2015

Genetic engineering of Agrobacterium tumefaciens to target chloroplasts and identification of a novel nuclear localization signal of VirD2 protein

Matsuoka, Aki

Lethbridge, Alta : University of Lethbridge, Dept. of Biological Sciences

GENETIC ENGINEERING OF *AGROBACTERIUM TUMEFACIENS* TO TARGET CHLOROPLASTS AND IDENTIFICATION OF A NOVEL NUCLEAR LOCALIZATION SIGNAL OF VIRD2 PROTEIN

AKI MATSUOKA
B.Sc. University of Lethbridge, 2010

A Thesis
Submitted to the School of Graduate Studies
of the University of Lethbridge
in Partial Fulfilment of the
Requirements for the Degree

DOCTOR OF PHILOSOPHY
IN
BIOMOLECULAR SCIENCE

Biological Sciences
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

GENETIC ENGINEERING OF *AGROBACTERIUM TUMEFACIENS*
TO TARGET CHLOROPLASTS AND
IDENTIFICATION OF A NOVEL NUCLEAR LOCALIZATION SIGNAL OF VIRD2 PROTEIN

AKI MATSUOKA

Date of Defence: August 26, 2015

Dr. A. Ziemienowicz Co-supervisor	Research Scientist	Ph.D.
Dr. I. Kovalchuk Co-supervisor	Professor	MD, Ph.D.
Dr. J. Thomas Thesis Examination Committee Member	Professor	Ph.D.
Dr. F. Eudes Thesis Examination Committee Member	Research Scientist	Ph.D.
Dr. S. Rood Internal Examiner	Professor	Ph.D.
Dr. P. Maliga External Examiner Waksman Institute, Rutgers, The State University of New Jersey, USA	Professor	Ph.D.
Dr. T. Burg Chair, Thesis Examination Committee	Associate Professor	Ph.D.

Dedication

To my family

Kunikazu, Miyo, Tadaaki, Takayuki, Nobuko

ABSTRACT

Agrobacterium tumefaciens has been adopted to become a greatest tool for plant genetic engineering. Since Agrobacteria predominantly integrate single copy of the transgene into the nuclear genome, the transgene expression is relatively low. In contrast, transgene expression from chloroplasts allows high expression. Therefore, my first objective was to develop a new method of chloroplast transformation by modifying *Agrobacterium* genetics. I attempted to transform tobacco chloroplasts by *Agrobacterium* carrying the VirD2 protein with inactivated nuclear localization signals (NLSs) fused to a chloroplast transit peptide. However, the transformation of tobacco chloroplasts was unsuccessful. It suggested that the transit peptide was overridden by a novel cryptic NLS of VirD2. My second objective was to determine a novel NLS in VirD2. I discovered the novel NLS in VirD2 by subcellular localization analyses of fusion proteins (fragment of VirD2-green fluorescent protein) in plant cells. My discovery contributed to characterization of VirD2.

ACKNOWLEDGEMENTS

I am truly grateful for what I have experienced under the supervision of Dr. Alicja Ziemienowicz and Dr. Igor Kovalchuk. I cannot describe how much I appreciate their support other than saying thank you. Thank you for giving me the opportunity to be your student. Thank you for all your time, effort, guidance and support in pushing me forward to complete my program. Thank you for being fantastic supervisors.

I would like to thank my committee members Dr. François Eudes and Dr. Jim Thomas for their advice and support. I always felt their sincere kindness and enthusiasm towards my success.

I sincerely appreciate Dr. Pal Maliga for agreeing to be my external examiner. Because of you, I had a chance to present my research at the Plant Transformation Technologies III conference in Vienna in 2014. Thank you for giving me the amazing opportunity. I would like to thank Dr. Stewart Rood for being my internal examiner. I hope my research is interesting to you. I would like to thank Dr. Theresa Burg for being the chair of my thesis defence. I hope I entertained you during my defence. I am truly thankful to Dr. Henry Daniell at the University of Pennsylvania, Philadelphia for providing me with the chloroplast transformation vector. Your kind gift led my project to a new direction. I am also gratefully thankful to Dr. Jonathan Monroe at James Madison University, Virginia, for providing me with the binary vectors. I was able to identify a novel NLS in VirD2 because of your kind, caring gift.

I would like to thank all past and present people at both Kovalchuks' labs for helpful discussions and technical assistance, including Dr. Andrey Golubov for his suggestion on the transit peptides, and Nina, Rommy and Vladimir for their green thumbs. Special thanks to Dr. Andriy Bilichak for our discussions which helped me to connect the dots in my head. I would like to thank Valentina Titova for correcting my thesis. I would like to thank Neema, Dr. Elizabeth Schultz and Doug Bray for assistance and suggestions with confocal microscopy.

I would like to thank the University of Lethbridge and the Alberta Innovates Technology Futures for the financial support, and the University Library for letting me work throughout my undergraduate and graduate programs.

Finally, I would like to thank everyone I met in Lethbridge including people on campus and off. I am so fortunate I met such amazing people. They are treasures in my life. Thank you all for being you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

aadA – aminoglycoside 3’ adenylyltransferase gene
ACE2 – angiotensin-converting enzyme 2
AGO2 – ARGONAUTE2
Ang-(1-7) – a cleaved product of Angiotensin II
aphA-6 – aminoglycoside (3’) transferase (APH(3’)-VI) gene
ATPases – adenosine triphosphatase
AvrRps4 – *Pseudomonas savastanoi* Type III effector
BABA – beta-aminobutyric acids
bar – *BIALOPHOS RESISTANCE* gene
BiFC – biomolecular fluorescence complementation
bp – base pair
BR – basic regions
BSA – bovine serum albumin
CAK– CDK-activating kinase
CaMV – cauliflower mosaic virus
CDKs – cyclin-dependent kinases
cDNA – complementary DNA
cNLS – classical nuclear localization signal
CRISPER/Cas system – regularly interspaced short palindromic repeats/CRISPR-associated system
crRNA – CRISPR RNA
CsA – cyclosporin A
CTB – cholera toxin B subunit
CTD – C-terminal domain
DAPI – 4′,6-diamidino-2-phenylindole
DCL-TP – tomato defective chloroplasts and leaves transit peptide
Δ – deletion
DNA – deoxyribonucleic acid
DNase – deoxyribonuclease
dNTP – deoxynucleotide 5′ triphosphate
EBV – Epstein-Barr virus
E. coli – *Escherichia coli*s
EGFP – enhanced green fluorescent protein
EPSP – 5-enolpyruvylshikimate-3-phosphate synthase gene
EBNA-1 – EBV nuclear antigen 1
GFP – green fluorescent protein
GOI – gene of interest
GUS – β-glucuronidase
HIV – human immunodeficiency virus
Hop – Hsp70/Hsp90 organizing protein
HPPD – 4-hydroxyphenylpyruvate dioxygenase
Hsp70 – heat shock protein 70
Hsp90 – heat shock protein 90
IgG – Immunoglobulin G
IPTG – isopropyl-beta-D-thiogalactopyranoside
IR region – Inverted repeat region
kDa – kilodaltons
LB – T-DNA left border
LB medium – Luria-Bertani broth medium
L-Gln – L-glutamine
LOH1– ceramide synthase
MBP – myelin basic protein
MS medium – Murashige and Skoog medium
mVirD2 – modified VirD2
NBT/BCIP – nitro-blue tetrazolium chloride and 5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt

List of Abbreviations (Continued)

Nef – *Negative Regulatory Factor*
NIa – a tobacco etch virus protein containing plant functional NLS
NLS – nuclear localization signal
NPC – nuclear pore complex
nptII – neomycin phosphotransferase II gene
NtGPDL – tobacco glycerophosphodiesterase-like protein
OD_{600} – the optical density of bacterial cell culture at λ=600 nm
ω domain – omega domain
OEP61 – outer envelope membrane protein 61
P5βR2 gene – progesterone 5β-reductase gene
PAT– phosphinotricin acetyl transferase
PCR – polymerase chain reaction
PEG – polyethylene glycol
PP2C – 2C Ser/Thr protein phosphatase
P*psbA* – promoter and 5' regulatory element of plastid D1 protein of photosystem II
prFdIII – non-photosynthetic ferredoxin III
prRBCS – ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit
Prrn – plastid RNA operon promoter
pTi – tumor inducing plasmid
RanGAP – Ran GTPase activating protein
RB – T-DNA right border
RbcS-TP – tobacco ribulose bisphosphate carboxylase (Rubisco) small subunit transit peptide
RIM – root inducing medium
RNase- ribonuclease
SCF – Skp1-Cdc53-F box protein
SDS-PAGE – Sodium dodecyl sulphate Polyacrylamide gel electrophoresis
SIM – shoot inducting medium
SNAT – serotonin *N*-acetyl transferase
SPP – the stromal processing peptidase
ssDNA – single stranded DNA
SV40 – simian virus 40
T1 – first generation transgenic plant
T4SS – type IV secretion system
TALE – transcription activator-like effectors
TALEN – transcription activator-like effector nuclease
Taq – *Thermus aquaticus*
TBE – Tris/borate/EDTA
TBP – TATA binding protein
T-complex – VirD2 covalently attached to the 5' end of the T-strand which is coated with VirE2 proteins
T-DNA – transferred DNA
T-DNA – VirD2 complex – VirD2 covalently attached to the 5' end of the T-strand
TIC – translocon at the inner envelope membrane of chloroplasts
TOC – translocon at the outer envelope membrane of chloroplasts
TP – transit peptide
TPR– tetratricopeptide repeat
T*psbA* – 3' regulatory element of plastid D1 protein of photosystem II
Tris – Tris(hydroxymethyl)aminomethane
trnA – plastid tRNA-Alanine gene
trnI – plastid tRNA-Isoleucine gene
T*rps16* – 3'untranslated region of the plastid ribosomal protein S16 gene
TRV – *Tobacco rattle virus*
TSP – total soluble protein
T-strand – a single stranded T-DNA
VBF – VIP1-binding F-box

List of Abbreviations (Continued)

VIGS – virus-induced gene silencing
VIP1 – VirE2 interacting protein1
vir region – virulence region
Vir protein – virulence protein
Y29 – tyrosine 29
YEB medium – yeast extract beef extract medium
YEP medium – yeast extract peptone medium
YFP – yellow fluorescent protein
ZFN – zinc-finger nuclease

Chapter 1: Introduction

1.1. *Agrobacterium*-mediated plant transformation

Agrobacterium tumefaciens, a Gram-negative phytopathogenic bacterium, is often called a *natural genetic engineer* because it can perform an intricate plant transformation to cause plant crown gall disease. Once the main principle of the transformation was understood, *Agrobacterium* was the most popular tool to introduce transgenes into the plant nuclear genome. It has been just over thirty years since first successful *Agrobacterium*-mediated plant transformation (Barton et al., 1983). Since then, extensive studies on *Agrobacterium*-mediated plant transformation have revealed that many bacterial proteins and plant host proteins are involved in the complex transformation processes. The newly acquired knowledge was applied to address the challenges associated with the *Agrobacterium*-mediated transformation. This resulted in an expansion of the host range, improvement of transformation efficiency and the ability to integrate the transgene precisely into the host genome.

1.1.1. Transformation overview

Agrobacterium carries an octopine or a nopaline tumor inducing plasmid (pTi) consisting of several regions including the transferred DNA (T-DNA) region and the virulence (*vir*) region (Gelvin, 2003). The T-DNA region is edged by 25 bp direct repeats called the right border (RB) and the left border (LB). The T-DNA contains genes encoding enzymes involved in biosynthesis of plant growth hormones including indole acetic acid (an auxin) from tryptophan and isopentenyl adenosine monophosphate (a cytokinin) from adenosine monophosphate, and genes encoding enzymes involved in opine synthesis and secretion (Ream, 2008). The *vir* region includes the *vir* genes encoding Vir proteins essential for the transformation process such as VirA, VirB, VirC, VirD, VirE, VirF and VirG. These proteins function in recognition of plant signals, processing and transfer of the T-DNA. The transformation process includes the following steps (Figure 1.1.1). First, phenolic compounds, which are secreted from the wounded plant, are sensed by the VirA/VirG signal transduction system, resulting in activation of the *vir* genes. Upon activation, the VirD1/VirD2 endonuclease cleaves the RB and LB sequences. VirD2 attaches to the 5' end of the cleaved T-DNA which is displaced as a single stranded DNA (T-strand) from the pTi, forming the T-

DNA-VirD2 complex. The complex and several other bacterial proteins such as VirD5, VirE2, VirE3 and VirF are translocated through the type IV secretion system consisting of VirB and VirD4 proteins into the plant cell. VirE2 proteins coat the T-strand of the complex forming the T-complex in the plant cell to protect it from the host nucleases. Finally, the T-complex is guided into the nucleus, where the T-DNA is integrated at random locations in the plant genome. Expression of the aforementioned genes involved in the plant hormone biosynthesis results in tumorigenesis (Gelvin, 2003, 2010b; Meyers et al., 2010; Ziemienowicz, 2001), while production of opines provides nutrition for the bacterium. For plant biotechnology purposes, T-DNA genes are replaced with genes of interest (transgenes). Vir proteins have been examined extensively; meanwhile the efficient *Agrobacterium* systems namely binary vector systems were developed to introduce transgenes into the plant genome. The *Agrobacterium*-mediated plant transformation methods have faced many challenges, and many of them were overcome by the tremendous efforts of researchers around the world.

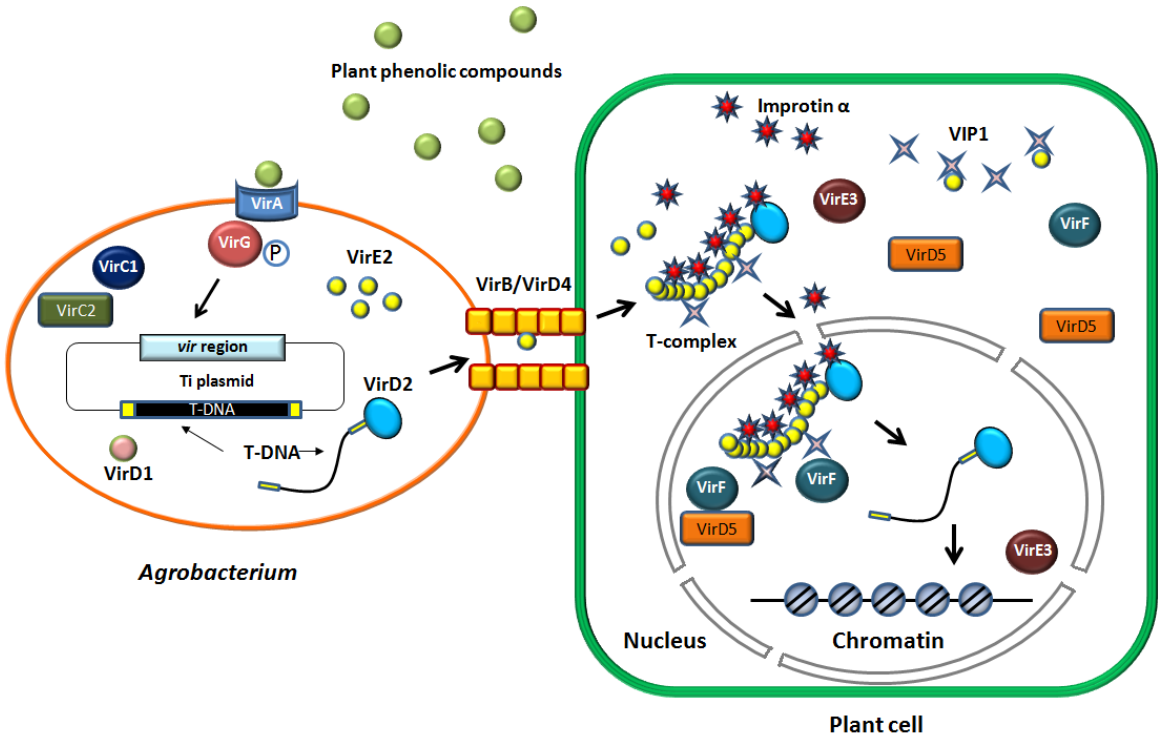


Figure 1.1.1. Schematic representation of *Agrobacterium*-mediated plant transformation.
Selected proteins are shown. P: phosphorylation. Arrows: movement of the complex.

1.1.2. Vir proteins

1.1.2.1. Sensory units: VirA and VirG

The membrane bound protein VirA and the cytoplasmic response regulator VirG function in the signal transduction system. VirA is composed of 7 domains: a short 18 amino acid N-terminal cytoplasmic domain, two membrane-spanning domains, a periplasmic sensor domain, the two linkers of kinase and receiver domains in the cytoplasmic C-terminal domain (Nair et al., 2011). Plant wounds release chemical stimuli including acidic pH, monosaccharides and certain phenolic compounds which are sensed by VirA. The VirA linker domain binds to phenolic compounds, and monosaccharides and pH signaling are recognized by the periplasmic domain. Upon binding of phenolic compounds to the linker domain, VirA autophosphorylates the phosphoacceptor histidine residue in the kinase domain. Subsequently, the phosphate from VirA is transferred to the aspartate residue of VirG. Phosphorylated VirG is suggested to be dimerized to activate transcription of *vir* genes by binding to the *vir* box which locates in all *vir* gene promoter regions (Gao et al., 2006; Lohrke et al., 1999; Nair et al., 2011). In the *Agrobacterium*-mediated

plant transformation, a phenolic compound acetosyringone can be added in the bacterial culture to artificially activate the VirA/VirG sensory unit resulting in the activation of the *vir* operons (Soltani et al., 2008).

1.1.2.2. Type IV secretion system: VirB1-11 and VirD4

After activation of the *vir* genes, the T-DNA is processed and transferred through the type IV secretion system (T4SS) from the bacterium to the plant cell. In addition, several Vir proteins are also transferred along with the T-DNA into the plant cell. T4SS consists of 12 Vir proteins notably VirB proteins and VirD4. VirB4, VirB11 and VirD4 are ATPases which supply the energy for T4SS biogenesis and its function (Ripoll-Rozada et al., 2013). VirB11 belongs to the traffic ATPase superfamily which is proposed to act as a switch between pilus biogenesis and DNA transport. VirB4 associates with the core complex at the inner membrane. VirB4 is assisted by VirB11 to dislocate VirB2 pilin. VirD4 is suggested to displace VirB4 from the base of the channel and to interact with Vir11 for substrate transport (Ripoll-Rozada et al., 2013).

Inner-membrane translocase components are VirB3, VirB6 and VirB8 (Atmakuri and Christie, 2008). VirB3 is an inner membrane protein containing one periplasmic domain and the C-terminal residues in the cytoplasm. It stabilizes VirB4, VirB7 and VirB8 (Mossey et al., 2010). VirB6 contains four or five transmembrane segments and is suggested to stabilize VirB3 and VirB5, and assist in formation of VirB7 and VirB9 complexes (Hapfelmeier et al., 2000; Jakubowski et al., 2004; Jakubowski et al., 2003; Judd et al., 2005a; Judd et al., 2005b). VirB8 is an inner membrane protein composed of a cytoplasmic N-terminal domain, a transmembrane helix and a C-terminal domain in the periplasm (Das and Xie, 1998). VirB8 stabilizes VirB1, VirB3, VirB4, VirB5, VirB6, VirB7 and VirB11. The interface of VirB8 homodimers is required for VirB2 to associate to the T-pilus assembly site (Sivanesan and Baron, 2011).

Outer membrane components called the core complex are VirB7, VirB9 and VirB10 (Ilangovan et al., 2015). VirB7 is the monotopic lipoprotein, which is exposed at the periplasmic surface (Fernandez et al., 1996; Jakubowski et al., 2003). VirB9 consists of several β strands, and its non-covalent interactions and the formation of the disulfide bridges between VirB7 are necessary to generate an active VirB7-VirB9 heterodimers (Bayliss et al., 2007). VirB10 was thought to be an inner membrane protein; however, it was

identified to have both the inner membrane and outer membrane parts (Chandran et al., 2009; Jakubowski et al., 2009). VirB10 contains several domains including the proline rich region which extends across the periplasm and the β-barrel domain which interacts with the VirB7-VirB9 channel complex (Jakubowski et al., 2009).

VirB1 localizes in the periplasm and belongs to a lytic transglycosylase family (Atmakuri and Christie, 2008). It is proposed to be required for the T4SS assembly. The N-terminal lytic transglycosylase domain would modify the peptidoglycan for the core complex assembly. VirB1*, the C-terminal third of the VirB1, is cleaved and secreted to the outside of the bacterial cell. VirB1* might promote the assembly of the T-pilus subunits via protein-protein interactions (Zupan et al., 2007). VirB5 localizes in the periplasm, and it is also a component of the pilus (Atmakuri and Christie, 2008). VirB5 is a single domain protein consisting of mostly α-helices and localizes in the periplasm and to the tip of the T-pilus, which might be important for host-cell recognition (Aly and Baron, 2007; Ilangovan et al., 2015; Yeo et al., 2003).

VirB2 is a major component of the cyclized T-pilus. A 12.3 kDa of an initial VirB2 protein is processed into a 7.2 kDa protein. The processed VirB2 protein consists of the hydrophilic region, which is likely responsible for lining the pore, and hydrophobic region, that is possibly required for the pilin-pilin interaction within the pilus assembly (Kerr and Christie, 2010; Lai and Kado, 1998).

1.1.2.3. Overdrive binding proteins: VirC1 and VirC2

VirC1 belongs to the ParA/MinD ATPase superfamily containing a deviant Walker A nucleotide triphosphate-binding motif; it binds to the overdrive sequence near the right border (Koonin, 1993; Toro et al., 1989). VirC1 is proposed to recruit the relaxosome composed of VirD1 and VirD2 at the T-DNA borders and to activate the VirD1/VirD2 nuclease activity. VirC1 then delivers the T-DNA-VirD2 complex to the VirD4 receptor (Atmakuri et al., 2007). VirC2 is a double stranded DNA specific DNA binding protein and interacts with VirC1 to stimulate assembly of the VirD1 and VirD2 relaxosome. VirC2 also binds through its ribbon-helix-helix domain to the double stranded DNA generated from folding of the single stranded T-DNA. Binding of VirC2 to the folded double stranded T-DNA results in protection of the T-DNA from the nucleolytic degradation and/or other functions for T-DNA transfer (Atmakuri et al., 2007; Lu et al., 2009).

1.1.2.4. Relaxosome: VirD1 and VirD2

VirD2 is one of the most well characterized Vir proteins essential for *Agrobacterium*-mediated plant transformation. VirD2 contains several domains and signal sequences, including a relaxase domain, nuclear localization signals (NLSs) and ω domain (Gelvin, 2012). VirD2 alone can cleave the single stranded DNA containing the T-DNA border sequences, but VirD2 requires VirD1, another bacterial protein, to cleave the double stranded T-DNA border sequences (Scheiffele et al., 1995; Steck et al., 1990). VirD2 attaches to the 5' end of the single stranded T-DNA via Y29 to form the T-DNA-VirD2 complex (Vogel and Das, 1992). NLSs of VirD2 are a monopartite type and a bipartite type at the N and C-terminus of the protein, respectively (Herrera-Estrella et al., 1990; Howard et al., 1992; Tinland et al., 1992). Previous studies that used reporter proteins fused to the C-terminal NLS demonstrated that this NLS was a strong nuclear localization signal (Howard et al., 1992; Tinland et al., 1992). Ziemienowicz et al. (2001) reported that VirD2 can import short fluorescently labelled single stranded DNA into the nuclei of tobacco protoplasts, while VirD2 lacking the C-terminal NLS was unable to direct single stranded DNA into the nucleus. Thus, the C-terminal NLS participates primarily in nuclear targeting of the T-complex, while the N-terminal NLS does not function due to its proximity to Y29 (Rossi et al., 1993; Ziemienowicz et al., 2001). VirD2 is imported into the nucleus by the importin α/β pathway through its NLSs. Importin α is an adaptor protein to recognize an NLS of a protein, and then it interacts with importin β. The protein cargo is targeted to the nuclear pore complex and translocated the cargo into the nucleus by importin β (Lange et al., 2007). Seven *Arabidopsis* importin α proteins (IMPa-1, IMPa-2, IMPa-3, IMPa4, IMPa-6, IMPa7 and IMPa-9) have been shown to interact with VirD2 in the yeast two-hybrid, plant biomolecular fluorescence complementation (BiFC) and *in vitro* protein-protein interaction assays (Bhattacharjee et al., 2008). The *impa-4* mutant *Arabidopsis* plants showed the rat (resistant to *Agrobacterium* transformation) phenotype, which was rescued by overexpression of other importin α isoforms. This indicated the importance of the IMPa-4 during the transformation (Bhattacharjee et al., 2008). The C-terminal end including the C-terminal NLS of VirD2 showed a strong interaction with IMPa-4. The C-terminal NLS and its adjacent amino acids of VirD2 were suggested to be important for nuclear targeting of the T-strand (Bhattacharjee et al., 2008).

The ω domain was suggested to play parts in T-DNA translocation from the bacterium to the plant cell and T-DNA integration into the plant genome; however, its function is the object of debate due to conflicting results (Bravo-Angel et al., 1998; Mysore et al., 1998; van Kregten et al., 2009).

1.1.2.5. Single stranded DNA binding protein: VirE2 and VirE1

VirE2 is another extensively studied Vir protein. This protein contains a single stranded DNA (ssDNA) binding domain and a bipartite NLS. VirE2 proteins are translocated into the plant cell and bind to the T-strand forming a solenoidal (telephone cord-like coiled) structure, which protects the T-strand from the host nucleases (Citovsky et al., 1997; Citovsky et al., 1989; Sundberg and Ream, 1999). Hence, VirE2 is suggested to maintain T-strand integrity and also shape the T-complex for translocation into the nucleus (Gelvin, 2012; Ziemienowicz et al., 2001). VirE2 is known to interact with plant proteins, including VirE2 interacting protein1 (VIP1) that also interacts with the host importin α proteins (Gelvin, 2010b). The role of VirE2 NLS is controversial due to conflicting results using the reporter protein fused to either the N or C-terminus of the protein (Gelvin, 2010b). VirE2 along with VirD2 were shown to interact with the abovementioned seven *Arabidopsis* importin α isoforms, and IMPa-4 was found to be more efficient than other isoforms in delivering VirE2 into the plant cell nucleus. Fusion proteins of VirE2 attached to the N or C-terminus of the yellow fluorescent protein (YFP) were constructed to observe their subcellular localization in the wild type and the *impa-4* mutant *Arabidopsis* and in wild type tobacco cells. The fusion proteins were localized in the cytoplasm as aggregates. Yet, the BiFC indicated that overexpression of IMPa-4 resulted in nuclear localization of VirE2 (Bhattacharjee et al., 2008). These results indicate importance of IMPa-4 in *Agrobacterium*-mediated plant transformation due to its interaction with both VirD2 and VirE2 (Bhattacharjee et al., 2008).

VirE1 is a small protein (7 kDa) without known enzymatic activity, although it interacts with VirE2. VirE2 oligomerizes to form filamentous aggregates *in vitro* in the absence of VirE1. In the bacterium, VirE1 interacts with VirE2 by electrostatic interaction to prevent oligomerization. VirE1 is likely removed from the VirE2 by mechanical action in bending the domains at the secretion channel, while VirE2 is translocated to the plant cell (Deng et al., 1999; Dym et al., 2008).

1.1.2.6. Adaptor and putative transcription activator

VirE3 is also translocated into the plant cell along with other Vir proteins and the T-DNA-VirD2 complex from *Agrobacterium*, and it binds to both VirE2 and importin α proteins acting as an adaptor molecule like the plant protein VIP1 (Lacroix et al., 2005). VirE3 also contains the nuclear localization signals, monopartite and bipartite types. In addition, interactions between VirE3 and pCsn5, a component of COP9 signalosome, VirE3 and pBrp, a plant specific general transcription factor IIB-related protein, were observed. VirE3 could induce the transcription of DNA through the GAL4-binding domain in yeast. This suggested that VirE3 may induce transcription of a transgene by interacting with the transcription factor pBrp (Garcia-Rodriguez et al., 2006).

1.1.2.7. F-box effector: VirF

VirF is a unique protein specific to the octopine type Ti plasmids, while it is absent in the nopaline types. This protein was first thought to be a host-range factor due to different transformation efficiency of *Nicotiana glauca* among octopine and nopaline strains of *Agrobacterium* (Melchers et al., 1990). Later, it was shown that VirF contains the F box motif which interacts with SCF (Skp1-Cdc53-F box protein) complex for ubiquitination and then proteolysis, and VirF is involved in proteolysis of VirE2 and VIP1 in the nucleus (Schrammeijer et al., 2001; Tzfira et al., 2004). Interestingly, a host plant protein F-box protein, VBF (VIP1-binding F-box), acts similarly to VirF. This VBF expression is induced by *Agrobacterium* (Magori and Citovsky, 2011c; Zaltsman et al., 2010).

1.1.2.8. Translocated effector: VirD5

VirD5 is another protein that is exported into the plant cell from the bacterium (Vergunst et al., 2005). It localizes in the nucleus due to four putative NLSs, and its function was found to be to stabilize VirF which is rapidly degraded by the host proteolytic enzymes (Magori and Citovsky, 2011a). VirD5 can form homodimers in the plant nucleus and bind to a specific DNA sequence, named VirD5 response element. In addition, VirD5 was proposed to be a transcriptional activator-like Vir protein because VirD5 was shown to act as a transcription factor in yeast. VirD5 can also interact with VIP1 in *Arabidopsis*, which

prevents binding of VBF to VIP1. Thus, the coat proteins might be protected by VirD5 from rapid degradation by the host proteins (Wang et al., 2014).

1.1.3. Development of an efficient transformation method: Binary vector systems

Because of the amazing ability of *Agrobacterium* to transform the plant genome, *Agrobacterium* was "disarmed" by removing the genes affecting plant hormone synthesis, and transgenes were initially cloned into the T-DNA region of the Ti plasmid. However, the large size and low copy number of Ti plasmids made it difficult to manipulate and isolate plasmids. The binary vector system has been developed to perform *Agrobacterium*-mediated plant transformation more efficiently. Two plasmids are essential in the system:

1) the helper plasmid originated from the *Agrobacterium* Ti plasmid containing the *vir* genes without a functional T-DNA region.

2) the binary vector composed of the gene of interest flanked by right and left T-DNA border sequences and the origin(s) of replication that can function in both *Escherichia coli* and *Agrobacterium* (Lee and Gelvin, 2008; Meyers et al., 2010). T-DNA also typically includes commonly used plant selectable marker genes, such as those encoding aminoglycoside antibiotic resistance such as kanamycin or hygromycin and those encoding herbicide resistance such as phosphinothricin/gluphosinate (Lee and Gelvin, 2008). Expression of both the gene of interest and plant selection marker gene is regulated by plant-specific transcription control elements; e.g., 35S promoter (Chen et al., 2006).

Several antibiotics are used for the bacterial selection in the binary system. Many *Agrobacterium* strains carry a mutation in their chromosomal DNA conferring resistance to rifampicin (Lee and Gelvin, 2008). These strains frequently harbor helper plasmids with genes for resistance to kanamycin, carbenicillin, erythromycin or gentamicin (Lee and Gelvin, 2008). Binary vectors with different selection markers from the abovementioned markers should be used.

Timentin, a mixture of ticarcillin and clavulanic acid, is used to suppress the growth of *Agrobacterium* in the plant culture (Ieamkhang and Chatchawankanphanich, 2005). Ticarcillin is a penicillin derivative which inhibits bacterial cell wall synthesis, and it is inactivated by β-lactamase produced by a wide variety of bacteria (Miller et al., 2001). Tricarcillin together with a β-lactamase

inhibitor clavulanic acid can enhance the activity of tricarcillin (Ieamkhang and Chatchawankanphanich, 2005).

1.1.4. Challenges and improvements of *Agrobacterium*-mediated plant transformation

The development of the binary vector systems provided many possibilities to transform plants more efficiently. However, due to the nature of *Agrobacterium*, this transformation method suffers from some limitations including the host range, transformation condition/efficiency and random integration of the T-DNA into the plant genome. Numerous different approaches have been taken to solve these problems.

1.1.4.1. Expansion of the host range

Agrobacterium can infect various types of dicotyledons, whereas monocotyledons are much less susceptible to *Agrobacterium* infection (Decleene, 1985). There are several reasons for difficulties with the monocot transformation via *Agrobacterium*. For example, composition of the cell wall differs in the dicots and monocots: β-linked glucose residues with interlocking chains of β-D-xyloglucans in the dicot cell wall, and the glucuronoarabinoxylans and linear chains of β-D-xylose with interlocking polysaccharides in monocots. Due to this monocot cell wall composition, the bacterial attachment to the monocot cell is inefficient (Sood et al., 2011). Monocots produce very few *vir* inducers such as a hydrophilic high molecular mass compound and a low molecular mass ethyl ferulate; moreover, they also release inhibitors of *Agrobacterium* virulence, including 2,4-dihydroxy-7-methoxy-2H-1,4-benzoxazin-3(4H)-one and 2-hydroxy-4,7-dimethoxybenzoxazin (Lacroix and Citovsky, 2013; Sood et al., 2011). Revealing the negative factors associated with transformation of monocots led to optimization of the transformation conditions. Successful rice transformation was reported in 1994 (Hiei et al., 1994). In this study, different explants including shoot apex, root segment, scutellum, immature embryo, root callus, scutellum callus and suspension cells were used for transformation with addition of acetosyringone. Scutellum callus showed a high transformation rate between 12.8% and 28.6% among three rice cultivars (Hiei et al., 1994). Maize was transformed using calli from immature embryos, and between 5 and 30% transformation efficiency was observed with the use of acetosyringone (Ishida et al., 1996). An updated maize transformation protocol has been established to achieve approximately 50% of the transformation efficiency using immature embryos

(Ishida et al., 2007). Also, freshly isolated immature embryos, precultured immature embryos and embryogenic calli of wheat were used for transformation using acetosyringone resulting in a transformation efficiency between 0.3% and 4.3% among three types of explants (Cheng et al., 1997). Barley can be transformed using immature embryos to have an average of 25% transformation efficiency (Bartlett et al., 2008). Interestingly, acetosyringone was found to be unnecessary to barley transformation (Bartlett et al., 2008). The host range of *Agrobacterium* has expanded to generate transgenic rice, maize and wheat among other monocots (James, 2014).

1.1.4.2. Improvement of transformation efficiency

Some plant species are highly susceptible to *Agrobacterium*-mediated transformation (An, 1985; Wroblewski et al., 2005). After the transformation, plant regeneration is a necessary step to obtain stably transformed plants which may be difficult and not optimized for some plant species. Addition of compounds such as 6-benzyladenine and 5-azacytidine and 2,4-dichlorophenoxyacetic acid in the growth media have been used to improve the regeneration of *Bambusa ventricosa* and *Acca sellowiana*, respectively (Fraga et al., 2012; Wei et al., 2015). More robust approach is the use of ammonium nitrate in the plant cultivation medium, which resulted in an increase of regeneration rate of explants and stable transformation rate of tobacco, *Triticum aestivum* and *Triticosecale* (Boyko et al., 2009; Greer et al., 2009). This approach was further modified and combined with potassium chloride or rare earth elements, namely, cerium and lanthanum (Boyko et al., 2011; Maheshwari and Kovalchuk, 2013). Different concentrations of each element and different combinations of elements were shown to improve transformation efficiency, and a more intact right border of the T-DNA was detected in the transgenic plants, suggesting improved quality of transgenic integration (Boyko et al., 2011; Maheshwari and Kovalchuk, 2013). The exact function of ammonium nitrate, potassium chloride and rare earth elements in the transformation process is unknown. Since they have been shown to affect the recombination frequency, it was suggested that these elements may have an effect on the host recombination machinery, which possibly promotes transformation efficiency (Maheshwari and Kovalchuk, 2013).

Manipulating host proteins involved in post-transcriptional gene silencing and transcriptional gene-silencing is a new approach to improve transformation. Bilichak et al. (2014) demonstrated enhanced

efficiency of *Agrobacterium*-mediated *Arabidopsis* transformation by transient down-regulation of the RNA silencing machinery namely AGO2 and NRPD1a (a subunit of RNA polymerase IV). Either *AGO2* or *NRPD1a* genes in the reproductive organs of *Arabidopsis* was transiently down-regulated using the *Tobacco rattle virus* (TRV)-based virus-induced gene silencing (VIGS) technique. Although TRV can enter the ovules of tobacco plants, it is not transmitted through seeds. Due to this unique feature of the virus, VIGS technique was used to transiently downregulate the *AGO2* or *NRPD1a* genes in *Arabidopsis*. After confirmation of down-regulation of the two genes by real-time PCR, those plants were used for *Agrobacterium*-mediated transformation via the floral dip method. Results showed an increase in the transformation efficiency by successfully down-regulating the two genes transiently (Bilichak et al., 2014). Downregulation of *AGO2* or *NRPD1a* genes decreased global DNA methylation in the plant genome, which led to an increase in DNA strand breaks (Bilichak et al., 2014). It is possible that higher DNA strand break levels resulted in efficient T-DNA integration by non-homologous end joining, by which T-DNA is most likely integrated into the plant genome (Magori and Citovsky, 2011b).

Alternative approaches were also used to improve plant transformation. Zhang et al. (2013a) reported that addition of a plant defense response inhibitor in the culture medium and cold treatment of calli increased transformation significantly. Pretreatment of calli of perennial ryegrass with a cold shock (20 min on ice) and addition of L-glutamine (L-Gln) prior to and during transformation, respectively, resulted in highest transformation efficiency (84%). L-Gln is known to inhibit beta-aminobutyric acids (BABA) which are produced by plants upon a bacterial infection, yet the molecular mechanism of BABA inhibition by L-Gln is still not understood (Zhang et al., 2013a). Attenuating plant defense responses was indeed effective to increase transformation efficiency. Additionally, exposure to abiotic stresses such as cold, aluminum salt and paraquat were shown to affect methylation status in the coding region of a glycerophosphodiesterase-like protein (NtGPDL) of the tobacco genome (Choi and Sano, 2007). This change in methylation status may cause genome instability as seen before (Bilichak et al., 2014), which could enhance the frequency of T-DNA integrations.

1.1.4.3. T-DNA integration at the precise genomic location

Precise integration of transgenes was a long time goal for the nuclear genome transformation via *Agrobacterium*-mediated plant transformation as well as other transformation methods. The T-DNA was thought to be usually integrated into the gene-rich areas of the genome in a non-random manner; however, this integration pattern is likely observed due to the use of a selection process of the transformants using exposure to herbicide or antibiotics, favoring regeneration of plants with high level of expression of transgenes (and antibiotic resistance markers) (Gelvin and Kim, 2007). Work by Gelvin and Kim (2007) demonstrated that T-DNA is integrated into the genome in a relatively random manner. Attempts to increase frequency of homologous recombination between T-DNA and the targeted site by introducing homologous sequences in T-DNA region were made, but without major success; such events were extremely rare (Gelvin, 2003). Targeting the T-DNA (or individual transgenes) into precise genome locations can now be achieved by zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeats/CRISPR-associated (CRISPER/Cas) system. ZFNs are fusions of newly engineered zinc-finger arrays of a DNA binding domain to a non-specific DNA-cleavage domain of the *Fok*I type II restriction endonuclease. The arrays consist of 3-6 or more C_2H_2 fingers, in which each finger consists of 30 amino acids folded into a ββα structure, and these DNA binding domains target specific DNA sequences 9-18 base pairs long (Liu et al., 2013; Porteus and Carroll, 2005). The ZFN system was successfully used to insert an herbicide resistance gene cassette flanked by short homologous sequences of the endogenous endochitinase gene into the tobacco genome (Cai et al., 2009).

TALENs also contain the *Fok*I nuclease domain but with transcription activator-like effectors (TALE) as DNA binding domains. TALEs are built in tandem arrays of 33-35 amino acid repeats, and each array recognizes a single base pair in the major groove. TALEs can be designed to bind to specific DNA sequences (Kim and Kim, 2014). TALEN was used to target the *Sur* genes encoding an enzyme involved in the amino acid biosynthesis in tobacco protoplasts transformed by the polyethylene glycol (PEG)-mediated method. The YFP gene was successfully inserted to the target site. Authors suggested that their data would be a framework for *Agrobacterium* or on other transformation methods to engineer plant genomes using TALENs (Zhang et al., 2013b).

The CRISPR/Cas system is more user friendly than ZFNs and TALENs which require complicated design and laborious assembly of the specific DNA binding domains for each target gene (Belhaj et al., 2013). In the CRISPR/Cas system, Cas9 nuclease is guided to the target genomic DNA adjacent to the protospacer adjacent motif sequence by two customizable small non-coding RNAs, CRISPR RNA (crRNA) and trans-activation crRNA or one synthetic RNA chimera; subsequently, Cas9 cleaves the target genomic DNA, which is repaired by both non-homologous end joining and homology-directed repair mechanisms leading to mutations inactivating the target gene (Belhaj et al., 2013). *Agrobacterium* is used as an efficient tool to introduce the CRISPR/Cas system in *Arabidopsis* and tobacco, and it can be used for rice and wheat (Belhaj et al., 2013). *Agrobacterium*-mediated plant transformation is also a useful tool in delivering ZFNs and TALENs because of the established transformation methods.

1.2. Chloroplast/plastid transformation

Although *Agrobacterium*-mediated plant transformation has been successful to introduce transgenes in the nuclear genomes of various plants, it still faces some issues. Since Agrobacteria integrates primarily single copies of transgenes at a single location in the nuclear genome, the level of a transgene expression is relatively low in the stably transformed plants (Meyers et al., 2010). Also, some recombinant proteins are toxic in the cytoplasm. For example, a cholera toxin B subunit (CTB) a candidate oral subunit vaccine for cholera, a trehalose, which provides abiotic stress tolerance, and an enzyme xylanase, which can be used in applications such as paper and fiber manufacture are harmful in the cytoplasm of transgenic plants (Daniell et al., 2005). In addition, due to the ability of plants to cross-pollinate, transgenes carried by pollen can escape and contaminate wild type plants (Daniell, 2002; Meyers et al., 2010). Chloroplast/plastid transformation provides solutions to these issues. Chloroplast transformation is a type of plastid transformation. Transformation of all types of plastids including chloroplast and other plastids such as proplastids, chromoplasts and amyloplasts are collectively referred to as plastid transformation.

1.2.1. Benefit of chloroplast/plastid transformation

Plastids exist in plant cells in various forms including the undifferentiated proplastids in meristematic tissues, the green chloroplasts in photosynthetic cells, the colored chromoplasts in flowers and

fruits and the starch storage amyloplasts in tubers and endosperm in seeds (Bock, 2007; Pyke and Pyke, 2009). All plastid types contain identical copies of the genome, yet their gene expression is influenced by their environment (Bock, 2007; Pyke and Pyke, 2009). Ten to 100 plastid genome copies exist in a plastid, and 10-100 plastids are in a plant cell and the number varies due to cell age and the tissue type (Daniell et al., 2002). This is an attractive aspect of plastid transformation because multiple copies of transgenes can be integrated, which would result in higher expression of the transgene. High expression of transgenes stemming from transformed chloroplasts/plastids (such plants are called transplastomics) is common; recombinant proteins accumulate to 5-25% of the total soluble protein (TSP) (Meyers et al., 2010). Even more than 70% TSP has been seen (Oey et al., 2009).

Plastids of most angiosperms are inherited maternally, since plastid exclusion occurs during the first pollen mitosis, plastid degradation in the generative cell and plastid exclusion during fertilization occur (Bock, 2007). Because of this, cross-pollination between transplastomic and transgenic plants can be prevented. Also, chloroplast compartmentalization provides protection to the cell. Transplastomic plants expressing CTB, a trehalose or a xylanase showed a normal phenotype (Daniell et al., 2005).

Multiple transgenes can be arranged in an operon-like structure which can be efficiently expressed in chloroplasts (Meyers et al., 2010; Zhou et al., 2007). The transgenes are flanked by the chloroplast homologous sequences which promote integration of the transgenes via homologous recombination. This prevents position effects (Meyers et al., 2010). Plastid transformation in higher plants has been significantly improved since the first successful transformation event (Svab et al., 1990), including optimized design of chloroplast transformation vectors.

1.2.2. Chloroplast/plastid transformation vectors

Chloroplast transformation vectors have been designed and efficiently used to transform several plant species. The common characteristic of these vectors is the presence of the chloroplast expression cassette which is flanked by chloroplast homologous sequences which are necessary to promote homologous recombination between the vector and the chloroplast genome (Lutz et al., 2007; Verma and Daniell, 2007; Verma et al., 2008). The order of the elements is shown in Figure 1.2.

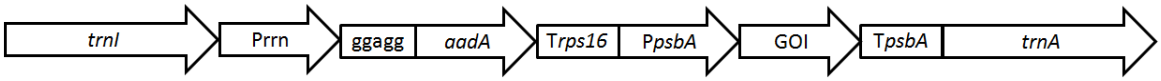


Figure 1.2. Chloroplast expression cassette for leaves (modified from Verma et al., 2008).
trnI and *trnA*: chloroplast homologous sequences; Prrn: promoter; P*psbA*; promoter and 5' regulatory element; ggagg : 5' regulatory element; T*rps16* and T*psbA*: 3' regulatory elements; *aadA* (conferring spectinomycin resistance): selectable marker; GOI: gene of interest.

The chloroplast homologous sequences *trnI* and *trnA* are commonly selected due to their location in inverted repeat (IR) regions in the genome. When the transgene is integrated into one of the IR region, the transgene is duplicated into the other IR region by a mechanism called copy correction (Verma and Daniell, 2007).

Several selectable markers have been used; commonly used markers in higher plants are antibiotic resistance genes including *aadA* (spectinomycin and streptomycin resistance), *nptII* and *aphA-6* (kanamycin resistance), and herbicide resistance genes including *bar* (phosphinothricin resistance), *EPSP* (glyphosate resistance), and *HPPD* (diketonitrile resistance) (Day and Goldschmidt-Clermont, 2011). These selectable markers can be removed efficiently by a recombination pathway mediated by DNA direct repeats and a site specific recombinase/integrase. Thus, a selectable marker needs to be flanked by these necessary sequences. The Cre site-specific recombinase recognizes 34 bp *loxP* sites, and the phi C31integrase recognizes non-identical 34 bp *attB* and 39-bp *attP*, resulting in excision of the region in between *loxP* sites and *attB-aatP* sites (Day and Goldschmidt-Clermont, 2011; Lutz et al., 2007).

1.2.3. Chloroplast/plastid transformation methods

Two methods are used to transform plastids; i.e., biolistic or PEG-mediated methods. The biolistic system requires tungsten or gold particles with 0.6 µm mean diameter and a biolistic device (Verma et al., 2008). The DNA-coated metal particles are accelerated toward the target by high-pressure helium to penetrate the cells, and a portion of the DNA may be stably integrated into the host genome (Kikkert et al., 2005).

In the PEG-mediated method, protoplasts are incubated in a solution containing various ions, PEG and the foreign DNA. The plasma membrane is affected by the solution allowing transgenes to penetrate and move into the cytoplasm (Kofer et al., 1998). The exact role of PEG in the transgene delivery through the chloroplast envelope is unclear (Kofer et al., 1998).

There are drawbacks associated with these methods. For example, the biolistic technique requires a costly device and is pricy because it involves gold particles. The regeneration protocols through somatic embryogenesis are required for chloroplast transformation, and the established protocols are available only for selected plant species (Daniell et al., 2005; Liu et al., 2013). The PEG-mediated method is typically used to transform protoplasts, preparation, maintenance, treatment and regeneration of which require complex protocols (Meyers et al., 2010). Regardless of these limitations, the biolistic technique is the most common method used to transform chloroplasts.

1.2.4. Growing research interest in chloroplast/plastid transformation

Chloroplast transformation has been getting attention for its possibility to be the most efficient method to produce various pharmaceutical proteins. Chloroplast-derived pharmaceutical proteins (bioencapsulated proteins) offer a significant reduction in costs associated with synthesis and purification of proteins (Davoodi-Semiromi et al., 2010). There are several successful cases of production of pharmaceutical proteins via chloroplast transformation. The CTB of *Vibrio cholerae* was fused to malarial vaccine antigens; apical membrane antigen-1 and merozoite surface protein-1 were expressed in lettuce and tobacco chloroplasts, which were used to immunize mice by subcutaneous injection using purified antigens or by oral administration with tobacco leaves (Davoodi-Semiromi et al., 2010). Both types of immunized mice elicited strong immunity against the malarial parasite and cholera toxin (Davoodi-Semiromi et al., 2010). Another interesting example is the chloroplast-derived human immunodeficiency virus (HIV)-1 capsid protein p24 and p24 fused to Nef (the negative regulatory protein). A high level of p24-Nef (about 40% TSP) was observed in transplastomic tobacco. Strong antigen-specific serum IgG response was observed after mice were immunized by subcutaneous injection with purified p24 and oral administration of partially purified p24-Nef after injection of p24 or Nef (Gonzalez-Rabade et al., 2011).

Bioencapsulated proteins produced in the transplastomic tobacco were also used to decrease ocular inflammation by angiotensin-converting enzyme 2 (ACE2) and Ang-(1-7), a cleaved product of Angiotensin II. The mice fed with chloroplast-derived ACE2/Ang-(1-7) showed significant reduction in pulmonary hypertension, a lung disease that causes a heart failure and death (Shil et al., 2014). Chloroplast-derived protein, myelin basic protein (MBP) fused with the transmucosal carrier CTB was also evaluated to

cross blood-brain and blood-retinal barriers for possible treatment of Alzheimer's disease (Kohli et al., 2014). Amyloid β42, can be cleaved to different fragments which then form aggregates and plaques – a major characteristic of Alzheimer's disease. The synaptic structural integrity and communication between axon and dendrites are lost by these newly formed aggregates and plaques. The chloroplast-derived MBT-CTB was orally administrated to the Alzheimer`s disease mice. The results showed that the MBT-CTB successfully crossed the blood-brain barrier and blood-retinal barrier, resulting in a decrease in formation of amyloid β42 in mice (Kohli et al., 2014). In addition, transplastomic plants can be administered directly for treating certain conditions. As a result, there is a growing interest in chloroplast-derived proteins and transplastomic plants to be used as pharmaceuticals.

1.2.5. Challenges of chloroplast/plastid transformation

Despite all the success in transforming chloroplasts of various plants for production of recombinant proteins, only rice proplastids were successfully transformed using rice calli among all the monocot plants (Lee et al., 2006). Proplastids are approximately 5 fold smaller than well-developed chloroplasts, which makes them a difficult target for biolistics (Lee et al., 2006). The smaller particle size was suggested to enhance the transformation of proplastids (Lee et al., 2006). Gold particles with an average between 0.07 and 0.3 μm were used to transform chloroplasts using tobacco leaf (Okuzaki et al., 2013). The highest chloroplast efficiency was observed with 0.3 μm particles (Okuzaki et al., 2013). Their findings have not been applied for plastid transformation in monocots. Biolistics is a robust transformation system to introduce transgenes in to plants and animals. However, the biolistics method is still not ideal for chloroplasts transformation, especially in monocots.

1.3. Intracellular protein trafficking

1.3.1. Nuclear import

Smaller proteins (<60 kDa) can diffuse through the nuclear pore complex (NPC), while bigger proteins require NLSs which are recognized and imported by the importin α/β pathway (Wang and Brattain, 2007). The NLSs typically contain stretches of basic amino acid residues (lysine and argentine) (Wang and Brattain, 2007). Importin α carries two NLS binding grooves; the major and minor binding sites are located

at N-terminal Armadillo (Arm) repeat 2-4 and the C-terminal Arm repeat 7 - 8, respectively (Kosugi et al., 2009a). Each repeat is composed of three α helices (Wang and Brattain, 2007). The six classes of NLSs have been identified (Kosugi et al., 2009a).

1) Class 1: A classical monopartite NLS containing a single cluster of basic amino acid residues which binds to the major binding groove of importin α. Consensus sequence: KR(K/R)R, K(K/R)RK.
2) Class 2: A classical monopartite NLS containing a different basic core than Class 1 NLS. The major binding groove of importin α binds to the class 2 NLSs. Consensus sequence: (P/R)*XX*KR(ˆDE)(K/R) (*X*: any amino acids; ˆDE: any amino acids but aspartate or glutamate.
3) Class 3: A non-canonical monopartite NLS which binds to the minor binding groove of importin α. Consensus sequence: KR*X*(W/F/Y)*XX*AF.
4) Class 4: A non-canonical monopartite NLS which binds to the minor binding groove of importin α. Consensus sequence : (R/P)*XX*KR(K/R)(ˆDE).
5) Class 5: A plant-specific NLS. Consensus sequence: LGKR(K/R)(W/F/Y).
6) Bipartite: Two imperfect monopartite NLSs. The class 3/class 4 NLS at the N-terminal region and the class 1/class 2 NLS at the C-terminal region. Consensus sequence: KRX_{10-12}K(KR)(KR) and KRX_{10-12}K(KR)*X*(K/R) ($X_{10\text{-}12}$: any 10-12 amino acids).

The protein nuclear import pathway is illustrated in Figure 1.3.1. The adaptor protein importin α directly binds the NLS in the protein in the cytoplasm. Importin β then binds to the importin α forming the trimeric complex to facilitate the interaction and translocate the complex through the NPC. Inside the nucleus, RanGTP binds to improtin β causing a conformational change to release the importin α-cargo complex. Association of an autoinhibitory region on the importin β-binding domain of importin α with the nucleoporin Nup2 (in yeast) causes dissociation of the cargo from importin α. A RanGTP-importin β complex and RanGTP-Nup2-importin α with the export receptor for importin α (Ces1 in yeast) are exported back to the cytoplasm. RanGAP at the cytoplasmic face of the NPC hydrolyzes RanGTP to RanGDP on the complexes to dissociate it to individual components (Lam and Dean, 2010; Lange et al., 2007).

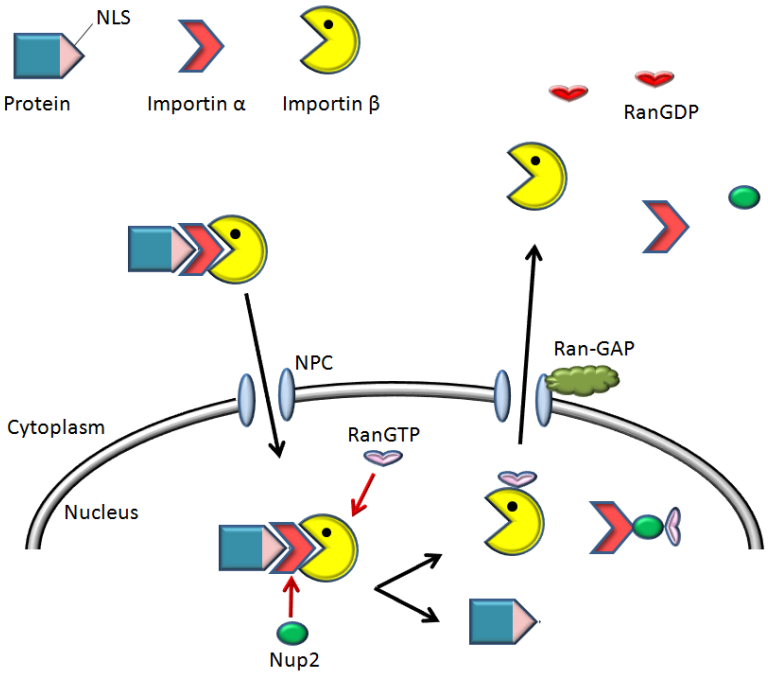


Figure 1.3.1. Nuclear protein import by the importin α/β pathway.
Modified from Lam and Dean, 2010; Lange et al., 2007.

1.3.2. Chloroplast protein import

More than 95% of chloroplast proteins are encoded in the nuclear genome, and majority of them contain chloroplast transit peptides at their N-terminus to be directed to the chloroplasts by the import machinery (Jarvis and Lopez-Juez, 2013; Lee et al., 2013; Shi and Theg, 2013). These proteins containing the transit peptide are called precursors. Although the transit peptides are known, their structure and interactions with the chloroplast import machinery has not yet completely understood. There is some new information about transit peptides based on sequence analysis. Transit peptides do not contain consensus sequences in their primary structure, and the length of transit peptides varies in size from about 20 to >100 residues. Some characteristics are commonly observed including a high content of hydroxylated amino acids and a formation of α-helical structures in hydrophobic environments as reviewed elsewhere (Jarvis and Lopez-Juez, 2013; Lee et al., 2013; Shi and Theg, 2013). Several transit peptides have been used to examine their efficiency in chloroplast targeting of the recombinant proteins to achieve the accumulation of

a particular recombinant protein or modification of a metabolic pathway (Gnanasambandam et al., 2007; Lee et al., 2008). Some transit peptides have preference for different plastid types. For example, the non-photosynthetic ferredoxin III (prFdIII) transit peptide and the ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit (prRBCS) transit peptide of wheat can direct a green fluorescent protein (GFP)-fusion protein into pollen amyloplasts and to plastids in endosperm by the prFdIII transit peptide and the prRBCS transit peptide, respectively (Li and Teng, 2013). Also, at least three transit peptide groups are defined to be imported into the plastids of a different age: Group 1) targeting young chloroplasts, 2) targeting any age of the chloroplasts, and 3) targeting older chloroplasts (Li and Teng, 2013).

Proteins which are imported into the chloroplasts must go through a TOC (translocon at the outer envelope membrane of chloroplasts) and a TIC (translocon at the inner envelope membrane of chloroplasts), illustrated in Figure 1.3.2. The core TOC complex is composed of Toc34, Toc75 and Toc159. Toc34 and Toc159 are GTPases and receptors for the precursors. Toc75 is a β-barrel membrane channel in which precursors can cross the outer membrane (Lee et al., 2013; Paila et al., 2015). Tic20, Tic21 and Tic110 are suggested to be channels across the inner membrane, and several other Tic proteins are involved in translocation of proteins (Li and Teng, 2013; Paila et al., 2015). Some cytosolic proteins are required for chloroplast import. The heat shock protein 90 (Hsp90) can bind both the transit peptide and the N-terminal region of the mature protein. The cytosolic Hsp90 has been shown to form a complex with Hsp70/Hsp90 organizing protein (Hop) and immunophilin FKBP. This complex binds to precursor protein and guides it to Toc64 that contains a cytosolic tetratricopeptide repeat (TPR) domain. In the cytoplasm, proteins having the TPR domain participate in the formation of Hsp70-Hsp90 chaperone complexes. Therefore, the TPR in Toc64 is thought to facilitate transfer of the precursor proteins from Hsp90 to the TOC complex (Lee et al., 2013; Paila et al., 2015). The cytosolic Hsp70 and 14-3-3 proteins can bind to the transit peptide. The 14-3-3 can also interact with the region of the mature protein. Binding of the guidance complex (Hsp70 and 14-3-3) to the transit peptide can escort the precursor protein to the TOC receptors. An outer envelope membrane protein 61 and an Hsp70 receptor at the outer envelope membrane have been proposed to facilitate the transfer of the precursor protein from the guidance complex to the TOC receptors (Lee et al., 2013; Paila et al., 2015). Interestingly, some precursor proteins do not require the guidance complexes to be directed to the TOC complex, yet the interactions of Hsp90 and Hsp70 with the precursor proteins could

potentially prevent misguiding or misfolding of the precursor proteins (Lee et al., 2013; Paila et al., 2015). During the translocation, the stromal processing peptidase cleaves off the transit peptide in the stroma, and mature proteins are released (Lee et al., 2013).

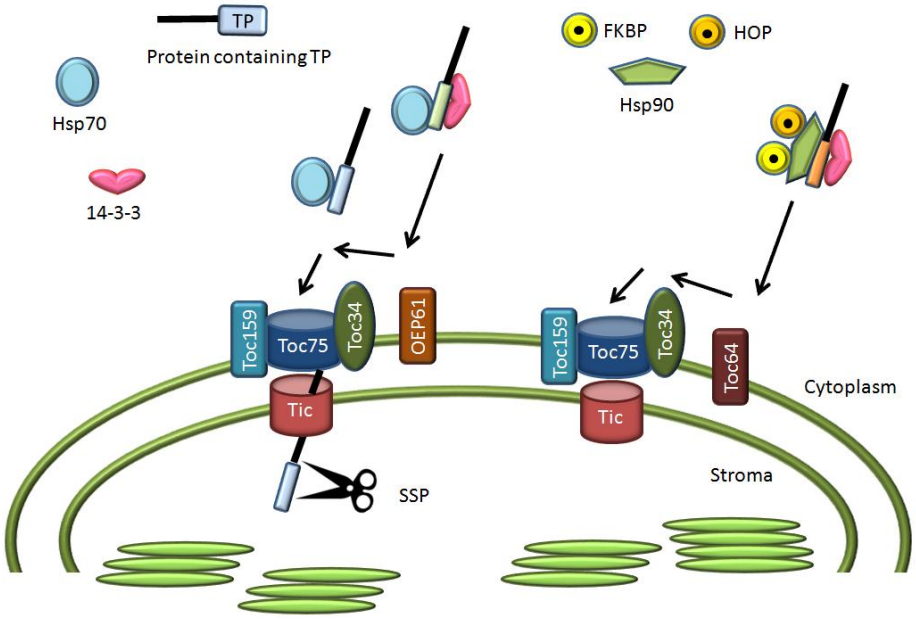


Figure 1.3.2. Chloroplast protein import.
Selected proteins are shown (modified from Lee et al., 2013; Paila et al., 2015). OEP61: outer envelope membrane protein 61; SPP: the stromal processing peptidase.

1.4. *Agrobacterium* VirD2 protein

Agrobacterium VirD2 protein is one of the best studied virulence proteins. It is essential for the transformation process because of its relaxase domain involved in cleavage of the T-DNA border sequences and its nuclear localization signals (Figure 1.4 containing all other motifs). Mutational analyses have been done on VirD2 to define its structural and functional role in the transformation for more than thirty years. Interactions between VirD2 and several plant proteins are also known. It seems that characterization of VirD2 is completed.

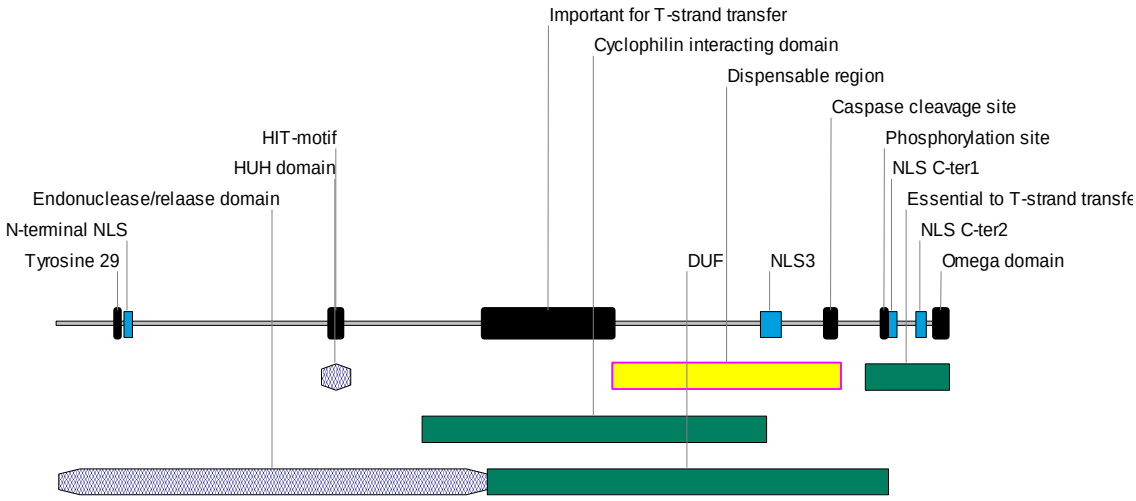


Figure 1.4. Schematic representation of VirD2 with domains and motifs.
NLS C-ter1 and NLS C-ter2: the C-terminal bipartite NLS. NLS3 (see Chapter 3).

1.4.1. Discovery of the *virD2* gene

An octopine (pTiAch5) and a nopaline (pTiC58) Ti plasmids of *Agrobacterium* were analyzed to construct detailed physical maps of the two plasmids by Southern blotting and by electron microscope heteroduplex analysis in 1981 (Engler et al., 1981). Based on the map of the plasmid, transposon insertion mutations were introduced into the regions of the plasmid to determine the change in virulence on *Kalanchoe daigremontiana* stems and leaves or carrot slices by Klee et al. (1983). Authors found five different loci (the virulence region) which affected tumor induction. These loci were named *virA, virB, virC, virD* and *virE* (Klee et al., 1983). Later on, *virG* along with *virA* were shown to be required to activate expression of the *vir* genes (Horsch et al., 1986; Stachel and Nester, 1986). The *virD* operon was suggested to contain two or more cistrons and encode a site specific endonuclease, which is involved in generation of the single stranded T-DNA (T-strand) (Stachel and Nester, 1986; Stachel et al., 1987; Yanofsky et al., 1986). Molecular characterization of the *virD* operon was performed to identify the individual *virD* genes, *virD1, virD2, virD3, virD4* and *virD5* (Porter et al., 1987). Since then, the structure and functions of VirD2 have been revealed. VirD2 contains several domains and signals such as a relaxome domain, the nuclear localization signals (NLSs) and ω domain. In addition, VirD2 has been shown to interact with plant host proteins to play an important role in the transformation process.

1.4.2. Domains and signal sequences in VirD2

1.4.2.1. Relaxase domain

VirD2 was first shown to be involved in cleavage of the T-DNA border sequences and attachment to the 5' end of the T-DNA (Herrera-Estrella et al., 1988; Ward and Barnes, 1988; Young and Nester, 1988). Later on, it was shown that the first 228 amino acids of the VirD2 proteins of an octopine and nopaline Ti plasmid had 82% homology (Steck et al., 1990). This region called the relaxase domain is required for endonuclease activity of VirD2. The relaxase domain was further minimized to 204 amino acids according to sequence comparison between different relaxases (van Kregten et al., 2009). The particular motif of the relaxase domain is critical for its activity. A comparison of amino acid sequences between VirD2 and other bacterial proteins presented that VirD2 contains a homologous region which is conserved in many replication initiation and mobilization proteins such as TraI, Mob and Rlx (Ilyina and Koonin, 1992). This region, called HxDxxx(P/u)HuHuuux (H: His; D: Asp; x: any residue; u: a bulky hydrophobic residue) motif, was analyzed in detail (Ilyina and Koonin, 1992). Relaxases are one of two major classes of the endonuclease superfamily, the HUH endonuclease. The HUH endonuclease contains the HUH motif (H: His; U: a bulky hydrophobic residue) consisting of two His residues separated by a bulky hydrophobic residue and a Y motif including one or two Tyr residues separated by several amino acids (Chandler et al., 2013). VirD2 contains the HUH motif in a form of HxDxxx(P/u)HuHuuux motif (residues 126-139 of VirD2: HVDRDHP**H**L**H**VVVN) (Vogel et al., 1995). When the conserved motif was substituted with different amino acids, the cleavage activity of VirD2 was abolished or reduced severely (Vogel et al., 1995). In VirD2, the His residues (133 and 135), which are absolutely conserved among the family of the virulence and mobilization proteins, were proposed to function as a metal binding motif (Vogel et al., 1995). In fact, HUH enzymes require a divalent metal ion to facilitate their cleavage activity (Chandler et al., 2013). Mg^{2+} was proven to be necessary for the cleavage activity of VirD2 (Scheiffele et al., 1995).

Tyr residues in VirD2 were suggested to be the place of attachment to the 5' end of the T-strand (Durrenberger et al., 1989). Two scenarios of the nicking reaction and the covalent attachment were proposed: 1) a one-step process in which two reactions occur simultaneously and 2) a two-step process in which the covalent attachment takes place after cleavage of the T-DNA border sequences (Vogel and Das, 1992). Therefore, alteration of the necessary Tyr residue would destroy the cleavage activity of VirD2 in

the first case, whereas in the latter case, a change in a Tyr residue would not affect the cleavage activity of VirD2, subsequently another Tyr may attach to the T-strand. Tyr residues at position 29, 68, 99, 119, 121, 160 and 195 within the relaxase domain of VirD2 were substituted with phenylalanine to observe the effects of these mutations (Vogel and Das, 1992). The results showed that a change in Tyr 29 and 121 to phenylalanine abolished the cleavage activity. Further mutations in these two Tyr residues showed that any change in Tyr 29 destroys the cleavage activity, while Tyr 121 can be replaced by a Trp residue (Vogel and Das, 1992). These results clearly demonstrated that Tyr 29 is essential for the nicking of the T-DNA border sequences which is accompanied by its covalent attachment.

1.4.2.2. Nuclear localization signals (NLSs)

VirD2 NLSs are the most characterized domains among other domains. Two types of NLSs exist in the protein: a monopartite type and a bipartite type at the N and C-terminal region, respectively. Initially, Herrera-Estrella et al. (1990) suggested the presence of the N-terminal NLS by examing the β-galactosidase activity in the plant cell fractions of transgenic tobacco expressing VirD2 (first 292 amino acids)-β-galactosidase fusion protein. The nuclear fraction showed approximately a three fold higher β-galactosidase activity than the cytosol fraction. Based on the results and the consensus sequences for nuclear targeting, a monopartite N-terminal NLS (RKGR) was suggested (Herrera-Estrella et al., 1990). Different VirD2 proteins among octopine and nopaline types have high homology in two short stretches of 4-5 amino acids in the C-terminal region of the proteins. These stretches have also extensive homology to the nuclear localization signals of other proteins, which led to a proposition of a bipartite C-terminal NLS of VirD2 (Howard et al., 1992). Truncated VirD2 was fused at the C-terminus of β-glucuronidase (GUS) to observe the GUS activity in transiently transformed tobacco protoplasts by electroporation. The results showed that GUS-truncated VirD2 with the absence of the N-terminal NLS and the presence of the C-terminal NLS was detected in the nucleus, whereas GUS-truncated VirD2 proteins, which contain the wild type N-terminal NLS and lack the C-terminal NLS with either deleted 158 C-terminal amino acids or only eliminated bipartite sequences, were detected in the cytosol (Howard et al., 1992). Therefore, the bipartite C-terminal NLS was confirmed, yet the N-terminal NLS was shown to be non-functional. This discrepancy regarding the N-terminal NLS of VirD2 was solved by other researchers, discussed below. Tinland et al (1992) used

several fragments of VirD2 fused to the N-terminus of β-galactosidase to study its NLSs. The fusion consisting of N-terminal NLS of VirD2 (11 amino acids) and β-galactosidase was detected by indirect immunofluorescence in the nucleus of transiently transformed tobacco protoplasts and yeast cells. The fusion of C-terminal NLS of VirD2 (20 amino acids) to β-galactosidase was also detected in the nucleus of protoplast and yeast cells (Tinland et al., 1992). These results supported the presence of functional N and C-terminal NLSs.

A different transformation system and reporter construct were used in the study by Relic et al. (1998). Subcellular localization of the GFP fused to the N-terminus of VirD2 variants was examined in transiently transformed mammalian cells. GFP-VirD2 carrying the N-terminal NLS and a deletion of the C-terminal NLS by *Nru*I (removal of 45 amino acids) was localized in the nucleus, while GFP-VirD2 lacking both NLSs were localized in the cytoplasm (Relic et al., 1998). Another approach was taken to observe subcellular localization of the fluorescently labelled VirD2 variants, which were directly introduced into the permeabilized tobacco protoplasts (Ziemienowicz et al., 2001). Confocal microscopy showed that VirD2 with an N-terminal NLS and lacking a C-terminal NLS deleted by *Nru*I was localized in the nucleus. However, when this VirD2 variant without the fluorescent label was introduced to the permeabilized protoplasts and mammalian cells with the fluorescently labelled single stranded DNA, the nuclear import of single stranded DNA was not observed (Ziemienowicz et al., 1999; Ziemienowicz et al., 2001). It was indicated that due to the proximity between the N-terminal NLS of VirD2 and Tyr 29, the N-terminal NLS is masked when the protein attaches to DNA which may cause a conformational change in VirD2 (Rossi et al., 1993; Ziemienowicz et al., 1999; Ziemienowicz et al., 2001).

1.4.2.3. Omega domain

The omega domain (DGRGG), which is located at the C-terminus of VirD2, was first proposed by Shurvinton et al. (1992). In their exerpiments, various mutations were introduced into VirD2 to analyze the effects on tumorigenesis on potato tuber discs. The results showed that deletion of three amino acids of the omega doamin and one adjacent amino acid (**bold underlined**) (**D**DGR) caused a decrease in tumorigenesis to 3%. The omega domain was thought to facilitate proper protein folding, which may indirectly affect nuclear targeting of the C-terminal NLS, or to be involved in transfer of the T-DNA-VirD2 complex from

the bacterium to the plant cell (Shurvinton et al., 1992). A few years later, two groups examined the function of the omega domain more closely. Mysore et al. (1998) mutated the omega domain of VirD2 by deleting **D**DGR and showed that *Agrobacterium* harbouring the VirD2 mutant had lower T-DNA transfer efficiency but more severe defect was observed in the T-DNA integration (Mysore et al., 1998). Thus, their conclusion was that the omega domain is important for T-DNA integration. Bravo-Angel et al. (1998) examined the function of the omega domain using two different VirD2 variants in tobacco seedling transformation: 1) precisely eliminating the omega domain (DGRGG) (strain AT*virD2*Δω) and 2) substituting Asp and Arg to Gly (**G**G**G**GG: substituted amino acids in **bold**) (strain AT*virD2*ω/G). Their results showed that both mutants had a reduction in T-DNA transfer efficiency. Strain AT*virD2*Δω decreased the efficiency to a third relative to the wild type strain, while strain AT*virD2*ω/G resulted in a transfer efficiency of 6% relative to the wild type strain. This suggested that pentaglycine likely had a more severe influence on the transfer step than the deletion mutation. Neither the T-DNA integration efficiency nor the T-DNA integration pattern was affected by these mutations. Thus, the omega domain was suggested to be important for the T-DNA transfer (Bravo-Angel et al., 1998). The discrepancy between the studies by Mysore el al. (1998) and Bravo-Angel et al. (1998) was suggested to be due to the use of different mutations in the omega domain (Gelvin, 2012). van Kregten et al. (2009) examined the domain of unknown function (DUF: amino acids 205-395) which is poorly conserved among VirD2. They constructed several fragments of VirD2, and *Agrobacterium* strains harboring the VirD2 variants were used to transiently and stably transform *Arabidopsis* roots. The results showed that the C-terminal 40 amino acids including the omega domain were necessary for the T-DNA translocation to the plant cell. Also, the region between amino acids 205 and 264 provided function in the T-DNA translocation into the plant cell, while amino acids 265-373 were thought dispensable (van Kregten et al., 2009). Despite of these results, the exact function of the omega domain is not yet clear.

1.4.3. Plant proteins that interact with VirD2

Since VirD2 is translocated into the plant cell, it is exposed to many host proteins for interactions. The interactions between VirD2 and several plant proteins including cytoplasmic and nuclear proteins have been examined. It was found that some interactions are necessary for the transformation process, whereas for

some the possible role in the process of transformation or some other unknown functions is yet to be discovered.

1.4.3.1. Cytoplasmic proteins

1.4.3.1.1. Cyclophilins

Cyclophilins are present in all subcellular compartments and involved in many cellular processes such as protein trafficking and maturation, stabilization of receptor complexes, apoptosis, receptor signaling, RNA processing and spliceosome assembly (Romano et al., 2004). Since the discovery of the plant cyclophilins (Gasser et al., 1990), 14 *Arabidopsis* cyclophilin genes have been identified. In addition, 15 genes encoding putative proteins containing cyclophilin-like domains were discovered (Romano et al., 2004). Some of them are known to localize in the cytoplasm including rotamase cyclophilin (ROC) 1-3 and 5-6, while ROC4 is targeted to the stromal compartment of the chloroplast (He et al., 2004; Romano et al., 2004). VirD2 was shown to interact with three cyclophilins ROC1, ROC4 and ROC5 (CrypA) by the yeast two-hybrid system and *in vitro* protein binding assay (Deng et al., 1998). ROC4 showed the weakest interaction with VirD2, suggesting that ROC4 was likely not involved in the T-complex trafficking in the cytoplasm (Deng et al., 1998). The immunosuppressive drug cyclosporin A (CsA), a cyclic 11 amino acid peptide produced by the fungus *Tolypocladium inflatum*, interacts with the cyclophilin to form a complex which binds to inhibit calcineurin, a calcium-calmodulin-activated Ser/Thr specific protein phosphatase. This prevents nuclear import of proteins which require activity of this type of phosphatise (Wang and Heitman, 2005). Addition of CsA during binding of VirD2 and Roc1 and Roc5 inhibited association of VirD2 with both types of cyclophilins (Deng et al., 1998). Moreover, CsA inhibited *Agrobacterium*-mediated transient transformation of tobacco cell suspension and *Arabidopsis* roots. The results of the yeast two-hybrid analysis using different regions of VirD2 determined that regions 174 to 337 and 274 to 337 are required for ROC1 and ROC5 interactions, respectively (Deng et al., 1998). Interestingly, however, *Agrobacterium* containing VirD2 with only the first 204 amino acids of VirD2 attached to SV40 NLS at its N-terminus and a translocation signal of VirF at its C-terminus was still able to transform plants both transiently and stably (van Kregten et al., 2009). Thus, interactions between VirD2 and cyclophilins do not likely have a specific role in the transformation step. Some of cyclophilin-like proteins (AtCYP59,

AtCYP63 and AtCYP95) are predicted to localize in the nucleus (He et al., 2004; Romano et al., 2004). It might be interesting to examine if VirD2 interacts with these putative cyclophilin in the nucleus, which might have some effects during the transformation.

1.4.3.1.2. Caspases

Caspases belong to a family of Cys proteases which cleave their substrates after an Asp residue. These proteases are well known to be associated with programmed cell death in mammalian cells, and they are also involved in the programmed cell death in plants through the hypersensitive response caused by pathogen invasion (Coll et al., 2011). Interestingly, VirD2 was found to be one of the caspase substrates (Chichkova et al., 2004). VirD2 was fragmented after incubation with the human caspase-3, which suggested that targeted Asp residues were located close to the C-terminus of VirD2. The C-terminal 86 amino acid region of VirD2 was fused to GFP at its N or C-terminus. The results of cleavage products by the caspase-3 showed that putative target sites were GEQD and TATD. Asp residues were substituted with **Ala** residues (GEQ**A** and TAT**A**), which did not allow the caspase-3 to cleave the sequence. A partially purified tobacco caspase specifically cleaved the TATD sequence (Chichkova et al., 2004). Based on this study, Reavy et al. (2007) substituted GEQ**D** to GEQ**A** and TAT**D** to TAT**A** in the nopaline VirD2. *Agrobacterium* harbouring this mutated VirD2 was used for transient expression of GFP in tobacco, spinach, barley, physalis and oil seed rape. The total area of GFP-expressed cells increased about 4-5 fold in tobacco, spinach and barley, and about 20 fold in physalis and oil seed rape relative to *Agrobacterium* harbouring the wild type VirD2 protein. Thus, these mutations in VirD2 were suggested to improve *Agrobacterium*-mediated plant transformation by preventing cleavage of VirD2 by plant caspase-like proteins (Reavy et al., 2007). Interestingly, two target sites (GEQD and TATD) exist in the nopaline type VirD2 protein (accession number: NC_003065), while the opine type VirD2 protein (accession number: CP007228) contains only GEQD. Since a tobacco caspase-like protein did not target the GEQD sequence, the substitution of an Asp residue may not have any influence on transformation using the octopine VirD2.

1.4.3.2. Cytoplasmic-nuclear proteins

1.4.3.2.1. Phosphatase

VirD2 has been shown to interact with phosphatase in the cytoplasm. Tomato DIG3, type 2C Ser/Thr protein phosphatase (PP2C) was tested for its interaction with VirD2 (Tao et al., 2004). A tomato cDNA library was first searched for the protein that binds specifically to VirD2. Tomato DIG3 was found to be strongly associated with the C-terminal region of VirD2 defined by the yeast two-hybrid analyses. The amino acid analysis of DIG3 suggested that its likely encodes PP2C. The results of DIG3 phosphatase activity assay indicated DIG3 indeed possesses the PP2C activity. The DIG3 potential target site was predicted to be Ser 394 in the proximity of the C-terminal NLS of VirD2. Effect of DIG3 on subcellular localization of GUS-cVirD2 (the C-terminal half of VirD2) was examined using electroporated tobacco protoplasts. GUS-cVirD2 in plants lacking DIG3 showed 80-90% of the GUS activity exclusively in the nucleus, while nuclear localization of GUS-cVirD2 in plants overexpressing DIG3 decreased to about 40%. A Ser 394 was substituted with an Ala residue in GUS-cVirD2 (GUS-mcVirD2) to observe the effect of DIG3. The GUS activity was about 50% regardless of DIG3. GUS-NIa (a tobacco etch virus protein containing plant functional NLS) was used to see any effects by DIG3, and DIG3 did not affect GUS-NIa at all. Thus, DIG3 is specific to a Ser at the C-terminal NLS of VirD2 (Tao et al., 2004). In conclusion, a phosphorylation/dephosphorylation target site is likely Ser 394 which can modulate VirD2 NLS activity. Phosphorylation of a Ser residue has been shown to up-regulate nuclear import by increasing its affinity for a specific importin. For example, phosphorylation of Ser 385 in NLS of the Epstein-Barr virus (EBV) nuclear antigen 1 (EBNA-1) protein enhances its nuclear import by promoting its binding affinity for importin α5, while replacement of serine 385 with an Asp residue decreases the binding affinity for this particular importin (Nardozzi et al., 2010). The NLS of the large tumor antigen of simian-virus 40 (SV40) is likely one of the best characterized NLSs. SV40 NLS (KKKRKV) does not contain a serine, but upstream of NLS has several Ser residues that can be phosphorylated. Although phosphorylation of serine residues of the classical monopartite NLS has been shown to accelerate the nuclear import of some proteins including SV40 large T-antigen, it does not increase the affinity of the NLS to the importin such as importin α1. Thus, phosphorylation upstream of NLS is suggested to alter the recognition of the NLS by the importin protein

(Nardozzi et al., 2010). It is possible that phosphorylation of VirD2 helps its nuclear import by the importin α/β pathway.

1.4.3.2.2. Importin proteins

The proteins containing NLSs are imported into the nucleus by the importin α/β pathways. NLSs are recognized by the adaptor protein importin α, which interacts with a heterodimeric import receptor composed of the β-karyopherin (importin β) that mediates interactions with the nuclear pore complex to release cargo into the nucleus (Walde and Kehlenbach, 2010). Since VirD2 contains a monopartite NLS and a bipartite NLS at its N and C-terminal regions, several importins have been identified. First, interaction between VirD2 and *Arabidopsis* AtKAPα (IMPa-1) was identified in 1997 by Ballas and Citovsky. VirD2 carrying the precise deletion (KRPR and RKRER of a nopaline type VirD2) of the C-terminal NLS (VirD2ΔNLS) did not interact with IMPa-1 *in vitro* and in the yeast two-hybrid system (Ballas and Citovsky, 1997). When fluorescently labeled VirD2ΔNLS was introduced into permeabilized yeast cells expressing IMPa-1, VirD2ΔNLS was not imported into the nucleus (Ballas and Citovsky, 1997). Bako et al. (2003) indicated interactions of VirD2 with IMPa-3, IMPa-4 and IMPa-6 in the yeast two-hybrid system. Later, Bhattacharjee et al. (2008) confirmed the interaction between VirD2 and the previously mentioned importin α proteins, and they also reported several other importin α proteins were able to bind to VirD2. In this study, seven importin α (IMPa-1, IMPa-2, IMPa-3, IMPa-4, IMPa-6, IMPa-7 and IMPa-9) were examined *in vitro*, by the yeast two-hybrid system and by biomolecular fluorescence complementation (BiFC) (Bhattacharjee et al., 2008). All examined importin α proteins were shown to interact *in vitro* and in yeast, while only IMPa-1, IMPa-4, IMPa-7 and IMPa-9 were detected to form a complex with VirD2 in tobacco protoplasts using the BiFC system (Bhattacharjee et al., 2008). Four VirD2 deletion mutants were constructed to examine IMPa-4 specifically: 1) ΔNLS: precise deletion of the C-terminal NLS (KRPR and KRAR) of an octopine VirD2, 2) Δ*Nru*I: deletion of 45 amino acids including the C-terminal NLS, 3) Δ*Kpn*I: a 70% internal deletion of VirD2 with presence of the N and C-terminal NLS, and 4) ATG-*Bam*HI: an N-terminal half of VirD2. IMPa-4 interaction was almost abolished in VirD2Δ*Nru*I and VirD2 ATG-*Bam*HI, whereas VirD2ΔNLS and VirD2Δ*Kpn*I still showed high interaction with IMPa-4 in the yeast two-hybrid system (Bhattacharjee et al., 2008). It would be interesting to see the interactions between VirD2 and

IMPa-5/IMPa-8, and to check the effect phosphorylation/dephosphorylation status of VirD2 may have on binding affinity to different importins.

1.4.3.3. Nuclear proteins

1.4.3.3.1. Cyclin-dependent kinase activating kinases and TATA box-binding protein

Cyclin-dependent kinases (CDKs) are in the Ser/Thr protein kinase family, which regulates a number of the cellular and developmental processes in eukaryotes. For example, CDKs can increase or decrease transcription by phosphorylating transcription factors or indirectly interacting with suppressors and activators. In addition, CDKs can associate with the C-terminal domain (CTD) of RNA polymerase II where a CDK consensus phosphorylation site exists (Doonan and Kitsios, 2009). CDKs are first phosphorylated by a CDK-activating kinase (CAK) to be activated to regulate cellular processes (Umeda et al., 2005). Alfalfa CAK (CAK2Ms) was found to phosphorylate VirD2 and also phosphorylate the C-terminal domain of RNA polymerase II large subunit where the TATA box-binding protein is recruited (Bako et al., 2003). VirD2 was reported to have several possible phosphorylation sites such as Ser and Thr motifs for Pro-directed protein kinases. First, abilities of a nuclear kinase and a cytoplasmic kinase to phosphorylate predicted sites of VirD2 were examined. The purified nuclear protein fraction and cytoplasmic protein fraction were used to examine which fractions contain the kinase that phosphorylates VirD2. The results showed that a high level of the phosphorylation on VirD2 was observed in the nuclear fraction, whereas a low level of the kinase activity was observed in the cytoplasmic protein fraction. Hence, a nuclear kinase plays a major role in phosphorylation of VirD2 (Bako et al., 2003). Subsequently, the regions between 2-109 and 248-447 amino acid of VirD2 were identified as phosphorylation sites. Due to amino acid sequence homology and the kinase activity on VirD2 and CTD, the alfalfa CAK2Ms kinase is a close homolog of *Arabidopsis* CAK2At and rice R2Os. Furthermore, the histidine-triad (HIT) motif containing (His-X-His-X-His: X as any amino acids) which associates with CAKs was found in the relaxase-domain of VirD2. This motif was identified as the binding site of the TATA binding protein (TBP), which is involved in tight association between VirD2 and TBP (Bako et al., 2003).

This low level of phosphorylation of VirD2 found in the cytoplasmic fraction might answer the question which kinase is involved to enhance nuclear import of phosphorylated VirD2, which might be

dephosphorylated by phosphatase. It might be interesting to see the levels of kinases and phosphatases in different organisms and different cell types in a given plant and to check whether these levels might affect VirD2 activity and change transformation efficiency.

1.4.3.3.2. Histones

Histones, especially H2A have been shown to be critical factors during the T-DNA integration process. *Arabidopsis* with an integration of the T-DNA into the 3' untranslated region of the histone *H2A-1* gene, one of the 13 member family of histone H2A, exhibited a rat (resistance to *Agrobacterium* transformation) phenotype (Mysore et al., 2000). Interestingly, overexpression of *H2A-1* enhanced the transformation (Mysore et al., 2000). *H2A-1* expression was examined closely to discover that the *H2A-1* gene was expressed in many non-dividing cells, namely in the elongation zone of the root, which are highly susceptible to *Agrobacterium* transformation (Yi et al., 2002). The association between *Agrobacterium* proteins and histones was described by Lacroix et al. (2008). VirE2 interacting protein 1 (VIP1) was shown to interact with mononucleosomes containing the core histone proteins (H2A, H2B, H3 and H4) of cauliflower florets *in vitro*, while VirD2 was unable to bind to any histones (Lacroix et al., 2008). Recently, Wolterink-van Loo et al. (2015) reported the interactions between VirD2 and histones of the yeast. BiFC was used to confirm interactions between VirD2 and yeast histones, H2A, H2B, H3 and H4. Among four histones, H2A and H2B showed more efficient interaction with VirD2 (Wolterink-van Loo et al., 2015). This study challenged the finding of Lacroix et al. (2008). It would be interesting to revisit VirD2 for its interaction with plant histones using BiFC or other approaches. Nevertheless, Wolterink-van Loo et al. (2015) showed the interaction of VirD2 with histones; they also suggested a possible role of VirD2 in T-DNA integration.

1.5. Objectives

Agrobacterium-mediated plant transformation is a well characterized process. Each Vir protein has been examined to elucidate their structure and function. Optimized transformation protocols are available for a number of plant species. Chloroplast transformation offers great potential to produce high levels of recombinant proteins, but only two methods, biolistic and PEG-mediated methods, are available.

Therefore, the first objective of my thesis was to develop a new method of transforming chloroplasts via *Agrobacterium* by modifying *Agrobacterium* genetics. I hypothesized that VirD2 carrying inactivated NLSs and an additional transit peptide (tomato defective chloroplasts and leaves transit peptide or tobacco ribulose bisphosphate carboxylase small subunit transit peptide) would direct the T-DNA complex to the chloroplast instead of the nucleus. I examined the cleavage activity of the modified VirD2 carrying the transit peptide to be sure that it was not disturbed by the presence of the transit peptide.

I performed *Agrobacterium*-mediated chloroplast transformation by tobacco leaf incision; however, the transformation of tobacco chloroplasts was unsuccessful. Although known functional NLSs were eliminated in the modified VirD2 containing a strong chloroplast transit peptide, the transit peptide was overridden by an unknown factor. Therefore, the second objective of my thesis was to identify a novel NLS(s) in the VirD2 protein.

Chapter 2

Modification of *Agrobacterium* VirD2 Protein for Chloroplast Transformation

2.1. Abstract

Agrobacterium is often referred to as a *natural genetic engineer* because of its ability to transform the plant nuclear genome. *Agrobacterium*-mediated plant transformation has been a springboard for the development of plant biotechnology. The transformation process has been examined extensively, and it was revealed that the *Agrobacterium* VirD2 protein is one of the key factors in the process of nuclear genome transformation due to its relaxase domain and nuclear localization signals (NLSs). VirD2 forms a complex with the transferred DNA (T-DNA) containing the transgene and pilots the complex termed the T-complex into the nucleus where the transgene is integrated into the nuclear genome. Knowledge of the *Agrobacterium* transformation mechanism has been applied to plant biotechnology. Various plant species have been transformed by the *Agrobacterium*-mediated method to introduce new traits such as abiotic/biotic resistance and herbicide tolerance. These transgenic plants are commercially available worldwide.

Chloroplast transformation is an alternative to plant nuclear genome transformation and offers attractive benefits. For example, chloroplast maternal inheritance prevents transgene escape through pollen. High transgene expression is usually observed due to a high copy number of chloroplast genomes. However, protocols for the regeneration through somatic embryogenesis are essential in chloroplast transformation, yet the established protocols are available only for selected plants species. Furthermore, the complete chloroplast genome sequence required for designing a chloroplast transformation vector is available for selected plant species. Since a natural genetic engineer for chloroplasts does not exist, biolistics is the most common method to transform chloroplasts. Interestingly, in the past, *Agrobacterium* has been used to transform chloroplasts without modifying *Agrobacterium* proteins, but it has never been repeated.

Based on the defined *Agrobacterium*VirD2 structure and function, we hypothesize that VirD2 which carries inactivated NLSs and is fused to a chloroplast transit peptide would guide the T-complex to chloroplasts instead of the nucleus. We generated *Agrobacterium* producing such modified VirD2, examined its cleavage activity, and transformed tobacco leaves by the leaf incision method. The results indicated that the addition of a transit peptide to the modified VirD2 protein did not affect the cleavage activity required at the early transformation step. However, the transformation of tobacco chloroplasts was unsuccessful. These puzzling results lead us to hypothesize a possibility of a novel cryptic NLS in VirD2.

2.2. Introduction

Agrobacterium tumefaciens, a Gram-negative phytopathogenic bacterium causing plant crown gall disease is often referred to as a *natural genetic engineer* because of the bacterial ability to transfer its DNA into the plant nuclear genome (Meyers et al., 2010; Pacurar et al., 2011). *Agrobacterium* possesses a tumor-inducing (Ti) plasmid containing a transferred DNA (T-DNA) that is integrated into the plant nuclear genome upon plant genetic transformation (Gelvin, 2003; Meyers et al., 2010). *Agrobacterium* virulence (Vir) proteins encoded in the Ti plasmid and their interactions with a plant during the transformation process have been analyzed in detail, yet it is still an ongoing progress (Gelvin, 2010b). One of the well-studied Vir proteins is the VirD2 protein which contains the relaxase domain that cleaves the T-DNA right and left border sequences at the early stage of transformation. Upon cleavage of the border sequences, VirD2 covalently attaches to the 5' end of the single-stranded T-DNA (T-strand) via Y29 to form the T-DNA-VirD2 complex. Subsequently, the complex and other Vir proteins such as VirE2 are transferred to a plant cell through a type IV secretion system consisting of VirB and VirD4 proteins. In the plant cell, VirE2 proteins bind to the T-strand of the complex forming the T-complex to protect it from degradation (Gelvin, 2010b, 2012).

Nuclear targeting of the T-complex is mediated via nuclear localization signals (NLSs) of VirD2 and possibly VirE2. VirD2 contains monopartite and bipartite types of NLSs at its N and C-terminal regions, respectively (Figure 2.2a) (Herrera-Estrella et al., 1990; Howard et al., 1992; Tinland et al., 1992). However, the N-terminal NLS is suggested to be non-functional in nuclear targeting of the T-complex due to its proximity to Y29 where the 5' end of T-DNA is attached (Vogel and Das, 1992). The C-terminal NLS of VirD2 is essential for nuclear targeting of the T-complex (Rossi et al., 1993; Shurvinton et al., 1992). In addition,VirE2 contains a bipartite NLS positioned in the central region, yet the role of its NLS is the subject to debate because of conflicting reports summarized in a review (Gelvin, 2012); nonetheless, VirE2 is important to maintain T-strand integrity.

Knowledge of the *Agrobacterium* infection mechanism has been applied to plant biotechnology by developing a method to manipulate the plant genome since 1983 (Barton et al., 1983). Over the years, the binary system has been developed to modify the plant nuclear genome more efficiently. Two plasmids are necessary in this system: 1) the helper plasmid originated from the *Agrobacterium* Ti plasmid containing

the *vir* genes but lacking a functional T-DNA region, and 2) the binary vector composed of the gene of interest flanked by right and left T-DNA border sequences and an origin(s) of replication that can function in both *Escherichia coli* and *Agrobacterium* (Lee and Gelvin, 2008; Meyers et al., 2010). Various types of genes have been placed in the T-DNA region for nuclear genome transformation. Furthermore, optimization of transformation conditions and plant regeneration protocols has made it possible to extend the range of the *Agrobacterium* host to include various dicot and monocot plants. Thus, currently transgenic plants with new traits such as abiotic/biotic tolerance and herbicide resistance which have been produced by *Agrobacterium*-mediated plant transformation are commercially available worldwide (James, 2014).

Chloroplast transformation is another way to modify the plant genome, namely, the chloroplast genome. Chloroplast transformed plants called transplastomic plants offer different benefits from plants with modified nuclear genomes. For example, chloroplast maternal inheritance prevents transgene transmission through pollen (Bock, 2014; Liu et al., 2013). High expression of the transgene, normally 5-25% TSP, is usually observed due to a high copy number of the chloroplast genome per chloroplast (Maliga, 2003; Meyers et al., 2010). In addition, transgenes are integrated into the chloroplast genome by homologous recombination at precise locations, and they are not silenced (Bock, 2014; Liu et al., 2013).The compartmentalization of chloroplasts helps minimize the toxicity of foreign proteins which are sometimes toxic in the cytosol (Daniell et al., 2002; Daniell et al., 2005).

Because a natural genetic engineer to transform the chloroplast genome does not exist, typically the biolistics system has been used to transform chloroplasts. The first successful experiment to transform chloroplasts was reported in 1988 using the unicellular green alga, *Chlamydomonas reinhardtii* (Boynton et al., 1988). Shortly after, chloroplasts in *Nicotiana tabacum* were transformed, and stable transplastomic plants were generated (Svab et al., 1990). To date, the DNA delivery system and regeneration protocol have been optimized for selected plant species (Verma and Daniell, 2007; Verma et al., 2008), yet chloroplast transformation still faces limitations. For example, regeneration protocols through somatic embryogenesis are required for chloroplast transformation, and established protocols are available only for selected plant species (Daniell et al., 2005; Liu et al., 2013). Furthermore, the chloroplast genome sequence is required to design the vector with a homologous sequence for recombination, but availability of the complete chloroplast genome sequence is limited (Liu et al., 2013; Verma and Daniell, 2007). Another challenge is to

obtain high expression of the gene of interest in non-green plastids because gene regulation differs; for example, it differs among chloroplasts in leaves, chromoplasts in tomato fruits, and amyloplasts in potato tuber (Kahlau and Bock, 2008; Valkov et al., 2009). Despite these obstacles, chloroplast transformation is an attractive alternative to nuclear genome transformation. Interestingly, two cases of *Agrobacterium*-mediated chloroplast transformation have been reported in the past (Block et al., 1985; Venkateswarlu and Nazar, 1991). In either study, none of the aforementioned *Agrobacterium* Vir proteins was modified to pilot the T-complex into chloroplasts. Unfortunately, *Agrobacterium*-mediated chloroplast transformation has never been reproduced since.

Because *Agrobacterium* is extremely useful for transformation of the plant nuclear genome in various plant species and regeneration protocols are established, our objective was to modify *Agrobacterium* genetics for *Agrobacterium*-mediated chloroplast transformation. We hypothesized that VirD2 carrying inactivated NLSs and fused to a chloroplast-targeting signal, a transit peptide, would facilitate directing the T-complex into the chloroplast instead of the nucleus (Figure 2.2b).

VirD2 NLSs have been previously modified. A precise deletion eliminating solely each part of the bipartite C-terminal NLS or a large deletion eliminating the C-terminus one third of VirD2 was fused to a reporter protein, and the fusion proteins were detected in cytosol despite the presence of the N-terminal NLS (Howard et al., 1992). Howard et al. (1992) proposed that fusing GUS to the N-terminus of VirD2 might have resulted in masking the N-terminal NLS. Moreover, VirD2 carrying a precise deletion or a broad deletion, which eliminated a region encoding 45 amino acids including the C-terminal NLS by *Nru*I, reduced the transient and stable transformation efficiency (Rossi et al., 1993; Shurvinton et al., 1992). Several point mutations were also introduced in the N-terminal NLS. Only one type of mutation (a change from arginine to threonine) in the N-terminal NLS reduced transient expression of the reporter gene. In addition, this N-terminal NLS mutation with a broad deletion for the C-terminal NLS led to low expression of the reporter gene (Rossi et al., 1993).

The mechanisms of chloroplast protein import have been described. More than 95% of chloroplast proteins are encoded in the nuclear genome, and they are post-translationally imported into the chloroplasts. Majority of the chloroplast proteins contain a transit peptide at their N-terminus which mediates chloroplast protein import. Cytosolic sorting factors such as 14-3-3 protein, Hsp70, Hsp90, FKBP, and HOP are

involved in recognition of the transit peptide and facilitate protein import through the translocon at the outer envelope membrane of chloroplasts and the translocon at the inner envelope membrane of chloroplasts. Upon import into the stroma of the chloroplast, the transit peptide is removed by the stromal processing peptidase, and then proteins are properly folded or directed to appropriate compartments within the chloroplast. Interestingly, transit peptides lack consensus sequences, and the length of transit peptides varies in size from about 20 to >100 residues. However, some common characteristics are defined including a high content of hydroxylated amino acids and formation of α-helical structures in hydrophobic environments reviewed elsewhere (Jarvis and Lopez-Juez, 2013; Lee et al., 2013; Shi and Theg, 2013). Several transit peptides have been used to examine their efficiency in chloroplast targeting of recombinant proteins to serve the purpose of accumulation of particular recombinant proteins or modification of metabolic pathways (Gnanasambandam et al., 2007; Lee et al., 2008).

According to the acquired knowledge of VirD2 and chloroplast protein import, we selected the following modifications on VirD2 and transit peptides. In order to ensure that none of the NLSs of VirD2 is functional, we chose an N-terminal modification (a change from arginine to threonine) and a broad deletion by *Nru*I or a precise deletion of the C-terminal NLS to generate two different types of modified VirD2 (mVirD2; Figure 2.4.2a). Furthermore, we selected the transit peptides (TPs) from tomato defective chloroplasts and leaves (DCL) and the tobacco ribulose bisphosphate carboxylase (Rubisco) small subunit (RbcS) based on a high efficiency of targeting the fusion protein of the transit peptide-green fluorescent protein (TP-GFP) to chloroplasts (Gnanasambandam et al., 2007). Consequently, four different TP-mVirD2 proteins were generated in this study (Figure 2.4.2a). Moreover, we constructed expression vectors carrying *TP-mvirD2* and determined the cleavage activity of the recombinant TP-mVirD2 protein to confirm that addition of the transit peptide to mVirD2 did not disrupt the cleavage activity of VirD2 required to process the T-DNA border sequences. Subsequently, we generated *Agrobacterium* producing the TP-mVirD2 fusion protein for *Agrobacterium*-mediated chloroplast transformation by the tobacco leaf incision method. Here, we report that we were successful to purify one of the TP-mVirD2 proteins, and it retained cleavage activity. However, the transformation of tobacco chloroplasts was unsuccessful, suggesting that transit peptides were not functional during *Agrobacterium*-mediated chloroplast transformation. Our results raise a possibility of the existence of a novel cryptic NLS in VirD2.

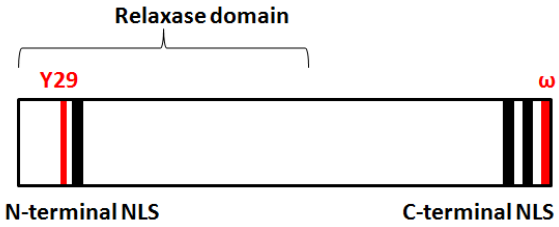


Figure 2.2a. Schematic representation of VirD2.
NLSs are indicated in black. Y29 and ω domain are indicated in red.

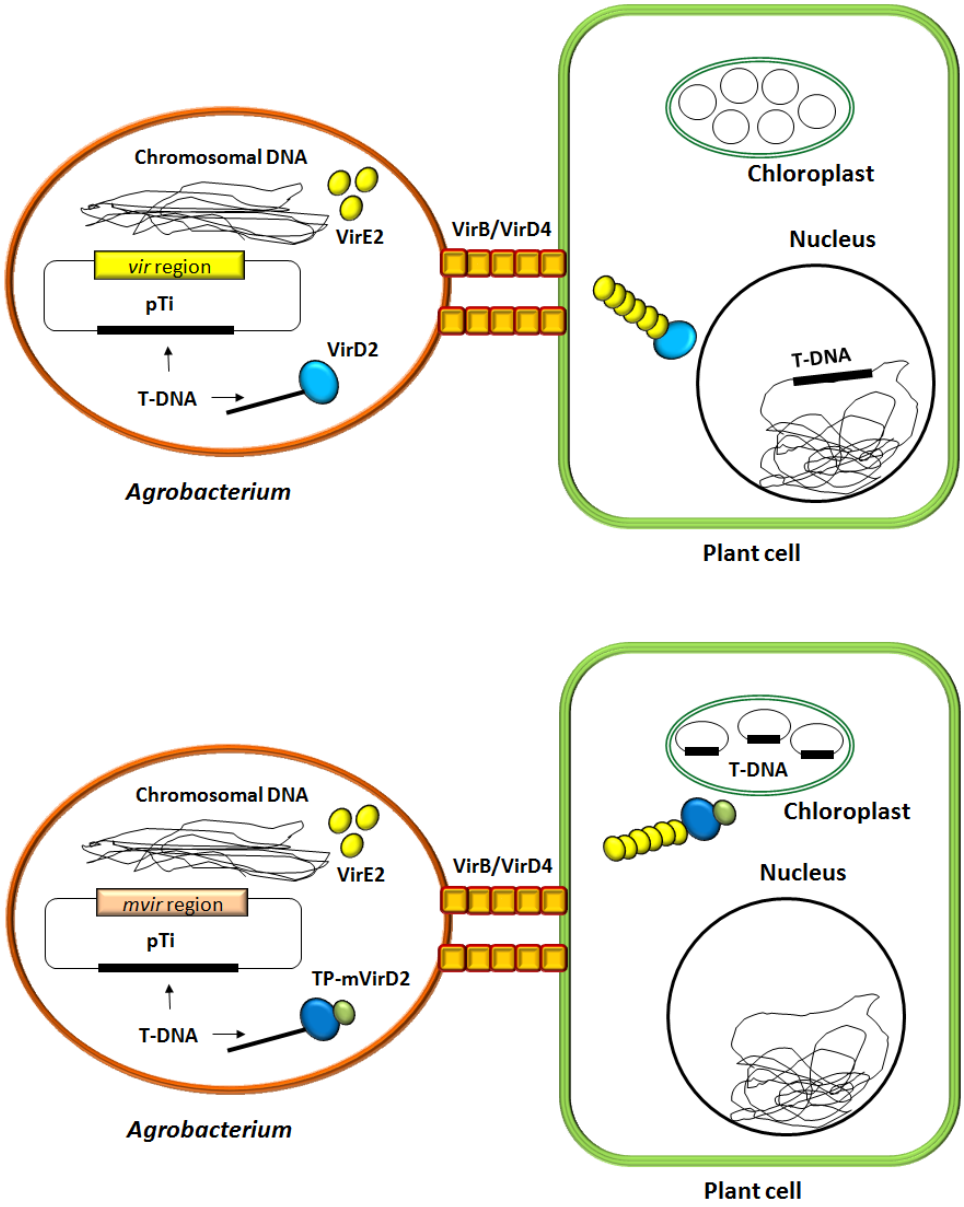


Figure 2.2b. Top – *Agrobacterium*-mediated plant nuclear genome transformation; Bottom –*Agrobacterium*-mediated chloroplast transformation.
vir region: *virulence* region encoding all Vir proteins; *mvir* region: the modified *vir* region encoding the modified VirD2 protein and all other wild type Vir proteins; VirB/VirD4: a type IV protein secretion system; VirD2: wild type VirD2; TP-mVirD2: VirD2 carrying inactivated NLS fused to a transit peptide.

2.3. Materials and Methods

A description of all initial and newly constructed plasmids and primer sequences are listed on Table 2.3a and 3.3b, respectively. The scheme of TP-m-VirD2 construction is illustrated in Figure 2.4.2a. The newly constructed plasmids illustrated in Figure 2.4.1a-e were confirmed by sequencing.

2.3.1. Construction of TP-mVirD2 vectors

2.3.1.1. Modifications of VirD2 NLSs to generate mVirD2

A *Cla*I-*Eco*RI fragment of pAVD44was cloned into the *Cla*I-*Eco*RI digested pVD43 resulting in pVD44 (Table 2.3a). The N-terminal NLS coding sequence was amplified as a *Sma*I and *Pst*I fragment using pVD43 to introduce a point mutation in the N-terminal NLS coding sequence by PCR. PCR was performed in a 20 µL reaction mixture containing 5 ng of pVD43, 0.5 µM of primers AZ44 and AZ45 (Table 2.3b), 0.02 U/µL of Phusion DNA Polymerase (Fermentas), 1xHF Buffer (Fermentas), 200 µM of each dNTP. PCR parameters were as follows: one cycle at 98^{o}C for 30 s, 30 cycles at 98^{o}C for 10 s, 66^{o}C for 30 s, 72^{o}C for 30s, and one cycle at 72^{o}C for 10 min. The *Sma*I-*Pst*I digested PCR product was cloned into the *Sma*I-*Pst*I cut pVD70 resulting in pVD80.

The *Sma*I-*Pst*I fragment of *mvirD2* carrying the modification in the N-terminal NLS coding sequence was amplified to introduce the *Age*I site just after the start codon of *virD2* by PCR. PCR was performed in a 20 µL reaction mixture containing 5 ng of pVD80, 0.5 µM of primers AZ44 and AZ47(Table 2.3b), 0.02 U/µL of Phusion DNA Polymerase (Fermentas), 1xHF Buffer (Fermentas), 200 µM of each dNTP. PCR parameters were as follows: one cycle at 98^{o}C for 30 s, 30 cycles at 98^{o}C for 10 s, 72^{o}C for 30 s, 72^{o}C for 30s, and one cycle at 72^{o}C for 10 min. The *Sma*I-*Pst*I digested PCR product was cloned into the *Sma*I-*Pst*I cut pVD44 and pVD80 resulting in pVD58-*Age*I and pVD80-*Age*I, respectively.

2.3.1.2. Construction of TP-mVirD2

Total genomic DNA of a commercially available tomato (*Solanum lycopersicum* cultivar Roma) was isolated using GenElute^{TM} Plant Genomic DNA Miniprep kit (Sigma) according to the manufacture's protocol. The sequence coding for DCL transit peptide plus a short amino acid sequence of the mature protein was amplified and flanked by *Age*I sites by PCR. PCR was performed in a 20 µL reaction mixture

containing 100 ng of tomato gDNA, 0.5 µM of primers AZ40 and AZ41(Table 2.3b), 0.125 U/µL of Taq DNA Polymerase (Fermentas), 1x Taq Buffer (Fermentas), 200 µM of each dNTP. PCR parameters were as follows: one cycle at 95^{o}C for 3 min, 30 cycles at 95^{o}C for 30 s, 58^{o}C for 30 s, 72^{o}C for 1 min, and one cycle at 72^{o}C for 10 min. The PCR product was purified using QIAquick® PCR Purification kit (Qiagen) according to the manufacturer's protocol. PCR was again performed using 1 ng of the purified PCR product, 0.5 µM of primers AZ48 and AZ50 (Table 2.3b), 0.125 U/µL of Taq DNA Polymerase (Fermentas), 1x Taq Buffer (Fermentas), 200 µM of each dNTP. PCR parameters were as follows: one cycle at 95^{o}C for 3 min, 30 cycles at 95^{o}C for 30 s, 66^{o}C for 30 s, 72^{o}C for 1min, and one cycle at 72^{o}C for 10 min. This PCR product was digested with *Age*I, and then cloned into the *Age*I-digested pVD58-*Age*I and pVD80-*Age*I resulting in pVD58-DCL-TP and pVD80-DCL-TP, respectively.

Total RNA was isolated from *Nicotiana tabacum* cultivar Big Havana wild-type plants using Tri Reagent® (Molecular Research Center, Inc.) according to the manufacturer's protocol. cDNA was synthesized from total RNA using RevertAid H Minus First Strand cDNA Synthesis kit (Fermentas) according to the manufacturer's protocol. PCR was performed to amplify the sequence coding for tobacco RbcS transit peptide and a short amino acid sequence of the mature protein using cDNA by PCR. PCR was performed in a 20 µL reaction mixture containing 2 µL of tobacco cDNA, 0.5 µM of primers AZ51 and AZ53 (Table 2.3b), 0.02 U/µL of Phusion DNA Polymerase (Fermentas), 1xHF Buffer (Fermentas), 200 µM of each dNTP. PCR parameters were as follows: one cycle at 98^{o}C for 30 s, 30 cycles at 98^{o}C for 10 s, 50.3^{o}C for 30 s, 72^{o}C for 30 s, and one cycle at 72^{o}C for 10 min. The *Age*I-digested PCR product was cloned into the *Age*I site of pVD58-*Age*I and pVD80-*Age*I resulting in pVD58-RbcS-TP and pVD80-RbcS-TP, respectively.

2.3.2. VirD2 and TP-mVirD2 cleavage activity assay

2.3.2.1. Construction of expression vectors

The *DCL-TP-mvirD2-58* and *DCL-TP-mvirD2-80* flanked by *Nhe*I and *Xho*I sites were amplified by PCR. PCR was performed in a 20 µL reaction mixture containing 5 ng of pVD58-DCL-TP or pVD80-DCL-TP, 0.5 µM of primers AZ61 and AZ68 (Table 2.3b), 0.02 U/µL of Phusion Hot Start DNA Polymerase (Fermentas), 1xHF Buffer (Fermentas), 200 µM of each dNTP. PCR parameters were as

follows: one cycle at 98°C for 30 s, 30 cycles at 98°C for 10 s, 68.5°C for 30 s, 72°C for 1 min, and one cycle at 72°C for 10 min. The *Nhe*I/*Xho*I-digested PCR fragments from pVD58-DCL-TP and pVD80-DCL-TP were cloned into *Nhe*I/*Xho*I-digested pET21b resulting in pET21b-DCL-TP-mVirD2-58 and pET21b-DCL-TP-mVirD2-80, respectively.

The *RbcS-TP-mvirD2-58* and *RbcS-TP-mvirD2-80* flanked by *Nhe*I and *Xho*I sites were amplified by PCR. PCR was performed in a 20 µL reaction mixture containing 5 ng of pVD58-RbcS-TP or pVD80-RbcS-TP, 0.5 µM of primers AZ67 and AZ68 (Table 2.3b), 0.02 U/µL of Phusion Hot Start DNA Polymerase (Fermentas), 1xHF Buffer (Fermentas), 200 µM of each dNTP. PCR parameters were as follows: one cycle at 98°C for 30 s, 30 cycles at 98°C for 10 s, 71.4°C for 30 s, 72°C for 1 min, and one cycle at 72°C for 10 min. The *Nhe*I/*Xho*I-digested PCR fragments from pVD58-RbcS-TP and pVD80-RbcS-TP were cloned into *Nhe*I/*Xho*I-digested pET21b resulting in pET21b-RbcS-TP-mVirD2-58 and pET21b-RbcS-TP-mVirD2-80, respectively.

2.3.2.2. The determination of optimal conditions for the overproduction of recombinant proteins

E. coli strain BL21[pLysS] was transformed with pET3a-VirD2 (pETD2-His6) encoding wild-type VirD2 (Pelczar et al., 2004), pET21b-DCL-TP-mVirD2-58, pET21b-RbcS-TP-mVirD2-58, pET21b-DCL-TP-mVirD2-80, and pET21b-RbcS-TP-mVirD2-80. The bacterial culture was plated on solid LB supplemented with 100 mg/L carbenicillin and 35 mg/L chloramphenicol. Transformants were used to inoculate 1 mL LB containing the aforementioned antibiotics. The bacterial cultures were incubated at 37°C for 2.5 h with vigorous shaking and then transferred into 10 mL LB containing the previously mentioned antibiotics. The cultures were incubated at either at 37 or 28°C till OD_{600} between 0.45 and 0.6 was reached. The *virD2*and *TP-mvirD2* genes were induced by 1 mM IPTG, and the bacterial cultures were incubated at either 37 or 28°C for 4 to 5 h. Culture aliquots containing the same number of bacterial cells according to OD_{600} readings of non-induced and induced cultures were centrifuged and then resuspended in 120 µL of 2x SDS-PAGE loading buffer. The samples were placed in the boiling water for 5 min. Twenty four µL of samples were loaded onto SDS-PAGE. Gels were stained by Coomassie blue staining.

Western blotting was performed using a primary antibody (rabbit polyclonal anti-VirD2 antibody in 1:2,000 dilution) and a secondary antibody (Abcam, goat polyclonal antibody to rabbit IgG in 1:10,000

dilution). The detection was carried out with NBT/BCIP solution according to the manufacturer's instructions (Roche).

Molecular weights of recombinant proteins were predicted using Science Gateway (http://www.sciencegateway.org/tools/proteinmw.htm).

2.3.2.3. Overproduction and purification of wild type VirD2

E. coli BL21[pLysS] was transformed with pET3a-VirD2. The bacterial culture was plated on solid LB supplemented with 100 mg/L carbenicillin and 35 mg/L chloramphenicol. Four individual colonies of transformants were used to inoculate 100 mL LB medium containing the aforementioned antibiotics. Cells were grown at 37^{o}C for 2.5h with shaking. Twenty five mL of 100 mL culture was transferred into four 250 mL LB media with the aforementioned antibiotics. Bacteria were grown at 37^{o}C with shaking till OD_{600} between 0.45 and 0.6 was reached. The *virD2* gene was induced by 1 mM IPTG, and the bacterial cultures were incubated at 37^{o}C for 4 to 5 h. The cells were harvested at 5,000xg for 25 min at 4^{o}C and resuspended in 40 mL of pre-chilled 1xLysis buffer [50 mM Tris-HCl pH 8.0, 150 mM KCl, 5 mM $MgCl_2$, 10% glycerol, 1 tablet of Protease inhibitor, cOmplete EDTA free (Roach) per 40 mL]. Subsequently, 1 mL of Lysozyme (40 mg/mL) was added to the bacterial suspension which was then incubated on ice for 30 min with inversion every 10 min. The bacterial suspension was incubated at 37^{o}C for 5 min and subsequently placed on ice. The cells were sonicated for 90 sec. Forty µL of 1 mg/mL DNase and 40 µL of 1 mg/mL RNase were added to the suspension, and the suspension was incubated on ice for 15 min. The cell suspension was centrifuged at 17,700xg for 30 min at 4^{o}C. The supernatant (Lysate I) was collected. The pellet was resuspended in 20 mL of Buffer A (50 mL Tris-HCl pH8.0, 150 mM KCl, 5 mM $MgCl_2$, 10 mM imidazole, 10% glycerol, 8 M urea) and incubated at room temperature for 2 h with rocking. The suspension was centrifuged at 17,700xg for 30 min at room temperature.

The supernatant (Lysate II) was collected and loaded onto 5 mL Ni column (GE Healthcare) which was equilibrated with 10 column volumes of Buffer A. The column was washed with 10 column volumes of Buffer A. Proteins were eluted with 100 mL linear gradient of 10-500 mM imidazole in Buffer A. Elution factions were pooled (Pool I) and then dialysed against 2 L of Buffer B [50 mM Tris-HCl pH8.0, 150 mM KCl, 5 mM $MgCl_2$, 10% glycerol, 2 tablets protease inhibitor (cOmplete, Roche) per 1 L] at room

temperature overnight. Dialysis was again carried out against 1 L cold Buffer B at 4^{o}C for 2 h. The dialysed sample was centrifuged at 17,700xg for 30 min at 4^{o}C. The supernatant (Load II) was loaded onto 5 mL Heparin column (GE Healthcare) which was equilibrated with 20 column volumes of Buffer B. The column was washed with 10 column volumes of Buffer B. Proteins were eluted with 10 column volumes of a linear gradient of 150 mM – 1 M KCl in buffer B. The elution fractions were pooled (Pool II) and then dialyzed against 1 L of Buffer B at 4^{o}C overnight. The sample was centrifuged at 16,100xg for 95 min at 4^{o}C, and the supernatant containing purified VirD2 was collected and quantified using the Bio-Rad Protein Assay (Bio-Rad) according to the manufacturer's protocol (Bio-Rad). The samples throughout the purification steps were quantified by the aforementioned Protein Assay and analyzed on SDS-PAGE. The gel was stained by GelCode™ Blue Stain (Thermo Scientific) according to the manufacturer's protocol.

2.3.2.4. Overproduction of RbcS-TP-mVirD2-58

E. coli BL21[pLysS] was transformed with pET21b-RbcS-TP-mVirD2-58. The bacterial culture was plated on solid LB supplemented with 100 mg/L carbenicillin and 35 mg/L chloramphenicol. Twenty individual colonies of transformants were used to inoculate 500 mL LB containing the aforementioned antibiotics. The cells were grown at 37^{o}C for 2.5h with shaking. One hundred mL of 500 mL culture was transferred into four 1 L LB media with the aforementioned antibiotics. Bacteria were grown at 37^{o}C with shaking till OD_{600} between 0.45 and 0.6. The *RbcS-TP-mvirD2-58* gene was induced by 1 mM IPTG, and the cultures were incubated at 28^{o}C for 4 to 5 h. A one litre culture was centrifuged at 5,000xg for 25 min at 4^{o}C and resuspended in 40 mL of pre-chilled 1xLysis buffer [50 mM Tris-HCl pH 8.0, 150 mM KCl, 5 mM $MgCl_2$, 10% glycerol, 1 tablet of protease inhibitor (cOmplete, Mini, EDTA-free, Roche) per 40 mL]. Afterwards, 1 mL of Lysozyme (40 mg/mL) was added to each bacterial suspension which was then incubated on ice for 30 min with inversion every 10 min. The bacterial suspension was incubated at 37^{o}C for 5 min and then placed on ice. The cells were sonicated for 90 sec and then centrifuged at 4,000xg for 5 min at room temperature. The supernatant (Lysate I) was kept, and the pellet was resuspended in 40 mL of B-PER® Bacterial Protein Extraction Reagent (Thermo Scientific) with 1 mL of Lysozyme (40 mg/mL), 40 µL of DNase (5 mg/mL), and 40 µL of RNase A (10 mg/mL). The cell suspension was rocked for 5 min at room temperature and then centrifuged at 17,700xg for 30 min at 4^{o}C.

The supernatant (Lysate II) was collected. The pellet was resuspended in 50 mL of BufferA (50 mM Tris-HCl pH8.0, 150 mM KCl, 5 mM $MgCl_2$, 10 mM imidazole, 10% glycerol, 8 M urea). The suspension was centrifuged at 17,700xg for 30 min at room temperature. The supernatant (Lysate III) was collected and loaded onto 5 mL Ni column (His-Trap HP, GE Healthcare) which was equilibrated with 20 column volumes of Buffer A. The column was washed with 10 column volumes of Buffer A and 10 column volumes of buffer A containing 40 mM imidazole. The proteins were eluted with 10 column volumes of a linear gradient of 40-300 mM imidazole in buffer A. The elution fractions were pooled (Pool I) and then dialysed against 2 L of Buffer C [50 mM Tris-HCl pH8.0, 20 mM KCl, 10% glycerol, 2 tablets protease inhibitor (cOmplete, Roche) per 1 L] at room temperature overnight. Dialysis was again carried out against 1 L cold Buffer C at 4^{o}C for 2 h. The dialysed sample was centrifuged at 17,700xg for 30 min at 4^{o}C. The supernatant (Pool II) was loaded onto 1 mL Q-Sepharose column (HiTra Q HP, GE Healthcare), pre-equilibrated with 10 column volumes of buffer C. The column was washed with 10 column volumes of Buffer C. The proteins were eluted with 10 column volumes of a linear gradient of 20 mM – 1 M NaCl in Buffer C. The elution factions were pooled (Pool III) and then dialyzed against 1 L of Buffer B at 4^{o}C overnight. The sample was centrifuged at 16,100xg for 30 min at 4^{o}C. The supernatant containing purified RbcS-TP-mVirD2-58 was collected. The purified protein was concentrated using ULTRAFREE® 0.5 Centrifugal Filter Device (MILLIPORE) according to the manufacturer's instructions and quantified using the Bio-Rad Protein Assay (Bio-Rad) according to the manufacturer's protocol (Bio-Rad). The samples throughout the purification steps were quantified by the aforementioned Protein Assay and analyzed on SDS-PAGE. The gel was stained by GelCode™ Blue Stain (Thermo Scientific) according to the manufacturer's protocol. The percent impurity was estimated using the Image J program (NIH, http://rsbweb.nih.gov/ij/).

2.3.2.5. Cleavage activity assay

The cleavage activity assay was performed using various amounts of purified VirD2 and RbcS-TP-VirD2-58 (0, 0.04, 0.08, 0.16, 0.3, 0.45, 0.6, and 0.9 µg) as described previously with the following modifications (Ziemienowicz et al., 1999). Recombinant proteins were incubated with 0.5 pmol of fluorescently labelled oligo AZ3 (Table 2.3b) in TKM buffer (50 mM Tris-HCl pH 8.0, 150 mM KCl, 1

mM $MgCl_2$) at 37°C for 1 h. The reactions were terminated by the addition of the stop solution (30% formamide, 1xDNA gel loading buffer, 10 mM EDTA pH 8.0) and incubated at 95°C for 3 min. Ten µL of the reaction were loaded on a 20% polyacrylamide and 8 M urea denaturing gel which was pre-run at 750 V for 30 min. The gel was covered with aluminum foil to protect the fluorescently labelled oligonucleotides during electrophoresis in 1xTBE buffer at 600 V for 1h and then 750 for 1 h. Fluorography was performed using a Typhoon system (GE Healthcare). The efficiency of the cleavage activity was calculated using the ratio between the cleavage product and the uncleaved substrate by the Image J program (NIH, http://rsbweb.nih.gov/ij/).

2.3.3. *Agrobacterium*-mediated chloroplast transformation by tobacco leaf incision

2.3.3.1. The binary vector system for *Agrobacterium*-mediated plant transformation

2.3.3.1.1. Construction of pAVD plasmids carrying *TP-mvirD2*

pVD58-DCL-TP, pVD58-RbcS-TP, pVD80-DCL-TP, and pVD80-RbcS-TP were digested with *Psi*I and *Eco*RI. The digestion product (1.5 kb) containing the*TP-mvirD2* fragment was used to replace the *virD2*gene in pAVD44 resulting in pAVD58-DCL-TP, pAVD58-RbcS-TP, pAVD80-DCL-TP, and pAVD80-RbcS-TP, respectively.

2.3.3.1.2. Generation of *Agrobacterium* carrying helper plasmids containing *TP-mvirD2*

pAVD derivatives were introduced into *Agrobacterium* strain GV3101[pPM6000K] (Rossi et al., 1993) by either freeze-thaw method (Hofgen and Willmitzer, 1988) or electroporation (Mattanovich et al., 1989). The bacterial cultures were plated on YEP medium (10 g peptone, 10 g yeast extract, 5 g $NaCl_2$, 10 mM $MgSO_4$, 20 g agar per litre, pH 7) containing rifampicin (100 mg/L), carbenicillin (25 mg/L), and kanamycin (50 mg/L). Colony PCR was performed to confirm the generation of GV3101[pPM6000AM] derivatives with following conditions: in a 20 µL reaction mixture containing a colony, 0.5 µM of primers AZ48 and AZ173 for GV3101[pPM6000AM2 or 4] and AZ182 and AZ173 for GV3101[pPM6000AM3 or 5] (Table 2.3b), 0.125 U/µL of Taq DNA Polymerase (Fermentas), 1x Taq Buffer (Fermentas), 200 µM of each dNTP. PCR parameters were as follows: one cycle at 95°C for 3 min, 40 cycles at 95°C for 30 s, 65°C

for GV3101[pPM6000AM2], 70°C for GV3101[pPM6000AM4], and 55°C for GV3101[pPM6000AM3 or 5] for 30 s, 72°C for 90 s, and one cycle at 72°C for 10 min.

2.3.3.1.3. Construction of a binary vector carrying the chloroplast expression cassette

A chloroplast transformation vector (pLD-CtV) was kindly provided by Dr. Daniell (University of Pennsylvania, Philadelphia, PA). The chloroplast expression cassette (*trnI*-Prrn-*aadA*-TpsbA-*trnA*) in pLD-CtV was amplified and flanked by *Pac*I and *Hind*III by PCR. PCR was performed in a 20 µL reaction mixture containing 10 µg of pLD-CtV, 0.5 µM of primers AZ269 and AZ270 (Table 2.3b), 0.02 U/µL of Phusion DNA Polymerase (Fermentas), 1xHF Buffer (Fermentas), 200 µM of each dNTP. PCR parameters were as follows: one cycle at 98°C for 30 s, 30 cycles at 98°C for 10 s, 55°C for 30 s, 72°C for 1 min, and one cycle at 72°C for 10 min. The spectinomycin resistance gene for bacterial selection in pPZP500 (Ali et al., 2012) was replaced with the chloramphenicol gene prior to this study. A region of the modified pPZP500 excluding the kanamycin resistance gene was amplified and flanked by *Hind*III and *Pac*I by PCR. PCR was performed in a 20 µL reaction mixture containing 10 µg of pPZP500, 0.5 µM of primers AZ273 and AZ274 (Table 2.3b), 0.02 U/µL of Phusion DNA Polymerase (Fermentas), 1xHF Buffer (Fermentas), 200 µM of each dNTP. PCR parameters were as follows: one cycle at 98°C for 30 s, 30 cycles at 98°C for 10 s, 65°C for 30 s, 72°C for 2 min, and one cycle at 72°C for 10 min. PCR fragment containing the chloroplast expression cassette and pPZP500 PCR product were digested with *Hind*III and *Pac*I, then they were ligated together resulting in pPZP500-aadA (Table 2.3a).

2.3.3.2. Growth conditions of *Nicotiana tabacum* plants

Seeds of the *N. tabacum* cultivar Big Havana were surface-sterilized with 10% sodium hypochloride for 10 min and then rinsed three times with distilled water for 5 min. The seeds were stored in distilled water at 4°C for 2 days. Subsequently, two seeds were placed in a glass jar (diameter 5 cm, height 10 cm) with solid MS medium (4.43 g MS with vitamins Phytotechnology) supplemented with 30 g of sucrose per litre, pH 5.3-5.6, and transferred to a growth chamber (Enconair, Winnipeg, MB, Canada); the light was provided by Octron T8 Fluorescent bulbs (Sylvania, Mississauga, ON, Canada). Upon

germination, one seedling was removed from the jar for growing (one plant per jar). After the plants grew to a sufficient size (about 2-3 month later), their leaves were used for *Agrobacterium*-mediated transformation.

2.3.3.3. *Agrobacterium*-mediated chloroplast transformation by tobacco leaf incision

Agrobacterium strain GV3101[pPM6000AM::pPZP500-aadA] series were used to inoculate 2 mL YEP medium containing rifampicin (100 mg/L), carbenicillin (25 mg/L), kanamycin (50 mg/L), and chloramphenicol (35 mg/L). The cultures were shaken at 28^{o}C overnight. One mL of the overnight culture was transferred into 100 mL YEP medium containing the appropriate antibiotics and incubated at 28^{o}C overnight with vigorous shaking. The bacterial cultures were centrifuged at 3,200xg for 20 min at room temperature. The bacterial pellet was resuspended in 25 mL of MS medium (Murashige and Skoog, 1962) by gentle vortex. The centrifugation and resuspension were repeated. After OD_{6oo} was adjusted to 0.5 using MS medium, acetosyringone (a final concentration of 100 µM) was added to the bacterial suspensions that were then incubated for 30 to 60 min at room temperature.

Tobacco leaves were submerged into the bacterial suspension in a Petri dish. Incisions were made by a scalpel on the adaxial side of leaves parallel to side veins and avoided on the midrib and leaf margins. The leaves were submerged in the bacterial suspension for 10 min and then were dried on a sterilized paper towel. The leaves were placed on solid MS medium with the abaxial side up and stored at 22^{o}C in the dark for co-cultivation. The leaves were well rinsed with distilled water and blotted dry. Then they were transferred onto the shoot-inducing medium (SIM) (4.43 g MS with vitamins (Phytotechnology), 30 g sucrose, 200 mg myo-inositol, 0.8 mg indole-3-acetic acid, 2 mg kinetin, 8 g agar per litre, pH 5.7 with NaOH) for callus and shoot regeneration with cefotaxime or timentin and spectinomycin in two trials.

In all trials, the leaves were transferred to fresh SIM with cefotaxime or timentin and spectinomycin every 3 - 4 weeks. The developed shoots were transferred to the root-inducing medium (4.43 g MS with vitamins Phytotechnology, 30 g sucrose, 200 mg myo-inositol, 0.5 mg α-naphthalene-acetic acid, 8 g agar per litre, pH 5.7 with NaOH) with the aforementioned antibiotics. After sufficient root development, the plants were transferred to soil.

2.3.3.3.1. The first trial

In the first trial, 100 mg/L cefotaxime, 100 mg/L timentin, and 500 mg/L spectinomycin were used in SIM. For the regeneration of negative controls, the untransformed leaves were placed on the aforementioned plates to examine the amount of spectinomycin that would prevent the regeneration of untransformed calli/shoots.

2.3.3.3.2. The second trial

In the second trial, four different treatments were conducted with leaves after co-cultivation: 1) the leaves were placed on SIM with 100 mg/L timentin, 2) the leaves were placed on SIM with 100 mg/L timentin, and then the regenerated shoots were later transferred onto SIM with 100 mg/L timentin and 100 mg/L spectinomycin, 3) the leaves were placed on SIM with 100 mg/L timentin and 100 mg/L spectinomycin, 4) the leaves were placed on SIM with 100 mg/L timentin and 250 mg/L spectinomycin. For the regeneration of negative controls, the untransformed leaves were placed on SIM containing 100 mg/L timentin and 100 mg/L spectinomycin, and SIM containing 100 mg/L timentin and 250 mg/L spectinomycin.

2.3.3.3.3. Transgene detection by PCR

Plant total DNA was isolated from 100 mg leaf samples using the cetyltrimethylammonium bromide method (Doyle and Dickson, 1987) with modifications according to Molecular Cytogenetics Research Group protocols (http://www.le.ac.uk/bl/phh4/dnaiso.htm). DNA was quantified using NanoDrop (Thermo Scientific).

In the second trial, PCR was performed in a 20 µL reaction mixture containing in average 250 ng of total DNA, 0.5 µM of primers AZ187 and AZ200, 0.125 U/µL of Taq DNA Polymerase (Fermentas), 1xFermentas Buffer (Fermentas), 200 µM of each dNTP. PCR parameters were as follows: one cycle at 95^{o}C for 3 min, 30 cycles at 95^{o}C for 1 min, 55^{o}C for 30 s, 72^{o}C for 3 min, and one cycle at 72^{o}C for 10 min.

Table 2.3a. *Agrobacterium* strains and plasmids used in this work.

Name	Characteristics	Reference
***Agrobacterium* strains**		
GV3101[pPM6000]	Cured C58 nopaline strain containing pPM6000 which is a derivative of pTiAch5 (octopine plasmid) and has no T-DNA	(Bonnard et al., 1989)
GV3101[pPM6000K]	Same as above but containing *virD2* deleted by *Kpn*I	(Rossi et al., 1993)
GV3101[pPM6000AM2]	Same as above but carrying pAVD58-DCL-TP in pPM6000K	This study
GV3101[pPM6000AM3]	Same as above but carrying pAVD58-RbcS-TP in pPM6000K	This study
GV3101[pPM6000AM4]	Same as above but carrying pAVD80-DCL-TP in pPM6000K	This study
GV3101[pPM6000AM5]	Same as above but carrying pAVD80-RbcS-TP in pPM6000K	This study
pVD derivatives	Contains the ampicillin resistance gene	
pVD43	pUC18 containing *virD2* under the control of the promoter and the Shine-Dalgarno sequence of *virD1*	(Rossi et al., 1993)
pVD44	Same as above but carrying *Nru*I deletion of the C-terminal NLS in *virD2*	(Rossi et al., 1993) Reconstructed in this study
pVD58	Same as pVD44 but carrying a point mutation in the N-terminal NLS	(Rossi et al., 1993)
pVD58-*Age*I	Same as pVD58 but carrying the *Age*I site before the *virD2* start codon	This study
pVD70	Same as pVD43 but carrying a precise deletion of the C-terminal NLS	(Rossi et al., 1993)
pVD80	Same as pVD70 but carrying a point mutation in the N-terminal NLS	(Rossi et al., 1993) Reconstructed in this study
pVD80-*Age*I	Same as pVD80 but carrying the *Age*I site before the *virD2* start codon	This study
pVD-TP derivatives		
pVD58-DCL-TP	Same as pVD58-*Age*I but carrying the *DCL-TP* sequence	This study
pVD58-RbcS-TP	Same as pVD58-*Age*I but carrying the *RbcS-TP* sequence	This study
pVD80-DCL-TP	Same as pVD80-*Age*I but carrying the *DCL-TP* sequence	This study
pVD80-RbcS-TP	Same as pVD80-*Age*I but carrying the *RbcS-TP* sequence	This study

Table 2.3a. Continued.

Name	Characteristics	Reference
pAVD derivatives	pARO 181, pUC derivative containing the kanamycin resistance gene and the *virD* promoter and various *virD2* genes	
pAVD44	Same as above but containing the *virD2* gene with *Nru*I deletion	(Daniell et al., 1998; Rossi et al., 1993)
pAVD58-DCL-TP	Same as above but containing the *DCL-TP-mvirD2-58* gene	This study
pAVD58-RbcS-TP	Same as above but containing the*RbcS-TP-mvirD2-58* gene	This study
pAVD80-DCL-TP	Same as above but containing the *DCL-TP-mvirD2-80*gene	This study
pAVD80-RbcS-TP	Same as above but containing the*RbcS-TP-mvirD2-80* gene	This study
Chloroplast expression vector		
pLD-CtV	The chloroplast expression vector carrying the chloroplast expression cassette: *aadA* driven by Prrn and TpsbA flanked by *trnI* and *trnA*	(Daniell et al., 1998)
Binary vectors		
pPZP500	A binary vector containing the chloramphenicol resistance gene* as a bacterial selection marker and the kanamycin resistance gene as a plant selection marker	(Ali et al., 2012) NCBI (Accession: JX025644)
pPZP500-aadA	A binary vector carrying the chloramphenicol resistance gene as a bacterial selection marker and the T-DNA containing the chloroplast expression cassette from pLD-CtV	This work
Protein expression vectors		
pET3a-VirD2 (pETD2-His6)	pET3a containing the *virD2-His* tag coding sequence	(Pelczar et al., 2004)
pET21b-DCL-TP-VirD2-58	pET21b containing the *DCL-TP-mvirD2-58-His* tag coding sequence	This work
pET21b-RbcS-TP-VirD2-58	pET21b containing the *RbcS-TP-mvirD2-58-His* tag coding sequence	This work
pET21b-DCL-TP-VirD2-80	pET21b containing the *DCL-TP-mvirD2-80-His* tag coding sequence	This work
pET21b-RbcS-TP-VirD2-80	pET21b containing the *RbcS-TP-mvirD2-80-His* tag gene	This work

* The spectinomycin resistance gene for bacterial selection in pPZP500 was replaced with the chloramphenicol gene prior to this study.

Table 2.3b. Primers used to construct vectors containing *mvirD2* or*TP-mvirD2*.

Primer	Sequence (5'-3')	Description
AZ3	CGGTATATATCCTG^CCAGTCAGGTCGACA GTTGC	34-mer, a substrate for VirD2 activity, 34-mer+VirD2=20-mer-VirD2+14-mer (^, cleavage site; underlined, the right border sequence)
AZ40	AGCAGCAGTTAGTTTTCAGCAC	The forward primer to amplify a 419 bp fragment coding for tomato DCL transit peptide and 6 amino acids of the mature protein (Gnanasambandam et al., 2007)
AZ41	CATCCGATTCCTCCTTCACCAA	The reverse primer to amplify a 419 bp fragment coding for tomato DCL transit peptide (Gnanasambandam et al., 2007)
AZ44	GAACG*CTGCAG*TTCCAGCTTTCCCTTCGTGG ACAGGTACTCC	The reverse primer to modify N-terminal NLS (R→T) of VirD2 and flank it by the *Pst*I site
AZ45	GGGCTAGACTTTAATCTGTCTA*CCCGGG*CA CATTTCGTGC	The forward primer to amplify a ~180 bp fragment containing the N-terminal NLS coding sequence and to introduce the *Sma*I site
AZ47	CTA*CCCGGG*CACATTTCGTGCTGGAGTATT CAGACCCTTCCGCTTTTTTTGGAGGAAGCT ATG*ACCGGT*CCCGATCGCGC	The forward primer to amplify ~180 bp fragment and to introduce the *Age*I site after ATG codon of *virD2, Sma*I site
AZ48	CAG*ACCGGT*ATGGCTTCAATTTGTACTTCAA	The forward primer to amplify sequence coding 50 amino acids of tomato DCL transit peptide + 6 amino acids of the mature protein (Gnanasambandam et al., 2007) and to introduce the *Age*I site
AZ50	CCAA*ACCGGT*CCCTCCTCCCTCTGAACC	The reverse primer to amplify sequence coding 50 amino acids of tomato DCL transit peptide + 6aa of the mature protein and to introduce the *Age*I site
AZ51	GAT*ACCGGT*ATGGCTTCCTCAGTTCTTTCCT CTG	The forward primer to amplify ~200 bp fragment coding for tobacco RbcS transit peptide (57 amino acids) + 5 amino acids of the mature protein and to introduce the *Age*I site
AZ53	CCG*ACCGGT*TGGCCACACCTGCAT	The reverse primer to amplify ~200 bp fragment coding for tobacco RbcS transit peptide (57 amino acids) + 5 amino acids of the mature protein and to introduce the *Age*I site
AZ61	CCACATATG*GCTAGC*ATGGCTTCAATTTGTA CTTCAAATTTTCAC	Forward primer to amplify *DCL-TP-mvirD2* and to introduce the *Nhe*I site
AZ67	CCA*CATATGGCTAGC*ATGGCTTCCTCAGTTC TTTCCTCTG	The forward primer to amplify *RbcS-TP-mvirD2* and to introduce the *Xho*I site
AZ68	CCG*CTCGAG*GGTCCCCCCGCGCCCATCG	The reverse primer to amplify *virD2* gene without STOP codon and to introduce the *Xho*I cloning site
AZ173	CTAGGTCCCCCCGCGC	The reverse primer which anneals to the end of the *virD2* gene
AZ182	ATGGCTTCCTCAGTTCTTTCCTCTGC	The forward primer which anneals to the beginning of *RbcS-TP* coding sequence

Table 2.3b. Continued.

Primer	Sequence (5'-3')	Description
AZ187	GGGCTATTAGCTCAGTGGTAGAGCG	The forward primer which anneals to *trnI* of the chloroplast genome
AZ200	CGCCGACTCCAACTACCGTC	The reverse primer which anneals to *trnA* of the chloroplast genome
AZ269	TTT*TTAATT*AAGGCCTCTTCGCTATTACGCC AG	The forward primer to amplify the chloroplast expression cassette in pLD-CtV and to introduce *Pac*I
AZ270	CGAT*AAGCTT*TGTATCGGCTAAGTTCACG	The reverse primer to amplify the chloroplast expression cassette from *Hind*III in pLD-CtV
AZ273	TTT*TTAATTAA*TGTTTACACCACAATATATCC	The forward primer to amplify the region excluding T-DNA in pPZP500 and to introduce *Pac*I
AZ274	ATA*TAAGCTT*GTAAACCTAAGAGAAAAGAG CG	The reverse primer to amplify the region excluding T-DNA in pZP500 and to introduce *Hind*III

The introduced restriction sites are indicated in *italics*; ^ indicates the VirD2 cleavage site.

2.4. Results

2.4.1. Construction of TP-mVirD2 vectors

Modifications of VirD2 to inactivate the N- and C-terminal NLSs were introduced according to a previous study (Rossi et al., 1993) using pVD43 as one of the initial plasmids (Figure 2.4.1a). Arginine in the N-terminal NLS was changed to threonine; additionally, two types of the C-terminal NLS deletion were selected: 1) a broad deletion of NLS (eliminating 45 amino acids including the C-terminal NLS by *Nru*I), and 2) a precise deletion of NLS (exclusively eliminating bipartite sequences) (Figure 2.4.2a). Initial plasmids: pAVD44 encoding a board deletion of the C-terminal NLS of VirD2, pVD43 encoding a wild-type VirD2, pVD80 encoding a precise deletion of the C-terminal NLS (Table 2.3a), and the PCR product containing the modification in the N-terminal NLS coding sequence were used to generate plasmids carrying the *mvirD2* gene encodingVirD2 with the inactivated N and C-terminal NLSs. As a result, two plasmids containing the NLS sequences with modifications were generated: 1) pVD58-AgeI encoding the modified N-terminal NLS, a board deletion of the C-terminal NLS, and the *Age*I site and 2) pVD80-AgeI encoding the modified N-terminal NLS, a precise deletion of the C-terminal NLS and the *Age*I site (Table 2.3a).

Tomato DCL and tobacco RbcS transit peptides were chosen due to their strong chloroplast targeting activity (Gnanasambandam et al., 2007). Tomato DCL-TP and tobacco RbcS-TP with short mature protein coding sequences were amplified by PCR using gDNA and cDNA, respectively. The restriction-digested PCR product encoding 50 amino acids of tomato DCL transit peptide and one amino acid of the mature protein were inserted into the beginning of the *mvirD2* genes in pVD58 and pVD80 derivatives, resulting in pVD58-DCL-TP and pVD80-DCL-TP, respectively (Table 2.3a). Similarly, the restriction-digested PCR product coding for 57 amino acids of tobacco RbcS transit peptide and five amino acids of the mature protein were inserted into the beginning of the modified *virD2* genes in pVD58 and pVD80 derivatives, resulting in pVD58-RbcS-TP and pVD80-RbcS-TP, respectively (Table 2.3a). As a result, these vectors encoding four different variants of the modified VirD2 protein were generated (Figure 2.4.2a and Figure 2.4.1b-e). The sequencing confirmed that all plasmids were successfully constructed. These four vectors were used to construct all expression vectors as well as helper plasmids.

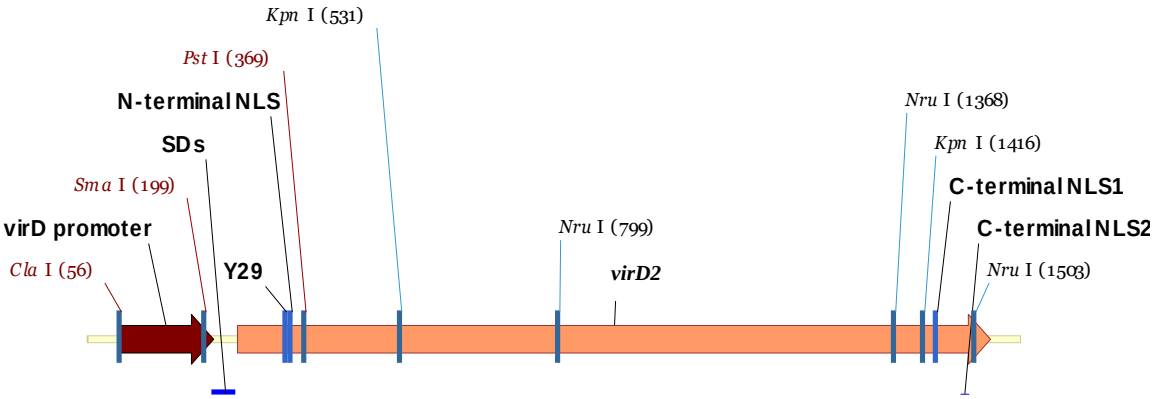


Figure 2.4.1a. Schematic representation of the virD2 gene in pVD43.
SD: the Shine-Dalgarno sequence; C-terminal NLS1 and 2: C-terminal bipartite NLS; some restriction sites are shown.

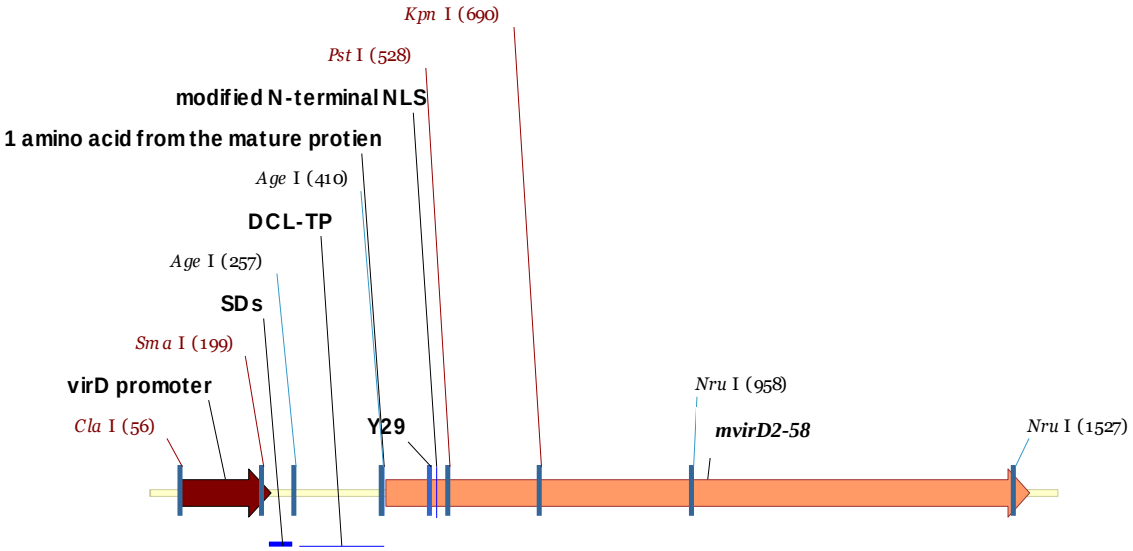


Figure 2.4.1b. Schematic representation of the *DCL-TP-mvirD2-58* gene in pVD58-DCl-TP.
SD: the Shine-Dalgarno sequence; some restriction sites are shown.

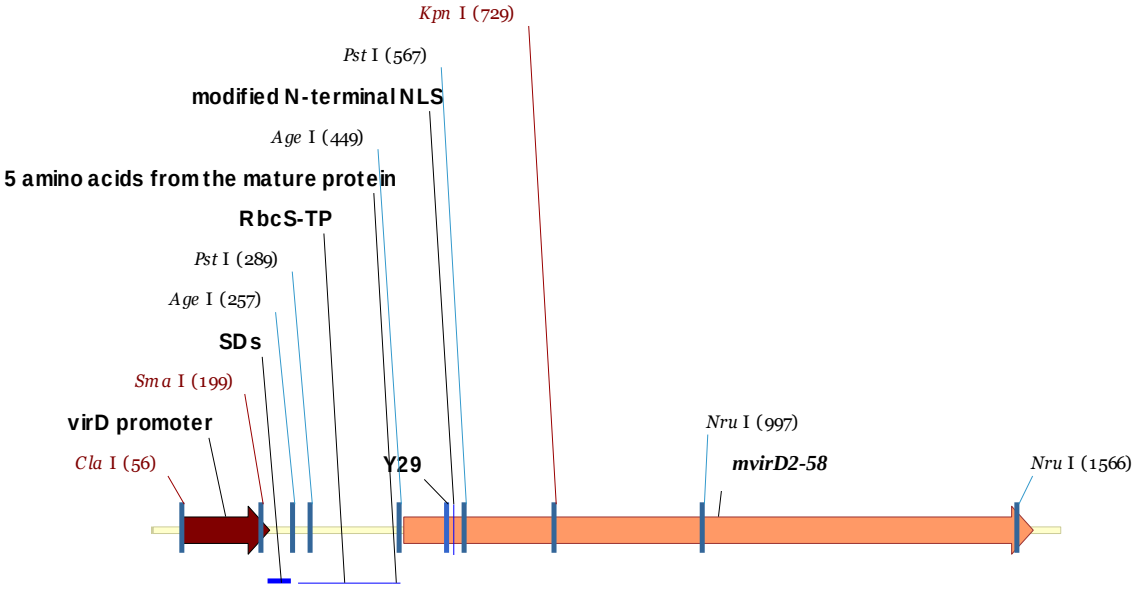


Figure 2.4.1c. Schematic representation of the *RbcS-TP-mvirD2-58* gene in pVD58-RbcS-TP.
SD: the Shine-Dalgarno sequence; some restriction sites are shown.

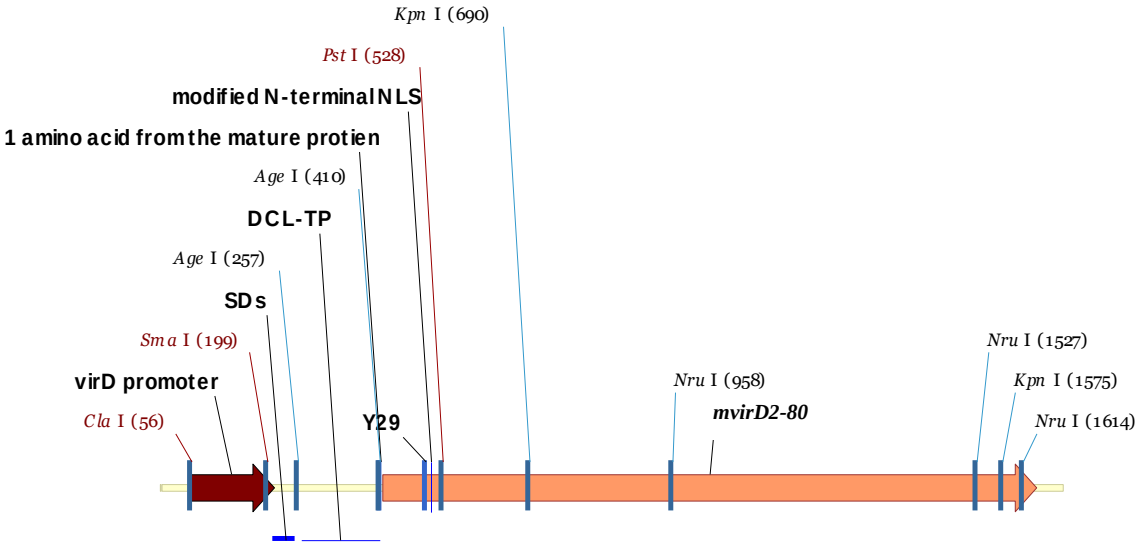


Figure 2.4.1d. Schematic representation of the *DCL-TP-mvirD2-80* gene in pVD80-DCl-TP.
SD: the Shine-Dalgarno sequence; some restriction sites are shown.

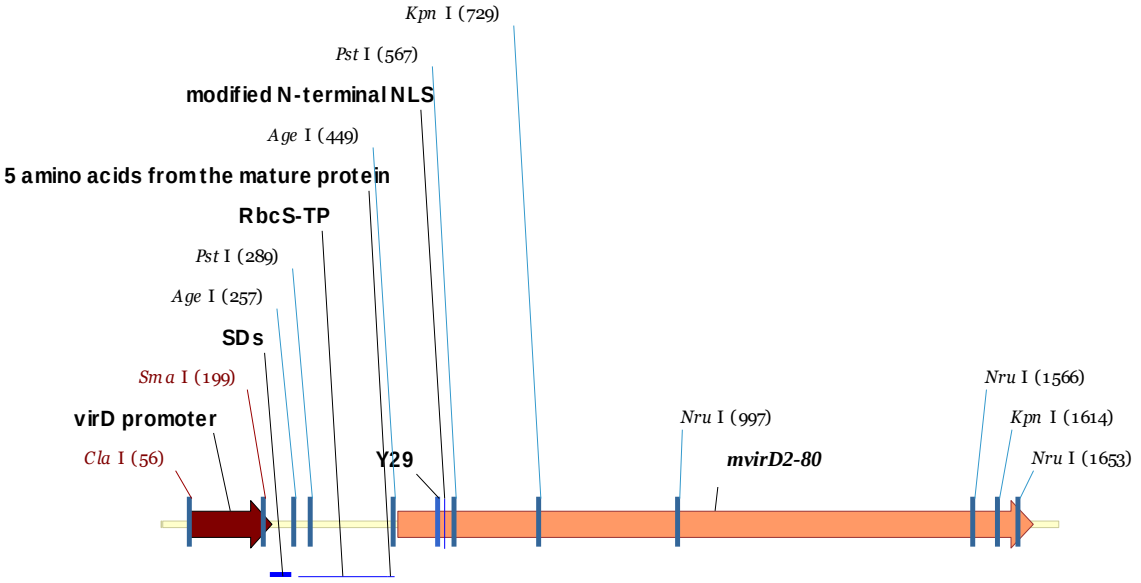


Figure 2.4.1e. Schematic representation of the *RbcS-TP-mvirD2-80* gene in pVD80-RbcS-TP.
SD: the Shine-Dalgarno sequence; some restriction sites are shown.

2.4.2. Overproduction of VirD2 and TP-mVirD2

The previously constructed pET3a-VirD2 encoding the wild-type VirD2 fused to His-tag was also used in this study (Pelczar et al., 2004). The *TP-mvirD2* genes of pVD58-DCL-TP, pVD58-RbcS-TP, pVD80-DCL-TP, and pVD80-RbcS-TP were cloned into an expression vector, pET21b, to fuse the *TP-mvirD2* with the His-tag coding sequence. As a result, all the TP-mVirD2 recombinant proteins illustrated in Figure 2.4.2a contain His-tag at the C-terminus. The presence of His-tag makes it easy to purify fusion proteins by the nickel affinity chromatography. At first, an optimal condition for the overproduction of the recombinant proteins in *E. coli* BL21[pLysS] carrying pET3a-VirD2 or pET21-TP-mVirD2 derivatives were examined. The bacterial cultures were incubated at 28 and 37^{o}C, and then *virD2* and *TP-mvirD2* were induced by IPTG when the suggested OD_{600} readings for the pET system by the manufacturer (Novagen) were reached. The non-induced and induced cultures were analyzed on SDS-PAGE and western blotting using anti-VirD2 antibody.

Polyacrylamide gels stained with Coomassie blue staining showed strong bands corresponding to 56 kDa and 55 kDa from the cultures grown at 37^{o}C with pET3a-VirD2 and pET21b-RbcS-TP-VirD2-58, respectively, whereas distinct bands were not observed from the cultures with any other plasmids (Figure 2.4.2b). However, western blot analysis using anti-VirD2 antibodies detected low amounts of the DCL-TP-mVirD2-58, DCL-TP-mVirD2-80 and RbcS-TP-mVirD2-80 proteins (Figure 2.4.2c). These amounts were unsatisfactory for efficient production of recombinant proteins. Interestingly, recombinant proteins from all plasmids were detected on the western blot at slightly larger size than the predicted protein weight (Figure 2.4.2c and Table 2.4.2).

In order to improve the production of the recombinant VirD2 proteins, a lower culture temperature (28^{o}C versus 37^{o}C) was tested. A large amount of wild type VirD2 from pET3a-VirD2 was produced as it was visible from the strongest bands at 56 kDa from 37 and 28^{o}C, and smaller proteins detected in western blot were likely from degraded VirD2 (Figure 2.4.2c). RbcS-TP-mVirD2-58 from pET21b-RbcS-TP-mVirD2-58 was also detected in a higher amount at 55 kDa as compared to other TP-mVirD2 (Figure 2.4.2c). The production of DCL-TP-mVirD2-58 remained unaffected by a lower culture temperature, and the production of DCL-TP-mVirD2-80 and RbcS-TP-mVirD2-80 was only slightly induced at 28C. Since the large amounts of VirD2 from both 37^{o}C and 28^{o}C cultures and RbcS-TP-mVirD2-58 from 28^{o}C culture

were detected on the Coomassie blue stained polyacrylamide gels and the western blot, only these two recombinant proteins were purified.

VirD2 forms inclusion bodies when it is overproduced in bacteria (Jasper et al., 1994). Therefore, inclusion bodies of VirD2 were solubilised in urea after cell lysis. Two-step affinity chromatography purification procedure was used to isolate recombinant wt VirD2: Ni column chromatography followed by protein renaturation by dialysis and then Heparin chromatography. During the purification process, five samples were collected: 1) Lysate I after the lysis of bacterial cells, 2) Lysate II after the denaturation of proteins in urea which was loaded onto the nickel column, 3) PoolI, the collection of elution factions from the Ni column, 4) Load II, a loading sample for the Heparin column after the renaturation by dialysis, and 5) Pool II, the collection of elution factions from the Heparin column. All samples were analyzed on SDS-PAGE (Figure 2.4.2d). VirD2 was purified successfully without any visible impurities on the Coomassie blue stained polyacrylamide gel (Figure 2.4.2d Left).

Rbc-TP-mVirD2-58 was also solubilised in urea after cell lysis. The purification procedure included affinity chromatography using the Ni column followed by protein renaturation by dialysis and then ion-exchange chromatography using the Q-sepharose column. Finally, the purified protein was concentrated by a centrifugal filter device. During the purification process, five samples were collected: 1) Lysate II after the lysis of bacterial cells and denaturation of protein, 2) Lysate III, a loading sample for the nickel column, 3) Pool I, the collection of elution factions from the Ni column, 4) Pool II, a loading sample for the Q-Sepharose column after the renaturation by dialysis, and 5) Pool III, the collection of elution factions from the Q-Sepharose column. All samples were analyzed on SDS-PAG (Figure 2.4.2d). A protein impurity with the size of about 45 kDa was detected in the purified RbcS-TP-mVirD2-58 protein (Figure 2.4.2d Right), and the percentage of this impurity was estimated as 20% according to the Image J analysis. Since a 80% purity of RbcS-TP-mVirD2-58 was obtained, further purification was not performed. The purified recombinant proteins were used for the cleavage activity assay.

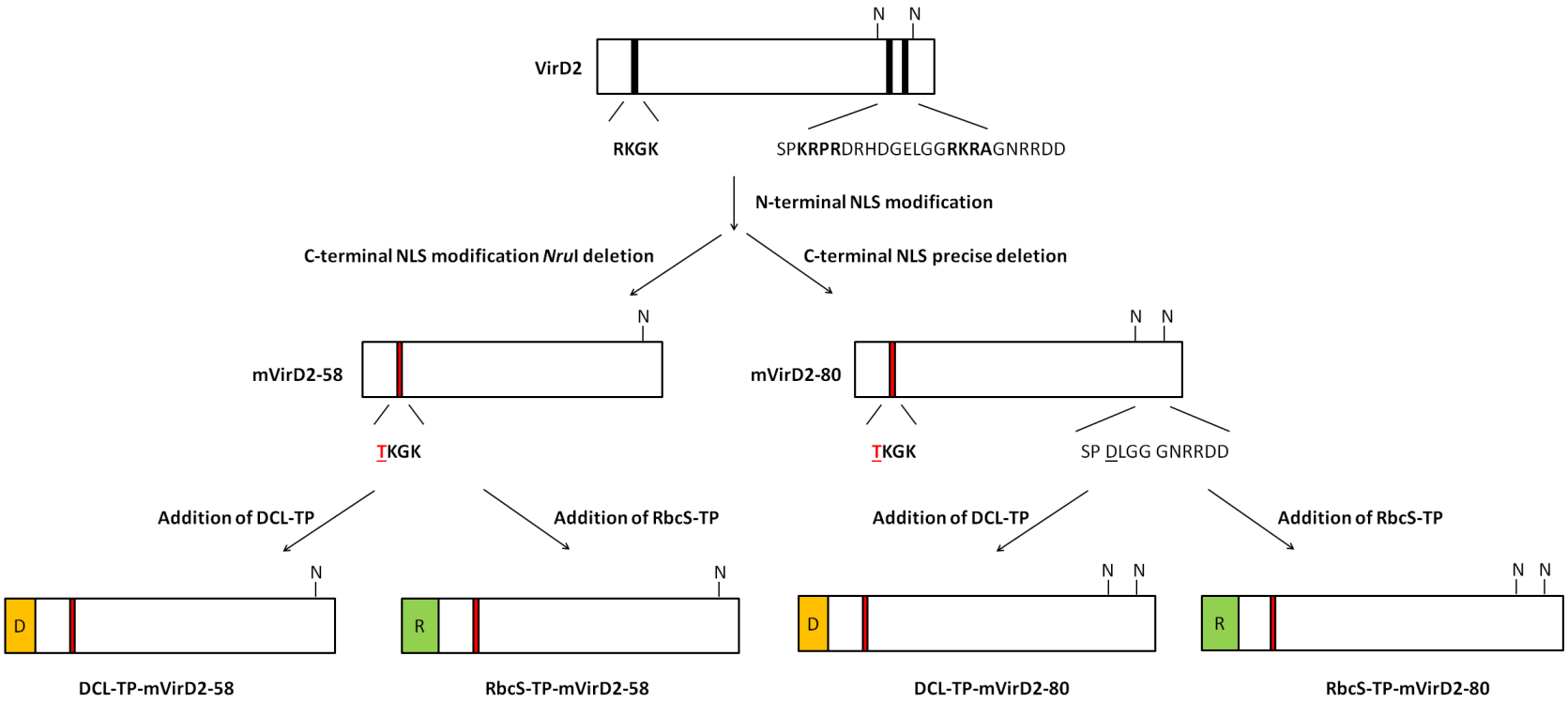


Figure 2.4.2a. Construction scheme of TP-mVirD2.

NLS are indicated in bold letter, and a modified amino acid in the N-terminal NLS is in red. In a precise deletion of the C-terminal NLS, one amino acid change (E→D) underlined was made previsouly, and it was found to have no influence at all (Rossi et al., 1993) (N:*Nru*I; D: DCL-TP with one amino acid from the mature protein; R: RbcS-TP with five amino acids from the mature protein).

Table 2.4.2. Description of recombinant proteins illustrating differences in the predicted protein weight and mobility in SDS-PAGE.

Plasmid	Recombinant protein	Amino acid length	Predicted protein weight (kDa)	Mobility in SDS-PAGE (kDa)
pET3a-VirD2 (pETD2-His6)	VirD2	430	48	56
pET21b-DCL-TP-mVirD2-58	DCL-TP-mVirD2-58	437	49	53
pET21b-RbcS-TP-mVirD2-58	RbcS-TP-mVirD2-58	450	50	55
pET21b-DCL-TP-mVirD2-80	DCL-TP-mVirD2-80	466	52	56
pET21b-RbcS-TP-mVirD2-80	RbcS-TP-mVirD2-80	479	53	56

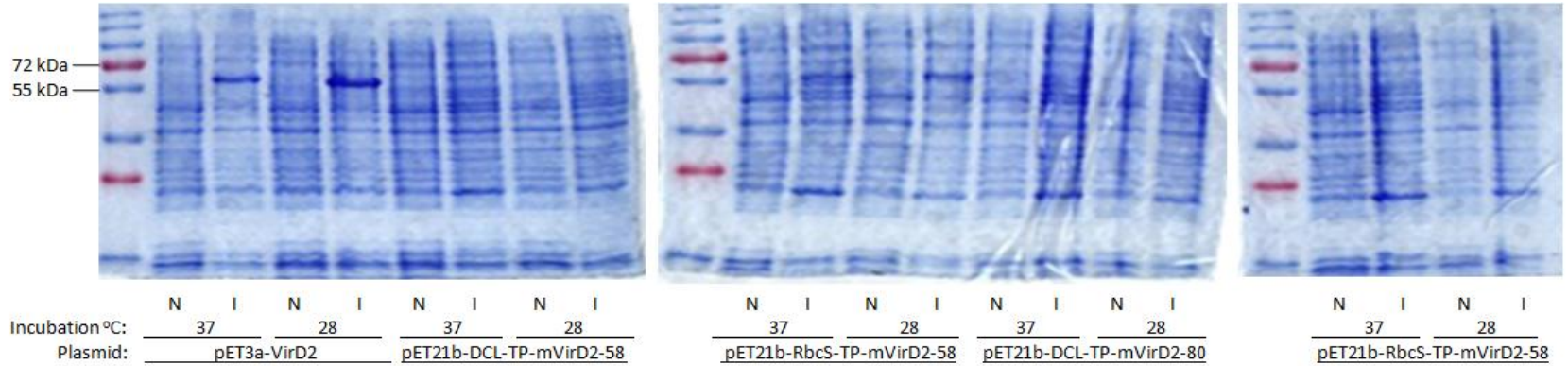


Figure 2.4.2b. Coomassie blue stained polyacrylamide gels containing non-induced (N) and induced (I) cultures of BL21b[pLysS] carrying expression vectors incubated at 37 or 28°C.

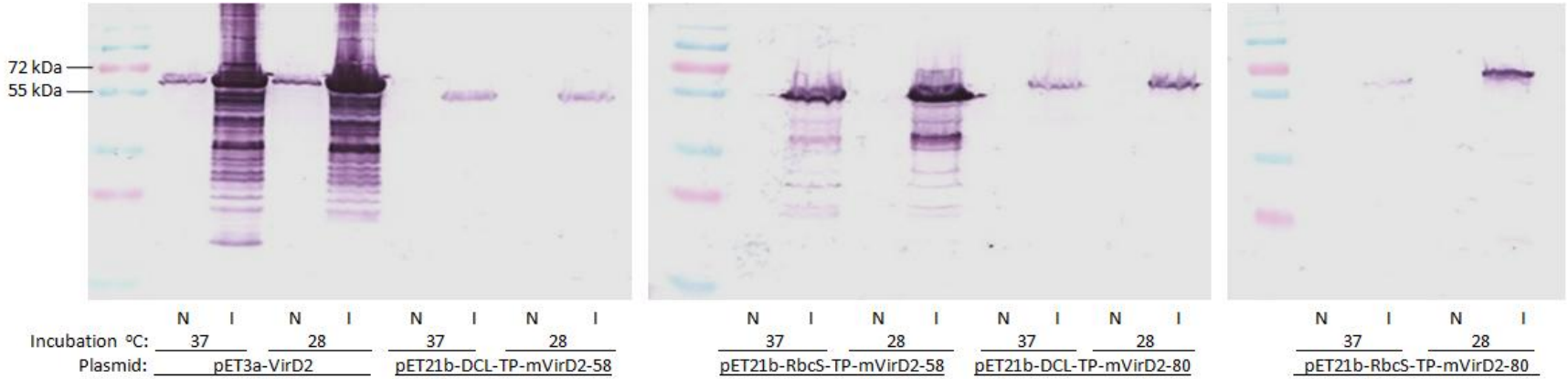


Figure 2.4.2c. Western blot detection of VirD2 and TP-mVirD2 series in non-induced (N) and induced (I) cultures of BL21b[pLysS] carrying expression vectors probed with anti-VirD2 antibody.

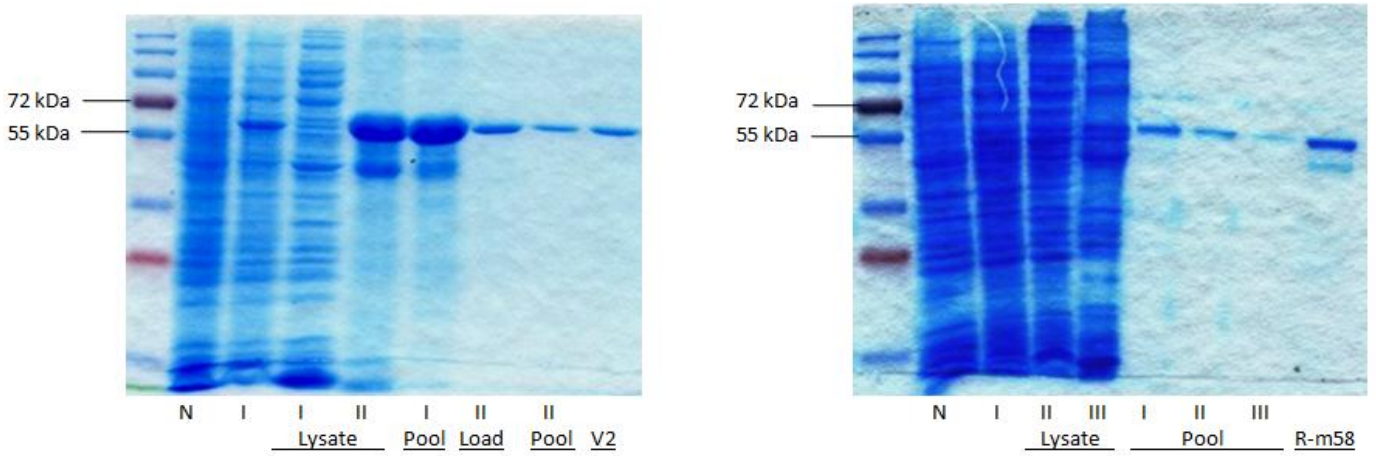


Figure 2.4.2d. Samples throughout the protein purification steps on polyacrylamide gels stained with GelCode™ Blue Stain.
Left – Purification of VirD2, N:20 µL of non-induced culture; I: 20 µL of induced culture; Lysate I: 25 µg; Lysate II: 25 µg; Pool I: 17.5 µg; Load II: 7.6 µg; Pool II: 0.26 µg; V2: 1 µg of purified VirD2.
Right – Purification of RbcS-TP-mVirD2-58; N: 20 µL of non-induced culture; I: 20 µL of induced culture; Lysate II: 25 µg; Lysate III: 25 µg; Pool I: 0.97 µg; Pool II: 0.25 µg; Pool III: 0.068 µg; R-m58:1 µg of purified RbcS-TP-mVirD2-58.

2.4.3. Cleavage activity analyses of VirD2 and RbcS-TP-mVirD2-58

In order to assess whether modifications of the VirD2 protein affect its activity crucial for the formation of the T-complex, the ability of the purified recombinant proteins VirD2 and RbcS-TP-mVirD2-58 to cleave the pTi border sequence was tested. The VirD2 site-specific ssDNA cleavage activity assay has been applied previously using the radioactively labelled ssDNA (Ziemienowicz et al., 1999; Ziemienowicz et al., 2001). We employed the same strategy using the fluorescently labelled ssDNA at the 5' end containing the right border sequence to detect the uncleaved substrate (34-mer) and the cleaved product (14-mer) generated by VirD2 and RbcS-TP-mVirD2 proteins. The reaction mixtures were separated on polyacrylamide/urea denaturing gel, and then the cleavage efficiency was determined based on a ratio between the uncleaved substrate and the cleaved product.

An electrophoretogram illustrates that VirD2 and RbcS-TP-mVirD2 were able to cleave 34-mer ssDNA and produce the cleaved product (14-mer) (Figure 2.4.3b). The quantity of cleavage product was comparable between both proteins. In addition, the maximum cleavage activity estimated by the Image J based on electrophoretograms was not significantly different among two proteins (VirD2: 77±4%, RbcS-TP-mVirD2-58: 73±4%) (Figure 2.4.3b). Three different phases of the cleavage efficiency were observed depending on the amount of the recombinant proteins: 1) a linear phase between 0 – 0.16 µg of protein, 2) an intermediate phase between 0.16 – 0.45 µg of protein, and 3) a plateau at > 0.45 µg of protein. It was concluded that changes in the VirD2 amino acid sequence resulting from mutations of NLSs and the addition of chloroplast targeting peptides; e.g., RbcS-TP, did not affect its cleavage activity.

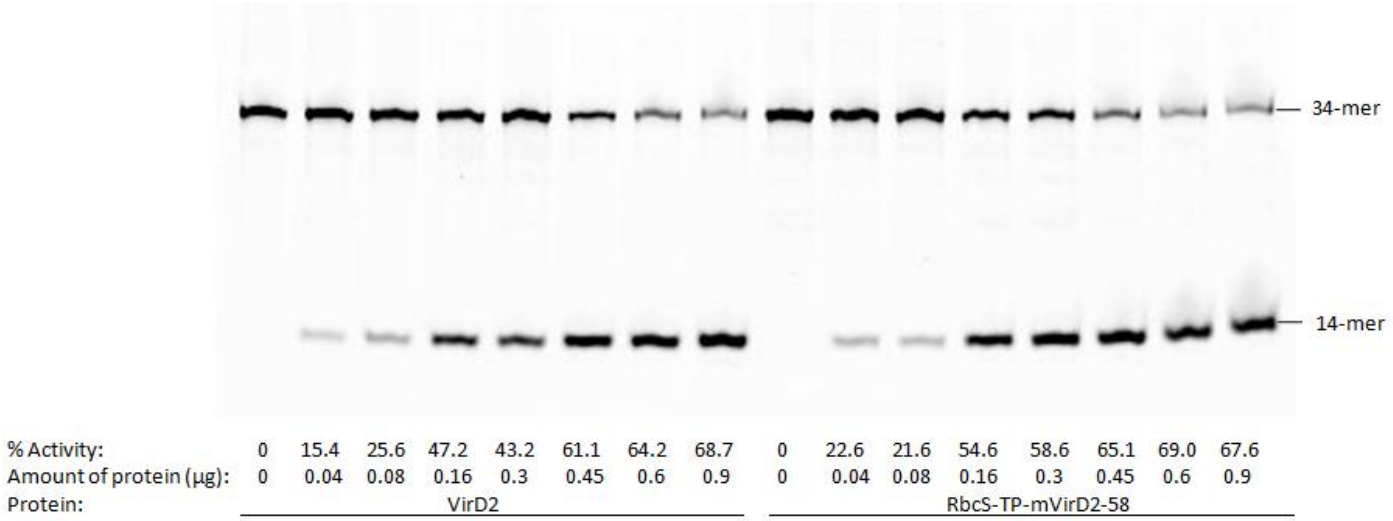


Figure 2.4.3a. An electrophoretogram of cleavage products by VirD2 and RbcS-TP-mVirD2-58 using the fluorescently labelled oligonucleotide.
34-mer: a cleavage substrate; 14-mer: the cleaved product; % activity: the cleavage activity estimated by Image J.

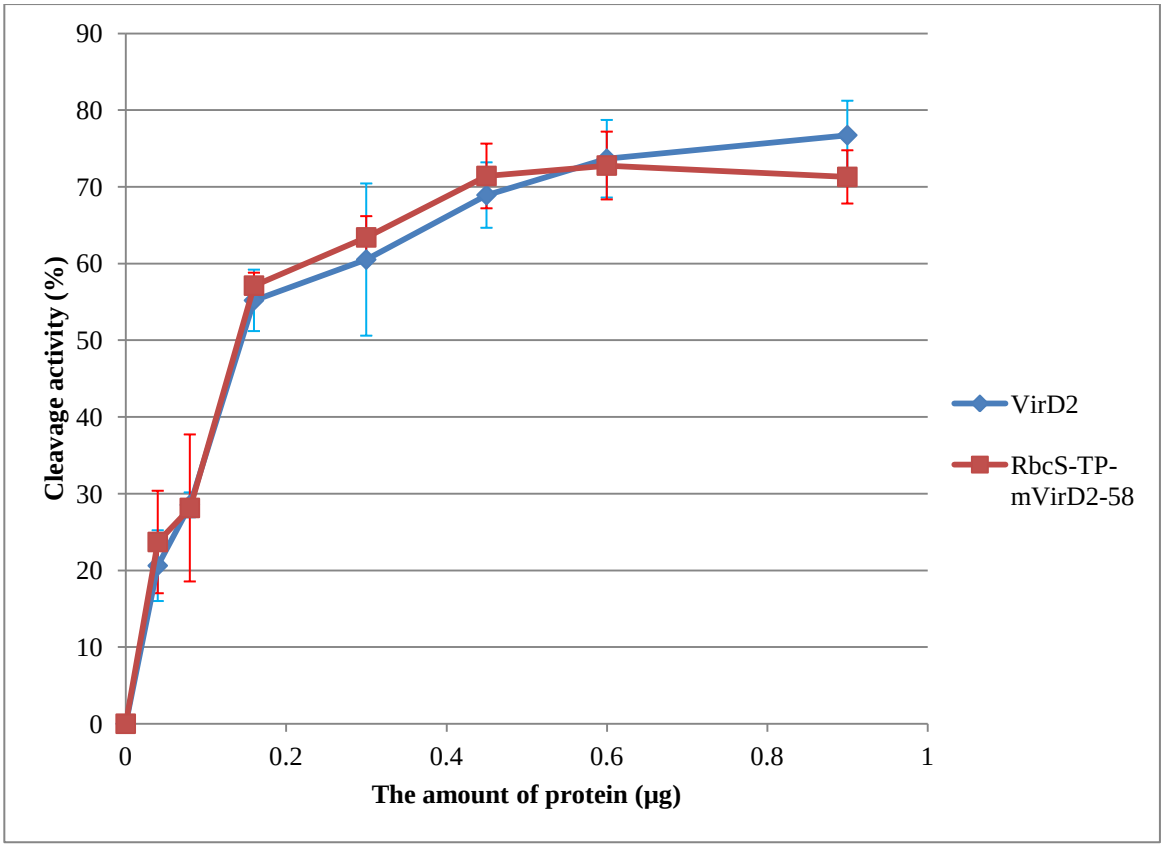


Figure 2.4.3b. Quantification of the cleavage activity of VirD2 and RbcS-TP-mVirD2 on the fluorescently labelled oligonucleotide substrate.
The data were generated using electrophoretograms (n=3±s.d.).

2.4.4. The generation of *Agrobacterium* strains for chloroplast transformation

2.4.4.1. The binary vector system for *Agrobacterium*-mediated chloroplast transformation

2.4.4.1.1. The generation of *Agrobacterium* carrying a helper plasmid, pPM6000AM containing *TP-mvirD2*

Since pVD derivatives are originated from pUC, they do not replicate in *Agrobacterium*. Thus, the *TP-mvirD2* genes from pVD derivatives need to be located in pAVD derivatives carrying an *Agrobacterium*-compatible origin of replication. Therefore, the *TP-mvirD2* genes in pVD derivatives were used to replace the *virD2* gene in pAVD44 resulting in new pAVD derivatives (pAVD58-DCL and RbcS-TP, and pAVD80-DCL and RbcS-TP). The newly constructed pAVD derivatives were then introduced into an *Agrobacterium* strain GV3101[pPM6000K] where the integration by homologous recombination between pAVD and a helper plasmid, pPM6000K, (Rossi et al., 1993) allowed for a replacement of the *virD2* gene with its modified versions (Figure 2.4.4.1.1). The helper plasmid, pPM6000K, carries an internal 70% deletion of the *virD2* gene by *Kpn*I. When pAVD derivatives with the *TP-mvirD2* gene are integrated into pPM6000K, the functional *TP-mvirD2* genes are present in *Agrobacterium* resulting in new helper plasmids, pPM6000AM derivatives (Table 2.3a). A stable pPM6000AM derivative in *Agrobacterium* was obtained through at least the eighth passage of replica plating.

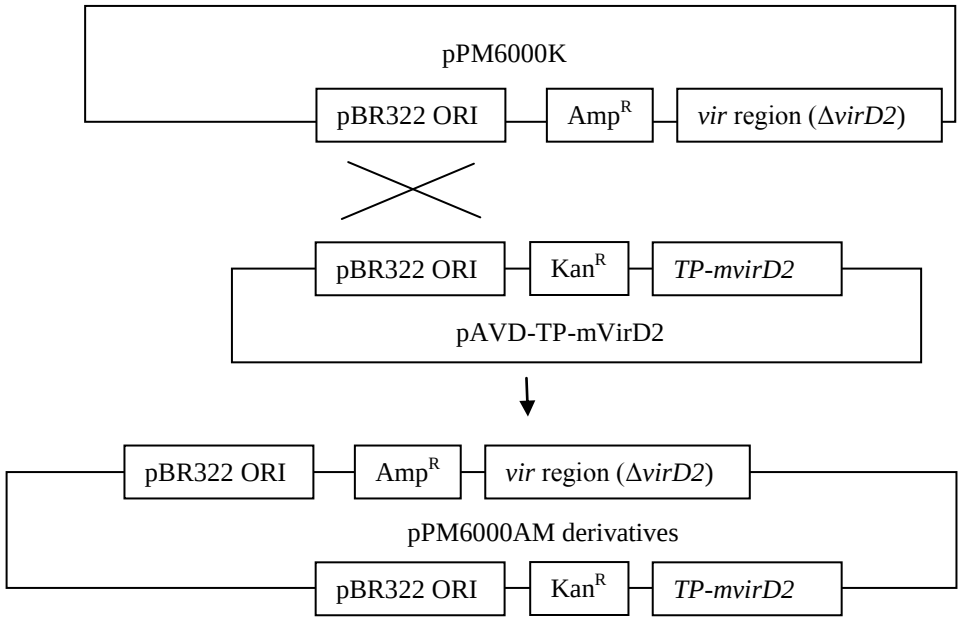


Figure 2.4.4.1.1. Construction of pPM6000AM derivatives.
Amp^R: the ampicillin resistance gene; Kan^R: the kanamycin resistance gene; pBR322 ORI: pBR322 origin of replication; cross lines: where homologous recombination occurs.

2.4.4.1.2. Construction of a binary vector carrying the chloroplast expression cassette

The chloroplast transformation vector (pLD-CtV) contains the chloroplast expression cassette carrying the *aadA* gene encoding the aminoglycoside 3' adenylyltransferase which confers spectinomycin and streptomycin resistance driven by the chloroplast regulatory elements, the 16S rRNA promoter (Prrn) and the *psbA* 3' terminator from a gene coding for a photosystem II reaction center component (TpsbA). The *aadA* gene expression cassette is flanked by chloroplast *trnI* and *trnA* regions required for a stable integration into the chloroplast genomes (Daniell et al., 1998). The kanamycin resistance gene in T-DNA of a modified binary vector pPZP500 was replaced with the chloroplast expression cassette (*trnI*-Prrn-*aadA*-TpsbA-*trnA*) resulting in pPZP500-aadA. Subsequently, pPZP500-aadA was introduced into *Agrobacterium* GV3101[pPM6000AM] derivatives to complete the binary vector system, GV3101[pPM6000AM::pPZP500-aadA] (Table 2.3a).

2.4.5. *Agrobacterium*-mediated chloroplast transformation

Leaves of the wild type *N. tabacum* cultivar Big Havana were transformed by the leaf incision method. A different concentration of spectinomycin between 0 and 500 mg/L was used in SIM medium in two different trials. Regeneration positive controls were provided using untransformed leaves on SIM without spectinomycin, and the shoots were regenerated from these leaves in all trials that will not be discussed any further. The untransformed leaves were placed on SIM containing spectinomycin served as regeneration-negative controls (selection controls).

2.4.5.1. The first trial

Tobacco leaves were transformed with derivatives of *Agrobacterium* strains GV3101[pPM6000AM::pPZP500-aadA], and the transformants were selected on SIM containing 500 mg/L spectinomycin. The untransformed leaves did not produce callus on SIM containing the aforementioned antibiotic as expected. Thus, 500 mg/L spectinomycin in SIM was sufficient to suppress the growth of untransformed plants. However, no regeneration of callus or shoot was also observed under these conditions from the leaves used for the transformation. Therefore, lower concentrations of the selection agent were used in the next transformation trials.

2.4.5.2. The second trial

Tobacco leaves were transformed with derivatives of *Agrobacterium* strains GV3101[pPM6000AM::pPZP500-aadA] and were grown on SIM without or with 100 mg/L or 250 mg/L spectinomycin. A total of 550 plants were regenerated on SIM without spectinomycin among all constructs (Treatment 1; Table 2.4.5.2). In Treatment 2, leaves were initially plated on SIM without spectinomycin to allow for the formation of calli/shoots. After shoots were emerged, they are placed on SIM containing 100 mg/L spectinomycin. Only six shoots from *Agrobacterium* carrying pPM6000AM4 survived on SIM with spectinomycin (Treatment 2; Table 2.4.5.2). In a total of five plants, three plants from *Agrobacterium* carrying pPM6000AM3 and two plants from the bacterium carrying pPM6000AM5 were regenerated when the leaves were first placed on SIM with 100 mg/L spectinomycin (Treatment 3). No regeneration of callus

or shoot was observed when the leaves were first placed on SIM with 250 mg/L spectinomycin (Treatment 4; Table 2.4.5.2).

PCR was performed to examine the presence of the transgene in the chloroplast genome. Primers AZ187 and AZ200 anneal to the *trnI* gene and the *trnA* gene, respectively (Figure 2.4.5). The total DNA of an untransformed plant would give a 1.1 kb PCR fragment (Figure 2.4.5, Top), while the transformed plant would produce a 2.5 kb PCR fragment due to the presence of an insert (Figure 2.4.5, bottom). Only a 1.1 kb PCR fragment was produced from all regenerated plants shown on the agarose gel (Figure 2.4.5.2). Thus, there was no integration of the transgene into the chloroplast genome. The untransformed leaves did not form callus and died on SIM with 100 mg/L spectinomycin, thus confirming that 100 mg/L spectinomycin provided a sufficient selection pressure.

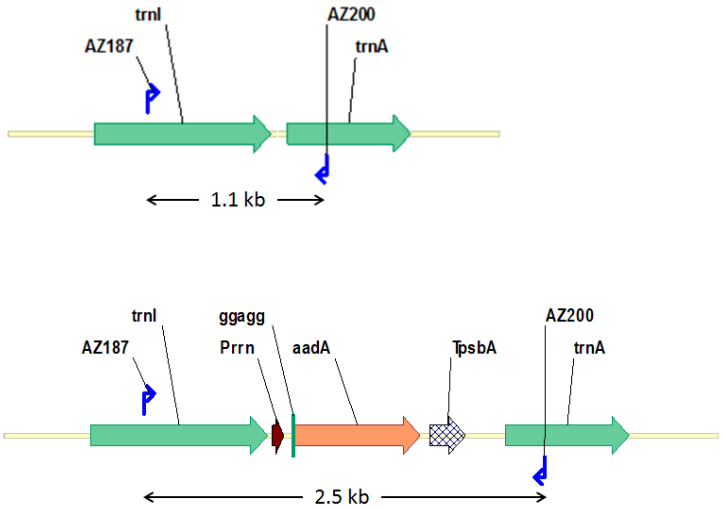


Figure 2.4.5. Top – The untransformed chloroplast genome. Bottom – the transformed chloroplast genome with the chloroplast expression cassette.
The chloroplast expression cassette: Prrn-*aadA*-TpsbA; Primers: AZ187 and AZ200; homologous recombination site: trnI-trnA.

Table 2.4.5.2. The number of regenerated plants from different treatments in the second trial.
Treatment 1: no spectinomycin; Treatment 2: leaves were initially placed on SIM without spectinomycin to allow for the growth of calli/shoots, and then shoots were transferred to SIM containing 100 mg/L spectinomycin; Treatment 3: SIM with 100 mg/L spectinomycin;
Treatment 4: SIM with 250 mg/mL spectinomycin.

	Treatment			
Construct	1	2	3	4
GV3101[pPM6000AM2::pPZP500-aadA]	33	0	0	0
GV3101[pPM6000AM3::pPZP500-aadA]	245	0	3	0
GV3101[pPM6000AM4::pPZP500-aadA]	173	6	0	0
GV3101[pPM6000AM5::pPZP500-aadA]	99	0	2	0
Total	550	6	5	0

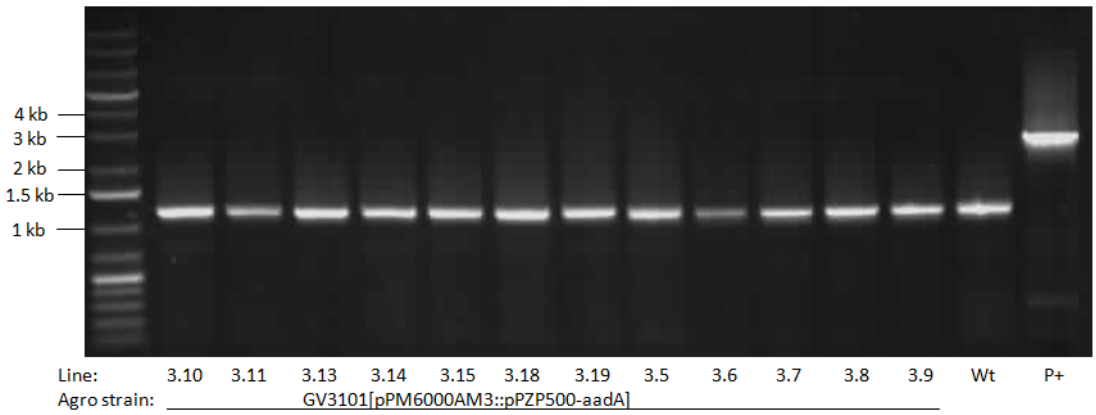


Figure 2.4.5.2. An electrophoretogram of PCR products detecting the presence of the transgene in the chloroplast genome from the second trial.
From left to right, 3.10 through 3.9 – selected transgenic lines, Wt: the untransformed wild type plant; P+: positive control pPZP500-aadA.

2.5. Discussion

Agrobacterium is an excellent tool to modify the plant nuclear genome. Many *Agrobacterium* proteins are involved in the transformation process. VirD2 plays important roles in T-DNA processing by its relaxase domain and in nuclear targeting of the T-complex by its NLSs (Gelvin, 2003, 2010b, 2012). However, *Agrobacterium* proteins have never been modified to transform the chloroplast genome. In theory, adding a chloroplast-targeting signal to VirD2 that lacks NLSs would direct the T-complex to the chloroplast, yet the transformation of tobacco chloroplasts was unsuccessful in our study. Thus, the presence of a strong transit peptide was overridden by an unknown factor.

2.5.1. Recombinant protein analysis on SDS-PAGE

The western blot of VirD2 and TP-mVirD2 using non-induced and induced cultures detected proteins with a slightly higher molecular weight than a predicted size (Figure 2.4.2c and Table 2.4.2). The predicted size of VirD2 is 48 kDa, but this protein is known to migrate as a 56 kDa polypeptide in SDS-PAGE, the phenomenon is called gel shifting (Jasper et al., 1994; Relic et al., 1998; Yanofsky et al., 1986). RbcS-TP-mVirD2-58 also migrated differently from the predicted size in SDS-PAGE (Figure 2.4.2c). Interestingly, this type of gel shifting is commonly seen in membrane proteins and cytosolic proteins with the amino acid substitution or post-translational modification (Rath et al., 2009; Shi et al., 2012). The gel shifting is proposed to occur due to detergent binding to hairpin structures in the proteins (Rath et al., 2009). In addition, the net charge of the protein is another factor; for example, a protein with a domain containing a high net negative charge migrates slowly in SDS-PAGE (Shi et al., 2012). These factors might lead to the gel shifting observed for VirD2 and TP-mVirD2 in SDS-PAGE.

The incubation temperature was critical for the induction of the *TP-mvirD2* genes. In previous studies, bacterial cultures were incubated at 16^{o}C or 37^{o}C after the induction of the *virD2* gene (Pelczar et al., 2004; Ziemienowicz et al., 1999). In our study, a high production of the VirD2 protein was observed even at 37 and 28^{o}C (Figure 2.4.2b and c). The lower temperature (28^{o}C) was more effective for the overproduction of RbcS-TP-mVirD2-58 than 37^{o}C. No difference was observed from non-induced and induced cultures containing DCL-TP-mVirD2-58, DCL-TP-mVirD2-80, and RbcS-TP-mVirD2-80, yet the western blot detected small amounts of these proteins (Figure 2.4.2b and c). The conditions for the

overproduction of these proteins still need to be optimized. A high production of a recombinant protein, plant progesterone 5β-reductase, was observed when the *P5βR2* gene was induced at early log-phase (OD_{600} 0.1) of the bacterial culture, and then the culture was incubated at 4^{o}C for 72 h (San-Miguel et al., 2013). These conditions might be effective for the overproduction of the DCL-TP-mvirD2-58, DCL-TP-mVirD2-80 and RbcS-TP-mVirD2-80 proteins.

2.5.2. Cleavage activity

The successfully purified VirD2 and RbcS-TP-mVirD2-58 were used for the cleavage activity assay. The endonuclease activity of VirD2 to cleave the T-DNA border sequences lies on the relaxase domain. The domain region is located within the N-terminal 228 amino acids where 82% homology is observed between pTiC58 and pTiA6, and it is necessary for T-strand processing (Scheiffele et al., 1995; Steck et al., 1990). However, based on a similarity between different relaxases, it is further minimized to 204 amino acids (van Kregten et al., 2009). In our study, the relaxase domain was not removed, and amino acid modifications were not introduced. Yet, since the DCL or RbcS transit peptide was fused to the N-terminus of mVirD2, we examined whether the additional amino acids would affect the protein folding and the cleavage activity of DCL- or RbcS-TP-mVirD2 fusion proteins. Figures 2.4.3a and b illustrate a similar pattern of cleavage products by VirD2 and RbcS-TP-mVirD2-58. Therefore, the addition of the RbcS transit peptide does not affect the cleavage activity of mVirD2. Only RbcS-TP-mVirD2-58 carrying 57 amino acids of the transit peptide and five amino acids from the mature protein out of four TP-mVirD2 proteins was purified and used for the cleavage activity assay in the study. Nevertheless, we assumed that these proteins would not differ from the tested one in their cleavage activity. RbcS-TP-mVirD2-80 is expected to show the same activity as the tested RbcS-TP-mVirD2-58 because they share the same transit peptide and differ only in the C-terminal part of VirD2 which is not involved in the cleavage activity. The DCL transit peptide in DCL-TP-mVirD2-58 and 80 would likely have no influence on the endonuclease activity of the proteins because it is a short domain containing 50 amino acids of the transit peptide and one amino acid of the mature protein.

Three phases of the cleavage efficiency were observed (Figure 2.4.3b). A linear phase indicated a high cleavage efficiency of both recombinant proteins at a lower amount of proteins, and the efficiency was

reduced in the intermediate phase. Finally, it reached a plateau when more than 0.45 µg of proteins were added (Figure 2.4.3b). A similar pattern of the cleavage efficiency was observed using the adeno-associated virus replication protein (Costello et al., 1997). Since the amount of the cleavage substrate and recombinant proteins is not a limiting factor in our case, it is possible that compartments of protein storage buffer might inhibit the cleavage activity of recombinant proteins, or a high amount of proteins may require more enzyme co-factors, Mg^{2+}, in the reaction mixture.

2.5.3. Chloroplast transformation

The *aadA* gene conferring resistance to spectinomycin and streptomycin is the most widely used selectable marker gene for chloroplast transformation (Miki and McHugh, 2004). In addition, the recommended concentration of spectinomycin is 100 to 500 mg/L for the selection of transplastomic plants (Verma et al., 2008). Therefore, 500 mg/L and 100-250 mg/L spectinomycin in SIM were used in the first and second trials, respectively. However, shoots were not regenerated at all on SIM containing 500 or 250 mg/L spectinomycin (Table 2.4.5.2). Only three and two plants were regenerated on SIM supplemented with 100 mg/L spectinomycin from the leaves transformed by GV3101[pPM6000K-AM3::pPZP500-aadA] and GV3101[pPM6000K-AM5::pPZP500-aadA], respectively (Table 2.4.5.2); however, the transgene was not detected in the chloroplast genome by PCR (Figure 2.4.5.2). These regenerated plants might be untransformed escape plants that are observed when a lower concentration of the selection agent is used (Ye et al., 2003). Moreover, these untransformed escapes could have spontaneous mutations in the *rrn16* gene in the chloroplast genome causing growth in the presence of spectinomycin (Dudas et al., 2012; Svab and Maliga, 1991).

The DCL and RbcS transit peptides in TP-mVirD2 were not recognized by the cytosolic chloroplast import machinery and did not facilitate chloroplast targeting of the T-complex at all. However, these transit peptides are strong chloroplast-targeting signals that have been shown to deliver the TP-GFP fusion protein to chloroplasts in turnip, radish, cabbage, cauliflower, broccoli, *Arabidopsis*, tobacco, sugarcane, wheat, lettuce, carrot and barley (Gnanasambandam et al., 2007). Furthermore, the subcellular localization prediction software available online predicted that DCL/RbcS-TP-mVirD2 proteins constructed in this study would localize in chloroplasts as shown in Table 2.5.3 (Briesemeister et al., 2010a, b;

Emanuelsson et al., 2000; Horton et al., 2007; Shatkay et al., 2007). Nevertheless, the DCL and RbcS transit peptides in DCL/RbcS-TP-mVirD2 were hampered for some unknown reason. Byeon et al. (2014) also observed a dysfunctional transit peptide of serotonin *N*-acetyltransferase (SNAT) in the fusion protein. When a rice SNAT transit peptide was fused to sheep SNAT, the fusion protein was localized to the cytoplasm instead of chloroplasts. Although the precise reason of the dysfunctional transit peptide in the fusion protein remains unknown, two possibilities were proposed: 1) the rice SNAT transit peptide may have a weak targeting activity and 2) the mature protein may not be suitable for chloroplast targeting because internal signals downstream of transit peptides is important for an additional targeting motif for chloroplast translocation (Byeon et al., 2014; Ha et al., 2004). Since DCL and RbcS transit peptides with an additional motif, a small part of the mature proteins, are strong chloroplast-targeting signals, both possibilities do not apply in the case of DCL/RbcS-TP-mVirD2. Our puzzling results lead to a question about nuclear targeting of the T-complex by VirD2 NLSs.

VirD2 NLSs have been examined in several studies. The first study that claimed the presence of the N-terminal NLS was conducted in 1990, and the C-terminal bipartite NLS was discovered in 1992 (Herrera-Estrella et al., 1990; Howard et al., 1992; Tinland et al., 1992). In these studies, a reporter protein was fused to VirD2 or a fragment of VirD2. The N-terminal NLS of VirD2 was examined by generating a fusion protein of first 292 amino acids of VirD2 excluding the C-terminal NLS to β-galactosidase that was detected in isolated tobacco nuclei (Herrera-Estrella et al., 1990). Thus, the presence of the N-terminal NLS of VirD2 was proposed. The experiments involving the C-terminal NLS of VirD2 were performed by constructing a fusion protein of β-glucuronidase-VirD2 carrying different deletions within VirD2 (Howard et al., 1992). Their results demonstrated that the fusion proteins containing the C-terminal NLS of VirD2 were localized in the nucleus of tobacco protoplasts. In the same year, Tinland et al (1992) fused only the N or C-terminal NLS to the N-terminus of β-galactosidase. The fusion proteins with either NLS were localized in the nuclei of tobacco protoplasts (Tinland et al., 1992). In addition, more recent studies reported that the fluorescently labelled VirD2 and the modified VirD2 protein with the active N-terminal NLS and lacking the C-terminal NLS by *Nru*I deletion were found to be localized to the nuclei of tobacco protoplasts (Ziemienowicz et al., 2001). Since then, VirD2 NLS study has not been conducted in the plant system.

Nevertheless, different approaches have been taken to observe the VirD2 function in various steps of plant transformation in addition to nuclear targeting. For example, the tumorigenesis was examined on potato tubers transformed by *Agrobacterium* strain carrying VirD2 with *NruI* deletion of the C-terminal NLS. The tumorigenesis was reduced to 1%, whereas a precise deletion of the C-terminal NLS decreased the tumorigenesis only to 60% (Shurvinton et al., 1992). On the other hand, a transient expression study on tobacco seedlings demonstrated that both a precise deletion and *Nru*I deletion of the C-terminal NLS caused a decrease in GUS activity to 4% (Rossi et al., 1993). This discrepancy suggested that the deleted domain might be involved in T-DNA integration which is necessary for a stable plant transformation but not for a transient expression of T-DNA (Janssen and Gardner, 1990; Lacroix et al., 2006; Nam et al., 1999; Rossi et al., 1993).

The involvement of VirD2 in the integration of T-DNA into the plant genome has been the subject of debate. The H-R-Y motif of the recombinase domain located at the N-terminal part of VirD2 was claimed to play a role in the integration of T-DNA, yet another study indicated that the VirD2 ω domain, present in the C-terminus near the C-terminal NLS may function in T-DNA integration, as reviewed elsewhere (Gelvin, 2012). Ziemienowicz et al. (2000) investigated the VirD2 fucntion in joining the 5' terminus of the T-DNA to plant DNA *in vitro* by the ligation-integration assay. Their results showed that VirD2 did not have a function as a DNA ligase in the reaction, while ligation between the 5' end of T-DNA from the artificial T-DNA-VirD2 complex and a partly double-stranded olignuceotide was observed in the reactoin with the presence of plant enzymes from plant extracts (Ziemienowicz et al., 2000). Although VirD2 does not act as a DNA ligase, VirD2 may still play a role in T-DNA integration through its interaction with plant proteins. VirD2 has been shown to be phosphorylated by the kinase CAK2M which also phosphorylates the C-terminal regulatory domain of RNA polymerase II that can recruit the TATA box-binding protein. Additionally, a tight association between VirD2 and the TATA binding-protein suggests that VirD2 interacts directly with the transcription machinery and help during the T-DNA integration process, as reviewed elsewhere (Ziemienowicz et al., 2008)

Moreover, van Kregten et al. (2009) investigated the region of VirD2 called the domain of unknown function (DUF) between the relaxase domain and the C-terminal NLS (amino acids 265 to 373) for its function during the translocation and integration of T-strands into host plant cells. Various

truncations were made within DUF, and a SV40 NLS was added to the N-terminus of all truncated VirD2 proteins. Also, the region containing the C-terminal NLS and ω domain was removed from some constructs, which then fused to a short translocation signal derived from VirF. Their results revealed that the DUF of VirD2 is not required for transformation processes, and the entire translocation domain or part of it is positioned in the C-terminal 40 amino acids of VirD2 including NLS and ω domain (van Kregten et al., 2009).

Although the function of VirD2 NLSs and other functions of VirD2 have been examined in various occasions, our results still cannot be explained by the previous findings, and they suggest the existence of a new domain of VirD2. In fact, a major finding of our study is that VirD2 with the inactivated NLSs in the presence of the transit peptide does not direct the T-complex to chloroplasts. Therefore, there must be an unknown factor that overrides a strong transit peptide not allowing the T-complex to enter the chloroplasts. Interestingly, the chloroplast-targeting signal was overridden in the fusion protein of the transit peptide to VirD2 containing an active C-terminal NLS which was localized into the nucleus (Barbara Hohn; personal communication). Although we were not able to clearly demonstrate the fact that modified complex still targeted the nucleus, it led us to hypothesize that VirD2 contains a novel NLS which could inhibit a function of the transit peptide. In conclusion, *Agrobacterium*-mediated chloroplast transformation by modifying the VirD2 protein was unsuccessful; nevertheless, our findings provide different perspectives on VirD2 NLSs.

Table 2.5.3. TP-mVirD2 protein sequences and their predicted subcellular localization using various prediction software.
DCL and RbcS transit peptide and additional mature protein sequences are underlined.

			Predicted Protein Subcellular Localization			
		The name of the website	Wolf Psort	TargetP	SherLoc	YLOC
		URL	http://www.genscript.com/psort/wolf_psort.html	http://www.cbs.dtu.dk/services/TargetP/	http://abi.inf.uni-tuebingen.de/Services/SherLoc/	http://abi.inf.uni-tuebingen.de/Services/YLoc/webloc.cgi
		Reference	(Horton et al., 2007)	(Emanuelsson et al., 2000)	(Shatkay et al., 2007)	(Briesemeister et al., 2010a, b)
Plasmid	Protein	Amino acid sequence				
pVD58-DCL-TP	DCL-TP-mVirD2-58	MASICTSNFHFLCRKNNSSPISHHLLLSPSSLSFSRCGGLRLCRCAAVKTGPDRAQVIIRIVPGGGTKTLQQIINQLEYLSTKGKLELQRSARHLDIPVPPDQIRELAQSWVTEAGIYDESQSDDDRQQDLTTHIIVSFPAGTDQTAAYEASREWAAEMFGSGYGGGRYNYLTAYHVDRDHPHLHVVVNRRELLGHGWLKISRRHPQLNYDGLRKKMAEISLRHGIVLDATSRAERGIAERPITYAEHRRLERMQAQKIQFEDTDFDETSPEEDRRDLSQSFDPFRSDPSTGEPDRATRHDKQPLEQHARFQESAGSSIKADARIRVSLESERSAQPSASKIPVIGHFGIETSYVAEASVRKRSGIFGTSRPVTDVAMHTVKRQQRSKRRNDEEAGPSGANRKGLKAAQVDSEANVGEQDTRDDGRGGT	Chloroplast	Chloroplast	Chloroplast	Chloroplast
pVD58-RbcS-TP	RbcS-TP-mVirD2-58	MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDITSIASNGGRVQCMQVWPTGPDRAQVIIRIVPGGGTKTLQQIINQLEYLSTKGKLELQRSARHLDIPVPPDQIRELAQSWVTEAGIYDESQSDDDRQQDLTTHIIVSFPAGTDQTAAYEASREWAAEMFGSGYGGGRYNYLTAYHVDRDHPHLHVVVNRRELLGHGWLKISRRHPQLNYDGLRKKMAEISLRHGIVLDATSRAERGIAERPITYAEHRRLERMQAQKIQFEDTDFDETSPEEDRRDLSQSFDPFRSDPSTGEPDRATRHDKQPLEQHARFQESAGSSIKADARIRVSLESERSAQPSASKIPVIGHFGIETSYVAEASVRKRSGIFGTSRPVTDVAMHTVKRQQRSKRRNDEEAGPSGANRKGLKAAQVDSEANVGEQDTRDDGRGGT	Chloroplast	Chloroplast	Chloroplast	Chloroplast

Table 2.5.3. Continued.

			Predicted Protein Subcellular Localization			
		The name of the website	Wolf Psort	TargetP	SherLoc	YLOC
		URL	http://www.genscript.com/psort/wolf_psort.html	http://www.cbs.dtu.dk/services/TargetP/	http://abi.inf.uni-tuebingen.de/Services/SherLoc/	http://abi.inf.uni-tuebingen.de/Services/YLoc/webloc.cgi
		Reference	(Horton et al., 2007)	(Emanuelsson et al., 2000)	(Shatkay et al., 2007)	(Briesemeister et al., 2010a, b)
Plasmid	Protein	Amino acid sequence				
pVD80-DCL-TP	DCL-TP-mVirD2-80	MASICTSNFHFLCRKNNSSPISHHLLLSPSSLSFSRCGGLRLCRCAAVKTGPDRAQVIIRIVPGGGTKTLQQIINQLEYLSTKGKLELQRSARHLDIPVPPDQIRELAQSWVTEAGIYDESQSDDDRQQDLTTHIIVSFPAGTDQTAAYEASREWAAEMFGSGYGGGRYNYLTAYHVDRDHPHLHVVVNRREELLGHGWLKISRRHPQLNYDGLRKKMAEISLRHGIVLDATSRAERGIAERPITYAEHRRLERMQAQKIQFEDTDFDETSPEEDRRDLSQSFDPFRSDPSTGEPDRATRHDKQPLEQHARFQESAGSSIKADARIRVSLESERSAQPSASKIPVIGHFGIETSYVAEASVRKRSGIFGTSRPVTDVAMHTVKRQQRSKRRNDEEAGPSGANRKGLKAAQVDSEANVGEQDTRDDSNKAADPVSASIGTEQPEASPDLGNRRDDGRGGT	Chloroplast	Chloroplast	Chloroplast	Chloroplast
pVD80-RbcS-TP	RbcS-TP-mVirD2-80	MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDITSIASNGGRVQCMQVWPTGPDRAQVIIRIVPGGGTKTLQQIINQLEYLSTKGKLELQRSARHLDIPVPPDQIRELAQSWVTEAGIYDESQSDDDRQQDLTTHIIVSFPAGTDQTAAYEASREWAAEMFGSGYGGGRYNYLTAYHVDRDHPHLHVVVNRRELLGHGWLKISRRHPQLNYDGLRKKMAEISLRHGIVLDATSRAERGIAERPITYAEHRRLERMQAQKIQFEDTDFDETSPEEDRRDLSQSFDPFRSDPSTGEPDRATRHDKQPLEQHARFQESAGSSIKADARIRVSLESERSAQPSASKIPVIGHFGIETSYVAEASVRKRSGIFGTSRPVTDVAMHTVKRQQRSKRRNDEEAGPSGANRKGLKAAQVDSEANVGEQDTRDDSNKAADPVSASIGTEQPEASPDLGNRRDDGRGGT	Chloroplast	Chloroplast	Chloroplast	Chloroplast

Chapter 3

Discovery of a Novel Nuclear Localization Signal in *Agrobacterium* VirD2 Protein

3.1. Abstract

Agrobacterium VirD2 protein is one of the best characterized bacterial proteins. It contains two nuclear localization signals: N-terminal NLS as a monopartite type and C-terminal NLS as a bipartite type. Mutations in the NLS coding sequences have been shown to decrease the efficiency of *Agrobacterium*-mediated plant transformation. Nuclear targeting of a VirD2 fusion protein to a reporter protein demonstrated that mutations in the NLS coding sequence inactivated its NLS activity. Therefore, we introduced a point mutation in the N-terminal NLS coding sequence and a broad or a precise deletion of the C-terminal NLS to generate the modified VirD2 (mVirD2) protein lacking both functional NLSs. This was done in attempt to engineer Agrobacteria that would deliver the T-DNA into chloroplasts rather than into the nucleus. For this purpose, a chloroplast transit peptide was fused to the N-terminus of mVirD2. Despite the aforementioned modifications, the T-complex was not delivered into chloroplasts. This indicated that the presence of a strong chloroplast transit peptide was likely overcome by an additional unknown VirD2 NLS or NLSs. Thus, we re-examined NLSs of VirD2 to clarify the dysfunction of transit peptides. We constructed several truncated VirD2 proteins fused to the enhanced green fluorescent protein (EGFP). The subcellular localization of recombinant proteins in transgenic *Arabidopsis* plants and transiently transformed tobacco leaves indicated that the C-terminal part of mVirD2 still showed a strong nuclear targeting activity. We used cNLS Map software to predict a putative NLS named NLS3. NLS3 with several additional amino acids were fused to EGFP resulting in the NLS3-EGFP fusion protein (33 kDa). This fusion protein was localized in the cytoplasm and in the nucleus of tobacco leaf cells likely due to its small size causing the diffusion of the protein through the nuclear pore. However, when two additional EGFPs were fused, this recombinant protein (NLS3-EGFPx3, 89 kDa) was localized exclusively in the nucleus of tobacco leaf cells. Therefore, we confirmed the presence of a novel NLS in VirD2.

3.2. Introduction

Agrobacterium tumefaciens has been a great tool in plant biotechnology because of the bacterial ability to transfer its DNA (transferred DNA, T-DNA) into the plant cell and then into the nucleus where the T-DNA is integrated. Although the journey of the T-DNA into the host nucleus is an intricate process, continuous efforts of researchers have shed light on the transformation mechanism and the interplay between *Agrobacterium* and plant proteins.

Agrobacterium carries the tumor inducing plasmid (pTi) wherein the virulence (*vir*) genes and the T-DNA are located. The *vir* genes encode several Vir proteins including VirA, VirB, VirC, VirD, VirE, VirF and VirG. These proteins play important roles in the transformation process. First, phenolic compounds which are secreted from the wounded plant are sensed by the VirA/VirG signal transduction system, resulting in an activation of the *vir* genes. Upon activation, the VirD1/VirD2 endonuclease cleaves the T-DNA border sequences, namely the right and left borders, which are edging T-DNA. In the meantime, VirD2 attaches to the 5' end of the cleavedT-DNAwhich is displaced as a single stranded DNA (T-strand) from the pTi forming the T-DNA-VirD2-complex. The complex is passing through the type IV secretion system consisting of VirB and VirD4 proteins into the plant cell. VirE2 proteins coat the T-strand of the complex forming the T-complex in the plant cell to protect it from host nucleases. Finally, the T-complex is guided into the nucleus where the T-DNA is integrated into the nuclear genome (Gelvin, 2003, 2010b; Meyers et al., 2010; Ziemienowicz, 2001).

VirD2 is an essential protein for T-DNA processing and transfer. This protein contains the relaxase domain which cleaves the left and right borders with a help of VirD1 (Scheiffele et al., 1995). This domain resides in the N-terminal half of the protein which is a highly (85%) conserved domain among different *Agrobacterium* strains (Howard et al., 1992). The ω domain at the C-terminus of the protein was suggested to be important but not essential for T-DNA translocation to the host cell (Bravo-Angel et al., 1998). VirD2 also carries the nuclear localization signals (NLS) to pilot the T-complex into the nucleus. A monopartite and a bipartite NLS are at the N-terminal and C-terminal regions of VirD2, respectively. At first, the presence of the N-terminal NLS was suggested by Herrera-Estrella et al. (1990). Tobacco leaf discs were transformed by *Agrobacterium* harboring a plasmid containing the sequence coding for the first 292 amino acids of VirD2 fused to the *lacZ* gene. A VirD2 (292 amino acids)-β-galactosidase fusion protein was

detected by the β-galactosidase activity in plant cell fractions. Approximately a three-fold higher β-galactosidase activity was observed in the nuclear fraction than in the cytosol fraction. According to the results and the proposed consensus sequences for nuclear targeting of proteins, an N-terminal NLS (RKGR) was suggested (Herrera-Estrella et al., 1990). In addition, the proposed bipartite C-terminal was tested by GUS activity using fusion proteins of β-glucuronidase (GUS) and several types of truncated VirD2 in tobacco protoplasts (Howard et al., 1992). The results showed that the GUS-truncated VirD2 protein in the absence of the N-terminal NLS and the presence of a C-terminal NLS was detected in the nucleus. On the contrary, the GUS-truncated VirD2 proteins that contain the wild type N-terminal NLS but lack a C-terminal NLS by removing the C-terminal 158 amino acids or eliminating only bipartite sequences were detected in the cytoplasm (Howard et al., 1992). This discrepancy regarding the N-terminal NLS of VirD2 was solved by other researchers.

Another study on NLSs of VirD2 was published by Tinland et al. (1992). In their experiments, VirD2-β-galactosidase fusion proteins were constructed to observe their subcellular localization in yeast and tobacco protoplasts. VirD2 missing the C-terminal 29 amino acids, including the C-terminal NLS fused to β-galactosidase, was localized in yeast nuclei, but it was not tested in the plant system. Eleven amino acids of VirD2 containing the N-terminal NLS fused to β-galactosidase and 20 amino acids of VirD2 carrying the C-terminal NLS fused to β-galactosidase were detected in the nucleus of tobacco protoplasts by indirect immunofluorescence (Tinland et al., 1992). These results reinforced the presence of the N and C-terminal NLSs. These findings were further confirmed by a confocal microscopy study of fluorescently labelled wild-type VirD2 and a mutant lacking the C-terminal NLS, both were localized in the nuclei isolated from tobacco protoplasts (Ziemienowicz et al., 2001). However, the discovery of a covalent linkage between the T-DNA and tyrosine 29 (Y29) of VirD2, just in a few amino acids proximity of the N-terminal NLS, questioned the efficiency of the N-terminal NLS activity in nuclear targeting of the T-complex (Vogel and Das, 1992). It was suggested that the N-terminal NLS was likely inactivated because the linkage between VirD2 and the T-DNA might cause a conformational change in VirD2 (Rossi et al., 1993).

While VirD2 NLSs have been examined using the reporter constructs, various types of modifications were made within VirD2 to observe the importance of VirD2 during a stable and transient transformation in several types of plant systems. For example, different regions of VirD2 were deleted to

examine the effect on tumorigenesis using potato tubers (Shurvinton et al., 1992). The deletion of a large portion of the relaxase domain led to the abolishment of tumorigenesis. Other mutations in VirD2 caused a severe or moderate reduction intumorigenesis. For example, removing 45 amino acids including the C-terminal NLS (a broad deletion of the C-terminal NLS by *Nru*I) of VirD2 reduced tumorigenesis to 1%. However, the elimination of exclusively the bipartite C-terminal NLS of VirD2 reduced tumorigenesis to 60% (Shurvinton et al., 1992). Furthermore, Rossi et al. (1993) observed an effect of modified VirD2 proteins on the transient expression of GUS by measuring the GUS activity in tobacco seedlings. Both the aforementioned broad and precise deletions of the C-terminal NLS of VirD2 reduced the transient expression to 1.9 and 4%, respectively. Four different point mutations were introduced to the N-terminal NLS, and only one point mutation (a change from arginine to threonine) decreased the transient expression. Three other mutations did not have a significant impact on GUS expression (Rossi et al., 1993). They concluded that the N-terminal NLS of VirD2 did not have a critical role in nuclear targeting of the T-DNA. Furthermore, a decrease in transient expression of GUS by a modification of the N-terminal NLS (a change from arginine to threonine) of VirD2 was suggested to be due to the interruption of T-DNA processing rather than due to the lower capacity for its nuclear targeting (Rossi et al., 1993). However, other studies reported that the deletion of the C terminal NLS did not abolish the cleavage activity of the modified VirD2 protein (Ziemienowicz et al., 1999), but it eliminated the NLS (Ziemienowicz et al., 2001).

Plant proteins interacting with VirD2 have been identified. *Arabidopsis thaliana* karyopherin α (AtKAPα renamed IMPa-1) of the importin α family was first discovered to interact with the C-terminal NLS of VirD2 (Ballas and Citovsky, 1997). Other importin α isoforms such as IMPa-2, IMPa-3 and IMPa-4 have been determined to interact with VirD2 (Bhattacharjee et al., 2008). Ballas and Citovsky (1997) showed that a precise deletion of the C-terminal NLS of VirD2 abolished nuclear import of VirD2 via KAPα, whereas Bhattacharjee et al. (2008) reported that VirD2 carrying a slightly different precise deletion was still able to be transferred into the nucleus via the importin α pathway. This discrepancy was suggested to be due to the difference in the type of deletions (Bhattacharjee et al., 2008). In addition, an interaction between IMPa-4 and VirD2 was abolished by a broad deletion of the C-terminal NLS of VirD2 via *Nru*I (Bhattacharjee et al., 2008).

VirE2 also contains a bipartite NLS in the central region of the protein, but its nuclear targeting is controversial. Studies on subcellular localization of VirE2 demonstrated either nuclear or cytoplasmic localization reviewed elsewhere (Gelvin, 2010a, 2012). Ziemienowicz et al. (2001) demonstrated that VirE2 was essential for the nuclear entry but not for nuclear targeting of the T-complex because VirE2 did not cause an accumulation of the fluorescently labelled ssDNA in the nucleus in the absence of VirD2, while the presence of VirD2 caused the accumulation of ssDNA in the nucleus of tobacco protoplasts. It was suggested that NLSs of ssDNA-bound VirE2 are unavailable to the import machinery (Ziemienowicz et al., 2001). However, the ssDNA-VirE2 complex was detected in the nucleus of stamen hair cells of *Tradescantia virginiana* (Zupan et al., 1996). The discrepancy between two studies was thought to be due to different systems used. Interestingly, another study showed that the ssDNA-VirE2 complex could still interact with importins, particularly IMPa-4 *in vitro* (Bhattacharjee et al., 2008). Subcellular localization of the ssDNA-VirE2 complex was not examined in this study. Thus, the function of VirE2 NLS in the nuclear import of the T-complex is still unclear.

Proteins targeted to the nucleus require NLS, while proteins imported into chloroplasts contain specific chloroplast transit peptides. More than 95% of chloroplast proteins are encoded in the nuclear genome. Therefore, most of these proteins possess transit peptides that enable protein import through the translocon at the outer chloroplast membrane and the translocon at the inner chloroplast membrane facilitated by cytosolic sorting factors such as 14-3-3 proteins, Hsp70, Hsp90, FKBP and HOP. Transit peptides lack consensus sequences and vary in length (Jarvis and Lopez-Juez, 2013; Lee et al., 2013; Shi and Theg, 2013). To find an efficient transit peptide to direct recombinant proteins into chloroplasts, Gnanasambandam et al. (2007) employed the tomato defective chloroplasts and leaves (DCL) and the tobacco ribulosebisphosphate carboxylase (Rubisco) small subunit (RbcS) transit peptides that were fused to the green fluorescent protein (GFP), and they demonstrated a high efficiency of chloroplast import of the recombinant proteins by these transit peptides in several monocot and dicot plants.

According to previous studies on VirD2 NLSs and findings of Gnanasambandam et al. (2007), we constructed modified VirD2 (mVirD2) proteins carrying inactivated NLSs (a change from arginine to threonine in the N-terminal NLS and a broad or a precise deletion of the C-terminal NLS) fused to a chloroplast transit peptide (TP) for *Agrobacterium*-mediated chloroplast transformation by tobacco incision

method (see Chapter 2). We tested the TP-mVirD2 cleavage activity to ensure that an addition of TP at the N-terminus of mVirD2 did not disturb its cleavage activity using fluorescently labelled oligos containing a right border sequence with the VirD2 cleavage site. The efficiency of cleavage activity of RbcS-TP-mVirD2-58 (VirD2 carrying the modified N-terminal NLS and a broad deletion of the C-terminal NLS) and wild type VirD2 was comparable (see Chapter 2). Thus, an additional TP did not affect the relaxase domain in TP-mVirD2. Furthermore, we used the chloroplast expression vector (pLD-CtV) (Daniell et al., 1998) to construct a vector containing the chloroplast expression cassette (*trnI*-Prrn-*aadA*-TpsbA-*trnA*). The newly constructed binary vector carrying the chloroplast expression cassette was used to transform tobacco leaves by *Agrobacterium* harboring the plasmid with *TP-mvirD2* genes. None of the regenerated plants contained transgene in the chloroplast, but possibly in the nucleus. These results led to a new hypothesis of the presence of a novel NLS.

Here, we constructed various types of truncated VirD2 fused to the enhanced GFP (EGFP) to observe the subcellular localization of fusion proteins in transgenic *Arabidopsis* plants and in transiently transformed tobacco leaves via agro-infiltration. We added extra EGFPs to the fragment of VirD2 to increase the protein size and avoid the diffusion of the fusion protein through the nuclear pore. Also, we used the cNLS Mapper software available online to predict a putative novel NLS named as NLS3 which may interact with importin α (Kosugi et al., 2009b). Our results showed that NLS3 fused to three EGFPs was localized exclusively in the nucleus. Therefore, the presence of NLS3 was confirmed. Interestingly, this NLS3 was not as strong as the C-terminal NLS of VirD2. It is possible that because of this weak NLS3 activity, it was overshadowed by the C-terminal NLS. Our finding added a new feature to VirD2.

3.3. Materials and Methods

3.3.1. Construction of pEZT-NL derivatives

All binary vectors to examine the subcellular localization of recombinant proteins were constructed using pEZT-NL (http://deepgreen.stanford.edu/) containing the analine polylinker and EGFP coding sequences, a kind gift from Dr. Monroe (James Madison University, Virginia). The generation of plasmid pVD derivatives was described in Chapter 2 (Table 2.3a). Primers used in this study are indicated in Table 3.3.1a. The names of VirD2 fragments are illustrated in Figure 3.3.1a.

3.3.1.1. Construction of pEZT-NL-DCL-TP and pEZT-NL-RbcS-TP

The *DCL-TP* sequence was amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 20 µL reaction mixture containing 5 ng of pVD58-DCL-TP (see Chapter 2, Table 2.3a) and primers AZ204 and AZ267 (Table 3.3.1), 0.025 U/µL of TakaraExTaq™ DNA Polymerase (Takara Bio USA), 1xExTaq™ Buffer (Takara Bio USA), 200 µM of each dNTP Mixture (Takara Bio USA). PCR parameters were as follows: one cycle at 98^{o}C for 30 s, 30 cycles at 98^{o}C for 10 s, 55^{o}C for 30 s, 72^{o}C for 30 s, and one cycle at 72^{o}C for 10 min. The *Xho*I-digested PCR product was cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-DCL-TP (Table 3.4.1).

The *RbcS-TP* sequence was amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 20 µL reaction mixture containing 5 ng of pVD58-RbcS-TP (see Chapter 2, Table 2.3a) and primers AZ205 and AZ268 (Table 3.3.1), 0.025 U/µL of TakaraExTaq™ DNA Polymerase (Takara Bio USA), 1xExTaq™ Buffer (Takara Bio USA), 200 µM of each dNTP Mixture (Takara Bio USA). PCR parameters were as follows: one cycle at 98^{o}C for 30 s, 30 cycles at 98^{o}C for 10 s, 55^{o}C for 30 s, 72^{o}C for 30 s, and one cycle at 72^{o}C for 10 min. The *Xho*I-digested PCR product was cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-RbcS-TP (Table 3.4.1).

3.3.1.2. Construction of pEZT-NL-DCL-TP-mVirD2-58 and pEZT-NL-RbcS-TP-mVirD2-58, pEZT-NL-DCL-TP-mVirD2-80 and pEZT-NL-RbcS-TP-mVirD2-80

The *DCL-TP-mvirD2-58* and *DCL-TP-mvirD2-80* genes were amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 20 µL reaction mixture containing 5 ng of pVD58-DCL-TP or pVD80-

DCL-TP (see Chapter 2, Table 2.3a), 0.5 µM of primers AZ204 and AZ203 (Table 3.3.1), 0.025 U/µL of TakaraExTaq™ DNA Polymerase (Takara Bio USA), 1xExTaq™ Buffer (Takara Bio USA), 200 µM of each dNTP Mixture (Takara Bio USA). PCR parameters were as follows: one cycle at 98^{o}C for 30 s, 30 cycles at 98^{o}C for 10 s, 62^{o}C for 30 s, 72^{o}C for 90 s, and one cycle at 72^{o}C for 10 min. The *Xho*I-digested PCR products were cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-DCL-TP-mVirD2-58 or pEZT-NL-DCL-TP-mVirD2-80, respectively (Table 3.4.1).

The *RbcS-TP-mvirD2-58* and *RbcS-TP-mvirD2-80* genes were amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 20 µL reaction mixture containing 5 ng of pVD58-RbcS-TP or pVD80-RbcS-TP (see Chapter 2, Table 2.3a), 0.5 µM of primers AZ245 and AZ203 (Table 3.3.1), 0.025 U/µL of TakaraExTaq™ DNA Polymerase (Takara Bio USA), 1xExTaq™ Buffer (Takara Bio USA), 200 µM of each dNTP Mixture (Takara Bio USA). PCR parameters were as follows: one cycle at 98^{o}C for 30 s, 30 cycles at 98^{o}C for 10 s, 70^{o}C for 30 s, 72^{o}C for 90 s, and one cycle at 72^{o}C for 10 min. The *Xho*I-digested PCR products were cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-RbcS-TP-mVirD2-58 or pEZT-NL-RbcS-TP-mVirD2-80, respectively (Table 3.4.1).

3.3.1.3. Construction of pEZT-NL-VirD2, pEZT-NL-mVirD2-58 and pEZT-NL-mVirD2-80

The *virD2* gene was amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 20 µL reaction mixture containing 5 ng of pVD43 (see Chapter 2, Table 2.3a), 0.5 µM of primers AZ202 and AZ203 (Table 3.3.1), 0.025 U/µL of TakaraExTaq™ DNA Polymerase (Takara Bio USA), 1xExTaq™ Buffer (Takara Bio USA), 200 µM of each dNTP Mixture (Takara Bio USA). PCR parameters were as follows: one cycle at 98^{o}C for 30 s, 30 cycles at 98^{o}C for 10 s, 69^{o}C for 30 s, 72^{o}C for 90 s, and one cycle at 72^{o}C for 10 min. The *Xho*I-digested PCR product was cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-VirD2 (Table 3.4.1).

The *mvirD2-58* and *mvirD2-80* genes were amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 25 µL reaction mixture containing 5 ng of pVD58-AgeI or pVD80 (see Chapter 2, Table 2.3a), 0.5 µM of primers AZ202 and AZ203 (Table 3.3.1), 0.02 U/µL of Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs), 1xQ5 Reaction Buffer (New England Biolabs), 200 µM of each dNTP Mixture. PCR parameters were as follows: one cycle at 98^{o}C for 30 s, 30 cycles at 98^{o}C for 10 s, 72^{o}C for

30 s, 72°C for 20 s, and one cycle at 72°C for 2 min. The *Xho*I-digested PCR products were cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-mVirD2-58 or pEZT-NL-mVirD2-80, respectively (Table 3.4.1).

3.3.1.4. Construction of pEZT-NL carrying the partial *virD2* or *mvirD2* gene encoding the N-terminal region of VirD2 or mVirD2

3.3.1.4.1. Construction of pEZT-NL-VirD2-N1/2 and pEZT-NL-mVirD2-58/80-N1/2

The sequences coding for the N-terminal halves of the proteins in the *virD2* and *mvirD2-58/*80 genes were amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 25 µL reaction mixture containing 5 ng of pVD43 or pVD80 (see Chapter 2, Table 2.3a), 0.5 µM of primers AZ202 and AM23 (Table 3.3.1), 0.02 U/µL of Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs), 1xQ5 Reaction Buffer (New England Biolabs), 200 µM of each dNTP Mixture. PCR parameters were as follows: one cycle at 98°C for 30 s, 30 cycles at 98°C for 10 s, 64°C for 30 s, 72°C for 10 s, and one cycle at 72°C for 2 min. The *Xho*I-digested PCR products were cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-VirD2-N1/2 or pEZT-NL-mVirD2-58/80-N1/2, respectively (Table 3.4.1).

3.3.1.4.2. Construction of pEZT-NL-VirD2-N1/3 and pEZT-NL-mVirD2-58/80-N1/3

The sequences coding for the N-terminal one third of VirD2 and mVirD2-58/80 in the *virD2* and m*virD2-58/80* genes were amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 25 µL reaction mixture containing 5 ng of pVD43 or pVD80 (see Chapter 2, Table 2.3a), 0.5 µM of primers AZ202 and AM25 (Table 3.3.1), 0.02 U/µL of Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs), 1xQ5 Reaction Buffer (New England Biolabs), 200 µM of each dNTP Mixture. PCR parameters were as follows: one cycle at 98°C for 30 s, 30 cycles at 98°C for 10 s, 72°C for 30 s for pVD43 and 60°C for pVD80, 72°C for 10 s, and one cycle at 72°C for 2 min. The *Xho*I-digested PCR products were cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-VirD2-N1/3 or pEZT-NL-mVirD2-80/80-N1/3, respectively (Table 3.4.1).

3.3.1.5. Construction of pEZT-NL carrying the partial *virD2* or *mvirD2* gene encoding the C-terminal region of VirD2 or mVirD2

3.3.1.5.1. Construction of pEZT-NL-VirD2-C1/2, pEZT-NL-mVirD2-58-C1/2 and pEZT-NL-mVirD2-80-C1/2

The sequences coding for the C-terminal half of the protein in the *virD2*, *mvirD2-58* and *mvirD2-80* genes were amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 25 µL reaction mixture containing 5 ng of pVD43, pVD58-AgeI or pVD80 (see Chapter 2, Table 2.3a), 0.5 µM of primers AM24 and AZ203 (Table 3.3.1), 0.02 U/µL of Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs), 1xQ5 Reaction Buffer (New England Biolabs), 200 µM of each dNTP Mixture. PCR parameters were as follows: one cycle at 98^{o}C for 30 s, 30 cycles at 98^{o}C for 10 s, 67^{o}C for 30 s, 72^{o}C for 10 s, and one cycle at 72^{o}C for 2 min. The *Xho*I-digested PCR products were cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-VirD2-C1/2, pEZT-NL-VirD2-58-C1/2 or pEZT-NL-VirD2-80-C1/2, respectively (Table 3.4.1).

3.3.1.5.2. Construction of pEZT-NL-VirD2-C1/3, pEZT-NL-mVirD2-58-C1/3, pEZT-NL-mVirD2-80-C1/3 and pEZT-NL-mVirD2-80-C1/3-EGFPx2

The sequences coding for the C-terminal one third of the protein in the *virD2*, *mvirD2-58* and *mvirD2-80* genes were amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 25 µL reaction mixture containing 5 ng of pVD43, pVD58-AgeI or pVD80 (see Chapter 2, Table 2.3a), 0.5 µM of primers AM28 and AZ203 (Table 3.3.1), 0.02 U/µL of Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs), 1xQ5 Reaction Buffer (New England Biolabs), 200 µM of each dNTP Mixture. PCR parameters were as follows: one cycle at 98^{o}C for 30 s, 30 cycles at 98^{o}C for 10 s, 72^{o}C for 30 s, 72^{o}C for 10 s, and one cycle at 72^{o}C for 2 min. The *Xho*I-digested PCR products were cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-VirD2-C1/3, pEZT-NL-mVirD2-58-C1/3 or pEZT-NL-mVirD2-80-C1/3, respectively (Table 3.4.1).

The sequences coding for the alanine linker and EGFP were amplified and flanked by *Eco*RI and *Sma*I sites by PCR. PCR was performed in a 20 µL reaction mixture containing 5 ng of pEZT-NL, 0.5 µM of primers AM42 and AM43 (Table 3.3.1), 0.025 U/µL of TakaraExTaq™ DNA Polymerase (Takara Bio

USA), 1xExTaq™ Buffer (Takara Bio USA), 200 µM of each dNTP Mixture (Takara Bio USA). PCR parameters were as follows: one cycle at 98°C for 30 s, 30 cycles at 98°C for 10 s, 70°C for 30 s, 72°C for 1 min, and one cycle at 72°C for 2 min. The *Eco*RI/*Sma*I-digested PCR product was cloned into the *Eco*RI/*Sma*I-digested pEZT-NL-mVirD2-80-C1/3 resulting in pEZT-NL-mVirD2-80-C1/3-EGFPx2 (Table 3.4.1).

3.3.1.5.3. Construction of pEZT-NL-VirD2-C1/3ΔNLS3, pEZT-NL-mVirD2-80-C1/3ΔNLS3 and pEZT-NL-mVirD2-80-C1/3ΔNLS3-EGFPx2

The sequences coding for the C-terminal part of the protein excluding NLS3 were amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 20 µL reaction mixture containing 5 ng of pVD43, pVD58-AgeI or pVD80 (see Chapter 2, Table 2.3a), 0.5 µM of primers AM37 and AZ203 (Table 3.3.1), 0.025 U/µL of TakaraExTaq™ DNA Polymerase (Takara Bio USA), 1xExTaq™ Buffer (Takara Bio USA), 200 µM of each dNTP Mixture (Takara Bio USA). PCR parameters were as follows: one cycle at 98°C for 30 s, 30 cycles at 98°C for 10 s, 62°C for 30 s, 72°C for 30 s for pVD43 and 10 s for pVD58-AgeI and pVD80, and one cycle at 72°C for 2 min. The *Xho*I-digested PCR products were cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-VirD2-C1/3ΔNLS3, pEZT-NL-mVirD2-58-C1/3ΔNLS3 or pEZT-NL-mVirD2-80-C1/3ΔNLS3, respectively (Table 3.4.1).

The sequences coding for the alanine linker and EGFP were amplified and flanked by *Eco*RI and *Sma*I sites (Section 3.3.1.5.2). The *Eco*RI/*Sma*I-digested PCR product was cloned into the *Eco*RI/*Sma*I-digested pEZT-NL-mVirD2-80-C1/3ΔNLS3 resulting in pEZT-NL-mVirD2-80-C1/3ΔNLS3-EGFPx2 (Table 3.4.1).

3.3.1.6. Construction of pEZT-NL-VirD2-M1/3 carrying the central region of the *virD2* gene

The central region of *virD2* encoding the central/middle part of VirD2 was amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 25 µL reaction mixture containing 5 ng of pVD43 (Chapter 2, Table 2.3a), 0.5 µM of primers AM26 and AM27 (Table 3.3.1), 0.02 U/µL of Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs), 1xQ5 Reaction Buffer (New England Biolabs), 200 µM of each dNTP Mixture. PCR parameters were as follows: one cycle at 98°C for 30 s, 30 cycles at 98°C for

10 s, 60°C for 30 s, 72°C for 10 s, and one cycle at 72°C for 2 min. The *Xho*I-digested PCR product was cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-VirD2-M1/3 (Table 3.4.1).

3.3.1.7. Construction of pEZT-NL-NLS3 derivatives: pEZT-NL-NLS3, pEZT-NL-NLS3-EGFPx2 and pEZT-NL-EGFPx3

The putative NLS3 coding sequence was amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 20 µL reaction mixture containing 5 ng of pVD43 (see Chapter 2, Table 2.3a), 0.5 µM of primers AM29 and AM30 (Table 3.3.1), 0.025 U/µL of TakaraExTaq™ DNA Polymerase (Takara Bio USA), 1xExTaq™ Buffer (Takara Bio USA), 200 µM of each dNTP Mixture (Takara Bio USA). PCR parameters were as follows: one cycle at 98°C for 30 s, 30 cycles at 98°C for 10 s, 60°C for 30 s, 72°C for 10 s, and one cycle at 72°C for 2 min. The *Xho*I-digested PCR product (84 bp) was cloned into the *Xho*I cut pEZT-NL resulting in pEZT-NL-NLS3 (Table 3.4.1).

The sequences coding for the alanine linker and EGFP were amplified and flanked by *Hind*III and *Sma*I sites by PCR. PCR was performed in a 25µL reaction mixture containing 10 ng of pEZT-NL, 0.5 µM of primers AM56 and AM57 (Table 3.3.1), 0.02 U/µL of Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs), 1xQ5 Reaction Buffer (New England Biolabs), 200 µM of each dNTP Mixture. PCR parameters were as follows: one cycle at 98°C for 30 s, 30 cycles at 98°C for 10 s, 67°C for 30 s, 72°C for 10 s, and one cycle at 72°C for 2 min. The *Hind*III/*Sma*I-digested PCR product was cloned into the *Hind*III/*Sma*I-cut pEZT-NL resulting in pEZT-NL-NLS3-EGFPx2 (Table 3.4.1).

The sequences coding for the alanine linker and EGFP were amplified and flanked by *Sma*I sites by PCR. PCR was performed in a 25 µL reaction mixture containing 10 ng of pEZT-NL, 0.5 µM of primers AM63 and AM64 (Table 3.3.1), 0.02 U/µL of Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs), 1xQ5 Reaction Buffer (New England Biolabs), 200 µM of each dNTP Mixture. PCR parameters were as follows: one cycle at 98°C for 30 s, 30 cycles at 98°C for 10 s, 68°C for 30 s, 72°C for 10 s, and one cycle at 72°C for 2 min. The *Sma*I-digested PCR product was cloned into the *Sma*I-cut pEZT-NL-NLS3-EGFPx2 resulting in pEZT-NL-NLS3-EGFPx3 (Table 3.4.1).

3.3.1.8. Construction of pEZT-NL-78bp from pCAMBIA1302 derivatives: pEZT-NL-78bp, pEZT-NL-78bp-EGFPx2 and pEZT-NL-78bp-EGFPx3

The first 78 bp of the *gfp* gene in pCAMBIA1302 (http://www.cambia.org/daisy/cambia/585), a kind gift from Dr. Monroe (James Madison University, Virginia), was amplified and flanked by *Xho*I sites by PCR. PCR was performed in a 25 µL reaction mixture containing 10 ng of pCAMBIA1302, 0.5 µM of primers AM70 and AM71 (Table 3.3.1), 0.02 U/µL of Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs), 1xQ5 Reaction Buffer (New England Biolabs), 200 µM of each dNTP Mixture. PCR parameters were as follows: one cycle at 98^{o}C for 30 s, 30 cycles at 98^{o}C for 10 s, 60^{o}C for 30 s, 72^{o}C for 10 s, and one cycle at 72^{o}C for 2 min. The *Xho*I-digested PCR product was cloned into the *Xho*I-cut pEZT-NL resulting in pEZT-NL-78bp (Table 3.4.1).

The sequences coding for the alanine linker and EGFP were amplified and flanked by *Hind*III and *Sma*I sites using pEZT-NL and primers AM56 and AM57 (Table 3.3.1). PCR parameters were mentioned previously (Section 3.3.1.7). The *Hind*III/*Sma*I-digested PCR product was cloned into the *Hind*III/*Sma*I-cut pEZT-NL-78bp resulting in pEZT-NL-78bp-EGFPx2 (Table 3.4.1).

The sequences coding for the alanine linker and EGFP were amplified and flanked by *Sma*I sites using pEZT-NL and primers AM63 and AM64 (Table 3.3.1) with the aforementioned PCR parameters (Section 3.3.1.7). The *Sma*I-digested PCR product was cloned into the *Sma*I-cut pEZT-NL-78bp-EGFPx2 resulting in pEZT-NL-78bp-EGFPx3 (Table 3.4.1).

3.3.2. Protein analyses using the online software

3.3.2.1. Identification of a novel NLS (NLS3)

Amino acid sequences of VirD2, mVirD2-58 and mVirD2-80 were analyzed on cNLS Mapper (Kosugi et al., 2009b).

3.3.2.2. Protein size prediction

Molecular weights of the recombinant proteins were predicted using Science Gateway (http://www.sciencegateway.org/tools/proteinmw.htm).

3.3.2.3. Sequence alignment

The VirD2 amino acid sequences from several octopine and nopaline type plasmids were aligned using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

3.3.3. Growth conditions of plants

3.3.3.1. The growth condition of *Arabidopsis thaliana*

Seeds of *Arabidopsis thaliana* (cultivar Columbia) were sowed on the mixture of *All purpose potting soil* (Plant *Etc*; Lethbridge, AB, Canada) and vermiculite (The Professional Gardener Co LTD, AB, Canada) in a 4:1 ratio in 2x2 inch square pots. Miracle-Gro fertilizer (Scotts Canada *Ltd.*, Mississauga, ON, Canada) was used to pre-soak the soil as described previously (Bilichak, 2013).The seeds were placed at 4^{o}C for 2 days to break the dormancy; subsequently, they were moved to a growth chamber (Biochambers, model SPRS-1115, Winnipeg, MB, Canada), set at high light conditions (32.8 $\mu Em^{-2}s^{-1}$) at 22°C under a 16-hour light regime and at 18°C under an 8-hour dark regime under a constant humidity of 65%. The soil was continuously moistened with tap water.

3.3.3.2. The growth condition of *Nicotiana benthamiana*

N. benthamiana seeds were placed on the previously mentioned soil mixture (Section 3.3.3.1) in 5x5 square pots which were then placed in a growth chamber (Biochambers, model SPRS-1115, Winnipeg, MB, Canada) with the aforementioned settings (Section 3.3.3.1). Five to six week-old plants were used for the following *Agrobacterium* infiltration.

3.3.4. *Agrobacterium*-mediated plant transformation

3.3.4.1. Stable transformation of *Arabidopsis* by the *Agrobacterium*-mediated floral dip method

Arabidopsis was stably transformed by the floral dip method as previously described with some modifications (Zhang et al., 2006). *Agrobacterium* strain GV3101[pPM6000] was transformed with pEZT-NL derivatives by electroporation (Mattanovich et al., 1989). Bacteria were plated on YEP medium (10 g peptone, 10 g yeast extract, 5 g $NaCl_2$, 10 mM $MgSO_4$, 20 g agar per litre, pH 7) containing rifampicin (100 mg/L), carbenicillin (25 mg/L), and gentamicin (25 mg/L). Subsequently, the bacteria were grown in 5 mL

YEP containing the aforementioned antibiotics at 28^{o}C with shaking overnight. One mL overnight culture was transferred to 200 mL YEP with the aforementioned antibiotics and then incubated at 28^{o}C for 16-24 h. The bacterial cells were harvested at 1,500xg for 20 min at room temperature. The bacterial pellet was re-suspended in the infiltration medium (5% sucrose) to OD_{600} 0.8. Subsequently, 0.005% Silwet L-77 (Lahile seeds, USA) was added to the infiltration medium. Plants with floral buds were submerged in the inoculum for 1 min. They were kept in the dark for 48 h supplemented with water in the greenhouse. Afterward, the plants were grown until siliques were developed. Seeds (T1) were harvested and then dried overnight at 37^{o}C.

3.3.4.2. Selection of transgenic *Arabidopsis* by glufosinate ammonium

The selection was made according to the previously published protocol (Bilichak, 2013). T1 seeds harvested from the floral dip were sowed on the soil mix pre-soaked with a fertilizer in 2-inch deep pots. The seeds were placed at 4^{o}C for 2 days and then moved to the growth chamber (Section 3.3.3.1). After two or three days of germination, seedlings were sprayed with the Liberty 150 CN Herbicide diluted 1000x and Crop Desiccant stock solution (Aventis CropSciense Canada Co, Canada, 150 g/L concentration of glufosinate ammonium in the stock solution) once a day for a minimum of 7 days continuously. Transgenic plants were green, while non-transgenic plants were smaller than transgenic plants and had bleached cotyledons. The transgenic plants were transferred into individual pots for further growth. T2 seeds were harvested.

3.3.4.3. DAPI staining of transgenic *Arabidopsis* seedlings

T2 *Arabidopsis* seeds were sowed on the soil mix pre-soaked with a fertilizer in 2x2 inch square pots. The seeds were placed at 4^{o}C for 2 days and then moved to the growth chamber (Section 3.3.3.1). Leaves of 2.5 week-old seedlings were submerged in the DAPI solution [1 µg/mL4,6-diamidino-2-phenylindole (SIGMA), 1% v/v Triton X-100 (EM Science)] and vacuum infiltrated for 5 min. The leaves were rinsed with distilled water three times and examined under the confocal microscope.

3.3.4.4. Transient transformation of tobacco leaves by the *Agrobacterium*-infiltration method

Transient expression of recombinant proteins in tobacco leaves by *Agrobacterium* infiltration was done according to the established protocol (Sparkes et al., 2006) with the following modifications. *Agrobacterium* GV3101[pPM6000] was transformed with pEZT-NL derivatives by electroporation (Mattanovich et al., 1989). The bacteria were plated on YEP medium containing rifampicin (100 mg/L), carbenicillin (25 mg/L), and gentamicin (25 mg/L). They were used to inoculate 5 mL YEB (5 g beef extract, 5 g peptone, 5 g sucrose, 1 g yeast extract, 2 mM $MgSO_4$ per litre, pH 7.5) with the aforementioned antibiotics, and the culture was incubated at 28^{o}C with shaking overnight. The overnight culture (1.5 mL) was centrifuged at 1,000xg for 10 min at room temperature. The bacterial pellet was resuspended in 1 mL of the infiltration medium (0.5% w/v D-glucose, 50 mM MES, 2 mM Na_3PO_4•$12H_2O$, 0.1 mM acetosyringone). The cell suspension was diluted to OD_{600} 0.05with the infiltration medium. The abaxial side between two veins of a leafwas infiltrated with the OD adjusted cell suspension. The plants were placed in the growth chamber with the previously mentioned settings. Two days post infiltration, a small piece of the infiltrated region of the leaf was examined under the confocal microscope described below.

3.3.4.5. DAPI staining of infiltrated tobacco leaves

Small pieces of the infiltrated region of tobacco leaves were examined under the confocal microscope to ensure the EGFP expression before DAPI staining. After the presence of EGFP was detected, the same leaf samples were incubated in the DAPI solution for 30 min at room temperature (Section 3.3.4.3). The leaves samples were rinsed with distilled water three times and then examined under the confocal microscope.

3.3.4.6. Confocal laser scanning microscopy

The leaf samples were mounted in water and analyzed under the Nikkon D-Eclipse C1 confocal microscope (Nikon) equipped with following lasers and detectors: a UV laser (408 nm) and its detector (450/35) for the DAPI stained nuclei, an argon laser (488 nm) and its detector (515/30) for the fluorescence from EGFP, a helium-neon laser (638 nm) and its detector (605/75) for the autofluorescence from chloroplasts. A differential interference contrast was also used. The laser and detector settings were

identical within *Arabidopsis* leaf samples. Tobacco leaves prior to the DAPI staining were examined at a constant setting. After staining, a detector (605/75) for a helium-neon laser (638 nm) was reduced to decrease the background noise. All other settings were the same before DAPI staining, and the settings were identical in all tobacco leaf samples. The color of some pictures was enhanced using Auto Correct in Microsoft Office Picture Manager to emphasize the subcellular localization of recombinant proteins. Original photos are presented in Figure 3.4.3.2e.

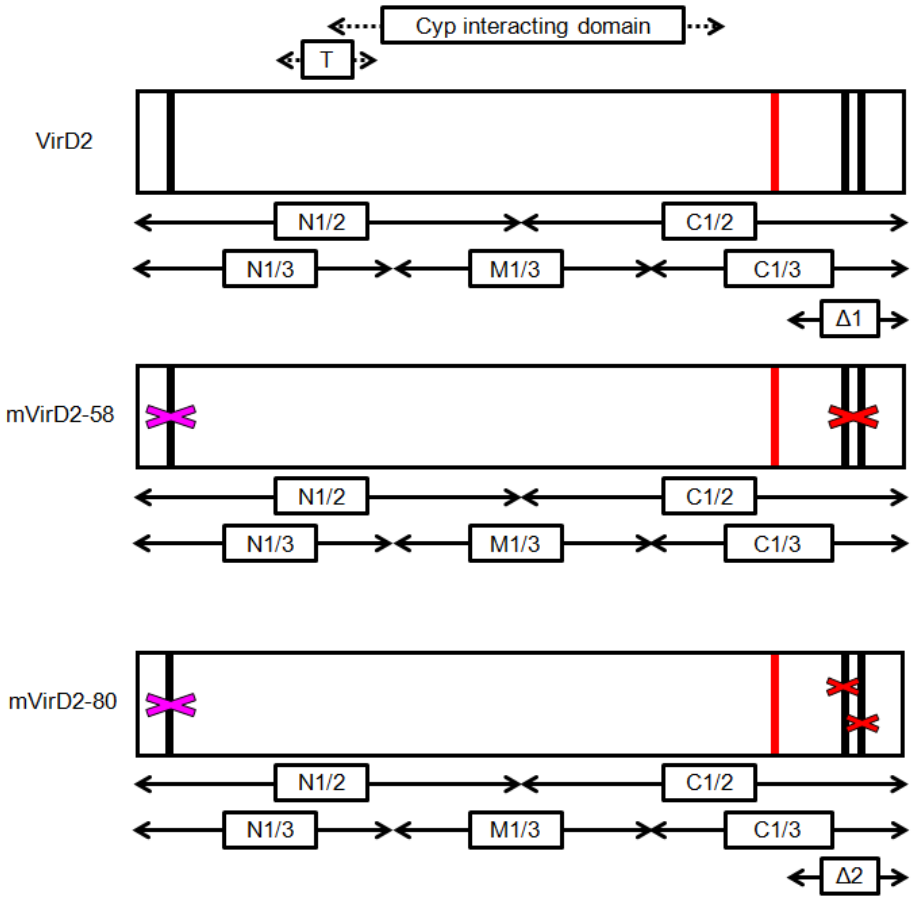


Figure 3.3.1. Fragments of VirD2, mVirD2-58 and mVirD2-80.
Black bars: N-terminal NLS and C-terminal NLS;magenta cross: modified N-terminal NLS; red cross in mVirD2-58: a broad deletion of the C-terminal NLS; red crosses in mVirD2-80: a precise deletion of the C-terminal NLS; red bar: NLS3; Δ1: VirD2-C1/3ΔNLS3; Δ2: mVirD2-80-C1/3ΔNLS3.N1/2: an N-terminal half of VirD2; C1/2: a C-terminal half; N1/3: an N-terminal one third of VirD2; M1/3: one third representing central/middle region of VirD2; C1/3: a C-terminal one third of VirD2.
Motif/Domains shown: T: HIT motif to interact with TATA box-binding protein; CyP interacting domain: cyclophilin interacting domain.

Table 3.3.1. Primers used to construct pEZT-NL derivatives.

Primer	Sequence (5'-3')	Description
AZ202	TATA*CTCGAG*ATGCCCGATCGCGCTC	The forward primer which anneals to the beginning of the *virD2* gene and to introduce the *Xho*I site
AZ203	ATAT*CTCGAG*GGTCCCCCCGCGC	The reverse primer which anneals to the end of the *virD2* gene and to introduce *Xho*I site
AZ204	ATAT*CTCGAG*ATGGCTTCAATTTGTAC	The forward primer which anneals to the beginning of the *DCL-TP* sequence and introduce the *Xho*I site
AZ205	ATAT*CTCGAG*ATGGCTTCCTCAGTTC	The forward primer which anneals to the beginning of the *RbcS-TP* sequence and introduce the *Xho*I site
AZ245	ATAT*CTCGAG*ATGGCTTCCTCAGTTCTTTC CTC	The forward primer which anneals to the beginning of the *RbcS-TP*sequence and introduce the *Xho*I site
AZ267	AAA*CTCGA*GACCGGTCTTCACCGCCGCGCA ACG	The reverse primer which anneals to the end of the *DCL-TP* sequence and introducethe*Xho*I site
AZ268	AAA*CTCGAG*ACCGGTTGGCCACACCTGCAT GCA	The reverse primer which anneals to the end of the *RbcS-TP* gene and introduce the *Xho*I site
AM23	TTA*CTCGAG*TTCGAATTGAATCTTTTG	The reverse primer which anneals to the middle of *virD2* gene and introduce the *Xho*I site
AM24	AAA*CTCGAG***ATG**GATACAGATTTTGATGA	The forward primer which anneals to the middle of *virD2* and introduce the *Xho*I site and a new **ATG** codon
AM25	TTT*CTCGAG*CCGACGATTGACCACGAC	The reverse primer to amplify the N-terminal 1/3 of VirD2 coding sequence and introduce the *Xho*I site
AM26	ATA*CTCGAG***ATG**GAACTTCTGGGGCACGG	The forward primer to amplify the middle of the *virD2* gene and introduce the *Xho*I site and a new **ATG** codon
AM27	AAA*CTCGAG*CTCGCTCTCCAATGATAC	The reverse primer to amplify the middle of the *virD2* gene and introduce the *Xho*I site
AM28	ATA*CTCGAG***ATG**CGGAGTGCCCAACCATC	The forward primer to amplify the C-terminal 1/3 of VirD2 coding sequence and introduce the *Xho*I site and a new **ATG** codon
AM29	ATA*CTCGAG***ATG**AAG CGCCAGCAGCG	The forward primer to amplify the *NLS3*coding sequence and introduce the *Xho*I site and a new **ATG** codon
AM30	AAA*CTCGAG*CTTCAATCCTTTACGGTTTG	The reverse primer to amplify the *NLS3* coding sequence and introduce the *Xho*I site
AM37	ATA*CTCGAG***ATG**GCTGCGCAAGTTGATTC	The forward primer which anneals to just after the *NLS3* coding sequence and introduce the *Xho*I site and a new **ATG** codon
AM42	ATA*GAATTC*CGGCTGCTGCCGCTGCC	The forward primer to amplify *Ala*(10)-*EGFP* sequence and introduce the *Eco*RI site

Table 3.3.1. Continued.

Primer	Sequence (5'-3')	Description
AM43	AAA*CCCGGG***C**CTTGTACAGCTCGTCCA	The reverse primer to amplify *Ala*(*10*)-*EGFP* sequence without a stop codon, introduce the *Sma*I site and add **C**to make in frame
AM56	TTT*AAGCTT*GCTGCTGCCGCTGCCGCTG	The forward primer to amplify *Ala*(*10*)-*EGFP* sequence and introduce the *Hind*III site
AM57	TATT*CCCGGG***C**CTTGTACAGCTCGTCCA	The reverse primer to amplify *Ala*(*10*)-*EGFP* sequence gene without a stop codon, introduce the *Sma*I site and add C to make in frame
AM63	ATA*CCCGGG***TA**GCTGCTGCCGCTGCCGC	The forward primer to amplify *Ala*(*10*)-*EGFP* sequence gene, and introduce the *Sma*I site and add **TA** to make in frame
AM64	AAT*CCCGGG***T**CTTGTACAGCTCGTCCATG	The reverse primer to amplify *Ala*(*10*)-*EGFP* sequence gene, and introduce the *Sma*I site and add **T** to make in frame
AM70	AAA*CTCGAG*ATGGTAGATCTGACTAGTAAA G	The forward primer which anneals to the beginning of the *gfp* sequence in pCAMBIA1302 and introduce the *Xho*I site
AM71	ATA*CTCGAG*AACATCACCATCTAATTCAAC	The reverse primer which anneals to a part of the *gfp* sequence in pCAMBIA1302 and introduce the *Xho*I site

The introduced restriction sites are indicated in *italics*. Added nucleotides are **bold**.

3.4. Results

In the previous chapter, we demonstrated that strong transit peptides were overridden by an unknown factor that did not allow the T-complex to reach the chloroplasts. Thus, we attempted to analyze this unknown factor. We examined the subcellular localization of fusion proteins, TP-mVirD2-EGFP and mVirD2-EGFP lacking known VirD2 NLSs. Several regions of VirD2 and mVirD2 were fused to EGFP to define a putative NLS.

3.4.1. Construction of pEZT-NL derivatives containing the *DCL-TP, RbcS-TP, virD2, mvirD2,* a partial *virD2* or partial *mvirD2* genes

The modifications of the N and C-terminal NLSs were introduced in our previous study (see Chapter 2). Previously constructed pVD58 derivative contains the *mvirD2-58* encoding the N-terminal modified NLS (change from arginine to threonine) and a broad deletion of the C-terminal NLS (removing 45 amino acids including the C-terminal NLS by *Nru*I). In addition, pVD80 derivative contains the *mvirD2-80* encoding the aforementioned N-terminal modified NLS and a precise deletion of the C-terminal NLS (exclusively eliminating bipartite sequences). The tomato *DCL-TP* or tobacco *RbcS-TP* sequence was cloned into pVD58 and pVD80 derivatives resulting in pVD58-DCL-TP, pVD58-RbcS-TP, pVD80-DCL-TP and pVD80-RbcS-TP (see Chapter 2, Table 2.3a).

The *DCL-TP*, *RbcS-TP*, *virD2*, and *mvirD2* genes in pVD derivatives were cloned into pEZT-NL resulting in various pEZT-NL derivatives (Table 3.4.1). The sequence coding for the N-terminal half of VirD2 was cloned from pVD43 carrying the wild type *virD2* gene into pEZT-NL resulting in pEZT-NL-VirD2-N1/2. pVD58 and pVD80 derivatives contain the same sequence coding for the N-terminal half; thus, the sequence coding for the N-terminal half or the N-terminal one third of mVirD2 was cloned into pEZT-NL resulting in pEZT-NL-mVirD2-58/80-N1/2 or pEZT-NL-mVirD2-58/80-N1/3, respectively.

The central (middle) regions of VirD2 and mVirD2 named M1/3 are identical in both proteins (Figure 3.3.1). This central coding sequence was cloned into pEZT-NL resulting in pEZT-NL-VirD2-M1/3 (Table 3.4.1).

The sequences coding for the C-terminal half or one third of the VirD2 and mVirD2 were cloned into pEZT-NL resulting in pEZT-NL-VirD2-C1/2, pEZT-NL-VirD2-C1/3, pEZT-NL-mVirD2-58-C1/2,

pEZT-NL-mVirD2-58-C1/3, pEZT-NL-mVirD2-80-C1/2 and pEZT-NL-mVirD2-80-C1/3 (Figure 3.3.1 and Table 3.4.1). In addition, an extra alanine linker and EGFP coding sequence were cloned into pEZT-NL containing the C-terminal VirD2 coding sequence; pEZT-NL-VirD2-C1/3-EGFPx2 and pEZT-NL-mVirD2-80-C1/3-EGFPx2 were successfully constructed (Table 3.4.1).

The amino acid sequences of VirD2 and mVirD2 were analyzed using cNLS Mapper (Figure 3.4.1a-c). The N-terminal NLS was not recognized by the program, while the C-terminal NLS of VirD2 was predicted to be a strong bipartite NLS (score 13.6 in Figure 3.4.1a). VirD2, mVirD2-58, and mVirD2-80 were predicted to have another monopartite type NLS named NLS3 (score 7 in Figure 3.4.1a-c).

After recognition of NLS3, we attempted to clone the sequence coding only for the NLS3 core region (QRSKRRNDEE, 10 amino acids) or shorter NLS3 (RSKRRN, 6 amino acids) into pEZT-NL, yet we were unsuccessful either with cloning or with transit expression in tobacco leaves (data not shown). Therefore, the sequence coding for 26 amino acids including the NLS3 core sequence with the flanking amino acids and a newly added methionine (**M**KRQ**QRSKRRNDEE**AGPSGANRKGLK, 26 aa) was cloned into pEZT-NL resulting in pEZT-NL-NLS3 (Table 3.4.1). Hereafter, NLS3 in our constructs designates the 26 amino acid. One or two extra alanine linker and EGFP sequences were cloned into pEZT-NL-NLS3 resulting in pEZT-NL-NLS3-EGFPx2 or pEZT-NL-NLS3-EGFPx3, respectively (Table 3.4.1).

The C-terminal parts of VirD2 and mVirD2 were revisited after the NLS3 recognition by cNLS Mapper. The sequences coding for the C-terminal part of VirD2, mVirD2-58, and mVirD2-80 excluding 25 amino acids containing NLS3 were cloned into pEZT-NL resulting in pEZT-NL-VirD2-C1/3ΔNLS3, pEZT-NL-mVirD2-58-C1/3ΔNLS3 and pEZT-NL-mVirD2-80-C1/3ΔNLS3 (Table 3.4.1).

Since pEZT-NL does not express EGFP well without addition of the coding sequence (http://deepgreen.stanford.edu/), the 78 bp coding for first 26 amino acids of GFP in pCAMBIA1302 was cloned into pEZT-NL to serve as the nuclear targeting negative control resulting in pEZT-NL-78bp. One or two extra alanine linker and EGFP coding sequences were cloned into pEZT-NL-78bp resulting in pEZT-NL-78bp-EGFPx2 or pEZT-NL-78bp-EGFPx3, respectively (Table 3.4.1).

Table 3.4.1. Description of pEZT-NL derivatives, recombinant proteins and their subcellular localization.

Binary vector	Recombinant protein	Protein size (kDa)	Subcellular localization	
			Arabidopsis (Stable)	Tobacco (Transient)
pEZT-NL carrying the *TP*sequence				
pEZT-NL-DCL-TP	DCL-TP-EGFP	36	Chloroplast	Chloroplast
pEZT-NL-RbcS-TP	RbcS-TP-EGFP	37	Chloroplast	Chloroplast
pEZT-NL carrying the *TP-mvirD2* gene				
pEZT-NL-DCL-TP-mVirD2-58	DCL-TP-mVirD2-58-EGFP	78	Nucleus	Nucleus
pEZT-NL-RbcS-TP-mVirD2-58	RbcS-TP-mVirD2-58-EGFP	79	Mainly in the nucleus Little in chloroplasts	Mainly in the nucleus Little in chloroplasts
pEZT-NL-DCL-TP-mVirD2-80	DCl-TP-mVirD2-80-EGFP	81	Nucleus	Mainly in the nucleus Some in the cytoplasm
pEZT-NL-RbcS-TP-mVirD2-80	RbcS-TP-mVirD2-80-EGFP	82	Mainly in the nucleus Some in chloroplasts	Mainly in the nucleus Some in chloroplasts
pEZT-NL carrying the *virD2* or *mvirD2* gene				
pEZT-NL-VirD2	VirD2-EGFP	78	Not viable	No expression
pEZT-NL-mVirD2-58	mVirD2-58-EGFP	73	Nucleus	Nucleus and aggregates in the cytoplasm
pEZT-NL-mVirD2-80	mVirD2-80-EGFP	76	Nucleus	Nucleus and aggregates in the cytoplasm
pEZT-NL carrying the sequence coding for the N-terminal partial VirD2 or partial mVirD2				
pEZT-NL-VirD2-N1/2	VirD2-N1/2-EGFP	54	Nucleus and cytoplasm	Aggregates in the cytoplasm
pEZT-NL-mVirD2-58/80-N1/2	mVirD2-58/80-N1/2-EGFP	54	Unsuccessful	Aggregates in the cytoplasm
pEZT-NL-VirD2-N1/3	VirD2-N1/3-EGFP	46	Unsuccessful	Nucleus and cytoplasm Aggregates in the cytoplasm
pEZT-NL-mVirD2-58/80-N1/3	mVirD2-58/80-N1/3-EGFP	46	Unsuccessful	Nucleus and cytoplasm Aggregates in the cytoplasm

Table 3.4.1. Continued.

Binary vector	Recombinant protein	Protein size (kDa)	Subcellular localization	
			Arabidopsis (Stable)	Tobacco (Transient)
pEZT-NL carrying the sequence coding forthe central region of VirD2				
pEZT-NL-mVirD2-M1/3	mVirD2-M1/3-EGFP	47	Unsuccessful	No expression
pEZT-NL carrying C-terminal portion				
pEZT-NL-VirD2-C1/2	VirD2-C1/2	54	Unsuccessful	Nucleus
pEZT-NL-mVirD2-58-C1/2	mVirD2-58-C1/2-EGFP	49	Unsuccessful	Nucleus
pEZT-NL-mVirD2-80-C1/2	mVirD2-80-C1/2-EGFP	52	Nucleus and cytoplasm	Nucleus
pEZT-NL-VirD2-C1/3	VirD2-C1/3-EGFP	46	Unsuccessful	Nucleus
pEZT-NL-mVirD2-58-C1/3	mVirD2-58-C1/3-EGFP	41	Strong in the nucleus and the cytoplasm	Strong in the nucleus and moderately the cytoplasm
pEZT-NL-mVirD2-80-C1/3	mVirD2-80-C1/3-EGFP	44	Strong in the nucleus and the cytoplasm	Strong in the nucleus and moderately the cytoplasm
pEZT-NL-mVirD2-80-C1/3-EGFPx2	mVirD2-80-C1/3-EGFPx2	71	Not tested	Nucleus
pEZT-NL carrying 1/3 C-terminal ΔNLS3				
pEZT-NL-VirD2-C1/3ΔNLS3	VirD2-C1/3ΔNLS3-EGFP	38	Not tested	Nucleus
pEZT-NL-mVirD2-58-C1/3ΔNLS3	mVirD2-58-C1/3 ΔNLS3-EGFP	34	Not tested	Nucleus and cytoplasm, more in cytoplasm than mVirD2-58-C1/3-EGFP
pEZT-NL-mVirD2-80-C1/3ΔNLS3	mVirD2-80-C1/3 ΔNLS3-EGFP	36	Not tested	Nucleus and cytoplasm, more in cytoplasm than mVirD2-80-C1/3-EGFP
pEZT-NL-mVirD2-80-C1/3ΔNLS3-EGFPx2	mVirD2-80-C1/3 ΔNLS3-EGFPx2	63	No tested	Nucleus and cytoplasm, similar to mVirD2-80-C1/3-EGFP localization

Table 3.4.1. Continued.

Binary vector	Recombinant protein	Protein size (kDa)	Subcellular localization	
			Arabidopsis (Stable)	Tobacco (Transient)
pEZT-NL carrying NLS3				
pEZT-NL-NLS3	NLS3-EGFP	33	Not tested	Nucleus and cytoplasm
pEZT-NL-NLS3-EGFPx2	NLS3-EGFPx2	60	Not tested	Mostly in the nucleus Some in the cytoplasm
pEZT-NL-NLS3-EGFPx3	NLS3-EGFPx3	89	Not tested	Nucleus
pEZT-NL carrying 78bp from pCAMBIA1302				
pEZT-NL-78bp	26aa-EGFP	33	Not tested	Nucleus and cytoplasm
pEZT-NL-78bp-EGFPx2	26aa-EGFPx2	60	Not tested	Nucleus and cytoplasm
pEZT-NL-78bp-EGFPx3	26aa-EGFPx3	88	Not tested	Nucleus and cytoplasm

MPDRAQVIIRIVPGGGTKTLQQIINQLEYLS**RKGK**LELQRSARHLDIPVPPDQIRELA
N-terminal NLS
QSWVTEAGIYDESQSDDDRQQDLTTHIIVSFPAGTDQTAAYEASREWAAEMFGSGYGGG
RYNYLTAYHVDRDHPHLHVVVNRRELLGHGWLKISRRHPQLNYDGLRKKMA
EISLRHGIVLDATSRAERGIAERPITYAEHRRLERMQAQKIQFEDTDFDETSPEEDRRD
LSQSFDPFRSDPSTGEPDRATRHDKQPLEQHARFQESAGSSIKADARIRVSLESERSAQ
PSASKIPVIGHFGIETSYVAEASVRKRSGIFGTSRPVTDVAMHTVKRQ**QRSKRRNDEE**
NLS3 (score 7)

AGPSGANRKGLKAAQVDSEANVGEQDTRDDSNKAADPVSASIGTEQPEASP**KRPR**D
C-terminal bipartite NLS
RHDGELGG**RKRAR**GNRRDDGRGGT
C-terminal bipartite NLS (score 13.6)

Figure 3.4.1a. VirD2 amino acid sequence analyzed by cNLS Mapper.
Letters in red: predicted NLSs. NLS3 was recognized as a monopartite type. The C-terminal NLS was recognized as a bipartite type. Bold letters: core NLS sequences. Yellow highlighted sequences (25 amino acids) were cloned into pEZT-NL. Score of >10 means that the protein is exclusively in the nucleus due to a strong NLS.

MPDRAQVIIRIVPGGGTKTLQQIINQLEYLS**TKGK**LELQRSARHLDIPVPPDQIRELA
Modified N-terminal NLS
QSWVTEAGIYDESQSDDDRQQDLTTHIIVSFPAGTDQTAAYEASREWAAEMFGSGY
GGGRYNYLTAYHVDRDHPHLHVVVNRRELLGHGWLKISRRHPQLNYDGLRKKMA
EISLRHGIVLDATSRAERGIAERPITYAEHRRLERMQAQKIQFEDTDFDETSPEEDRRD
LSQSFDPFRSDPSTGEPDRATRHDKQPLEQHARFQESAGSSIKADARIRVSLESERSAQ
PSASKIPVIGHFGIETSYVAEASVRKRSGIFGTSRPVTDVAMHTVKRQ**QRSKRRNDEE**
NLS3 (score 7)
AGPSGANRKGLKAAQVDSEANVGEQDTRDDGRGGT

Figure 3.4.1b. mVirD2-58 amino acid sequence analyzed by cNLS Mapper.
Letters in red: predicted NLSs. NLS3 was recognized as a monopartite type. Yellow highlighted sequences (25 amino acids) were cloned into pEZT-NL. Score of 7 means that the protein is mostly in the nucleus and some in the cytoplasm.

MPDRAQVIIRIVPGGGTKTLQQIINQLEYLS**TKGK**LELQRSARHLDIPVPPDQIRELA
Modified N-terminal NLS

QSWVTEAGIYDESQSDDDRQQDLTTHIIVSFPAGTDQTAAYEASREWAAEMFGSGY
GGGRYNYLTAYHVDRDHPHLHVVVNRRELLGHGWLKISRRHPQLNYDGLRKKMA
EISLRHGIVLDATSRAERGIAERPITYAEHRRLERMQAQKIQFEDTDFDETSPEEDRRD
LSQSFDPFRSDPSTGEPDRATRHDKQPLEQHARFQESAGSSIKADARIRVSLESERSAQ
PSASKIPVIGHFGIETSYVAEASVRKRSGIFGTSRPVTDVAMHTVKRQ**QRSKRRNDEE**
NLS3 (score 7)
AGPSGANRKGLKAAQVDSEANVGEQDTRDDSNKAADPVSASIGTEQPEASPDLGNR
RDDGRGGT

Figure 3.4.1c. mVirD2-80 amino acid sequence analyzed by cNLS Mapper.
Letters in red: predicted NLSs. NLS3 was recognized as a monopartite type. Yellow highlighted sequences (25 amino acids) were cloned into pEZT-NL. Score of 7 means that the protein is mostly in the nucleus and some in the cytoplasm.

3.4.2. Confocal microscopy of untransformed leaves of *Arabidopsis* and tobacco leaves

The same settings of the laser and its detector levels were used for all *Arabidopsis* leaves. Wild type *Arabidopsis* leaf was stained with DAPI and then examined under the confocal microscope. No EGFP was detected, as expected (Figure 3.4.2). DAPI stained nuclei and autofluorescence from chloroplasts were detected (Figure 3.4.2). Thus, 5 min of vacuum infiltration was sufficient to stain *Arabidopsis* nuclei. Transmission between different channels was not observed.

In the case of transiently transformed tobacco leaves, recombinant protein expression was first examined under the confocal microscope. Settings of the laser and its detector levels were kept constant for all leaf samples. With this setting, EGFP was not detected in the untransformed leaf (data not shown). The leaf samples were examined prior to the DAPI staining to check the expression of the recombinant proteins and to ensure that the staining would not alter the localization of the recombinant proteins. After DAPI staining using the same leaf samples, the samples were again observed under the confocal microscope. The DAPI stained nuclei and chloroplast autofluorescence were detected, but EGFP was not observed as expected (Figure 3.4.2). The staining procedure did not influence the localization of the recombinant proteins in all other transiently transformed leaves. The laser and its detector levels were optimally set to prevent transmission between the channels (Figure 3.4.2).

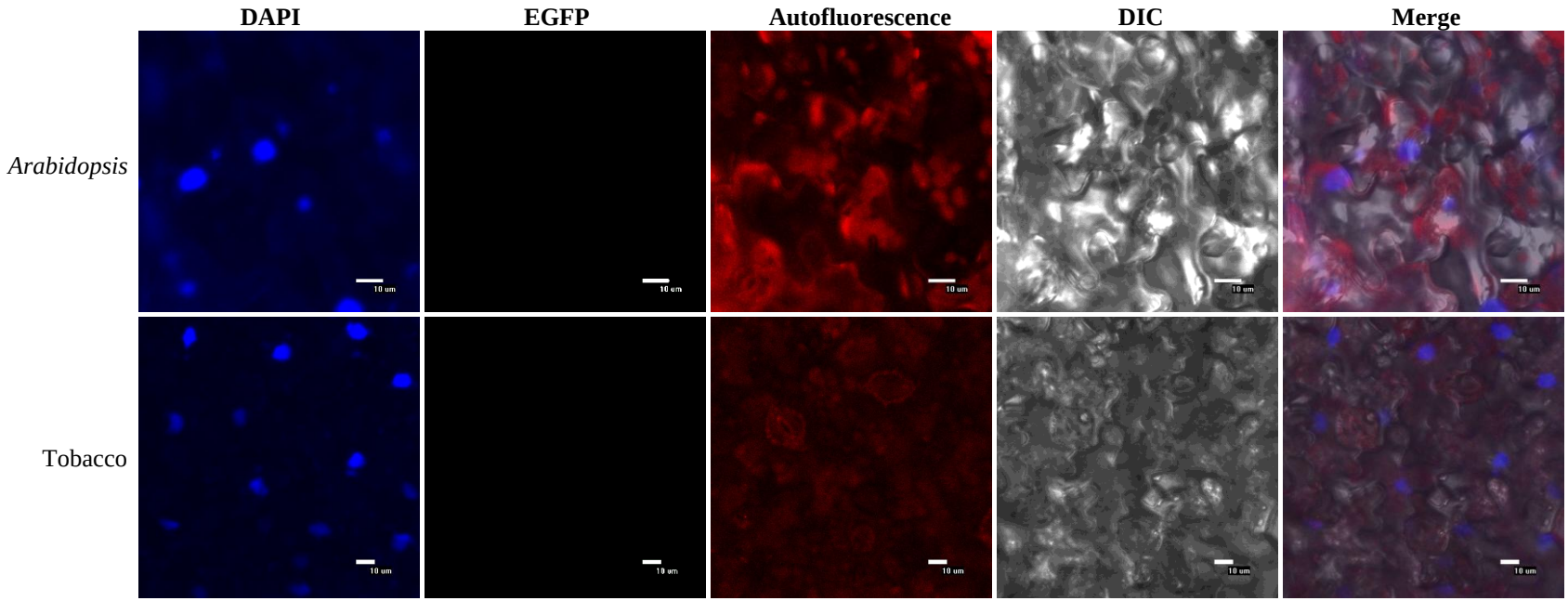


Figure 3.4.2. Untransformed wild type 2.5 weeks old transgenic *Arabidopsis* leaves (Top panels) and tobacco leaves (Bottom panels) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC. Scale bar = 10 µm.

3.4.3. Subcellular localization of the recombinant proteins in leaves in the stably transformed *Arabidopsis* plants and the transiently transformed tobacco leaves

pEZT-NL vector carries the T-DNA containing the bacterial *BIALOPHOS RESISTANCE* gene (*BAR*) encoding the enzyme phosphinotricin acetyl transferase (PAT) (http://deepgreen.stanford.edu/), which provides resistance to glufosinate ammonium (Weigel and Glazebrook, 2006). The genes of interest were cloned upstream of the alanine poly linker and EGFP coding sequence. The polylinker gives the flexibility between the protein of interest and EGFP to fold a fusion protein in a proper conformation. pEZT-NL derivatives were introduced to *Agrobacterium* strain GV3101[pPM6000]. *Arabidopsis* was transformed by the floral dip method using some constructs (Table 3.4.1). Transgenic *Arabidopsis*plants were screened using glufosinateammonium. The leaves of T2 seedlings were stained with DAPI and analyzed under the confocal microscope. In addition, *N.benthamiana* leaves were agro-infiltrated to see transient expression of the gene of interest. Subcellular localization of the recombinant proteins was observed 48 hours post infiltration.

3.4.3.1. Subcellular localization of DCL-TP-EGFP and RbcS-TP-EGFP in *Arabidopsis* and tobacco leaves

Tomato DCL and tobacco RbcS transit peptides were shown to direct the recombinant proteins into chloroplasts of various plants (Gnanasambandam et al., 2007). However, in our previous experiments, use of TP-mVirD2 derivatives did not result in the transformation of the chloroplast genome (see Chapter 2). Therefore, efficiency of DCL-TP and RbcS-TP chloroplast targeting was re-examined in our recombinant proteins and the transformation system.

The leaves of stably transformed *Arabidopsis* and agroinfiltrated tobacco leaves exhibited localization of DCL-TP-EGFP and RbcS-TP-EGFP in chloroplasts (Figure 3.4.3.1a and b). Despite the small size of the fusion proteins, less than 40 kDa (Table 4.4.1), diffusion of the recombinant proteins through the nuclear pore was not observed (Figure 3.4.3.1a and b). These results confirmed that DCL-TP and RbcS-TP were highly efficient to import recombinant proteins into chloroplasts.

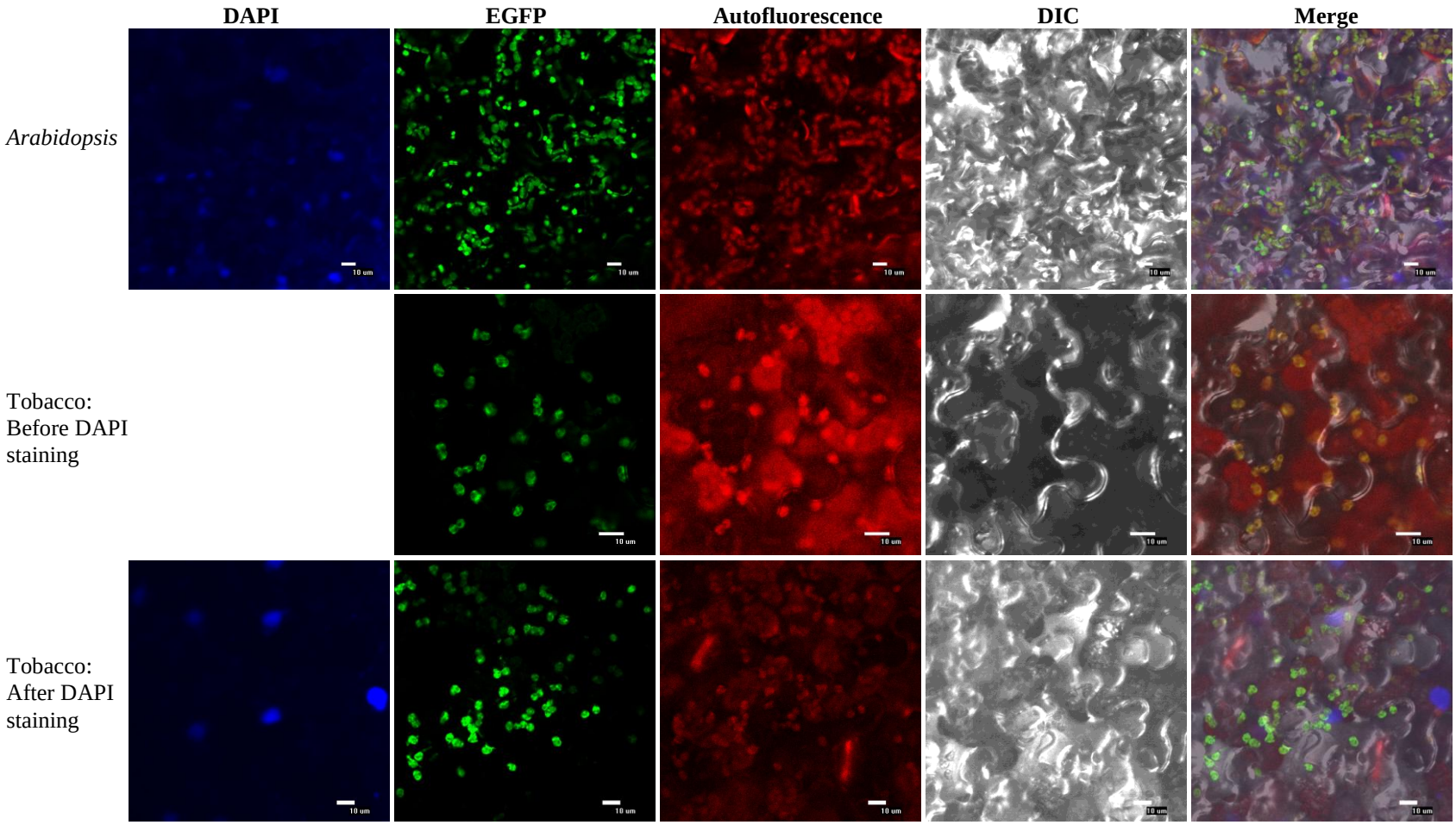


Figure 3.4.3.1a. DCL-TP-EGFP subcellular localization in transgenic *Arabidopsis* leaf (Top row) and in the tobacco leaves 48 h after agro-infiltration (Middle row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.
DAPI: DAPI stained nuclei, EGFP: recombinant protein fused to EGFP, Autofluorescence: chloroplast autofluorescence, DIC: differential interference contrast, Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC. Scale bar = 10 µm.

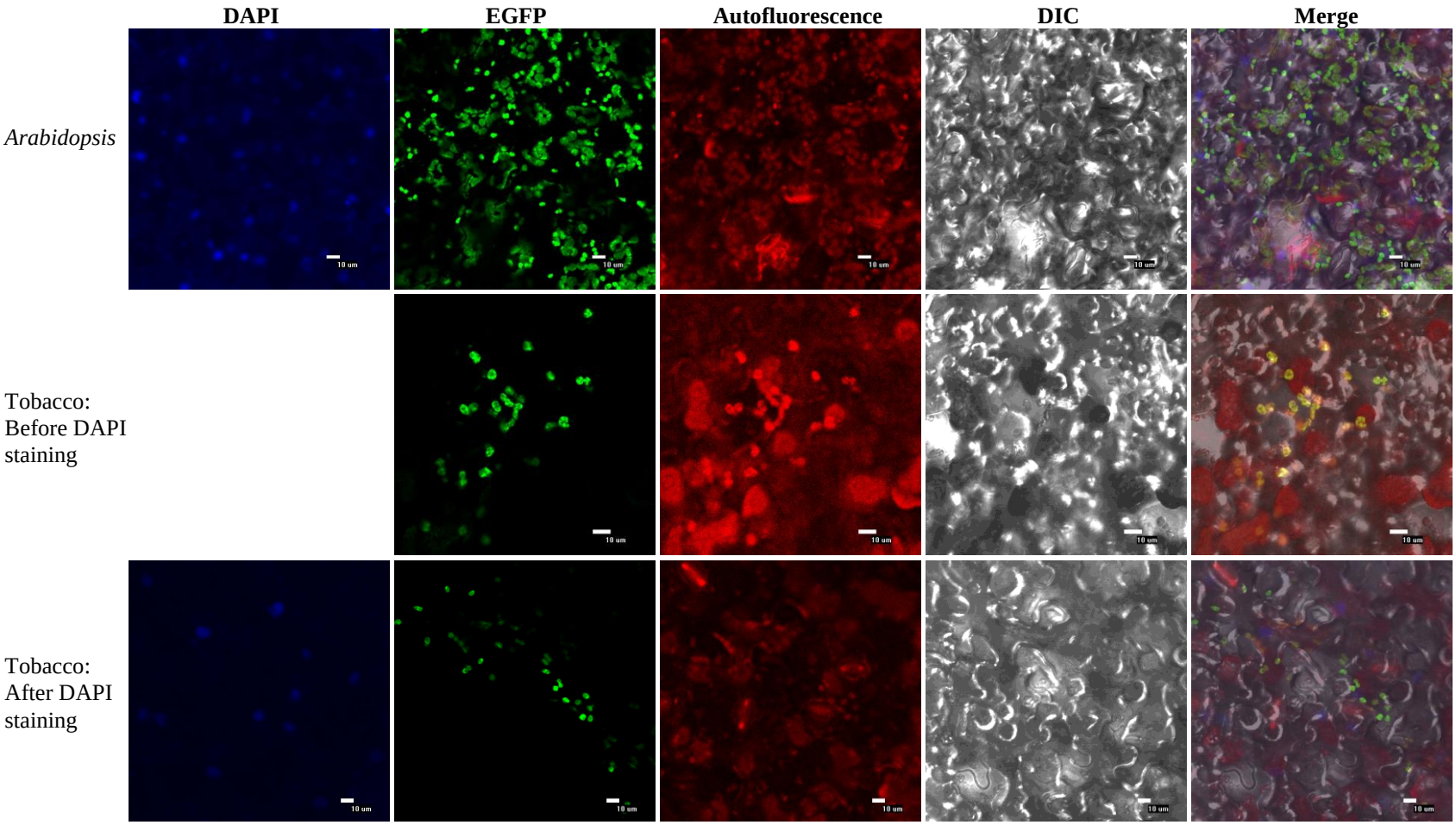


Figure 3.4.3.1b. RbcS-TP-EGFP subcellular localization in transgenic *Arabidopsis* leaf (Top row) and in the tobacco leaves 48 h after agro-infiltration (Middle row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC. Scale bar = 10 µm.

3.4.3.2. Subcellular localization of DCL-TP-mVirD2-58-EGFP, RbcS-TP-mVirD2-58-EGFP, DCL-TP-mVirD2-80-EGFP and RbcS-TP-mVirD2-80 in *Arabidopsis* and tobacco leaves

DCL-TP and RbcS-TP proved to be strong chloroplast-targeting signals in both *Arabidopsis* and tobacco leaves. Thus, these signals should be functional in fusion proteins of DCL-TP or RbcS-TP fused to mVirD2-58 carrying the modified N-terminal NLS and a broad deletion of the C-terminal NLS and mVirD2-80 carrying the modified N-terminal NLS and a precise deletion of the C-terminal NLS.

DCL-TP-mVirD2-58-EGFP was detected exclusively in the nucleus of *Arabidopsis* and tobacco leaf cells (Figure 3.4.3.2a). There was no transmission between EGFP and autofluorescence channels in the *Arabidopsis* leaf, whereas EGFP channel transmitted to autofluorescence detector in the tobacco leaf (* mark in Figure 3.4.3.2a). However, since the DAPI stained nucleus merged with EGFP, DCL-TP-mVirD2-58-EGFP was localized in the nucleus.

In the *Arabidopsis* and tobacco leaves, RbcS-TP-mVirD2-58-EGFP was localized mainly in the nucleus and very little in chloroplasts (arrows in Figure 3.4.3.2b). Also in the tobacco leaf, chloroplasts containing the recombinant protein were not detected in most areas (Figure 3.4.3.2b); thus, most of the recombinant protein was not imported into chloroplasts. Instead, it clearly localized in the tobacco leaf nuclei. Some nuclei did not show transmission from the EGFP to autofluorescence channel in the *Arabidopsis* leaf (** mark in Figure 3.4.3.2b). There was some transmission between EGFP and autofluorescence channels in tobacco leaves (* mark in Figure 3.4.3.2b). Despite transmission between the channels, overlap of the DAPI stained nuclei and EGFP suggested that RbcS-TP-mVirD2-58-EGFP was mainly localized in the nucleus and very little in chloroplasts of the *Arabidopsis* and tobacco leaves.

The expression of DCL-TP-mVirD2-80-EGFP in the *Arabidopsis*leaf was low. The recombinant protein was detected in the nucleus (Figure 3.4.3.2c); also very low intensity signals of EGFP were overlapping with the nuclei (open arrow in Figure 3.4.3.2c). There was no transmission between EGFP and autofluorescence channels in the *Arabidopsis* leaf. The recombinant protein was also detected mostly in the nucleus and little in the cytoplasm of the tobacco leaf cell (Figure 3.4.3.2c). Transmission occurred between EGFP and autofluorescence channels in the tobacco leaf (* mark in Figure 3.4.3.2c). Although slightly different localization of DCL-TP-mVirD2-80-EGFP was observed in the *Arabidopsis* and tobacco leaves, the recombinant protein was localized mainly in the nucleus.

RbcS-TP-mVirD2-80-EGFP was localized largely in the nucleus and slightly in chloroplasts in the *Arabidopsis* leaf (arrows in Figure 3.4.3.2d). However, there was a larger amount of the recombinant protein localized in the chloroplast as compared to RbcS-TP-mVirD2-58 protein in *Arabidopsis*. In the agro-infiltrated tobacco leaf, RbcS-TP-mVirD2-80-EGFP was also localized mainly in the nucleus and slightly in the chloroplasts (arrows in Figure 3.4.3.2d). DAPI was unable to stain the same cell as the observed one before the DAPI staining in Figure 3.4.3.2d, yet EGFP fluorescence was detected strongly in the nucleus as the DAPI staining merged with EGFP in both before and after DAPI stained cells (Figure 3.4.3.2d). EGFP did not transmit to the autofluorescence channel in tobacco before DAPI staining, but there was slight transmission between EGFP and autofluorescence channels in the *Arabidopsis* and tobacco leaves after the DAPI staining (* mark in Figure 3.4.3.2d). Regardless of the slight transmission between two channels, overlay of DAPI stained nuclei and EGFP suggested that RbcS-TP-mVirD2-EGFP was localized mainly in the nucleus and slightly in chloroplasts in the *Arabidopsis* and tobacco leaves.

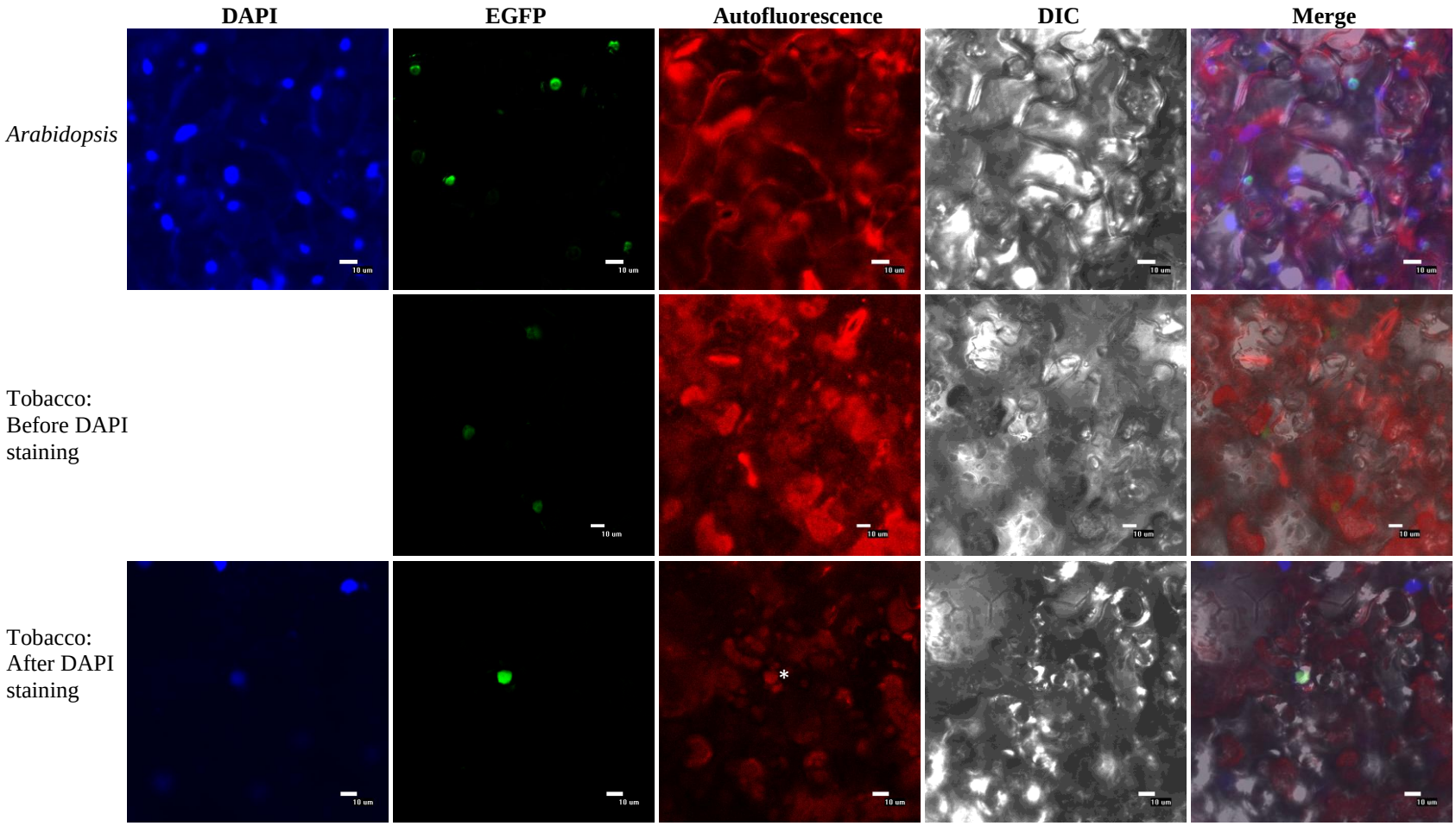


Figure 3.4.3.2a. DCL-TP-mVirD2-58-EGFP subcellular localization in transgenic *Arabidopsis* leaf (Top row) and in the tobacco leaves 48 h after agro-infiltration (Middle row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; * mark: transmission from EGFP. Scale bar = 10 µm.

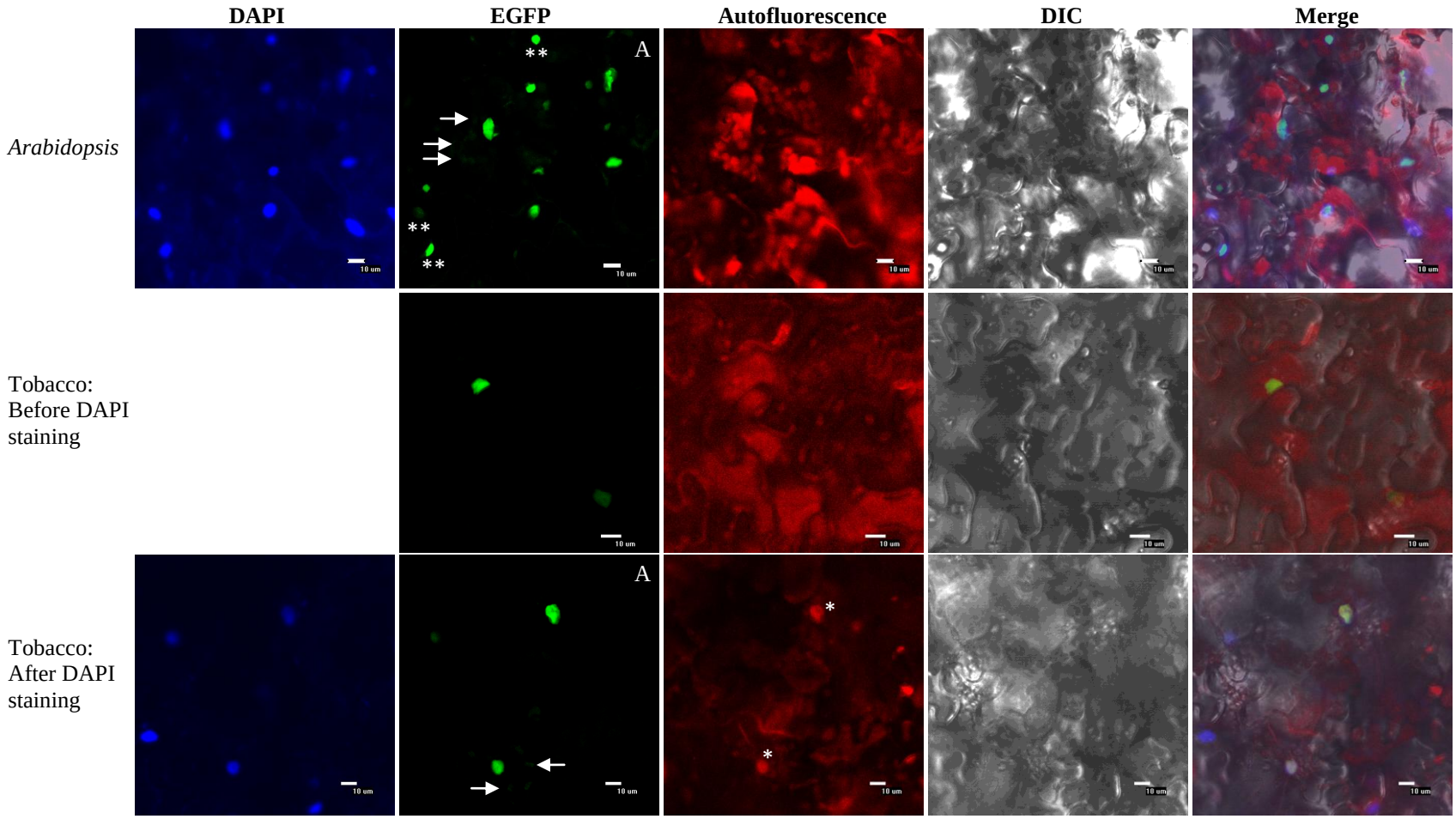


Figure 3.4.3.2b. RbcS-TP-mVirD2-58-EGFP subcellular localization in transgenic *Arabidopsis* leaf (Top row) and in the tobacco leaves 48 h after agro-infiltration (Middle row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI; EGFP, Autofluorescence and DIC, arrows: chloroplasts; * mark: transmission from EGFP; ** mark: no transmission to Autofluorescence; A: auto corrected. Scale bar = 10 µm.

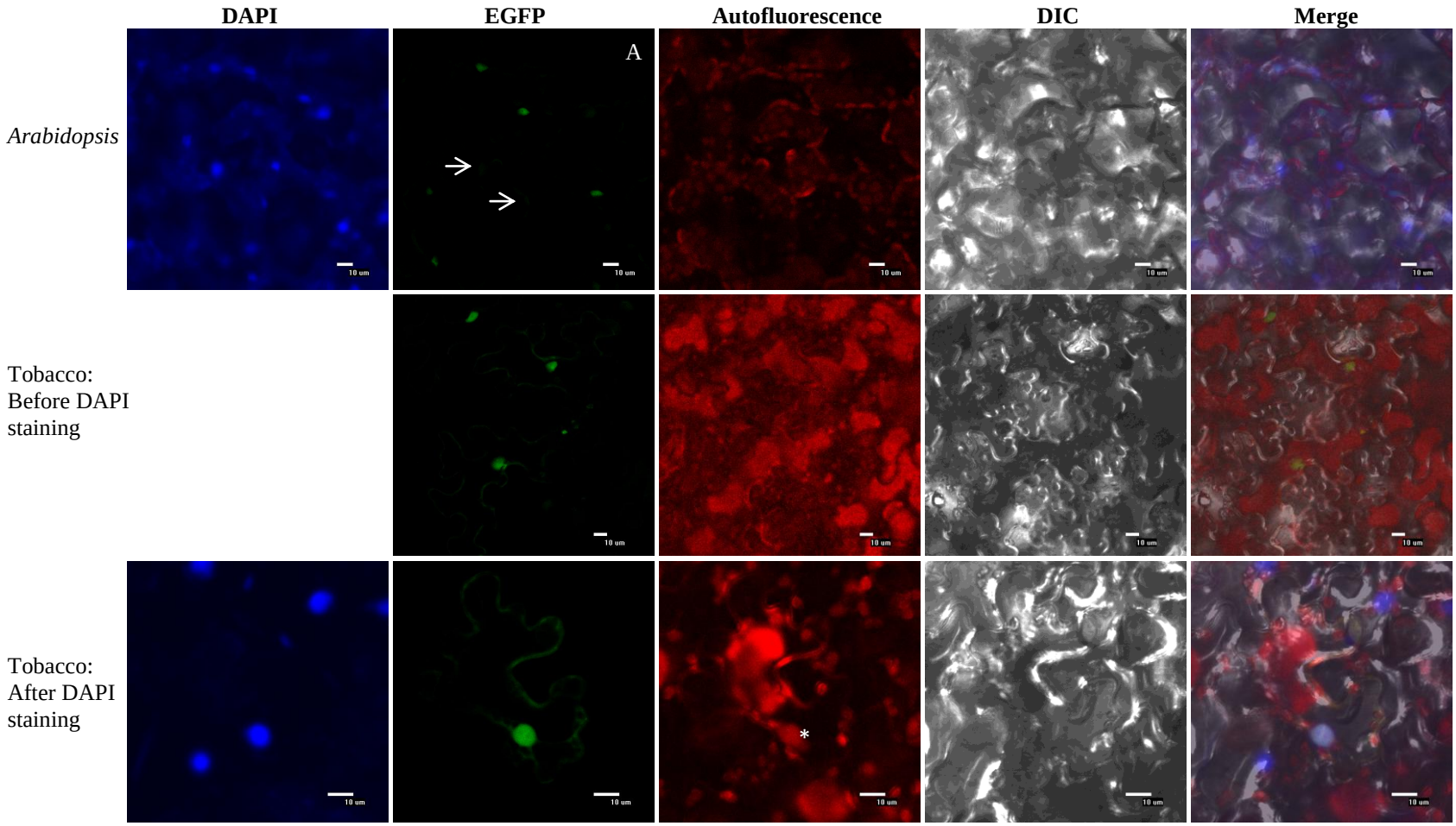


Figure 3.4.3.2c. DCL-TP-mVirD2-80-EGFP subcellular localization in transgenic *Arabidopsis* leaf (Top row) and in the tobacco leaves 48 h after agro-infiltration (Middle row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; open arrows: nuclei; * mark: transmission from EGFP; A: auto corrected. Scale bar = 10 µm.

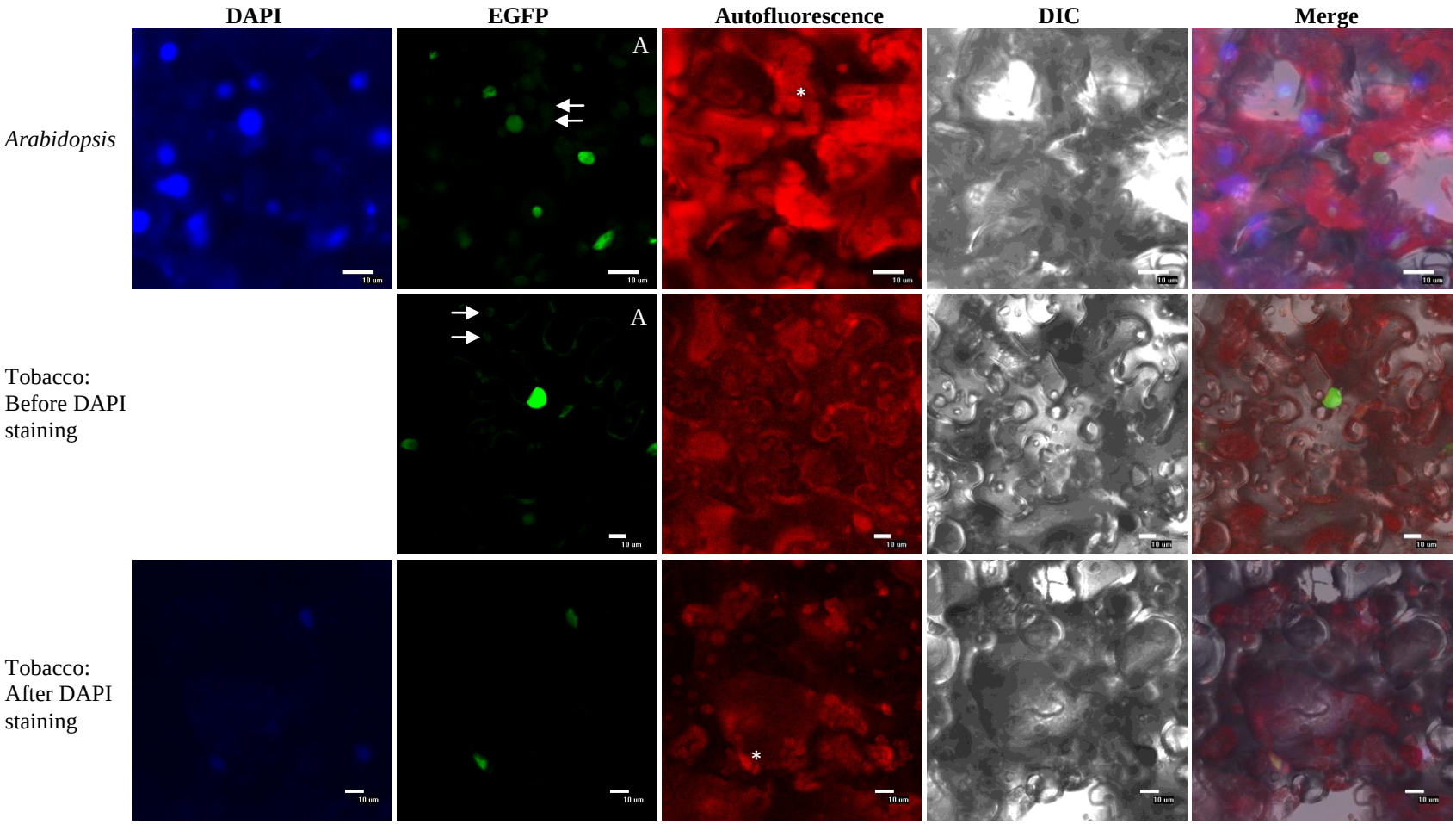


Figure 3.4.3.2d. RbcS-TP-mVirD2-80-EGFP subcellular localization in transgenic *Arabidopsis* leaf (Top row) and in the tobacco leaves 48 h after agro-infiltration (Middle row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; arrows in *Arabidopsis* EGFP: EGFP in chloroplasts; * mark: transmission from EGFP; A: auto corrected. Scale bar = 10 μm.

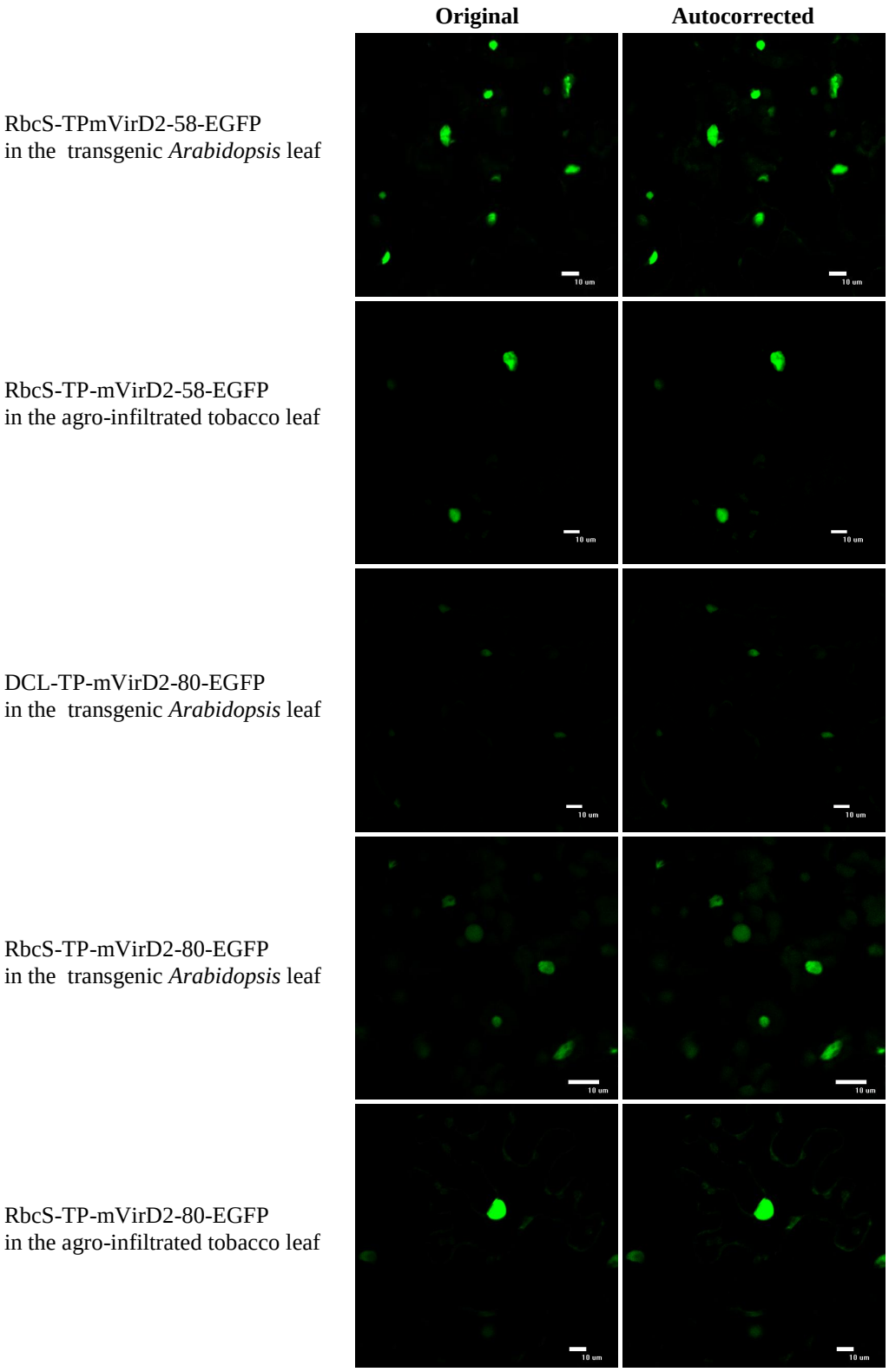


Figure 3.4.3.2e. Original and autocorrected pictures indicating subcellular localization of recombinant proteins.
Autocorrection was done to enhance only green colour to make the subcellular localization of the recombinant proteins more explicit.

3.4.3.3. Subcellular localization of VirD2-EGFP, mVirD2-58-EGFP and VirD2-80-EGFP in *Arabidopsis* and tobacco leaves

The subcellular localization of TP-mVirD2 suggested that DCL and RbcS-TP were not strong chloroplast targeting signals when they were fused to mVirD2. Thus, the subcellular localization of mVirD2-58-EGFP and mVirD2-80-EGFP was examined in leaves of stably transformed *Arabidopsis* and transiently transformed tobacco leaves to confirm the active nuclear targeting of these proteins.

The *mvirD2-58* and *mvirD2-80* genes in pVD58-AgeI and pVD80 were cloned into pEZT-NL resulting in pEZT-NL-mVirD2-58 and pEZT-NL-mVirD2-80, respectively (Table 3.4.1). Furthermore, the *virD2* gene was cloned into pEZT-NL giving rise to pEZT-NL-VirD2 (Table 3.4.1). VirD2-EGFP was planned to be used as a nuclear targeting control. However, neither generation of transgenic *Arabidopsis* transformants nor transient transformation of tobacco leaves was successful for VirD2-EGFP in any of three attempts.

mVirD2-58-EGFP and mVirD2-80-EGFP were localized exclusively in the nucleus of the *Arabidopsis* leaves, while in the tobacco leaf the recombinant proteins were localized mainly in the nucleus and also formed aggregates in the cytoplasm (circles in Figure 3.4.3.3a and b). There was no transmission between EGFP and autofluorescence channels in the *Arabidopsis* leaves. In the tobacco leaf, transmission between two channels was observed from some nuclei (* marks in Figure 3.4.3.3a and b), yet some other nuclei did not show transmission between the channels (** marks in Figure 3.4.3.3a and b). The overlay of the DAPI stained nuclei and EGFP suggested that mVirD2-58-EGFP and mVirD2-80-EGFP were localized in the nucleus of the tobacco leaf cell. These results demonstrated that recombinant proteins without known VirD2 NLSs could still be imported largely to the nucleus in *Arabidopsis* and tobacco leaves.

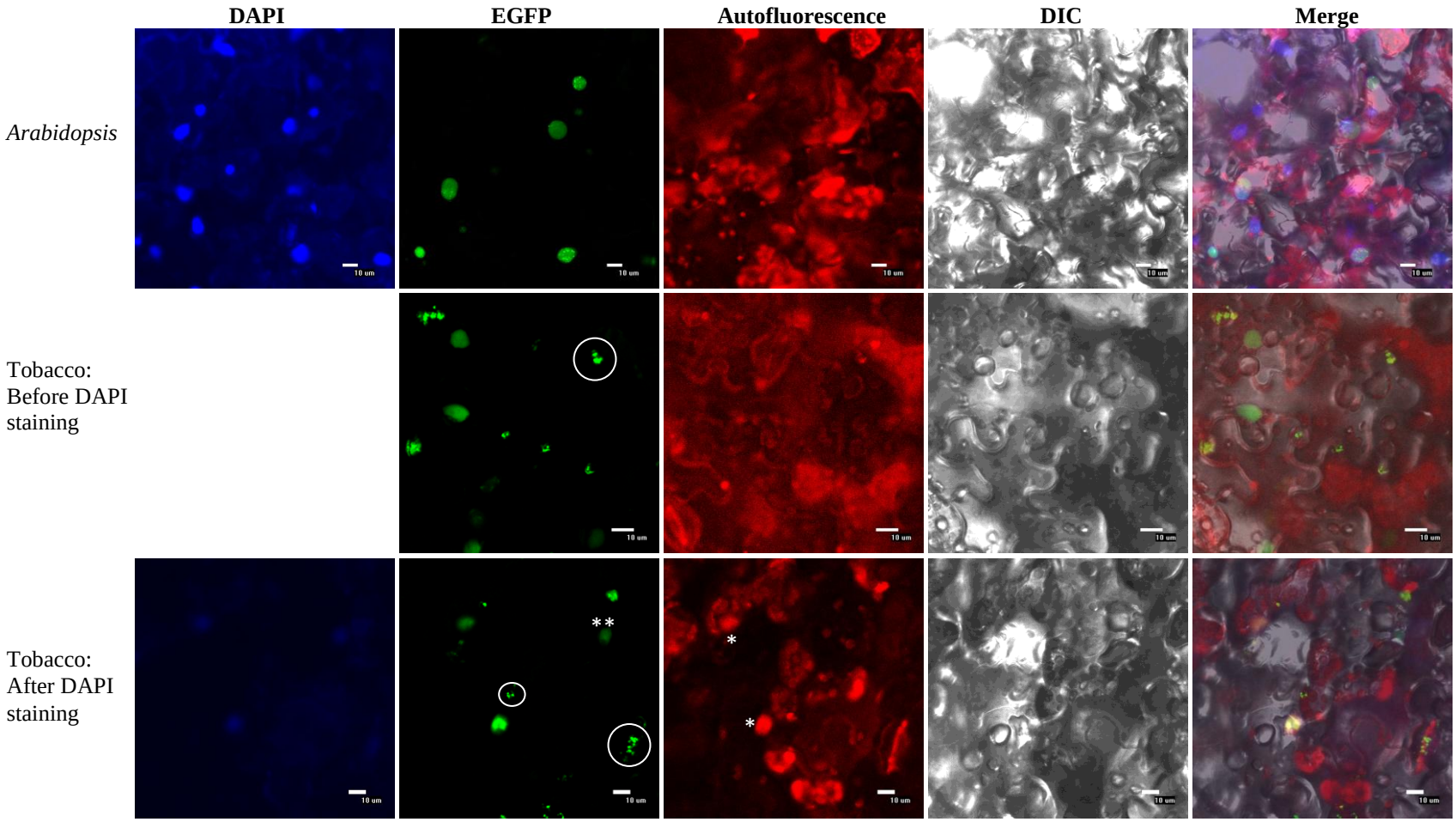


Figure 3.4.3.3a. mVirD2-58-EGFP subcellular localization in transgenic *Arabidopsis* leaf (Top row) and in the tobacco leaves 48 hours after agro-infiltration (Middle row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; circles: EGFP aggregates; * mark: transmission from EGFP; ** mark: no transmission to Autofluorescence. Scale bar = 10 µm.

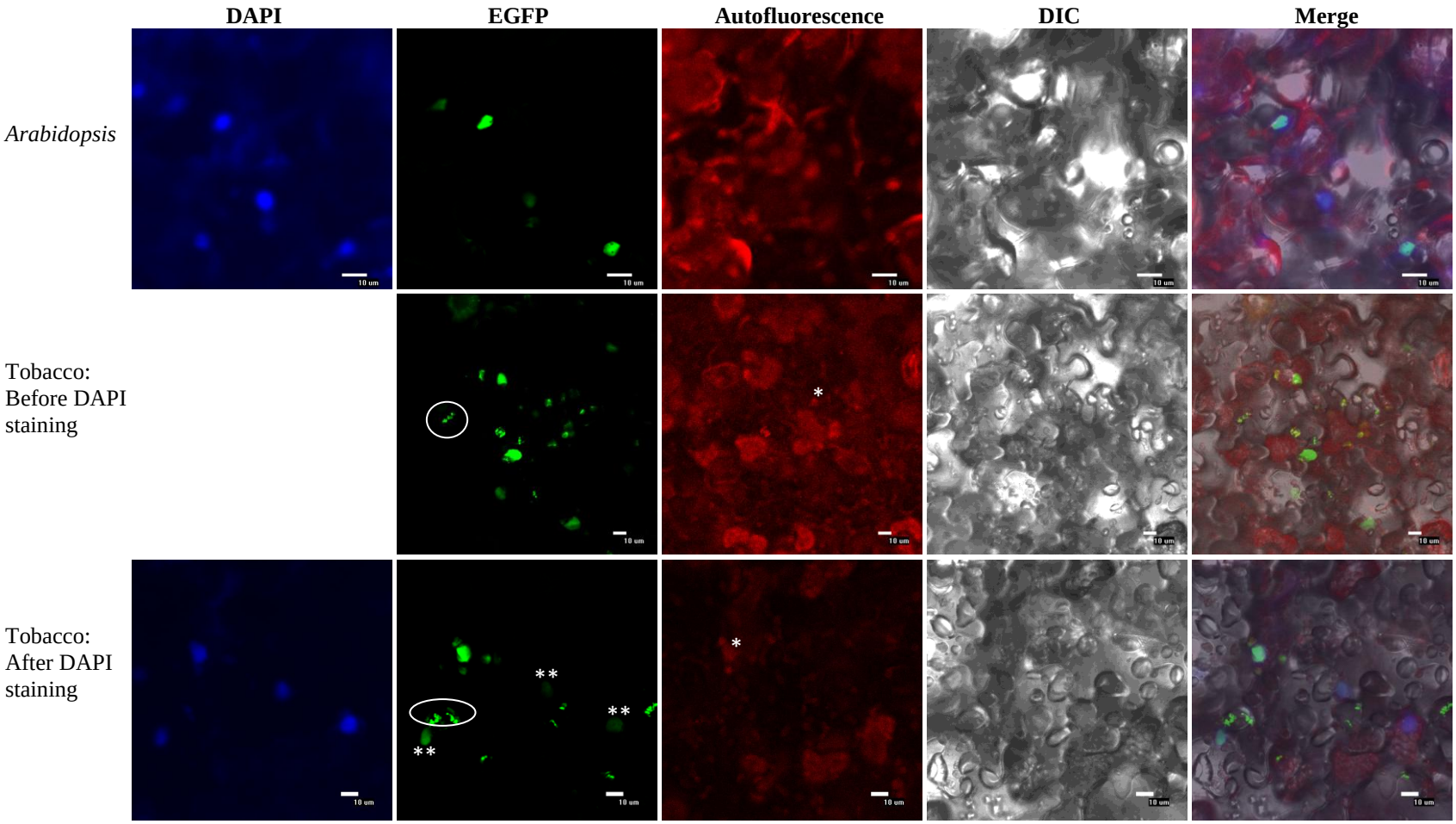


Figure 3.4.3.3b. mVirD2-80-EGFP subcellular localization in transgenic *Arabidopsis* leaf (Top row) and in the tobacco leaves 48 h after agro-infiltration (Middle row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; circles: EGFP aggregates; * mark: transmission from EGFP; ** mark: no transmission to Autofluorescence. Scale bar = 10 µm.

3.4.4. Subcellular localization of fusion proteins carrying the N-terminal parts of VirD2 and mVirD2in stably transformed *Arabidopsis* and transiently transformed tobacco leaves

3.4.4.1. Subcellular localization of VirD2-N1/2-EGFP, mVirD2-58/80-N1/2-EGFP, VirD2-N1/3-EGFP and mVirD2-58/80-N1/3-EGFP in *Arabidopsis* and tobacco leaves

To examine an effect of a change from arginine to threonine in the N-terminal NLSin mVirD2, the N-terminal half of mVirD2-58/80 were fused to EGFP. Additionally, the N-terminal half of VirD2 was fused to EGFP to serve as a nuclear targeting control. The sequences coding for the N-terminal half of VirD2 and mVirD2-58/80 were cloned into pEZT-NL resulting in pEZT-NL-VirD2-N1/2 and pEZT-NL-mVirD2-58/80-N1/2 (Table 3.4.1a). In addition, smaller fragments of the N-terminal one third of VirD2 and mVirD2-58/80 were fused to EGFP resulting in VirD2-N1/3-EGFP and mVirD2-58/80-N1/3-EGFP, to observe if there would be the different subcelluar localization from VirD2-N1/2-EGFP and mVirD2-58/80-N1/2-EGFP. The sequences coding for the N-terminal one third of VirD2 and mVirD2-58/80 were cloned into pEZT-NL resulting in pEZT-NL-VirD2-N1/3 and pEZT-NL-mVirD2-58/80-N1/3, respectively (Table 3.4.1a).

VirD2-N1/2-EGFP was found in the nucleus and little in the cytoplasm of *Arabidopsis* leaf cells (Figure 3.4.4.1a). However, this recombinant protein formed aggregates in the cells of the tobacco leaves, what most likely prevented its targeting to a specific subcellular location (Figure 3.4.4.1a). In the tobacco leaves, transmission from EGFP to autofluorescence channel was observed (* mark in Figure 3.4.4.1a).

Arabidopsis transformation was unsuccessful using *Agrobacterium* carrying pEZT-NL-mVirD2-58/80-N1/2-EGFP in three transformation attempts; however, mVirD2-58/80-N1/2-EGFP was successfully detected in the transiently transformed tobacco leaves. The subcellular localization of the recombinant protein formed aggregates in the cell (Figure 3.4.4.1b), similar to VirD2-N1/2-EGFP.

Arabidopsis transformation was unsuccessful in the case of VirD2-N1/3-EGFP and mVirD2-58/80-N1/3-EGFP in three transformation attempts. In the tobacco leaves, both proteins were detected in the cytoplasm (some in aggregates) and weakly in the nucleus (Figure 3.4.4.1c and d). This localization was most likely caused by diffusion of the proteins through the nuclear pore due to their moderate size (54kDa;

Table 3.4.1). Transmission of EGFP to autofluorescence channels was not observed in *Arabidopsis* or tobacco leaves.

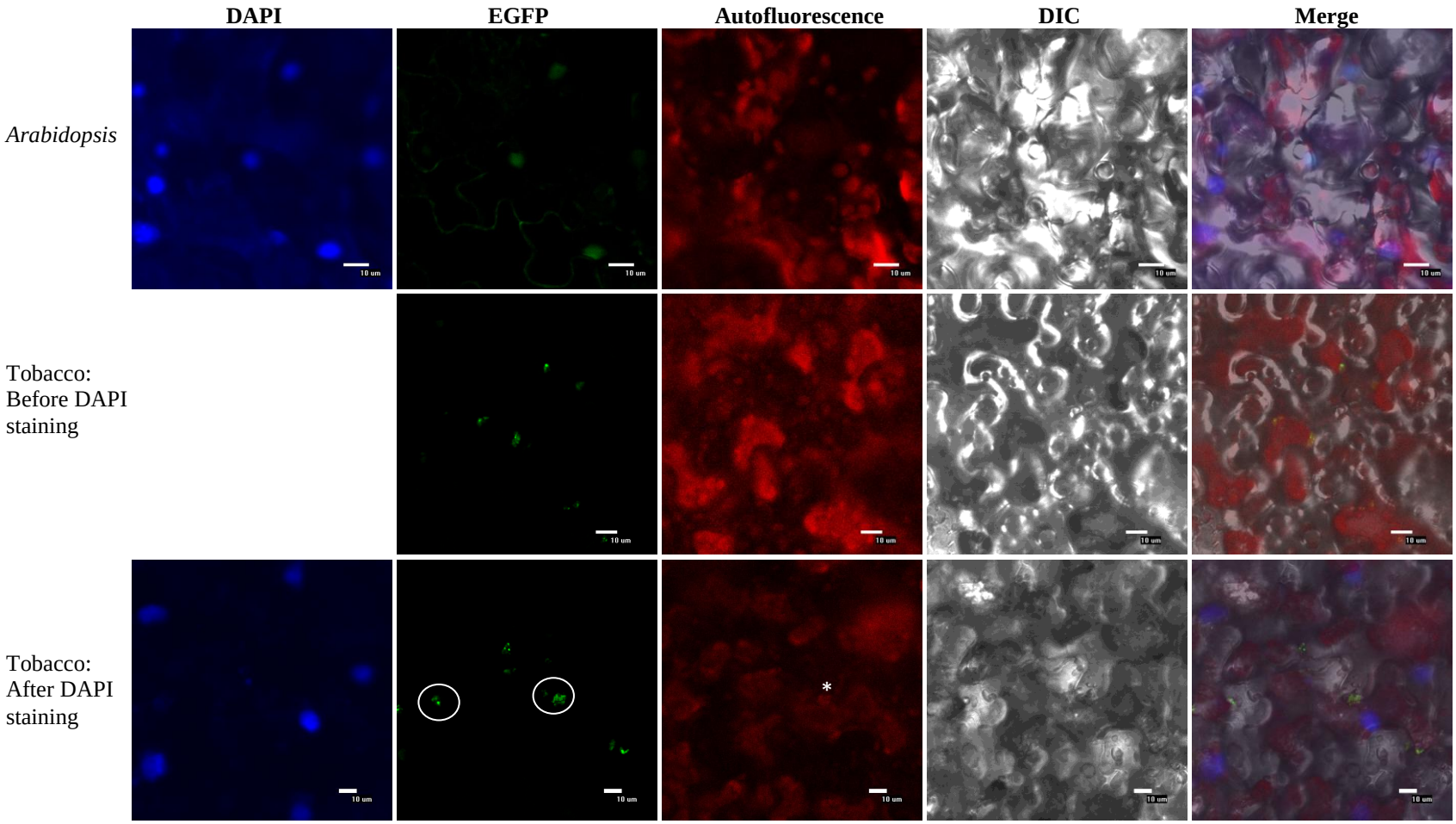


Figure 3.4.4.1a. VirD2-N1/2-EGFP subcellular localization in transgenic *Arabidopsis* leaf (Top row) and in the tobacco leaves 48 h after agro-infiltration (Middle row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei, EGFP: recombinant protein fused to EGFP, Autofluorescence: chloroplast autofluorescence, DIC: differential interference contrast, Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC, circles: EGFP aggregates, * mark: transmission from EGFP. Scale bar = 10 µm.

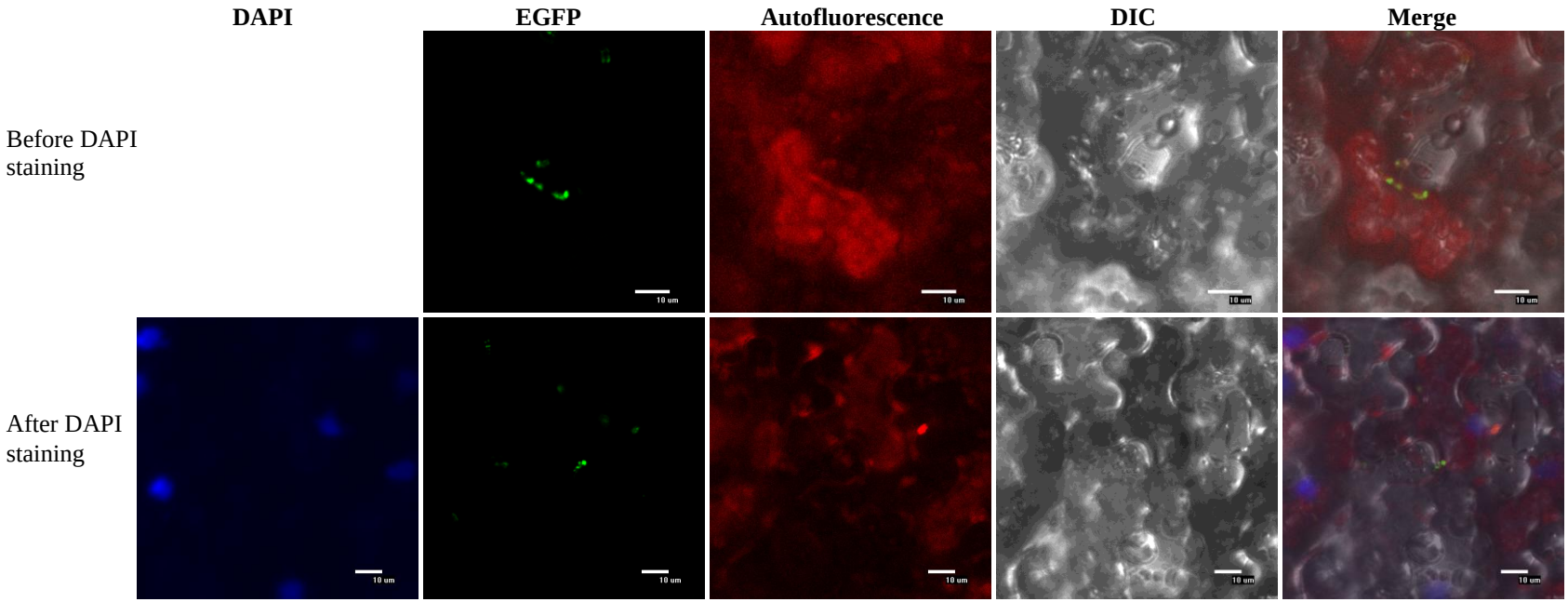


Figure 3.4.4.1b. mVirD2-58/80-N1/2-EGFP subcellular localization in the tobacco leaves 48 h after agro-infiltration (Top row: before DAP staining, Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei, EGFP: recombinant protein fused to EGFP, Autofluorescence: chloroplast autofluorescence, DIC: differential interference contrast, Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC. Scale bar = 10 µm.

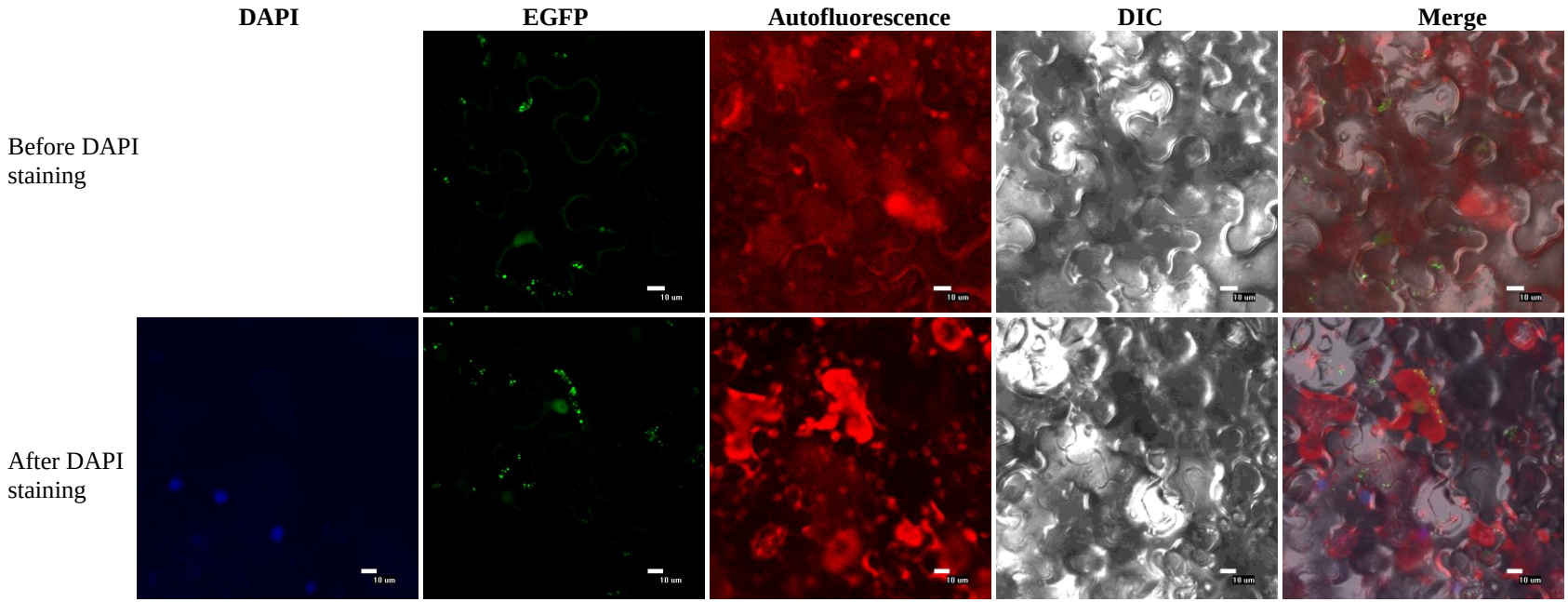


Figure 3.4.4.1c. Subcellular localization of VirD2-N1/3-EGFP in the tobacco leaves 48 h after agro-infiltration (Top row: before DAP staining, Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei, EGFP: recombinant protein fused to EGFP, Autofluorescence: chloroplast autofluorescence, DIC: differential interference contrast, Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC. Scale bar = 10 µm.

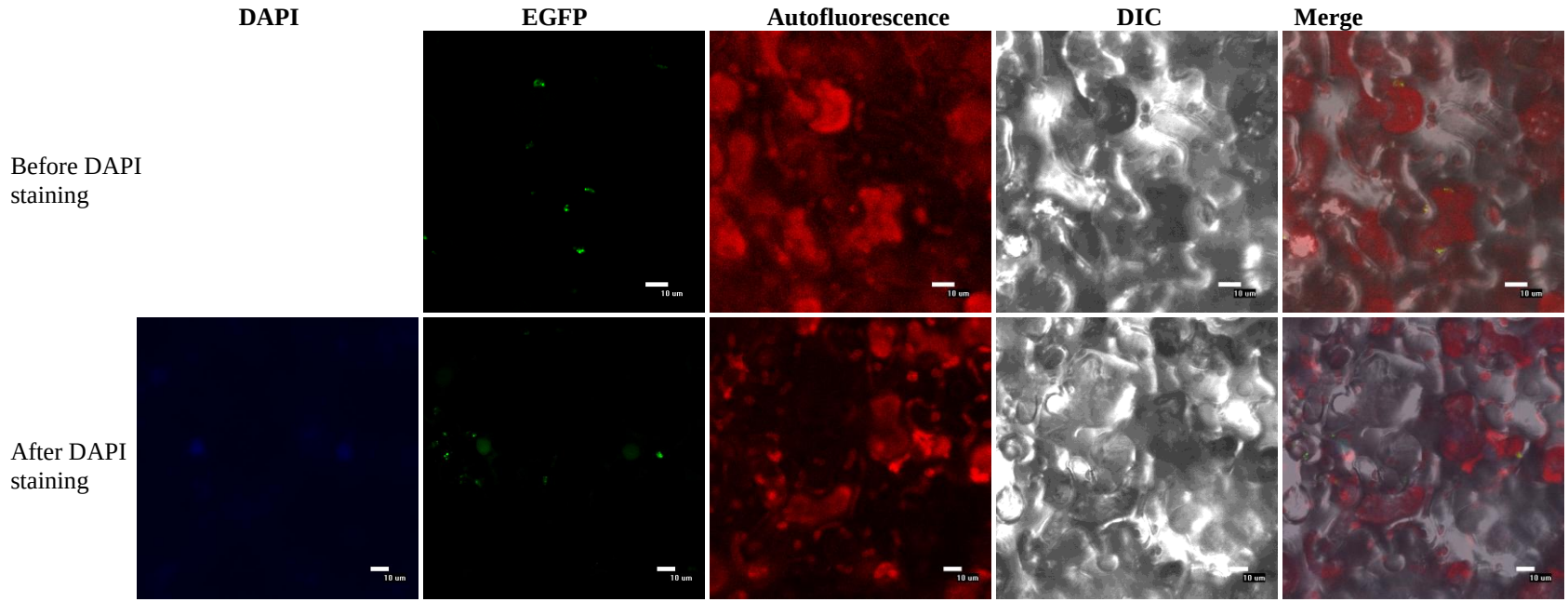


Figure 3.4.4.1d. Subcellular localization of mVirD2-58/80-N1/3-EGFP in the tobacco leaves 48 h after agro-infiltration (Top row: before DAP staining, Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei, EGFP: recombinant protein fused to EGFP, Autofluorescence: chloroplast autofluorescence, DIC: differential interference contrast, Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC. Scale bar = 10 µm.

3.4.5. Subcellular localization of the fusion proteins carrying the central region of VirD2 in stably transformed *Arabidopsis* and transiently transformed tobacco leaves

The sequence coding for the central region of VirD2 illustrated in Figure 3.3.1 was cloned from pVD43 containing the wild type *virD2* gene into pEZT-NL. Neither stable transformation of *Arabidopsis* nor transient expression in tobacco leaves was successful.

3.4.6. Subcellular localization of the C-terminal parts of VirD2 and mVirD2-EGFP in stably transformed *Arabidopsis* and/or transiently transformed tobacco leaves

3.4.6.1. Subcellular localization of VirD2-C1/2, mVirD2-58-C1/2, mVirD2-80-C1/2, VirD2-C1/3, mVirD2-58-C1/3, mVirD2-80-C1/3, and mVirD2-80-C1/3-EGFPx2 in *Arabidopsis* and tobacco leaves

mVirD2-58 and mVirD2-80 lack the C-terminal NLS by broad and precise deletions, respectively. Thus, when the C-terminal halves of mVirD2-58 and mVirD2-80 were fused to EGFP, it was predicted that the fusion proteins would not be localized in the nucleus. The sequences coding for the C-terminal halves and the C-terminal one third of VirD2, mVirD2-58 and mVirD2-80 were cloned into pEZT-NL resulting in pEZT-NL-VirD2-C1/2, pEZT-NL-mVirD2-58-C1/2, pEZT-NL-mVirD2-80-C1/2, pEZT-NL-VirD2-C1/3, pEZT-NL-mVirD2-58-C1/3 and pEZT-NL-mVirD2-80-C1/3 (Table 3.4.1).

Arabidopsis transformation was successful only for pEZT-NL-mVirD2-80-C1/2, pEZT-NL-mVirD2-58-C1/3 and pEZT-NL-mVirD2-80-C1/3, and unsuccessful when pEZT-NL-VirD2-C1/2, pEZT-NL-mVirD2-58-C1/2 and pEZT-NL-VirD2-C1/3 were used in three transformation attempts (Table 3.4.1). Transient expression in tobacco leaves was successful for all constructs. VirD2-C1/2-EGFP was localized exclusively in the nucleus of the tobacco leaf, as expected, due to its C-terminal NLS (Figure 3.4.6.1a). However, mVirD2-58-C1/2-EGFP and mVirD2-80-C1/2 lacking the C-terminal NLS were still localized exclusively in the nucleus of tobacco leaves (Figure 3.4.6.1b and c). mVirD2-80-C1/2-EGFP was detected in the nucleus and partially in the cytoplasm in the transgenic *Arabidopsis* leaf (Figure 3.4.6.1c). In tobacco leaves, transmission of EGFP to autofluorescence was observed for mVirD2-58-C1/2-EGFP and mVirD2-80-C1/2, but not for VirD2-C1/2-EGFP (* mark in Figure 3.4.6.1b and c). Overlay of the DAPI stained nuclei and EGFP indicated that the recombinant proteins were in the nuclei.

VirD2-C1/3-EGFP was localized exclusively in the nucleus of the tobacco leaf (Figure 3.4.6.1d), whereas mVirD2-58-C1/3-EGFP and mVirD2-80-C1/3-EGFP were found in the cytoplasm and strongly in the nuclei of *Arabidopsis* and tobacco leaves (Figure 3.4.6.1e and f). The observed localization of mVirD2-58-C1/3-EGFP and mVirD2-80-C1/3-EGFP might be caused either by their diffusion through the nuclear pore due to their small size (41 and 44 kDa; Table 3.4.1) or by the presence of a cryptic NLS, or both.

In order to exclude the possibility of diffusion of the recombinant proteins through the nuclear pore, an extra alanine linker and EGFP was fused to mVirD2-80-C1/3-EGFP giving rise to mVirD2-80-C1/3-EGFPx2. Only transient expression in tobacco leaves was detected. mVirD2-80-C1/3-EGFPx2 was localized exclusively in the nucleus (Figure 3.4.5.1g). An increase in size from 44 kDa (mVirD2-80-C1/3-EGFP) to 71 kDa (mVirD2-80-C1/3-EGFPx2) restricted the subcellular localization of the recombinant protein. These results suggested that mVirD2 contains a cryptic NLS.

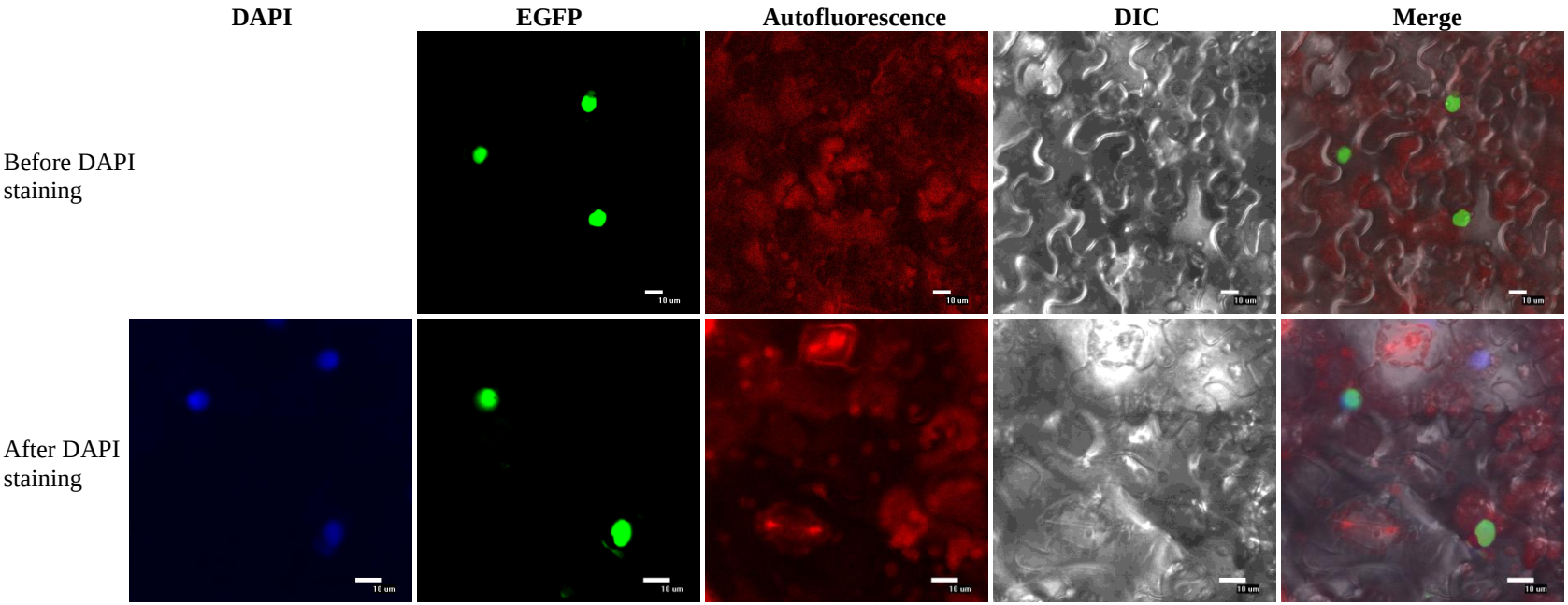


Figure 3.4.6.1a. Subcellular localization of VirD2-C1/2-EGFP in the tobacco leaf 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC. Scale bar = 10 µm.

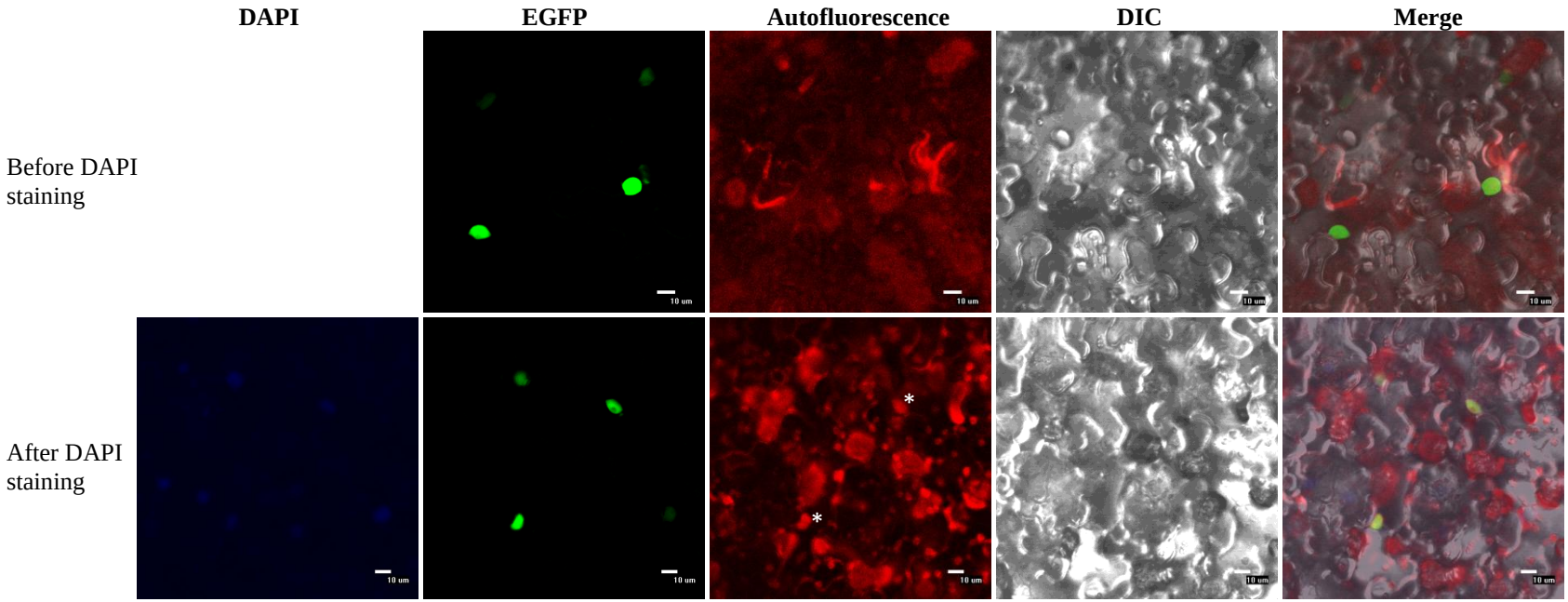


Figure 3.4.6.1b. Subcellular localization of mVirD2-58-C1/2-EGFP in the tobacco leaf 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; * mark: transmission from EGFP. Scale bar = 10 μm.

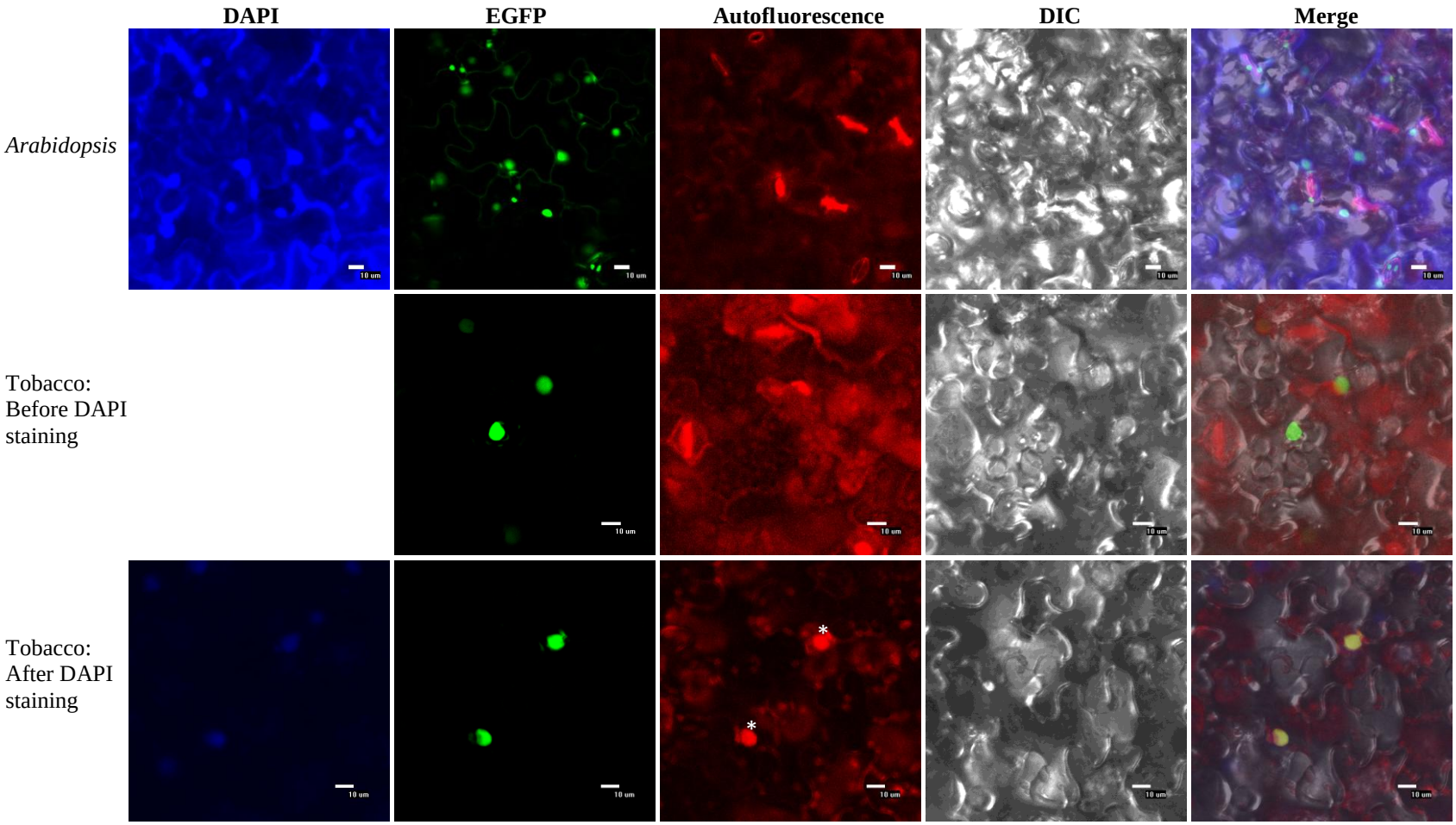


Figure 3.4.6.1c. mVirD2-80-C1/2-EGFP subcellular localization in transgenic *Arabidopsis* leaf (Top row) and in the tobacco leaf 48 h after agro-infiltration (Middle row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; * mark: transmission from EGFP. Scale bar = 10 µm.

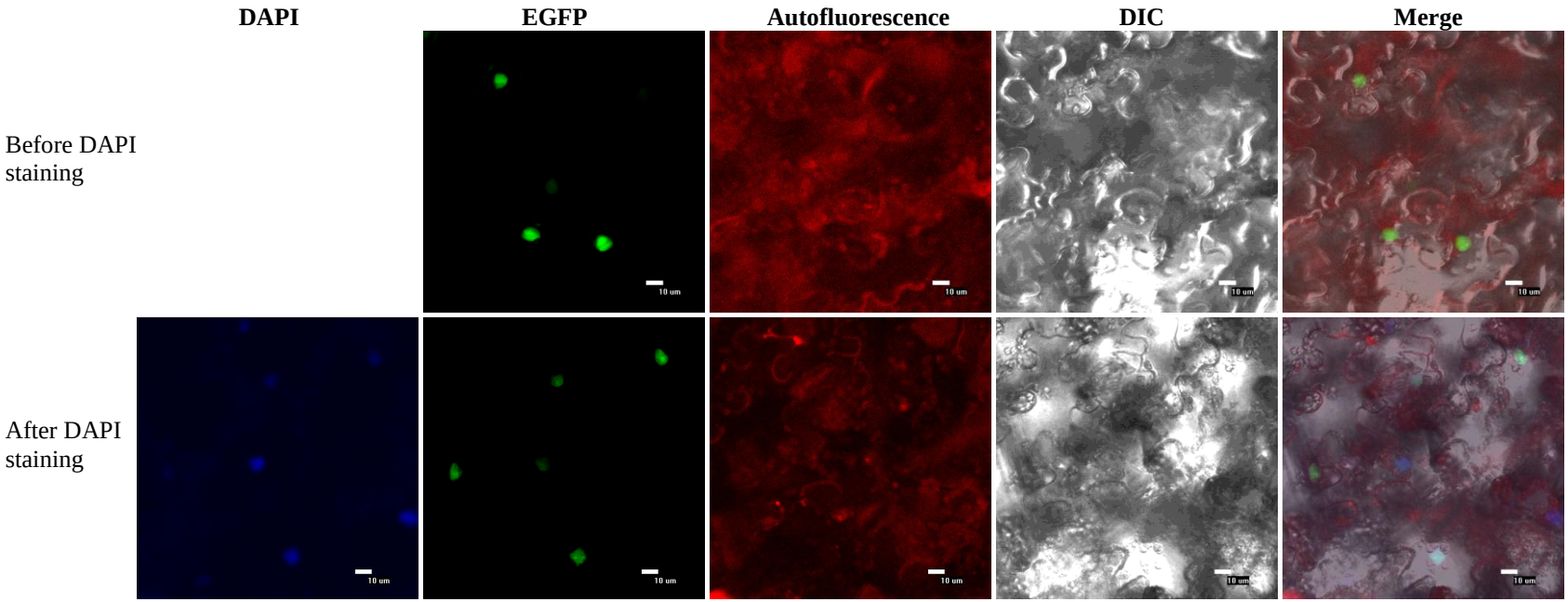


Figure 3.4.6.1d. VirD2-C1/3-EGFP subcellular localization in the tobacco leaf 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC. Scale bar = 10 µm.

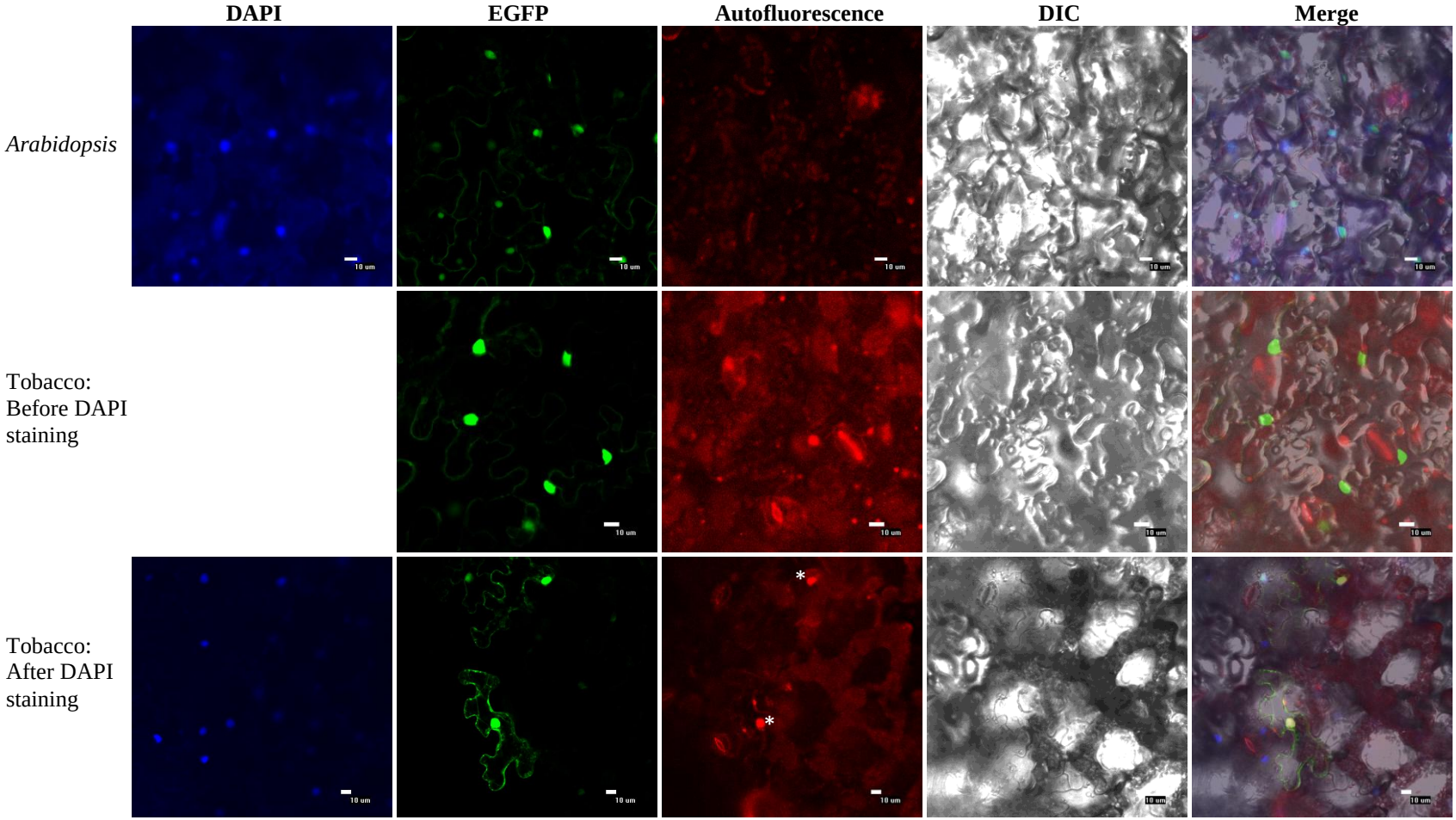


Figure 3.4.6.1e. mVirD2-58-C1/3-EGFP subcellular localization in transgenic *Arabidopsis* leaf (Top row) and in the tobacco leaf 48 h after agro-infiltration (Middle row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; * mark: transmission from EGFP. Scale bar = 10 µm.

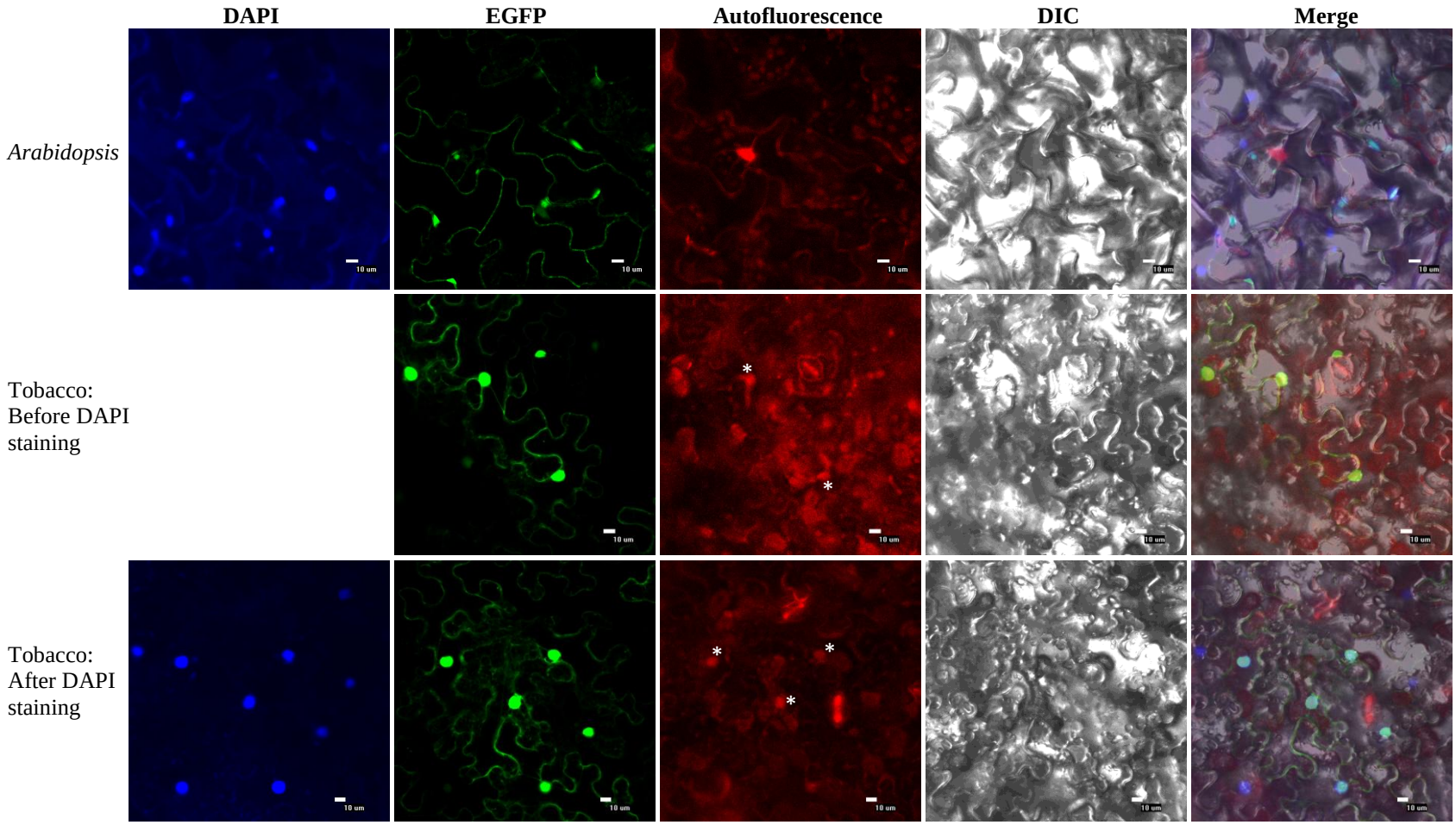


Figure 3.4.6.1f. mVirD2-80-C1/3-EGFP subcellular localization in transgenic *Arabidopsis* leaf (Top row) and in the tobacco leaf 48 h after agro-infiltration (Middle row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; * mark: transmission from EGFP. Scale bar = 10 µm.

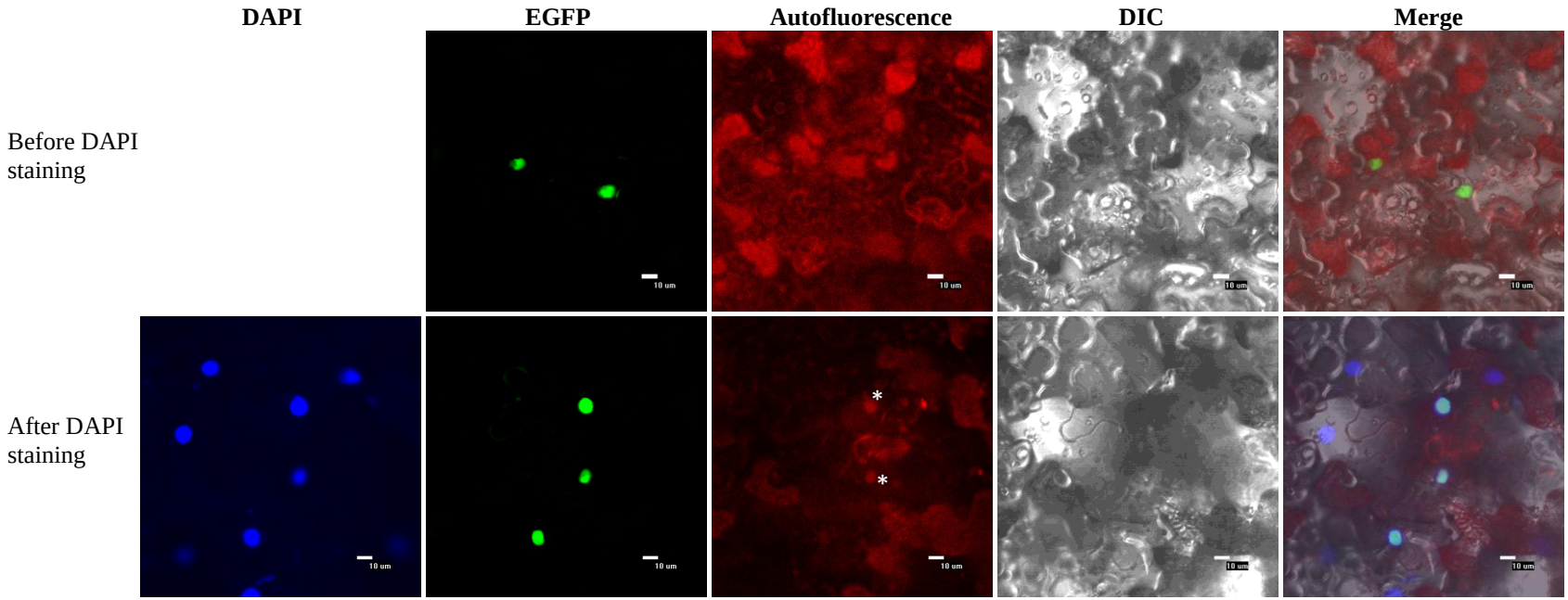


Figure 3.4.6.1g. mVirD2-80-C1/3-EGFPx2 subcellular localization in the tobacco leaf 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; * mark: transmission from EGFP. Scale bar = 10 µm**.**

3.4.6.2. Subcellular localization of VirD2-C1/3ΔNLS3-EGFP, mVirD2-58-C1/3ΔNLS3-EGFP and mVirD2-80-C1/3ΔNLS3-EGFP in tobacco leaves

VirD2 and mVirD2 amino acid sequences were analyzed using cNLS Mapper, and the predicted relative levels of NLS activity (score) were recorded (Figure 3.4.1a-c). Analysis showed a putative NLS (**QRSKRRNDEE**) named NLS3 (score 7: mainly in the nucleus, partially in the cytoplasm) which was about a half of the strength of the predicted bipartite of the C-terminal NLS (score 13.6: exclusively in the nucleus) (Figure 3.4.1a-c). To examine the nuclear targeting function of NLS3, the region containing sequence coding for NLS3 and the flaking amino acids (KRQ**QRSKRRNDEE**AGPSGANRKGLK) was deleted, indicated as ΔNLS3, from VirD2-C1/3-EGFP, mVirD2-58-C1/3-EGFP and mVirD2-80-C1/3-EGFP resulting in recombinant proteins VirD2-C1/3ΔNLS3-EGFP, mVirD2-58-C1/3ΔNLS3-EGFP and mVirD2-80-C1/3ΔNLS3-EGFP. The extra NLS3 flanking sequence was included because cloning of the shorter sequence cloned into pEZT-NL was not successful and/or transient transformation was not successful in the tobacco leaf as explained in Section 3.5.6 (data not shown). The sequence coding for the aforementioned proteins with ΔNLS3 in pVD43, pVD58-AgeI and pVD80 were cloned into pEZT-NL resulting in pEZT-NL-VirD2-C1/3ΔNLS3-EGFP, pEZT-NL-mVirD2-58-C1/3ΔNLS3-EGFP and pEZT-NL-mVirD2-80-C1/3ΔNLS3-EGFP (Table 3.4.1).

VirD2-C1/3ΔNLS3-EGFP was detected exclusively in the nucleus as expected due to the presence of the C-terminal NLS (Figure 3.4.6.2a), whereas mVirD2-58-C1/3ΔNLS3-EGFP and mVirD2-80-C1/3ΔNLS3-EGFP were found in the nucleus and in the cytoplasm. The cytoplasmic localization was slightly stronger than in the case of mVirD2-58-C1/3 and mVirD2-80-C1/3 (Figure 3.4.6.2b-c). EGFP transmission to autofluorescence was observed only in mVirD2-80-C1/3ΔNLS3-EGFP (Figure 3.4.6.2c). The EGFP distribution pattern was not largely different between mVirD2-58-C1/3-EGFP and mVirD2-58-C1/3ΔNLS3-EGFP as well as between mVirD2-80-C1/3-EGFP and mVirD2-80-C1/3ΔNLS3-EGFP (Figure 3.4.6.1e-f and Figure 3.4.6.2b-c). This observation might be due to the small protein size: mVirD2-58-C1/3-EGFP (41 kDa), mVirD2-58-C1/3ΔNLS3-EGFP (34 kDa), mVirD2-80-C1/3-EGFP (44 kDa), and mVirD2-80-C1/3ΔNLS3-EGFP (36 kDa) (Table 3.4.1a). The absence of NLS3 in mVirD2-58-C1/3ΔNLS3-EGFP and mVirD2-80-C1/3ΔNLS3-EGFP likely caused diffusion of the recombinant proteins through the nuclear pore. An extra alanine linker and EGFP were added to mVirD2-80-C1/3ΔNLS3 resulting in

mVirD2-80-C1/3ΔNLS3-EGFPx2 (63 kDa) to prevent passive translocation of the proteins through the nuclear pore. The sequence coding for the alanine linker and EGFP was cloned into pEZT-NL-mVirD2-80-C1/3ΔNLS3 resulting in pEZT-NL-mVirD2-80-C1/3ΔNLS3-EGFPx2.

Despite a larger protein size, mVirD2-80-C1/3ΔNLS3-EGFPx2 (63 kDa) localization showed a similar pattern to mVirD2-80-C1/3ΔNLS3-EGFP (36 kDa) in the tobacco leaf (Figure 3.4.6.2b and d). Nevertheless, the results of localization of mVirD2-80-C1/3-EGFPx2 in the nucleus (Section 3.4.6.1) and localization of mVirD2-80-C1/3ΔNLS3-EGFPx2 showed that the absence of NLS3 altered the localization of the recombinant protein.

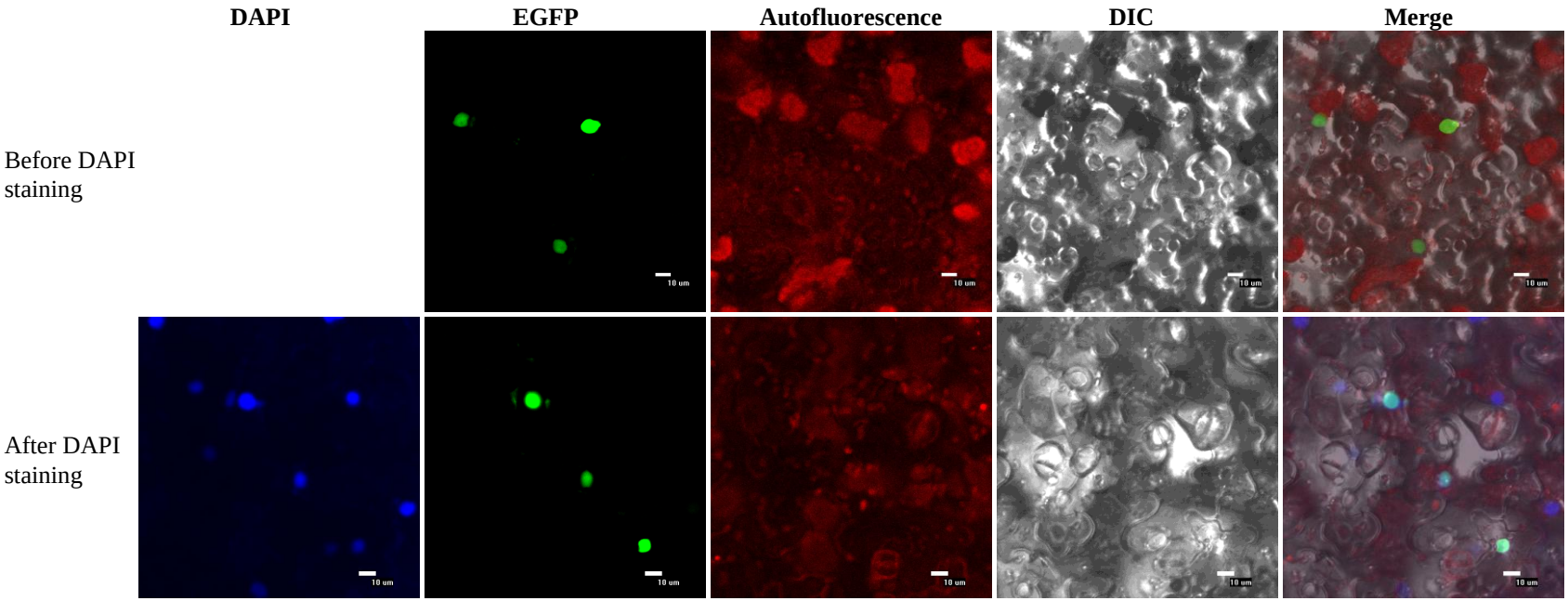


Figure 3.4.6.2a. VirD2-C1/3ΔNLS3-EGFP subcellular localization of in the tobacco leaf 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC. Scale bar = 10 µm.

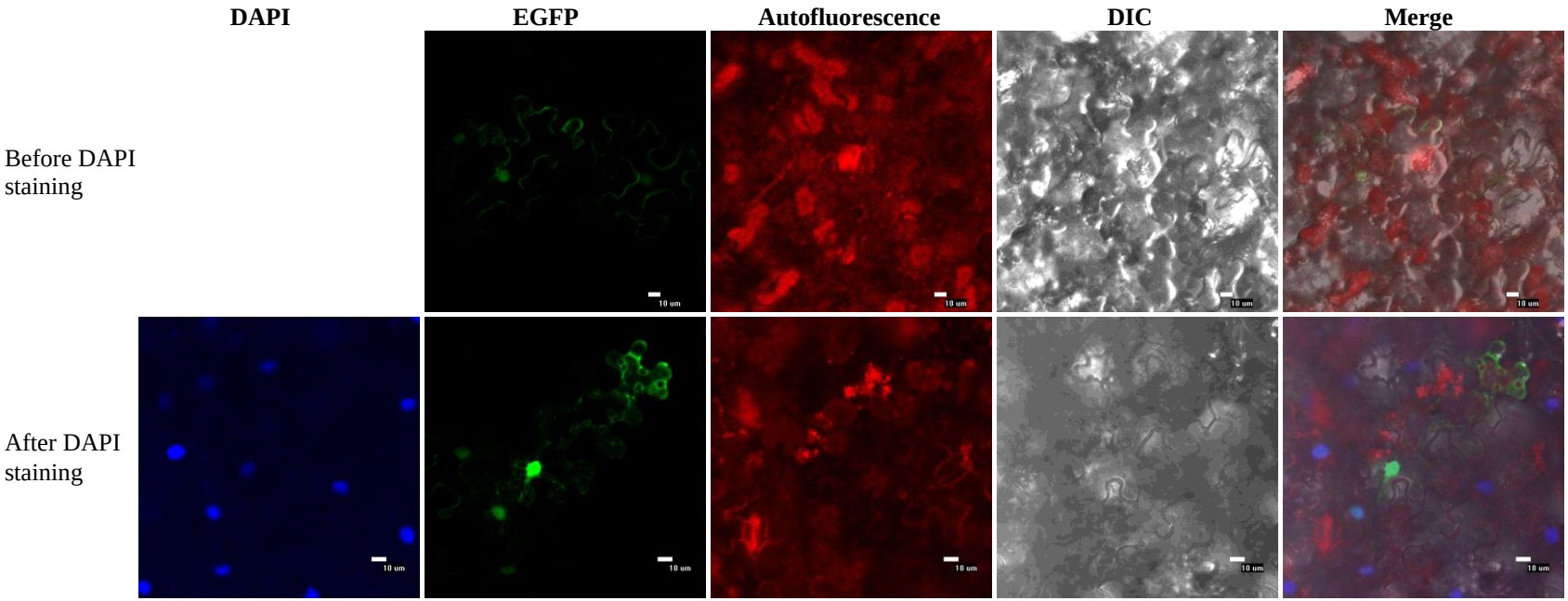


Figure 3.4.6.2b. mVirD2-58-C1/3ΔNLS3-EGFP subcellular localization of in the tobacco leaf 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC. Scale bar = 10 µm.

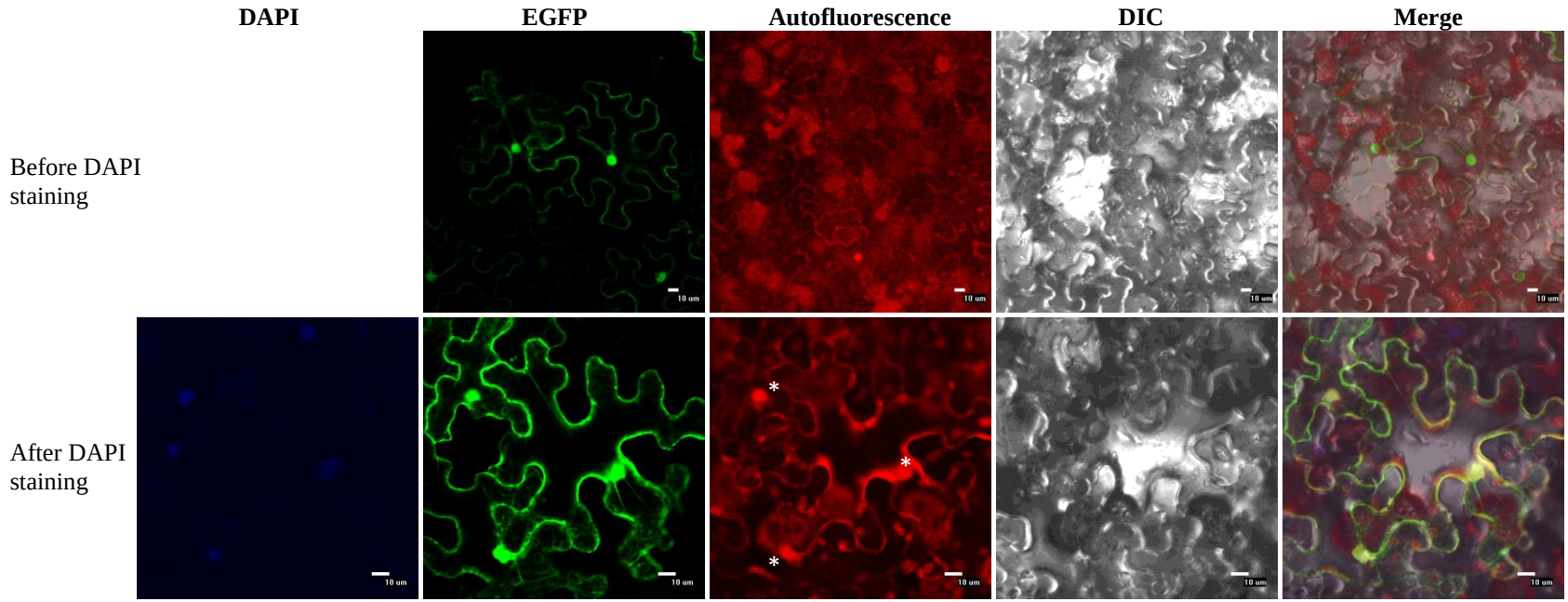


Figure 3.4.6.2c. mVirD2-80-C1/3ΔNLS3-EGFP subcellular localization of in the tobacco leaf 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; * mark: transmission from EGFP. Scale bar = 10 μm.

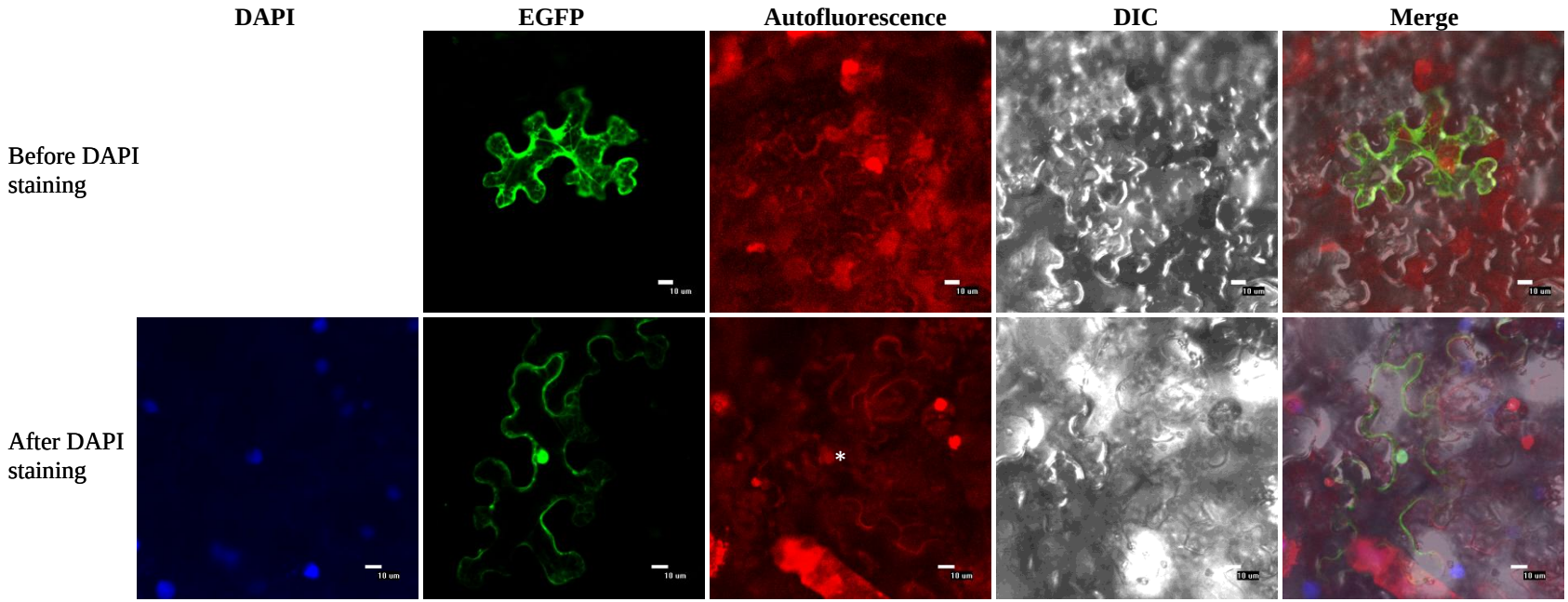


Figure 3.4.6.2d. mVirD2-80-C1/3ΔNLS3-EGFPx2 subcellular localization of in the tobacco leaf 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; * mark: transmission from EGFP. Scale bar = 10 µm.

3.4.7. Subcellular localization of NLS3-EGFP and 26aa-EGFP in tobacco leaves

Since the deletion of the NLS3 with the extra sequence had a negative impact on nuclear targeting of mVirD2-80-C1/3ΔNLS3-EGFPx2, we examined the nuclear targeting function of NLS3. The sequence coding for NLS3 and its flanking amino acids (KRQ**QRSKRRNDEE**AGPSGANRKGLK) was amplified using pVD43 carrying the wild type *virD2* and then cloned into pEZT-NL, resulting in pEZT-NL-NLS3 (Table 3.4.1). A nuclear targeting negative control was also constructed to ensure that addition of the same number of amino acids would not cause nuclear localization of the recombinant protein. For this purpose, the 78 bp long-sequence coding for first 26 amino acids of GFP in pCAMBIA1302 was cloned into pEZT-NL resulting in pEZT-NL-78bp. The recombinant protein was named as 26aa-EGFP (Table 3.4.1). One or two alanine and EGFP coding sequences were added into pEZT-NL-NLS3 and pEZT-NL-78bp resulting in pEZT-NL-NLS3-EGFPx2, pEZT-NL-NLS3-EGFPx3, pEZT-NL-78bp-EGFPx2 and pEZT-NL-78bp-EGFPx3. These additions were made to increase the size of recombinant proteins (Table 3.4.1).

Both NLS3-EGFP and 26aa-EGFP were detected in the nucleus and in the cytoplasm of tobacco leaf cells (Figure 3.4.7a and b). Transmission of EGFP to the autofluorescence channel was observed in NLS3-EGFP (* mark in Figure 3.4.7a). NLS3-EGFPx2 was found mainly in the nucleus and slightly in the cytoplasm where it formed some aggregates (circle in Figure 3.4.7c). Transmission was observed from EGFP in the one nucleus to the autofluorescence channel, whereas EGFP from other nuclei was not transmitted to the autofluorescence channel (* and ** marks in Figure 3.4.7c). Therefore, it was clear that NLS3-EGFPx2 was localized mainly in the nucleus. A high expression of 26aa-EGFPx2 was observed, and it was detected in the nucleus and the cytoplasm (Figure 3.4.7d). Finally, NLS3-EGFPx3 was localized exclusively in the nucleus, whereas 26aa-EGFPx3 was still found in the nucleus and the cytoplasm (Figure 3.4.7e and f), likely due to diffusion through the nuclear pore despite its size (88 kDa; Table 3.4.1). Transmission between the channels was not observed in either 26aa-EGFPx3 or NLS3-EGFPx3. It was concluded that NLS3 (KRQ**QRSKRRNDEE**AGPSGANRKGLK) was able to direct the recombinant protein into the nucleus. This confirmed the presence of a new NLS in VirD2.

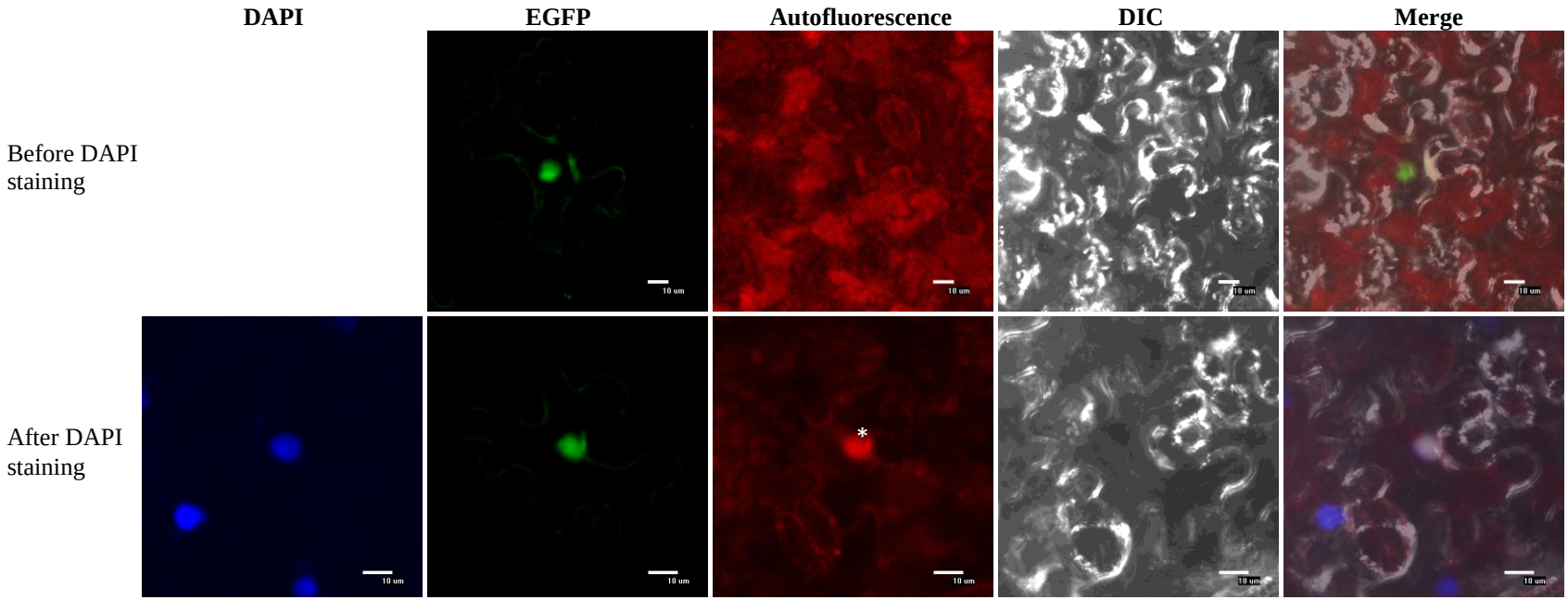


Figure 3.4.7a. NLS3-EGFP subcellular localization in the tobacco leaves 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; * mark: transmission from EGFP. Scale bar = 10 μm.

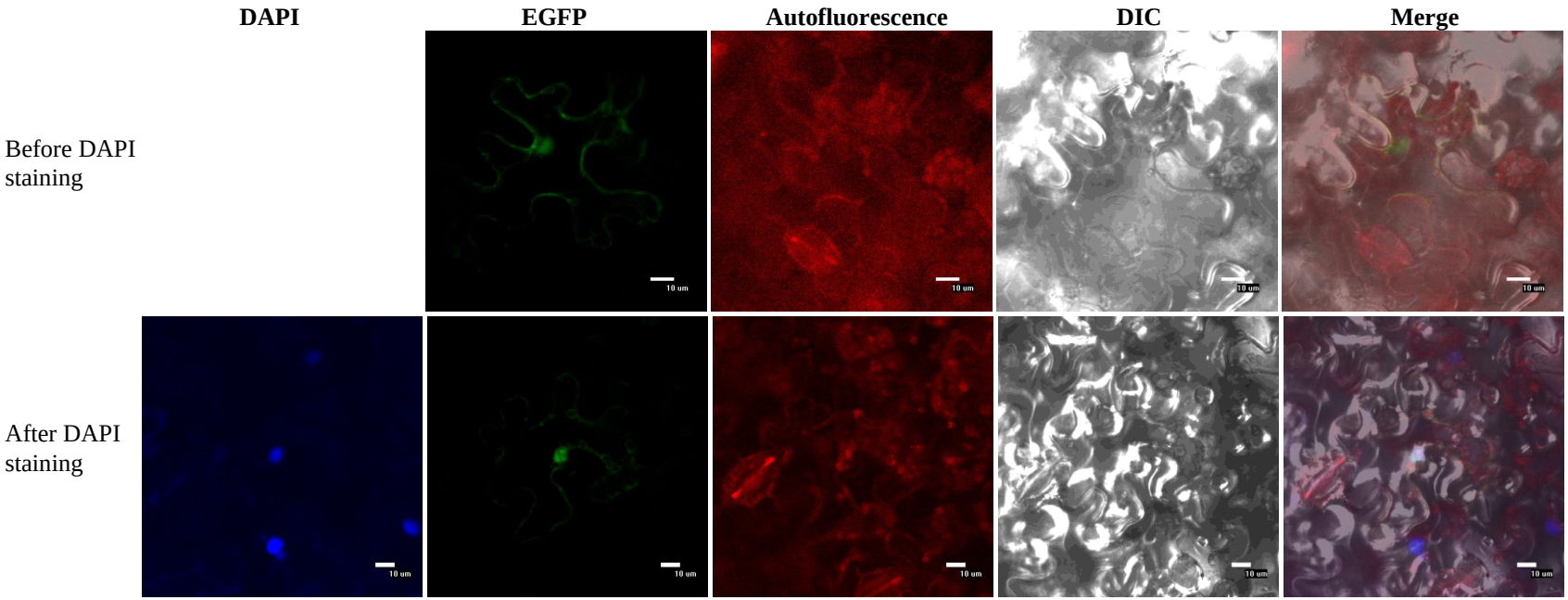


Figure 3.4.7b. 26aa-EGFP subcellular localization of in the tobacco leaf 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; * mark: transmission from EGFP. Scale bar = 10 μm.

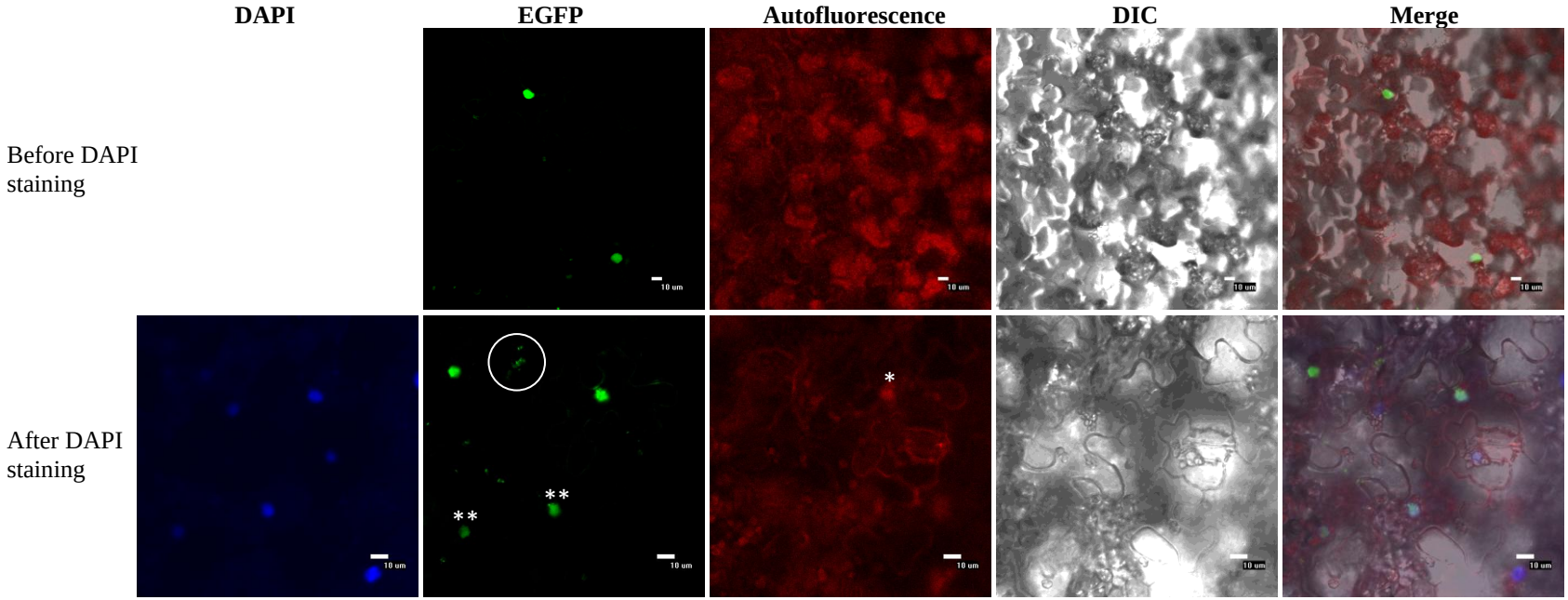


Figure 3.4.7c. NLS3-EGFPx2 subcellular localization in the tobacco leaves 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; * mark: transmission from EGFP; ** mark: no transmission to Autofluorescence; circle: aggregates. Scale bar = 10 μm.

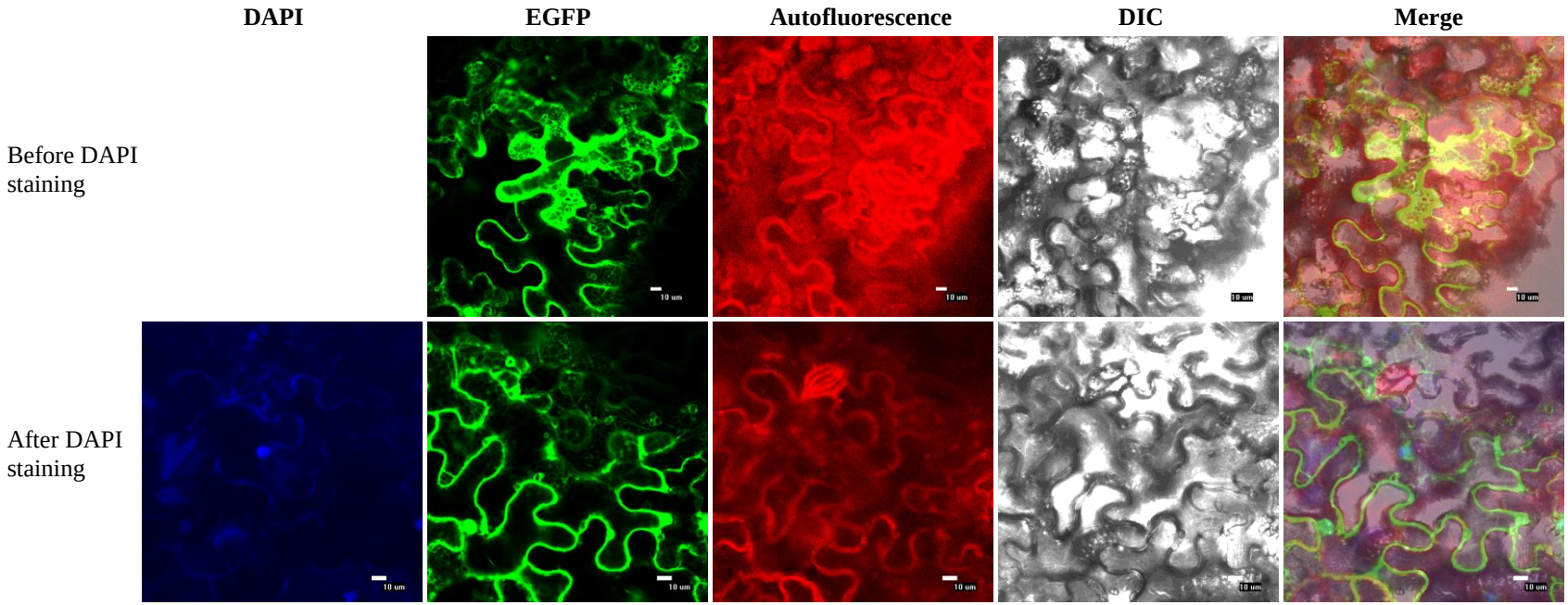


Figure 3.4.7d. 26aa-EGFPx2 subcellular localization in the tobacco leaves 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC. Scale bar = 10 µm.

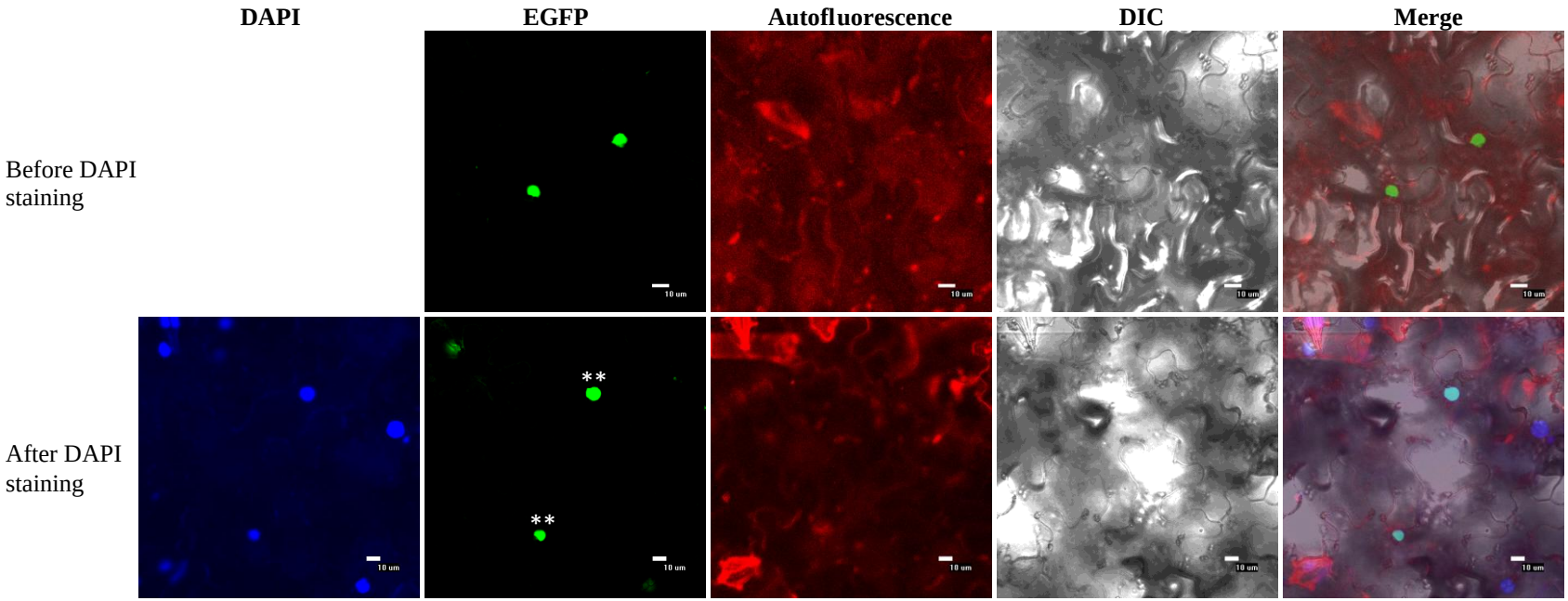


Figure 3.4.7e. NLS3-EGFPx3 subcellular localization in the tobacco leaves 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC; ** mark: no transmission to Autofluorescence. Scale bar = 10 µm.

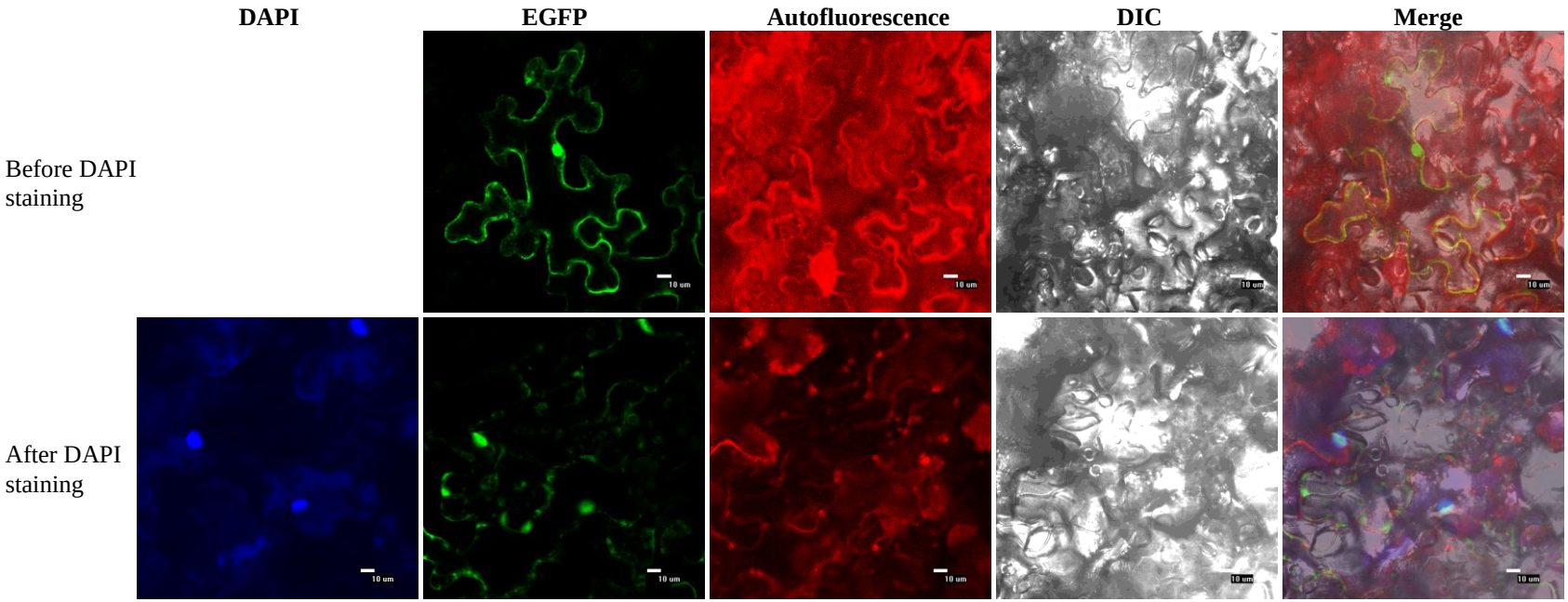


Figure 3.4.7f. 26aa-EGFPx3 subcellular localization in the tobacco leaves 48 h after agro-infiltration (Top row: before DAP staining; Bottom row: after DAPI staining) examined under the confocal microscope.

DAPI: DAPI stained nuclei; EGFP: recombinant protein fused to EGFP; Autofluorescence: chloroplast autofluorescence; DIC: differential interference contrast; Merge: merged photo of DAPI, EGFP, Autofluorescence and DIC. Scale bar = 10 µm.

3.5. Discussion

In our previous study on *Agrobacterium*-mediated chloroplast transformation, we introduced the modifications to VirD2 NLSs, a point mutation (from arginine to threonine) to the N-terminal NLS and a broad or a precise deletion of the C-terminal NLS, to inactivate its nuclear targeting according to the previous study (Rossi et al., 1993); additionally, DCL or RbcS chloroplast transit peptide was added at the N-terminus of the mVirD2 proteins to direct the T-complex into the chloroplast instead of the nucleus. Our results indicated that the chloroplast genome was not transformed despite all the modifications (see Chapter 2). We hypothesized that a novel unidentified NLS was still driving the modified VirD2 to the nucleus in spite of the presence of the transit peptide. We generated various fragments of the VrD2 or mVirD2 fused to EGFP to investigate the presence of a novel cryptic NLS.

3.5.1. Chloroplast targeting by DCL-TP and RbcS-TP

The tomato DLC or tobacco RbcS transit peptide was fused to the modified VirD2 proteins with inactivated NLSs (see Chapter 2). However, these transit peptides were not functional in *Agrobacterium*-mediated chloroplast transformation; therefore, we re-examined DCL and RbcS transit peptides to observe their chloroplast targeting efficiency in our plant systems. DCL-TP-EGFP and RbcS-TP-EGFP were exclusively localized in chloroplasts in leaf cells of transgenic *Arabidopsis* and transiently transformed tobacco plants (Figure 3.4.3.1a and b). Therefore, DCL-TP and RbcS-TP are effective chloroplast targeting signals for the recombinant proteins.

Various transit peptides have been examined for their efficiency in chloroplast targeting of recombinant proteins. The RbcS transit peptides of different plant species have been tested for their efficiency in chloroplast targeting. Interestingly, the presence of a short amino acid sequence from the mature protein can influence the import of the recombinant proteins into the chloroplasts. For example, the presence of first 24 amino acids from the mature pea RbcS transit peptide was necessary for chloroplast import of the fusion protein of RbcS-TP fused to 5-enolpyruvyl 3-phosphoshikimate synthase of *Salmonella typhimurium*. The recombinant protein accumulated in the chloroplasts of tobacco protoplasts and transgenic tobacco (Comai et al., 1988). The recombinant protein without the fragment of the mature

protein failed to be detected in the chloroplasts (Comai et al., 1988). However, another recombinant protein, the pea RbcS transit peptide fused to the bacterial neomycin phosphotransferase II (NPT-II), was constructed only with the first methionine of mature RbsS and seven amino acid linker between the RbcS transit peptide and NPT-II. This recombinant protein successfully accumulated in chloroplasts of transgenic tobacco plants (Vandenbroeck et al., 1985). This approach was employed using maize or pea RbcS transit peptide fused to the yellow fluorescent protein or GFP, respectively. Fusion proteins were localized in chloroplasts of transgenic maize (Sattarzadeh et al., 2010). Tobacco RbcS transit peptide is another robust chloroplast targeting signal in monocot and dicot plants. For example, tobacco RbcS transit peptide with five amino acids from mature RbcS protein was fused to GFP. The recombinant protein was detected in chloroplasts of monocot and dicot plants (Gnanasambandam et al., 2007). DCL transit peptidewith six amino acids from mature DCL protein fused to GFP was also employed, and was also localized to chloroplasts of the monocot and dicot plants (Gnanasambandam et al., 2007). In our study, RbcS-TP-EGFP contains the transit peptide plus five amino acids from the mature RbcS protein, while DCL-TP-EGFP includes only one amino acid from the mature DCL protein. The length of the mature protein in the recombinant proteins was not critical for DCL-TP-EGFP and RbcS-TP-EGFP to be imported into chloroplasts in our case.

3.5.2. Chloroplast/nuclear targeting of TP-mVirD2

DCL-TP and RbcS-TP showed a strong chloroplast targeting in DCL-TP-EGFP and RbcS-TP-EGFP; however, when these transit peptides were fused to mVirD2, they did not function as efficient as TP-EGFP. DCL-TP-mVirD2-58-EGFP was detected only in the nucleus of both transgenic *Arabidopsis* and transiently transformed tobacco leaves (Figure 3.4.3.2a). DCL-TP-mVirD2-80-EGFP was also observed mainly in the nucleus in *Arabidopsis*, whereas a small amount of the recombinant protein was detected in the cytoplasm of tobacco leaves (Figure 3.4.3.2c). Therefore, DCL-TP was not functional to direct mVirD2 into chloroplasts. Interestingly, RbcS-TP-mVirD2-58-EGFP and RbcS-TP-mVirD2-80-EGFP were mainly localized in the nuclei and weakly in chloroplasts of *Arabidopsis* and tobacco leaves (Figure 3.4.3.2 b and d). These proteins contain five amino acids of the mature RbcS-TP which might have a positive impact on chloroplast targeting. Otherwise, it is possible that RbcS-TP can be more effective than DCL-TP for

chloroplast targeting of the recombinant proteins. Regardless of the presence of the strong transit peptides, TP-mVirD2 derivatives were largely localized in the nucleus.

Dysfunctional transit peptide has been observed previously. A rice serotonin *N*-acetyl transferase (SNAT) transit peptide was fused to the N-terminus of the sheep SNAT. This recombinant protein was localized in the cytoplasm of the transiently transformed tobacco leaves (Byeon et al., 2014). The weak chloroplast targeting of rice SNAT transit peptide and unsuitability of sheep SNAT as a mature protein were suggested as possibilities why the tested transit peptide was dysfunctional (Byeon et al., 2014). However, a definitive reason why the aforementioned transit peptide was dysfunctional is still not known. Since TP-mVirD2 was localized mainly in the nucleus, there must be another explanation other than the weak chloroplast targeting activity.

3.5.3. Lack of expression of full-length VirD2 and the central part of VirD2 in *Arabidopsis* or tobacco

In order to have a nuclear targeting control, wild type VirD2 was fused to EGFP (VirD2-EGFP). The central part of VirD2 was also fused to EGFP (VirD2-M1/3-EGFP) to investigate the presence of a putative NLS in the region. *Arabidopsis* plants stably expressing VirD2-EGFP and VirD2-M1/3-EGFP could not be obtained. Also, VirD2-EGFP and VirD2-M1/3-EGFP were not detected when they were transiently expressed in tobacco leaves. The constitutive expression of the full-length *virD2* gene from a cauliflower mosaic virus (CaMV) double 35S promoter was observed to be toxic to *Arabidopsis*, yet the reason of toxicity was unknown (Hwang et al., 2006). In our experiments, the *virD2* and *virD2-M1/3* genes were also expressed from the CaMV double 35S promoter in pEZT-NL. The expression of these proteins could also be toxic to plants. This toxicity might be caused by interaction with plant proteins. Cyclophilins and the TATA box-binding protein have been shown to interact with VirD2 (Bako et al., 2003; Deng et al., 1998). Figure 3.3.1 illustrates overlap of these domains with VirD2-M1/3. It is possible that a high expression of VirD2-EGFP and VirD2-M1/3-EGFP may cause unavailability of these plant proteins for normal plant function which leads to dysfunction in cellular regulations. Recently, VirD2 was revealed to interact with yeast histones H2A, H2B, H3 and H4 *in vitro* (Wolterink-van Loo et al., 2015). The biomolecular fluorescence complementation experiments confirmed the interaction between VirD2 and histones (Wolterink-van Loo et al., 2015). Continuous expression and/or transient expression of VirD2-

EGFP and VirD2-M1/3-EGFP might cause instability of the chromatin structure which may prevent viable seed production when floral dip infiltration is used for transformation.

3.5.4. Nuclear targeting of mVirD2

NLSs of both mVirD2-58 and mVirD2-80 have been modified by introducing a point mutation (from arginine to threonine) to the N-terminal NLS and by removing the C-terminal NLS by a broad or a precise deletion. mVirD2-58-EGFP and mVirD2-80-EGFP were exclusively localized in the nucleus in *Arabidopsis* and tobacco leaves, and aggregates were observed only in the transiently transformed tobacco leaves (Figure 3.4.3.3a and b). Aggregates were likely because a very high level of expression from the vector leads to the saturation and formation of aggregates upon transient transformation of plants (Escobara et al., 2003). Moreover, aggregates may have been formed because of changes in isoelectric point.

The subcellular localization of a recombinant protein may differ in transgenic plants and the transiently transformed plant system. For example, a fusion protein (AvrRps4-GFP) was detected in the cytoplasm and in the nucleus, while the recombinant protein was localized in chloroplasts in transgenic *Arabidopsis* (Heidrich et al., 2011; Li et al., 2014; Sohn et al., 2012). Since we did not observe the different subcellular localization of mVirD2-58-EGFP and mVirD2-80-EGFP in transgenic *Arabidopsis* and transiently transformed tobacco leaves, it was highly likely that these recombinant proteins without known NLSs localized in the nucleus of both plant systems. The nuclear localization of mVirD2-58 and mVirD280 was surprising because it was expected that mutation of both known NLSs of VirD2 would abolish nuclear targeting of the modified protein. This expectation was supported by previous studies on VirD2 with mutated N and C-terminal NLSs which showed that these mutations resulted in a dramatic reduction of plant transformation (Rossi et al., 1993) and in the cytoplasmic localization of GFP-mVirD2 in mammalian cells (Relic et al., 1998). This discrepancy may be due to use of different systems, since a nuclear targeting efficiency of modified NLS sequences is known to vary in yeast, plant and mammalian cells (Kosugi et al., 2009a). Our results suggested two possibilities: 1) modification of the N-terminal NLS did not inactivate its NLS activity and 2) a novel NLS exists in VirD2. The mVirD2 fragment-EGFP fusion proteins were analyzed for their subcellular localization to consider these possibilities.

3.5.5. Nuclear import of the N-terminal fragments of VirD2 and mVirD2

Although the wild type N-terminal NLS is present in VirD2-N1/2-EGFP, the recombinant protein was detected in the nucleus and cytoplasm of transgenic *Arabidopsis*, yet only aggregates were observed in the tobacco leaf (Figure 3.4.4.1a). These aggregates occur in the cells of tobacco leaves likely due to the high expression form the vector (Escobara et al., 2003). Additionaly, changes in isoelectric point may have caused aggregation. It was generally thought that proteins with a maximum size of 60 kDa could diffuse through the nuclear pore, yet proteins up to 90 - 110 kDa were found to enter passively the nuclei of mammalian cells (Wang and Brattain, 2007). The size of VirD2-N1/2-EGFP (54 kDa; Table 3.4.1) may cause diffusion of the protein into the nucleus in transgenic *Arabidopsis*.

mVirD2-58/80-N1/2-EGFP contains a point mutation (arginine to threonine) in the N-terminal NLS. Transgenic *Arabidopsis* carrying this construct could not be obtained, and transiently expressed protein only formed aggregates in the tobacco leaf (Figure 3.4.4.1a). Thus, the effect of this modification on the nuclear targeting activity of the N-terminal NLS could not be assessed, unfortunately.

VirD2-N1/3-EGFP (46 kDa) and mVirD2-58/80-N1/3-EGFP (46 kDa) mainly formed aggregates, and weak signals were detected in the nucleus and cytoplasm of tobacco leaf cells (Figure 3.4.4.1c and d). Since VirD2-N1/3-EGFP was not detected exclusively in the nucleus and mVirD2-58/80-N1/3-EGFP was not only in the cytoplasm, we cannot exclude the possible diffusion of the proteins through the nuclear pore due to their small sizes (Wang and Brattain, 2007). It is possible that small sizes of VirD2-N1/3-EGFP (46 kDa) and mVirD2-58/80-N1/3-EGFP (46 kDa) caused diffusion of these proteins into the nucleus easier compared to VirD2-N1/2-EGFP (54 kDa) and mVirD2-58/80-N1/2-EGFP (54 kDa) that were larger in size (Table 3.4.1).

VirD2-N1/2-EGFP and VirD2-N1/3 contain the wild type N-terminal NLS. Regardless of the size of these proteins, the presence of the N-terminal NLS should contribute to a strong nuclear targeting of these proteins as shown previously (summarized in Table 3.5a) (Relic et al., 1998; Tinland et al., 1992; Ziemienowicz et al., 2001). Different transformation systems and reporter proteins used in our and previous studies might contribute to this difference. VirD2 (11 amino acids containing the N-terminal NLS) fused to the β-galactosidase fusion protein was detected in the nucleus of *N. plumbaginifolia* mysophyll protoplasts (Tinland et al., 1992). GFP-VirD2 carrying the *Nru*I deletion of the C-terminal NLS was detected in the

nuclei of mammalian cells (Relic et al., 1998). Fluorescently labelled VirD2 carrying the *Nru*I deletion of the C-terminal NLS was detected in the nucleus of permeabilized *N. tabacum* (BY2) protoplasts from the cell suspension culture (Ziemienowicz et al., 2001). It has been shown that an addition of the reporter protein at the N or C-terminus of the protein of interest could alter the subcellular localization of fusion proteins (Marion et al., 2008). For example, when GFP was fused to the N or C-terminus of the *Arabidopsis* ceramide synthase (LOH1), the subcellular localization of fusion proteins was different in transiently transformed tobacco leaves: LOH1-GFP in the Golgi apparatus and GFP-LOH1 in the Golgi apparatus, the nucleus and the cytoplasm (Marion et al., 2008). A different transformation system might be another factor to alter the subcellular localization as seen before with a fusion protein AvrRps4-GFP (Heidrich et al., 2011; Li et al., 2014; Sohn et al., 2012). Thus, despite the presence of the N-terminal NLS in VirD2-N1/2-EGFP and VirD2-N1/3-EGFP, it was not strong enough to direct the fusion protein exclusively into the nucleus in transiently transformed tobacco leaves.

mVirD2-58/80-N1/2-EGFP and mVirD2-58/80-N1/2-EGFP have a modification in the N-terminal NLS (a change from arginine to threonine). Similar amino acid substitution was done in the unique region of in the capsid proteins of porcine parvovirus (Boisvert et al., 2014). This unique region is called five basic regions (BR1 through BR5). Various amino acid substitutions were introduced into those regions. When changes from arginine to threonine and from lysine to asparagines were introduced in BR2, the EGFP-BR2 mutant was detected throughout porcine fibroblast cells. Meanwhile, EGFP with other BRs containing mutations were slightly more enriched in the nucleus (Boisvert et al., 2014). Hence, it is possible that a change from arginine to threonine in the N-terminal NLS of VirD2 had an influence on its activity. However, since our results were ambiguous, we cannot conclude this amino acid change in fact inactivated the nuclear localization activity of the N-terminal NLS.

3.5.6. Nuclear import of the C-terminal fragments of VirD2 and mVirD2

VirD2-C1/2-EGFP containing the wild type C-terminal NLS of VirD2 was localized exclusively in the nucleus of the tobacco leaf (Figure 3.4.6.1a). Surprisingly, mVirD2-58-C1/2-EGFP and mVirD2-80-C1/2-EGFP carrying a broad and precise deletion of the C-terminal NLS were still exclusively localized in the nucleus (Figure 3.4.6.1b and c). In transgenic *Arabidopsis* cells, mVirD2-80-C1/2-EGFP was localized

largely in the nucleus and some in the cytoplasm (Figure 3.4.6.1c). This slight inconsistency could be caused by a stable transformation versus transient transformation as well as different plant systems used (Li et al., 2014), but it does not diminish the main finding that VirD2 lacking the known functional NLSs targeted the GFP fusion protein into the nucleus of the plant cell.

VirD2-C1/3-EGFP (46 kDa) still possesses the active C-terminal NLS; thus, its nuclear localization in transgenic *Arabidopsis* and transiently transformed tobacco was expected (Figure 3.4.6.1d). mVirD2-58-C1/3-EGFP and mVirD2-80-C1/3-EGFP still showed strong EGFP signals in the nucleus and moderate level of EGFP from the cytoplasm were detected in the transgenic *Arabidopsis* and the tobacco leaf cells (Figure 3.4.6.1e and f). It is inevitable that the size of mVirD2-58-C1/3-EGFP (41 kDa) and mVirD2-80-C1/3-EGFP (44 kDa) may cause the diffusion into the nucleus. Therefore, mVirD2-80-C1/3-EGFPx2 containing an extra alanine linker-EGFP was generated to increase the size of the protein to 71 kDa which may prevent a passive translocation of the recombinant proteins through the nuclear pore. The result presented that mVirD2-80-C1/3-EGFPx2 was exclusively localized in the nucleus (Figure 3.4.6.1g). This was surprising because when GUS was fused to the VirD2 lacking the C-terminal NLS by precise deletion, it was localized in the cytoplasm in tobacco protoplasts regardless of the N-terminal NLS (Howard et al., 1992). Furthermore, when GFP was fused to the N-terminus of VirD2 lacking both N-terminal NLS and the C-terminal NLS by *Nru*I deletion, the fusion protein was detected only in the cytoplasm of mammalian cells (Relic et al., 1998). Our analysis of subcellular localization was performed using two different systems - a stable transformation of *Arabidopsis* and transient transformation of tobacco. We constructed a recombinant protein of EGFP fused to the C-terminus of the VirD2 variants. It is possible that the difference in the transformation system and a reporter protein used for fusion in our case compared to previous studies might have affected the subcellular localization of the recombinant proteins (Heidrich et al., 2011; Li et al., 2014; Marion et al., 2008; Sohn et al., 2012). Since our results strongly suggested the presence of a novel NLS, we examined it more closely.

3.5.7. Discovery of a novel NLS3 in VirD2

According to the cNLS Map analysis (Figure 3.4.1a-c), there was a putative monopartite NLS, named NLS3 (QRSKRRNDEE). Due to technical reasons including a lack of success with cloning of

smaller fragments and a failure of transient expression in tobacco using shorter fragments fused to EGFP, we used NLS3 together with extra amino acid sequences from around NLS3 to construct the fusion protein. Thus, NLS3 indicates the following sequence hereafter: **M**KRQ**QRSKRRNDEE**AGPSGANRKGLK (**M** is newly added methionine). In addition, the peptide of the same length (26 amino acids) from GFP in pCAMBIA1302 was fused to EGFP to serve as a nuclear targeting negative control. Figure 3.4.7a and b shows that both NLS3-EGFP and 26aa-EGFP were detected in the nucleus and cytoplasm. These proteins are only 33 kDa, which can cause diffusion through the nuclear pore (Wang and Brattain, 2007; Xu and Meier, 2008). We applied the strategy used in previous studies to increase proteins by additional two and three alanine linker-EGFPs (Crawford and Zambryski, 2000; Wang and Brattain, 2007). NLS3-EGFPx2 and 26aa-EGFPx2 containing an extra alanine linker-EGFP, resulting in 60 kDa proteins, formed aggregates and localized mainly in the nucleus and some in the cytoplasm (Figure 3.4.7c). 26aa-EGFP showed a high expression and still was found in the nucleus and cytoplasm (Figure 3.4.7d). NLS3-EGFPx3 and 26aa-EGFPx3 are 88 kDa and 89 kDa, respectively, due to additional three alanine-EGFPs (Table 3.4.1). The results indicated clearly that NLS3-EGFPx3 was localized exclusively in the nucleus (Figure 3.4.7e), yet 26aa-EGFPx3 still likely diffused through the nuclear pore by passitve transport (Figure 3.4.7f). Therefore, presence of NLS3 was confirmed, and this sequence can be indicated in the updated scheme of the VirD2 protein (Figure 3.5.5a). NLS3 is not a strong nuclear targeting signal as the C-terminal NLS because VirD2-C1/3-EGFP containing the wild type C-terminal NLS (46 kDa) was found exclusively in the nucleus (Figure 3.4.6.1d), while NLS3-EGFP (33 kDa) and NLS3-EGFPx2 (60 kDa) were still detected in the nuclei and the cytoplasm (Figure 3.4.7a and c). Other studies did not detect the presence of NLS3; VirD2 lacking the C-terminal NLS was not able to direct the fluorescently labelled single stranded DNA into the nuclei of permeabilized tobacco protoplasts and mammalian cells (Table 3.5a) (Ziemienowicz et al., 1999; Ziemienowicz et al., 2001). It is not clear why this modified VirD2 that should have contained NLS3 was not functional to direct the complex into the nucleus in the aforementioned experimental systems. Moreover, GFP-VirD2 with the absent N- and C-terminal NLSs localized in the cytoplasm of mammalian cells (Relic et al., 1998), despite that it contained NLS3. It is possible that weak NLS3 was not recognized before in the different constructs and transformation systems. However, NLS3 was strong enough to override DCL-TP and RbcS-TP in the TP-mVirD2 proteins to direct the T-complex into the nucleus. This is

likely one of the reasons why *Agrobacterium*-mediated chloroplast transformation did not work (see Chapter 2).

We analyzed the VirD2 sequence from octopine and nopaline pTi types (Figure 3.5.5b). NLS3 is not highly conserved between octopine and nopalineplasmids, yet cNLS Mapper also predicted the presence of a putative monopartite NLS at the same region (RQRRAKRPHDD, score of 8, Figure 3.5.5b). Six classes of NLS and their consensus sequences have been identified. Class 1 and 2 are classical monopartite types and bind to a major binding site of importin α. Class 3 and 4 are noncanonical NLSs and bind to a minor binding site of importin α. Class 5 is the plant-specific NLS. The last class, class 6 is bipartite NLSs (Kosugi et al., 2009a). According to class 1 consensus sequences KR(K/R)R and K(K/R)RK (Kosugi et al., 2009a), NLS3 belongs to this class, and a more specific sequence of NLS3 is predicted to be KRRN. The presence of asparagines after arginine likely decreases the nuclear targeting activity. Interestingly, when asparagines in NLS3 (QRSKRR**N**DEE, score 7) is substituted to lysine and arginine, the scores become 9 and 12, respectively (Table 3.5b). It would be interesting to prove changes in strength of NLS activity upon single amino acid substitution experimentally. Due to technical reasons, NLS3 with extra sequences was fused in this study; nuclear targeting of NLS3 without the extra sequences should be tested in the future to confirm the specific sequence of NLS3. However, these sequences beside NLS3 may influence the efficiency of nuclear targeting (Kosugi et al., 2009a).

Class 5 NLS which is functional only in plant cells suggested that the efficiency of NLS activity vary in yeast, plant and mammalian cells (Kosugi et al., 2009a). Different modifictions in NLSs alter the strength of NLS acitivty in yeast, plant and mammalian cells (Kosugi et al., 2009a). For example, GUS-GFP-A**KR**I**R**VT (class 2 NLS; **Bold underlined)** NLS3 (score 7 by cNLS Mapper) is about a half of the strength of the C-terminal NLS (score 13.6 by cNLS Mapper). The predicted score of NLS3 and NLS sequence suggests that this NLS might not be efficient in the mammalian cells in which GFP-VirD2 with the absent N and C terminal NLSs but the present NLS3 was localized in the cytoplasm (Relic et al., 1998).

The identification of NLS3 can answer to some puzzling results from previous studies. Rossi et al. (1993) also suggested a cryptic NLS in VirD2 after they still observed a low transient expression (lowest 0.3%) of the transgene from VirD2 lacking the C-terminal NLS. In fact, the presence of a putative NLS (**KRR**NDEEAGPSGAN**RR**GL**K** – bold letters for nuclear targeting) was proposed by Shurvinton et al. in

1992. In their stduy, VirD2 proteins carrying different C-terminal trunctions were used for tumorigenesis on tobacco stems and potato discs (Shurvinton et al., 1992). A precise deletion of the C-terminal NLS and the putative NLS reduced the tumorigenesis to 60% (Shurvinton et al., 1992), but the deletion of the putative NLS resulted only in a 10% reduction of tumorigenesis Therefore, the putative additional NLS was never taken into account to test its nuclear targeting activity. The identification of NLS3 determined that their prediction on a putative NLS was correct, yet it cannot be explained why a high tumorigenesis efficiency was still observed. The involvement of VirE2 in nuclear targeting of the T-complex was proposed (Shurvinton et al., 1992). The importance of VirE2 proteins in T-DNA transfer was analyzed by Gelvin (1998). Transgenic tobacco overexpressing VirE2 was used for leaf disc transformation. *Agrobacterium virE2-virD2*ΔNLS (a precise deletion of the C-terminal NLS) double mutation was capable of forming tumors on such transgenic tobacco leaves, while leaves of wild-type plants were not transformed (Gelvin, 1998). Hence, the importance of VirE2 was suggested, yet its nuclear tareting of the T-complex and its subcellular localization remain controversial due to conflicting results, as reviewed elsewhere (Gelvin, 2012).

Despite the fact that many VirD2 domains including a relaxase domain, NLSs, and ω domain are well studied, there is a large domain of unknown function (DUF) that resides between the relaxase domain and the C-terminal NLS. DUF was examined thoroughly using several VirD2 truncations possessing a different combination of the C-terminal NLS, ω domain, a translocation signal from VirF, and parts of DUF supplemented with NLS of a simian virus 40 (SV40) large tumor (T)-antigen. These VirD2 proteins were used in transient and stable transformation in the *Arabidopsis* by root transformation assay. Translocation domain was found to be in the region of the C-terminal 40 amino acids of VirD2 including the C-terminal NLS and ω domain (van Kregten et al., 2009). Importantly, a region of DUF (204-264 amino acids) strongly enhances the T-strand transfer, whereas a section of amino acids 265-373 was said to be dispensable (van Kregten et al., 2009). Another function of VirD2 is mediating DNA substrate docking with VirD4 via its C-terminus; this interaction is necessary for the activation of VirB10 in type IV secretion system (Cascales et al., 2013). Unfortunately, our NLS3 is in the region of 335-344 amino acids within the so-called dispensable region (265-373 amino acids). Since van Kregten et al. (2009) supplemented VirD2

variants with SV40 NLS, it would be intersting to study the transformation using VirD2 carrying only the relaxase domain, NLS3, DUF (204-264 amino acid) and the translocation domain of VirF.

Although the previous study indicated that NLS3 did not have a significant impact on transformation (Shurvinton et al., 1992), the presence of NLS3 is evident. Our finding provides a new feature of VirD2.

3.5.8. Protein size limit to diffuse through the nuclear pore

It was thought that the maximal size of a protein that can diffuse through the nuclear pore was 60 kDa which was predicted by the diameter (about 9-12 nm) of the nuclear pore complex (Ma et al., 2012; Wang and Brattain, 2007). However, Wang and Brattain (2007) reported the passive nuclear import of proteins much larger than 60 kDa in mammalian cells. Plasmids encoding several GFP fusion proteins (GFP1-GFP5) with a protein range of 28 kDa to 140 kDa were constructed, and the subcellular localization of these GFP proteins was observed 8 h after transfection of HeLa cells (Wang and Brattain, 2007). GFP3 with the size of around 90 kDa showed an efficient passive diffusion through the nuclear pore, while GFP4 with the size of around 110 kDa was passively imported in a less efficient manner. However, GFP5 around 140 kDa was predominantly localized in the cytoplasm (Wang and Brattain, 2007). Wiegert et al. (2007) reported that the different transformation system affected the size limit for passive protein import through the nuclear pore. A plasmid encoding the extracellular signal-regulated kinase was fused to the N-terminus of GFP (about 70 kDa), and this fusion protein was localized in the nucleus and cytoplasm of the transfected hippocampal neurons due to the passive nuclear import (Wiegert et al., 2007). When bovine serum albumin (BSA 40 kDa) was introduced to hepatoma cells by microinjection, BSA was excluded from the nucleus. This suggested the 40 kDa was an exclusion limit (Lang et al., 1986). These previous studies showed the importance of the type of the system used to observe the protein subcellular localization; the system used can affect the size limit of the passive nuclear import of proteins. Most importantly, the maximal size of proteins to diffuse through the nuclear pore was suggested to be from 90 to 110 kDa, and this can also depend of the structure of the protein (Wang and Brattain, 2007).

In plant systems, the size exclusion limit (up to 60 kDa) of the passive diffusion through the nuclear pore was thought to be the same as in the mammalian system. Our finding and previous studies

raise a question on the passive nuclear import in plants. For example, when dextrans (70 kDa) were introduced to permeabilized tobacco protoplasts by the nuclear import assay, dextran molecules did not diffuse through the nuclear pore, which suggested the nuclear exclusion limit was 70 kDa (Merkle et al., 1996). In this study, we presented that 26aa-EGFPx3 (88 kDa) was still able to be passively translocated through the nuclear pore in the transiently transformed tobacco cells (Figure 3.4.7f). GFP4 (110 kDa) has a compact structure (Wang and Brattain, 2007). 26aa-EGFPx3 propably has a compact structure as well, which made the passive nuclear import easier. It would be useful to have information on the passive nuclear import of proteins in different protein and transformation systems.

In conclusion, we did not only discover a novel NLS3 in VirD2 but also provided useful information about protein diffusion through the nuclear pore using this particular transformation system.

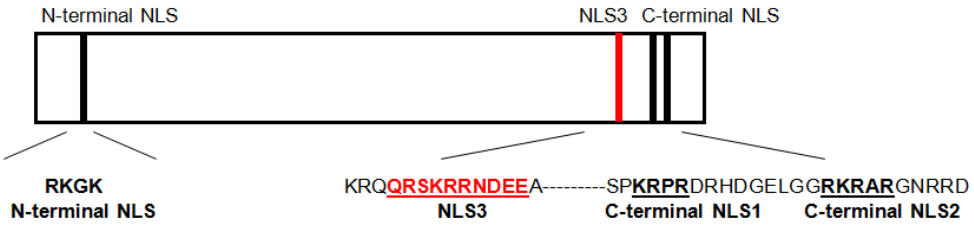


Figure 3.5.5a. Schematic representation of VirD2 with the proposed VirD2 NLS3.
Dash line: omitted amino acids.

Table 3.5a. Summary of previous studies on the nuclear targeting properties of VirD2 NLSs.

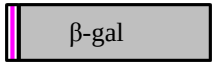
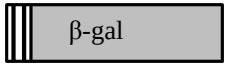
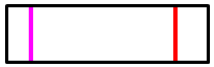
Organisms	Transformation methods	Stable/ Transient	Detection methods	References	VirD2 variants	Subcellular localization
N. tabacum (W38) Leaf discs	*Agro*-mediated	Stable	β-galactosidase activity using cellular factions	(Herrera-Estrella et al., 1990)	 First 292 amino acids of VirD2	Nucleus
N. tobacum (XD) Protoplasts from cell suspension culture	Electroporation	Transient	GUS activity in protoplasts	(Howard et al., 1992)		Nucleus
						Cytoplasm
					 Precise deletion of the C-terminal NLS	Cytoplasm
N. plumbaginifolia Mesophyll protoplast	PEG	Transient	Indirect immunofluorescence	(Tinland et al., 1992)	 11 amino acids including the N-terminal NLS	Nucleus
					 20 amino acids including the C-terminal NLS	Nucleus
Mammalian cells: HEK 293 cells HeLa cells	Calcium phosphate method-transfection	Transient	Indirect Immunofluorescence Staining	(Relic et al., 1998)		Nucleus
					 *Nru*I deletion of the C-terminal NLS	Nucleus
					 Deletion of the N-terminal NLS *Nru*I deletion of the C-terminal NLS	Cytoplasm
Mammalian cells: HeLa cells	Permeabilized cell	Transient	Confocal microscopy: Fluorescently labelled single stranded DNA	(Ziemienowicz et al., 1999)	 + ssDNA *Nru*I deletion of the C-terminal NLS	Cytoplasm (ssDNA)

Table 3.5a. Continued.

Organisms	Transformation methods	Stable/ transient	Detection methods	References		Subcellular localization
N. tabacum (BY2) Protoplasts from cell suspension	Permeabilized protoplasts	Transient	Confocal microscopy: Fluorescently labelled protein	(Ziemienowicz et al., 2001)	 *Nru*I deletion of the C-terminal NLS	Nucleus
			Confocal microscopy: Fluorescently labelled single stranded DNA		 + ssDNA *Nru*I deletion of the C-terminal NLS	Cytoplasm (ssDNA)

Pink line: N-terminal NLS; Red line: NLS3; Black lines: C-terminal NLS.

Table 3.5b. cNLS Mapper prediction.

Name	Predicted NLS sequence	Score
NLS3	QRSKRRN**N**DEE	7
NLS3 (N→K)	QRSKRR**K**KDEE	9
NLS3 (N→R)	QRSKRR**R**DEE	13

Changed amino acid is indicated in red.

```
pTi15955_octopine      MPDRAQVIIRIVPGGGTKTLQQIINQLEYLSRKGKLELQRSARHLDIPVPPDQIRELAQS 60
pAch5_octopine         MPDRAQVIIRIVPGGGTKTLQQIINQLEYLSRKGKLELQRSARHLDIPVPPDQIRELAQS 60
pTiBo542_octopine      MPDRAQVIIRIVPGGGTKTLQQIINQLEYLSRKGKLELQRSARHLDIPVPPDQIRELAQS 60
pTiC58_nopaline        MPDRAQVIIRIVPGGGTKTLQQIINQLEYLSRKGRLELQRSARHLDIPLPPDQIHELARS 60
                       **********************************:*************:*****:***:*

pTi15955_octopine      WVTEAGIYDESQSDDDRQQDLTTHIIVSFPAGTDQTAAYEASREWAAEMFGSGYGGGRYN 120
pAch5_octopine         WVTEAGIYDESQSDDDRQQDLTTHIIVSFPAGTDQTAAYEASREWAAEMFGSGYGGGRYN 120
pTiBo542_octopine      WVTEAGIYDESQSDDDRQQDLTTHIIVSFPAGTDQTAAYEASREWAAEMFGSGYGGGRYN 120
pTiC58_nopaline        WVQETGTYDESQPDEERQQELTTHIIVSFPAGTSQVAAYAASREWAAEMFGSGAGGGRYN 120
                       ** *:* *****.*::***:*************.*.*** ************* ******

pTi15955_octopine      YLTAYHVDRDHPHLHVVVNRRELLGHGWLKISRRHPQLNYDGLRKKMAEISLRHGIVLDA 180
pAch5_octopine         YLTAYHVDRDHPHLHVVVNRRELLGHGWLKISRRHPQLNYDGLRKKMAEISLRHGIVLDA 180
pTiBo542_octopine      YLTAYHVDRDHPHLHVVVNRRELLGQGWLKISRRHPQLNYDGLRKKMAEISLRHGIVLDA 180
pTiC58_nopaline        YLTAFHIDRDHPHLHVVVNRRELLGHGWLKISRRHPQLNYDALRIKMAEISLRHGIALDA 180
                       ****:*:*******************:***************.** ***********.***

pTi15955_octopine      TSRAERGIAERPITYAEHRRLERMQAQKIQFEDTDFDETSPEEDRRDLSQSFDPFRSDPS 240
pAch5_octopine         TSRAERGIAERPITYAEHRRLERMQAQKIQFEDTDFDETSPEEDRRDLSQSFDPFRSDPS 240
pTiBo542_octopine      TSRAERGIAERPITYAEYRRLERMQAQKIQFEDTDFDETSPEEDRRDLSQSFDPFRSDAS 240
pTiC58_nopaline        SRRAERGITERPITYAQYRRLEREQARQIRFEDADLEQSSPQGDHPEFSQPFDTSPFEAS 240
                       : ******:*******::***** **::*:***:*::::**: *: ::**.**.   :.*

pTi15955_octopine      TGEPDRATRHDKQPLEQHARFQESAGSSIKADARIRVSLESERSAQPSASKIPVIGHFGI 300
pAch5_octopine         TGEPDRATRHDKQPLEQHARFQESAGSSIKADARIRVSLESERSAQPSASKIPVIGHFGI 300
pTiBo542_octopine      AGEPDRATRHDKQPLEPHARFQEPAGSSIKADARIRVPLESERGAQPSASKIPVTGHFGI 300
pTiC58_nopaline        AGGPEDMPRPNNRQNESQVHLQEPAGVSNEAGVLVRVALETERLAQPFVSETILADDIGS 300
                       :* *:  .* :::  * :.::**.** * :*.. :**.**:** *** .*:  : ..:*

pTi15955_octopine      ETSYVAEASVRKRSGIFGTSRPVTDVAMHTVK-RQQRSKRRNDEEAGPSGANRKGLKAAQ 359
pAch5_octopine         ETSYVAEASVRKRSGIFGTSRPVTDVAMHTVK-RQQRSKRRNDEEAGPSGANRKGLKAAQ 359
pTiBo542_octopine      ETSYVAEASVPKQSGNSDTSRPVTDVAMHTVE-RQQRSKRRHDEEAGPSGANRKRLKAAQ 359
pTiC58_nopaline        GSSRVAEGRVESANRTPDIPRAATEAATHTTHDRQRRAKRPHDDDGGPSGAKRVTLEGIA 360
                        :* ***. * . .   . .*..*:.* **.. **:*:** :*::.*****:*  *:.
```

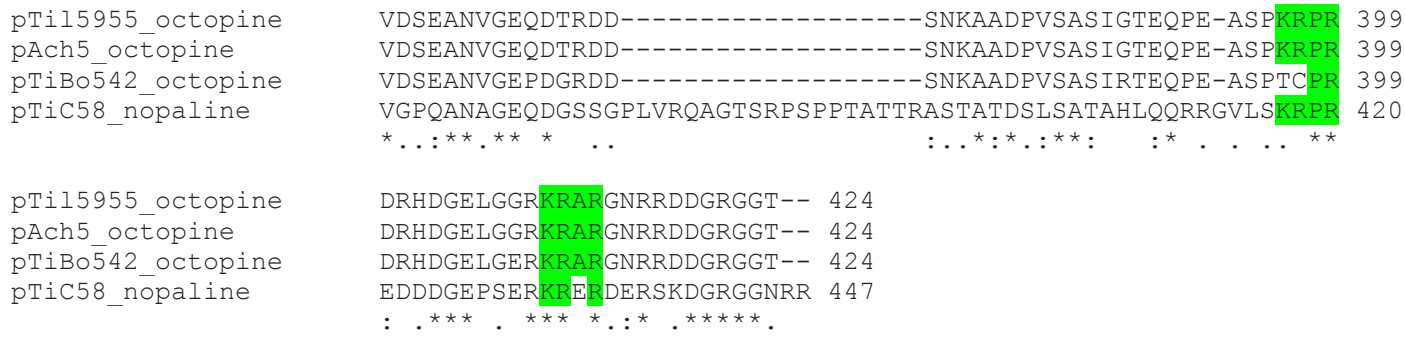


Figure 3.5.5b. Alignment of VirD2 amino acid sequences encoded by octopine- and nopaline-type plasmids.
Sequences are retrieved from NCBI (Accession number - pTil5955: NG_035347; pAch5: CP007228; pTiBo542: DQ058764; pTiC58: NC_003065) and aligned using ClustalW2, (Letters highlighted in yellow: NLS3; Letters highlighted in green: C-terminal NLS;
Red box: sequence used for the recombinant protein; Bold letters: predicted NLS by cNLS Mapper).

Chapter 4: Conclusions and Future directions

In this project, VirD2 was modified to inactivate NLSs (mVirD2): a point mutation (arginine to threonine) in the N-terminal NLS and removal of the C-terminal NLS broadly or precisely based on the previous study (Rossi et al., 1993). The strong chloroplast transit peptide either tomato DCL or tobacco RbcS was fused to the N-terminus of mVirD2. As a result, four types of TP-mVirD2 were constructed: DCL-TP-mVirD2-58, RbcS-TP-mVirD2-58, DCL-TP-mVirD2-80 and RbcS-TP-mVirD2-80. RbcS-TP-mVirD2-58 was successfully purified in this study, and it was shown to retain the cleavage activity. This indicated that the presence of the transit peptide did not disturb the cleavage activity of TP-mVirD2. I attempted to transform tobacco chloroplasts by *Agrobacterium* carrying the *TP-mvirD2* gene and a binary vector containing the chloroplast expression cassette to achieve the initial objective of development of an *Agrobacterium*-mediated chloroplast transformation method. Results indicated that *Agrobacterium*-mediated chloroplast transformation was unsuccessful using TP-mVirD2. These undesirable results, however, created an interesting question: "Why are strong chloroplast transit peptides dysfunctional in the TP-mVirD2 proteins without any known functional NLSs?" We hypothesized the presence of an unknown NLS in VirD2. We tested our hypothesis in the second project.

Several fragments of VirD2 and mVirD2 were fused to EGFP. These fusion proteins were used to observe their subcellular localization in stably transformed *Arabidopsis* and transiently transformed tobacco leaf cells. The fusion proteins containing the N-terminal one third of VirD2/mVirD2 containing either the wild type NLS or the modified NLS were localized in the nucleus and the cytoplasm of the tobacco cells. Since these fusion proteins were small in size (46 kDa), it probably allowed protein diffusion through the nuclear pore. Therefore, it is inconclusive that the point mutation indeed inactivated the N-terminal NLS, and it requires further investigation. Fusion proteins containing the C-terminal part of mVirD2 without any known functional NLS were localized exclusively in the nucleus. The amino acid NLS prediction software (Kosugi et al., 2009b) indicated the putative NLS (NLS3) upstream of the C-terminal NLS. We confirmed the presence of the novel NLS (NLS3) in VirD2 by showing exclusive nuclear localization of the NLS3-EGFPx3 fusion protein in tobacco cells.

Another interesting finding was that fusion protein 26aa-EGFPx3 (88 kDa) serving as negative control for the nuclear import was still able to be passively imported through the nuclear pore. The size

exclusion of the protein is up to 110 kDa in the mammalian cells. This was in contrast to the previously published data demonstrating that the protein size exclusion of diffusion through the nuclear pore was 70 kDa in the evacuolated permeablized tobacco cells (Merkle et al., 1996). This suggested that the type of transformation method, organisms and reporter proteins likely affect the protein diffusion through the nuclear pore. Our results suggest that in the agro-infiltrated tobacco leaf, 88 kDa (26aa-EGFPx3) can be passively imported into the nucleus through the nuclear pore easily. In the future, it would be useful to obtain data indicating the diffusion limit according to the type of transformation method, organism and reporter proteins. For this, we can construct larger proteins by adding one or two EGFPs to 26aa-EGFPx3 resulting in 26aa-EGFPx4 (117 kDa) and 26aa-EGFPx5 (145 kDa). Subcellular localization of these proteins can be determined by confocal microscopy using agro-infiltrated tobacco leaves. Western blotting needs to be performed to ensure the proper size of the recombinant proteins are produced in the plant cells. These data would help to define the diffusion limit of EGFP constructs in the agro-infiltrated tobacco leaves.

I was able to achieve the second objective by discovering a new NLS in VirD2 protein - NLS3. Although I could not develop a method of *Agrobacterium*-mediated chloroplast transformation, information provided in Table 3.5b could be useful to improve nuclear targeting of the T-complex in *Agrobacterium*-mediated plant transformation in general. It might be interesting to investigate if a change in NLS3 from aspartate to lysine or arginine (score 9 or 13, respectively in Table 3.5b), improves the plant nuclear genome transformation efficiency. Moreover, interaction of NLS3 with the nuclear import machinery could be studied. *Arabidopsis* contains nine genes encoding importin α proteins (Bhattacharjee et al., 2008). Bimolecular fluorescence complementation (BiFC) can be used to determine which importin α proteins strongly interact with NLS3. In addition, a change from aspartate to lysine or arginine in NLS3 may alter the strength of interactions between NLS3 and importin α.

According to van Kregten et al. (2009), NLS3 is in the "dispensable" region of VirD2, since it is not a part of the minimal region of VirD2 required for transformation of the nuclear genome. Hence, for *Agrobacterium*-mediated chloroplast transformation, VirD2 can be modified to a short VirD2 (1-204 amino acids of the relaxase domain) supplemented with the translocation signal from VirF at its C-terminus and a transit peptide at its N-terminus. In my study, I used tomato DCL and tobacco RbcS transit peptides because they are strong transit peptides to import the recombinant proteins into chloroplasts. The non-

photosynthetic ferredoxin III precursor (prFdIII) and root $NADP^+$-oxidoreductase (prFNR) precursor were imported into root plastids (Li and Teng, 2013). Transit peptides of these proteins might also be useful for fusion to a short VirD2 for targeting the T-complex to root plastids.

Another important factor VirE2 carries a bipartite NLS (Citovsky et al., 1992). It is likely that this protein also has to be modified to avoid the possibility of nuclear targeting of the T-complex. However, it is inevitable to abolish VirE2-VirE2 interaction while the bipartite NLS is removed because it overlaps with one cooperativity domain which is necessary for cooperative interaction between VirE2 and single stranded DNA (Citovsky et al., 1992; Sundberg and Ream, 1999). A few amino acids insertion between the bipartite NLS of VirE2 retained the single stranded DNA binding activity of VirE2 but caused a significant reduction in tumorigenesis (Dombek and Ream, 1997). At first, subcellular localization of the modified VirE2 protein needs to be analyzed. VirE2-YEP (yellow fluorescent protein) and YEP-VirE2 fusion proteins showed cytoplasmic aggregates in *Arabidopsis* root, while BiFC showed interactions between VirE2 and several importin α proteins (Bhattacharjee et al., 2008). Thus, BiFC would be useful to determine if modified VirE2 interacts with importin α proteins.

Protocols for *Agrobacterium*-mediated plant transformation are optimized for several monocot species (see Chapter 1). It might be possible that these optimized protocols are effective with use of *Agrobacterium* carrying the modified VirD2 and VirE2 for chloroplast transformation.

My initial attempt to develop a new method to transform chloroplast by modifying *Agrobacteirum* genetics was unsuccessful at this time. However, because of the results from these first experiments, an endeavor was taken to discover a novel NLS3. My results provided useful information on VirD2.

References

Ali, M.A., Shah, K.H., and Bohlmann, H. (2012). pMAA-Red: a new pPZP-derived vector for fast visual screening of transgenic *Arabidopsis* plants at the seed stage. BMC biotechnology *12*, 11.

Aly, K.A., and Baron, C. (2007). The VirB5 protein localizes to the T-pilus tips in *Agrobacterium tumefaciens*. Microbiol-Sgm *153*, 3766-3775.

An, G.H. (1985). High-efficiency transformation of cultured tobacco cells. Plant Physiol *79*, 568-570.

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 J *26*, 2540-2551.

Atmakuri, K., and Christie, P.J. (2008). Translocation of oncogenic T-DNA and effector proteins to plant cells (Springer, 233 Spring Street, New York, Ny 10013, United States).

Bako, 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. Proceedings of the National Academy of Sciences of the United States of America *100*, 10108-10113.

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

Bartlett, J.G., Alves, S.C., Smedley, M., Snape, J.W., and Harwood, W.A. (2008). High-throughput *Agrobacterium*-mediated barley transformation. Plant Methods *4*, 12.

Barton, K.A., Binns, A.N., Matzke, A.J.M., and Chilton, M.D. (1983). Regeneration of intact tobacco plants containing full length copies of genetically engineered T-DNA, and transmission of T-DNA to R1-progeny. Cell *32*, 1033-1043.

Bayliss, R., Harris, R., Coutte, L., Monier, A., Fronzes, R., Christie, P.J., Driscoll, P.C., and Waksman, G. (2007). NMR structure of a complex between the VirB9/VirB7 interaction domains of the pKM101 type IV secretion system. Proceedings of the National Academy of Sciences of the United States of America *104*, 1673-1678.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., and Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. Plant Methods *9*, 10.

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

Bilichak, A. (2013). The role of epigenetics in the maintenance of plant genome stability. In Department of Biological Sciences (University of Lethbridge, Canada).

Bilichak, A., Yao, Y.L., and Kovalchuk, I. (2014). Transient down-regulation of the RNA silencing machinery increases efficiency of *Agrobacterium*-mediated transformation of *Arabidopsis*. Plant Biotechnol J *12*, 590-600.

Block, M.D., Schell, J., and Montagu, M.V. (1985). Chloroplast transformation by *Agrobacterium tumefaciens*. EMBO J *4*, 1367-1372.

Bock, R. (2007). Structure, function, and inheritance of plastid genomes. In Topics in Current Genetics, R. Bock, ed. (Springer-Verlag Berlin Heidelberg).

Bock, R. (2014). Genetic engineering of the chloroplast: novel tools and new applications. Curr Opin Biotech *26*, 7-13.

Boisvert, M., Bouchard-Levesque, V., Fernandes, S., and Tijssen, P. (2014). Classic nuclear localization signals and a novel nuclear localization motif are required for nuclear transport of porcine parvovirus capsid proteins. J Virol *88*, 11748-11759.

Bonnard, G., Tinland, B., Paulus, F., Szegedi, E., and Otten, L. (1989). Nucleotide sequence, evolutionary origin and biological role of a rearranged cytokinin gene isolated from a wide host range biotype III *Agrobacterium* strain. Mol Gen Genet *216*, 428-438.

Boyko, A., Matsuoka, A., and Kovalchuk, I. (2009). High frequency *Agrobacterium tumefaciens*-mediated plant transformation induced by ammonium nitrate. Plant Cell Rep *28*, 737-757.

Boyko, A., Matsuoka, A., and Kovalchuk, I. (2011). Potassium chloride and rare earth elements improve plant growth and increase the frequency of the *Agrobacterium tumefaciens*-mediated plant transformation. Plant Cell Rep *30*, 505-518.

Boynton, J.E., Gillham, N.W., Harris, E.H., Hosler, J.P., Johnson, A.M., Jones, A.R., Randolphanderson, B.L., Robertson, D., Klein, T.M., Shark, K.B., *et al.* (1988). Chloroplast transformation in *Chlamydomonas* with high-velocity microprojectiles. Science *240*, 1534-1538.

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.

Briesemeister, S., Rahnenfuhrer, J., and Kohlbacher, O. (2010a). Going from where to why-interpretable prediction of protein subcellular localization. Bioinformatics *26*, 1232-1238.

Briesemeister, S., Rahnenfuhrer, J., and Kohlbacher, O. (2010b). YLoc-an interpretable web server for predicting subcellular localization. Nucleic Acids Res *38*, W497-W502.

Byeon, Y., Lee, H.Y., Lee, K., and Back, K. (2014). A rice chloroplast transit peptide sequence does not alter the cytoplasmic localization of sheep serotonin N-acetyltransferase expressed in transgenic rice plants. Journal of pineal research *57*, 147-154.

Cai, C.Q., Doyon, Y., Ainley, W.M., Miller, J.C., DeKelver, R.C., Moehle, E.A., Rock, J.M., Lee, Y.L., Garrison, R., Schulenberg, L., *et al.* (2009). Targeted transgene integration in plant cells using designed zinc finger nucleases. Plant Mol Biol *69*, 699-709.

Cascales, E., Atmakuri, K., Sarkar, M.K., and Christie, P.J. (2013). DNA substrate-induced activation of the *Agrobacterium* VirB/VirD4 type IV secretion system. Journal of Bacteriology *195*, 2691-2704.

Chandler, M., de la Cruz, F., Dyda, F., Hickman, A.B., Moncalian, G., and Ton-Hoang, B. (2013). Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nat Rev Microbiol *11*, 525-538.

Chandran, V., Fronzes, R., Duquerroy, S., Cronin, N., Navaza, J., and Waksman, G. (2009). Structure of the outer membrane complex of a type IV secretion system. Nature *462*, 1011-U1066.

Chen, Q.J., Zhou, H.M., Chen, J., and Wang, X.C. (2006). A Gateway-based platform for multigene plant transformation. Plant Mol Biol *62*, 927-936.

Cheng, M., Fry, J.E., Pang, S., Zhou, H., Hironaka, C.M., Duncan, D.R., Conner, T.W., and Wan, Y. (1997). Genetic transformation of wheat mediated by *Agrobacterium tumefaciens*. Plant Physiol *115*, 971-980.

Chichkova, N.V., Kim, S.H., Titova, E.S., Kalkum, M., Morozov, V.S., Rubtsov, Y.P., Kalinina, N.O., Taliansky, M.E., and Vartapetian, A.B. (2004). A plant caspase-like protease activated during the hypersensitive response. Plant Cell *16*, 157-171.

Choi, C.S., and Sano, H. (2007). Abiotic-stress induces demethylation and transcriptional activation of a gene encoding a glycerophosphodiesterase-like protein in tobacco plants. Mol Genet Genomics *277*, 589-600.

Citovsky, V., Guralnick, B., Simon, M.N., and Wall, J.S. (1997). The molecular structure of *Agrobacterium* VirE2-single stranded DNA complexes involved in nuclear import. Journal of Molecular Biology *271*, 718-727.

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. Proceedings of the National Academy of Sciences of the United States of America *86*, 1193-1197.

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

Coll, N.S., Epple, P., and Dangl, J.L. (2011). Programmed cell death in the plant immune system. Cell Death Differ *18*, 1247-1256.

Comai, L., Larsonkelly, N., Kiser, J., Mau, C.J.D., Pokalsky, A.R., Shewmaker, C.K., McBride, K., Jones, A., and Stalker, D.M. (1988). Chloroplast transport of a ribulose bisphosphate carboxylase small subunit-5-enolpyruvyl 3-phosphoshikimate synthase chimeric protein requires part of the mature small subunit in addition to the transit peptide. Journal of Biological Chemistry *263*, 15104-15109.

Costello, E., Saudan, P., Winocour, E., Pizer, L., and Beard, P. (1997). High mobility group chromosomal protein 1 binds to the adeno-associated virus replication protein (Rep) and promotes Rep-mediated site-specific cleavage of DNA, ATPase activity and transcriptional repression. Embo J *16*, 5943-5954.

Crawford, K.M., and Zambryski, P.C. (2000). Subcellular localization determines the availability of non-targeted proteins to plasmodesmatal transport. Curr Biol *10*, 1032-1040.

Daniell, H. (2002). Molecular strategies for gene containment in transgenic crops. Nat Biotechnol *20*, 581-586.

Daniell, H., Datta, R., Varma, S., Gray, S., and Lee, S.B. (1998). Containment of herbicide resistance through genetic engineering of the chloroplast genome. Nat Biotechnol *16*, 345-348.

Daniell, H., Khan, M.S., and Allison, L. (2002). Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. Trends Plant Sci *7*, 84-91.

Daniell, H., Kumar, S., and Dufourmantel, N. (2005). Breakthrough in chloroplast genetic engineering of agronomically important crops. Trends Biotechnol *23*, 238-245.

Das, A., and Xie, Y.H. (1998). Construction of transposon Tn3phoA: its application in defining the membrane topology of the *Agrobacterium tumefaciens* DNA transfer proteins. Mol Microbiol *27*, 405-414.

Davoodi-Semiromi, A., Schreiber, M., Nalapalli, S., Verma, D., Singh, N.D., Banks, R.K., Chakrabarti, D., and Daniell, H. (2010). Chloroplast-derived vaccine antigens confer dual immunity against cholera and malaria by oral or injectable delivery. Plant Biotechnol J *8*, 223-242.

Day, A., and Goldschmidt-Clermont, M. (2011). The chloroplast transformation toolbox: selectable markers and marker removal. Plant Biotechnol J *9*, 540-553.

Decleene, M. (1985). The susceptibility of monocotyledons to *Agrobcterium-tumefaciens*. Journal of Phytopathology *113*, 81-89.

Deng, W.Y., Chen, L.S., Peng, W.T., Liang, X.Y., 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*, 1795-1807.

Deng, W.Y., Chen, L.S., Wood, D.W., Metcalfe, T., Liang, X.Y., Gordon, M.P., Comai, L., and Nester, E.W. (1998). *Agrobacterium* VirD2 protein interacts with plant host cyclophilins. Proceedings of the National Academy of Sciences of the United States of America *95*, 7040-7045.

Dombek, P., and Ream, W. (1997). Functional domains of *Agrobacterium tumefaciens* single-stranded DNA-binding protein VirE2. Journal of Bacteriology *179*, 1165-1173.

Doonan, J.H., and Kitsios, G. (2009). Functional Evolution of Cyclin-Dependent Kinases. Mol Biotechnol *42*, 14-29.

Doyle, J.J., and Dickson, E.E. (1987). Preservation of Plant-Samples for DNA Restriction Endonuclease Analysis. Taxon *36*, 715-722.

Dudas, B., Jenes, B., Kiss, G.B., and Maliga, P. (2012). Spectinomycin resistance mutations in the *rrn16* gene are new plastid markers in *Medicago sativa*. Theor Appl Genet *125*, 1517-1523.

Durrenberger, F., Crameri, A., Hohn, B., and Koukolikovanicola, Z. (1989). Covalently bound VirD2 protein of *Agrobacterium tumefaciens* protects the T-DNA from exonucleolytic degradation. Proceedings of the National Academy of Sciences of the United States of America *86*, 9154-9158.

Dym, O., Albeck, S., Unger, T., Jacobovitch, J., Brainzburg, A., Michael, Y., Frenkiel-Krispin, D., Wolf, S.G., and Elbaum, M. (2008). Crystal structure of the *Agrobacterium* virulence complex VirE1-VirE2 reveals a flexible protein that can accommodate different partners. Proceedings of the National Academy of Sciences of the United States of America *105*, 11170-11175.

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of Molecular Biology *300*, 1005-1016.

Engler, G., Depicker, A., Maenhaut, R., Villarroel, R., Vanmontagu, M., and Schell, J. (1981). Physical mapping of DNA-base sequence homologies between an octopine and a nopaline Ti plasmid of *Agrobacterium tumefaciens*. Journal of Molecular Biology *152*, 183-208.

Escobara, N.M., Haupt, S., Thow, G., Boevink, P., Chapmana, S., and Oparka, K. (2003). High-throughput viral expression of cDNA-Green fluorescent protein fusions reveals novel subcellular addresses and identifies unique proteins that interact with plasmodesmata. Plant Cell *15*, 1507-1523.

Fernandez, D., Dang, T.A.T., Spudich, G.M., Zhou, X.R., Berger, B.R., and Christie, P.J. (1996). The *Agrobacterium tumefaciens virB7* gene product, a proposed component of the T-complex transport apparatus, is a membrane-associated lipoprotein exposed at the periplasmic surface. Journal of Bacteriology *178*, 3156-3167.

Fraga, H.P.F., Vieira, L.N., Caprestano, C.A., Steinmacher, D.A., Micke, G.A., Spudeit, D.A., Pescador, R., and Guerra, M.P. (2012). 5-Azacytidine combined with 2,4-D improves somatic embryogenesis of Acca sellowiana (O. Berg) Burret by means of changes in global DNA methylation levels. Plant Cell Rep *31*, 2165-2176.

Gao, R., Mukhopadhyay, A., Fang, F., and Lynn, D.G. (2006). Constitutive activation of two-component response regulators: characterization of VirG activation in *Agrobacterium tumefaciens*. J Bacteriol *188*, 5204-5211.

Garcia-Rodriguez, F.M., Schrammeijer, B., and Hooykaas, P.J.J. (2006). The *Agrobacterium* VirE3 effector protein: a potential plant transcriptional activator. Nucleic Acids Res *34*, 6496-6504.

Gasser, C.S., Gunning, D.A., Budelier, K.A., and Brown, S.M. (1990). Structure and expression of cytosolic cyclophilin peptidyl-prolyl cis-trans isomerase of higher-plants and production of active tomato cyclophilin in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America *87*, 9519-9523.

Gelvin, S.B. (1998). *Agrobacterium* VirE2 proteins can form a complex with T strands in the plant cytoplasm. Journal of Bacteriology *180*, 4300-4302.

Gelvin, S.B. (2003). *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. Microbiol Mol Biol Rev *67*, 16-37, table of contents.

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

Gelvin, S.B. (2010b). Plant proteins involved in *Agrobacterium*-mediated genetic transformation. Annu Rev Phytopathol *48*, 45-68.

Gelvin, S.B. (2012). Traversing the cell: *Agrobacterium* T-DNA's journey to the host genome. Front Plant Sci *3*, 11.

Gelvin, S.B., and Kim, S.I. (2007). Effect of chromatin upon *Agrobacterium* T-DNA integration and transgene expression. Biochim Biophys Acta-Gene Struct Expression *1769*, 410-421.

Gnanasambandam, A., Polkinghorne, I.G., and Birch, R.G. (2007). Heterologous signals allow efficient targeting of a nuclear-encoded fusion protein to plastids and endoplasmic reticulum in diverse plant species. Plant Biotechnol J *5*, 290-296.

Gonzalez-Rabade, N., McGowan, E.G., Zhou, F., McCabe, M.S., Bock, R., Dix, P.J., Gray, J.C., and Ma, J.K. (2011). Immunogenicity of chloroplast-derived HIV-1 p24 and a p24-Nef fusion protein following subcutaneous and oral administration in mice. Plant Biotechnol J *9*, 629-638.

Greer, M.S., Kovalchuk, I., and Eudes, F. (2009). Ammonium nitrate improves direct somatic embryogenesis and biolistic transformation of *Triticum aestivum*. New Biotech *26*, 44-52.

Ha, S.B., Lee, S.B., Lee, Y., Yang, K., Lee, N., Jang, S.M., Chung, J.S., Jung, S., Kim, Y.S., Wi, S.G., *et al.* (2004). The plastidic *Arabidopsis* protoporphyrinogen IX oxidase gene, with or without the transit sequence, confers resistance to the diphenyl ether herbicide in rice. Plant Cell Environ *27*, 79-88.

Hapfelmeier, S., Domke, N., Zambryski, P.C., and Baron, C. (2000). VirB6 is required for stabilization of VirB5 and VirB3 and formation of VirB7 homodimers in *Agrobacterium tumefaciens*. Journal of Bacteriology *182*, 4505-4511.

He, Z.Y., Li, L.G., and Luan, S. (2004). Immunophilins and parvulins. Superfamily of peptidyl prolyl isomerases in *Arabidopsis*. Plant Physiol *134*, 1248-1267.

Heidrich, K., Wirthmueller, L., Tasset, C., Pouzet, C., Deslandes, L., and Parker, J.E. (2011). *Arabidopsis* EDS1 connects pathogen effector recognition to cell compartment-specific immune responses. Science *334*, 1401-1404.

Herrera-Estrella, A., Chen, Z.M., Vanmontagu, M., and Wang, K. (1988). VirD proteins of *Agrobacterium tumefaciens* are required for the formation of a covalent dna protein complex at the 5' terminus of T-strand molecules. Embo J *7*, 4055-4062.

Herrera-Estrella, A., Vanmontagu, M., and Wang, K. (1990). A bacterial peptide acting as a plant nuclear targeting signal - the amino-terminal portion of *Agrobacterium* VirD2 protein directs a beta-galactosidase fusion protein into tobacco nuclei. Proceedings of the National Academy of Sciences of the United States of America *87*, 9534-9537.

Hiei, Y., Ohta, S., Komari, T., and Kumashiro, T. (1994). Efficient transformation of rice (*Oryza sativa* l) mediated by *Agrobacterium* and sequence-analysis of the boundaries of the T-DNA. Plant J *6*, 271-282.

Hofgen, R., and Willmitzer, L. (1988). Storage of competent cells for *Agrobacterium* transformation. Nucleic Acids Res *16*, 9877-9877.

Horsch, R.B., Klee, H.J., Stachel, S., Winans, S.C., Nester, E.W., Rogers, S.G., and Fraley, R.T. (1986). Analysis of *Agrobacterium tumefaciens* virulence mutants in leaf-disks. Proceedings of the National Academy of Sciences of the United States of America *83*, 2571-2575.

Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Res *35*, W585-W587.

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-localization signal - implications for nuclear uptake of DNA in plant-cells. Cell *68*, 109-118.

Hwang, H.H., Mysore, K.S., and Gelvin, S.B. (2006). Transgenic *Arabidopsis* plants expressing *Agrobacterium tumefaciens* VirD2 protein are less susceptible to *Agrobacterium* transformation. Mol Plant Pathol *7*, 473-484.

Ieamkhang, S., and Chatchawankanphanich, O. (2005). Augmentin (R) as an alternative antibiotic for growth suppression of *Agrobacterium* for tomato (*Lycopersicon esculentum*) transformation. Plant Cell Tiss Org *82*, 213-220.

Ilangovan, A., Connery, S., and Waksman, G. (2015). Structural biology of the Gram-negative bacterial conjugation systems. Trends in microbiology *23*, 301-310.

Ilyina, T.V., and Koonin, E.V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA-replication encoded by diverse replicons from eubacteria, eukaryotes and archaebacteria. Nucleic Acids Res *20*, 3279-3285.

Ishida, Y., Hiei, Y., and Komari, T. (2007). *Agrobacterium*-mediated transformation of maize. Nat Protoc *2*, 1614-1621.

Ishida, Y., Saito, H., Ohta, S., Hiei, Y., Komari, T., and Kumashiro, T. (1996). High efficiency transformation of maize (*Zea mays* L) mediated by *Agrobacterium tumefaciens*. Nat Biotechnol *14*, 745-750.

Jakubowski, S.J., Kerr, J.E., Garza, I., Krishnamoorthy, V., Bayliss, R., Waksman, G., and Christie, P.J. (2009). *Agrobacterium* VirB10 domain requirements for type IV secretion and T pilus biogenesis. Mol Microbiol *71*, 779-794.

Jakubowski, S.J., Krishnamoorthy, V., Cascales, E., and Christie, P.J. (2004). *Agrobacterium tumefaciens* VirB6 domains direct the ordered export of a DNA substrate through a type IV secretion system. Journal of Molecular Biology *341*, 961-977.

Jakubowski, S.J., Krishnamoorthy, V., and Christie, P.J. (2003). *Agrobacterium tumefaciens* VirB6 protein participates in formation of VirB7 and VirB9 complexes required for type IV secretion. Journal of Bacteriology *185*, 2867-2878.

James, C. (2014). Global Status of Commercialized Biotech/GM Crops (Ithaca, NY.: ISAAA).
Janssen, B.J., and Gardner, R.C. (1990). Localized transient expression of gus in leaf-disks following cocultivation with *Agrobacterium*. Plant Mol Biol *14*, 61-72.

Jarvis, P., and Lopez-Juez, E. (2013). Biogenesis and homeostasis of chloroplasts and other plastids. Nat Rev Mol Cell Biol *14*, 787-802.

Jasper, F., Koncz, C., Schell, J., and Steinbiss, H.H. (1994). *Agrobacterium* T-strand production in-vitro - sequence-specific cleavage and 5' protection of single-stranded-dna templates by purified VirD2 protein. Proceedings of the National Academy of Sciences of the United States of America *91*, 694-698.

Judd, P.K., Kumar, R.B., and Das, A. (2005a). The type IV secretion apparatus protein VirB6 of *Agrobacterium tumefaciens* localizes to a cell pole. Mol Microbiol *55*, 115-124.

Judd, P.K., Mahli, D., and Das, A. (2005b). Molecular characterization of the *Agrobacterium tumefaciens* DNA transfer protein VirB6. Microbiol-Sgm *151*, 3483-3492.

Kahlau, S., and Bock, R. (2008). Plastid transcriptomics and translatomics of tomato fruit development and chloroplast-to-chromoplast differentiation: Chromoplast gene expression largely serves the production of a single protein. Plant Cell *20*, 856-874.

Kerr, J.E., and Christie, P.J. (2010). Evidence for VirB4-mediated dislocation of membrane-integrated VirB2 pilin during biogenesis of the *Agrobacterium* VirB/VirD4 type IV secretion system. Journal of Bacteriology *192*, 4923-4934.

Kikkert, J.R., Vidal, J.R., and Reisch, B.I. (2005). Stable transformation of plant cells by particle bombardment/biolistics. In Methods in Molecular Biology, L. Peña, ed., pp. 61-78.

Kim, H., and Kim, J.S. (2014). A guide to genome engineering with programmable nucleases. Nat Rev Genet *15*, 321-334.

Klee, H.J., White, F.F., Iyer, V.N., Gordon, M.P., and Nester, E.W. (1983). Mutational analysis of the virulence region of an *Agrobacterium tumefaciens* Ti plasmid. Journal of Bacteriology *153*, 878-883.

Kofer, W., Eibl, C., Steinmuller, K., and Koop, H.U. (1998). PEG-mediated plastid transformation in higher plants. In Vitro Cellular & Developmental Biology-Plant *34*, 303-309.

Kohli, N., Westerveld, D.R., Ayache, A.C., Verma, A., Shil, P., Prasad, T., Zhu, P., Chan, S.L., Li, Q., and Daniell, H. (2014). Oral delivery of bioencapsulated proteins across blood-brain and blood-retinal barriers. Molecular therapy : the journal of the American Society of Gene Therapy *22*, 535-546.

Koonin, E.V. (1993). A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. Journal of Molecular Biology *229*, 1165-1174.

Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M., and Yanagawa, H. (2009a). Six classes of nuclear localization signals specific to different binding grooves of importin alpha. Journal of Biological Chemistry *284*, 478-485.

Kosugi, S., Hasebe, M., Tomita, M., and Yanagawa, H. (2009b). Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proceedings of the National Academy of Sciences of the United States of America *106*, 10171-10176.

Lacroix, B., and Citovsky, V. (2013). The roles of bacterial and host plant factors in *Agrobacterium*-mediated genetic transformation. Int J Dev Biol *57*, 467-481.

Lacroix, B., Li, J., Tzfira, T., and Citovsky, V. (2006). Will you let me use your nucleus? How *Agrobacterium* gets its T-DNA expressed in the host plant cell. Canadian Journal of Physiology and Pharmacology *84*, 333-345.

Lacroix, B., Loyter, A., and Citovsky, V. (2008). Association of the *Agrobacterium* T-DNA-protein complex with plant nucleosomes. Proceedings of the National Academy of Sciences of the United States of America *105*, 15429-15434.

Lacroix, B., Vaidya, M., Tzfira, T., and Citovsky, V. (2005). The VirE3 protein of *Agrobacterium* mimics a host cell function required for plant genetic transformation. Embo J *24*, 428-437.

Lai, E.M., and Kado, C.I. (1998). Processed VirB2 is the major subunit of the promiscuous pilus of *Agrobacterium tumefaciens*. Journal of Bacteriology *180*, 2711-2717.

Lam, A.P., and Dean, D.A. (2010). Progress and prospects: nuclear import of nonviral vectors. Gene therapy *17*, 439-447.

Lang, I., Scholz, M., and Peters, R. (1986). Molecular mobility and nucleocytoplasmic flux in hepatoma-cells. J Cell Biol *102*, 1183-1190.

Lange, A., Mills, R.E., Lange, C.J., Stewart, M., Devine, S.E., and Corbett, A.H. (2007). Classical nuclear localization signals: Definition, function, and interaction with importin alpha. Journal of Biological Chemistry *282*, 5101-5105.

Lee, D.W., Jung, C., and Hwang, I. (2013). Cytosolic events involved in chloroplast protein targeting. Biochim Biophys Acta-Mol Cell Res *1833*, 245-252.

Lee, D.W., Kim, J.K., Lee, S., Choi, S., Kim, S., and Hwang, I. (2008). *Arabidopsis* nuclear-encoded plastid transit peptides contain multiple sequence subgroups with distinctive chloroplast-targeting sequence motifs. Plant Cell *20*, 1603-1622.

Lee, L.Y., and Gelvin, S.B. (2008). T-DNA binary vectors and systems. Plant Physiol *146*, 325-332.

Lee, S.M., Kang, K., Chung, H., Yoo, S.H., Xu, X.M., Lee, S.B., Cheong, J.J., Daniell, H., and Kim, M. (2006). Plastid transformation in the monocotyledonous cereal crop, rice (Oryza sativa) and transmission of transgenes to their progeny. Mol Cells *21*, 401-410.

Li, G.Y., Froehlich, J.E., Elowsky, C., Msanne, J., Ostosh, A.C., Zhang, C., Awada, T., and Alfano, J.R. (2014). Distinct Pseudomonas type-III effectors use a cleavable transit peptide to target chloroplasts. Plant J *77*, 310-321.

Li, H.M., and Teng, Y.S. (2013). Transit peptide design and plastid import regulation. Trends Plant Sci *18*, 360-366.

Liu, W.S., Yuan, J.S., and Stewart, C.N. (2013). Advanced genetic tools for plant biotechnology. Nat Rev Genet *14*, 781-793.

Lohrke, S.M., Nechaev, S., Yang, H., Severinov, K., and Jin, S.J. (1999). Transcriptional activation of *Agrobacterium tumefaciens* virulence gene promoters in Escherichia coli requires the *A. tumefaciens* RpoA gene, encoding the alpha subunit of RNA polymerase. J Bacteriol *181*, 4533-4539.

Lu, J., den Dulk-Ras, A., Hooykaas, P.J.J., and Glover, J.N.M. (2009). *Agrobacterium tumefaciens* VirC2 enhances T-DNA transfer and virulence through its C-terminal ribbon-helix-helix DNA-binding fold. Proceedings of the National Academy of Sciences of the United States of America *106*, 9643-9648.

Lutz, K.A., Azhagiri, A.K., Tungsuchat-Huang, T., and Maliga, P. (2007). A guide to choosing vectors for transformation of the plastid genome of higher plants. Plant Physiol *145*, 1201-1210.

Ma, J., Goryaynov, A., Sarma, A., and Yang, W.D. (2012). Self-regulated viscous channel in the nuclear pore complex. Proceedings of the National Academy of Sciences of the United States of America *109*, 7326-7331.

Magori, S., and Citovsky, V. (2011a). *Agrobacterium* counteracts host-induced degradation of its effector F-Box protein. Sci Signal *4*, 7.

Magori, S., and Citovsky, V. (2011b). Epigenetic control of *Agrobacterium* T-DNA integration. Biochim Biophys Acta-Gene Regul Mech *1809*, 388-394.

Magori, S., and Citovsky, V. (2011c). Hijacking of the host SCF ubiquitin ligase machinery by plant pathogens. Front Plant Sci *2*, 8.

Maheshwari, P., and Kovalchuk, I. (2013). Combination of ammonium nitrate, cerium chloride and potassium chloride salts improves *Agrobacterium tumefaciens*-mediated transformation of *Nicotiana tabacum*. Plant Biotechnol Rep *7*, 147-154.

Maliga, P. (2003). Progress towards commercialization of plastid transformation technology. Trends Biotechnol *21*, 20-28.

Marion, J., Bach, L., Bellec, Y., Meyer, C., Gissot, L., and Faure, J.D. (2008). Systematic analysis of protein subcellular localization and interaction using high-throughput transient transformation of *Arabidopsis* seedlings. Plant J *56*, 169-179.

Mattanovich, D., Ruker, F., Machado, A.D., Laimer, M., Regner, F., Steinkellner, H., Himmler, G., and Katinger, H. (1989). Efficient transformation of *Agrobacterium* spp by electroporation. Nucleic Acids Res *17*, 6747-6747.

Melchers, L.S., Maroney, M.J., Dendulkras, A., Thompson, D.V., Vanvuuren, H.A.J., Schilperoort, R.A., and Hooykaas, P.J.J. (1990). Octopine and nopaline strains of *Agrobacterium tumefaciens* differ in virulence - molecular characterization of the VirF locus. Plant Mol Biol *14*, 249-259.

Merkle, T., Leclerc, D., Marshallsay, C., and Nagy, F. (1996). A plant in vitro system for the nuclear import of proteins. Plant J *10*, 1177-1186.

Meyers, B., Zaltsman, A., Lacroix, B., Kozlovsky, S.V., and Krichevsky, A. (2010a). Nuclear and plastid genetic engineering of plants: comparison of opportunities and challenges. Biotechnol Adv *28*, 747-756.

Miki, B., and McHugh, S. (2004). Selectable marker genes in transgenic plants: applications, alternatives and biosafety. Journal of biotechnology *107*, 193-232.

Miller, L.A., Ratnam, K., and Payne, D.J. (2001). Beta-lactamase-inhibitor combinations in the 21st century: current agents and new developments. Current opinion in pharmacology *1*, 451-458.

Mossey, P., Hudacek, A., and Das, A. (2010). *Agrobacterium tumefaciens* type IV secretion protein VirB3 is an inner membrane protein and requires VirB4, VirB7, and VirB8 for stabilization. Journal of Bacteriology *192*, 2830-2838.

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plant *15*, 473-497.

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.

Mysore, K.S., Nam, J., and Gelvin, S.B. (2000). An *Arabidopsis* histone H2A mutant is deficient in *Agrobacterium* T-DNA integration. Proceedings of the National Academy of Sciences of the United States of America *97*, 948-953.

Nair, G.R., Lai, X., Wise, A.A., Rhee, B.W., Jacobs, M., and Binns, A.N. (2011). The integrity of the periplasmic domain of the VirA sensor kinase is critical for optimal coordination of the virulence signal response in *Agrobacterium tumefaciens*. J Bacteriol *193*, 1436-1448.

Nam, J., Mysore, K.S., Zheng, C., Knue, M.K., Matthysse, A.G., and Gelvin, S.B. (1999). Identification of T-DNA tagged *Arabidopsis* mutants that are resistant to transformation by *Agrobacterium*. Molecular and General Genetics *261*, 429-438.

Nardozzi, J.D., Lott, K., and Cingolani, G. (2010). Phosphorylation meets nuclear import: a review. Cell Commun Signal *8*, 17.

Oey, M., Lohse, M., Kreikemeyer, B., and Bock, R. (2009). Exhaustion of the chloroplast protein synthesis capacity by massive expression of a highly stable protein antibiotic. Plant J *57*, 436-445.

Okuzaki, A., Kida, S., Watanabe, J., Hirasawa, I., and Tabei, Y. (2013). Efficient plastid transformation in tobacco using small gold particles (0.07-0.3 mu m). Plant Biotechnology *30*, 65-72.

Pacurar, D.I., Thordal-Christensen, H., Pacurar, M.L., Pamfil, D., Botez, C., and Bellini, C. (2011). *Agrobacterium tumefaciens*: From crown gall tumors to genetic transformation. Physiol Mol Plant Pathol *76*, 76-81.

Paila, Y.D., Richardson, L.G.L., and Schnell, D.J. (2015). New insights into the mechanism of chloroplast protein import and its integration with protein quality control, organelle biogenesis and development. Journal of Molecular Biology *427*, 1038-1060.

Pelczar, P., Kalck, V., Gomez, D., and Hohn, B. (2004). *Agrobacterium* proteins VirD2 and VirE2 mediate precise integration of synthetic T-DNA complexes in mammalian cells. EMBO Rep *5*, 632-637.

Porter, S.G., Yanofsky, M.F., and Nester, E.W. (1987). Molecular characterization of the *virD* operon from *Agrobacterium tumefaciens*. Nucleic Acids Res *15*, 7503-7517.

Porteus, M.H., and Carroll, D. (2005). Gene targeting using zinc finger nucleases. Nat Biotechnol *23*, 967-973.

Pyke, K., and Pyke, K. (2009). Different types of plastids and their structure (Cambridge: Cambridge Univ Press).

Rath, A., Glibowicka, M., Nadeau, V.G., Chen, G., and Deber, C.M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proceedings of the National Academy of Sciences of the United States of America *106*, 1760-1765.

Ream, W. (2008). Production of a mobile T-DNA by *Agrobacterium tumefaciens* (Springer, 233 Spring Street, New York, Ny 10013, United States).

Reavy, B., Bagirova, S., Chichkova, N.V., Fedoseeva, S.V., Kim, S.H., Vartapetian, A.B., and Taliansky, M.E. (2007). Caspase-resistant VirD2 protein provides enhanced gene delivery and expression in plants. Plant Cell Rep *26*, 1215-1219.

Relic, B., Andjelkovic, 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. Proceedings of the National Academy of Sciences of the United States of America *95*, 9105-9110.

Ripoll-Rozada, J., Zunzunegui, S., de la Cruz, F., Arechaga, I., and Cabezon, E. (2013). Functional interactions of VirB11 traffic ATPases with VirB4 and VirD4 molecular motors in type IV secretion systems. Journal of Bacteriology *195*, 4195-4201.

Romano, P.G.N., Horton, P., and Gray, J.E. (2004). The *Arabidopsis* cyclophilin gene family. Plant Physiol *134*, 1268-1282.

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.

San-Miguel, T., Perez-Bermudez, P., and Gavidia, I. (2013). Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature. SpringerPlus *2*, 89.

Sattarzadeh, A., Fuller, J., Moguel, S., Wostrikoff, K., Sato, S., Covshoff, S., Clemente, T., Hanson, M., and Stern, D.B. (2010). Transgenic maize lines with cell-type specific expression of fluorescent proteins in plastids. Plant Biotechnol J *8*, 112-125.

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. Journal of Biological Chemistry *270*, 1269-1276.

Schrammeijer, B., Risseeuw, E., Pansegrau, W., Regensburg-Tuink, T.J.G., Crosby, W.L., and Hooykaas, P.J.J. (2001). Interaction of the virulence protein VirF of *Agrobacterium tumefaciens* with plant homologs of the yeast Skp1 protein. Curr Biol *11*, 258-262.

Shatkay, H., Hoglund, A., Brady, S., Blum, T., Donnes, P., and Kohlbacher, O. (2007). SherLoc: high-accuracy prediction of protein subcellular localization by integrating text and protein sequence data. Bioinformatics *23*, 1410-1417.

Shi, L.X., and Theg, S.M. (2013). The chloroplast protein import system: From algae to trees. Biochim Biophys Acta-Mol Cell Res *1833*, 314-331.

Shi, Y.H., Mowery, R.A., Ashley, J., Hentz, M., Ramirez, A.J., Bilgicer, B., Slunt-Brown, H., Borchelt, D.R., and Shaw, B.F. (2012). Abnormal SDS-PAGE migration of cytosolic proteins can identify domains and mechanisms that control surfactant binding. Protein Sci *21*, 1197-1209.

Shil, P.K., Kwon, K.C., Zhu, P., Verma, A., Daniell, H., and Li, Q. (2014). Oral delivery of ACE2/Ang-(1-7) bioencapsulated in plant cells protects against experimental uveitis and autoimmune uveoretinitis. Molecular therapy : the journal of the American Society of Gene Therapy *22*, 2069-2082.

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. Proceedings of the National Academy of Sciences of the United States of America *89*, 11837-11841.

Sivanesan, D., and Baron, C. (2011). The dimer interface of *Agrobacterium tumefaciens* VirB8 is important for type IV secretion system function, stability, and association of VirB2 with the core complex. Journal of Bacteriology *193*, 2097-2106.

Sohn, K.H., Hughes, R.K., Piquerez, S.J., Jones, J.D.G., and Banfield, M.J. (2012). Distinct regions of the Pseudomonas syringae coiled-coil effector AvrRps4 are required for activation of immunity. Proceedings of the National Academy of Sciences of the United States of America *109*, 16371-16376.

Soltani, J., van Heusden, G.P.H., and Hooykaas, P.J.J. (2008). *Agrobacterium*-mediated transformation of non-plant organisms (Springer, 233 Spring Street, New York, Ny 10013, United States).

Sood, P., Bhattacharya, A., and Sood, A. (2011). Problems and possibilities of monocot transformation. Biol Plant *55*, 1-15.

Sparkes, I.A., Runions, J., Kearns, A., and Hawes, C. (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. Nat Protoc *1*, 2019-2025.

Stachel, S.E., and Nester, E.W. (1986). The genetic and transcriptional organization of the vir region of the A6-Ti plasmid of *Agrobacterium tumefaciens*. Embo J *5*, 1445-1454.

Stachel, S.E., Timmerman, B., and Zambryski, P. (1987). Activation of *Agrobacterium tumefaciens vir* gene expression generates multiple single-stranded T-strand molecules from the pTiA6 T-region - requirement for 5' *virD* gene products. Embo J *6*, 857-863.

Steck, T.R., Lin, T.S., and Kado, C.I. (1990). *VirD2* gene product from the nopaline plasmid pTiC58 has at least 2 activities required for virulence. Nucleic Acids Res *18*, 6953-6958.

Sundberg, C.D., and Ream, W. (1999). The *Agrobacterium tumefaciens* chaperone-like protein, VirE1, interacts with VirE2 at domains required for single-stranded DNA binding and cooperative interaction. Journal of Bacteriology *181*, 6850-6855.

Svab, Z., Hajdukiewicz, P., and Maliga, P. (1990). Stable transformation of plastids in higher plants. Proc Natl Acad Sci U S A *87*, 8526-8530.

Svab, Z., and Maliga, P. (1991). Mutation proximal to the tRNA binding region of the *Nicotiana* plastid 16S ribosomal RNA confers resistance to spectinomycin. Mol Gen Genet *228*, 316-319.

Tao, Y., Rao, P.K., Bhattacharjee, S., and Gelvin, S.B. (2004). Expression of plant protein phosphatase 2C interferes with nuclear import of the *Agrobacterium* T-complex protein VirD2. Proceedings of the National Academy of Sciences of the United States of America *101*, 5164-5169.

Tinland, B., Koukolikovanicola, Z., Hall, M.N., and Hohn, B. (1992). The T-DNA linked VirD2 protein contains 2 distinct functional nuclear localization signals. Proceedings of the National Academy of Sciences of the United States of America *89*, 7442-7446.

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. Journal of Bacteriology *171*, 6845-6849.

Tzfira, T., Vaidya, M., and Citovsky, V. (2004). Involvement of targeted proteolysis in plant genetic transformation by *Agrobacterium*. Nature *431*, 87-92.

Umeda, M., Shimotohno, A., and Yamaguchi, M. (2005). Control of cell division and transcription by cyclin-dependent kinase-activating kinases in plants. Plant Cell Physiol *46*, 1437-1442.

Valkov, V.T., Scotti, N., Kahlau, S., MacLean, D., Grillo, S., Gray, J.C., Bock, R., and Cardi, T. (2009). Genome-wide analysis of plastid gene expression in potato leaf chloroplasts and tuber amyloplasts: transcriptional and posttranscriptional control. Plant Physiol *150*, 2030-2044.

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*, 1356-1365.

Vandenbroeck, G., Timko, M.P., Kausch, A.P., Cashmore, A.R., Vanmontagu, M., and Herreraestrella, L. (1985). Targeting of a foreign protein to chloroplasts by fusion to the transit peptide from the small subunit of ribulose 1,5-bisphosphate carboxylase. Nature *313*, 358-363.

Venkateswarlu, K., and Nazar, R.N. (1991). Evidence for T-DNA mediated gene targeting to tobacco chloroplasts. Biotechnology (N Y) *9*, 1103-1105.

Vergunst, A.C., van Lier, M.C.M., den Dulk-Ras, A., Stuve, T.A.G., 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*. Proceedings of the National Academy of Sciences of the United States of America *102*, 832-837.

Verma, D., and Daniell, H. (2007). Chloroplast vector systems for biotechnology applications. Plant Physiology (Rockville) *145*, 1129-1143.

Verma, D., Samson, N.P., Koya, V., and Daniell, H. (2008). A protocol for expression of foreign genes in chloroplasts. Nat Protoc *3*, 739-758.

Vogel, A.M., and Das, A. (1992). Mutational analysis of *Agrobacterium tumefaciens* VirD2 - tyrosine-29 is essential for endonuclease activity. Journal of Bacteriology *174*, 303-308.

Vogel, A.M., Yoon, J., and Das, A. (1995). Mutational analysis of a conserved motif of *Agrobacterium tumefaciens* VirD2. Nucleic Acids Res *23*, 4087-4091.

Walde, S., and Kehlenbach, R.H. (2010). The part and the whole: functions of nucleoporins in nucleocytoplasmic transport. Trends Cell Biol *20*, 461-469.

Wang, P., and Heitman, J. (2005). The cyclophilins. Genome Biol *6*, 6.

Wang, R.W., and Brattain, M.G. (2007). The maximal size of protein to diffuse through the nuclear pore is larger than 60 kDa. FEBS Lett *581*, 3164-3170.

Wang, Y.F., Peng, W., Zhou, X., Huang, F., Shao, L.Y., and Luo, M.Z. (2014). The putative *Agrobacterium* transcriptional activator-like virulence protein VirD5 may target T-complex to prevent the degradation of coat proteins in the plant cell nucleus. New Phytol *203*, 1266-1281.

Ward, E.R., and Barnes, W.M. (1988). VirD2 protein of *Agrobacterium tumefaciens* very tightly linked to the 5' end of T-strand DNA. Science *242*, 927-930.

Wei, Q., Cao, J.J., Qian, W.J., Xu, M.J., Li, Z.R., and Ding, Y.L. (2015). Establishment of an efficient micropropagation and callus regeneration system from the axillary buds of *Bambusa ventricosa*. Plant Cell Tiss Org *122*, 1-8.

Weigel, D., and Glazebrook, J. (2006). Glufosinate ammonium selection of transformed *Arabidopsis*. CSH protocols *2006*.

Wiegert, J.S., Bengtson, C.P., and Bading, H. (2007). Diffusion and not active transport underlies and limits ERK1/2 synapse-to-nucleus signaling in hippocampal neurons. J Biol Chem *282*, 29621-29633.

Wolterink-van Loo, S., Escamilla Ayala, A.A., Hooykaas, P.J.J., and van Heusden, G.P.H. (2015). Interaction of the *Agrobacterium tumefaciens* virulence protein VirD2 with histones. Microbiology (Reading, England) *161*, 401-410.

Wroblewski, T., Tomczak, A., and Michelmore, R. (2005). Optimization of *Agrobacterium*-mediated transient assays of gene expression in lettuce, tomato and *Arabidopsis*. Plant Biotechnol J *3*, 259-273.

Xu, X.M., and Meier, I. (2008). The nuclear pore comes to the fore. Trends Plant Sci *13*, 20-27.

Yanofsky, M.F., Porter, S.G., Young, C., Albright, L.M., Gordon, M.P., and Nester, E.W. (1986). The virD operon of *Agrobacterium tumefaciens* encodes a site-specific endonuclease. Cell *47*, 471-477.

Ye, G.N., Colburn, S.M., Xu, C.W., Hajdukiewicz, P.T.J., and Staub, J.M. (2003). Persistence of unselected transgenic DNA during a plastid transformation and segregation approach to herbicide resistance. Plant Physiol *133*, 402-410.

Yeo, H.J., Yuan, Q., Beck, M.R., Baron, C., and Waksman, G. (2003). Structural and functional characterization of the VirB5 protein from the type IV secretion system encoded by the conjugative plasmid pKM101. Proceedings of the National Academy of Sciences of the United States of America *100*, 15947-15952.

Yi, H.C., Mysore, K.S., and Gelvin, S.B. (2002). Expression of the *Arabidopsis* histone H2A-1 gene correlates with susceptibility to *Agrobacterium* transformation. Plant J *32*, 285-298.

Young, C., and Nester, E.W. (1988). Association of the VirD2 protein with the 5' end of T strands in *Agrobacterium tumefaciens*. Journal of Bacteriology *170*, 3367-3374.

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*, 197-209.

Zhang, W.J., Dewey, R.E., Boss, W., Phillippy, B.Q., and Qu, R.D. (2013a). Enhanced *Agrobacterium*-mediated transformation efficiencies in monocot cells is associated with attenuated defense responses. Plant Mol Biol *81*, 273-286.

Zhang, X.R., Henriques, R., Lin, S.S., Niu, Q.W., and Chua, N.H. (2006). *Agrobacterium*-mediated transformation of *Arabidopsis thaliana* using the floral dip method. Nat Protoc *1*, 641-646.

Zhang, Y., Zhang, F., Li, X.H., Baller, J.A., Qi, Y.P., Starker, C.G., Bogdanove, A.J., and Voytas, D.F. (2013b). Transcription activator-like effector nucleases enable efficient plant genome engineering. Plant Physiol *161*, 20-27.

Zhou, F., Karcher, D., and Bock, R. (2007). Identification of a plastid intercistronic expression element (IEE) facilitating the expression of stable translatable monocistronic mRNAs from operons. Plant J *52*, 961-972.

Ziemienowicz, A. (2001). Odyssey of *Agrobacterium* T-DNA. Acta Biochimica Polonica *48*, 623-635.

Ziemienowicz, A., Gorlich, D., Lanka, E., Hohn, B., and Rossi, L. (1999). Import of DNA into mammalian nuclei by proteins originating from a plant pathogenic bacterium. Proceedings of the National Academy of Sciences of the United States of America *96*, 3729-3733.

Ziemienowicz, A., Merkle, T., Schoumacher, F., Hohn, B., and Rossi, L. (2001). Import of *Agrobacterium* T-DNA into plant nuclei: Two distinct functions of VirD2 and VirE2 proteins. Plant Cell *13*, 369-383.

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*, 6317-6322.

Ziemienowicz, A., Tzfira, T., and Hohn, B. (2008). Mechanisms of T-DNA integration (Springer, 233 Spring Street, New York, Ny 10013, United States).

Zupan, J., Hackworth, C.A., Aguilar, J., Ward, D., and Zambryskl, P. (2007). VirB1*promotes T-Pilus formation in the vir-type IV secretion system of *Agrobacterium tumefaciens*. Journal of Bacteriology *189*, 6551-6563.

Zupan, J.R., Citovsky, V., and Zambryski, P. (1996). *Agrobacterium* VirE2 protein mediates nuclear uptake of single-stranded DNA in plant cells. Proc Natl Acad Sci U S A *93*, 2392-2397.