# Agrobacterium infection : translocation of virulence proteins and role of VirF in host cells

Jurado Jácome, E.

# CHAPTER 4

## Analysis of protein interaction domains of the *Agrobacterium tumefaciens* effector protein VirF

Esmeralda Jurado-Jácome, Paul J.J. Hooykaas and Annette C.Vergunst

**ABSTRACT**

In eukaryotic cells, F-box proteins (FBPs) act as the specificity factors of SCF complexes that recognize proteins destined for protein degradation. The *Agrobacterium tumefaciens* VirF protein interacts through its N-terminal F-box domain with the *Arabidopsis* ASK1 protein, which is part of the SCF E3 ligase complex. Therefore, VirF may target host proteins with its C-terminal domain for ubiquitin-mediated proteolysis. Using GST pull down assays several putative VirF interactors were identified earlier (Chapter 3), which bound *in vitro* to VirFΔFbox. To characterize possible substrate binding domains in VirF, we used a series of deletion mutants of VirF in GST-binding assays with His-tagged versions of four of the identified putative interactors of VirF (PIF), namely 3-deoxy-D-arabino-heptosulonate 7-phosphate (DAHP) synthase 2 (DHS2, PIF1), Lon protease-like protein (PIF2), *Arabidopsis thaliana* calmodulin binding protein (pirin-like protein, PIF3), and vacuolar H+ ATPase subunit B3 (VHA-B3, PIF4). Our results showed that three of the putative partners interacted with the same region in VirF (amino acids 146 to 166) but also that the central region of VirF (amino acids 75-100) is important for the interaction with the PIF proteins. By subsequent *in silico* analysis no particular FBPs domains such WD40, LRR or Kelch repeats were identified on VirF and also in this *in silico* analysis of the PIFs no common motif that is involved in the recognition of these PIF proteins by VirF was identified.

## INTRODUCTION

As part of SCF E3 Ubiquitin Ligases, F-box proteins (FBPs) play a central role in targeting substrates for degradation via the 26S proteasome (Hershko and Ciechanover, 1998). To be structurally linked to the SCF core, FBPs interact with SKP1 via the N-terminal F-box region, while the C-terminal protein-protein interaction domain facilitates recruitment of the target protein (Bai *et al*., 1996), followed by its proteasome-mediated degradation. In mammals, FBPs are grouped in three classes according to their substrate-binding domains: FBWs, harboring WD-40 repeats; FBLs, containing leucine reach repeats (LRR) and FBXs, with no WD-40 or LRR but showing different protein-protein interaction domains such as CASH (carbohydrate-interacting), cyclin box, CH (calponin homology), TDL (traf-domain-like), Sec 7, zinc-finger and proline rich-domains. Also, some FBXs proteins are post-translationally modified by methylation, ribosylation, farnesylation or acetylation, which may affect their binding to substrates (reviewed by Cardozo and Pagano, 2004). Therefore, the interaction of the FBP C-terminal region with the substrate recognition site may differ among groups of FBPs.

The *A. tumefaciens* VirF protein interacts with the *Arabidopsis* SKP1 proteins ASK1 and ASK2 (Schrammeijer *et al*., 2001, Chapter 6). Therefore, VirF most likely plays a role in plant cell mediated protein degradation as part of a SCF complex. Using yeast two-hybrid analysis we identified several putative VirF interacting proteins (PIF), for five of which interaction was confirmed using *in vitro* binding assays (Chapter 3). Several of the identified proteins seem to be involved in plant defense response (Chapter 3), which suggests that VirF may be involved in dampening the host response to infection. Here we analyzed whether VirF interacts with these proteins through specific domains for each substrate, or whether a common region in VirF is responsible for recognition. To identify the VirF sequence involved in substrate interaction, progressive N-terminal deletions of VirF were tested in *in vitro* GST pull down assays for interaction with four of the identified Protein Interactors of VirF (PIF): 3-deoxy-D-arabino-heptosulonate 7-phosphate (DAHP) synthase 2 (DHS2, PIF1), Lon protease-like protein (PIF2), *Arabidopsis thaliana* calmodulin binding protein (pirin-like protein, PIF3) and putative vacuolar H+ATPase subunit B3 (VHA-B3, PIF4). To obtain more information about the PIF interacting regions and determine possible similarities among them we made use of sequence prediction analysis.

Our results suggest that one region in VirF, between amino acids 146 to 166, is essential for interaction with the Lon protease-like protein, pirin-like protein and VHA-B3. The region between amino acids 100-125 is necessary for interaction with DHS2 synthase. However, this result is not conclusive, and further mutational analysis is required. Though we predict that a similar region in VirF is used for recognition of some putative target proteins, our results show that similar to other FBPs, the VirF substrate-binding region might also differ among targets and probably require additional modifications for proper protein recognition. In addition, were unable to detect a common binding motif in the PIFs for recognition by the FBP VirF.

## EXPERIMENTAL PROCEDURES

### Construction of GST-VirF deletions

Construction of a *virFΔ42N* translational fusion to Glutathione S-transferase (*GST*) is described in Chapter 3 (pSDM 3031). Additional N-terminal deletions (*virFΔ75, Δ100, Δ125, Δ145, Δ166*) were translationally fused to GST (Figure 1A) in the pGEX-KG vector (Guan and Dixon, 1991) as well as the DNA sequence encoding the region between amino acids (aa) 42 and 166 (*virFΔ42-166*). Plasmid pSDM3192 (Schrammeijer *et al*., 2001) was used as a template to amplify portions of *virF* using several different forward primers (see Table 1) in combination with VirF18 as reverse primer. The region of *virF* (between aa 42 and 166) was amplified with primers VirF126 and VirF15. The PCR products were cloned in pGEX-KG (Guan and Dixon, 1991) as *Sal*I-*Hind*III (resulting in pSDM3519) and *EcoR*I-*Sac*I fragments for the other fusions, respectively (resulting in pSDM3543, pSDM3544, pSDM3545, pSDM3546 [Table 1]). For the *virFΔ166* N-terminal deletion, a *Sal*I-*Hind*III *virF* fragment from pSDM3214 (Chapter 3) was subcloned in pME6010 (Heeb *et al*., 2000; Bloemberg *et al*., 2000) with *Xho*I-*Hind*III, resulting in pSDM3212 (den Dulk-Ras, A., *unpublished*). An *Xho*I-*Bgl*II segment of pSDM3212, containing *virFΔ166,* was cloned as *Xho*I-*BamH*I fragment into pIC19H (Alting-Mees and Short, 1989). Finally, a 111 bp *Xho*I-*Hind*III fragment was then inserted into pGEX-KG (Guan and Dixon, 1991) giving origin to pSDM3033 carrying the *GST::virFΔ166* fusion.

Table 1. Primers used for amplification of VirF deletions

| Primers | Sequence | Deletion | pSDM |
|---|---|---|---|
| Forward: | | | |
| VirF126 | 5'- AGCGTCGACCTGTCGAGTCGGCTGAG -3' | *virFΔ42-166* | 3519 |
| VirF19 | 5'- GGAATTCTAAGATCTCTCAAAATCCTCTC GCGAAAC -3' | *virFΔ75* | 3543 |
| VirF20 | 5'- GGAATTCTAGGTACCCAATTAAGTGAAGC GCAACTG -3' | *virFΔ100* | 3544 |
| VirF21 | 5'- GGAATTCTAACTAGTCATCAAGACGGTCA ACCAGAAG -3' | *virFΔ125* | 3545 |
| VirF22 | 5'- GGAATTCTAGCATGCAATATCCCTAACTT GGTCTTCAACG -3' | *virFΔ145* | 3546 |
| Reverse: | | | |
| VirF15 | 5'- GCTCTAGAAGCTTCAGAGAGCGGAGATT TCATTGAA -3' | | 3519 |
| VirF18 | 5'- CGAGCTCTCTCATAGACCGCGCGTTGAT CG -3' | | 3543<br>3544<br>3545<br>3546 |

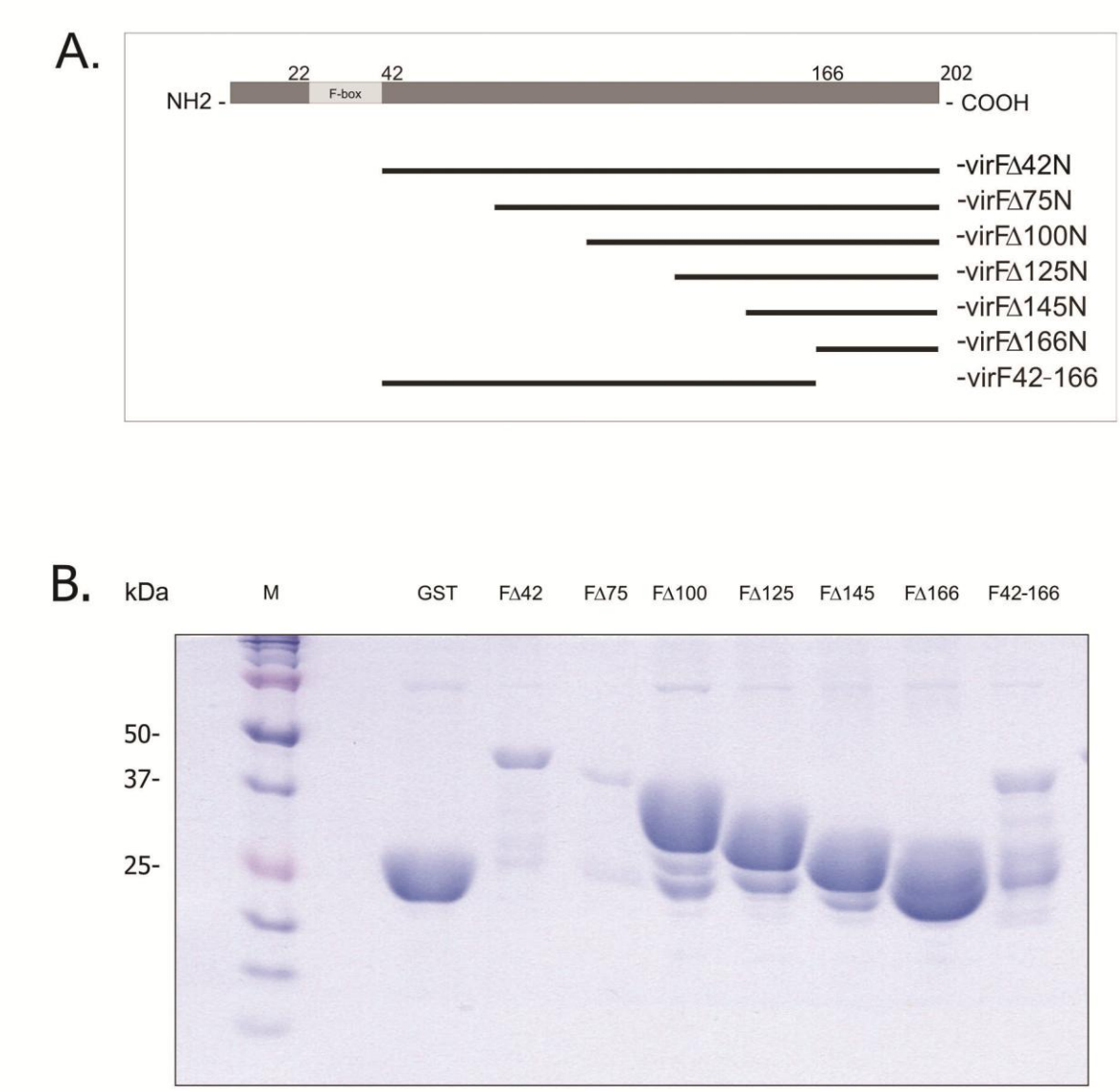


**Figure 1. VirF deletions. A)** Graphical representation of the 42, 75, 100, 125, 145, 166 N-terminal VirF deletions and the intermediate 42-166 aminoacids region. **B)** SDS-PAGE gel representation (Coomassie Brilliant Blue staining) of the progressive N-terminal VirF deletions and VirF42-166 region as glutathione S-transferase (GST) fusions in the heterologous expression vector pGEX-KG (pSDM3031, pSDM3543 to pSDM3546 and pSDM3519 respectively). *E. coli* BL21 cells were IPTG induced and translational fusions to GST detected by SDS-PAGE, resulting in proteins of 43, 40, 37, 34, 32, 30, and 39 kDa. GST: control pGEX-KG vector, M: BioRad dual color protein marker.

**Protein binding assays**

To study the binding of the different GST::VirF deletions to the His-tagged PIFs [DAHP synthase 2 (DHS2, At4g33510, PIF1), Lon protease-like protein (At1g75460, PIF2), *A. thaliana* calmodulin binding protein (pirin-like protein, At2g43120, PIF3) and ($VH^+$-ATPase B3 subunit (VHA-B3, At1g20260, PIF4)] *in vitro*, protein fusions were expressed in *E. coli* strain BL21-DE3 (Novagen). *In vitro* GST pull-down assays were performed as described in Chapter 3. Glutathione beads with bound GST::VirF fusion proteins were incubated with the different His-tagged PIFs. The beads were resuspended in sodium dodecyl sulfate

polyacrylamide gel electrophoresis (SDS-PAGE) loading buffer and proteins were separated in a 12.5% polyacrylamide gel, including protein extracts of the corresponding bacterial culture as His::PIF fusion positive control, and extract of GST alone, *virFΔ42N* vs His alone cultures as negative controls. Proteins were blotted on a polyvinylidine fluoride (PVDF)-Immobilon P-Millipore membrane using a semidry blotting apparatus (2117 Multiphor II electrophoresis unit from LKB [Upsala]), and blotting buffer (39 mM glycine, 48 mM Tris-HCL, 0.0375% SDS, 20% Methanol). Detection of His-tagged interactors was performed by Western blot analysis using penta-His antibodies in a 1:2000 dilution (Qiagen, Benelux B.V.) and visualized by chemiluminiscence using HRP substrate (Lumiglu reagent-Cell Signaling Technology, Westburg, The Netherlands).

**Protein sequence analysis**

Sequences of PIF proteins were previously submitted to DNA and protein BLAST analysis using the *Arabidopsis thaliana* MATDB (http://mips.gsf.de/proj/thal/db/Index.html) and GenBank (http://ncbi.nlm.nih.gov/BLAST/) databases (Chapter 3). The search for protein domains and motifs was done using the Eukaryotic Linear Motif database (http://elm.eu.org/); predictions of subcellular localization and presence of nuclear localization signal (NLS) were made by LOCtree analysis (http://cubic.bioc.columbia.edu/predictNLS/).

**Multiple alignment**

Fasta files of the four protein sequences obtained from NCBI (National Center for Biotechnology Information, USA) data base [PIF1 (NP_195077.1), PIF2 (AAK76613.1), PIF3 (NP_850385.1) and PIF4 (BAD44678)] were submitted to ClustalW (http://www.ebi.ac.uk/clustalw/index.html, Chenna *et al*., 2003; Larkin *et al.,* 2007) and T-Coffee analysis (http://www.ebi.ac.uk/t-coffee/; Notredame, Higgins and Heringa, 2000). Phylogram representation was obtained after multiple alignments of all four sequences in the ClustalW site. Fasta files were then submitted to the MEME motif searching site (http://meme.sdsc.edu/meme/meme.html) and consensus sequences were obtained by the MEME linked MAST analysis as explained in Pinzon A, 2006. Particular Eukaryotic Linear Motifs were searched at the ELM database (http://elm.eu.org/).

**RESULTS AND DISCUSSION**

Earlier, we identified several host proteins that interacted with VirF in yeast two hybrid, as well as *in vitro* in the absence of its F-box motif (VirFΔ42N) (Chapter 3). In the present work, we searched for common C-terminal binding motifs in VirF essential for interaction with each of the four PIF binding partners. To this end we first performed a deletion analysis, and secondly in an attempt to predict functional domains, we performed an *in silico* analysis of VirF using ELM (eukaryotic linear motif search; http://elm.eu.org/). In order to obtain more information about the interacting regions of the PIF proteins previously obtained (Chapter 3) and to determine possible similarities among them, we describe the protein fragments found that actually showed binding to VirFΔ42N (regions encoded by the

pACT-Gal4 AD cDNA clones). Besides an ELM search, we performed ClustalW (http://www.ebi.ac.uk/clustalw/index.html, Chenna *et al*., 2003; Larkin *et al.,* 2007), T-Coffee (http://www.ebi.ac.uk/t-coffee/; Notredame, Higgins , and Heringa, 2000), and MEME motif/ MAST *in silico* predictions (http://meme.sdsc.edu/meme/meme.html) on each of these PIF sequences.

**Expression of VirF deletions and interaction with PIFs**

DNA sequences encoding progressive 5' deletions of *virF* were fused translationally to glutathione S-transferase (GST) in the heterologous expression vector pGEX-KG, resulting in expression of VirF-GST fusion proteins with deletions of VirF amino acids (aa) 1-42, 1-75, 1-100, 1-125, 1-145 and 1-166; also the region of VirF between aa 42 to 166 was fused to GST (pSDM3031, pSDM3543 to pSDM3546 and pSDM3519 respectively, Table 1, Figure 1A). After IPTG induction in *E. coli* BL21 cells, the VirF-GST fusion proteins (with molecular weights of 43, 40, 37, 34, 32, 30, and 39 kDa respectively) were all detected by SDS-PAGE (Figure 1B). To search for C-terminal regions in VirF essential for protein recognition, these VirF deletions were tested by GST-pull down analysis in their binding capacity to some of the previously identified PIF proteins (Chapter 3). Figure 2 shows that interaction with PIF2 and PIF3 was maintained after deletion of the first 75 N-terminal residues (VirFΔ75N) and after deletion of the C-terminal 36 aa (VirF42-166), but disappeared in deletions VirFΔ100N, VirFΔ125N and VirFΔ166N. We conclude that an important target recognition domain is lost upon deletion of the N-terminal 100 aa, and thus must at least partially be located between aa 75 and aa 100. The stronger binding of the VirFΔ42N fusion compared to VirFΔ75N indicates that aa 42 to 75 are also partially involved in binding. This could be either by a direct interaction of this sequence region with a target protein or due to the requirement of some or all amino acids there contained for a conformational change needed for efficient binding. It was remarkable that binding of PIF2 and PIF3 (weaker) was found with VirFΔ145N but not with VirFΔ100N or VirFΔ125N. This suggests that deletion of the N-terminal 145 aa (but not of the N-terminal 125 aa), releases a second target interaction domain in VirF that must be at least partially located between aa 145 and aa 166, as the loss of the N-terminal 166 aa does not allow binding any more. Whether this second binding interface of VirF is only an artifact due to the deletion or can become available for binding in the complete VirF protein *in vivo*, for instance due to conformational changes as a consequence of protein modification, will require further study. Alternatively, the region between aa 145 and 166 may be the sole important interaction domain, that is not recognized in the somewhat longer versions VirFΔ100N and VirFΔ125N, due to conformational changes of the half truncated proteins that may mask the recognition site. In addition, the C-terminal 36 aa seem to have some role in the efficiency of binding of PIF3.

A quite different outcome was observed for the other two VirF interactors. DHS2 (PIF1) interacted with VirFΔ42N, VirFΔ100N and ViF42-166, but showed no binding to VirFΔ75N and ViFΔ166N, and only very weak binding to VirFΔ125N and VirFΔ145N (Figure 2). This suggests that a main interaction domain is located at least partially between aa 42 and aa 75. A second interaction domain seems to become exposed after deletion of the N-

terminal 100 aa. Alternatively, the interaction domain may lay between aa 100 and 166, with the most important site of interaction between aa 100-125 that is not recognized in the VirFΔ75N fusion due to masking of the site by conformational changes. Protein VHA-B3 (PIF4) showed some binding to the negative control (GST). The binding of the deletions was therefore interpreted according to the basal signal shown by the GST negative control. Considering this, it was clear that strong binding to VirF was lost upon deletion of the area between aa 42 and aa 75 (like for the other PIFs). Upon deletion of the N-terminal 145 aa strong binding to VirF becomes apparent again, suggesting that there is a binding domain in

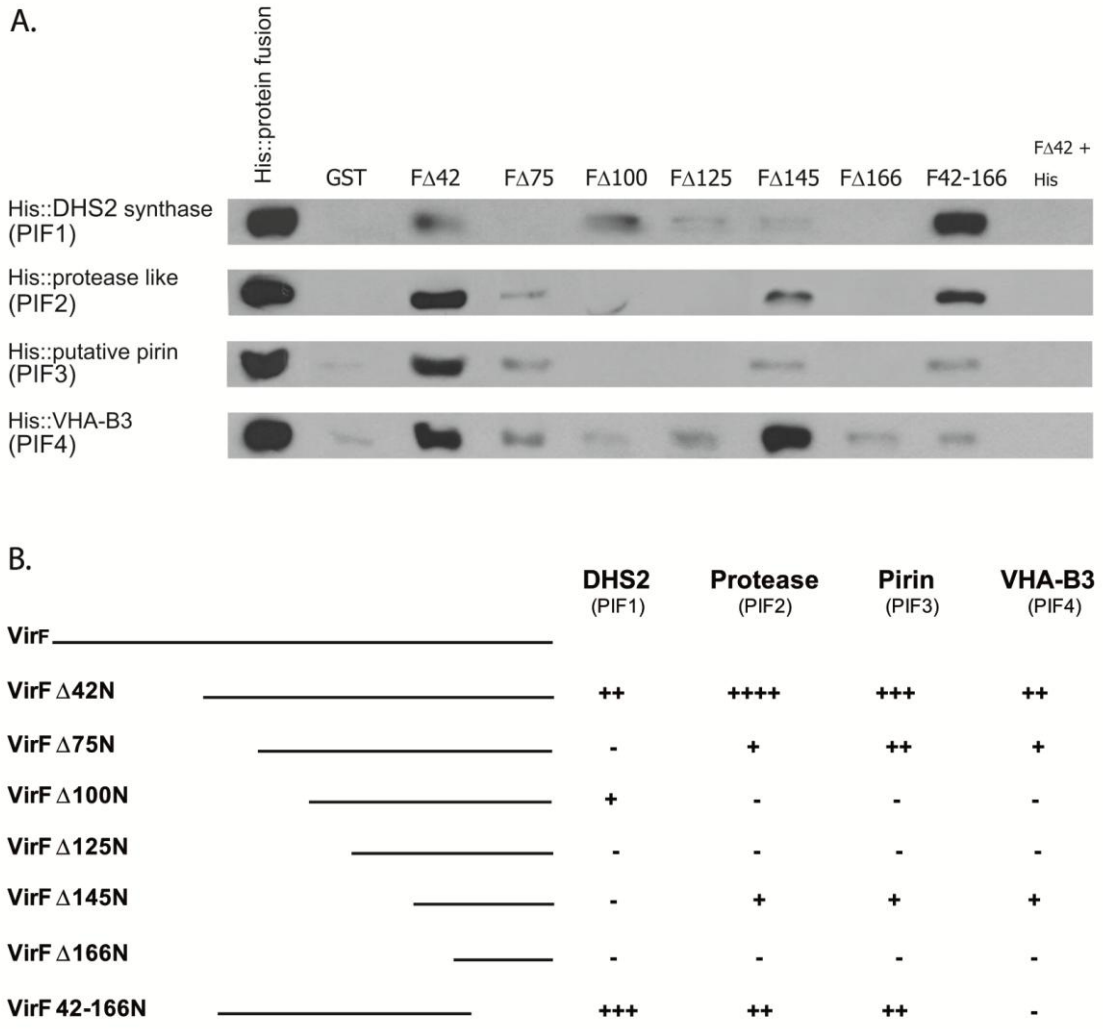


**Figure 2. A)** Immunodetection of GST pull down analysis between VirF and its interacting partners DAPH synthase 2/DHS2 (PIF1), Lon protease-like protein (PIF2), pirin-like protein (PIF3) and putative VHA-B3 synthase subunit (PIF4). **His::protein fusion**: His-fusion expression positive control consisting of extract of IPTG induced *E. coli* cultures of each His::PIF fusion before binding assay (Input); **GST**: Binding reaction GST control consisting in glutathione-matrix after binding reaction between cells carrying pGEX-KG lacking *virF* fusions and cells expressing the different His::PIF's fusions; **FΔ42 till FΔ166**: Glutathione-matrix after binding reaction of GST::virF 42, 75, 100, 125, 145 and 166 N-terminal deletions with His::PIF's fusions; **F42-166**: Glutathione-matrix after binding reaction of GST::virF42-166 fusion with His::PIF's fusions; **FΔ42+His**: Binding reaction His- expression negative control consisting in Glutathione-matrix after binding reaction of GST::virFΔ42 IPTG induced cells carrying empty pET16H vector. **B)** Graphical representation of the GST pull down analysis between different VirF N-terminal deletions (FΔ42 till FΔ166) and its protein interacting partners (PIFs).

the C terminal part of VirF that is shielded or becomes inactive after deletion of the N-terminus beyond 42 aa. Deletion of the C-terminal 37 aa removes this binding site as can be seen in the VirF42-166 protein.

Summarizing, the data suggest that an important target recognition domain may partially be located between aa 75 and aa 100. The stronger binding of the VirFΔ42N fusion compared to VirFΔ75N and VirFΔ100N indicates that aa 42 to 75 are also partially involved in binding. When the deleted regions were located in the middle part of the VirF sequence, more precisely inside the region located between aa 75-125, there was only weak interaction left with PIF1 (DHS2). On the other hand, the simultaneous elimination in VirF of the first N-terminal 42 aa and last C-terminal 37 aa (VirF42-166) led to a reduction in the strength of the interaction of VirF with all four PIFs and, disappearance of the VirF residues 100-125 nullified the VirF recognition of all the four PIFs. The weaker binding of PIF2 and PIF3 found with VirFΔ145N but not with VirFΔ100N or VirFΔ125N suggests that deletion of the N-terminal 145 aa (but not of the N-terminal 125 aa), releases a second target interaction domain in VirF that must be at least partially located between aa 145 and aa 166, as the loss of the N-terminal 166 aa does not allow binding any more. In conclusion, the described results indicate that the protein domains centrally located may display higher importance for the mediation of protein-protein interaction than those present at either the N-terminal or C-terminal borders of the VirF sequence.

Considering that the binding patterns found for PIF2, PIF3 and PIF4 indicate that some domains located in the central region of VirF may have some influence in the interaction with the PIF proteins, we performed ELM analysis to identify domains or motifs that could be located inside this region. The ELM analysis predicted motifs such as RXXL (APCC-binding destruction motifs, aa 116-121), phosphothreonine motifs that bind a subset of forkhead associated domains (FHA, aa 93-99, 114-120, 123-129), glycogen synthase kinase (GSK3, aa 105-112, 109-116) and, phosphoinositide-3-OH-kinase related kinases (PIKK, aa 105-111) phosphorylation sites. Putative RXXL and phosphothreonine motifs are also located in the C-terminal region, 186-191 and 179-185, respectively. However, none of these predicted motifs is located in an area with significance for the interaction with the PIF proteins.

Post translational modification can be important for substrate binding (Cardozo and Pagano, 2004; Xiaolu *et al*., 2008; Glenn *et al*., 2008) of FBPs. Although the analysis predicted sites for post translational modifications, such as phosphorylation and N-glycosylation, for VirF, our *in vitro* assays showed such post-translational modifications of VirF were not necessary to detect VirF interacting proteins (Schrammeijer *et al.,* 2001; Chapter 3). *In silico* analysis of VirF showed the presence of a LIG_APCC_Dbox (RXXL , destruction box) region towards the C-terminus of VirF, which may suggest that VirF could be targeted for degradation via the ubiquitin-proteasome system inside the host cell, as described for several eukaryotic proteins carrying this domain (Peters, 2006). Prediction of interaction with Traf domain like (TDL) deubiquitinating enzymes LIG_USP7_1 (9-13) and LIG_USP7_2 (131-137), suggests that VirF levels may be regulated by the host degradation control system. However, further experimental evidence is needed to verify this.

***In silico* analysis of PIFs**

In Chapter 3 of this thesis we described the identification by yeast two-hybrid analysis of the four PIFs here used for interaction studies with VirF. Not for all PIFs the complete cDNA clone was identified (Table 2). This indicated for instance for PIF1, that only the cloned region was important for interaction with VirF. We used this knowledge to simplify the *in silico* analysis for possible domains that may be involved in interaction with VirF. ProDkin binding sites, which mediate protein-protein interaction of short proline rich regions within proteins and WW domains, are predicted in PIF2, 3 and 4. WW domain-containing proteins are involved in many cellular processes such as ubiquitin mediated protein degradation and regulation of mitotic domains (Chen and Matesic, 2007). However, absence in VirF of PPXY and PPLP motifs, characteristic of WW domain interacting proteins, suggests that association of WW domains with VirF interaction is unlikely. Figure 3 depicts the results of a domain search (ELM) for the four PIFs. Summarizing, for most predicted motifs it is difficult to determine whether the prediction is significant, and additional experimentation is needed to confirm a possible role.

Table 2. Description of VirF interacting amino acid regions of PIF proteins

| | **PIF1** (DHS2) | **PIF2** (Lon protease-like) | **PIF3** (Pirin-like) | **PIF4** (VHB3) |
|---|---|---|---|---|
| Total protein Size (aa) | 507 | 278 | 321 | 487 |
| Protein size of the cDNA encoded binding region (aa) | 243 | 266 | 157 | 446 |
| Amino acid position of cDNA encoded region that bound to VirF | 176-419 (Middle region) | 1-266 (Almost all protein except the last 12 C-ter aa) | 150-307 (Middle region) | 41-487 (Almost all protein except the first N- term 40 aa) |

Comparison of the predicted ELM sites for the four PIF proteins indicated that six predicted domains were present in the four PIFs. However, predicted phosphorylation sites may not have a relevant meaning considering that consensus sites are so small that they may be found in virtually all proteins; in addition to the fact that plant kinases have probably different specificity than the mammalian ones. Most interesting, the prediction of an APCC-D-box (RXXL, destruction box) in all four PIFs suggests the ubiquitin-proteasome mediated degradation of these eukaryotic proteins (Peters, 2006), in line with a role for VirF in degradation of these proteins.

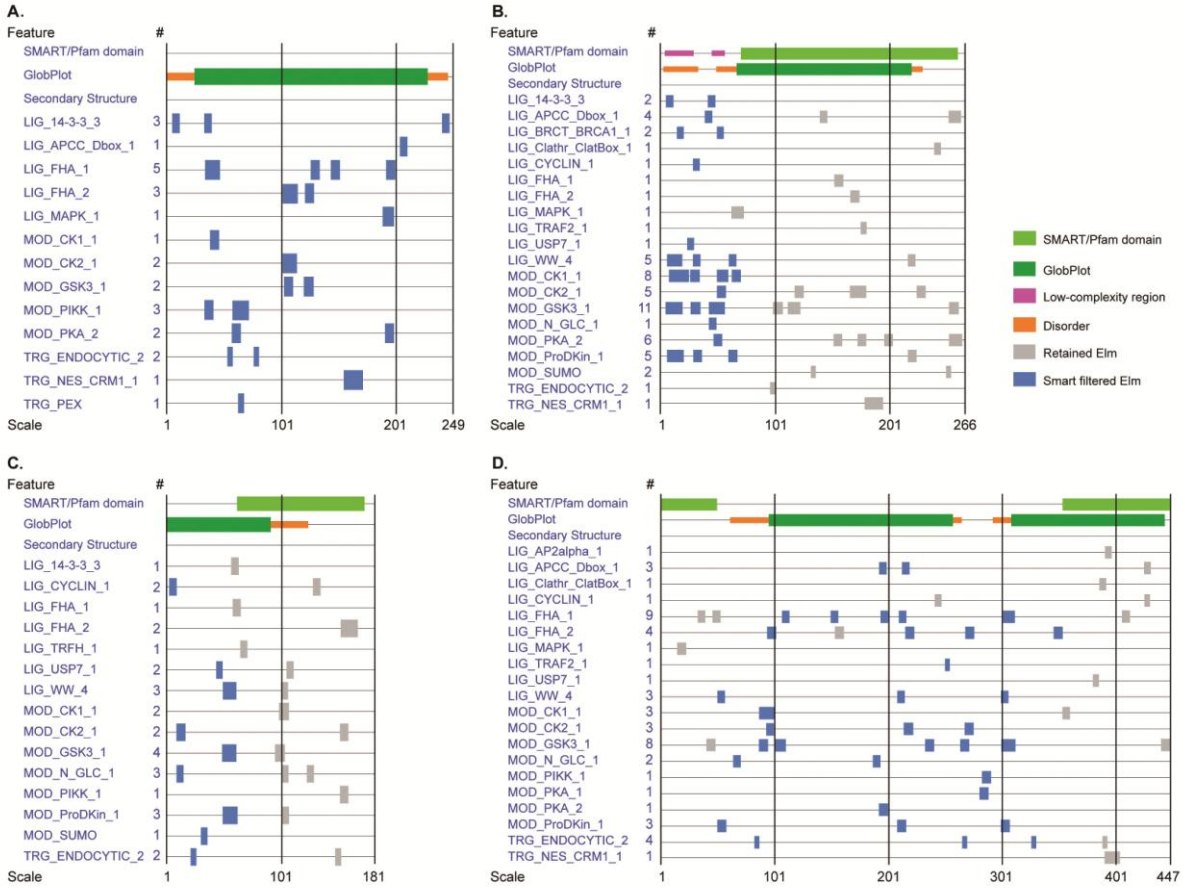


**Figure 3. Eukaryotic Linear Motifs (ELM) profile for predicted domains of protein interactors of VirF (PIF's) interactin region** (http://elm.eu.org). **A)** DAPH synthase 2 (DHS2, PIF1). **B)** Protease like protein (PIF2). **C)** Pirin like protein (PIF3). **D)** Putative VHA-B3 synthase subunit (PIF4). For PIF1 Pfam domains are not predicted; predicted SMART/Pfam Domains for PIF2 LON DOMAIN, for PIF 3 Pirin-C terminal domain and for PIF4 predicted ATP synthase alpha/beta family, beta-barrel N and C terminal domains. GlobPlot:: globular conformation associated to secondary structure surrounded by small disordered/unorganized linear regions; LIG_14-3-3-1/2/3: binding regions for 14-3-3- proteins; LIG_AP2alpha: motifs of regulatory/accessory proteins involved in the control of clathrin coated vesicle formation that bind to the alpha-subunit of the adaptor protein AP-2; LIG_APCC_Dbox_1: APCC-binding destruction motifs; LIG_BRCT_BRCA1_1: breast cancer susceptibility protein phosphopeptide ligands; LIG_Clathr_ClatBox: clathin box motifs recognition site; LIG_CYCLIN: substrate recognition site that interacts with cyclin and thereby increases phosphorylation by cyclin/cdk complexes; LIG_FHA_1/2: forkhead associated domains; LIG_MAPK_1: mitogen-activated protein kinase docking motif; LIG_PP1: conserved PP1c (protein phosphatase 1 catalytic subunit)- binding motif; LIG_TRAF2: TNF (tumor necrosis factor) receptor-associated factor 2 (TRAF2)-binding domain; LIG_TRFH_1: TRFH domain binding motifs found in proteins recruited to the shelterin complex by TRF1 and TRF2; LIG_USP7_1: ubiquitin-specific portease binding motif; LIG_WW_4: class IV WW phosphorylation-dependent interaction motif; MOD_CK1_1/2: casein kinase 1 and 2 phosphorylation site; MOD_GSK3_1: site recognised by GSK3 for Ser/Thr Phosphorylation; MOD_N-GLC_1: N-glycosylation site; MOD_PIKK_1: phosphoinositide-3-OH-kinase related kinases phosphorylation site; MOD_PKA_1/2: cAMP-dependent protein kinases-A phosphorylation sites; MOD_ProDKin: proline-directed kinases; MOD_SUMO: small ubiquitin-related modifiers (SUMO) target site; TRG_ENDOCYTIC_2: tyrosine-based sorting signal responsible for the interaction with mu subunit of AP (Adaptor Protein) complex; TRG_NES_CRM1_1: nuclear export signal that binds to the CRM1 exportin protein; TRG_PEX: peroxisome targeting signal-1 receptor.

**Prediction of cellular localization**

In order for VirF to play a role during infection in degradation of host proteins, these proteins must have the same subcellular localization as VirF. Apparently, the bacterial ontogeny of VirF predicts a mitochondrial or chloroplastic localization (http://cubic.bioc.columbia.edu/services/loctree/out/3417767.html); however experimental evidence for this is lacking. Tzfira and colleagues (2004) showed that the VirF protein localizes in the plant cell nucleus, although no putative NLS has been predicted (http://cubic.bioc.columbia.edu/cgi/var/nair/resonline.pl). In chapter 6 of this thesis we provide evidence that VirF localizes both in the cytoplasm and in the cell nucleus in stably transformed plant cells. Therefore, the exact subcellular localization of VirF is currently not clear.

LOCtree prediction analysis indicated that none of the PIF proteins contains an NLS sequence (http://cubic.bioc.columbia.edu/predictNLS/). However, experimental evidence has shown a nuclear localization for PIF3 and interaction with Bcl-3 and Nuclear Factor 1 (NF1) proteins. It is presumed that proteins of the CUPIN family are involved in several processes involved in transcription and replication regulation (Dechend *et al*., 1999), positioning them in the nucleus. For proteins like PIF2 and PIF4 a mitochondrial localization is predicted (http://cubic.bioc.columbia.edu/services/loctree/out/4524131.html), or as for PIF4, which is part of a plant structural V-ATPase, a vacuolar organization (MADB, 2002). Proteasome degradation does not take place in the chloroplast and mitochondrion. However, interaction with VirF may result in relocalization of the protein, resulting in proteolysis. Though not seen as evidence, the predicted subcellular localization patterns are not in disagreement with a role for VirF in the proteolytic degradation of these proteins.

**Multiple alignment analysis of PIFs**

To determine whether the predicted targets of VirF contained common sequence regions or particular motifs associated with recognition by FBPs that were not detected using ELM analysis, we performed alignment of the four full length PIF sequences using ClustalW (http://www.ebi.ac.uk/clustalw/index.html; Chenna *et al.,* 2003; Larkin *et al.,* 2007) and T-Cofee analysis (http://www.ebi.ac.uk/t-coffee/; Notredame, Higgins and Heringa, 2000). No characteristic common pattern could be determined. Multiple alignment of all four sequences by ClustalW showed 16% of similarity between PIF1 (DHS2) and PIF3 (pirin-like protein), while PIF2 (Lon protease-like) and PIF4 (VHA-B3 subunit) did not present similarities. T-Cofee analysis provides higher alignment precision. Using this programme we obtained (21%) similarity between PIF1 and PIF3. To find shorter sequence regions among these proteins that share amino acid identity or similarity, we submitted the PIFs fasta formatted sequences to the MEME motif searching site (http://meme.sdsc.edu/meme/meme.html). We identified three motifs in PIF1 that were also separately present in some of the additional PIF proteins (Table 3). Using the applications offered by MAST analysis, a more precise consensus sequence among the PIFs could be determined for each pattern found by MEME. In order to predict if these motifs were involved in any relevant protein interaction, the obtained consensus sequences were analyzed at the ELM site (http://elm.eu.org/) using *A.*

*thaliana* as source organism. However, in addition to predictions obtained via ELM analysis, no particular interacting regions could be identified that could be directly correlated to VirF.

Table 3. Motif analysis (MEME) in PIF´s sequences

| Motif Pattern | Motif in PIFs | Eukaryotic Linear Motif (ELM)* | ELM Pattern |
|---|---|---|---|
| **1.** | | | |
| [AY]-Y-[DE]-C-[GS]-[AK]-H-[MV]-L-[VW] | PIF1: YYDCSAHMLW | | |
| | PIF4: AYECGKHVLV | | |
| **MAST Consensus:** | **AYDCGAHMLV** | **Not identified** | **Not identified** |
| **2.** | | | |
| [HY]-H-[NT]-[AH]-[CK]-[DS]-P | PIF1: YHTHCDP | | |
| | PIF2: HHNAKSP | | |
| **MAST Consensus:** | **HHNACDP** | **Not identified** | **Not identified** |
| **3.** | | | |
| A-[HQ]-[IP]-[HI]-[MQ]-[NT]-[TV]-[PT]-E-[FS]-W-[NT]-[AS]-[FH]-[AE]-[CY] | PIF1: AHPIMTTTEFWTSHEC | | |
| | PIF3: AQIHQNVPESWNAFAY | | |
| **MAST Consensus:** | **AHIHMNTPEFWNAFAC** | **MOD_PRODKin_1**<br>**LIG_WW_4** | **...([ST])P..**<br>**...[ST]P]..** |

**Predicted functional site description:**

| | |
|---|---|
| LIG_WW_4: | Class IV WW domains interaction motif; serine phosphorylation-dependent associated to ubiquitin mediated protein degradation and mitotic regulation (http://elm.eu.org/elmPages/ LIG_WW_4.html) |
| MOD_PRODKin_1: | Proline-Directed Kinase (e.g. MAPK) phosphorylation site *in* higher eukaryotes (http://elm.eu.org/elmPages/MOD_ProDKin_1.html) |

With the alignment of the PIFs protein sequences we sought to predict a certain degree of similarity between these proteins. However, these *in silico* findings only indicate that the function of several predicted motifs is far to be determined in the context of VirF interaction due to the absence of a particular sequence that could offer clear evidence of a common region involved in the recognition of these proteins by VirF.

## ACKNOWLEDGMENTS

The authors would like to thank Maximiliano Corredor-Adamez for the construction of pSDM3543 to 3546, Miranda van Lier for construction of pSDM3033 and Cielo R. Jurado-Jácome and Martin L. Brittijn for preparing the figures. This work was supported by the Research Council for Earth and Life Sciences (ALW) with financial aid from the Netherlands Organization for Scientific Research (NWO).

## REFERENCES

**Alting-Mees MA and Short JM. 1989.** pBluescript II: gene mapping vectors. *Nucleic Acids Res* **17:** 9494.

**Ang XL, Seeburg DP, Sheng M and Harper JW. 2008**. Regulation of postsynaptic RapGAP SPAR by polo-like kinase 2 and the SCF_-TRCP ubiquitin ligase in hippocampal neurons. *J Biol Chem* **283:** 29424–29432.

**Bai C, Sen P, Hofmann K, Ma L, Goebl M, Harper JW and Elledge SJ. 1996.** SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* **86:** 263-274.

**Bloemberg GV, Wijfes AH, Lamers GE, Stuurman N and Lugtenberg BJ. 2000.** Simultaneous imaging of *Pseudomonas fluorescens* WCS365 populations expressing three different autofluorescent proteins in the rhizosphere: new perspectives for studying microbial communities. *Mol Plant Microbe Interact* **13:** 1170-6.

**Boyer L and Lemichez E. 2004.** Targeting of host-cell ubiquitin and ubiquitin-like pathways by bacterial factors. *Nat Rev Microbiol* **2:** 779-788.

**Cardozo T and Pagano M. 2004.** The SCF ubiquitin ligase: insights into a molecular machine. *Nat Rev Mol Cell Biol* **5:**739-51**.**

**Chen C and Matesic LE. 2007.** The Nedd4-like family of E3 ubiquitin ligases and cancer. *Cancer Metastasis Rev* **26:**587-604.

**Chenna RS, Hideaki KT, Lopez R, Gibson TJ, Higgins DG and Thompson JD. 2003.** Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**:3497-500.

**Dechend R, Hirano F, Lehmann K, Heissmeyer V, Ansieau S, Wulczyn FG, Scheidereit C and Leutz A. 1999.** The Bcl-3 oncoprotein acts as a bridging factor between NF-kappaB/Rel and nuclear co-regulators. *Oncogene* **18:** 3316-3323.

**Ebel W, Skinner MM, Dierksen KP, Scott JM and Trempy JE. 1999.** A conserved domain in *Escherichia coli* Lon Protease is involved in substrate discriminator activity. *J Bacteriol* **181:** 2236-2243.

**Glenn KA, Nelson RF, Wen HM, Mallinger AJ and Paulson HL. 2008.** Diversity in tissue expression, substrate binding and SCF complex formation for a lectin family of ubiquitin ligases. *J Biol Chem* **283:**12717-29.

**Guan KL and Dixon JE. 1991.** Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Anal.Biochem.* **192:** 262-267.

**Heeb S, Itoh Y, Nishijyo T, Schnider U, Keel C, Wade J. WalshU, O'Gara F and Haas D. 2000.** Small, stable shuttle vectors based on the minimal pVS1 replicon for use in gram-negative, plant-associated bacteria. *Mol Plant Microbe Interact* **13**(2): 232-7.

**Hershko A and Ciechanover A. 1998.** The ubiquitin system. *Annu Rev Biochem* **67:** 425-479.

**Jin J, Ang XL, Shirogane T and Wade Harper J. 2005.** Identification of substrates for F-box proteins. *Met Enzymol* 287-309.

**Keith B, Dong X, Ausubel FM and Fink GR. 1991.** Differential induction of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase genes in *Arabidopsis thaliana* by wounding and pathogenic attack. *Proc Natl Acad Sci USA.* **88:** 8821-8825.

**Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG. 2007.** ClustalW and ClustalX version 2. *Bioinformatics* **23:** 2947-2948.

**Liu HT, Li B, Shang ZL, Li XZ, Mu RL, Sun DY and Zhou RG. 2003.** Calmodulin is involved in heat shock signal transduction in wheat. *Plant Physiol* **132:** 1186-1195.

**Lu PJ, Zhou XZ, Shen M and Lu KP. 1999.** Function of WW domains as phosphoserine- or phosphothreonine-binding modules. *Science* **283:** 1325–1328.

**Notredame C, Higgins DG and Heringa J. 2000.** T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302:** 205-17.

**Orzaez D, de Jong AJ and Woltering EJ. 2001.** A tomato homologue of the human protein PIRIN is induced during programmed cell death. *Plant Mol Biol* **46:** 459-468.

**Peters, JM. 2006.** The anaphase promoting complex/cyclosome: a machine designed to destroy.*Nat Rev Mol Cell Biol* **7:**644-56.

**Pinzón VA. 2006.** Introducción a la Bioinformática: Práctica 2: Alineamiento múltiple e dentificación y búsqueda de Motivos. Centro de Bioinformática del Instituto de Biotecnología en la Universidad Nacional de Colombia y del Laboratorio de Micología y Fitopatología de la Universidad de los Andes. Introducción a la Bioinformática. Creative Commons, Bogotá, Colombia.

**Satou M, Toyoda T, Konagaya A, Carninci P, Kawai J, Hayashizaki Y and Shinozaki K. 2005.** Large-scale analysis of RIKEN *Arabidopsis* full-length (RAFL) cDNAs. EMBL/GenBank/DDBJ Databases.

**Schrammeijer B. 2001.** Functional analysis of the virulence protein VirF from *Agrobacterium tumefaciens*. Doctoral Thesis, Leiden University, The Netherlands.

**Tempe D, Casas M, Karaz S, Blanchet-Tournier MF and Concordet JP. 2006.** Multisite protein kinase a and glycogen synthase kinase 3{beta} phosphorylation leads to Gli3 ubiquitination by SCF{beta}TrCP. *Mol and Cell Biol* **26:** 4316-4326.

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

**Walker GE, Dunbar B, Hunter IS, Nimmo HG and Coggins JR. 1996.** Evidence for a novel class of microbial 3-deoxy-D-arabino-heptulosonate- 7-phosphate synthase in *Streptomyces coelicolor* A3(2), *Streptomyces rimosus* and *Neurospora crassa*. *Microbiology* **142:** 1973-1982.

**Watanabe N, Arai H, Iwasaki Ji, Shiina M, Ogata K, Hunter T and Osada H. 2005.** Cyclin-dependent kinase (CDK) phosphorylation destabilizes somatic Wee1 via multiple pathways. *Proc Natl Acad Sci USA* **102:** 11663-11668.