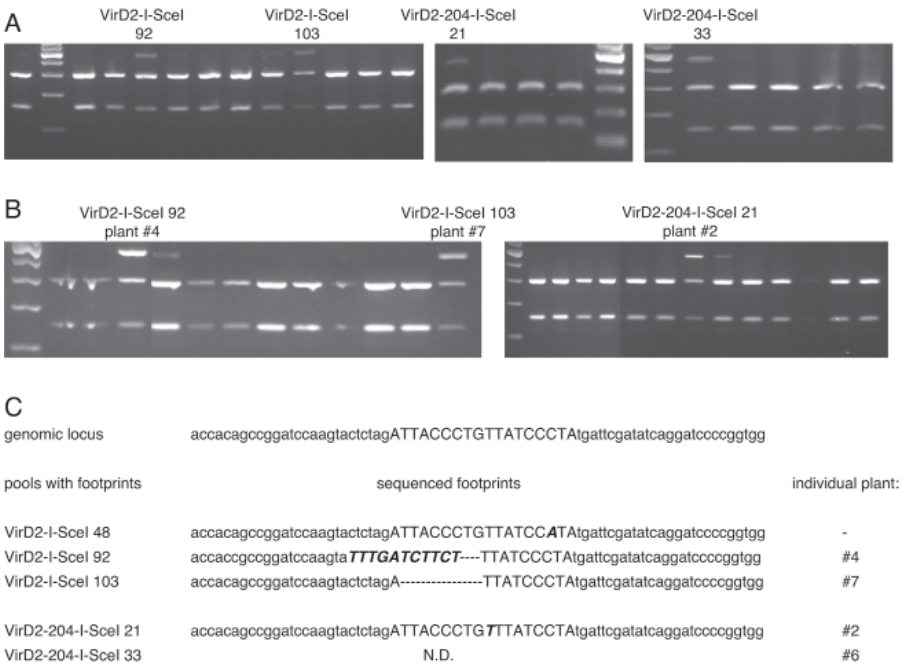

Figure 2: A: pools VirD2-I-SceI 92 and 103, and pools VirD2-204-I-SceI 21 and 33 are representative examples of pools that show an I-SceI-resistant band after digestion, B: individual plants of pools VirD2-I-SceI 92 (plant #4) and 103 (plant #7) and of pool VirD2-204-I-SceI 21 (plant #2) show an I-SceI-resistant 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.

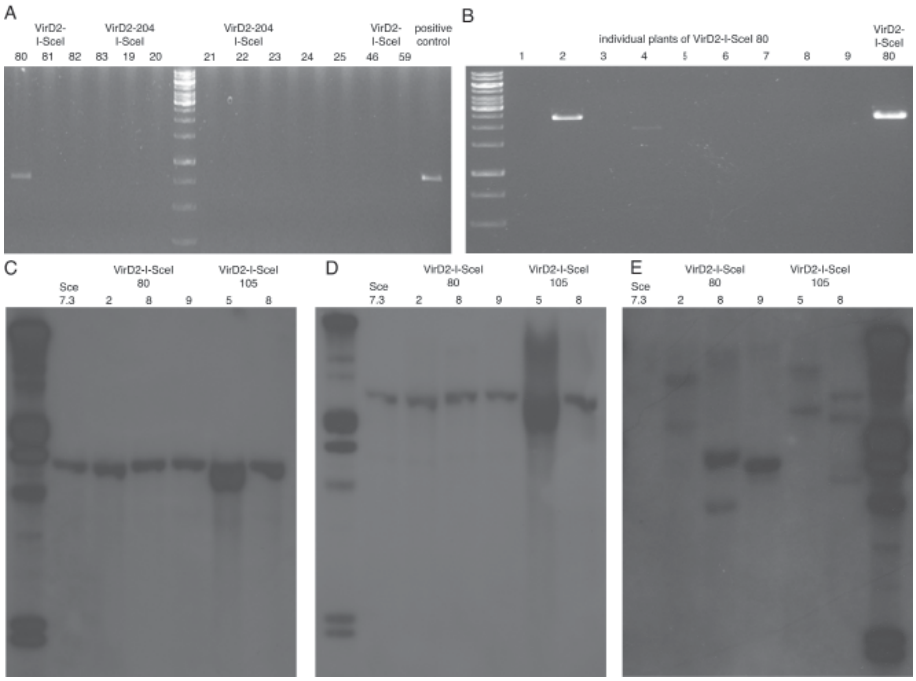

Figure 3: A: PCR with primers in *PPT* and *HPT* on DNA of different pools, indentifying VirD2-I-SceI 80 as a candidate for GT, and (B) on individual plants of pools VirD2-I-SceI 80, identifying plant #2 as the candidate for GT in pool 80. C: Southern blot analysis on individual plants #2, 8 and 9 of 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.