Structure/function studies of effectors from the potato late blight and rice blast pathogens

Freya A. Varden

Thesis submitted to the University of East Anglia
for the degree of Doctor of Philosophy

Department of Biological Chemistry
John Innes Centre

February 2019

Abstract

In a world with an increasing population and changing climate, the availability of food is a growing concern. Crop plants struggle in the face of increasing stresses, including the pressure of diseases that are spread by microbial pathogens. Understanding how plants defend themselves against disease is vital for finding solutions to food shortages.

Microbial pathogens deploy effector proteins as tools to promote infection, and plants evolve to recognise effectors and launch an immune response accordingly. This PhD project focuses on structure/function studies of two effectors from pathogens of important crop plants: the *Magnaporthe oryzae* (rice blast) effector AVR-Pia, and the *Phytophthora infestans* (potato late blight) effector PexRD24.

This work shows that a rice immune receptor pair, Pikp-1/Pikp-2, is able to recognise AVR-Pia, even though this effector is genetically characterised as being recognised by a different receptor. The structural basis of Pikp-1 interacting with AVR-Pia through an integrated domain is shown, and attempts are made to engineer this integrated domain to recognise AVR-Pia more strongly in planta, with some success.

For PexRD24, the interaction with a potato host target enzyme, protein phosphatase 1, is explored. While many issues were encountered with the production of stable, soluble protein, it is shown that PexRD24 can interact with the enzyme in vitro. The enzyme still retains function in complex with the effector, indicating that the role of the effector is not to inhibit phosphatase activity, and may enhance it.

This research uses both biochemical and structural techniques, alongside in planta assays, to broaden the field of knowledge in molecular plant-microbe interactions. By gaining a detailed understanding of how such interactions take place, it will be possible to start engineering plants with a more robust immune system. Through protecting crops against devastating losses by microbial diseases, we move a step closer to global food security.

Acknowledgements

I'd like to start by thanking Mark Banfield, my supervisor, for all his guidance, help and support over the last four years, both when things went well, and even more so when things went wrong! Also to Christine Faulkner, my secondary supervisor, for her valuable insight in review meetings.

The Banfield lab have been a tremendous source of wisdom, technical know-how and inspiration over the last four years. Special thanks go to Richard and Marina, who helped design some of the experiments that are discussed in this thesis and taught me many of the techniques. Collaborators in this work, particularly Hiromasa Saitoh, Paul Birch and Mark Youles, have provided materials and great discussions throughout the project.

There have also been many outside the lab who have provided scientific support – from the crystallography (Clare Stevenson and Dave Lawson) and proteomics (Gerhard Saalbach) platforms, to horticultural and media services. Your help has been completely invaluable. Also thanks should go to everyone in the wider Biological Chemistry department for loaning me your advice, experience and occasionally equipment whenever I asked!

I am hugely grateful to my family who have been an amazing source of support through all the challenges of the last few years. And finally to my friends in Norwich and beyond, I owe a great deal of thanks for encouragement, laughter and helping me keep a sense of perspective. I hope that you know who you are, but I won't namecheck you all here - instead, I will buy you a drink and thank you in person!

Abbreviations

A_{280}	Absorbance at 280 nm
AGF	Analytical gel filtration
AIM	Autoinduction media
AU	Absorbance units
AVR	Avirulence protein
bp	Base pair
CC	Coiled coil
CD	Circular dichroism
cv.	cultivar
DAMP	Damage-associated molecular pattern
dATP	Deoxyadenosine triphosphate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dpi	Days post infiltration
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ETI	Effector-triggered immunity
ETS	Effector-triggered susceptibility
EV	Empty vector
GF	Gel filtration
GTEN	TEN buffer + glycerol
HEPES	4-(2hydroxyethyl)-1-piperazineethanesulfonic acid
HMA	Heavy metal-associated
HR	Hypersensitive response
HRP	Horseradish peroxidase
IMAC	Immobilised metal ion affinity chromatography
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth

LRR	Leucine-rich repeat
MAMP	Microbe-associated molecular pattern
MAX	*M. oryzae* Avrs and ToxB
MBP	Maltose binding protein
MOPS	3-(N-morpholino)propanesulfonic acid
MTI	MAMP-triggered immunity
MW	Molecular weight
NB-ARC	Nucleotide-binding domain found in Apaf1, R proteins and CED4
NLR	Nucleotide binding leucine-rich repeat protein
NMR	Nuclear magnetic resonance
OD_{600}	Optical density at 600 nm
PCR	Polymerase chain reaction
PDB	Protein Data Bank
pNPP	*p*-Nitrophenyl phosphate
POI	Protein of interest
PP1c	Protein phosphatase 1 (catalytic subunit)
PRR	Pattern recognition receptor
PVPP	Polyvinylpolypyrrolidone
RATX1	Related to ATX1
RU	Response units
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPR	Surface plasmon resonance
SUMO	Small ubiquitin-like modifier
TAE	Tris-acetate-EDTA
TBS-T	Tris-buffered saline with Tween®-20
T_m	Melting temperature
UV	Ultraviolet
v/v	Volume to volume
WT	Wild-type
w/v	Weight to volume
Y2H	Yeast two-hybrid

Table of Contents

1. General Introduction 15

1.1 The Plant Immune System 16

1.1.1 Responding to stress 16

1.1.2 MAMPs and cell surface immunity 17

1.1.3 Effectors and effector-triggered immunity 18

1.1.4 Modelling the plant immune system 20

1.2 Oomycete and Fungal Pathogens 21

1.2.1 The threat to food security 21

1.2.2 Pathogen life cycle 23

1.2.2.1 *Magnaporthe oryzae* 23

1.2.2.2 *Phytophthora infestans* 23

1.3 Pathogen Effectors 24

1.3.1 RxLR effectors 26

1.3.2 MAX effectors 27

1.3.3 Effector targets 28

1.4 Immune Receptors 29

1.4.1 Structure of NLRs 30

1.4.2 Models for effector recognition 32

1.4.3 Integrated domains 33

1.5 Project Aims and Objectives 35

2. Materials and Methods 37

2.1.1 Chemicals and Reagents 38

2.2 DNA Techniques 38

2.2.1 Plasmid vectors 38

2.2.2 Agarose gel electrophoresis 38
2.2.3 Polymerase Chain Reaction 39
2.2.3.1 Standard PCR for cloning 39
2.2.3.2 Overlap extension PCR 39
2.2.3.3 Colony PCR 40
2.2.3.4 Purification of PCR products 41
2.2.4 DNA fragment synthesis 41
2.2.5 Cloning 41
2.2.5.1 In-Fusion® cloning 41
2.2.5.2 TOPO® cloning 42
2.2.5.3 Golden Gate cloning 42
2.2.6 Plasmid purification 45
2.2.7 DNA sequencing 45
2.3 Bacterial Techniques 45
2.3.1 Bacterial growth media 45
2.3.1.1 Lysogeny broth 46
2.3.1.2 Auto-induction media 46
2.3.1.3 PowerBroth™ 46
2.3.1.4 EnPresso® 46
2.3.2 Antibiotics 46
2.3.3 Bacterial strains used 46
2.3.3.1 Cloning strains 46
2.3.3.2 Expression strains 46
2.3.4 Bacterial glycerol stocks 47
2.3.5 Making chemically competent *E. coli* 47
2.3.6 Transforming chemically competent *E. coli* 47
2.3.7 Transforming competent *A. tumefaciens* 48
2.4 SDS-PAGE 48
2.5 *E. coli* Expression Screening 49
2.5.1 Small-scale expression testing 49
2.5.2 Medium-throughput expression screening 50
2.6 Protein Purification from *E. coli* 50

2.6.1 Large-scale expression 50
2.6.2 Purification 51
2.6.3 Measuring protein concentration 53
2.6.4 Intact mass spectrometry 53
2.7 Analytical Gel Filtration 53
2.8 Circular Dichroism 54
2.9 Differential Scanning Fluorimetry 54
2.9.1 Assessing stability effect of Mn 54
2.9.2 Screening buffer conditions 55
2.10 Surface Plasmon Resonance 55
2.11 Phosphatase Enzyme Assays 56
2.11.1 pNPP substrate 56
2.11.2 Commercial peptide substrate 56
2.12 Crystallography 57
2.12.1 Commercial crystallisation screens 57
2.12.2 Data collection 57
2.12.3 Data processing 58
2.12.4 Molecular replacement 58
2.12.5 Structure refinement 58
2.13 Transient Expression in *N. benthamiana* 59
2.13.1 Preparation of *A. tumefaciens* 59
2.13.2 Growth of plants 59
2.13.3 Infiltration 60
2.14 *N. benthamiana* Cell Death Assays 60
2.14.1 Scoring cell death assays 60
2.14.2 Confirming protein expression 60
2.15 Co-immunoprecipitation 61
2.16 Western Blotting 61

3. Cross-reactivity in the Pik/Pia systems 63

3.1 Introduction 64

3.2 Pikp responds to AVR-Pia in planta 68

3.2.1 In rice 68

3.2.2 In *N. benthamiana* 69

3.3 Pikp can interact directly with AVR-Pia 73

3.3.1 in vitro 73

3.3.1.1 Purification of AVR-Pia 73

3.3.1.2 In vitro interactions 74

3.3.2 in planta 77

3.4 Pikp-HMA binds AVR-Pia at an unusual interface 79

3.5 Addition of the N-terminal 'arm' 85

3.6 Testing pCambia for AVR-Pia in planta assays 94

3.7 Investigating the 'interface' mutants 97

3.8 A Pikp-HMA/AVR-Pia/AVR-PikD complex is not formed in vitro 100

3.9 Discussion 102

4. Engineering an enhanced response to AVR-Pia 107

4.1 Introduction 108

4.2 Structure-guided point mutations 109

4.2.1 Design of mutants 109

4.2.2 Testing in planta 111

4.3 RGA5/Pikp-HMA secondary structure swaps 112

4.4 Combining swaps enhances the response 116

4.5 RGA5-HMA in Pikp-1 background is autoactive 119

4.6 Purifying swaps in vitro 123

4.7 Swap2+6 has increased affinity for AVR-Pia 125

4.8 Discussion 127

5. PexRD24 targets a host phosphatase 131

5.1 Introduction 132
5.2 PP1c isoforms can be expressed from *E. coli* 134
5.3 PexRD24 can be expressed from *E. coli* 143
5.4 Interaction of PexRD24 and PP1c 147
5.5 Effect of PexRD24 on PP1c enzyme activity 154
5.6 Discussion 159

6. General Discussion 163

Appendix 1 168

Introduction 169
Designing the pOPIN/GG vector system 170
Vector development and testing 172
Mep1 protein expression using pOPIN/GG 175
Discussion 176

Appendix 2 177

References 181

Table of Figures

Figure 1.1-1: The plant immune system. 20

Figure 1.2-1: Photographs of disease symptoms on important crop plants. 22

Figure 1.4-1: Models for effector perception by NLRs. 33

Figure 3.1-1: Gene schematic of Pik-1 and Pik-2 65

Figure 3.1-2: Sequence alignment of Pikp and Pikm HMA domains. 65

Figure 3.1-3: Comparison of the Pikp-1/Pikp-2 and RGA5/RGA4 systems. 68

Figure 3.2-1: Pikp confers partial resistance to *M. oryzae* expressing AVR-Pia. 69

Figure 3.2-2: Pikp shows partial cell death response to AVR-Pia in the *N. benthamiana* model system. 71

Figure 3.2-3: Pikm does not respond to AVR-Pia in the *N. benthamiana* model system. 72

Figure 3.3-1: Purification of AVR-Pia. 74

Figure 3.3-2: Qualitative binding analysis for AVR-Pia with Pikp-HMA and Pikm-HMA. 75

Figure 3.3-3: Quantitative binding analysis for AVR-Pia with Pikp-HMA and Pikm-HMA. 77

Figure 3.3-4: Co-immunoprecipitation assay investigating Pik-1 interaction with AVR-Pia in planta. 78

Figure 3.4-1: Images of protein crystals used for collecting AVR-Pia/Pikp-HMA structural data. 80

Figure 3.4-2: Structure of AVR-Pia in complex with Pikp-HMA. 84

Figure 3.5-1: Addition of N-terminal extension enhances Pikp response to AVR-Pia in rice. 86

Figure 3.5-2: Purification of AVR-Pia$^{NAVR\text{-}PikD}$. 87

Figure 3.5-3: Purification issues for truncated versions of AVR-PikD. 88

Figure 3.5-4: Pikp does not respond to AVR-Pia$^{NAVR\text{-}PikD}$ or AVR-PikD$^{\Delta 22\text{-}52}$ in the *N. benthamiana* model system. 89

Figure 3.5-5: Pikm does not respond to AVR-Pia$^{NAVR\text{-}PikD}$ or AVR-PikD$^{\Delta 22\text{-}52}$ in the *N. benthamiana* model system. 90

Figure 3.5-6: Qualitative binding analysis for AVR-Pia$^{NAVR\text{-}PikD}$ with Pikp-HMA and Pikm-HMA. 91

Figure 3.5-7: Quantitative binding analysis for AVR-Pia$^{NAVR\text{-}PikD}$ with Pikp-HMA and Pikm-HMA. 92

Figure 3.5-8: In vitro interactions of AVR-Pia$^{NAVR\text{-}PikD}$ with Pikm-HMA under high salt conditions with/without C-terminal 6xHis tag. 93

Figure 3.6-1: Pikp NLRs expressed from the pCambia vector trigger cell death even in presence of negative control in *N. benthamiana*. 95

Figure 3.6-2: The impact of expressing AVRs from pCambia in the *N. benthamiana* model system. 96

Figure 3.7-1: Locations of Ala211 and Glu230 on both the AVR-PikD/Pikp-HMA and AVR-Pia/Pikp-HMA structures. 98

Figure 3.7-2: The Golden Gate system cannot be used to investigate the response of the Pikp-HMA^{A211E} and Pikp-HMA^{E230R} mutants to AVR-Pia. 99

Figure 3.7-3: Using the Pikp-HMA^{A211E} and Pikp-HMA^{E230R} mutants to investigate AVR-Pia and AVR-Pia$^{NAVR\text{-}PikD}$ binding sites in planta 100

Figure 3.8-1: Analytical gel filtration indicates that Pikp-HMA cannot simultaneously bind AVR-Pia and AVR-PikD. 102

Figure 4.2-1: Locations of A220 and G214 mutation sites in AVR-Pia/Pikp-HMA structure. 110

Figure 4.2-2: Pikp-HMA^{A220E} and Pikp-HMA^{G214E} do not enhance response to AVR-Pia. 111

Figure 4.3-1: Design of RGA5/Pikp-HMA secondary structure swaps. 113

Figure 4.3-2: RGA5/Pikp-HMA swaps can enhance response to AVR-Pia. 115

Figure 4.4-1: Design of RGA5/Pikp-HMA secondary structure swap combinations. 116

Figure 4.4-2: RGA5/Pikp-HMA swap combinations can enhance response to AVR-Pia. 118

Figure 4.5-1: RGA5-HMA in Pikp-1 background triggers autoactive cell death in *N. benthamiana* 119

Figure 4.5-2: Some RGA5/Pikp-HMA swaps show an autoactive phenotype (see figure on previous page). 123

Figure 4.6-1: SDS-PAGE showing RGA5/Pikp-HMA swaps purified from *E. coli*. 123

Figure 4.6-2: Phyre2 models of RGA5/Pikp-HMA Swap5 and Swap2+5. 125

Figure 4.7-1: In vitro interactions of RGA5/Pikp-HMA swaps with AVR-Pia 127

Figure 5.1-1: Diagram highlighting sequence features of PexRD24 134

Figure 5.2-1: Intact mass spectrometry analysis of PP1c-3 137

Figure 5.2-2: Comparison of PP1c isoforms (see figure on previous page). 137

Figure 5.2-3: Purified PP1c is stabilised by the addition of manganese 139

Figure 5.2-4: Generating truncated versions of PP1c-3. 141

Figure 5.2-5: Generating a phosphatase-dead mutant of PP1c-3 143

Figure 5.3-1: Purification and characterisation of PexRD24 145

Figure 5.3-2: Expression and partial purification of PexRD24mut. 146

Figure 5.4-1: Purification of a PP1c-3^{1-293}/PexRD24 complex 148

Figure 5.4-2: The PP1c-3^{1-293}/PexRD24 complex is more stable in a phosphate-based buffer. 150

Figure 5.4-3: Stabilised proteins allow PP1c-3^{1-293}/PexRD24 complex formation in vitro. 152

Figure 5.4-4: Images of sample PP1c-3/PexRD24 crystals. 153

Figure 5.5-1: Enzyme assays suggest that PexRD24 does not inhibit PP1c-3 activity. 158

Figure 0-1: Features of the pOPIN/GG vectors. 172

Figure 0-2: Testing expression from the pOPIN/GG vectors. 173

Figure 0-3: Expression of Mep1 from the pOPIN/GG vector F5. 175

Table of Tables

Table 1: Vectors used for cloning the constructs generated in this project. 38

Table 2: PCR parameters for PfuUltra II. 39

Table 3: PCR parameters for PfuTurbo. 40

Table 4: PCR parameters for colony PCR. 40

Table 5: Parts used to generate effector transcriptional units by Golden Gate cloning. 43

Table 6: Thermocycling conditions for the digestion-ligation reaction. 43

Table 7: Parts used to generate Pik-1 NLR transcriptional units by Golden Gate cloning. 44

Table 8: Parts used to generate pOPIN/GG transcriptional units by Golden Gate cloning. 45

Table 9: Data collection and refinement statistics for the AVR-Pia/Pikp-HMA complex structure. 83

Table 10: Comparison between the expected and actual molecular weights of purified swap proteins, measured using intact mass spectrometry. 124

General Introduction

When people look outside at the world around them, most don't give plants a second thought; the 'green background' lining our roads and parks is dismissed as rather banal. On reflection, we appreciate plants as being a major food source and key to life on our planet, but only recently have we begun to understand just how complex their interactions with other organisms can be. These complex interactions are at the very centre of this project. The plants discussed here are crops that are responsible for feeding millions of people around the globe, and the relationship between humans and the plants we grow is increasingly elaborate as we engineer and breed plants specifically for our own use. On the other side of this tug of war are the insects, fungi, bacteria, oomycetes and viruses that rely on the plants for their own survival - these interactions are not necessarily destructive, as symbiotic relationships between plants and microorganisms can be found in many ecosystems. In the middle are the plants trying to defend themselves from attack on all sides and promote their own growth and survival.

In this project, we focus on the interaction of pathogenic microorganisms with important crop plants. Due to the breadth of research in this field, a fully comprehensive review of plant immunity is beyond the scope of this introduction, so the focus will be on pathogen effectors and effector-triggered immunity. This chapter outlines firstly how plants use a complex immune system to defend themselves from attack, with a focus on their response to effector proteins delivered by disease-causing microorganisms. The two pathogens of interest are *Phytophthora infestans*, the potato late blight pathogen, and *Magnaporthe oryzae*, the rice blast fungus; the deadly diseases caused by these pathogens are widely studied and highly relevant to global food security. Finally, the chapter will delve deeper into the details of the plant-pathogen interactions, examining the different component proteins at the molecular level.

1.1 The Plant Immune System

1.1.1 Responding to stress

Plants encounter stresses in many different forms. Abiotic stresses such as temperature fluctuations, drought, salt stress and soil pollutants are encountered to different extents all over the world and are responsible for reducing the productivity of many different plants (dos Reis et al., 2012). Biotic stresses such as herbivores, feeding insects and disease-causing microorganisms are also highly destructive, and it is against these stresses that plants have developed a complex defensive immune system (figure 1.1-1). Some of these defences are 'passive', such as a preformed waxy outer cuticle and

secondary metabolite deterrents (Dangl & Jones, 2001; Heath, 2000), and these outer defences are designed to be non-specifically effective against any pathogens in the environment. Disease-causing microorganisms such as bacteria and fungi that are able to breach the surface will encounter further levels of inducible resistance. This introduction will focus primarily on how the plant detects a threat and recognises the presence of a pathogen. Following activation of the immune system, many downstream signalling responses are triggered, which include hormonal responses (Pieterse et al., 2012), ion fluxes (Seybold et al., 2014), triggering of MAP kinase cascades (Thulasi Devendrakumar et al., 2018) and transcriptional changes.

1.1.2 MAMPs and cell surface immunity

If a microbial pathogen can overcome the passive outer defences and breach the plant cell wall, they will encounter the first active line of defence, which is at the cell surface plasma membrane (Spoel & Dong, 2012). Microorganisms contain conserved characteristic features that can be recognised as potential threats by a plant and trigger an immune response. These are known as microbe-associated molecular patterns (MAMPs) but are also sometimes called PAMPs (pathogen-associated molecular patterns). However, as discussed elsewhere (Boller & Felix, 2009) the term 'MAMPs' will be used throughout this thesis because these signature features can be found in a wide range of microbes regardless of whether they are pathogenic to the plant, and therefore the term 'pathogen-associated' seems somewhat misleading. MAMPs are found in all classes of microbes and are essential to organism fitness – typical examples include bacterial flagellin, chitin from fungi and β-glucan from oomycetes (Newman et al., 2013). In addition to MAMPs, endogenous molecules (deriving from the plant cells) known as DAMPs (damage-associated molecular patterns) can also be responsible for inducing an immune response when they are released into the extracellular space from cells that have been damaged by wounding or infection (Choi & Klessig, 2016). There are several different defined classes of plant DAMPs, which may be peptide-based or polysaccharide-based, and even extracellular ATP was recently identified as acting as a DAMP in plants (Choi & Klessig, 2016). However, the focus in this section will be on the recognition of the 'non-self' MAMP danger signals.

MAMPs are detected by receptors on the cell surface, which are known as PRRs (pattern-recognition receptors). In plants, PRRs are located at the cell surface and consist of an extracellular domain, a single-pass transmembrane domain and a cytosolic domain that may possess kinase activity (Frescatada-Rosa et al., 2015). The possession of an

intracellular kinase domain determines the class of PRR. RLKs (receptor-like kinases) possess the kinase domain while RLPs (receptor-like proteins) do not. Given that RLPs lack this signalling domain, they likely act together with RLKs to induce downstream signalling following recognition of the MAMP (Zipfel, 2014). Recognition of non-self danger signals leads to MAMP-triggered, cell surface immunity, also known as MTI (MAMP-triggered immunity), which is characterised by downstream responses that include the production of ROS (reactive oxygen species), the activation of MAP kinase pathways and callose deposition (Zipfel, 2008). These responses either directly provide protection for the plant (e.g. callose deposition reinforces cell walls) or indirectly assist by modifying transcription of defence-related genes.

One of the most well-characterised examples of MAMP recognition by a PRR in plants is the flg22/FLS2 interaction (Trdá et al., 2015). Flg22 is a short epitope (twenty-two amino acids) of the flagellin protein, which is the major component of bacterial flagella. Flg22 is recognised by the PRR FLS2 (flagellin-sensitive 2) (Gómez-Gómez & Boller, 2000). Upon binding of flg22, FLS2 rapidly forms a complex with the protein BAK1 (brassinosteroid insensitive 1-associated receptor kinase 1), which leads to downstream signalling and immunity (Sun et al., 2013).

1.1.3 Effectors and effector-triggered immunity

One characteristic of MAMPs is that they are essential to the structure or life cycle of the pathogen. This means that parts of the molecule which are essential for folding or function are under strong negative selection. However, immunogenic regions of the MAMP can be under diversifying selection, in an attempt to generate new MAMPs that can still carry out their essential function and yet evade host recognition (McCann et al., 2012). However, pathogens have also developed an array of specialist molecules known as effectors, which are not generally considered essential for the basic survival of the pathogen. These effectors are virulence molecules that facilitate infection by manipulating the structure and function of the host cells. Effectors can have a wide range of functions, and are targeted spatially and temporally for maximum potency. Pathogens will frequently produce a whole arsenal of effectors that can be deployed. For example, *Pseudomonas syringae* can secrete a repertoire of around thirty effectors (Chang et al., 2005), while fungi and oomycetes appear to have many more effector candidates that are still being identified (Sonah et al., 2016). Effectors may be targeted to the plant apoplast, or pathogens may use a specialised translocation system to deliver them to the cell cytoplasm, where they can target diverse organelles such as the nucleus, chloroplast

and vesicle trafficking pathways (Deslandes & Rivas, 2012; Petre et al., 2015). Further details about effectors, including their different characteristics, methods of translocation and techniques for identification will be discussed later.

One common function of effector proteins is to disrupt components of the MAMP-triggered, cell-surface signalling pathway. If the pathogen is able to overcome this important first line of active defence, the plant must launch a counter-attack, or risk being left defenceless. Plants have therefore developed an additional defence against these pathogen effectors, which is an intracellular immune response, traditionally known as effector-triggered immunity (also called ETI). The immune components responsible for recognising the effectors and triggering this response are R or resistance proteins, of which the most abundant are the NLR (Nucleotide-binding leucine-rich repeat) proteins, which can recognise effectors from diverse pathogens (Dodds & Rathjen, 2010). The architecture and function of NLR proteins will be discussed further in section 1.4. The hallmark feature of effector-triggered immune signalling is the hypersensitive response, which involves programmed cell death around the site of infection that helps to isolate and halt the spread of the pathogen (Spoel & Dong, 2012). When an effector is recognised by a host resistance protein, its role becomes one of avirulence. The pathogen then faces pressure to delete or mutate an effector to avoid recognition and regain the virulence activity of that effector molecule (Bent & Mackey, 2007).

Many features of the immune signalling pathway are shared between responses that are triggered at the cell surface and those triggered inside the cell, but effector-triggered responses appear to be stronger and extended compared to those triggered by MAMPs (Cui et al., 2015). Despite these distinctions, it has been argued that the boundary between MAMP-triggered and effector-triggered immunity is not as clear-cut as sometimes depicted. Although these definitions of 'MTI' and 'ETI' can sometimes be helpful for simplicity, the responses to MAMPs and effectors may have varying intensities, and some MAMPs may have a narrow distribution while some effectors might be so widespread that they could easily be defined as MAMPs – therefore the boundary between intracellular and cell surface signalling might be somewhat 'blurred' (Thomma et al., 2011).

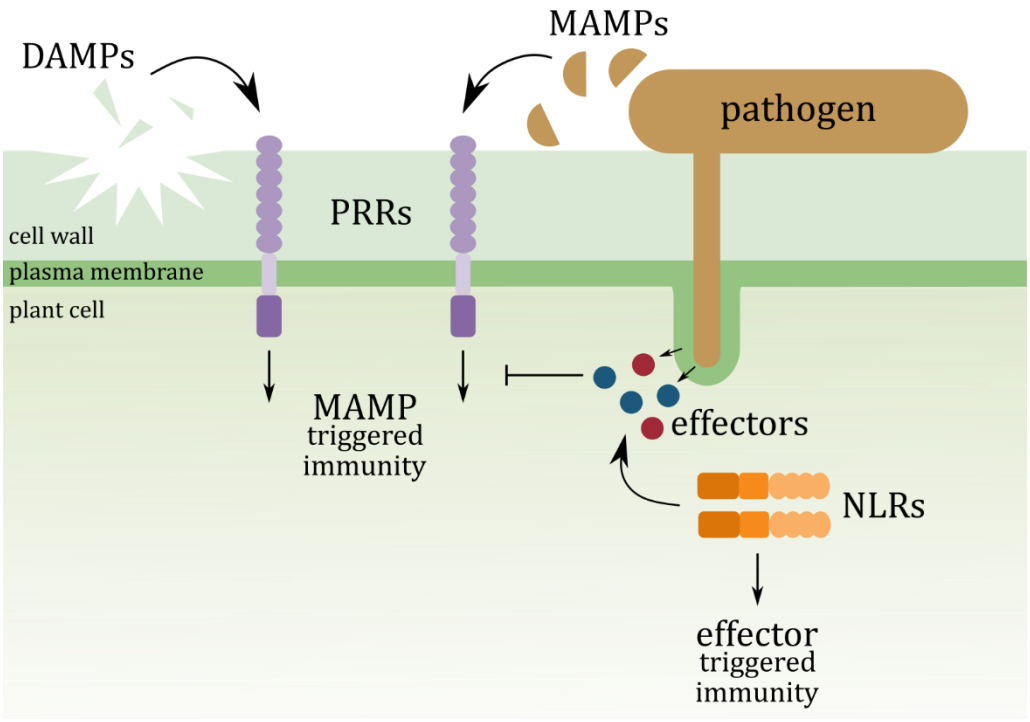


Figure 1.1-1: The plant immune system.

Diagram showing the interactions of the different components of the molecular plant immune system. DAMPs = damage-associated molecular patterns, MAMPs = microbe-associated molecular patterns, PRRs = pattern-recognition receptors, NLRs = nucleotide-binding leucine-rich repeat receptors.

1.1.4 Modelling the plant immune system

Over the last decade or so, there have been a number of models proposed to help conceptualise the plant immune system.

One such proposal is the 'zigzag model' developed by Jones and Dangl in 2006, which describes the hierarchy of different levels of immunity (Jones & Dangl, 2006). In brief, the model shows that plants are able to recognise MAMPs using their cell surface PRRs, then the pathogen deploys effectors to undermine this layer of defence (leading to effector-triggered susceptibility), before the plant evolves intracellular NLR proteins that are able to recognise specific pathogen effectors and launch an immune response. An 'arms race' analogy is often used to describe this process, where pathogen and host plant are in a race to evolve new ways to undermine defences or recognise invasion.

Other, more recent, models have expanded upon the original zigzag concept, arguing that it has limitations, and does not show the full scope of immunity. Pritchard and Birch (Pritchard & Birch, 2014) discuss that the model only includes MAMPs, effectors and their corresponding immune receptors, excluding the contribution from other molecules

that are involved in the plant/pathogen interaction, or environmental factors. The authors also argue that the model cannot be quantified in any way, and they propose a new model with the scope to add in new processes and modify parameters to address some limitations of the zigzag model. Other authors have suggested further modifications to the classic zigzag model, including the 'invasion model' that replaces the immune receptors (either PRRs or NLRs) with invasion pattern receptors (IPRs) and replaces effectors and MAMPs with the term invasion patterns (IPs), which can also be extended to include DAMPs, as they are simply ligands that indicate the chance of an invasion event (Cook et al., 2015). This view of immunity becomes somewhat more generalised but can also prevent constraining a particular pathogen-derived molecule to be either an 'effector' or a 'MAMP'. Finally, the 'multicompetent model' (Andolfo & Ercolano, 2015) attempts to take into consideration the lifecycle and feeding habit of the plant pathogen, arguing that PRRs and NLRs cannot distinguish specifically how the plant is being fed upon; this model attempts to integrate other aspects of the plant defence response that involve a change in primary metabolism.

In reality, the ever-expanding breadth and depth of research on plant immune signalling means that a comprehensive model does not seem feasible. Nevertheless, it remains useful to develop simple models that help to conceptualise and understand generalised plant/pathogen interactions.

1.2 Oomycete and Fungal Pathogens

1.2.1 The threat to food security

The number of people on the planet is increasing, and is predicted to reach 9.6 billion by 2050, up from a current population of around 7.2 billion (Gerland et al., 2014). Adding to this pressure on land and resources is the fact that trade, global movement and agricultural practises are spreading crop pests and pathogens to new regions, where they can infect new populations and thrive. There is also speculation that global warming may play a part in the spread of pathogens to hitherto unaffected areas (Bebber, 2015). Global food security is therefore threatened not just by the growing population and changing climate, but also by crop pathogens that are increasing in potency.

It has long been known that fungal and oomycete pathogens are a huge problem in agriculture, and this project focuses on two globally important crop diseases – potato late blight disease, caused by the oomycete pathogen *Phytophthora infestans*, and rice blast disease caused by the fungal pathogen *Magnaporthe oryzae* (figure 1.2-1).

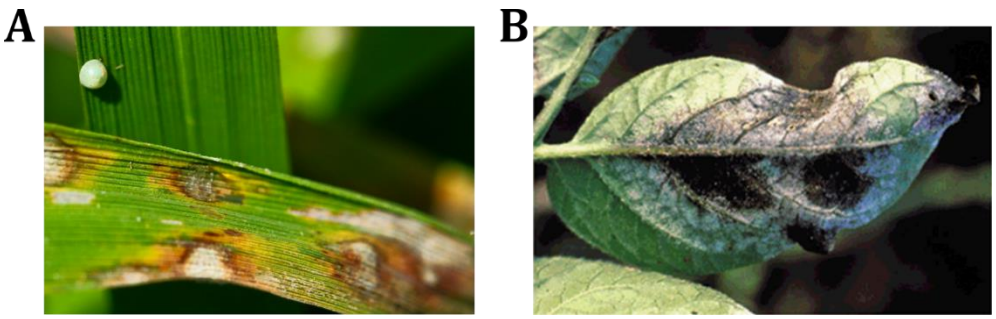


Figure 1.2-1: Photographs of disease symptoms on important crop plants.

A) Rice blast disease lesions (caused by *Magnaporthe oryzae*) seen on rice leaves. Photo sourced from the International Rice Research Institute at http://www.knowledgebank.irri.org/training/fact-sheets/pest-management/diseases/item/blast-leaf-collar. **B**) Late blight disease lesions (caused by *Phytophthora infestans*) seen on potato leaves. Photo sourced from (Fry, 2008)(modified figure).

Potato late blight disease spread to Europe in the nineteenth century, and its effects were particularly devastating in Ireland, where the population was heavily dependent on potatoes as a food source. In the Great Famine (also called the Irish Potato Famine), it is estimated that around one million people died, and late blight remains the most destructive disease of potato, which is one of the world's most important food crops (Yoshida et al., 2013). In a recent survey, scientists agreed that *P. infestans* was currently the most important oomycete pathogen for both its scientific and economic impact (Kamoun et al., 2015). *P. infestans* has been widely studied, and with the sequencing of its genome completed in 2009 (Haas et al., 2009), it is hoped that more can be learnt about its life cycle and method of infection, so that the scientific field will be able to propose better methods of control for agriculture.

Rice blast disease is caused by the ascomycete fungal pathogen *Magnaporthe oryzae*. Rice is a vital food source for billions of people worldwide, and *M. oryzae* is responsible for annual crop losses of an estimated 10-35 % of this total harvest. In context, this means the rice that is destroyed by this pathogen could feed between 212 and 742 million additional people (Fisher et al., 2012). With genome sequence information also available for the pathogen, it has become an important model organism in recent years (Xue et al., 2012).

1.2.2 Pathogen life cycle

Oomycetes and fungi share some apparent similarities in morphology and life cycle, but they in fact derive from different kingdoms; oomycetes are part of the Stramenopiles, whereas fungi form their own taxonomic kingdom (Birch et al., 2006).

1.2.2.1 *Magnaporthe oryzae*

When a fungal spore of *M. oryzae* lands on a rice leaf, it generates a germ tube that forms an infection structure known as an appressorium on the surface. When the appressorium matures, rapid synthesis of glycerol and similar compounds creates turgor pressure inside the structure. A thick inner layer of melanin helps to maintain rigidity and prevents the efflux of glycerol, thus creating an inflexible structure that ruptures the cuticle of the infected cell, allowing hyphae to spread out from the site of infection (Ryder & Talbot, 2015). The initial stages of infection occur in the living rice tissue, but once visible brown disease lesions are present on the rice leaf, the pathogen adopts a more necrotrophic lifestyle, making it a hemibiotrophic pathogen (Wilson & Talbot, 2009). During infection, the host cells are sequentially invaded by specialised invasive hyphae, which allow colonisation of the plant and secretion of effector proteins. Apoplastic effectors are distributed throughout the extrainvasive hyphal membrane (EIHM) compartment, which surrounds the hyphae. Cytoplasmic effectors are localised in the biotrophic interfacial complex (BIC), a plant-derived membrane-rich structure that is initially seen at the primary hyphal tips and remains in place when the hyphae move to a different cell (Giraldo et al., 2013). From the BIC, the cytoplasmic effectors are translocated into the host cell using this pathogen-host interface, although the mechanism for this process is not yet fully understood.

1.2.2.2 *Phytophthora infestans*

Like *Magnaporthe oryzae*, *P. infestans* is hemibiotrophic (Fry, 2008) and can reproduce both sexually and asexually (Goss et al., 2014), but the asexual lifecycle allows rapid spreading of the pathogen across large areas. This life cycle involves sporangia germinating on the plant leaf surface to produce a germ tube that penetrates the plant tissue via formation of an appressorium. Following this initial infection, hyphae spread through the apoplast and projections from the hyphae known as haustoria enter the plant cells by invagination of the surface plasma membrane (Whisson et al., 2011). Several days after infection, the symptoms of *P. infestans* can be seen all over the potato plant, from lesions on the leaf and stem to brown necrotic tissue in the tubers (Fry, 2008). Haustoria are the key sites of exchange between the pathogen and the plant, and upon

invasion of the host cell the host plasma membrane undergoes significant alteration and rearrangement to form a different structure known as the extrahaustorial membrane. However, it is still not clear how effectors are translocated across these barriers into the host cell, and there has been significant debate on the process (Petre & Kamoun, 2014), which is outside the scope of this thesis. However, recent live cell imaging has shown the delivery of a cytoplasmic *P. infestans* effector into the host cell via the pathogen haustoria and through the extrahaustorial matrix (Wang et al., 2017).

1.3 Pathogen Effectors

As discussed in previous sections, microbial pathogens use effector proteins to undermine host defences and promote pathogen colonisation. Effectors can have a wide range of different structures and functions, and new effector properties are being uncovered on a regular basis.

Gram-negative bacterial pathogens use a Type III secretion system for effector delivery, and recently the translocation of *Pseudomonas syringae* and *Ralstonia solanacearum* effectors into the host cell using this system was visualised (Henry et al., 2017). It is interesting to note that *P. syringae* and *R. solanacearum* only secrete around thirty to seventy effectors into the host cell (Henry et al., 2017), whereas the fungal and oomycete pathogens discussed here are predicted to have several hundreds of effector proteins, many of which appear to be important for virulence. This subject has been recently discussed by Thordal-Christensen et al., who proposed that the complexity of infection strategy of filamentous pathogens compared to bacteria might be a factor in the disparity of effector numbers (Thordal-Christensen et al., 2018). Additionally, the considerably larger genome of filamentous pathogens allow for greater diversity of effectors compared to bacterial pathogens.

With so many putative effectors found in fungal and oomycete pathogens, the identification and characterisation of these large catalogues can be a challenge. Although bacterial effectors are not the focus of this thesis, an outline of computational methods used to predict effectors delivered from the Type III secretion system in gram-negative bacteria can be found in McDermott et al. (McDermott et al., 2011). For filamentous pathogens, the in silico prediction of novel effectors can be beneficial for directing further in vitro and in planta studies (Jones et al., 2018). Some effector families (such as the RxLR effectors described below) have signature motifs or conserved regions that can assist in identification. For the RxLR effector family in *P. infestans*, Haas et al. were able

to use the highly conserved RxLR amino acid motif to predict 563 RxLR genes in the pathogen genome, although only around 79 were shown to be significantly upregulated during host infection (Haas et al., 2009). Another defined class of oomycete effectors are the Crinkler (CRN) proteins, which also contain conserved N-terminal motifs (Schornack et al., 2010), allowing them to be more readily identified in sequenced genomes. A number of tools have been generated that use genome and transcriptome data to identify putative effectors with conserved motifs and particular sequences (reviewed by Sonah et al. (Sonah et al., 2016)).

For those effectors, particularly fungal effectors, that do not contain conserved sequence motifs, accurate identification can be more difficult. Effectors are commonly small proteins and often lack sequence similarity to proteins of known function; possibly a result of the pathogen's need to rapidly evolve and diversify its effector repertoire to evade detection (Franceschetti et al., 2017). EffectorP is a tool that was developed specifically for the purpose of identifying such effector proteins (Sperschneider, Dodds, Gardiner, et al., 2018; Sperschneider et al., 2016). Using machine learning, this method can use information gathered from the fungal effectors discovered to date, such as size and amino acid content, and apply these parameters to give sensitive results for effector identification.

As well as specific sequence features, other effector characteristics have been exploited to identify new candidates. As discussed in later sections (sections 1.3.1 and 1.3.2), some pathogen effectors have been shown to have conserved structural elements, including the WY-domain, which is enriched in RxLR effectors (Boutemy et al., 2011), and the MAX fold, found in sequence divergent *M. oryzae* effectors (de Guillen et al., 2015). Using knowledge of these conserved structural elements, it has been possible to identify new putative effectors.

In addition to using bioinformatic approaches to identify effectors, new tools are being developed to assist with effector research in other areas. For example, the LOCALIZER tool predicts the subcellular localisation of effector and plant proteins (Sperschneider et al., 2017). Another machine learning approach, ApoplastP, has been developed to help distinguish whether an effector will be apoplastic or cytoplasmic using known protein features (Sperschneider, Dodds, Singh, et al., 2018). In conjunction with other research methods, such tools could help develop understanding of effector function and identification of host targets.

Despite great advances being made in pathogen genome sequencing and identification of novel effectors, in silico predictions may identify false positives or false negatives in their searches. For putative effectors identified, it is important to use in planta and/or in vitro methods to characterise the proteins fully and understand their biological relevance.

In the following sections, the effector classes of most relevance to this project will be discussed in more detail.

1.3.1 RxLR effectors

RxLR effectors describe a class of oomycete effectors that are comprised of two domains. The C-terminal domain is responsible for the effector's function, and the N-terminal domain contains a highly conserved R-x-L-R motif (where R is arginine, x is any amino acid and L is leucine) often followed by an E-E-R motif (where E is glutamic acid) within thirty amino acids (Grouffaud et al., 2010). The structure of the N-terminal domain of an RxLR effector is predicted to be disordered (Boutemy et al., 2011). The RxLR motif is thought to be involved in translocation of effectors (Whisson et al., 2007), but an understanding of exactly how this occurs is still being developed (Wawra et al., 2017). The C-terminal domain of RxLR effectors is highly varied and generally does not share similarity with proteins of known function (Boutemy et al., 2011). As previously discussed, these domains are likely under high selection pressure to counteract plant defences, and in *P. infestans*, they are located in dynamic, repetitive regions of the genome, which may contribute to rapid diversification (Haas et al., 2009). Despite this diversity, conserved features have been identified in the C-terminal domain, and it has been suggested that many RxLR effectors derive from a common ancestor (Jiang et al., 2008). Both the RxLR effectors AVR3a11 and PexRD2 were characterised structurally and were shown to contain a core α-helical fold, called the WY domain, which was predicted to be present in around 44 % of oomycete effectors (Boutemy et al., 2011). It was hypothesised that the WY domain could allow for significant variation upon this core structural fold, including additions and deletions at protein termini and within loop regions, oligomerisations and repeats of the WY-domain fold; provided the hydrophobic core of the helical bundle is maintained (Boutemy et al., 2011; Win et al., 2012). To date, all except one of the RxLR effectors to have been structurally characterised contain the WY-domain (Franceschetti et al., 2017).

Studies over a number of years have revealed the diversity of RxLR effector function and subcellular localisation. A number of RxLR effectors can translocate to the nucleus (Du

et al., 2015; Wang et al., 2015), while the *P. infestans* effector Pi03192 functions by preventing two host transcription factors (known as NTP1 and 2) from re-localising to the nucleus, where they assist in inducing defence responses (McLellan et al., 2013). AVRblb2, in contrast, localises to the plasma membrane around the pathogen haustoria, and prevents a defensive host protease from being secreted into the apoplast (Bozkurt et al., 2011). In terms of effector function, only one RxLR effector is known to have enzymatic activity itself (Dong et al., 2011) but there are a number that are known to inhibit or alter host enzyme activity in some way (for a recent example, see Fan et al. (Fan et al., 2018)). The manipulation of enzyme activity by effector proteins is an interesting strategy given the relative abundance of effector proteins and their targets. For a relatively small complement of effectors to successful impact all or sufficient of the of active plant enzyme molecules seems difficult to achieve. Other functions include the suppression of host RNA silencing activity (Vetukuri et al., 2017; Xiong et al., 2014), and the targeting of host MAP kinases, either to disrupt signalling (King et al., 2014) or to potentially enhance or direct the MAPK activity elsewhere (Murphy et al., 2018). Finally, a group of RxLR effectors from *P. infestans* have been shown to suppress signalling responses to the MAMP flg22. Although *P. infestans* does not contain the flg22 epitope, it appears that it must contain a MAMP that also activates the same signalling pathway (Zheng et al., 2014).

These examples show the diversity of RxLR effector function. This project will investigate the function of another RxLR effector, PexRD24, which will be discussed in detail in chapter 5.

1.3.2 MAX effectors

Although fungal effectors do not contain obvious conserved sequence motifs like the RxLR effectors, families of effectors are emerging.

The first *M. oryzae* effector to be structurally characterised was AVR-Pizt (Zhang et al., 2013). The core of the effector folds into a six-stranded β-sandwich structure, with a disulphide bond linking strands β-4 and β-5. The N- and C-termini of the protein were disordered (Zhang et al., 2013). The authors of the study noted that a similar β-sandwich fold was present in other known effector protein structures, namely AvrL567 from the flax rust pathogen (Wang et al., 2007) and ToxA from the wheat tan spot fungal pathogen, *Pyrenophora tritici-repentis* (Sarma et al., 2005). Another host-selective toxin from *P. tritici-repentis*, ToxB, was found to be a structural homolog to AVR-Pizt, despite the proteins only sharing around 20 % sequence identity (Nyarko et al., 2014). A number of

other *M. oryzae* effectors that have been structurally characterised also share this β-sandwich fold, while lacking any sequence similarity – AVR-PikD, AVR-Pia and AVR1-CO39 (de Guillen et al., 2015; Maqbool et al., 2015; Ose et al., 2015).

Although previous authors had noted the similarity in these effector structures, it was a study by de Guillen et al. that first named this conserved three-dimensional structural fold the MAX (*Magnaporthe* Avrs and ToxB like) fold. The MAX fold is comprised of a six-stranded β-sandwich fold, made up of two β-sheets stabilised by a disulphide bond (de Guillen et al., 2015). A key feature of the MAX effectors was that they had high sequence diversity, allowing substantial variation in parts of their structure, such as variation in the length of β-strands, the size of exposed loops and the size and charge of surface-exposed residues (de Guillen et al., 2015).

This study (de Guillen et al., 2015) analysed the prevalence of the MAX effector family in fungal pathogens and found that it was greatly expanded in *M. oryzae* and *M. grisea*, accounting for 5-10 % of effector candidates in these species, suggesting that MAX effectors may derive from a common ancestor of these two *Magnaporthe* species. The authors suggest that effectors which are forced to adapt and evolve quickly, leading to diversity in sequence and function, maintain a core structural fold that facilitates translocation into the host and preserves protein stability (de Guillen et al., 2015).

Work to understand the function of MAX effectors from *M. oryae* is still ongoing, but some have been characterised. For example, AVR-Pizt is known to interact with host E3 ligases that play a role in immunity (Park et al., 2016).

The MAX effectors AVR-PikD and AVR-Pia form a major part of this project, and their structure and function will be discussed in more detail in chapter 3.

1.3.3 Effector targets

Effectors often have little sequence similarity to proteins of known function, so it can be difficult to predict their host targets or interacting partners from sequence information. However, to understand the function and biological relevance of an effector, it is important to identify and characterise its interaction with host cellular components. Yeast two-hybrid is a high-throughput method to identify interactors, by screening effectors against bespoke libraries made from infected pathogen tissue (Lu, 2012). This technique has been used to successfully identify interactors for effectors from a range of pathogens, including the effector AVR3a from *P. infestans* (Bos et al., 2010), SnTox3 from *Parastagonospora nodorum* (a disease of wheat) (Breen et al., 2016) and HopD1 from

Pseudomonas syringae (Block et al., 2014). In the context of this thesis, yeast two-hybrid was used to identify the host target of PexRD24, the *P. infestans* effector that is the focus of chapter 5 (Boevink, Wang, et al., 2016). Although not an interaction with a host target, the direct interaction of the effector and corresponding NLR was also determined by yeast two-hybrid for the AVR-Pik/Pik-1 interaction (Kanzaki et al., 2012) and AVR-Pia/RGA5 interaction (Cesari et al., 2013). These interactions will be discussed in more detail in chapter 3.

Another common method to identify effector targets is to purify the effectors from plant tissue using immunoprecipitation, which should pull down interactors from the plant. The co-immunoprecipitation is coupled with mass spectrometry for identification of purified targets (Win et al., 2011). Examples of using this technique to successfully identify effector targets include Petre et al. (Petre et al., 2015) and Chaparro-Garcia et al. (Chaparro-Garcia et al., 2015).

Following identification of potential effector targets, these must be validated using a range of techniques in vitro and in planta. Pairwise yeast two-hybrid and co-immunoprecipitation can be used as confirmatory methods, as well as techniques such as bimolecular fluorescence complementation (BiFC). In vitro techniques are also powerful methods to characterise protein interactions, including qualitative methods such as analytical gel filtration, and quantitative methods such as surface plasmon resonance. These techniques will be discussed in more detail in chapter 3, as these in vitro methods are central to assessing effector interactions in this project.

1.4 Immune Receptors

NLR immune receptors are responsible for detecting effectors deployed from plant pathogens. Since their discovery, much has been learned about the modular structure of these receptors, and different models have been proposed for how they are able to detect and respond to effectors. However, there are still many details to be understood, particularly relating to the downstream signalling that leads to an immune response following effector detection.

Their important role in immunity means that scientists have deployed NLR proteins in crop plants as a way to mediate resistance to important plant pathogens such as *P. infestans*. One setback in these trials is the fact that microbial pathogens can evolve at a rapid rate to undermine the defensive advantage conferred by addition of new NLRs into the plant (Fry, 2008). However, with an increasing understanding of how immune

receptors function, it is becoming plausible that this technique could provide a route to more durable resistance (Jones et al., 2014).

1.4.1 Structure of NLRs

NLR proteins have a multidomain architecture, that comprises an N-terminal domain of either a coiled coil (CC) or toll interleukin-1-like (TIR) type, a nucleotide binding domain known as the NB-ARC domain (nucleotide-binding domain found in Apaf1, R proteins and CED4) and a C-terminal leucine-rich repeat domain (LRR) (Sukarta et al., 2016). The structural and functional analysis of these modular proteins is a major area of current research, but many features have been elucidated in recent years. There are many similarities between plant NLRs and immune receptors from the animal kingdom, meaning that it may be possible to gain further insight by sharing research between the two systems (Bentham et al., 2017).

The N-terminal domain of the receptor is thought to have a signalling function (Bai et al., 2012; Swiderski et al., 2009). Several structures of plant TIR type domains have been solved, and the domain can self-associate through different interfaces (Zhang et al., 2017). Recently the structures of several coiled coil domains have been uncovered, although these revealed some interesting contradictions. It was shown that the CC domain of Sr33, which confers resistance to the wheat stem rust pathogen, had a more similar structure to the CC domain of the unrelated NLR Rx than to its orthologue MLA10, which confers resistance to barley powdery mildew (Casey et al., 2016). One possibility is that the different structures are actually showing different functional states for the CC domain – the MLA10 structure (a dimer of two elongated antiparallel helix-turn-helix monomers) may be showing the active form of the domain, which is dimeric, whereas the Sr33 and Rx CC domains (both found to be four helix bundles) are in the inactive monomeric state (El Kasmi & Nishimura, 2016). However, further work is needed to uncover the exact mechanism of signalling and function of the NLR N-terminal domain. It should be noted that not all NLR proteins have an N-terminal domain of the class TIR or CC. Other types of N-terminal domain have been identified and defined, although less abundant that the two main classes, and will not be discussed further (Shao et al., 2016).

The NB-ARC domain is responsible for the ATPase activity of the NLR and is made up of three subunits that form a nucleotide binding pocket (Sukarta et al., 2016). The nucleotide exchange function of this domain is thought to be important in the activation of the NLR for immune signalling.

The C-terminal domain of the NLR protein is a leucine-rich repeat domain. The LRR domain is thought to have a role in effector recognition specificity and mediating effector interaction (Krasileva et al., 2010). No crystal structure of an LRR domain from a plant NLR has yet been solved, but structures of LRR domains from other proteins have been determined, including the animal NLR protein NLRC4 (Hu et al., 2013). These studies have indicated that the LRR domain may play a role in maintaining the autoinhibited state of the receptor in the absence of pathogen effectors. LRR domains are narrow and arc-shaped, and contain repeating patterns of leucine residues that form a hydrophobic core (Padmanabhan et al., 2009).

One commonly accepted model for NLR activation is that the immune receptor functions as a molecular switch, whereby recognition of an effector disrupts the autoinhibited state of the NLR (which is stabilised by the LRR domain) allowing the formation of an 'open' conformation, which involves nucleotide exchange (exchange of ADP for ATP) in the NB-ARC domain. As the protein undergoes a conformational change, it initiates interaction with downstream signalling partners to trigger an immune response (Lukasik & Takken, 2009). More recent studies have proposed a modification to this model, indicating that the NLR might exist in an equilibrium between the closed and open state, but that the recognition of a pathogen effector stabilises the open state, leading to immune signalling (Bernoux et al., 2016).

Dimerisation, or indeed oligomerisation, of NLRs in immune signalling also appears important. In animals, NLRs become active through the formation of an apoptosome, a heptamer formed by intermolecular interactions that enables signalling. However, similar oligomeric structures formed by the entire NLR have not been conclusively observed in plant immune signalling (Wróblewski et al., 2018). The N-terminal (generally CC or TIR) domain of plant NLRs is capable of forming homo- and hetero-dimers (e.g. (M. Bernoux et al., 2011; Williams et al., 2014)) that are involved in cell death signalling. The formation of such heteromers aligns with the concept of 'sensor' and 'helper' NLRs discussed in section 1.4.3. During evolution, it may be important that these N-terminal domains retain the ability to oligomerise with each other by conserving particular features (Wróblewski et al., 2018). However, dimerisation of these N-terminal domains is frequently not sufficient to enable downstream signalling, and other NLR domains appear to play a role in the transduction of the signal following association (El Kasmi et al., 2017). Further structural and functional studies are required to determine the exact nature of NLR oligomerisation involved in immune signalling.

Despite extensive research into NLR structure and function, and much recent progress, the exact details of how the conformational change and downstream signalling events occur are still lacking.

1.4.2 Models for effector recognition

Aside from ongoing research regarding the downstream signalling of the NLR proteins, there has also been much debate regarding how these receptors recognise effectors. Perhaps the simplest explanation for this question is a solution of one resistance gene having the responsibility of recognising one pathogen effector protein. This was proposed as the 'gene-for-gene' concept (Flor, 1971), and examples of direct binding between NLR and effector proteins have been discovered (Maud Bernoux et al., 2011). However, other observed effector/NLR interactions suggest that the gene-for-gene concept is unlikely to be the only method of detection.

Other models have been proposed for effector recognition (figure 1.4-1) that involve the presence of an additional protein functioning as a mediator between the resistance protein and the effector. These models are known as the 'guard' and 'decoy' models. The guard hypothesis is that NLR proteins monitor the perturbation of other plant proteins by pathogen effectors and induce defence responses when they notice the activity of the effector. Using the example of bacterial speck disease of tomato, it was proposed that the bacterial effector AvrPto was interacting with a host Ser/Thr kinase Pto, which was responsible for interaction with a number of defence-related transcription factors. The host NLR Prf is thought to monitor the Pto kinase protein and initiate downstream defence signalling when it recognises the formation of the complex with AvrPto (van der Biezen & Jones, 1998). Other examples of the guard hypothesis include the host target RIN4, which is targeted by multiple effectors and monitored by two NLR proteins RPS2 and RPM1 (Jones & Dangl, 2006). The decoy model proposes a modification to the guard model, whereby the host factor that is targeted by the effector is actually functioning as a decoy, and plays no part in plant development or immunity, but is required for function of the resistance protein (van der Hoorn & Kamoun, 2008). Later research suggested that the AvrPto interaction system might also fit within this decoy model. It was found that although AvrPto inhibits the kinase activity of Pto, this inhibition was not required for initiating Prf-mediated defence responses, indicating that AvrPto might have other host virulence targets. FLS2 was identified as another such target. This evidence suggests that Pto might be mimicking FLS2 and acting as a decoy, whereby the binding of AvrPto triggers a strong defence response via Prf, but that the inhibition of Pto itself has no effect

on plant fitness (Zipfel & Rathjen, 2008). Other types of decoy systems are present in host/pathogen interactions, including pathogen-derived decoys that are designed to prevent recognition by the host (Paulus & van der Hoorn, 2018).

Another strategy employed by the host immune system is to integrate the 'decoy' protein into the NLR itself, as an additional domain. This model has been previously known as the 'integrated decoy' hypothesis (Cesari, Bernoux, et al., 2014), but in fact there has been some debate whether this term is accurate. For many such situations, it is not clear whether the integrated domain is a true decoy, or whether it has biochemical activity alongside its role in immune signalling - for this reason, the term 'sensor domain' is preferred by many (Wu et al., 2015). To maintain clarity, these domains will simply be referred to as 'integrated domains' in this thesis. The nature and function of these domains will be discussed in section 1.4.3.

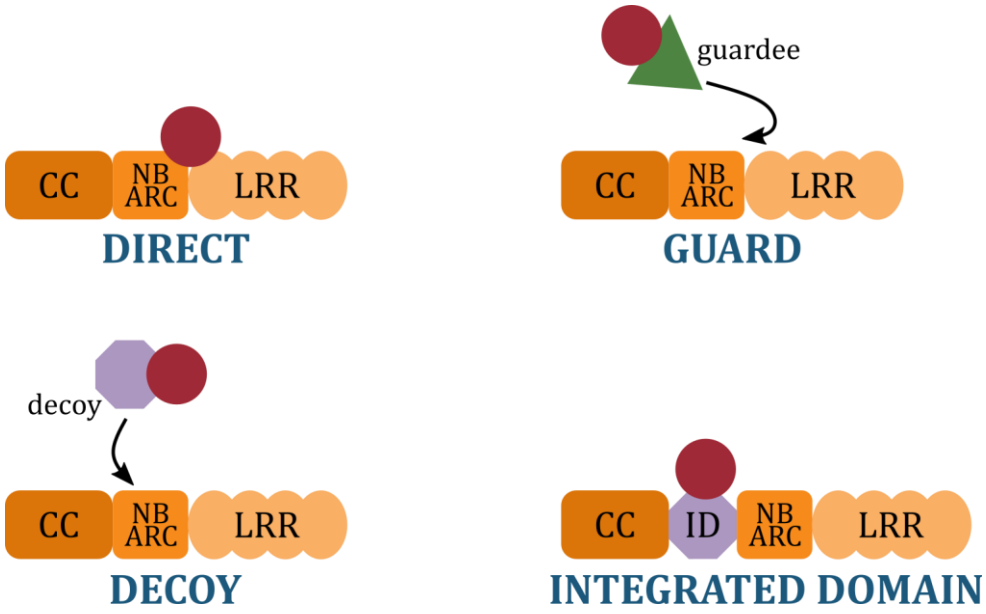


Figure 1.4-1: Models for effector perception by NLRs.

Diagrams highlighting the four different models for effector perception by NLR proteins. Individual domains of NLR proteins are labelled, namely: CC = coiled coil, NB-ARC = Nucleotide-binding, LRR = leucine-rich repeat, ID = integrated domain. Effector is represented by a red circle. Note that CC is being used as a representative N-terminal domain in this diagram.

1.4.3 Integrated domains

One emerging theme for NLR proteins is that they frequently function in pairs (Eitas & Dangl, 2010), where both NLR genes are found at a single locus, tightly linked and transcribed in opposite directions (Cesari, Kanzaki, et al., 2014). It was found that both NLRs in these pairs are vital for immune signalling, and they often appear to have

different architecture, suggesting that they might play separate roles in initiating the defence response. One proposed mechanism was that one NLR would function as an effector sensor, and the other NLR would function as a 'helper' that was required for immune signalling but not specifically for detection of the effector (Bonardi et al., 2011). Recent work has uncovered more about the concept of 'helper' NLRs in plant immunity to reveal complex signalling networks, for example the NRC (NLR required for cell death) family in *N. benthamiana*. Members of the NRC family act as helpers for a variety of sensor NLRs, and it is believed that by diversifying from genetically linked pairs into an unlinked network, the immune system has the potential to accelerate gain-of-immunity to a variety of pathogens. These NRC helpers are not thought to be involved in direct contact with pathogen proteins, and therefore evolve at a slower rate. The functional redundancy in the family may help sensor NLRs to evolve more rapidly and gain new functionality (Wu et al., 2017).

NLR proteins that contain novel integrated domains have been shown in several examples to function in pairs with an additional NLR. The concept of the 'sensor' and 'helper' NLRs fits this model well, as the NLR that contains the integrated domain could function to recognise the effector through its 'decoy' or 'sensor' integrated domain, and the second member of the NLR pair could function as the helper NLR that translates this recognition into an immune signalling response.

One striking feature of these integrated domains is that they can be of varied size and function, and are widespread. In one study, all 31 plant species tested were shown to contain NLRs with integrated domains; on average 3.5 % of the NLR complement for each genome (Kroj et al., 2016). Sarris et al. discovered 265 unique integrated domains from 40 plant genomes (Sarris et al., 2016) but the authors predict this to be an underestimate. Separately to Kroj et al., Sarris et al. predicted that around 10 % of NLRs in each genome contain integrated domains. Both studies used a range of plant genomes in their study, including monocots, dicots and mosses. The discovery of such a wide repertoire of integrated domains may aid the discovery of new effector targets for functional analysis. However, if only 3.5 % of NLR proteins from a typical plant genome contain an integrated domain, this raises questions about the spread and specialisation of these integrations. If they confer a significant advantage onto the resistance capabilities of a plant, it might be expected that this percentage would be considerably larger.

The advantage of integrating these domains into the NLR protein itself, rather than maintaining them as separate proteins, has been discussed (Ellis, 2016). It seems possible that by integrating into the NLR itself, the domain is free to evolve maximum

interaction strength with the pathogen effector (or multiple effectors), rather than maintaining a focus on its original function (Ellis, 2016) (which is not to say that the integrated domain does not retain any biochemical activity). However, the actual evolutionary advantage of integration is still open to debate and conjecture. One logical advantage of integration over maintaining the NLR and pathogen-binding domain separately is that proteins will always be maintained in a physically linked fashion. This ensures identical stoichiometry and identical subcellular localisation, which should aid effector detection and translate binding to signalling even in the presence of a weak interaction. Similarly, by genetically linking the NLR with its partner domain, the two will coexpress more readily and are less likely to be disrupted by genetic recombination (Baggs et al., 2017). From an evolutionary perspective, this genetic linkage might cause the integration to be maintained longer term in the host. This could perhaps be disadvantageous if a pathogen was able to overcome recognition by that particular integrated domain – as the host would have to modify or lose an entire immune receptor, rather than simply by modifying a guard or decoy protein.

Two of the most widely studied NLR pairs containing integrated domains are the rice NLRs Pik-1/Pik-2 and RGA5/RGA4; both these pairs contain an integrated heavy metal-associated (HMA) domain and will be discussed in more detail in chapter 3. Other NLRs known to contain integrated domains include the RRS1/RPS4 pair, which utilises an integrated WRKY domain, a known target of the *Ralstonia solanacearum* effector PopP2 (Le Roux et al., 2015).

1.5 Project Aims and Objectives

The overall aim of this project is to better understand the function and recognition of plant pathogen effectors using biochemical and structural methods. In doing this, it is hoped that this work will become part of a growing field of research that aims to tackle the issue of global food security. Although engineering crop plants to convey greater resistance to disease has its challenges, both scientifically and ethically, it may become essential as we face a growing population and fluctuating climatic conditions. The first step towards engineering of crop plants is to gain a thorough understanding of how the plant and pathogen are interacting at the molecular level.

The first part of the project will investigate the recognition of the *Magnaporthe oryzae* effector AVR-Pia by the mismatched Pik immune receptor. The two aims of this section are:

- To understand how the integrated heavy metal-associated domain of Pikp-1 binds to AVR-Pia (using X-ray crystallography and biophysical protein-protein interaction techniques). Given that the overall structure of AVR-Pia is known to be similar to AVR-PikD, we hypothesise that the interaction of the effector with the HMA domain will take a similar form in both cases.
- To engineer the NLR protein to strengthen or extend the recognition in planta. We hypothesise that it will be possible to modify Pikp-HMA to bind AVR-Pia more strongly in vitro, and that this would lead to a stronger immune response in planta. By rational structure-based design, we predict that it will be feasible to maintain an interaction with the 'matched' effector AVR-PikD

The second part of the project will investigate the *Phytophthora infestans* effector PexRD24, and its interaction with a host enzyme protein phosphatase 1. The aims of this part of the project are:

- To gain structural insight into how PexRD24 binds to the catalytic subunit of its target. We predict that this interaction will involve the conserved RVxF motif found in PexRD24, which is known to be important in the interaction of phosphatase regulatory subunits with PP1c.
- To investigate whether effector binding modulates the enzymatic activity of the phosphatase. Based on work from a collaborating group, we hypothesise that the phosphatase activity of PP1c towards a generic substrate will not be inhibited by the interaction of PexRD24, as PexRD24 appears to be acting as a relocaliser of the catalytic subunit, rather than simply as an inhibitor.

Overall, this project will contribute new insights into the structure/function diversity of filamentous plant pathogen effectors.

Materials and Methods

2.1.1 Chemicals and Reagents

Chemicals were purchased from Sigma-Aldrich (now Merck), VWR (now Avantor), Melford Laboratories, or Thermo Fisher Scientific unless otherwise specified.

2.2 DNA Techniques

2.2.1 Plasmid vectors

The following plasmid vectors were used in this thesis (table 1). Note that a complete list of primers used is shown in Appendix 2.

Table 1: Vectors used for cloning the constructs generated in this project.

Vector	Features	Antibiotic Resistance	Cloning Method	Expression System	Source
pOPINS3C	6xHis-SUMO (N)	Carb	In-Fusion®	*E. coli*	OPPF
pOPINF	6xHis (N)	Carb	In-Fusion®	*E. coli*	OPPF
pOPINM	6xHis-MBP (N)	Carb	In-Fusion®	*E. coli*	OPPF
pOPINE	6xHis (C)	Carb	In-Fusion®	*E. coli*	OPPF
pOPINA	Untagged	Kan	In-Fusion®	*E. coli*	OPPF
pOPIN/GG F1-F6	Allows insertion of various tags for *E.coli* expression (C or N)	Carb/Kan	Golden Gate	*E. coli*	TSL SynBio
pCR™8/GW/TOPO®	Efficient cloning of A-tailed DNA (used as holding vector before GG cloning)	Spec	TOPO®	-	Invitrogen
pICH47751	Lvl 1 acceptor vector, position 3	Carb	Golden Gate	*N. benthamiana*	TSL SynBio
pICH47742	Lvl 1 acceptor vector, position 2	Carb	Golden Gate	*N. benthamiana*	TSL SynBio

2.2.2 Agarose gel electrophoresis

Agarose gel electrophoresis is used to separate DNA fragments according to size. Agarose gels were made by adding ethidium bromide at a final concentration of 0.5 µg/ml to the required volume of molten 1.0 % w/v agarose made up in TAE buffer (40 mM Tris-acetate pH 8.0, 1.0 mM EDTA), and setting in a mould with well-comb. DNA samples were mixed with 4x loading dye (12 % ficoll 400 and 0.25 % w/v Orange G) and loaded into the wells with an accompanying Quick-Load® 1 kb DNA ladder (New England Biolabs, NEB). Electrophoresis was carried out at 90V in 1x TAE buffer. DNA bands were then visualised using a UV transilluminator.

2.2.3 Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) was used for several different purposes in this project.

2.2.3.1 Standard PCR for cloning

PCR was used to amplify DNA for cloning into the vector of interest. Unless specified otherwise, PfuUltra II Fusion HS DNA Polymerase (Agilent Technologies) was used for amplification in a 50 µl reaction volume. 1x PfuUltra II buffer, 100 ng template DNA, 0.2 µM each primer, 0.25 mM dNTP mix and 1 µl PfuUltra II polymerase (at supplied concentration) were mixed on ice and immediately loaded into a T3000 Thermocycler (Biometra) for PCR. Parameters are shown in table 2, below.

Table 2: PCR parameters for PfuUltra II.

PCR Stage	Temperature (°C)	Time (sec)	Cycles
Activation	95	120	x1
Denaturation	95	20	x40
Annealing	57	20	
Extension	72	30	
Final extension	72	180	x1
Hold	12	Hold	x1

Following PCR, 45 µl reaction mixture was mixed with 4x loading dye and used for agarose gel electrophoresis as described in 2.2.2.

2.2.3.2 Overlap extension PCR

Overlap extension PCR was used to generate the H129A point mutation in PP1c-3 (Chapter 5). The protocol for this method was adapted from S.King, Ph.D. thesis, 2013. Four primers were designed to generate two overlapping sections of PP1c-3 DNA that contained the required base changes to code for the H129A point mutation. Two standard PCR reactions were carried out (using PfuUltra II, with a 55 °C annealing temperature) to amplify the fragments – i.e. each PCR product is generated from one flanking primer (from outside the mutation site) and one mutation primer (containing the point mutation). The template DNA for the reaction was PP1c-3 in pDONR201 vector (Supplied by Paul Birch). Although this vector is not used elsewhere in the project, it was used as the template for this reaction because it has a different antibiotic resistance (kanamycin) to the cloning vector (pOPINS3C is carbenicillin resistant), which will prevent contamination with parent DNA. PCR products were purified as described in section 2.2.3.4.

Overlap extension PCR was then carried out by mixing 1 µl of each purified PCR product (above) and diluting 100-fold. PCR was then carried out in a 47.5 µl reaction, containing 1x PfuTurbo® reaction buffer, 0.2 mM dNTP mix, 0.5 µl mixed PCR template and 0.5 µl PfuTurbo® DNA polymerase (Agilent Technologies). The mixture was subjected to 7 cycles of PCR (table 3) to cause extension from the overlapping regions of PCR template.

Table 3: PCR parameters for PfuTurbo.

PCR Stage	Temperature (°C)	Time (sec)	Cycles
Activation	95	120	x1
Denaturation	95	30	x7 then x40
Annealing	60	30	
Extension	72	60	
Final extension	72	600	x1
Hold	12	Hold	x1

0.1 mM dNTP mix and 0.1 µM of each flanking primer were then added to the PCR mixture to a final volume of 50 µl. 40 more cycles of PCR (table 3) were carried out, with the annealing temperature reduced to 55 °C. The resultant full-length PCR product containing the desired mutations was purified and used for In-Fusion cloning (section 2.2.5.1) to clone into pOPINS3C.

2.2.3.3 Colony PCR

Table 4: PCR parameters for colony PCR.

PCR Stage	Temperature (°C)	Time (sec)	Cycles
Activation	95	120	x1
Denaturation	95	15	x30
Annealing	60	15	
Extension	72	15-70	
Hold	12	Hold	x1

To test for the success of generating plasmids containing the correct insert DNA, colony PCR was used to investigate transformed *Escherichia coli*. MyTaq™ Red (Bioline) was used for colony PCR because the red dye element allows direct loading for agarose gel electrophoresis, allowing quick and efficient PCR experiments. For each colony PCR reaction, 1x MyTaq™ Red buffer and 0.15 µM appropriate primers were mixed to a final volume of 19 µl. A single colony was then picked from the agar plate and partially inoculated into the PCR mixture, and 1 µl MyTaq™ polymerase was added. The remainder of the colony was then inoculated into 1 ml Lysogeny broth (LB) media and shaken at 37 °C. Thermal cycling was performed as described in table 4. Extension time was varied depending on the length of construct to be amplified, using 30 seconds per kb DNA.

Success of colony PCR was assessed using agarose gel electrophoresis (10 µl PCR reaction was loaded directly onto an agarose gel). For colonies containing inserts of the correct size, the 1 ml LB culture was transferred to 10 ml LB media (supplemented with appropriate antibiotics) and grown up overnight at 37 °C, 200 rpm. These cultures were then used for DNA preparation and subsequent sequencing.

2.2.3.4 Purification of PCR products

Following electrophoresis, bands of interest were excised from the agarose gel and purification of PCR products (except for colony PCR) was carried out using the NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel) according to the kit protocol.

2.2.4 DNA fragment synthesis

Some DNA fragments used in this project were commercially synthesised by Integrated DNA Technologies as gBlocks® Gene Fragments, which are sequence-verified double-stranded linear chains of nucleic acids. Fragments were designed with correct overhangs and BsaI sites for direct cloning into the appropriate vector (pCR™8 in this case). Upon receipt, synthesised DNA was resuspended in 2.5 mM Tris pH 8.5 buffer to a final concentration of 20 ng/µl. Resuspended DNA was heated at 50 °C for 10 minutes then vortexed to ensure full resuspension.

2.2.5 Cloning

2.2.5.1 In-Fusion® cloning

In-Fusion® cloning was used for generating constructs in pOPIN vectors. pOPIN vectors, developed by the Oxford Protein Production Facility (OPPF), are a suite of vectors designed for expression of constructs with different purification and solubility tags in a range of expression hosts. To clone into the vectors, specific primer extensions are used, allowing direct cloning of PCR product (with appropriate extensions) into the linearised vector (Berrow et al., 2007). Prior to cloning, purified pOPIN vectors must be cut with the appropriate restriction enzymes to allow insertion of the PCR fragment. pOPINF, pOPINS3C and pOPINM were treated with restriction enzymes KpnI and HindIII, pOPINE with NcoI and PmeI, and pOPINA with NcoI and DraI. Cut vector is then purified by gel extraction (section 2.2.3.4). In-Fusion® cloning was carried out by mixing 50-100 ng linearised pOPIN vector with 50-100 ng purified PCR product (generated as described in 2.2.3.1), and 5x In-Fusion® HD Enzyme Premix (Clontech) in a 5 µl reaction. The reaction was incubated at 42 °C for 30 minutes, before 2 µl of the completed In-Fusion® reaction

was transformed into competent *E. coli* DH5α cells (see section 2.3). Transformed cells were plated onto LB agar supplemented with the appropriate antibiotic (carbenicillin or kanamycin), 1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) and 0.02 % w/v X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) for blue/white selection. Due to high efficiency of In-Fusion® cloning, only two white colonies from transformed plates were taken and grown up in LB media for DNA preparation and subsequent sequencing.

2.2.5.2 TOPO® cloning

Commercially-synthesised DNA fragments (section 2.2.4) for Golden Gate cloning were cloned into the pCR™8 vector using TOPO® cloning. These constructs were subsequently used for cloning into level 1 acceptor vectors by Golden Gate cloning. DNA fragments were first modified by A-tailing, to allow ligation into the pCR™8 vector. 6.8 µl resuspended DNA fragment was mixed with 1x Standard Taq Reaction buffer (NEB), 0.2 mM dATP and 0.2 µl Taq DNA polymerase (NEB) in a final volume of 10 µl. Mixture was incubated at 72 °C for 20 minutes. 4.5 µl A-tailed DNA from the reaction was then mixed with 0.5 µl of pCR™8 vector and 1 µl salt solution (both from the Invitrogen pCR™8/GW/TOPO® kit). Mixed components were incubated at room temperature for 30-60 minutes, before 3 µl was transformed into One Shot™ TOP10 competent *E. coli* cells (section 2.3). Transformed *E. coli* were plated onto LB agar supplemented with spectinomycin, and plates were incubated at 37 °C overnight. Colony PCR and DNA sequencing were used to confirm insertion of the correct construct.

2.2.5.3 Golden Gate cloning

Golden Gate cloning was used to generate constructs for in planta transient expression. Several constructs used had been generated previously (de la Concepcion et al., 2018) for other projects, but several effector proteins and Pik-1 mutants were generated specifically for this work.

Golden Gate cloning is carried out using a 'one pot one step' approach in a digestion-ligation reaction (Engler et al., 2008), where restriction enzymes are used to cut out the required fragments for cloning, and ligase enzymes then join the fragments up in the specified order in the acceptor vector. Only level 1 cloning was used in this project, meaning that only BsaI restriction sites were used, and all constructs were cloned into level 1 acceptor vectors. Standard parts (such as promoters and tags) are supplied by TSL SynBio.

To clone the effectors AVR-Pia, AVR-Pia$^{NAVR\text{-}PikD}$ and AVR-PikD$^{\Delta 22\text{-}52}$, PCR was first carried out (as described in section 2.2.3.1) to amplify the effector coding sequence. Primers

were designed to flank the coding sequence with BsaI restriction sites and a four-base overhang that defines how the pieces ligate together in the acceptor vector. Each standard part used is derived from a level 0 module and is also flanked with BsaI restriction sites and overhangs. The overhangs are pre-defined as described in Patron et al. (Patron et al., 2015) . Complete effector transcriptional units are made using the following parts (table 5).

Table 5: Parts used to generate effector transcriptional units by Golden Gate cloning.

Standard Part	Code (TSL SynBio)	Description	5' overhang	3' overhang	Source
Acceptor vector	pICH47751	Level 1 Position 3	CGCT	GGAG	Icon Genetics
Promoter + 5' UTR	pICSL13005	Ubi10	GGAG	CCAT	Mark Youles
N-terminal tag	pICSL30009	4xMyc	CCAT	AATG	Mark Youles
Coding sequence	-	PCR product	AATG	GCTT	-
3' UTR + Terminator	pICH41414	35S	GCTT	CGCT	Icon Genetics

All parts are mixed together in a 2:1 molar ratio of part:acceptor, where 100 ng acceptor vector is used. In a final volume of 20 µl (in H_20), DNA parts were mixed with 1.5 µl 10x BSA (bovine serum albumin), 1.5 µl 10x T4 DNA ligase buffer (NEB), 200 U T4 DNA ligase (0.5 µl) (NEB) and 5 U BsaI (0.5 µl) (Thermo Scientific). The mixture was then subjected to the digestion-ligation reaction (table 6) carried out in the T3000 Thermocycler (Biometra).

Table 6: Thermocycling conditions for the digestion-ligation reaction.

Temperature (°C)	Time (sec)	Cycles
37	20	x1
37	180	x27
16	240	
50	300	x1
80	300	x1
12	Hold	x1

Following the digestion-ligation reaction, 2 µl of reaction mixture was transformed into Stellar™ competent *E. coli* cells (see section 2.3) and plated onto LB agar supplemented with carbenicillin, 1 mM IPTG and 0.02 % w/v X-gal for blue/white selection. White

colonies were then selected for colony PCR (section 2.2.3.3), and subsequent DNA sequencing.

For cloning of mutant Pik-1 NLRs, a similar procedure was used. However, for simplicity of cloning, Pik-1 had previously been divided into separate domains, and Golden Gate standard parts created for each domain (by M. Franceschetti, J. C. de la Concepcion). Therefore, as the mutations were only created inside the HMA domains, only new HMAs needed to be created. Mutant HMA domains were made synthetically (section 2.2.4) and were created with BsaI sites and appropriate overhangs in place. The synthetic HMA domains were cloned into the pCR™8 vector (section 2.2.5.2), which generated them as standard parts for cloning into the level 1 NLR transcriptional unit (table 7).

Table 7: Parts used to generate Pik-1 NLR transcriptional units by Golden Gate cloning.

Standard Part		Code (TSL SynBio)	Description	5' overhang	3' overhang	Source
Acceptor vector		pICH47742	Level 1 Position 2	CGCT	GGAG	Icon Genetics
Promoter + 5' UTR		pICH85281	Mas	GGAG	AATG	Icon Genetics
Coding Sequence	CC	-	Domain Pikp-1	AATG	CAGA	M. F.
	HMA	-	Domain Pikp-1	CAGA	GACG	-
	NB-ARC	-	Domain Pikp-1	GACG	CTTT	M. F.
	LRR	-	Domain Pikp-1	CTTT	TTCG	M. F.
C-terminal tag		pICSL50001	Hellfire (6xHis+3xFLAG)	TTCG	GCTT	Panos Sarris
3' UTR + Terminator		pICH77901	Mas	GCTT	CGCT	Icon Genetics

Golden Gate cloning was then carried out exactly as described above for effector transcriptional units, using parts in a 2:1 molar ratio of parts:acceptor as before. Digestion-ligation reaction and colony PCR were carried out as described.

Appendix 1 describes a new technique tested as part of this PhD project, known as pOPIN/GG; a hybrid of pOPIN based and Golden Gate based cloning. pOPIN/GG constructs were also generated using a Golden Gate cloning procedure. This process was exactly the same as described above but using different standard parts. The coding sequence for the protein of interest was amplified using PCR with BsaI site and overhangs as appropriate, then the PCR product was cloned directly into the pOPIN/GG vector.

Table 8: Parts used to generate pOPIN/GG transcriptional units by Golden Gate cloning.

Standard Part	Code (TSL SynBio)	Description	5' overhang	3' overhang	Source
pOPIN/GG F5 Vector	-	Acceptor (for N-term tag)	GCTT	CCAT	Mark Youles
N-terminal tag	pICSL30019	C3_6xHis	CCAT	AATG	Mark Youles
Coding sequence	-	PCR product	AATG	GCTT	-

Table 8 shows the pOPIN/GG F5 version of the vector, here used for generating constructs with an N-terminal cleavable 6xHis tag. Digestion-ligation reaction was carried out as described previously.

2.2.6 Plasmid purification

To prepare purified DNA, a single colony of transformed *E. coli* (usually the cloning strains DH5α or Stellar™) was grown up overnight at 37 °C, 200 rpm in a 10 ml culture. 10 ml culture was harvested by centrifugation at 3,400 x*g* for 10 minutes to pellet cells. DNA was purified from the cell pellet using the ISOLATE II Plasmid Mini Kit (Bioline) or the NucleoSpin® Plasmid Kit (Macherey-Nagel) according to kit protocols. Concentration of purified DNA was measured using the NanoDrop™ ND-1000 or NanoDrop™ One Spectrophotometers (Thermo Fisher).

For purifications of larger quantities of DNA of higher quality, DNA was prepared using the Plasmid Midi Kit (QIAGEN) according to the kit protocol. Concentration of purified DNA was measured as above.

2.2.7 DNA sequencing

DNA sequencing was performed by LightRun sequencing (GATC Biotech, now Eurofins) using 500 ng DNA mixed with appropriate sequencing primer at 2.5 µM concentration.

2.3 Bacterial Techniques

2.3.1 Bacterial growth media

All media are prepared in de-ionised H_20 and autoclaved prior to use.

2.3.1.1 Lysogeny broth

Lysogeny broth (LB) media (LB Broth Miller, Formedium) is composed of 1.0 % w/v tryptone, 0.5 % w/v yeast extract and 1.0 % w/v sodium chloride, at pH 7.0. For making LB agar (used to culture bacteria on plates), 1.0 % w/v agar was also added, and plates were poured while agar was molten.

2.3.1.2 Auto-induction media

Auto-induction media LB base (AIM) (Formedium) is composed of 1.0 % w/v tryptone, 0.5 % w/v yeast extract, 0.33 % w/v $(NH_4)_2SO_4$, 0.68 % w/v KH_2PO_4, 0.71 % w/v Na_2HPO_4, 0.05 % w/v glucose, 0.2 % w/v α-lactose, 0.015 % w.v $MgSO_4$, trace elements.

2.3.1.3 PowerBroth™

PowerBroth™ media (Molecular Dimensions) was made up from a powder preparation, using 5.2 % w/v PowerBroth™ powder and 0.4 % v/v glycerol.

2.3.1.4 EnPresso®

EnPresso® B Media (BioSilta, now Merck) Starter Pack was used according to the manufacturer's protocol to test for protein expression.

2.3.2 Antibiotics

Antibiotics were used for plasmid and bacterial selection throughout this project. In each case, stock solutions were made at 1000x concentration and used as required. Final concentrations used were: carbenicillin 100 µg/ml, kanamycin 30 µg/ml, spectinomycin 100 µg/ml, rifampicin 50 µg/ml, gentamycin 20 µg/ml, chloramphenicol 34 µg/ml. Most antibiotic stocks were dissolved in de-ionised H_2O and filter sterilised using 0.22 µm Minisart® syringe filters. Rifampicin is dissolved in 100 % DMSO, and chloramphenicol is prepared in 100 % ethanol.

2.3.3 Bacterial strains used

2.3.3.1 Cloning strains

Escherichia coli DH5α (Hanahan, 1983), Stellar™ (Clontech), One Shot™ TOP10 (Invitrogen).

2.3.3.2 Expression strains

Escherichia coli cell strains used were BL21(DE3) (NEB), BL21-AI™ (Invitrogen), SHuffle® T7 (NEB) (Lobstein et al., 2012), Rosetta™ 2(DE3)pLacI (Merck), SoluBL21™

(Genlantis). *Agrobacterium tumefaciens* GV3.101 (Larebeke et al., 1974) was used for transient expression in planta.

2.3.4 Bacterial glycerol stocks

For bacteria transformed with plasmids of interest, saturated overnight liquid culture (with appropriate antibiotics) was mixed with 50 % v/v sterile glycerol and stored at -80 °C.

2.3.5 Making chemically competent *E. coli*

For each cloning and expression strain used, the bacteria were either purchased as chemically competent from commercial suppliers or made chemically competent before use.

To generate chemically competent *E. coli*, a glycerol stock of the strain of interest was inoculated into 10 ml LB media containing appropriate antibiotics and shaken overnight at either 30 or 37 °C (strain dependent) at 200 rpm. 1 ml of overnight culture was diluted into 100 ml fresh LB media and shaken as before until OD_{600} 0.3. Cells were chilled on ice for 5 minutes before being harvested by centrifugation at 2,800 x*g* for 7 minutes at 4 °C. Supernatant was discarded and the cell pellet was then resuspended in 40 ml TfbI buffer (100 mM rubidium chloride, 10 mM calcium chloride, 50 mM manganese chloride, 30 mM potassium acetate, 15 % v/v glycerol, pH 5.8) and incubated on ice for 5 minutes. After a further centrifugation step (2,800 x*g*, 7 minutes, 4 °C) the supernatant was discarded and cell pellets were resuspended in 4 ml TfbII buffer (10 mM MOPS, pH 6.5, 10 mM rubidium chloride, 75 mM calcium chloride, 15 % v/v glycerol). Cells were incubated on ice for 20 minutes and then aliquoted into pre-chilled, sterile Eppendorf tubes (50 µl aliquots), flash frozen in liquid nitrogen and stored at -80 °C.

2.3.6 Transforming chemically competent *E. coli*

To transform chemically competent *E. coli* with the construct of interest, an aliquot of competent cells was thawed on ice, and 50-100 ng of DNA was added directly to the cells. After incubating on ice for 30 minutes, cells were heat-shocked for 30 seconds at 42 °C. 450 µl LB media was added and cells were recovered for 60 minutes at 37 °C with shaking at 200 rpm. Typically, 50-200 µl recovered cells were plated onto an LB agar plate (with appropriate antibiotics) and incubated at 37 °C overnight.

2.3.7 Transforming competent *A. tumefaciens*

Electrocompetent *Agrobacterium tumefaciens* are prepared as described in S.King, Ph.D. thesis, 2013. To transform the cells, a 50 µl aliquot of electrocompetent *A. tumefaciens* GV3.101 was thawed on ice and added to a sterile, cold electroporation cuvette. 50-200 ng DNA was added to the cuvette and placed on ice. Cells were immediately shocked at 2,500 V and 500 µl LB media was added. Cells were recovered for 60 minutes at 28 °C with shaking at 180 rpm. 20-50 µl recovered cells were plated onto an LB agar plate (with rifampicin and gentamycin along with appropriate antibiotics) and incubated at 28 °C overnight.

2.4 SDS-PAGE

Two different SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) systems were used for this project.

Method 1: SDS-PAGE gels were made within the lab and stored at 4 °C prior to use. Resolving gel was made with either 12 or 17 % w/v polyacrylamide, diluted in 375 mM Tris-HCl, pH 8.8 stock with 0.1 % w/v SDS. 0.1 % w/v ammonium persulfate and 0.04 % v/v N,N,N',N'-tetramethylethylenediamine was added immediately prior to pouring into Mini-PROTEAN (Bio-Rad) 1.0 mm glass gel casting plates, up to 2 cm below the top of the shorter plate. Water-saturated butanol was added to the top of the resolving gel during the polymerisation process to ensure a level boundary. After setting, the butanol was removed and the stacking gel was added (5 % w/v polyacrylamide, 63 mM Tris-HCl, pH 6.8 with 0.1 % w/v SDS, 0.1 % w/v ammonium persulfate and 0.1 % v/v N,N,N',N'-tetramethylethylenediamine). Gel casting combs were added to the stacking gel and allowed to set. After gels were set, they were fitted into the Mini-PROTEAN Tetra Cell assembly (Bio-Rad), which was filled with 1x Tris/tricine SDS running buffer (25 mM Tris-HCl, 250 mM NaCl, 0.1 % w/v SDS). Protein samples were prepared by heating at 95 °C for 5 minutes in SDS loading dye (50 mM Tris-HCl, pH 6.8, 100 mM dithiothreitol (DTT), 2.0 % w/v SDS, 0.1 % bromophenol blue, 10 % glycerol). Combs were removed from gels and denatured samples were loaded along with a molecular weight marker ladder (RunBlue™ Prestained Molecular Weight marker (Expedeon) or SeeBlue™ Plus2 Prestained Protein Standard (Invitrogen)). Gels were run at 140-180 V until dye front was near the bottom edge of the gel. Gels were then stained with InstantBlue™ Coomassie Protein Stain (Expedeon) at room temperature for at least 60 minutes, before being rinsed and stored in deionised H_2O.

Method 2: All gels and associated equipment were purchased pre-made from Expedeon, with a range of resolving percentages (12, 16, or 4-20 %). The RunBlue™ Teo-Tricine SDS Mini Protein gels were fitted into the RunBlue™ gel running tank and the tank was filled with 1x RunBlue™ Teo-Tricine SDS running buffer. Protein samples were prepared by heating at 95 °C for 5 minutes in RunBlue™ 4x LDS Sample Buffer, with addition of 100 mM DTT. Denatured samples were loaded along with a molecular weight marker ladder (RunBlue™ Prestained Molecular Weight marker (Expedeon), or PageRuler™ Plus prestained protein ladder (Thermo Fisher) if used for subsequent Western blotting). Gels were run at 140-180 V (or 70-100 V if used for subsequent Western blotting), until dye front was near the bottom edge of the gel. Unless being transferred for Western blot analysis, gels were then stained with InstantBlue™ Coomassie Protein Stain (Expedeon) at room temperature for at least 60 minutes, before being rinsed and stored in deionised H_2O.

2.5 *E. coli* Expression Screening

2.5.1 Small-scale expression testing

For testing expression on a small scale with no purification steps, a single colony was taken from an agar plate or a scraping was taken from a glycerol stock and inoculated into 10 ml LB media (supplemented with appropriate antibiotics). Cultures were shaken overnight at 30 or 37 °C (depending on bacterial strain) at 200 rpm. 200-500 µl overnight bacterial culture were then subcultured into 10 ml fresh LB media and shaken until OD_{600} 0.4-0.8. Expression was then induced with the appropriate reagent, either 1 mM IPTG, or 0.2 % L-arabinose if using BL21-AI™ cells. Following induction of expression construct, cells were shaken at 200 rpm at a reduced temperature of 18 °C for around 16 hours, or at 30/37 °C for 4 hours. Cells were then harvested by taking 1 ml culture and spinning at 5,000 x*g* for 1 minute in an Eppendorf tube. Supernatant was fully removed, and cell pellet was resuspended in 500 µl 1x BugBuster® Protein Extraction Reagent (supplemented with 0.25 µl benzonase) (Merck). Cells were lysed in extraction reagent, with gentle agitation, for 20 minutes at room temperature, and a sample (crude cell lysate) was taken for SDS-PAGE analysis. Lysed cells were then centrifuged at 17,000 x*g*, 4 °C for 15 minutes to pellet insoluble material. A sample of the supernatant (soluble fraction) was taken for SDS-PAGE analysis.

2.5.2 Medium-throughput expression screening

To perform a medium-throughput expression screen, constructs were first transformed into the appropriate expression strain. 4 ml LB media supplemented with appropriate antibiotics was added to each well of a 24-deep-well block and single colonies were added to individual wells. Plate was sealed with a gas-permeable seal and shaken overnight at 37 °C, 200 rpm. Fresh 24-deep-well blocks were filled with expression media (LB or PowerBroth™) supplemented with antibiotics and chemical chaperones dimethyl sulfoxide (DMSO) or ethanol (3 % v/v) as needed, and 200 µl of overnight samples were subcultured into the fresh media. Plates were shaken at 37 °C until OD_{600} 0.4-0.8 for each culture. Cells were then induced with either 1 mM IPTG or 0.2 % L-arabinose as appropriate. Plates were covered with gas-permeable seals and shaken at 18 °C, 200 rpm overnight. Cells were harvested by centrifugation at 3,800 x*g* for 10 minutes in the 24-deep-well block and supernatant was discarded. Cells were resuspended in varying quantities of lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole, 5 % v/v glycerol, 50 mM glycine, 1 protease inhibitor tablet (per 50 ml buffer) (Roche), 1 mM $MgCl_2$, 0.5 µl benzonase (per ml buffer) and 0.25 mg lysozyme (per ml buffer)) to give approximately the same visual density of cells in solution. Cells were then sonicated using a Vibra-Cell™ (Sonics) sonicator fitted with a 24-well probe on ice (40 % amplitude, with sonication frequency 1 second on followed by 3 seconds off for 2 minutes). 30 µl of each lysed cell culture was added to SDS-PAGE loading dye (with DTT) in a 96-well PCR plate in a known orientation. 24-deep-well block containing the lysed cells was centrifuged for 40 minutes at 4,800 x*g* and 4 °C. 30 µl clarified supernatant was added to SDS-PAGE loading dye in the same 96-well PCR block as the crude cell lysate samples. PCR block was heated for 5 minutes at 95 °C, sealed with a foil seal to prevent evaporation. Using a multichannel pipette, cell lysate and soluble fraction samples were loaded into the required number of SDS-PAGE gels and success of expression screen was analysed using SDS-PAGE.

2.6 Protein Purification from *E. coli*

2.6.1 Large-scale expression

Once an expression condition had been shown to be successful using a small-scale expression technique, the expression was scaled up to obtain purified protein. A scraping of glycerol stock for *E. coli* transformed with the construct of interest was inoculated into 50 ml LB media supplemented with appropriate antibiotics. The culture was then grown

overnight (with shaking at 200 rpm) at 30 or 37 °C depending on the strain of *E. coli*. 20 ml of overnight culture was then subcultured into 1 L of expression media supplemented with appropriate antibiotics. For LB expression media, the culture was then grown for 3-6 hours until OD_{600} 0.4-0.8. Expression was then induced with 1 mM IPTG or 0.2 % L-arabinose (depending on *E. coli* strain) and shaken (200 rpm) at 18 °C for another 16 hours or overnight. For AIM expression media (auto-induction media) the induction step is not necessary. When cells are grown in AIM, they metabolise glucose in preference to lactose. When the glucose supplies are exhausted (as cells reach log phase), the cells convert lactose to allolactose, release the lac repressor and induce expression of the T7 RNA polymerase (Studier, 2005). Therefore, when grown in AIM, cultures are shaken at 30 or 37 °C for around 3-5 hours, until OD_{600} 0.4-0.8, then the temperature is simply reduced to 18 °C for a further 16 hours as before.

Cells were harvested at 5,410 x*g* for 7 minutes to pellet cells. Cell pellets were then frozen at -80 °C until use.

2.6.2 Purification

Cell pellets of *E. coli* containing expressed protein were thawed from -80 °C at room temperature. Cells were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole, 5 % v/v glycerol, 50 mM glycine, 1 protease inhibitor tablet (per 60 ml buffer) (Roche)), using 60 ml lysis buffer for each cell pellet (derived from 2L expression culture). Once homogenous, cells were sonicated on ice using the Vibra-Cell™ (Sonics) sonicator, with a single 10-12 mm probe. Sonicator was used at 40 % amplitude, with sonication frequency 1 second on followed by 3 seconds off for 4-8 minutes, depending on cell density. Lysed cells were then centrifuged at 36,250 x*g* for 30 minutes at 4 °C to pellet insoluble material. Soluble fraction supernatant was poured into a fresh tube and placed on ice. If necessary, a sample was taken for SDS-PAGE analysis to confirm production of soluble protein. Automated protein purification was carried out using the ÄKTAxpress system (GE Healthcare). In this procedure, soluble fraction was first loaded onto a nickel-charged 5 ml HisTrap™ HP IMAC column (GE Healthcare) and was washed with binding buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM imidazole, 5 % v/v glycerol, 50 mM glycine) to remove any unbound proteins from the column. The His-tagged protein was then eluted from the Ni-IMAC column in elution buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 500 mM imidazole, 5 % v/v glycerol, 50 mM glycine) and immediately loaded onto a Superdex™ 75 HiLoad™ 26/600 gel filtration column (GE Healthcare) that was pre-equilibrated in gel filtration buffer. For standard protein

preparations, a HEPES-based gel filtration buffer was used (20 mM HEPES pH 7.5, 150 mM NaCl), but for certain purifications, a phosphate-based gel filtration buffer was used (100 mM sodium phosphate pH 7.5, 250 mM NaCl). Protein was flowed through and eluted from the gel filtration column in fractions, that were analysed by SDS-PAGE to determine which fractions contained the proteins of interest. Purified, tagged protein was then pooled and 3C protease enzyme (10 µg per mg of protein) was added. Protein and enzyme were incubated overnight at 4 °C to cleave the His/solubility tag. Untagged protein was then separated from its tag by manually loading onto a pre-equilibrated Ni-IMAC column in binding buffer. Once loaded, Ni-IMAC column was washed with 15 ml binding buffer, to wash out untagged protein. Untagged protein of interested was then loaded onto the Superdex™ 75 HiLoad™ 26/600 column in the ÄKTAxpress system and a further cycle of gel filtration provided further separation from contaminants. Fractions containing purified protein were analysed by SDS-PAGE and pooled. Purified protein was then concentrated by ultrafiltration using Vivaspin® centrifugal concentrators (Sartorius) at 7,500 x*g* and 4 °C. Vivaspin® concentrators of varying sizes and molecular weight cut offs (MWCO) were used, depending on the size of the protein of interest. A MWCO would be chosen to be less than half the molecular weight of the protein e.g. a protein of 8 kDa in size would be concentrated using a concentrator of 3 kDa MWCO. When at an appropriate concentration (depending on intended use) purified protein was either used immediately or flash frozen in liquid nitrogen, followed by storage at – 80 °C.

For smaller scale partial purifications, soluble cell lysate fraction (from a 100 ml expression cell culture) would be loaded manually onto a Ni-IMAC column (5 ml HisTrap™ HP IMAC column) pre-equilibrated in binding buffer. 25 ml binding buffer would then be flowed through the column to wash out any unbound protein. 20-30 ml elution buffer was then run through the column to elute tagged proteins. SDS-PAGE samples were taken at each stage of the process and analysed to ensure solubility of tagged proteins.

For purifications of AVR-Pia in complex with a partner, a modified purification process was used. Both proteins were purified separately, until the stage following tag cleavage. When the complex partners were run through the Ni-IMAC column to separate them from their tags, both proteins were eluted simultaneously into the same tube, and subsequently treated as the same sample for the final gel filtration stage.

2.6.3 Measuring protein concentration

For proteins containing aromatic residues, the concentration was measured by absorbance at 280 nm using NanoDrop™ ND-1000 or NanoDrop™ One Spectrophotometers (Thermo Fisher). Absorbance readings were corrected using the calculated extinction coefficient for each protein, calculated using the ExPASy ProtParam online tool (Gasteiger et al., 2005). For proteins that contained no or very few aromatic residues, protein concentrations were measured using the Direct Detect® Infrared Spectrometer (Merck), which relies on the detection of amide bonds rather than a specific residue composition. Both these techniques enable rapid quantification of protein concentration in a sample, without the need for additional reagents or standard curves. When protein concentration was measured for an experiment, the same protein concentration measuring technique was used for all proteins in the experiment, to ensure consistency.

2.6.4 Intact mass spectrometry

Protein intact masses were determined by staff at the JIC Proteomics Platform, using the Synapt G2-Si mass spectrometer coupled to an Acquity UPLC system (Waters). Spectra were generated and analysed by Dr. Gerhard Saalbach.

2.7 Analytical Gel Filtration

Analytical gel filtration was performed at 4 °C by loading 120 µl of the protein of interest onto a Superdex™ 75 10/300 GL column (GE Healthcare) that was pre-equilibrated in the required gel filtration buffer. Proteins were diluted to 50 or 100 µM concentration in gel filtration buffer prior to loading, and a flow rate of 0.5 ml/min was used. For experiments assessing complex formation, proteins were mixed in a known ratio (dependent on expected oligomerisation state) and incubated on ice for at least 2 hours prior to loading on the gel filtration column. A total of 1.5 column volumes was eluted from the column (36 ml), 0.5 ml gel filtration fractions were collected, and eluent samples were analysed by SDS-PAGE. The Superdex™ 75 10/300 GL column has a void volume of 7.4 ml.

2.8 Circular Dichroism

Circular Dichroism (CD) was performed using a Chirascan™ Plus CD Spectrometer (AppliedPhotophysics), which was fully purged of oxygen (using nitrogen gas) prior to use. Protein samples were diluted into CD buffer (20 mM potassium phosphate, pH 7.2) and used at 0.05-0.2 mg/ml; concentration was adjusted for each individual protein to maximise data quality. Generally, a high-tension voltage (HTV) measurement below 600 V and an absorbance reading of less than 2 AU indicates sufficient quality. Data was collected from 180-260 nm, in 0.5 nm steps at 20 °C. Machine background and buffer blank runs were performed, and subsequent protein measurements were subtracted from the blank run. 4 traces were acquired per sample and averaged to give a final dataset. Chirascan™ software was used to convert readings from milidegrees (machine units) to mean residue molar ellipticity (MRME), which corrects the measurements to account for protein concentration. DichroWeb was used to assign secondary structure features (see Chapter 5).

2.9 Differential Scanning Fluorimetry

Differential Scanning Fluorimetry (also known as the thermofluor or thermal shift assay) is used to assess the stability of proteins, by measuring unfolding at increasing temperatures. This technique was used to assess the effect of metal addition to PP1c-3, and also for a commercial buffer screen.

2.9.1 Assessing stability effect of Mn

In a 50 µl total volume, 5 µg protein, $MnCl_2$ (at 0.5, 1, 2, 4 mM final concentrations) and 2.5x SYPRO® Orange (Invitrogen) were mixed in A4 buffer, in a white 0.2 ml 96-well PCR plate (Thermo Fisher) and sealed with an optical adhesive seal. Plates were spun briefly at 500 x*g* for 1 minute to mix contents and immediately loaded into a CFX96 Q-PCR machine (BioRad). Temperature was increased from 20 to 95 °C in 0.2 °C increments, and fluorescence intensity measurements were taken at each step. The melt curve first derivatives were calculated by machine software and plotted in Excel. Melting temperature (T_m) was calculated from the lowest point of the derivative curve.

2.9.2 Screening buffer conditions

The MD1-96 RUBIC Buffer Screen (Molecular Dimensions) was carried out according to the kit protocol, and DSF was carried out in the CFX96 Q-PCR machine (BioRad). Melt curves were plotted in Excel and examined visually to assess the buffer conditions that conferred the greatest increase in T_m.

2.10 Surface Plasmon Resonance

Protein-protein interactions were measured by SPR (surface plasmon resonance) using the Biacore T200 system (GE Healthcare). All experiments were run at 25 °C analysis temperature, in SPR running buffer (20 mM HEPES pH 7.5, 860 mM NaCl, 0.1 % Tween®-20). 30 µl of the his-tagged effectors were immobilised onto a NTA sensor chip (GE Healthcare) which was activated with 30 µl of 0.5 mM $NiCl_2$. Concentrations of immobilised effector were adjusted in each case to give a response of 250 RU ± 30 upon loading. For R_{max} experiments, HMA was flowed over the immobilised effector at a flow rate of 30 µl/min, at 4, 40 and 100 nM concentrations. HMA was flowed over both the sample cell and the reference cell (a separate flow cell that did not have nickel or effector immobilised on the chip) for 360 sec contact time and 180 sec dissociation time. After HMA dissociation, 30 µl of 350 mM ethylenediaminetetraacetic acid (EDTA), pH 8.0 was flowed over both sample and reference flow cells to regenerate the chip. Following the regeneration step, SPR buffer was flowed over for 30 sec (15 µl) to remove any EDTA from the system prior to the next nickel loading. For experiments to determine the equilibrium dissociation constant, a greater range of HMA concentrations were used, and HMA was flowed over the sample and reference cells at 60 µl/min, for 350 sec contact time and 180 sec dissociation time. Chip was regenerated as before. Reference cell response units were subtracted for each measurement. Response for a buffer blank run (no HMA) was also subtracted from each measurement prior to analysis. For R_{max} measurements, raw data was exported from machine software, and R_{max} was calculated and plotted using Microsoft Excel. For kinetics data, the K_D was determined from the multicycle kinetics curves using the Biacore machine software, with a fit model allocated depending on expected stoichiometry.

2.11 Phosphatase Enzyme Assays

2.11.1 pNPP substrate

Protocol for phosphatase assay using *p*-nitrophenyl phosphate (pNPP) as substrate was adapted from Kelker et al. (Kelker et al., 2009). In an assay volume of 25 µl, phosphatase enzymes at a final concentration of 0.4 µM were mixed with pNPP substrate at final concentrations of 0.5, 1, 2, 4 and 8 mM. The pNPP substrate (NEB) had been made up as appropriate stock solutions in 150 mM Bis-Tris pH 6.5, 150 mM NaCl. The phosphatase assay buffer (used to dilute the enzymes and make up final volume in assay) consisted of 50 mM Tris pH 8.0, 5 mM imidazole, 700 mM NaCl, 1 mM $MnCl_2$. pNPP was added to start the reaction, and the substrate and enzyme were incubated for 15 minutes at room temperature. 25 µl 1 M NaOH was added to quench the reaction, and the absorbance at 405 nm was measured immediately using the CLARIOstar® plate reader (BMG Labtech). Samples were tested in triplicate, and buffer blank samples were run for each pNPP concentration. For this assay, a commercially available phosphatase enzyme was used as a positive control. This enzyme was recombinant human PPP1A, expressed from *E. coli* as a full-length enzyme with an N-terminal His tag (Abcam, ab113150). Raw data i.e. absorbance vs. substrate concentration was plotted in Excel.

2.11.2 Commercial peptide substrate

An alternative method to assess phosphatase enzyme activity is to detect levels of free phosphate using malachite green. This was carried out using the Ser/Thr Phosphatase Assay Kit 1 (Merck) as described in the kit protocol, with some modifications. The reaction volume used was 5 µl, and was quenched with 20 µl malachite green. The final enzyme concentration used in the assay was 0.4 µM. The assay buffer used was standard gel filtration buffer (20 mM HEPES pH 7.5, 150 mM NaCl). Reactions were incubated for 10 minutes before being quenched. Following quenching and colour development, absorbance at 620 nm was measured in the CLARIOstar® plate reader (BMG Labtech). Samples were tested in triplicate, and the assay was repeated three times. Raw data i.e. absorbance vs. substrate concentration was plotted in Excel.

2.12 Crystallography

2.12.1 Commercial crystallisation screens

For crystallisation, freshly purified protein was concentrated to above 5 mg/ml (up to 20 mg/ml) and either flash frozen and stored at -80 °C, or used immediately. Prior to crystallography, any precipitate or dust contamination was pelleted by centrifugation at 17,300 x*g*, 4 °C for 15 minutes. Supernatant soluble protein was transferred to a fresh tube.

A number of commercial crystallisation screens are available, containing a wide range of different buffer compositions, salts and additives. For this project, the Morpheus® (Molecular Dimensions), PACT premier™ (Molecular Dimensions), JCSG-*plus*™ (Molecular Dimensions), Structure (Molecular Dimensions), BCS (Molecular Dimensions), ProPlex™ (Molecular Dimensions), PGA™ (Molecular Dimensions), MIDAS™ (Molecular Dimensions), PEGs suite (Qiagen), $AmSO_4$ suite (Qiagen) commercial screens were used. In addition, a custom screen designed by the PX platform at JIC (Clare Stevenson and Dave Lawson) was used – known as the KISS screen.

Crystallisation screens were set up as sitting drop vapour diffusion experiments. 40 µl of screen solution was transferred to a MRC 2 Well Crystallisation Plate (Swissci) using the Rainin Liquidator™ 96 pipetting system (Mettler Toledo) to aliquot all 96 wells simultaneously. Sitting drops were composed of 0.3 µl purified protein with 0.3 µl reservoir solution, and drops were dispensed using the Oryx Nano or Oryx8 crystallisation robots (Douglas Instruments). In certain cases, where protein precipitated when aliquoted using the robot, only the reservoir solution was dispensed by the robot (0.5 µl) and 0.5 µl purified protein was added to the sitting drop by manual pipetting. Crystallisation plates were sealed immediately, and experiments were incubated at 20 °C. To monitor crystal formation, plates were imaged at regular intervals with the Minstrel Crystallisation Imager (Rigaku), which automatically takes images under visible and UV light at specified timepoints.

2.12.2 Data collection

Crystals of the AVR-Pia/Pikp-HMA complex were fished by Clare Stevenson from the Morpheus® screen. The Morpheus® screen contains cryoprotectant conditions within the screen composition, meaning that crystals obtained from this screen do not need to be treated with an additional cryoprotecting solution. Cryoprotectant is required to

minimise crystal damage due to the formation of ice when the crystal is cooled (Garman & Owen, 2006). Crystals were frozen immediately in liquid nitrogen, and shipped to the Diamond Light Source synchrotron in Oxfordshire, where X-ray diffraction experiments were performed. Data was collected remotely on beamline DLS-i03. The dataset was collected at a wavelength of 0.9763 Å.

2.12.3 Data processing

Crystallographic data was processed using the Xia2 pipeline (Winter, 2010) and AIMLESS (Evans & Murshudov, 2013), as implemented in the CCP4 software suite (Winn et al., 2011).

2.12.4 Molecular replacement

For this project, the calculated electron density was interpreted by using molecular replacement to solve the crystallographic phase problem. This was a suitable solution in this case because both partners of the complex (AVR-Pia and Pikp-HMA) had already been structurally characterised elsewhere. The structure of AVR-Pia had been solved by NMR spectroscopy (PDB file 2MYW) and the structure of Pikp-HMA had been solved by X-ray crystallography (PDB file 5A6P). A single model from the ensemble of AVR-Pia and a monomer structure of Pikp-HMA were used for molecular replacement using PHASER (McCoy et al., 2007).

2.12.5 Structure refinement

Following molecular replacement, a refinement process was required to improve the fit between the model and the experimental data. COOT (Emsley et al., 2010) was used to visualise the model and experimentally-derived electron density, and manual rebuilding was carried out to generate a good fit. Water molecules were also added into appropriate regions of electron density. Successive rounds of manual rebuilding were followed by rounds of refinement using REFMAC5 (Murshudov et al., 2011), until the model and observed data were in close agreement, as assessed by R/R_{free} values. Structure was validated using tools provided in COOT, and finally assessed by MolProbity (Chen et al., 2010). Finalised structures were visualised in the CCP4 molecular graphics program (CCP4MG) (Winn et al., 2011) to generate figures for presentation in this thesis.

2.13 Transient Expression in *N. benthamiana*

2.13.1 Preparation of *A. tumefaciens*

Following transformation of *A. tumefaciens* GV3.101 with the constructs of interest, glycerol stocks were made as described previously and stored at -80 °C. Glycerol stocks were then used to grow up bacterial cultures for infiltration. Two different methods were used for growing up the bacteria, but within each experiment only one method was used to grow up all bacteria, to ensure consistency.

Method 1: A small amount of glycerol stock was transferred, with a pipette tip, into LB media (10 ml in a screw-top glass Universal) containing rifampicin, gentamycin and the appropriate antibiotic for the transformed plasmid (generally carbenicillin for Golden Gate level 1 constructs, kanamycin for P19). 10 ml cultures were set up for each construct, and were shaken at 28 °C and 180 rpm for approximately 40-48 hours. Each culture was then harvested by centrifugation at 3,400 x*g* for 20 minutes, or until cells had formed a pellet. The cell pellet was then resuspended in infiltration buffer (see below).

Method 2: 20 µl glycerol stock (thawed) was spread evenly over a LB agar plate containing rifampicin, gentamycin and the appropriate antibiotic for the transformed plasmid. Plates were then incubated at 28 °C for 40-48 hours. A sterile plastic loop was used to scrape cells from the surface of the agar plate and resuspend them in infiltration buffer (see below).

Following growth of agrobacterium cultures, cells were resuspended in infiltration buffer (10 mM MES, pH 5.6, 10 mM $MgCl_2$, 150 mM acetosyringone) to give a concentrated solution of cells. The OD_{600} of a 40x dilution of this cell solution was measured and recorded. The OD_{600} measurement was then used to dilute cells in infiltration buffer to give stock solutions of known densities. These were then mixed in appropriate ratios for the experiment. For each infiltration site, the components were: NLR proteins (at OD_{600} 0.4 final concentration), effectors (at OD_{600} 0.6 final concentration and P19 (at OD_{600} 0.1 final concentration). Once cells were resuspended in infiltration buffer, they were incubated at room temperature for 2-5 hours prior to infiltration.

2.13.2 Growth of plants

Nicotiana benthamiana for cell death assays were grown in controlled environment rooms with 22 °C temperature, 80 % humidity and a 16/8 hour light/dark photoperiod.

Plants that were infiltrated for protein extraction and immunoprecipitation only were grown in glasshouses with similar conditions. Plants were infiltrated at four weeks old.

2.13.3 Infiltration

For each *Nicotiana benthamiana* plant, the two youngest leaves of a usable size were infiltrated. A blunt syringe was used to infiltrate bacterial solution into the abaxial surface of the leaf.

2.14 *N. benthamiana* Cell Death Assays

2.14.1 Scoring cell death assays

Cell death assays were carried out by infiltrating different combinations of NLRs and effectors (and P19) in spots on each leaf. The position of each spot was rotated around the leaf, to account for variation in different areas. Following infiltration, plants were left in the same controlled environment conditions for 5 days. 5 dpi, infiltrated leaves were harvested from the plant and photographed under UV and daylight. Photographs were then used to assign cell death scores from 0 to 6, using the developed scoring system (Maqbool et al., 2015). Scores for three independent repeats were displayed as box and whisker plots, generated using R (Team, 2008) and graphics package ggplot2 (Wickham, 2016).

2.14.2 Confirming protein expression

Following photography of leaves, 3x 1 cm leaf discs were taken for each sample from the set of infiltrated leaves (leaves were chosen at random). These leaf discs were flash frozen in liquid nitrogen and ground up using a pre-chilled micropestle. 300 µl plant protein extraction buffer was added to the ground leaf tissue. Extraction buffer is made from GTEN (25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 10 % v/v glycerol) and supplemented with 10 mM DTT, 2 % w/v PVPP, 0.1 % Tween®-20, 1x plant protease inhibitor cocktail (Sigma). Leaf tissue was then pelleted by centrifugation at 20,000 x*g*, 4 ^{o}C for 5 minutes. Supernatant was transferred to a fresh pre-chilled Eppendorf tube and centrifuged for a further 2 minutes. 20 µl of clarified supernatant (soluble protein extract) was mixed with 8 µl SDS-PAGE loading dye, and used for SDS-PAGE and subsequent Western blotting to confirm expression of tagged proteins of interest in the leaf tissue.

2.15 Co-immunoprecipitation

For co-immunoprecipitation (co-IP) analysis, whole-leaf infiltrations were carried out using the same constructs as for cell death assays, but without expression of Pikp-2, to prevent cell death in the tissue. 4 dpi, whole leaves were harvested from the plants with removal of leaf midrib. 3 leaves for each sample were combined and flash frozen in liquid nitrogen. Tissue was then ground up in liquid nitrogen in a pre-chilled pestle and mortar, before being resuspended in 2x tissue weight of ice-cold plant protein extraction buffer. Tissue was fully resuspended in the buffer by vortexing. All subsequent steps were carried out at 4 °C with pre-chilled equipment and consumables. Samples were centrifuged at 4,200 x*g* for 30 minutes, and supernatant was decanted and filtered using a 0.45 µm Minisart® syringe filter. 20 µl of this soluble protein extract (co-IP input) was taken and mixed with 8 µl SDS-PAGE loading dye.

1 ml of soluble plant protein extract was then mixed with 20 µl of α-FLAG® M2 magnetic beads (Sigma), which had been equilibrated in IP buffer (GTEN + 0.1 % Tween®-20) by washing 4 times, and turned end-over-end for 60-90 minutes at 4 °C. Using a magnetic rack to separate beads and liquid, the supernatant was discarded, and magnetic beads were washed 5 times with ice-cold IP buffer. Bound protein was then eluted from the magnetic beads by adding 30 µl SDS-PAGE loading dye and heating at 70 °C for 10 minutes. Dye was then separated from magnetic beads before loading onto a gel, along with the co-IP input samples.

2.16 Western Blotting

Following SDS-PAGE of plant protein extract (either for expression testing or co-IP), Western blots were carried out to visualise the proteins. Gels were transferred using the Trans-Blot® Turbo™ Transfer System (Bio-Rad) to a PVDF (polyvinylidene difluoride) membrane, according to the manufacturers protocol, using the 'High molecular weight' programme. Membranes were then blocked in 20 ml TBS-T (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1 % Tween®-20) supplemented with 5 % w/v dried milk powder for at least 60 minutes at 4 °C with gentle agitation. Following blocking, membranes were incubated in TBS-T + 5 % w/v dried milk + 1° antibody overnight at 4 °C with gentle agitation.

Antibody solution was removed, and the membrane was washed 3 times in TBS-T, with 20 minutes incubation in TBS-T following each wash. To visualise the blots, excess liquid was removed, and 125 µl each of LumiBlue ECL Extreme reagents (Expedeon) were

mixed and immediately applied to the membrane to provide a thin coverage. Chemiluminescence was detected using the ImageQuant LAS 500 spectrophotometer (GE Healthcare).

Following membrane visualisation, Western blot was stripped by covering with a thin layer (approximately 15 ml) of Restore™ Western Blot Stripping Buffer (Thermo Fisher) and incubating with gentle rocking at room temperature for at least 60 minutes. Membranes were then washed 3 times with TBS-T and blocked again for further addition of antibody.

Following application and visualisation of all antibodies, staining of total protein for each blot was achieved by incubation of each membrane with 15 ml Ponceau S stain (0.1 % w/v Ponceau S in 5 % v/v acetic acid). Excess background staining was then washed off with dH_20, and membranes were imaged by scanning using a generic document scanner.

In this project, the following antibodies are used for Western blot analysis. All are primary, HRP (horse radish peroxidase) conjugated antibodies.

Antibody	Supplier	Dilution Used
α-FLAG HRP	Generon	1:5000
α-Myc HRP	Santa Cruz Biotechnologies	1:1000
α-HA HRP	Thermo Fisher	1:3000

Cross-reactivity in the Pik/Pia systems

Investigating the cross-reactivity of Pik integrated domains with the *M. oryzae* effector AVR-Pia

3.1 Introduction

Integrated domains in NLRs can take many forms, and this chapter will focus on a pair of NLR proteins, Pik-1/Pik-2 (Zhai et al., 2011), that contain an integrated heavy metal-associated domain between the coiled-coil (CC) and nucleotide-binding (NB-ARC) domains of Pik-1. As discussed for the integrated domain hypothesis, this suggests that an HMA domain-containing protein might have been the target of a pathogen effector. Functional HMA domain proteins contain a C-x-x-C motif (where C is cysteine and x is any amino acid) that is able to coordinate heavy metals (Abreu-Neto et al., 2013) and such metal-binding proteins are important in all organisms, including plants. The rice protein Pi21, which contains a heavy metal-associated domain, is known to be a susceptibility factor for *M. oryzae*, suggesting that these types of proteins are involved in immunity, and could be possible effector targets (Fukuoka et al., 2009). However, no specific examples of host HMA domain-containing effector targets have yet been published. In a recent transcriptome profiling study, it was found that four genes encoding proteins containing HMA domains were expressed at higher levels in Pi21-silenced rice lines compared to wild-type Nipponbare, when the plants were treated with *M. oryzae*. This could suggest that these HMA domain-containing proteins play a role in the resistance to *M. oryzae* that is mediated by the loss-of-function of Pi21 (Zhang et al., 2016). Aside from this involvement of Pi21 in rice immunity, metals and metal transport are known to be important in other plant defences, including iron redistribution in the response of wheat to *Blumeria graminis f. sp. tritici* (Liu et al., 2007).

The Pik locus is found on the long arm of rice chromosome 11 (Zhai et al., 2011) and comprises seven different Pik alleles (Ariya-anandech et al., 2018). The Pik-1/Pik-2 NLRs are genetically linked in a head-to-head orientation and share a common promoter (figure 3.1-1). Pik-1 functions as the 'sensor' NLR that recognises its cognate effector via the integrated domain, while Pik-2 is the helper NLR that assists in mediating signalling (de la Concepcion et al., 2018).

The *M. oryzae* effector recognised by the Pik pair of NLRs is AVR-Pik (Yoshida et al., 2009). It has been shown that direct interaction occurs between the Pik HMA domain and AVR-Pik and that this is required for triggering an immune response to the effector (Maqbool et al., 2015). Both the Pik NLRs and AVR-Pik effectors are found as allelic series in their respective organisms. Pik-1 contains two residues in the HMA domain that are

highly polymorphic and are used to readily distinguish the different alleles; it is likely that these different alleles evolved as part of the arms race between rice and the fungal pathogen (Kanzaki et al., 2012). Details of these different alleles have been discussed elsewhere (de la Concepcion et al., 2018; Maqbool et al., 2015) but for this project, only two alleles of the Pik NLRs and two alleles of AVR-Pik effectors will be discussed. These are the Pikp and Pikm NLR alleles and AVR-PikD and AVR-PikC effector alleles.

Figure 3.1-1: Gene schematic of Pik-1 and Pik-2.

Diagram showing the orientation of Pik-1 and Pik-2 at the Pik locus, including shared promoter and similar domain architecture.

The Pikp pair (Yuan et al., 2011) and Pikm pair (Ashikawa et al., 2008) show the greatest pairwise sequence variation within the large Pik-1 family (Costanzo & Jia, 2010). This variation is largely due to the highly polymorphic HMA domain between the two alleles (62 % sequence identity) (de la Concepcion et al., 2018). While the other domains within the two Pik-1 proteins are very similar (96-100 % protein sequence identity throughout), the HMA domains of the different Pik alleles differ considerably. For illustrative purposes, an alignment of Pikp-HMA and Pikm-HMA is shown in figure 3.1-2.

```
Pikp-HMA -GLKQKIVIKVAMEGNNCRSKAMALVASTGGVDSVALVGDLRDKIEVVGYGIDPIKLISALRKKVGDAELLQVSQANKD
Pikm-HMA GGEMQKIVFKIPMVDDKSRTKAMSLVASTVGVHSVAIAGDLRDQVVVVGDGIDSINLVSALRKKVGPAMFLEVSQVKED
          *  ****:*: * .::.*:***:***** **.***:.*****:: *** *** *:*:******** * :*:***.::*
```

Figure 3.1-2: Sequence alignment of Pikp and Pikm HMA domains.

Sequence alignment of Pikp-HMA and Pikm-HMA amino acid sequence. Shaded residues indicate the two well-characterised and highly polymorphic sites in the Pik-1 family.

Pikp and Pikm have different recognition specificities for different AVR-Pik alleles; Pikp is only able to recognise AVR-PikD, whereas Pikm can recognise AVR-PikD and other additional AVR-Pik alleles. It has therefore been proposed that the Pikp/AVR-PikD interaction is older in terms of evolutionary time, and that this recognition has subsequently been broken by other AVR-Pik effectors. This would lead to the development of other Pik NLRs, such as Pikm, in accordance with the plant/pathogen arms race. Different rice accessions studied have shown that the different Pik alleles are present at different frequencies across the species, and more accessions appear to contain Pikp than Pikm (Kanzaki et al., 2012). One AVR-Pik effector allele that is currently unrecognised by any Pik NLR is AVR-PikC, and this allele will act as a negative

control in several experiments described in the following chapter. AVR-PikD will function as a positive control as it is recognised by both Pikp and Pikm.

Another pair of rice NLRs that contain a heavy metal-associated domain is the RGA5/RGA4 pair. RGA5/RGA4 NLRs are able to respond to the *M. oryzae* effectors AVR-Pia (Okuyama et al., 2011) and AVR1-CO39 (Cesari et al., 2013). The HMA domain is found in the C-terminus of RGA5, and is known as RATX1, meaning 'related to ATX1' where ATX1 is a known copper-binding protein (Cesari et al., 2013). Both AVR-Pia and AVR1-CO39 have been shown to physically interact with the RATX1 domain (which will be called the RGA5-HMA domain in this project for consistency and clarity). Cesari et al. showed that physical interaction of AVR-Pia was required for triggering resistance by studying AVR-Pia variants with two nonsynonymous polymorphisms, which prevented an interaction with RGA5-HMA and also prevented pathogen recognition (Cesari et al., 2013). Ortiz et al. used NMR titration and yeast two-hybrid experiments to identify a candidate interaction surface between AVR-Pia and the RGA5-HMA domain. The authors also found that AVR-Pia could interact with other regions of RGA5 outside of the HMA domain, but that interaction with RGA5-HMA was required for activation of immunity (Ortiz et al., 2017).

Despite the similarities between the RGA5/RGA4 and Pik-1/Pik-2 systems, it appears that the mechanisms of activation are very different between the two pairs of NLRs. While Pik-1/Pik-2 appear to use a cooperative mechanism, where the recognition of effector causes a positive change that allows signalling to occur via Pik-2, the RGA5/RGA4 pair appears to function via negative regulation, where recognition of the effector through RGA5-HMA derepresses signalling through RGA4, meaning that RGA4 is constitutively active unless repressed by RGA5 in the absence of effector (Bialas et al., 2018; Cesari, Kanzaki, et al., 2014). However, the details of the NLR interactions and the resultant downstream signalling effects remain to be understood.

As discussed previously, AVR-Pia and AVR-PikD are both members of the MAX effector family and share a common structural fold. The crystal structure of AVR-PikD (de la Concepcion et al., 2018; Maqbool et al., 2015) and the solution structure of AVR-Pia (de Guillen et al., 2015; Ose et al., 2015) show the similarities between these effectors. However, the structure of AVR-Pia has so far only been published in isolation, whereas AVR-PikD has been structurally characterised in complex with the HMA domain of both Pikp-1 and Pikm-1. In fact, recent research has shed light on the interactions of the AVR-Pik effectors and Pik-HMA domains in some detail. In 2015, the first crystal structure of Pikp-HMA in complex with AVR-PikD was solved. Pikp-HMA has an α/β sandwich

structure comprising a four-stranded antiparallel β-sheet and two α-helices, and its overall structure does not change when a complex is formed with AVR-PikD. The HMA domain cysteines required for metal coordination are not conserved in Pikp-HMA, so no metal is bound in the structure (Maqbool et al., 2015). When AVR-PikD and Pikp-HMA form a complex, hydrogen bonding between β-strands leads to the formation of a continuous β-sheet throughout the global complex structure. A second site of interaction involves Asp224 from Pikp-HMA, which forms salt bridge and hydrogen bonding interactions with AVR-PikD β-2. The final major site of interaction between Pikp-HMA and AVR-PikD is formed by the N-terminal extension of AVR-PikD, which protrudes out from the core β-sandwich structure of the effector and partially wraps around Pikp-HMA. Of particular importance within this region is the AVR-PikD residue His46, which forms interactions with Ser218, Glu230 and Val232 from Pikp-HMA (Maqbool et al., 2015). His46 is one of the polymorphic residues that distinguishes AVR-PikD from the other AVR-Pik effector alleles (Kanzaki et al., 2012). It was found that mutation of this histidine residue to a glutamate (which disrupts the interaction through incorporation of a large, charged residue) abolishes both AVR-PikD/Pikp-HMA binding in a yeast two-hybrid assay and recognition of the effector in planta. For this reason, the AVR-PikD^{H46E} mutant has become an established negative control that is used extensively in this project. One observation of note is that the Glu230 residue from Pikp-HMA, which is important in forming a hydrogen bond with AVR-PikD^{H46}, is replaced by a valine in Pikm, which results in loss of this hydrogen bond formation. The His46 residue therefore becomes less crucial in the AVR-PikD/Pikm-HMA interaction, and when the same AVR-PikD^{H46E} mutant is tested against Pikm, an interaction in yeast two-hybrid screens is observed, and a weak cell death response is seen in planta (de la Concepcion et al., 2018).

Other structures of Pik-HMA domains in complex with effectors have been solved and shown to have a similar conformation (de la Concepcion et al., 2018). One point of note is that a new construct of Pik-HMA (both Pikp and Pikm) has been made that contains an extension of five residues at the C-terminus compared with the previously published structure of AVR-PikD/Pikp-HMA in Maqbool et al. (Maqbool et al., 2015). This construct was made for reasons of improved stability in vitro and is used throughout this project.

Finally, the work carried out thus far on the Pik-HMA domains has shown that, in solution, Pikp-HMA exists as a dimer (whether in isolation or in complex with AVR-PikD) whereas Pikm-HMA behaves as a monomer (de la Concepcion et al., 2018; Maqbool et al., 2015). The authors of these studies have proposed that the dimer formation of Pikp-HMA is simply an artefact of protein purification in vitro (see further discussion in section 3.7).

The research described so far has demonstrated that there is extensive precedent for studying these NLR pairs with integrated HMA domains. Despite containing the same type of integrated domain, the mechanism of activation appears to differ significantly for RGA5/RGA4 and Pik-1/Pik-2. A comparison of the two systems is shown in figure 3.1-3, indicating the overall global domain structures and sequence similarity. A number of tools have been produced for structural and biochemical characterisation of these proteins, and model systems have been developed for understanding in planta interactions. This project describes the overlap and cross-reactivity of the two systems and describes how the effector AVR-Pia can be bound by Pikp-HMA and produce a response in planta.

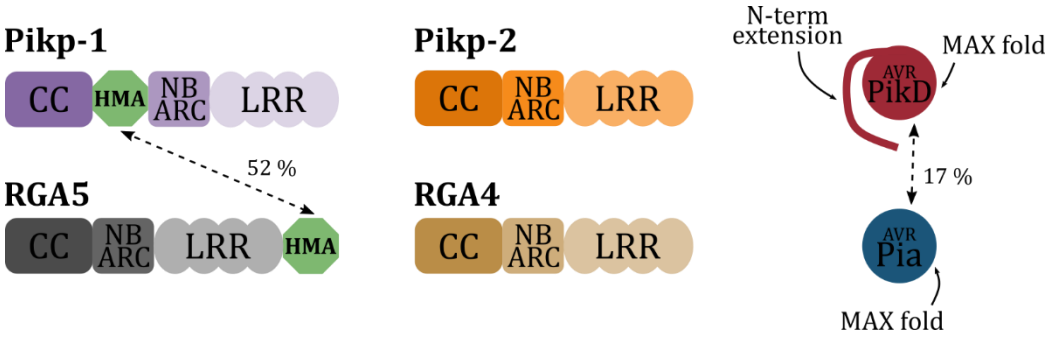


Figure 3.1-3: Comparison of the Pikp-1/Pikp-2 and RGA5/RGA4 systems.

Diagram showing a comparison between the domain architecture of Pikp-1/Pikp-2 and RGA5/RGA4. Domains are labelled individually, where CC = coiled coil, NB-ARC = nucleotide binding domain, LRR = leucine-rich repeat and HMA = heavy metal-associated. Dashed arrows show the percentage amino acid sequence similarity between the two domains or proteins indicated by the arrow. A comparison of the global structure of AVR-PikD and AVR-Pia is shown, indicating the N-terminal extension as the main structural difference between the effectors. Note that only the effector domain is indicated, the signal peptide is not considered in the structure or sequence comparison.

3.2 Pikp responds to AVR-Pia in planta

3.2.1 In rice

The trigger for this novel project was an observation made by Professor Hiromasa Saitoh, formerly at the Iwate Biotechnology Research Center, and now at the Tokyo University of Agriculture in Japan. During punch-inoculations of different rice cultivars, he observed consistent, partial resistance of cv. K60 (containing Pikp-1/Pikp-2) to *M. oryzae* Sasa2 expressing AVR-Pia (figure 3.2-1). The same effect was not present in cv. Tsuyuake (containing Pikm-1/Pikm-2), suggesting that the effect was specific for the Pikp allele. One example of the results is shown in figure 3.2-1, which also suggests that the cross-

reactivity does not occur in reverse, i.e. rice cultivars containing RGA5/RGA4 pair of NLRs (cv. Sasanishiki) do not respond to *M. oryzae* strains expressing AVR-PikD.

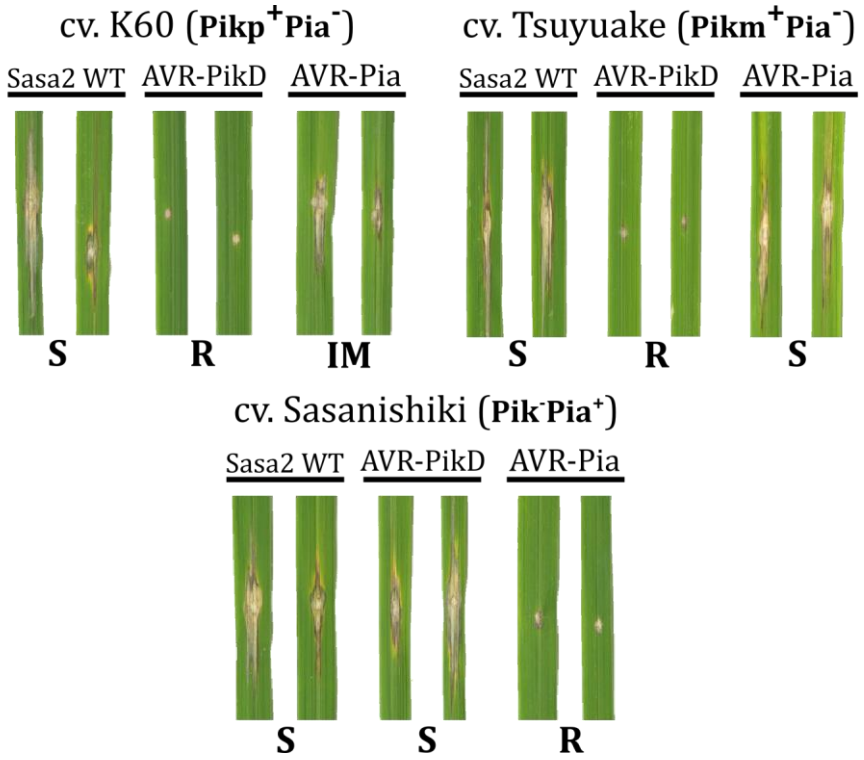


Figure 3.2-1: Pikp confers partial resistance to *M. oryzae* expressing AVR-Pia.

Images of sample results from rice inoculation assays by H. Saitoh. *M. oryzae* Sasa2 strains expressing either no effectors (WT), AVR-PikD or AVR-Pia were punch-inoculated onto rice cultivars containing either Pikp-1/Pikp-2 (cv. K60), Pikm-1/Pikm-2 (cv. Tsuyuake) or RGA5/RGA4 (cv. Sasanishiki). Leaf samples were taken and photographed 10 dpi. R = resistant phenotype, S = susceptible phenotype, IM = intermediate phenotype (appearance between resistant and susceptible).

3.2.2 In *N. benthamiana*

Following these observations in rice, I took up the project, and first investigated whether the effects could be also seen in the model system *Nicotiana benthamiana*. In the Banfield lab (and others), there is precedent for using the *N. benthamiana* system as a simple and efficient model to investigate the responses of NLRs to different AVR proteins, rather than using the host plant and pathogen. Good correlation has been shown between the rice host and *N. benthamiana* model for the Pik/AVR-Pik system (Maqbool et al., 2015). In this model system, *Agrobacterium tumefaciens* containing genes of interest are infiltrated into *N. benthamiana* leaves, leading to transient in planta expression of the genes encoded – this process is known as agroinfiltration.

Firstly, Pikp-1 and Pikp-2 were transiently expressed, along with either AVR-PikD, AVR-PikD^{H46E} or AVR-Pia. After 5 days (5dpi), there were visible signs of necrotic tissue at the infiltration site for Pikp-1/Pikp-2/AVR-PikD. When viewed under UV light (figure 3.2-2), there was auto-fluorescence seen due to the accumulation of phenolic compounds, characteristic of the hypersensitive cell death response (Kim et al., 2003). This indicates that the Pikp NLRs are able to trigger an immune response to AVR-PikD. This Pikp-1/Pikp-2 response to AVR-PikD has been observed numerous times (Maqbool et al., 2015) and acts as the positive control for this, and other, experiments in the project. Conversely, the AVR-PikD^{H46E} mutant acts as a negative control, as the Pik NLRs are unable to respond to this mutant (as detailed in section 3.1), and therefore no HR-like cell death is visible at the infiltration site.

For the test sample (Pikp-1/Pikp-2/AVR-Pia), there appears to be a slight yellowing of the tissue at the infiltration site, and a weak fluorescence under UV light (figure 3.2-2). This indicates that there could be a weak immune response triggered by the Pikp NLRs to AVR-Pia, correlating with the observations seen in rice. This assay was repeated 70 times over three biological repeats to confirm the observation. Each repeat was then scored visually using a simple 0-6 HR index scale ((Maqbool et al., 2015) and shown in figure 3.2-2) and the scores were assembled in a box plot (figure 3.2-2). To confirm that each protein was correctly expressed within the leaf, protein was extracted from leaf tissue samples taken from the infiltration sites and Western blot analysis was used to confirm protein accumulation. Figure 3.2-2 shows that each protein was expressed, although the amount of each protein in the Pikp-1/Pikp-2/AVR-PikD sample appears lower than the others (as indicated in the Ponceau image for total loading). This is likely because the stronger immune response in this sample leads to greater cell death and hence lower total protein accumulation in these samples. This effect was consistently seen in all three biological repeats (one representative blot is shown in figure 3.2-2) and across other similar experiments. Cumulative cell death scores robustly suggest that the

Pikp NLRs can respond to AVR-Pia, although the response is considerably weaker than the response to their 'matched' effector AVR-PikD.

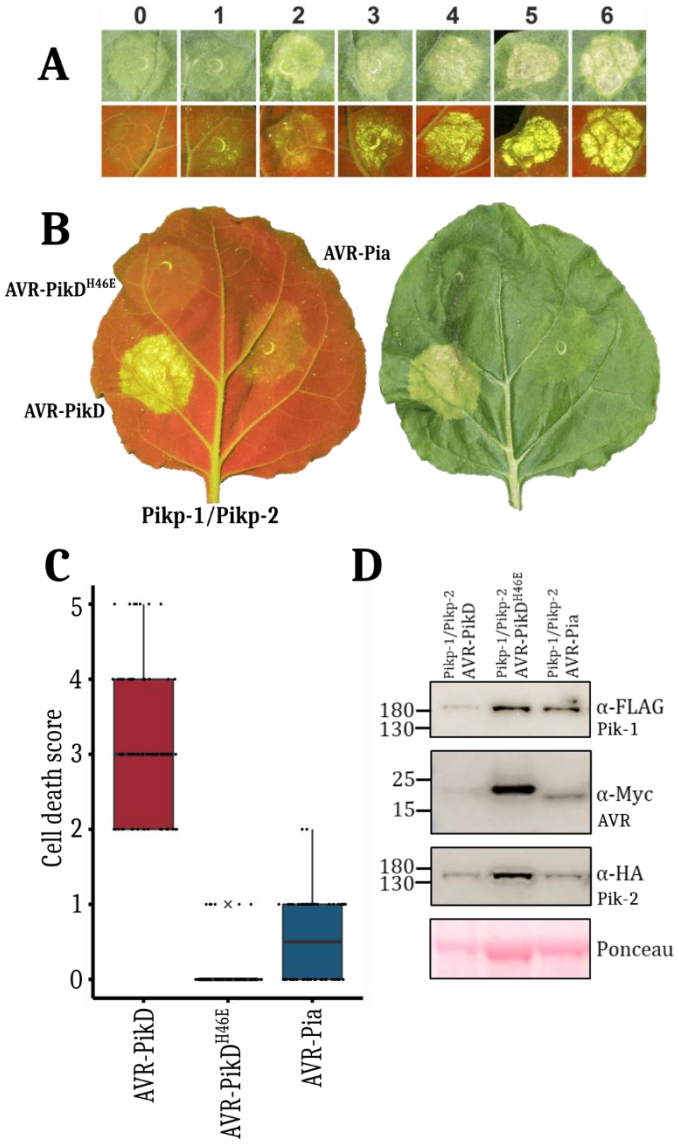


Figure 3.2-2: Pikp shows partial cell death response to AVR-Pia in the *N. benthamiana* model system.

A) Scale used for visual scoring of HR-like cell death from 0 to 6 (taken from (Maqbool et al., 2015)) used throughout this work. **B**) Representative leaf images taken under UV and daylight 5 dpi showing cell death progression at infiltration sites. **C**) Box plot showing cell death scores for 70 technical repeats over 3 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles. **D**) Western blot showing the expression of proteins at the infiltration sites. Blot shown is representative of 3 experiments showing similar results.

To further confirm that the observations made in rice correlate with those in the model *N. benthamiana* system, the assay was repeated with expression of Pikm-1 and Pikm-2 with AVR-PikD, AVR-PikD^{H46E} and AVR-Pia. Despite confirmed expression of all proteins in the leaf tissue, the Pikm NLRs were unable to respond to AVR-Pia in planta (figure 3.2-3). There is, in fact, a weak response to the AVR-PikD^{H46E} negative control, as previously observed, due to differences in the AVR-PikD His46 interface with Pikm-HMA compared with Pikp-HMA (de la Concepcion et al., 2018). Overall, this result is in agreement with the rice assays, suggesting that the weak immune response to AVR-Pia is specific for the Pikp allele.

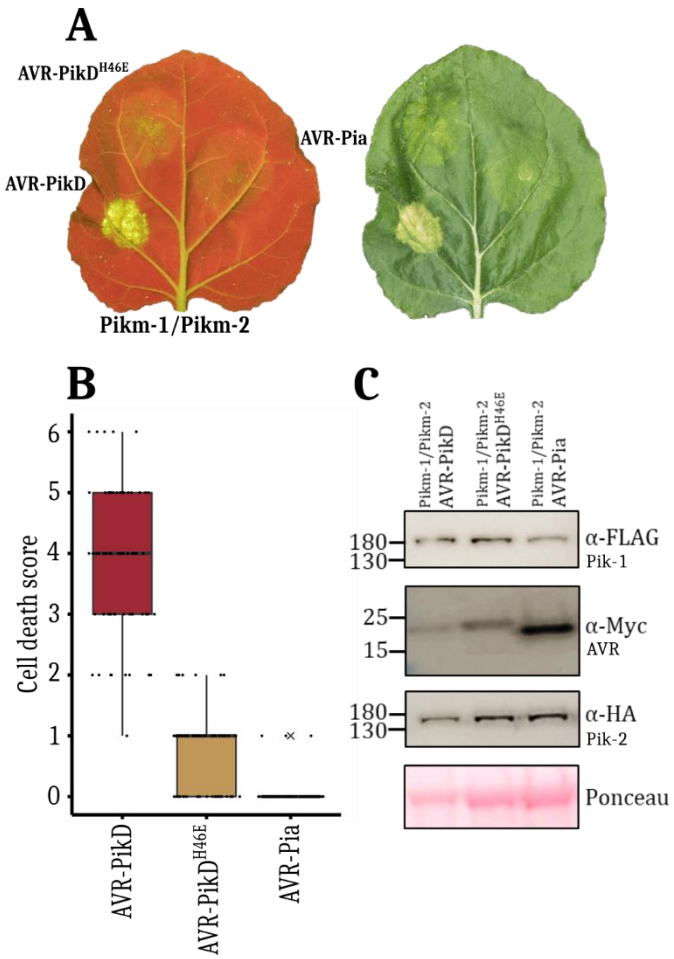


Figure 3.2-3: Pikm does not respond to AVR-Pia in the *N. benthamiana* model system.

A) Representative leaf images taken under UV and daylight 5 dpi showing cell death progression at infiltration sites. **B**) Box plot showing cell death scores for 70 technical repeats over 3 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles. **C**) Western blot showing the expression of proteins at the infiltration sites. Blot shown is representative of 3 experiments showing similar results.

3.3 Pikp can interact directly with AVR-Pia

3.3.1 in vitro

3.3.1.1 Purification of AVR-Pia

In previous biochemical and structural studies of the Pik/AVR-Pik system, the AVR-Pik effectors and the HMA domain of Pik-1 have been expressed using *E. coli* as a heterologous expression system. The isolated Pik-HMA domain is used because of well-documented issues with expressing and purifying full-length NLR proteins from *E. coli* for in vitro study (Askari et al., 2012). It has previously been shown that Pikp-HMA is able to directly interact with AVR-PikD in vitro but not with the AVR-PikC allele (de la Concepcion et al., 2018; Maqbool et al., 2015).

AVR-Pia was previously unstudied within the group (although the protein has been heterologously expressed elsewhere (e.g. (de Guillen et al., 2015)) but was readily purified from *E. coli* SHuffle cells, following methods developed for AVR-Pik proteins (Maqbool et al., 2015). It should be noted that as with other similar effectors studied in vitro, only the effector domain of AVR-Pia was used, and the expressed construct did not contain the N-terminal signal peptide, which comprises residues 1-19 of the full-length protein (Maqbool et al., 2015). The protein purification procedure involved a straightforward two-step IMAC/GF purification from cell lysate, followed by cleavage of a His-SUMO solubility tag and a final gel filtration step for purity. Untagged AVR-Pia eluted as a single monodisperse peak from the gel filtration column (figure 3.3-1); the small irregularity on the trace is due to a pause and restart step in the gel filtration stage, which is necessary for changing the fractionation plate. AVR-Pia appears to show a smearing pattern when analysed by SDS-PAGE (figure 3.3-1), even when an increased concentration of reducing agent (DTT) was added to the SDS loading dye (result not shown). The smearing results in the appearance of an apparent additional band on the gel, but no additional protein is seen at this molecular weight. This appearance was maintained throughout, whenever AVR-Pia purified from *E. coli* was analysed by SDS-PAGE. However, despite the unclear SDS-PAGE analysis, intact mass spectrometry confirmed a single, full-length, species at 7534.2 Da. The expected mass for AVR-Pia is 7536.6 Da, and the reduction of 2 Da in the actual measured mass indicates the formation of a disulphide bond, as would be expected for effectors containing the MAX fold (de Guillen et al., 2015).

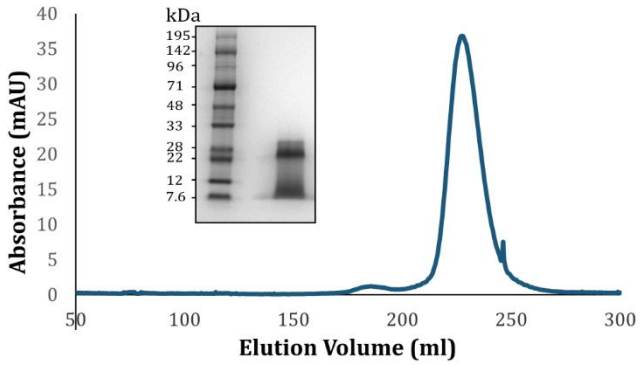


Figure 3.3-1: Purification of AVR-Pia.

Following cleavage of SUMO tag by 3C protease, AVR-Pia was flowed down a Superdex 75 26/600 gel filtration column, and showed a major peak on the trace at 228 ml elution volume. Accompanying SDS-PAGE gel shows purified AVR-Pia protein, with characteristic smear pattern.

3.3.1.2 In vitro interactions

Following successful purification of AVR-Pia, interactions with Pikp-HMA and Pikm-HMA were assessed using both a qualitative and quantitative method. Analytical gel filtration is a size exclusion chromatography technique where proteins are flowed through a resin containing pores of different sizes. Larger particles cannot enter the pores and therefore elute from the column earlier, whereas smaller particles (smaller proteins in this case) can enter the pores, so their passage through the resin is slowed. The elution volume of the proteins can be detected by their absorbance at 280 nm. If two individual proteins are mixed and are able to form a complex, their cumulative size will lead to an earlier elution from the resin – seen as a peak shift to the left on the trace. When AVR-Pia and Pikp-HMA are mixed, there is a distinct shift, as AVR-Pia alone elutes at 15.0 ml, but when mixed with Pikp-HMA, a new peak is seen at 12.3 ml (figure 3.3-2A). SDS-PAGE confirms that both proteins can be seen under this peak. Note that Pikp-HMA absorbs poorly at 280nm, but SDS-PAGE analysis of fractions suggests elution volume is around 12.8 ml. By contrast, when AVR-Pia is mixed with Pikm-HMA, no peak shift is seen on the trace and SDS-PAGE shows that the proteins are eluting separately (at 13.0 ml for Pikm-HMA and 15.4 ml for AVR-Pia) (figure 3.3-2B).

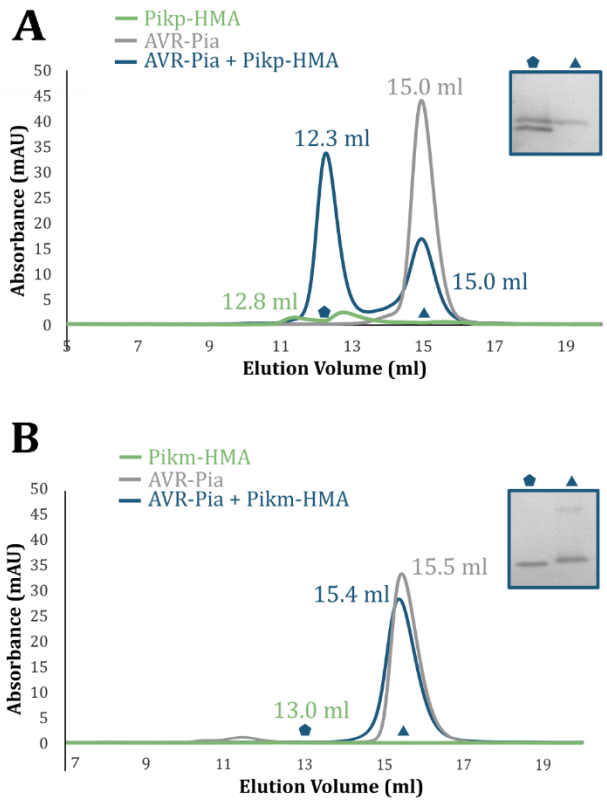


Figure 3.3-2: Qualitative binding analysis for AVR-Pia with Pikp-HMA and Pikm-HMA.

A) Analytical gel filtration traces showing the elution volume of AVR-Pia alone (grey), Pikp-HMA alone (green) and an AVR-Pia/Pikp-HMA mixture (blue). SDS-PAGE analysis shows the fractions taken where indicated on the trace. Volumes shown indicate elution volumes. **B)** Analytical gel filtration traces showing the elution volume of AVR-Pia alone (grey), Pikm-HMA alone (green) and an AVR-Pia/Pikm-HMA (blue). SDS-PAGE analysis shows the fractions taken where indicated on the trace. Volumes shown indicate elution volumes. Baselines corrected to zero.

To further investigate these binding events, we used surface plasmon resonance (SPR), a biophysical technique that measures the change in refractive index when one binding partner (the analyte, in this case the HMA protein) is flowed over a chip covered with the second immobilised binding partner (the ligand, in this case the effector protein). In order to immobilise the effector onto the chip, each effector was purified with a non-cleavable C-terminal 6xHis tag, which bound to the Ni-NTA chip surface. The output of the analyte binding to ligand is shown in a change of response units (RU), and is proportional to the molecular mass on the surface of the chip. The maximum binding capacity of the chip is known as R_{max}, and is dependent on the molecular weight (MW) of both the analyte and ligand, the level of ligand immobilisation onto the chip (R_{ligand}) and the stoichiometry of the interaction. Once the theoretical R_{max} has been calculated, the experimental RU value can be expressed as a percentage of R_{max}, which gives an indication of the strength of binding; if the interaction between the HMA and the effector

is strong, there will be a greater proportion of HMA bound to the immobilised effector.

$$R_{max} = \frac{MW\ analyte\ (HMA)}{MW\ ligand\ (AVR)} \times R_{ligand} \times Stoichiometry$$

In the work conducted here, both Pikp-HMA and Pikm-HMA have high affinities for AVR-PikD, in agreement with previously published work (de la Concepcion et al., 2018). For both HMAs, the $\%R_{max}$ value is near to 80 % at 100 nM (figure 3.3-3). For Pikp-HMA, the $\%R_{max}$ for AVR-Pia at 100 nM HMA concentration is 58 % (figure 3.3-3A), which is higher than might be expected given the apparently weak interaction suggested from the low immune response in planta. However, the in vitro system clearly differs quite considerably from the plant cell, not only in terms of environmental conditions, but also because only the isolated HMA domain is being considered in vitro, outside of its full-length NLR context. In contrast, the $\%R_{max}$ value for Pikm-HMA binding to AVR-Pia at 100 nM is 2 %. Taking into consideration that the $\%R_{max}$ for Pikm-HMA at the same concentration for AVR-PikC, which is acting as a negative control in these experiments, is 8 %, this suggests that there is no meaningful interaction between Pikm-HMA and AVR-Pia (figure 3.3-3B).

As well as displaying binding affinities as a percentage of R_{max}, SPR is commonly used to calculate the equilibrium dissociation constant (K_D) for a given interaction. This is calculated by using a greater range of analyte concentrations and fitting the responses to a steady-state affinity model. Previously published work has determined the K_D for Pikp-HMA/AVR-PikD to be 5.9 nM and for Pikm-HMA/AVR-PikD to be 4.7 nM (de la Concepcion et al., 2018). Despite repeated attempts, the K_D for the Pikp-HMA/AVR-Pia interaction could not be determined because the data failed quality tests set by the Biacore software. This could be due to a lower binding affinity for Pikp-HMA/AVR-Pia,

and may require a change of experimental conditions, such as a change of buffer or flow rate, to enable accurate values to be obtained.

The overall conclusion from these in vitro experiments is that the interactions between the Pik HMAs and AVR-Pia in vitro correlate well with the response of full-length Pik NLRs to AVR-Pia in planta.

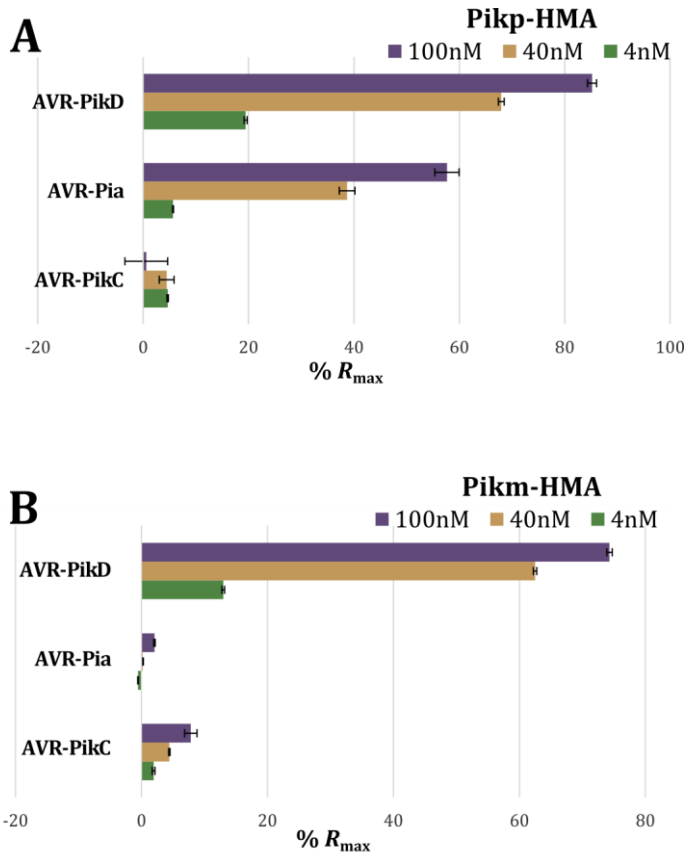


Figure 3.3-3: Quantitative binding analysis for AVR-Pia with Pikp-HMA and Pikm-HMA.

Surface plasmon resonance bar charts showing %R_{max}, the percentage of theoretical maximum response for HMA binding to immobilised effector, at three different concentrations of HMA protein. Each measurement represents an average of three results, with error bars showing standard deviation. Chart shown is representative of three separate experiments. **A**) Binding of Pikp-HMA to AVR-PikD (positive control), AVR-PikC (negative control) and AVR-Pia. **B**) Binding of Pikm-HMA to AVR-PikD (positive control), AVR-PikC (negative control) and AVR-Pia.

3.3.2 in planta

Figure 3.2-2 shows that there was an increased amount of HR-like cell death in leaves that had been agroinfiltrated with Pikp-1/Pikp-2/AVR-Pia compared to the negative control Pikp-1/Pikp-2/AVR-PikD^{H46E}. However, it was not demonstrated that this increased cell death was due to a direct interaction in planta between the NLRs and the

effector. Therefore, co-immunoprecipitation (co-IP) experiments were carried out to investigate direct binding in the plant tissue. FLAG antibody was used to immunoprecipitate Pik-1 NLRs from plant extract (tagged with a C-terminal 6xHis/3xFLAG tag) and it was determined whether Pik-1 had been able to co-purify with the effector in each case (tagged with an N-terminal 4xMyc tag). It should be noted that Pik-2 was not agroinfiltrated for the co-immunoprecipitation experiments, as its presence would have resulted in cell death at the infiltration site, impeding the successful extraction of protein. While Pik-1 was clearly able to pull down AVR-PikD, its ability to bind directly with AVR-Pia was less clear. For three separate repeats of the co-IP experiment (biological repeats using different leaf tissue samples), it was shown that in 2 out of 3 cases Pikp-1 did not co-purify with AVR-Pia, as shown in figure 3.3-4.

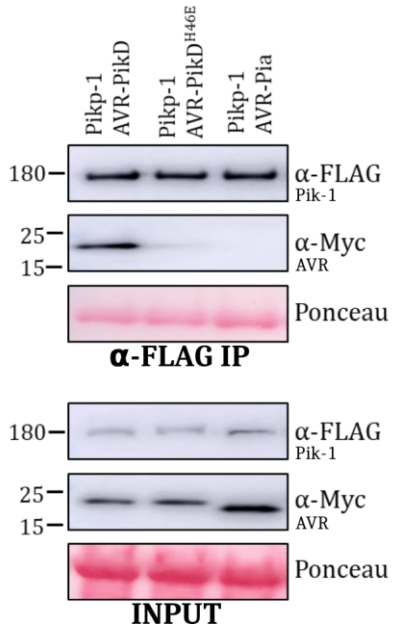


Figure 3.3-4: Co-immunoprecipitation assay investigating Pik-1 interaction with AVR-Pia in planta.

Western blot showing protein input from plant extract (bottom panel) and the outcome of immunoprecipitation of Pikp-1 with FLAG magnetic beads (top panel), visualised with α-FLAG and α-Myc antibodies. Protein loading is shown by Ponceau stain. This blot is representative of two out of three biological repeats.

However, in one experimental repeat, it appeared that AVR-Pia did co-immunoprecipitate with Pikp-1, due to the presence of a faint band in the AVR-Pia sample when the eluate was visualised using the α-Myc antibody (data not shown). In both situations, the positive (AVR-PikD) and negative (AVR-PikD^{H46E}) controls behaved as expected. This discrepancy could be a symptom of a weak interaction between Pikp-1 and AVR-Pia or could be due to a difference in the leaf tissue itself. To resolve this, more repeats of the co-IP experiment would need to be carried out to reach a consensus. From the data obtained thus far, no proof of a direct interaction between AVR-Pia and full-length Pikp-1 in planta has been found, but it appears that if any interaction did occur it would be very weak, as expected from the weak HR-like cell death. As discussed above, the decision was taken to not include Pik-2 in the co-IP experiments, to avoid cell death

and loss of protein. However, this does preclude any Pik-2-dependent interactions from taking place. For example, if the Pik-1/effector interaction was dependent on Pik-2, this would be excluded from the results. However, precedent within the lab suggests that this is unlikely, as previous co-IP experiments with the Pik proteins have shown an NLR/effector interaction in the absence of Pik-2.

3.4 Pikp-HMA binds AVR-Pia at an unusual interface

Following confirmation that AVR-Pia can bind Pikp-HMA in vitro, the next step was attempting to understand and visualise the binding interface between the two proteins. Solving the structure of the two proteins in complex would provide information about the similarities and differences between Pikp-HMA binding AVR-PikD and AVR-Pia. In addition, this information could be used to make informed mutations within Pikp-HMA to generate a stronger response to AVR-Pia. Previous structural characterisation of Pik-HMA and AVR-Pik complexes has been carried out by co-transformation of *E. coli* SHuffle cells with an MBP-tagged Pik-HMA construct and an untagged AVR-Pik construct. Both partners are co-expressed and the tagged HMA domain pulls down its interacting effector during IMAC purification of cell lysate. The MBP tag is then cleaved from the HMA domain and a second gel filtration step yields purified complex for crystallography (Maqbool et al., 2015). When this method was attempted for the AVR-Pia/Pikp-HMA complex, the yield from purification was low, and the protein was prone to precipitation. One hypothesis is that the weak interaction between AVR-Pia and Pikp-HMA makes complex formation difficult in the presence of the bulky MBP tag on Pikp-HMA, leading to excess unbound HMA in the final sample, which is prone to precipitation upon concentration. Instead, a new purification method was developed to overcome these issues. Both AVR-Pia and Pikp-HMA were purified separately (with SUMO and MBP solubility tags respectively) and immediately following cleavage and removal of the tags from both partners, the proteins were mixed together and purified for a final time using preparative gel filtration. This new technique allows easier complex formation (using untagged proteins) but still allows removal of any un-complexed protein (via the final gel filtration step) – causing increased yields and reduced protein precipitation.

Following successful purification of AVR-Pia/Pikp-HMA, the complex was used for crystallisation trials. Three different commercially available crystallisation screens were used; the PEGs suite (Qiagen), Morpheus® and JCSG-*plus*™ (both Molecular Dimensions). Two different protein concentrations were used, of approximately 10 and 20 mg/ml, although accurate protein concentration was measured before each screen was set up.

Best quality crystals were obtained in the Morpheus® screen, using protein at a measured concentration of 18 mg/ml. The crystals were found in well D2 of the screen, and the conditions in this well were: 0.12 M Alcohols (0.2 M 1,6-Hexanediol; 0.2 M 1-Butanol; 0.2 M 1,2-Propanediol; 0.2 M 2-Propanol; 0.2 M 1,4-Butanediol; 0.2 M 1,3-Propanediol), 0.1 M Buffer System 1 (1.0 M imidazole; MES monohydrate (acid), pH 6.5) and 50 % v/v Precipitant Mix 2 (40 % v/v Ethylene glycol; 20 % w/v PEG 8000). Images of the crystals used for data collection are shown in figure 3.4-1.

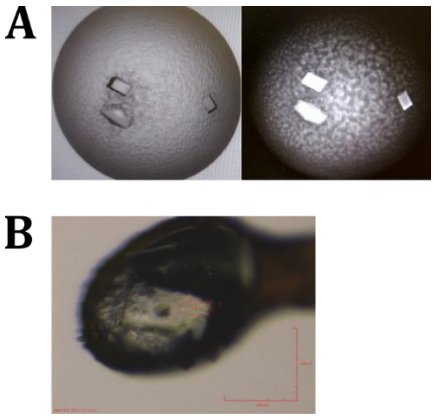


Figure 3.4-1: Images of protein crystals used for collecting AVR-Pia/Pikp-HMA structural data.

A) Images of protein crystals taken under visible (left) and UV light (right) by the Minstrel Crystallisation Imager (Rigaku). Images were taken 5 days after crystallisation trays were set up. **B**) Photo of a single protein crystal mounted on a loop prior to data collection at the Diamond Light Source.

X-ray diffraction data was collected at the Diamond Light Source synchrotron in Oxfordshire on beamline DLS-i03. Crystals diffracted to 1.9 Å resolution. The structure was solved by molecular replacement using PDB files 5A6P (Pikp-HMA, (Maqbool et al., 2015)) and 2MYW (AVR-Pia, (de Guillen et al., 2015)), followed by rounds of manual rebuilding and refinement. Due to the use of both members of the complex themselves as models for the molecular replacement, there were few residues that required significant changes during refinement. The loop containing the non-functional metal binding site between β-1 and α-1 in Pikp-HMA was found to be disordered as described previously (Maqbool et al., 2015) so was not included during refinement. Similarly, the two-residue scar at the N-terminus of Pikp-HMA ('G-P…') deriving from the 3C protease cleavage of the solubility tag was disordered. Additionally, the three residues at the C-terminus of Pikp-HMA were disordered in this structure. 89 water molecules were positioned during refinement. Table 9 shows statistics for the data collection and

refinement, and further details about the programs used for structure refinement and building can be found in Chapter 2 (Materials and Methods).

Each partner in the complex adopts a similar overall shape to the previously solved structures. Pikp-HMA (de la Concepcion et al., 2018; Maqbool et al., 2015) comprises two adjacent α-helices opposite a four-stranded β-sheet. Currently available structures of AVR-Pia (de Guillen et al., 2015; Ose et al., 2015) were both solved in solution using NMR spectroscopy, and this represents the first AVR-Pia structure solved by X-ray crystallography. AVR-Pia has previously been shown to contain the six-stranded β-sandwich that is characteristic of MAX effectors (de Guillen et al., 2015), but in the structure described here (figure 3.4-2) β-5 is not well-defined and appears as a loop joining β-4 and β-6. Within this loop, the overall shape appears similar to the effector structure determined by NMR spectroscopy (de Guillen et al., 2015), and residues are positioned in a largely similar way. One noticeable difference is a slight alteration in the hydrogen bonding, which may contribute to the lack of defined β-strand in the X-ray crystallography structure. Both structures contain a hydrogen bond between the oxygen atom in the backbone of Leu70 and the backbone nitrogen of Gly73. However, while the published NMR structure also contains a hydrogen bond between the backbone nitrogen of Leu70 and oxygen of Lys74, in the structure solved here, this Leu70 hydrogen bond is formed with a water molecule, rather than the amino acid chain. However, despite this region not being fully defined as a β-strand, it still forms the same conserved structure as seen elsewhere. As expected from previous structural characterisation, the effector is stabilised by a disulphide bond between Cys25 and Cys66.

Despite the two proteins in the complex adopting essentially identical folds to their published structures (in isolation), there are striking differences between the AVR-Pia/Pikp-HMA complex and the complexes of AVR-Pik effectors with Pikp-HMA. Figure 3.4-2A shows global representations of AVR-Pia/Pikp-HMA and AVR-PikD/Pikp-HMA (PDB 6G10, (de la Concepcion et al., 2018)), with the monomer of Pikp-HMA displayed in the same orientation for both structures. Broadly, Pikp-HMA binds AVR-PikD opposite its β-sheet face, whereas AVR-Pia is bound adjacent to α-1 (Asn201–Ser212) and β-2 (Val216–Val212), creating a very different interface. Figure 3.4-2B shows this interaction from an alternate viewpoint. Here, it is possible to see how the β-sheet of Pikp-HMA aligns itself with β-3 of AVR-PikD, and how the N-terminal arm of AVR-PikD wraps around the other side of the HMA β-sheet. On the other hand, AVR-Pia is positioned alongside Pikp-HMA α-1 and β-2. In both cases, the positioning of the effector relative to Pikp-HMA allows the formation of a continuous anti-parallel β-sheet. In the

case of AVR-PikD, the β-strands from Pikp-HMA form a sheet with β-strands 3-5 of AVR-PikD. For AVR-Pia, the β-strands involved are 1,2 and 6. Another striking feature is that while Pikp-HMA is present as a dimer in the crystal structure with AVR-PikD (as discussed in section 3.1), it is only a monomer in the structure with AVR-Pia. Indeed, AVR-Pia occupies the same binding surface as the Pikp-HMA dimer in the AVR-PikD/Pikp-HMA structure, which explains the lack of Pikp-HMA dimerisation in this case.

PISA analysis (Krissinel, 2015) indicates that the interface area for the AVR-Pia/Pikp-HMA structure is just 460 $Å^2$, compared to the much larger 986 $Å^2$ for the AVR-PikD/Pikp-HMA interaction (de la Concepcion et al., 2018). Additionally, the interface between AVR-Pia and Pikp-HMA is stabilised mainly by hydrogen bonds between the peptide backbone. The main contributors to these interface hydrogen bonds are Val219, Tyr41, Asp217 and Arg43 (figure 3.4-2C). The backbone oxygen from Leu38 in AVR-Pia also forms a hydrogen bond with the side chain of Arg226 in Pikp-HMA. There are fewer interactions between amino acid side chains – a hydrogen bond/salt bridge interaction is formed between the side chains of Arg43 (AVR-Pia) and Asp217 (Pikp-HMA), and the hydroxyl group on the C-terminal residue of AVR-Pia, Tyr85, also forms a hydrogen bond with Ser212 (figure 3.4-2C). An indirect interaction, mediated by a water molecule, is also seen between the side chains of Tyr41 and Ser204 (figure 3.4-2C). These weak intermolecular interactions and small interface area appear to provide an explanation for the weaker binding affinity seen in vitro for AVR-Pia with Pikp-HMA compared to AVR-PikD with Pikp-HMA (section 3.3).

Table 9: Data collection and refinement statistics for the AVR-Pia/Pikp-HMA complex structure.

*The highest resolution shell is shown in parenthesis.
**As calculated by MolProbity
R_{merge} is calculated as described in (Evans, 2006).

Data collection statistics	
Wavelength (Å)	0.9763
Space group	$P22_12_1$
Cell dimensions	
a, *b*, *c* (Å)	34.84, 53.44, 117.81
α, β, γ (◦)	90.00, 90.00, 90.00
Resolution (Å)*	48.67-1.90 (1.94-1.90)
R_{merge} **(%)**	5.7 (122.9)
$I/\sigma I$	19.7 (2.4)
Completeness (%)	
Overall	100 (100)
Anomalous	100 (100)
Unique reflections	18107 (1151)
Redundancy	
Overall	12.6 (13.3)
Anomalous	6.8 (7.0)
CC(1/2) (%)	99.9 (80.9)
Refinement and model statistics	
Resolution (Å)	48.72-1.90 (1.95-1.90)
R_{work}/R_{free} **(%)**	20.3/24.5 (35.8/41.8)
No. atoms	
Protein	2113
Water	89
B-factors	
Protein	54.1
Water	58.1
R.m.s deviations	
Bond lengths (Å)	0.0117
Bond angles (º)	1.501
Ramachandran plot (%)**	
Favoured	98.5
Allowed	1.5
Outliers	0
MolProbity Score	1.52 (95^{th} percentile)

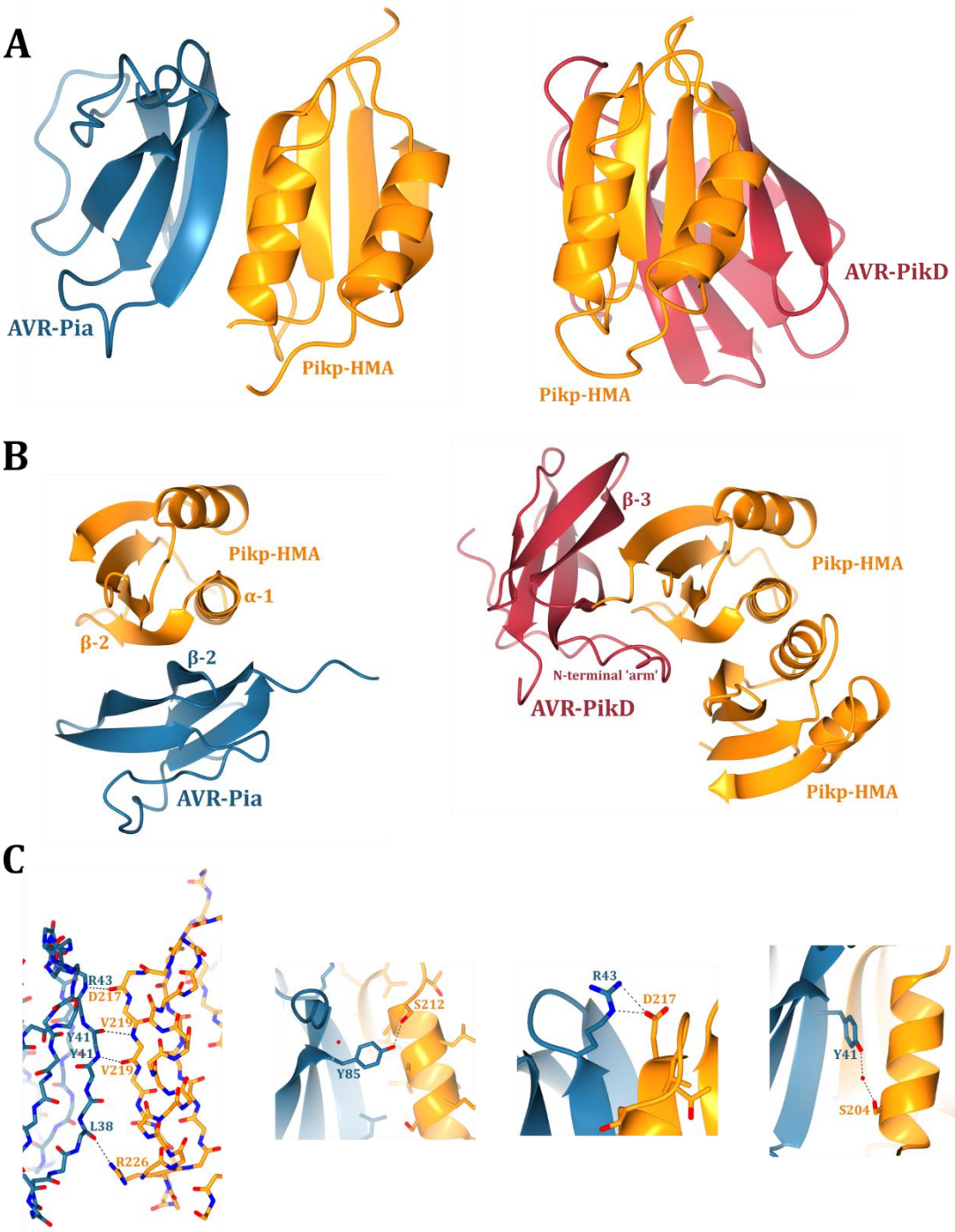


Figure 3.4-2: Structure of AVR-Pia in complex with Pikp-HMA.

A) Structure of AVR-Pia in complex with Pikp-HMA refined to 1.9 Å resolution by X-ray crystallography, compared to the structure of AVR-PikD in complex with Pikp-HMA (de la Concepcion et al., 2018). AVR-Pia is shown in blue, AVR-PikD in red and Pikp-HMA in gold. The Pikp-HMA monomer is shown in the same orientation for both structures. **B**) An alternative view of the AVR-Pia/Pikp-HMA and AVR-PikD/Pikp-HMA structures shown in **A**, with secondary structure features labelled. Pikp-HMA monomer shown in the same orientation for both structures. **C**) Details of the interface between AVR-Pia and Pikp-HMA, showing interactions at the peptide backbone, and side-chain interactions. Dotted lines show hydrogen bonds, red spheres represent water molecules. Carbons are coloured according to the chain (AVR-Pia in blue, Pikp-HMA in gold) with oxygen atoms shown in red and nitrogen in dark blue. Labels show the single letter amino acid code with position in the peptide chain.

3.5 Addition of the N-terminal 'arm'

Pikp-HMA is able to bind AVR-Pia only weakly in vitro (section 3.3.1), and it was hypothesised that this binding might be strengthened if AVR-Pia more closely resembled the 'matched' binding partner to Pikp, AVR-PikD. Although these effectors share only 17 % sequence identity, their structures have shown that they share a core six-stranded β-sandwich fold, which is known as the MAX fold (de Guillen et al., 2015). However, one noticeable difference between these two effectors is the presence of an N-terminal extension on AVR-PikD (comprising Arg31 to Pro52), that partially wraps around, and is held in place by, the core structure (Maqbool et al., 2015). This extension, or 'arm', plays an important part in the interaction of AVR-PikD and Pikp-HMA, as it includes the His46 residue that is known to form hydrogen bonds/salt bridge interactions with Ser218 and Glu230 in Pikp-HMA ((Maqbool et al., 2015) and see section 3.1).

Prior to solving the crystal structure of AVR-Pia in complex with Pikp-HMA, it was envisaged that the HMA would bind AVR-Pia in a similar orientation to AVR-PikD. Under this assumption, it was considered that genetically grafting the N-terminal extension of AVR-PikD onto AVR-Pia, might recapitulate some of the binding strength conveyed onto AVR-PikD by this arm. Conversely, would removing the arm from AVR-PikD weaken or disrupt the binding of Pikp-HMA?

Initially, this theory was tested by Hiromasa Saitoh using *Magnaporthe oryzae* Sasa2 expressing the chimera AVR-Pia$^{NAVR\text{-}PikD}$ and the rice cv. K60 containing Pikp NLRs (figure 3.5-1). His preliminary results suggest that the Pikp NLRs respond more strongly to the AVR-Pia$^{NAVR\text{-}PikD}$ chimera than they do to AVR-Pia alone, although it appeared that the presence of the arm on AVR-Pia decreased the ability of the RGA5/RGA4 NLRs to trigger an immune response, suggesting that the extra extension on the effector might be hindering recognition.

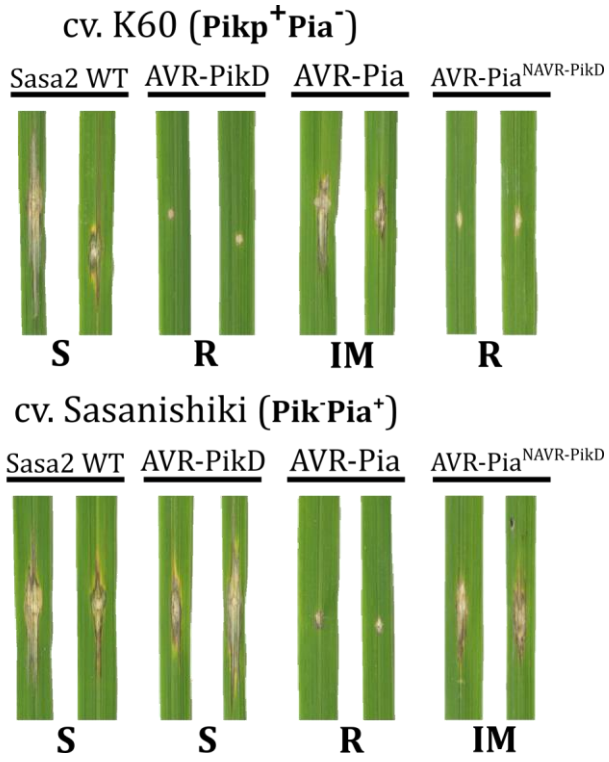


Figure 3.5-1: Addition of N-terminal extension enhances Pikp response to AVR-Pia in rice.

Images of sample results from rice inoculation assays by H. Saitoh. *M. oryzae* Sasa2 strains expressing either no effectors (WT), AVR-PikD, AVR-Pia or AVR-Pia$^{NAVR\text{-}PikD}$ were punch-inoculated onto rice cultivars containing either Pikp1/Pikp2 (cv. K60) or RGA5/RGA4 (cv. Sasanishiki). Leaf samples were taken and photographed 10 dpi. R = resistant phenotype, S = susceptible phenotype, IM = intermediate phenotype (appearance between resistant and susceptible).

This provisional data was taken forward using our model *N. benthamiana* and in vitro systems. The chimeric protein AVR-Pia$^{NAVR\text{-}PikD}$ was made by Hiromasa Saitoh in the pOPIN vector suite for *E. coli* expression and I then generated the chimera in the Golden Gate vector system for in planta experiments.

Initially, AVR-Pia$^{NAVR\text{-}PikD}$ was tested for stability in vitro. The protein was expressed with a cleavable SUMO tag in *E. coli* SHuffle cells, according to the normal protocol for AVR-Pik effectors. Figure 3.5-2 shows a purification gel filtration trace for AVR-Pia$^{NAVR\text{-}PikD}$ along with SDS-PAGE analysis of the purified protein. The protein exhibits the same smearing gel pattern as seen for AVR-Pia, which is also included in figure 3.5-2 for comparison. Despite the unusual appearance when analysed by SDS-PAGE, AVR-Pia$^{NAVR\text{-}PikD}$ elutes from the gel filtration column as a single, symmetrical peak, indicating that it is monodisperse and stable during purification. Intact mass spectrometry analysis showed a single peak at 10733.2 Da, which (as for AVR-Pia) is 2 Da lower than the expected mass of 10735.2 Da. This indicates that even with the addition of the N-terminal

extension, the AVR-Pia portion of the chimera is likely still able to form the disulphide bond to stabilise the core structure.

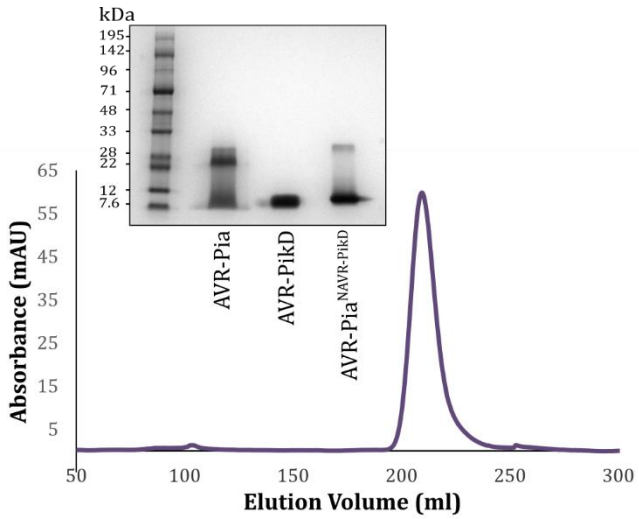


Figure 3.5-2: Purification of AVR-Pia$^{NAVR-PikD}$.

Following cleavage of SUMO tag by 3C protease, AVR-Pia$^{NAVR-PikD}$ was flowed down a Superdex 75 26/600 gel filtration column, and showed a major peak on the trace at 209 ml elution volume. Accompanying SDS-PAGE gel shows purified protein, with the same characteristic smear pattern as for AVR-Pia. Purified AVR-PikD and AVR-Pia are shown alongside for comparison.

In order to investigate the response of Pikp to AVR-PikD without the influence of the effector's N-terminal extension, truncated versions of the protein were cloned. Firstly, AVR-PikD$^{\Delta 22-52}$ was generated, whereby the entire N-terminal arm was removed, leaving only the core structure residues remaining. However, upon further consideration, two additional (less severe) truncations were made. This was in light of the fact that a cysteine residue at position 54 is involved in creating a disulphide bond with Cys70 that stabilises the core effector structure (Maqbool et al., 2015). It is possible that trimming the protein so close to Cys54 may affect its ability to correctly form the disulphide bond, so AVR-PikD$^{\Delta 22-47}$ and AVR-PikD$^{\Delta 22-44}$ were generated, that still remove a significant portion of the extension, but do not trim so close to the core structure. Each of these three truncations were cloned for expression from *E. coli* and purified as normal for AVR-Pik effectors. It soon became apparent that removal of the N-terminal extension of AVR-PikD destabilised the protein as, despite expression of each construct, the final yields were very low (as seen by very faint protein bands in SDS-PAGE analysis), and it was clear that the versions with more of the arm residues intact were more stable (figure 3.5-3). It is noteworthy that the most stable of the truncations, AVR-PikD $^{\Delta 22-44}$, is also the only version that leaves Asp45 intact, a residue that is known to help anchor the arm to the core structure via a salt-bridge interaction with Arg110 (Maqbool et al., 2015). Given the

destabilisation caused by truncating AVR-PikD, it was decided not to pursue this investigation any further in vitro, as working with unstable proteins can cause results to be less reliable. However, it is possible that in a plant cell environment, these instability issues would be lessened, so AVR-PikD$^{\Delta 22\text{-}52}$ was also cloned for in planta transient expression.

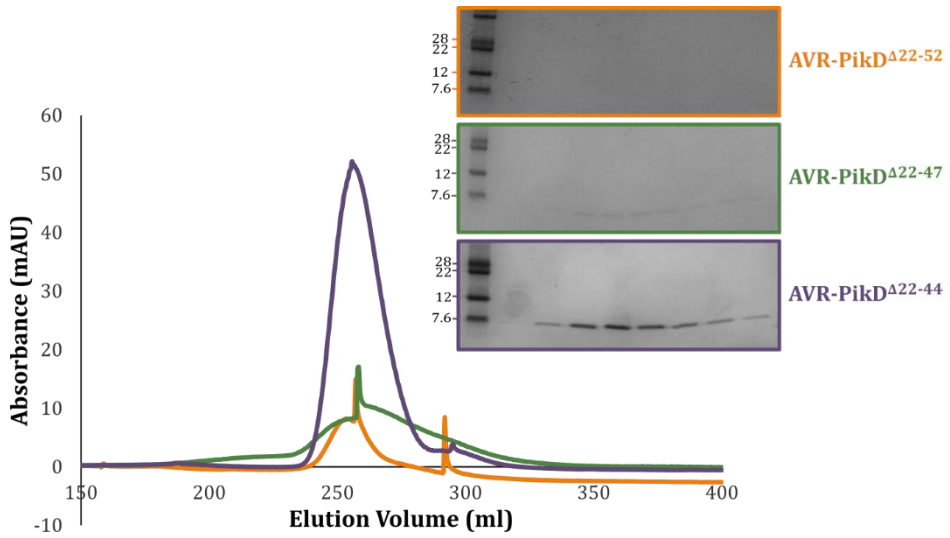


Figure 3.5-3: Purification issues for truncated versions of AVR-PikD.

Overlaid traces of truncated AVR-PikD variants, flowed down a Superdex 75 26/600 gel filtration column following cleavage of purification tag, shown with SDS-PAGE analysis of the fractions from under the peak in each case. Truncated variants are AVR-PikD$^{\Delta 22\text{-}52}$ (orange), AVR-PikD$^{\Delta 22\text{-}47}$ (green) and AVR-PikD$^{\Delta 22\text{-}44}$ (purple).

Having generated the appropriate tools, i.e. AVR-Pia$^{\text{NAVR-PikD}}$ and AVR-PikD$^{\Delta 22\text{-}52}$, the constructs were tested in the model *N. benthamiana* system for eliciting an HR-like cell death response by both Pikp and Pikm NLRs. Figure 3.5-4B and figure 3.5-5B show that neither Pikp nor Pikm are able to respond to AVR-Pia$^{\text{NAVR-PikD}}$ or AVR-PikD$^{\Delta 22\text{-}52}$, as cell death is not observed in either of these cases. Expression testing of infiltrated leaf tissue (figure 3.5-4C and 3.5-5C) indicates that AVR-Pia$^{\text{NAVR-PikD}}$ is expressed and stable in the leaf tissue, suggesting that the lack of cell death response is not due to insufficient protein production. It is possible that the addition of the N-terminal extension has disrupted the binding of Pikp-HMA to AVR-Pia, either to the extent that it cannot bind the effector at all, or that it cannot bind in the correct orientation to trigger an immune response. In the case of Pikm, it appears that the addition of the N-terminal extension onto AVR-Pia has not enabled the NLR to bind and respond to the effector. For AVR-PikD$^{\Delta 22\text{-}52}$, the accumulation of protein in the infiltration site appears very low (figure 3.5-4C and 3.5-5C), suggesting that the instability and expression issues encountered in vitro are also

present in planta. For this reason, the truncated versions of AVR-PikD were not investigated any further.

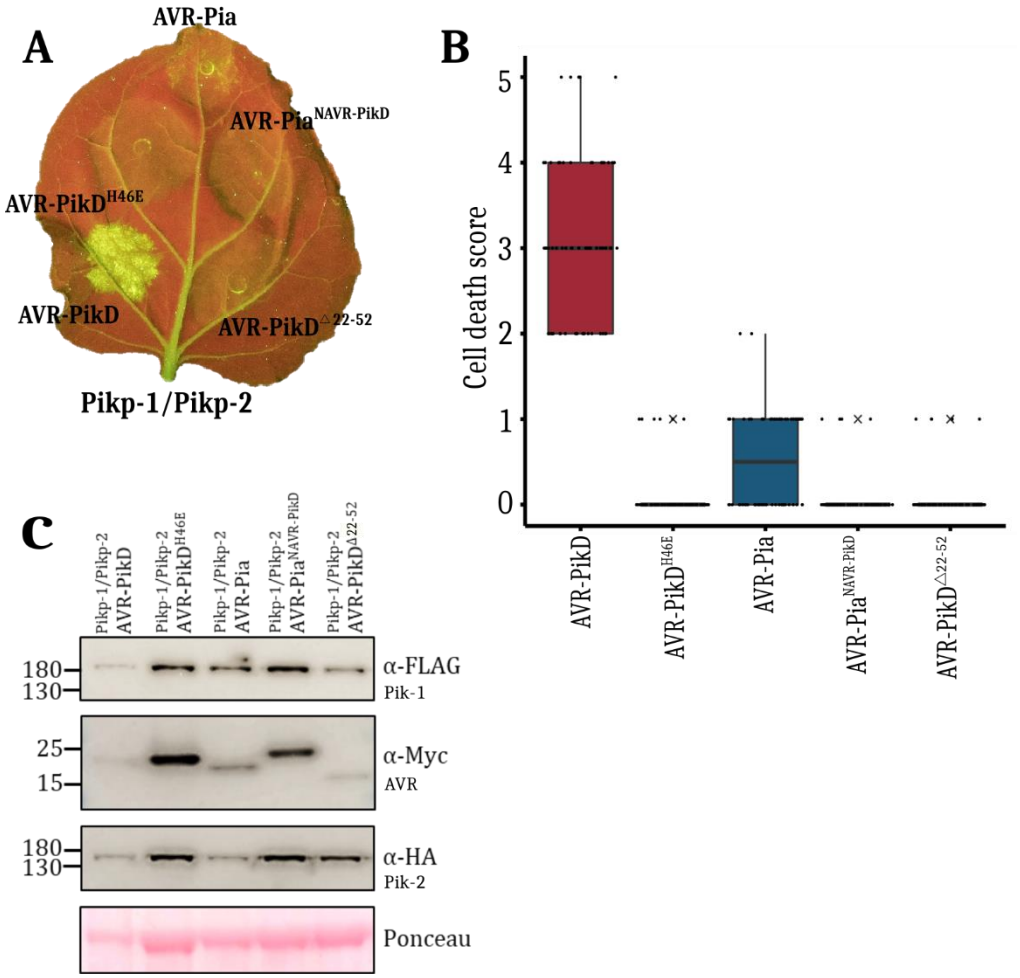


Figure 3.5-4: Pikp does not respond to AVR-Pia$^{NAVR\text{-}PikD}$ or AVR-PikD$^{\Delta 22\text{-}52}$ in the *N. benthamiana* model system.

A) Representative leaf images taken under UV light 5 dpi showing cell death progression at infiltration sites. **B**) Box plot showing cell death scores for 70 technical repeats over 3 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles. **C**) Western blot showing the expression of proteins at the infiltration sites. Blot shown is representative of 3 experiments showing similar results. Note: For **B**) and **C**) the data for AVR-PikD, AVR-PikD^{H46E} and AVR-Pia (also shown in figure 3.2-2) is shown alongside for comparison.

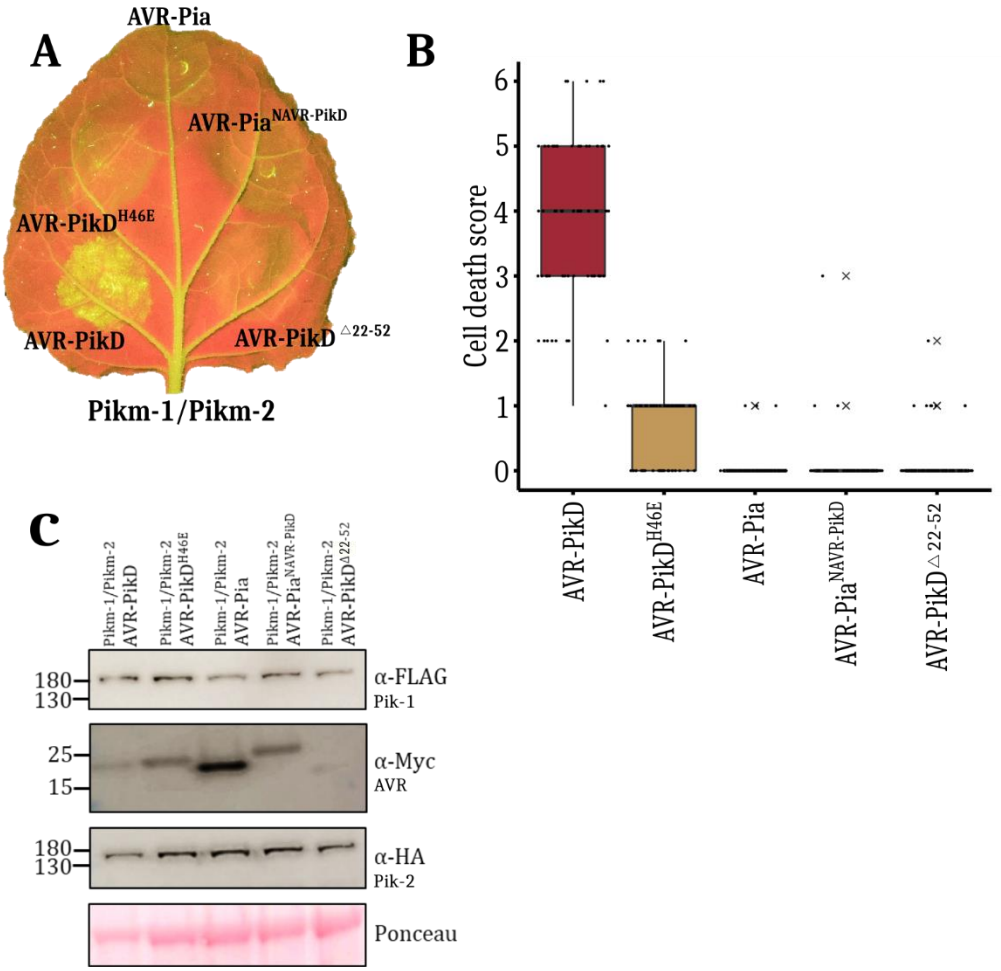


Figure 3.5-5: Pikm does not respond to AVR-Pia$^{NAVR\text{-}PikD}$ or AVR-PikD$^{\Delta 22\text{-}52}$ in the *N. benthamiana* model system.

A) Representative leaf images taken under UV light 5 dpi showing cell death progression at infiltration sites. **B**) Box plot showing cell death scores for 70 technical repeats over 3 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles. **C**) Western blot showing the expression of proteins at the infiltration sites. Blot shown is representative of 3 experiments showing similar results. Note: For **B**) and **C**) the data for AVR-PikD, AVR-PikD^{H46E} and AVR-Pia (also shown in figure 3.2-3) is shown alongside for comparison.

To further explore the ability of Pikp-HMA and Pikm-HMA to bind to the chimera AVR-Pia$^{NAVR\text{-}PikD}$, we used in vitro techniques as described in section 3.3. It was found that Pikp-HMA could bind the chimera in analytical gel filtration studies (figure 3.5-6A), evidenced by a peak shift on the trace when the two proteins were mixed. In SPR (figure 3.5-7A) it was found that Pikp-HMA could bind AVR-Pia$^{NAVR\text{-}PikD}$, and the %R_{max} for this interaction was very similar to that of AVR-Pia (also shown in figure 3.5-7A for comparison); 60 % at 100nM HMA concentration.

The fact that both effectors are binding to Pikp-HMA with similar affinity implies that the N-terminal extension has not had any impact on the strength of binding. When investigating Pikm-HMA, the results were more surprising. Initially, there was a peak shift seen from 13.2 ml to 12.4 ml in the analytical gel filtration trace when Pikm-HMA was mixed with AVR-Pia$^{NAVR\text{-}PikD}$ (figure 3.5-6B). This result suggested that the presence of the N-terminal extension has caused gain-of-binding for Pikm-HMA, as Pikm-HMA cannot bind to AVR-Pia (see figure 3.3-3). However, the SPR experiments indicated that the $\%R_{max}$ for the AVR-Pia$^{NAVR\text{-}PikD}$/Pikm-HMA interaction was very low (figure 3.5-7B), and that binding is effectively zero when compared to the negative control AVR-PikC.

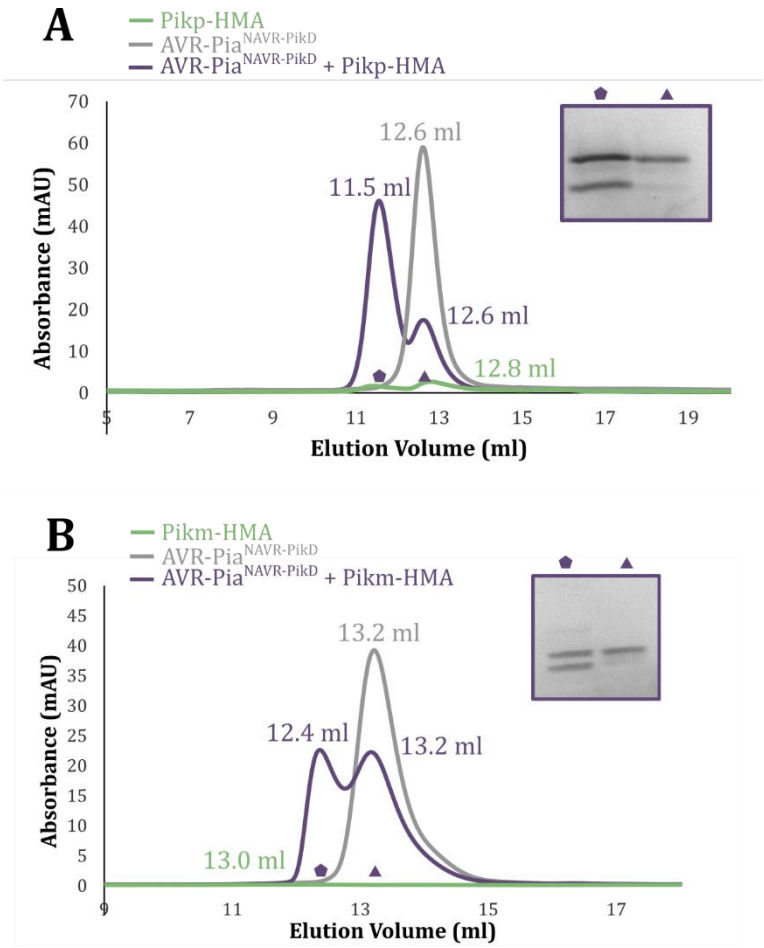


Figure 3.5-6: Qualitative binding analysis for AVR-Pia$^{NAVR\text{-}PikD}$ with Pikp-HMA and Pikm-HMA.

A) Analytical gel filtration traces showing the elution volume of AVR-Pia$^{NAVR\text{-}PikD}$ alone (grey), Pikp-HMA alone (green) and an AVR-Pia$^{NAVR\text{-}PikD}$ /Pikp-HMA mixture (blue). SDS-PAGE analysis shows the fractions taken where indicated on the trace. Volumes shown indicate elution volumes. **B**) Analytical Gel Filtration traces showing the elution volume of AVR-Pia$^{NAVR\text{-}PikD}$ alone (grey), Pikm-HMA alone (green) and an AVR-Pia$^{NAVR\text{-}PikD}$ /Pikm-HMA (blue). SDS-PAGE analysis shows the fractions taken where indicated on the trace. Volumes shown indicate elution volumes. Baselines corrected to zero.

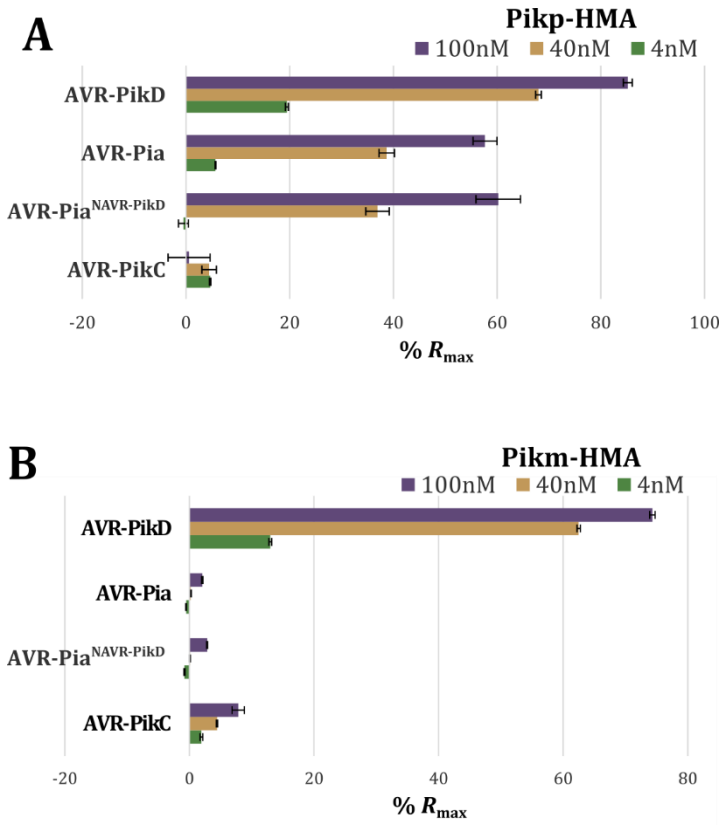


Figure 3.5-7: Quantitative binding analysis for AVR-Pia$^{NAVR\text{-}PikD}$ with Pikp-HMA and Pikm-HMA.

Surface plasmon resonance bar charts showing %R_{max}, the percentage of theoretical maximum response for HMA binding to immobilised effector, at three different concentrations of HMA protein. Each measurement represents an average of three results, with error bars showing standard deviation. Chart shown is representative of three separate experiments. **A**) Binding of Pikp-HMA to AVR-PikD (positive control), AVR-PikC (negative control), AVR-Pia and AVR-Pia$^{NAVR\text{-}PikD}$. **B**) Binding of Pikm-HMA to AVR-PikD (positive control), AVR-PikC (negative control), AVR-Pia and AVR-Pia$^{NAVR\text{-}PikD}$.

The disparity in these results could be because of the very different nature of the techniques, given that in analytical gel filtration, the proteins are mixed together and incubated at 4°C for several hours prior to assessment of complex formation, whereas in SPR, one partner is flowed over the top of the other, in a much more fleeting interaction. Another factor to take into consideration is that in SPR, the effector proteins are tagged with an additional C-terminal 6xHis tag for immobilisation on the chip, whereas this is not present for the proteins being tested by analytical gel filtration. Finally, the buffer conditions for the two experiments differ, as the buffer for SPR contains a high concentration of NaCl (860 mM) to prevent non-specific interactions between the HMA proteins and the chip reference cell. To take these factors into consideration, the analytical gel filtration experiment shown in figure 3.5-6B was repeated, comparing AVR-Pia$^{NAVR\text{-}PikD}$ with and without the C-terminal 6xHis tag, in a running buffer containing 860 mM NaCl (figure 3.5-8). The interaction between Pikm-HMA and AVR-Pia$^{NAVR\text{-}PikD}$

was still maintained under these conditions, suggesting that the difference in results is due to an innate difference between the techniques.

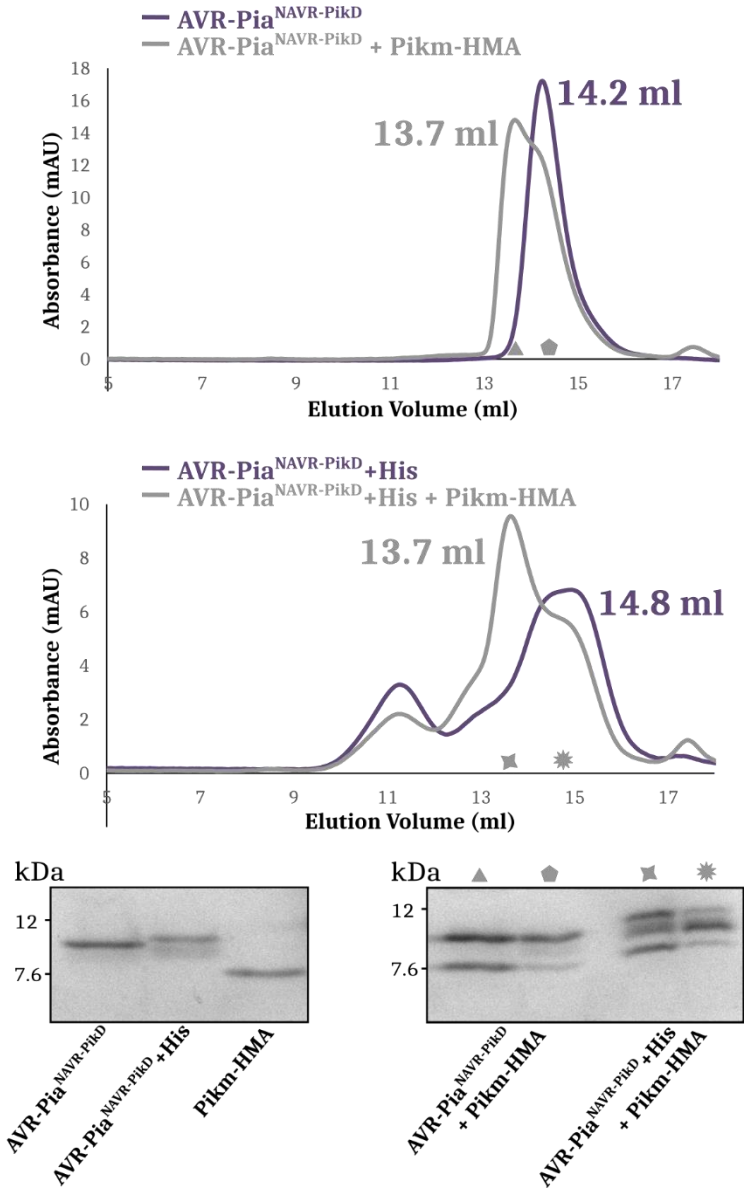


Figure 3.5-8: In vitro interactions of AVR-Pia$^{NAVR\text{-}PikD}$ with Pikm-HMA under high salt conditions with/without C-terminal 6xHis tag.

Analytical gel filtration traces showing the elution volume of AVR-Pia$^{NAVR\text{-}PikD}$ alone (purple) and an AVR-Pia$^{NAVR\text{-}PikD}$/Pikm-HMA mixture (grey). Top panel shows AVR-Pia$^{NAVR\text{-}PikD}$ without the additional C-terminal 6xHis tag, bottom panel shows the same experiment using tagged protein. SDS-PAGE analysis shows the fractions taken where indicated by corresponding symbols on the trace. Volumes shown indicate elution volumes. Baselines corrected to zero.

To understand how AVR-Pia$^{NAVR\text{-}PikD}$ is interacting with Pikp-HMA, and indeed if the interaction with Pikm-HMA is robust, the effector was co-purified with both HMAs in the same manner as for the AVR-Pia/Pikp-HMA complex (section 3.4) and put into numerous

crystallisation trials. Unfortunately, although several protein crystals were obtained, they all failed to diffract X-rays when tested.

From these results, it appears that the addition of the N-terminal extension of AVR-PikD onto AVR-Pia does not improve the binding and response of Pikp-HMA or Pikm-HMA. Both NLRs failed to respond to the chimera in planta, and the binding affinities did not appear to be increased in SPR experiments. This could suggest that the N-terminal arm remains disordered in the chimera, and does not anchor to the core structure as seen in AVR-PikD. This putative disorder could be one factor in the poor crystal formation for this protein. Additionally, the solving of the AVR-Pia/Pikp-HMA structure provided alternative, structure-informed, methods to improve Pikp response to AVR-Pia.

3.6 Testing pCambia for AVR-Pia in planta assays

In recent years, the Golden Gate method of cloning has gained popularity, and is now used widely for generating expression constructs (see Chapter 6). All the in planta experiments discussed thus far in the chapter have used this method of cloning for generating transient expression constructs. However, the pCambia system is also commonly used, and has been employed previously in the lab (Maqbool et al., 2015). The pCambia system is based on the pPZP family of agrobacterium vectors (Hajdukiewicz et al., 1994) and can drive high levels of expression, leading to stronger responses in the cell death assays (M. Franceschetti, personal communication).

Given that the cell death response to AVR-Pia is quite low using the Golden Gate system, it was decided to try using pCambia vectors, in case this boosted the cell death to a more visible level. All constructs were provided by Hiromasa Saitoh, including both Pikp-1 and Pikp-2 NLRs in pCambia, along with AVR-PikD, AVR-PikD^{H46E}, AVR-PikD$^{\Delta 22\text{-}52}$, AVR-Pia and AVR-Pia$^{NAVR\text{-}PikD}$. As well as using a different vector system, all constructs used alternative tags compared to those in the Golden Gate system (simply due to precedent in the different labs).

Initial cell death experiments were carried out with these constructs, using the agrobacterium ratios previously optimised for pCambia (Maqbool, 2015). Initial results showed some cell death in response to the negative control AVR-PikD^{H46E} (figure 3.6-1), suggesting that the NLRs expressed from the pCambia vectors were causing a slight immune response without the presence of a recognisable effector. Additionally, there

was a lot of variation between cell death scores of the same sample even within the same experiment.

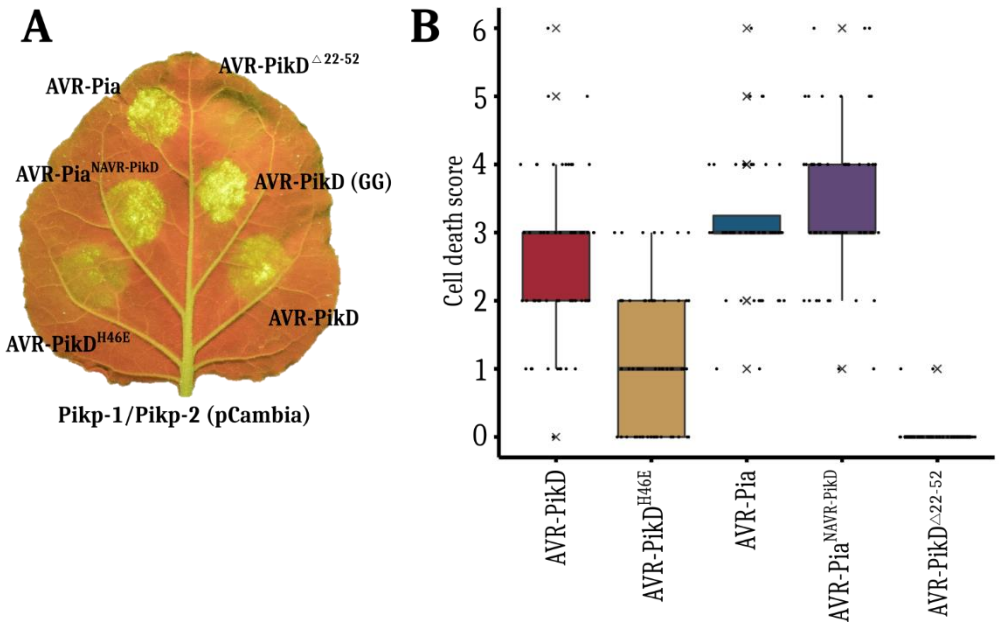


Figure 3.6-1: Pikp NLRs expressed from the pCambia vector trigger cell death even in presence of negative control in *N. benthamiana*.

A) Representative leaf images taken under UV light 5 dpi showing cell death progression at infiltration sites. **B)** Box plot showing cell death scores for 60 technical repeats over 2 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles.

To prevent these issues, the experiment was redesigned using the NLRs expressed from Golden Gate vectors and only the effectors in pCambia. This regained the specificity of response, and the Pikp NLRs were no longer responding to the negative control (figure 3.6-2B). Nevertheless, the results were still somewhat surprising. Firstly, the response to AVR-Pia increased dramatically, but this was alongside a decrease in response to AVR-PikD, the positive control. In fact, the response to AVR-PikD expressed from pCambia was lower than the response to AVR-PikD expressed from the Golden Gate vector, which was not expected. Additionally, there was a strong immune response seen to the chimera AVR-Pia$^{NAVR\text{-}PikD}$ (figure 3.6-2B), where there had been no response to it when expressed from the Golden Gate vector (figure 3.5-4). This could be due to a stronger level of expression driven by the pCambia promoter, or it could be because the tag on the pCambia construct is found at the C-terminus of the protein, rather than the N-terminus as in the Golden Gate system. Given that the N-terminus of the chimera could be disordered, having the tag at the N-terminus of the protein in the Golden Gate system may be causing an issue for recognition of the effector in planta.

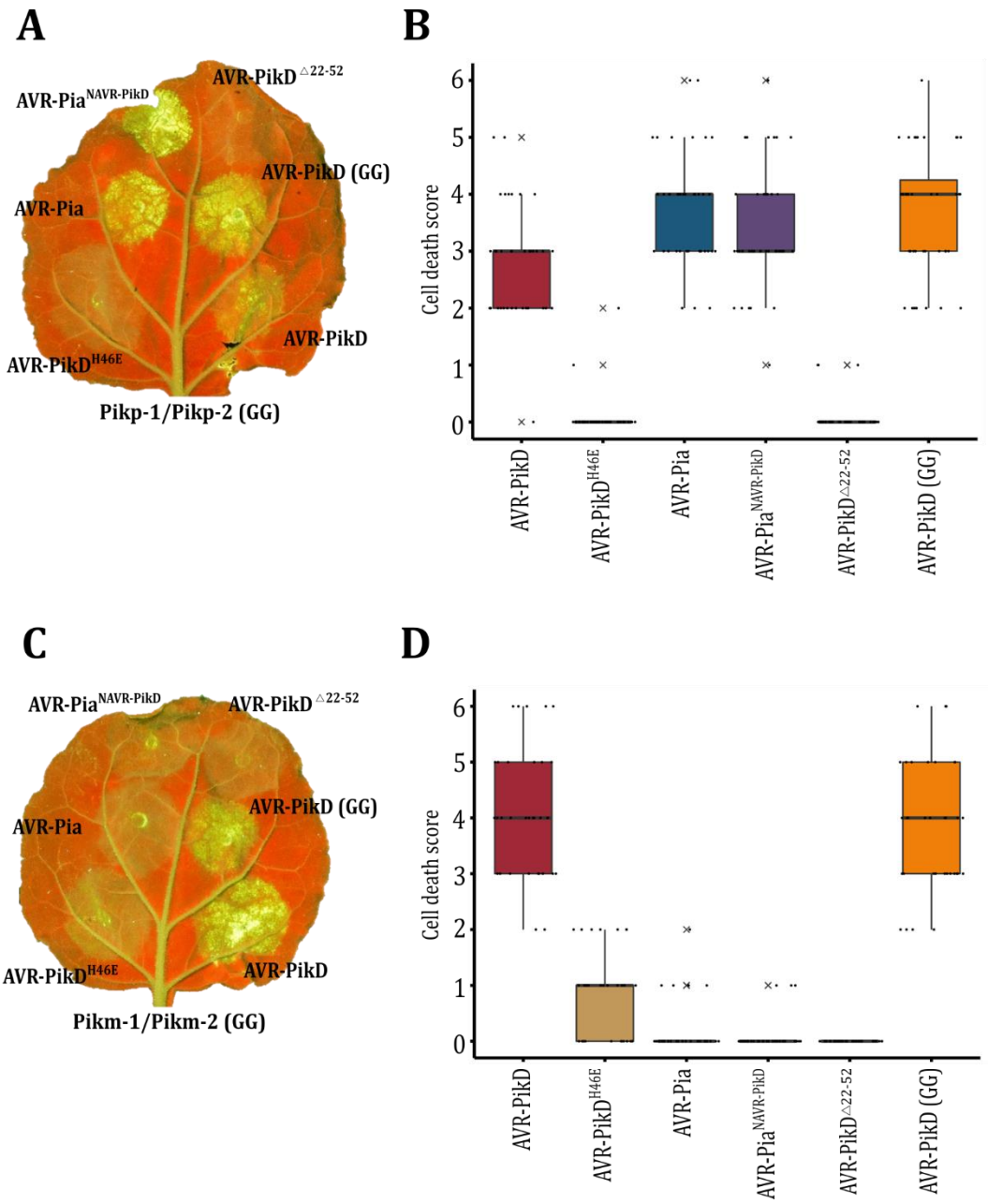


Figure 3.6-2: The impact of expressing AVRs from pCambia in the *N. benthamiana* model system.

A) Representative leaf images taken under UV light 5 dpi showing cell death progression at infiltration sites for Pikp. **B)** Box plot (Pikp) showing cell death scores for 40 technical repeats over 2 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles. **C)** Representative leaf images taken under UV light 5 dpi showing cell death progression at infiltration sites for Pikm. **D)** Box plot (Pikm) showing cell death scores for 34 technical repeats over 2 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles.

Whereas the responses of the Pikp NLRs differ greatly when effectors are expressed from the pCambia system compared to the Golden Gate system, the pattern of response for Pikm is broadly similar (figure 3.6-2D). While the response to AVR-PikD is still weaker

than expected, Pikm NLRs still cannot respond to either AVR-Pia or AVR-Pia$^{NAVR\text{-}PikD}$, in agreement with the Golden Gate system (see figures 3.2-3 and 3.5-5).

Overall, the pCambia system appears to provide a tool by which we can see a much stronger Pikp response to both AVR-Pia and the AVR-Pia$^{NAVR\text{-}PikD}$ chimera, but the fact that the response of Pikp to the 'matched' effector AVR-PikD is lower than the response to AVR-Pia suggests that this system is less representative of the true rice/*M. oryzae* system than the Golden Gate vector system. There was much more variability in the pCambia-based results compared to the equivalent experiments from the Golden Gate system. Another difficulty encountered was that some proteins expressed from the pCambia vectors were harder to detect via Western blot when extracted from leaf tissue. This issue requires troubleshooting that was not possible during the timescale of the project, but means that when pCambia vectors are used for cell death assays in this work, it has not been possible to confirm protein expression. Therefore, despite some benefits to using the pCambia system, it does not appear to be the most suitable system in my hands for this project.

3.7 Investigating the 'interface' mutants

As discussed in section 3.4, Pikp-HMA binds AVR-Pia at a different interface to AVR-PikD. Following determination of the AVR-PikD/Pikp-HMA structure (Maqbool et al., 2015), a number of mutants had been made in the lab (by A. Maqbool, M. Franceschetti) to confirm and investigate the AVR-PikD binding interface in planta. One of these mutations in Pikp-HMA was at position Glu230 (figure 3.7-1), a key residue involved in binding His46 on the N-terminal extension of AVR-PikD. Glu230 was mutated to an arginine – causing a change from a negatively to positively charged amino acid. It was found that this mutation prevented Pikp from recognising AVR-PikD in planta (M. Franceschetti, personal communication). A further point mutation was made at the Pikp-HMA dimerisation interface (figure 3.7-1), to investigate whether dimer formation was of biological importance. Ala211 was mutated to the bulkier, negatively charged residue glutamate, to disrupt the close contact of the two HMAs. This mutation had no effect on the response to AVR-PikD in planta (M. Franceschetti, personal communication), suggesting that dimer formation of Pikp-HMA is an in vitro artefact, rather than having biological relevance (de la Concepcion et al., 2018).

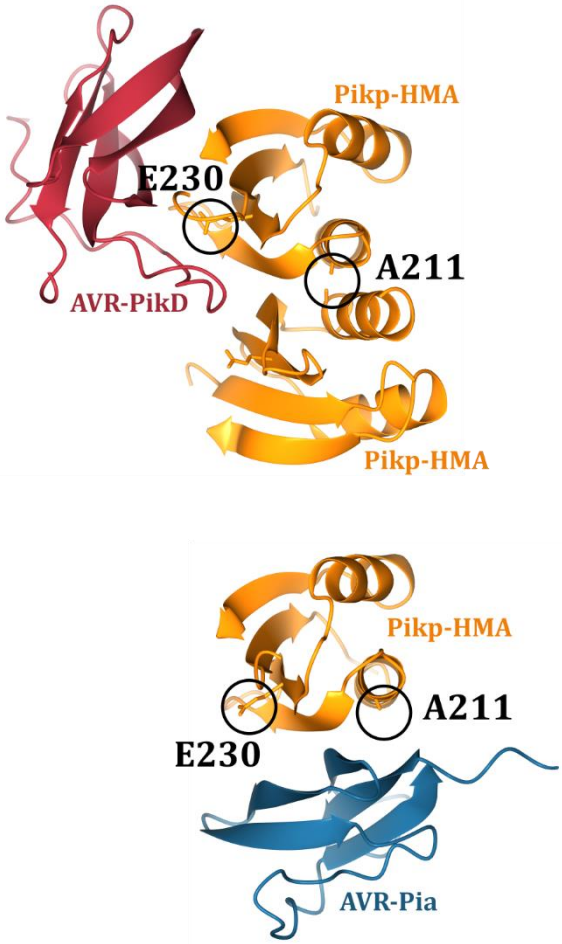


Figure 3.7-1: Locations of Ala211 and Glu230 on both the AVR-PikD/Pikp-HMA and AVR-Pia/Pikp-HMA structures.

Diagrams showing the locations for the Pikp-HMA^{A211E} and Pikp-HMA^{E230R} mutations in both the AVR-PikD/Pikp-HMA and AVR-Pia/Pikp-HMA crystal structures.

Given the difference in in vitro binding interfaces (see section 3.4) for AVR-Pia and AVR-PikD, it is expected that that these two mutants (Pikp-HMA^{A211E} and Pikp-HMA^{E230R}) will respond differently when challenged with AVR-Pia in planta. Unfortunately, when this was investigated, the low cell death response of Pikp to AVR-Pia made it difficult to obtain meaningful results (figure 3.7-2) – as there was little difference in the response of the two mutants compared to WT Pikp. However, the results for the positive control AVR-PikD were as expected and as observed by others in the lab, in that Pikp-HMA^{E230R} was no longer able to respond to AVR-PikD, but Pikp-HMA^{A211E} behaved the same as WT.

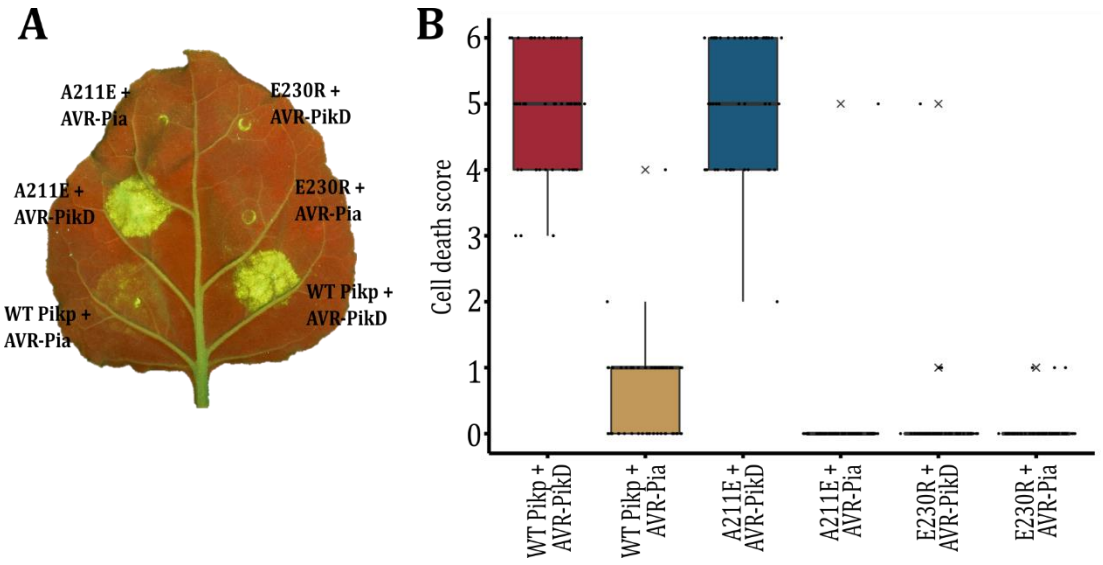


Figure 3.7-2: The Golden Gate system cannot be used to investigate the response of the Pikp-HMA^{A211E} and Pikp-HMA^{E230R} mutants to AVR-Pia.

A) Representative leaf images taken under UV light 5 dpi showing cell death progression at infiltration sites. **B)** Box plot showing cell death scores for 50 technical repeats over 2 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles. **Note:** For brevity, 'Pikp-HMA^{A211E}' is represented as 'A211E' and Pikp-HMA^{E230R} is represented as 'E230R'.

It was discussed in section 3.6 that although using pCambia vectors did not seem to be a good model for the native rice/*M. oryzae* system, it provided a technique to obtain a stronger response for Pikp to AVR-Pia. This could be a useful tool to help study the Pikp-HMA^{A211E} and Pikp-HMA^{E230R} mutants. Additionally, using the pCambia system, it is possible to see a response for WT Pikp to AVR-Pia$^{NAVR\text{-}PikD}$, so this could be a method to investigate whether the chimera is binding at the 'AVR-Pia-like' interface or the 'AVR-PikD-like' interface. Therefore, the WT and two mutant Pikp-1 NLRs (in the Golden Gate vectors as before) were tested against AVR-PikD, AVR-Pia and the chimera in the pCambia system. The results (figure 3.7-3) suggest that the cell death response in planta supports the binding interfaces determined in the crystal structures. The A211E mutation (that breaks the Pikp-HMA dimer interface) does not affect the response to AVR-PikD, but completely arrests the response to AVR-Pia or AVR-Pia$^{NAVR\text{-}PikD}$. In contrast, the E230R mutation prevents response to AVR-PikD, but leaves the response to AVR-Pia and the chimera unaffected. It is observed that the cell death responses triggered by the Pikp-HMA^{E230R} mutant are slightly lower than the WT response, which could indicate either that the point mutation has hindered the response to AVR-Pia and AVR-Pia$^{NAVR\text{-}PikD}$, or may indicate that Pikp-HMA^{E230R} is expressed at a lower level in the leaf tissue. As mentioned in section 3.6, it was not possible to get good quality Western blots for these experiments, so this could not be confirmed. Nevertheless, these initial

observations suggest that Pikp-HMA binds AVR-Pia$^{NAVR\text{-}PikD}$ at the same interface as AVR-Pia, suggesting that the N-terminal extension has not had an impact on the binding site.

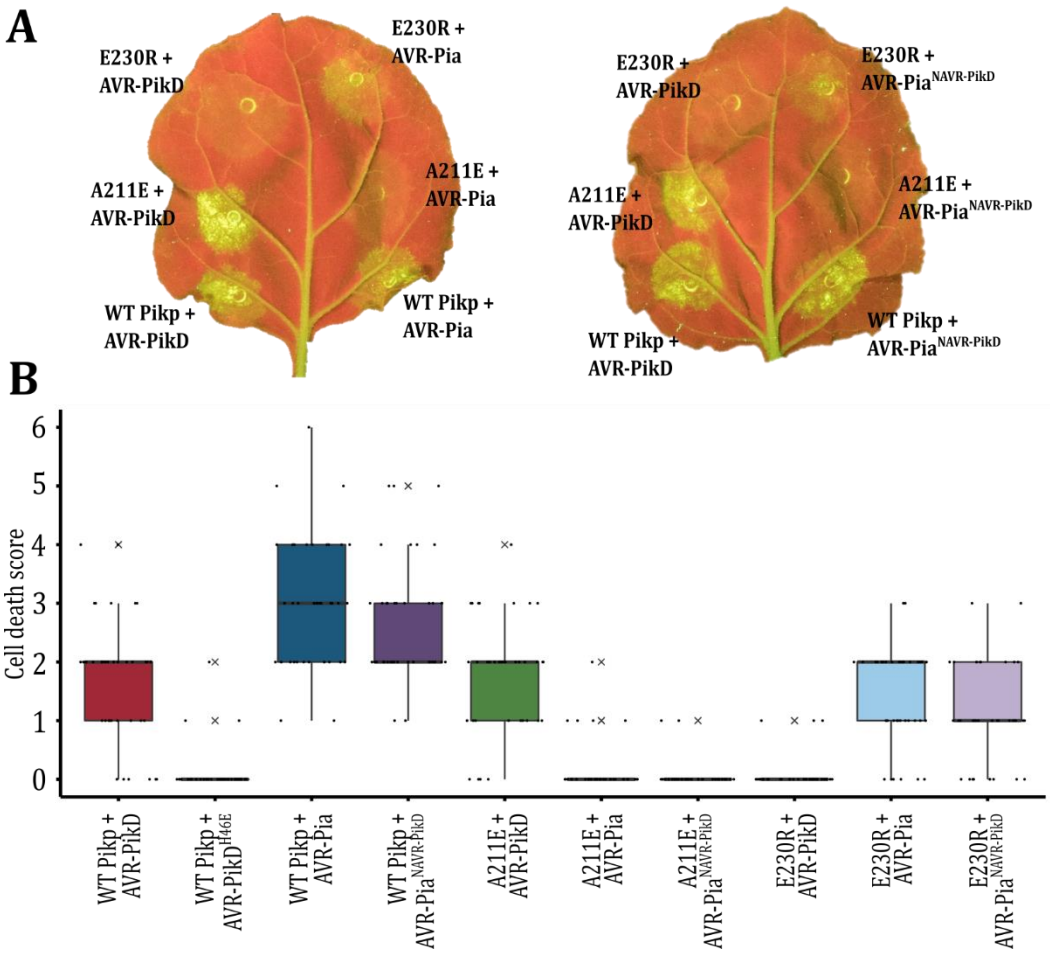


Figure 3.7-3: Using the Pikp-HMA^{A211E} and Pikp-HMA^{E230R} mutants to investigate AVR-Pia and AVR-Pia$^{NAVR\text{-}PikD}$ binding sites in planta.

A) Representative leaf images taken under UV light 5 dpi showing cell death progression at infiltration sites. **B)** Box plot showing cell death scores for 40 technical repeats over 2 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles. **Note:** For brevity, 'Pikp-HMA^{A211E}' is represented as 'A211E' and Pikp-HMA^{E230R} is represented as 'E230R'.

3.8 A Pikp-HMA/AVR-Pia/AVR-PikD complex is not formed in vitro

Evidence obtained thus far has indicated that Pikp-HMA binds AVR-Pia and AVR-PikD at different interfaces both in vitro and in planta. It could be hypothesised that Pikp-HMA can bind both effectors at the same time, which if true, would lead to interesting questions about the effect on immune response in planta. To investigate this, a simple competition assay was visualised using analytical gel filtration (figure 3.8-1). A pre-

purified complex of AVR-Pia and Pikp-HMA was incubated with purified AVR-PikD. If AVR-PikD was able to bind to Pikp-HMA alongside AVR-Pia, a triple complex would be formed, accompanied by a peak shift on the analytical gel filtration trace. Both components (the AVR-Pia/Pikp-HMA complex and AVR-PikD) were first run separately down the analytical column to determine their elution volumes (figure 3.8-1) before the complex and AVR-PikD were mixed in a 1:1 molar ratio and incubated for 2 hours on ice. This mixture was then run down the column. No significant peak shift was observed on the trace (figure 3.8-1). Instead, when fractions were analysed by SDS-PAGE, it appeared that the peak at 13.5 ml (labelled d) was actually now composed entirely of AVR-PikD and Pikp-HMA, and the secondary peak (e and f) contained AVR-Pia and AVR-PikD. This suggests that the AVR-PikD has fully displaced AVR-Pia from a complex with Pikp-HMA. The free AVR-Pia now elutes later, co-eluting with excess AVR-PikD. The excess AVR-PikD is likely present because it had only been mixed in a 1:1 complex with the Pikp-HMA, and in vitro the AVR-PikD forms a 1:2 complex due to Pikp-HMA dimerisation. Overall, this experiment shows further evidence for the stronger binding affinity of Pikp-HMA for AVR-PikD compared to AVR-Pia, and indicates that it is not possible for both effectors to be bound by Pikp-HMA simultaneously.

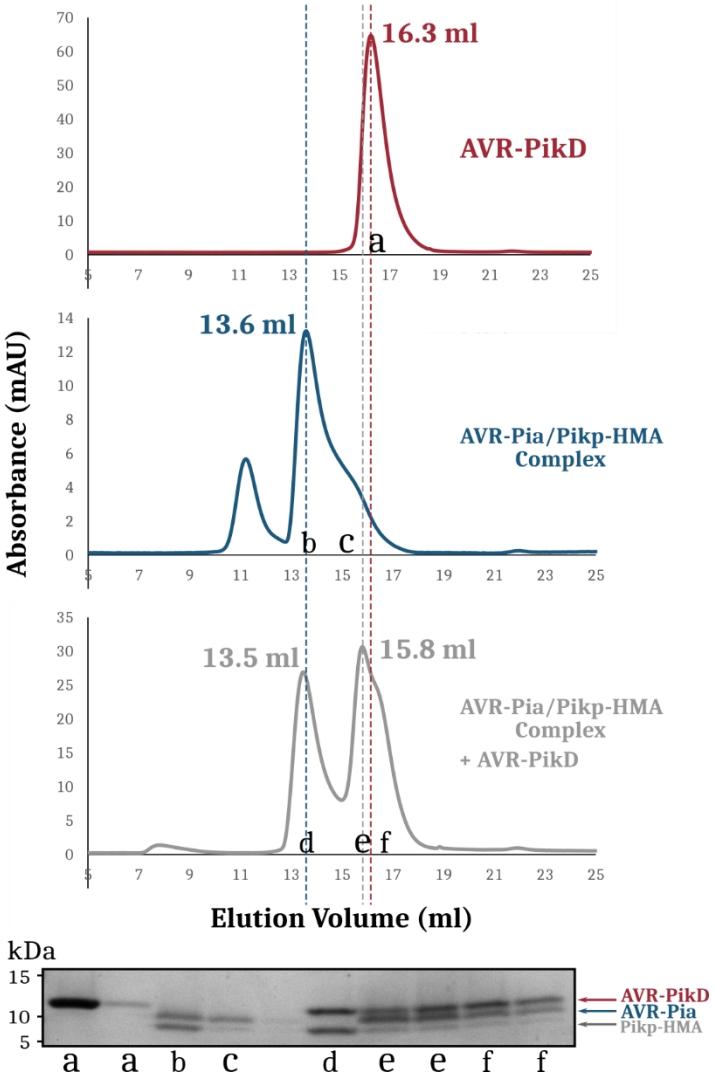


Figure 3.8-1: Analytical gel filtration indicates that Pikp-HMA cannot simultaneously bind AVR-Pia and AVR-PikD.

Analytical gel filtration traces showing the elution volume of AVR-PikD alone (red), a pre-formed complex of AVR-Pia/Pikp-HMA (blue) and a 1:1 mixture of AVR-PikD and AVR-Pia/Pikp-HMA (grey). SDS-PAGE analysis shows the fractions taken where indicated by letters a-f on the trace. Volumes shown indicate elution volumes. Baselines corrected to zero.

3.9 Discussion

Building upon initial observations using *M. oryzae* infecting rice, the work described here has shown that Pikp is able to respond weakly to its non-matched effector AVR-Pia and that the isolated HMA domain is able to bind to AVR-Pia in vitro, although more weakly than to AVR-PikD. This observation is perhaps not surprising, given that these effectors are structurally highly conserved and are both recognised by integrated HMA domains in rice NLRs. What is more surprising is the discovery of a novel effector binding interface for Pikp-HMA, where AVR-Pia binds at the site normally held by a second Pikp-HMA molecule in solution. The purification strategy designed to form the complex for

crystallographic studies, whereby AVR-Pia and Pikp-HMA are purified separately and then mixed at the final stage of purification, suggests that the AVR-Pia must be able to break this Pikp-HMA dimer, (which forms when the HMA is heterologously expressed from *E. coli*) and take its place in the complex. However, given that the dimerisation of Pikp-HMA in vitro is thought to have no biological relevance, it is likely that AVR-Pia breaking this dimer is also not biologically important. Interestingly, a previous study had used docking models to compare the expected orientation of binding for AVR-PikD/Pikp-HMA and AVR-Pia/RGA5-HMA. The models predicted that RGA5-HMA and Pikp-HMA would bind their respective effectors at different interfaces (Ortiz et al., 2017), which is in agreement with the results shown here for AVR-Pia binding to Pikp-HMA. This could suggest that the change in interface is due to a property of the effector, rather than a property of the HMA domain. In addition, the same study by Ortiz et al. used NMR titration experiments coupled with validation by yeast two-hybrid and co-immunoprecipitation to identify residues that were likely to be involved in the interaction between AVR-Pia and the HMA domain of RGA5. The candidate interaction surface of AVR-Pia identified in the study does not fully match the interaction interface identified here between AVR-Pia and Pikp-HMA, but some of the key residues found in the study are positioned at the interface between AVR-Pia and Pikp-HMA. For example, both AVR-Pia residues Phe24 and Thr46, which were shown to be important for effector recognition by the RGA5/RGA4 NLRs are at the interface with Pikp-HMA. Ortiz et al. found that mutation of residue Arg43 to a glycine resulted in abolished immunoprecipitation with the isolated HMA domain of RGA5. In the structure shown here (figure 3.4-2), Arg43 forms a hydrogen bond with Asp217 from Pikp-HMA, indicating that the some of the same residues are important for AVR-Pia to interact with both RGA5-HMA and Pikp-HMA. This suggests that the interfaces might be similar in both cases, although there is currently no crystallographic or NMR structural evidence to show how the AVR-Pia/RGA5-HMA interface is formed.

While this interaction of AVR-Pia and Pikp-HMA appears to be robust in planta and in vitro, no such interaction has been observed for AVR-Pia binding to the Pikm allele. Within the different Pik alleles, the HMA domain is the most polymorphic region (de la Concepcion et al., 2018) and Pikm-HMA shares only 62 % sequence identity with Pikp-HMA, although they share a similar structure (de la Concepcion et al., 2018). Further tests would be needed to investigate why Pikm-HMA does not form a complex with AVR-Pia. One notable difference is that the important Asp217 residue in the AVR-Pia/Pikp-HMA interface is replaced by a histidine residue in Pikm-HMA, which would no longer be able to form the strong hydrogen bond/salt bridge interaction with Arg43 in AVR-Pia. This

could significantly weaken the binding between the two proteins, although seems unlikely to fully account for a lack of interaction.

In an attempt to strengthen the interaction between AVR-Pia and Pikp-HMA, and perhaps introduce binding to Pikm-HMA, a chimera was made whereby the N-terminal extension of AVR-PikD (the greatest structural disparity between the two effectors) was grafted onto AVR-Pia. Although this protein was stable when expressed from *E. coli*, it could not be determined whether the N-terminal arm was making a significant difference to the overall structure of the protein. Attempts at generating diffraction quality crystals of the chimera in complex with Pikp-HMA failed, which could be due to disorder. In planta, the addition of this N-terminal extension appears to prevent Pikp from responding to the effector, even though the protein is stably expressed, which could suggest that the arm is in some way preventing a response – possibly by blocking the Pikp-HMA binding site. In vitro, the binding affinity of the chimera for Pikp-HMA is very similar to AVR-Pia – an indication that the N-terminal arm is unable to effect any change in binding to the HMA. Overall, this strategy to strengthen the interaction of AVR-Pia and Pikp-HMA was unsuccessful. Even though in analytical gel filtration experiments, Pikm-HMA was able to bind to the chimera, where it had previously been unable to bind AVR-Pia, this did not hold true in SPR experiments or in planta, suggesting that the gel filtration results were misleading rather than of biological relevance.

While the pCambia system did not provide a reliable model in my hands to replace the Golden Gate cloning method, it was interesting to observe that when the effectors were expressed using the pCambia vectors, the cell death response by Pikp was greatly increased for AVR-Pia, and found to be almost the same level for AVR-Pia$^{\text{NAVR-PikD}}$, where there was no response seen when using the Golden Gate system. One hypothesis for this is that the pCambia promoter drives much higher expression than the Ubiquitin promoter used in the Golden Gate vector. This could be further investigated by testing the expression level of the effectors in planta when expressed from the pCambia system, which was unfortunately not possible within the timescale of this project. However, despite these issues, experiments carried out using the 'interface mutants' Pikp-HMA$^{\text{A211E}}$ and Pikp-HMA$^{\text{E230R}}$ strengthened the idea that the addition of the N-terminal arm of AVR-PikD onto AVR-Pia does not change the binding interface to the 'AVR-PikD-like' interface. This assay also suggests that the AVR-Pia/Pikp-HMA complex is in a different orientation compared to AVR-PikD/Pikp-HMA in the plant cell environment as well as in vitro (as seen by X-ray crystallography) because the interface mutant that disrupts the AVR-Pia binding site is still able to respond to AVR-PikD and the mutant that

disrupts the AVR-PikD binding site is still able to respond to AVR-Pia. This lends weight to the idea that the AVR-Pia/Pikp-HMA crystal structure shows the biologically relevant interface in *N. benthamiana*, despite its small interface area and weak intermolecular bonding.

Despite the novel results gained from in planta study, some further improvements could be made to these experiments. Although negative controls (generally the non-binding AVR-PikD variant with a H46E mutation) were used throughout, an additional experiment using an empty vector control in place of the effector would be needed to ensure that the cell death observed was effector-dependent. By using an empty vector control, it would ensure that no part of the effector construct plays a role in inducing cell death in *N. benthamiana* (whether due to recognition by plant immune receptors or not). On a related level, the expression levels of the NLR proteins themselves should be considered. This transient expression of proteins in the model plant (under the *A. tumefaciens* Mas promoter) does not accurately mimic the expression levels that would be found in the native rice under the native promoter. It has previously been shown that the expression levels of NLRs correlate with the level of cell death observed, and that using different promoters for transient expression can significantly alter the observations of hypersensitive response (Zhang et al., 2004). Although this is perhaps not surprising, it does indicate that effector-independent changes in observed cell death can be an issue in such transient expression assays.

Additionally, it would be beneficial to test RGA5-HMA under the same conditions, to be used as a positive control for AVR-Pia cell death. However, due to agreements in place between different research groups, it was not possible to work with the RGA5/RGA4 pair in this project.

As described in section 3.3, it was not possible to prove a direct interaction between Pik-1 and AVR-Pia in planta, as the co-IP experiments gave conflicting results. As discussed, it is hoped that further repeats of these experiments might provide more conclusive answers, but it is also possible that the weak interaction in this particular case might always lead to a varied outcome. It might be necessary to try other techniques, such as chemical cross-linking (Fukao, 2012) to consistently visualise the weak interaction. Alternatively, other techniques could be used to assess the interaction between NLR and effector, such as fluorescent-based in planta methods (Kerppola, 2006), although the additions of fluorescent tags can cause additional steric difficulties in obtaining biologically relevant protein complexes.

Finally, a preliminary in vitro assay indicates that Pikp-HMA cannot bind both effectors at the same time, despite their different interfaces. AVR-PikD is readily able to displace AVR-Pia from the Pikp-HMA complex, presumably due to its much higher binding affinity in vitro. This suggests that in planta, Pikp would likely recognise AVR-PikD in preference to AVR-Pia if both effectors were present, due to the stronger affinity of this interaction. Although this provisional analytical gel filtration test was not pursued further, an alternative method to investigate this AVR-Pia/AVR-PikD competition would have been to use SPR. If AVR-Pia had been immobilised to the chip and Pikp-HMA was flowed over, the mass on the chip would have increased as the HMA binds. If AVR-PikD was then flowed over (without chip regeneration), the mass would have decreased as AVR-PikD pulled Pikp-HMA away from AVR-Pia. This assay would provide a different method to assess the effector binding competition, although it was not possible within the timeframe of this study.

The work described here has shown how Pikp can respond to AVR-Pia weakly in planta and shows an interesting example of cross-reactivity for rice NLRs to structurally similar effectors. Detailed studies of how NLR domains can interact with pathogen effectors have important implications for agriculture, as it extends our current understanding of how plants combat disease. Based upon the observations described here, Chapter 4 shows how the project was extended into engineering an NLR that can respond more strongly to AVR-Pia. Although not directly applicable to agriculture at this stage, this may provide an interesting proof-of-concept study that could be applied to other NLR/effector pairs.

Engineering an enhanced response to AVR-Pia

Modifying the Pikp HMA domain to trigger a stronger immune response to AVR-Pia in planta

4.1 Introduction

Understanding the ways in which host NLR proteins recognise pathogen effectors is key to unravelling the complexities of immunity and the constant evolutionary battle between plant and pathogen. Aside from satisfying simple curiosity, the ability to interpret the causes and symptoms of plant disease at the molecular level affords us the opportunity to try tipping the balance of immunity in favour of the plant. In recent years, there have been several examples of NLRs being engineered in different ways to change or extend their recognition ability (Cesari, 2018). One example of an NLR being engineered to recognise additional effectors is the case of the R3a NLR, which was mutated to recognise a new isoform of the *P. infestans* effector AVR3a (Segretin et al., 2014). Other cases of engineering have involved the modification of a decoy protein, rather than the NLR itself. One such example is the *A. thaliana* host protein PBS1, which is targeted by a protease effector AVRPphB. Cleavage of PBS1 causes pathogen recognition and immune response through the NLR RPS5. Modification of the cleavage site of PBS1 to recognition sites for other protease effectors allowed RPS5 to mediate a response to other effectors (Kim et al., 2016). As the scientific field learns more about the mechanisms behind NLR-mediated plant immunity, there is capacity for further advances in this area.

In the context of the Pik NLR system, recently solved structures of Pik-HMA domains in complex with different AVR-Pik alleles (de la Concepcion et al., 2018) have provided promising insights that may help rationally design modified Pik NLRs to recognise a broader range of AVR-Pik effectors, including the previously unrecognised allele AVR-PikC. Using structural and biophysical techniques, as well as model plant systems, it is possible to assess whether engineered NLRs are likely to have changed or enhanced recognition when tested in their native system – in this case the rice/*M. oryzae* host/pathogen system. Whether or not the immune engineering would be robust and long-lasting in the field is a question that is not easy to answer. As previously discussed, the rate of pathogen evolution means that creating durable, broad-range resistance is a challenge (Zhang & Coaker, 2017). Nevertheless, modifying the amino acid sequence of plant NLRs and thus enabling greater resistance to disease is a powerful technique that is worthwhile exploring.

Chapter 3 describes how the Pikp NLR pair is able to weakly respond to AVR-Pia in planta, despite being responsible for recognising the AVR-Pik effectors. This observation was extensively characterised in vitro and in planta, and the structural basis of this

interaction was analysed. Although there is another pair of NLRs with an integrated HMA domain (RGA5/RGA4) that is able to recognise AVR-Pia in planta, it would be interesting to be able to engineer a single NLR that was able to respond robustly to both AVR-Pia and AVR-PikD. The AVR-Pia/Pikp-HMA (Chapter 3) and AVR-PikD/Pikp-HMA (de la Concepcion et al., 2018; Maqbool et al., 2015) crystal structures showed that the effectors bind to Pikp-HMA at different interfaces, so it may be possible to strengthen the interaction at the AVR-Pia binding site without significantly disrupting the binding of AVR-PikD at a different location. The work described in this chapter shows how different strategies were employed to engineer the Pikp HMA domain to respond more strongly to AVR-Pia.

4.2 Structure-guided point mutations

4.2.1 Design of mutants

Chapter 3 describes how the structure of AVR-Pia in complex with Pikp-HMA was determined. By studying this structure in detail, it might be possible to identify regions in Pikp-HMA that could be mutated to convey stronger binding to AVR-Pia. If the proteins can interact more strongly in vitro, precedent suggests that this would correlate with the NLR triggering a stronger immune response in planta (de la Concepcion et al., 2018). The AVR-Pia/Pikp-HMA crystal structure revealed a 1:1 complex with the primary interface centred at β-2 in Pikp-HMA. Many of the interactions at the interface are hydrogen bonds that form between the peptide backbone of the two partners in the complex. This creates limited options for generating point mutations; because while changing the character of amino acid side chains is relatively straightforward, the innate properties of the peptide backbone cannot be changed. However, upon close examination of the structure, two residues were identified that had the potential to strengthen the AVR-Pia/Pikp-HMA interaction.

The first of these mutations is an alanine positioned on β-2 (figure 4.2-1). The small alanine residue does not extend far into the space between the two partners of the complex, but it was noted that in RGA5 (the 'matched' NLR to AVR-Pia) the corresponding residue at this position in the sequence is a glutamate, which is much larger than alanine and negatively charged. It is possible that in the 'matched' interaction, this residue is forming contacts with AVR-Pia that helps to stabilise a complex. Therefore, the alanine in Pikp-HMA (Ala220) was mutated to a glutamate – to give a A220E mutation.

The second point mutation to be generated is at a loop of Pikp-HMA (residues Ser212 – Val216) that extends into the space slightly beyond the core structure. Opposite this loop in the complex is a positively charged surface patch on AVR-Pia generated by the side chains of two arginine residues (Arg23 and Arg43), that point towards the loop of Pikp-HMA (figure 4.2-1). The residue in the centre of the Pikp-HMA loop is a glycine, and it was hypothesised that replacing this glycine with a glutamate would create a long, negatively charged side chain that might be able to reach across the short gap between the two partners in the complex and form an interaction with the positively charged patch in AVR-Pia. In the WT complex, Arg43 is already forming a stabilising interaction with Pikp-HMA residue Asp217, but there is potential to strengthen the bonding in this area if Arg23 could form an additional interaction with the Ser212 – Val216 loop region.

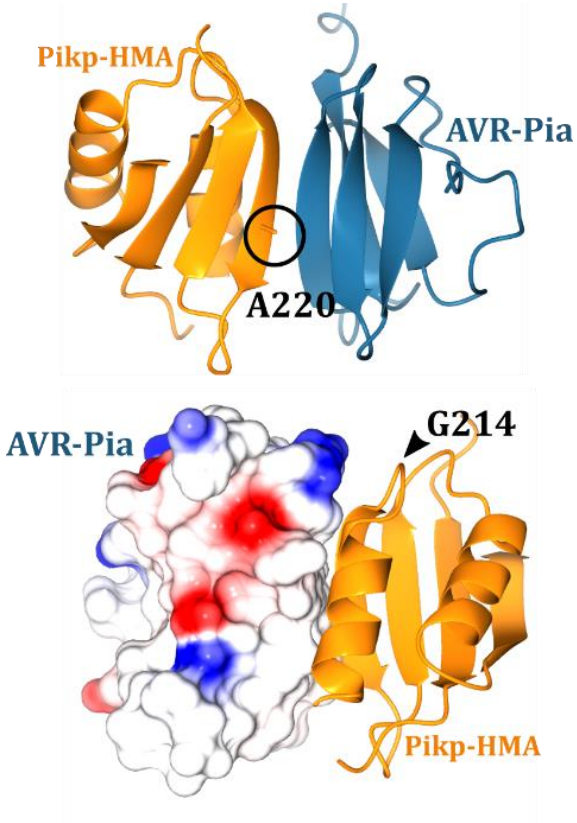


Figure 4.2-1: Locations of A220 and G214 mutation sites in AVR-Pia/Pikp-HMA structure.

Top panel indicates the location of Ala220 in Pikp-HMA (circled). Bottom panel indicates the location of Gly214 in Pikp-HMA (arrow). Pikp-HMA is shown in gold, AVR-Pia is shown in blue (top panel) or as a surface representation showing electrostatic potential (bottom panel).

4.2.2 Testing in planta

Both point mutations were generated by ordering synthetic DNA (IDT gBlocks® Gene Fragments) for the HMA domain and cloning back into the full-length Pikp NLR in the Golden Gate vector system. These mutant Pikp-1 NLRs were then agroinfiltrated into *N. benthamiana* with WT Pikp-2 and AVR-Pia, and the HR-like cell death response was scored as described in Chapter 3 (see figure 3.2-2). This experiment shows that the Pikp-HMA^{A220E} mutant cannot respond to AVR-Pia, as no cell death was observed in planta (figure 4.2-2).

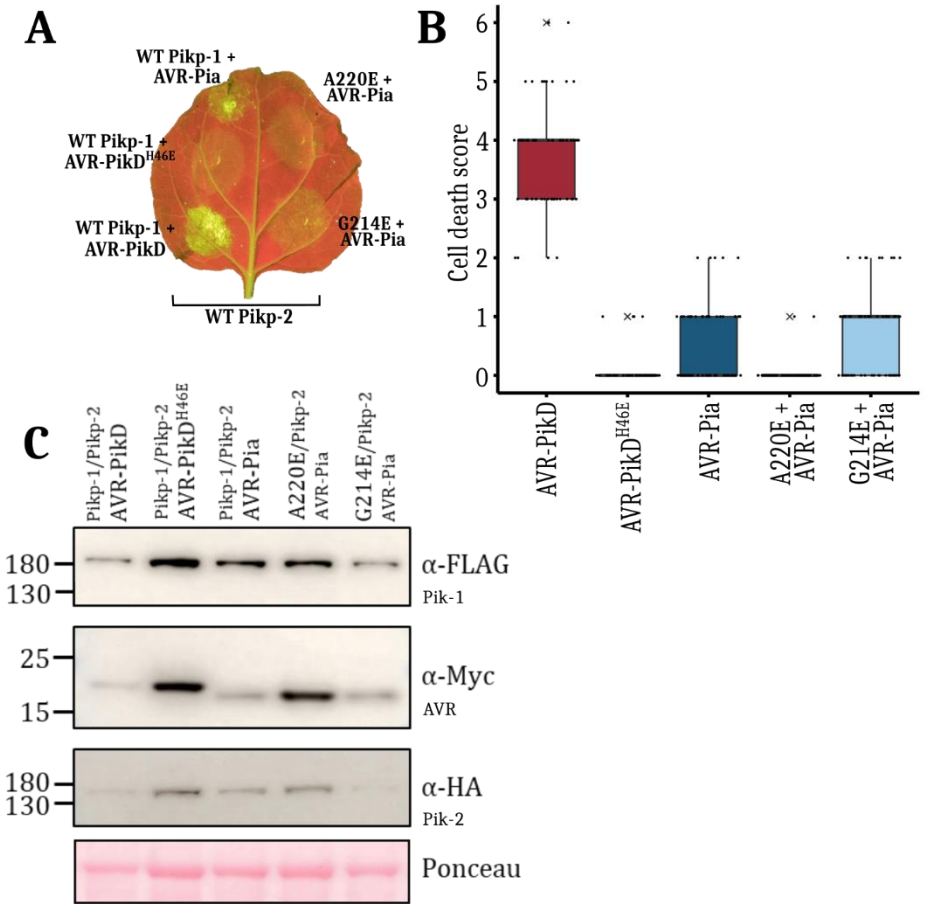


Figure 4.2-2: Pikp-HMA^{A220E} and Pikp-HMA^{G214E} do not enhance response to AVR-Pia.

A) Representative leaf images taken under UV light 5 dpi showing cell death progression at infiltration sites. **B)** Box plot showing cell death scores for 60 technical repeats over 3 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles. **C)** Western blot showing the expression of proteins at the infiltration sites. Blot shown is representative of 3 experiments showing similar results. **Note:** For brevity, Pikp-HMA^{A220E} is represented as 'A220E' and Pikp-HMA^{G214E} is represented as 'G214E'.

This could be because the addition of a bulkier residue at that position has caused steric hindrance - pushing the effector away from the HMA domain, rather than strengthening the interaction. The Pikp-HMA^{G214E} mutant yields cell death scores that are almost identical to WT Pikp-1 (figure 4.2-2), indicating that the G214E mutation has had neither a positive or negative effect. Possibly the glutamate side chain is unable to extend far enough across the gap to form the putative salt-bridge interaction with AVR-Pia Arg23, or even if this interaction is occurring, it may not affect the response in planta. In both cases, the mutant Pikp-1 NLRs were successfully expressed in planta and detected by Western blot analysis (figure 4.2-2C).

4.3 RGA5/Pikp-HMA secondary structure swaps

Using the crystal structure of AVR-Pia/Pikp-HMA to strategically design point mutations that could strengthen the interaction between the two partners was not successful. An alternative strategy is to make bigger global changes to the HMA domain to discover whether there is a more general region of the protein that could be engineered for a stronger response. If segments of Pikp-HMA are mutated to the equivalent residues from RGA5-HMA, it might be possible to identify key residues that allow RGA5-HMA to respond strongly to AVR-Pia. To achieve these modifications in a methodical manner, segments of RGA5-HMA were swapped into Pikp-HMA in a step-wise method according to secondary structure features, to help maintain the overall structure of Pikp-HMA. Given that the structure of RGA5-HMA is not available, Phyre2 (Kelley et al., 2015) was used to predict the secondary structure features of RGA5-HMA from the known structure of Pikp-HMA (PDB accession 5A6P, (Maqbool et al., 2015)) via the one-to-one threading approach. Using a sequence alignment of the two proteins and the results from Phyre2, the RGA5-HMA was divided up into six sections according to the secondary structure features (figure 4.3-1). Swap boundaries were positioned at residues that were

conserved between the two HMA domains, to minimise global structure disruption as far as possible.

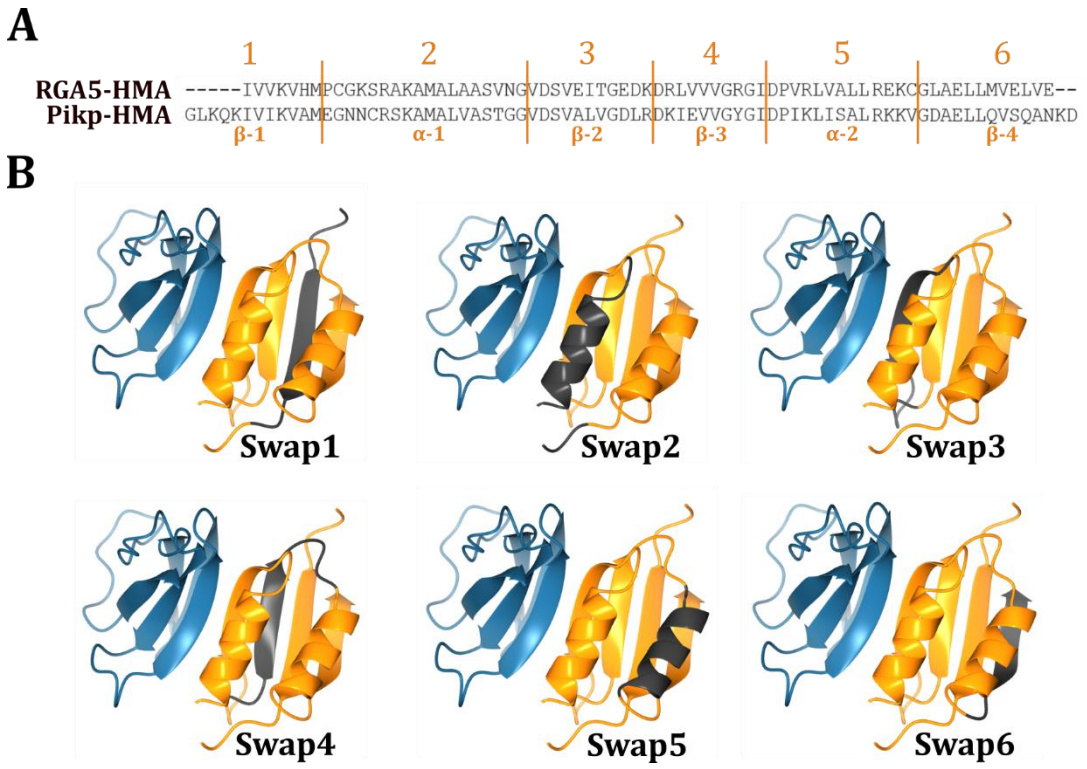


Figure 4.3-1: Design of RGA5/Pikp-HMA secondary structure swaps.

A) Sequence alignment of RGA5-HMA with Pikp-HMA showing the boundaries for secondary structure swaps. **B)** Diagram showing the location of each secondary structure swap mapped onto the AVR-Pia/Pikp-HMA structure. AVR-Pia is shown in blue, Pikp-HMA is shown in gold, and swapped sections from RGA5-HMA are shown in grey.

These swaps (numbered 1-6) were synthesised as described in section 4.2.2 and tested for response to AVR-Pia in planta. The cell death scores (figure 4.3-2B) indicate that Swap1, 3 and 4 have lost the ability to respond to AVR-Pia, even though they are expressed successfully in the leaf tissue (figure 4.3-2C). Swap2 appears to respond to AVR-Pia with a similar magnitude to WT Pikp, while both Swap5 and 6 appear to have an increased response to the effector. In each case, the NLRs and effector were correctly expressed within the leaf tissue (figure 4.3-2C).

To confirm whether the swaps were able to interact with AVR-Pia in the leaf tissue, co-immunoprecipitation was carried out for Swap2, 5 and 6. The results are not completely clear, but it appears that the mutations have not increased strength of binding between NLR and effector. For the WT (Chapter 3, section 3.3.2) it appeared that in some cases AVR-Pia could be pulled down by the NLR, but in other repeats it was not. For each of

Swap2, 5 and 6, and for WT Pikp-1, figure 4.3-2D appears to show a very faint band for the α-Myc antibody when the tagged NLRs are pulling down the effectors. This suggests that there could be a weak interaction between Swap2, 5 and 6 and AVR-Pia in planta, but the binding does not appear to have been strengthened compared to WT Pikp-1, and is still considerably weaker than the interaction between Pikp-1 and AVR-PikD.

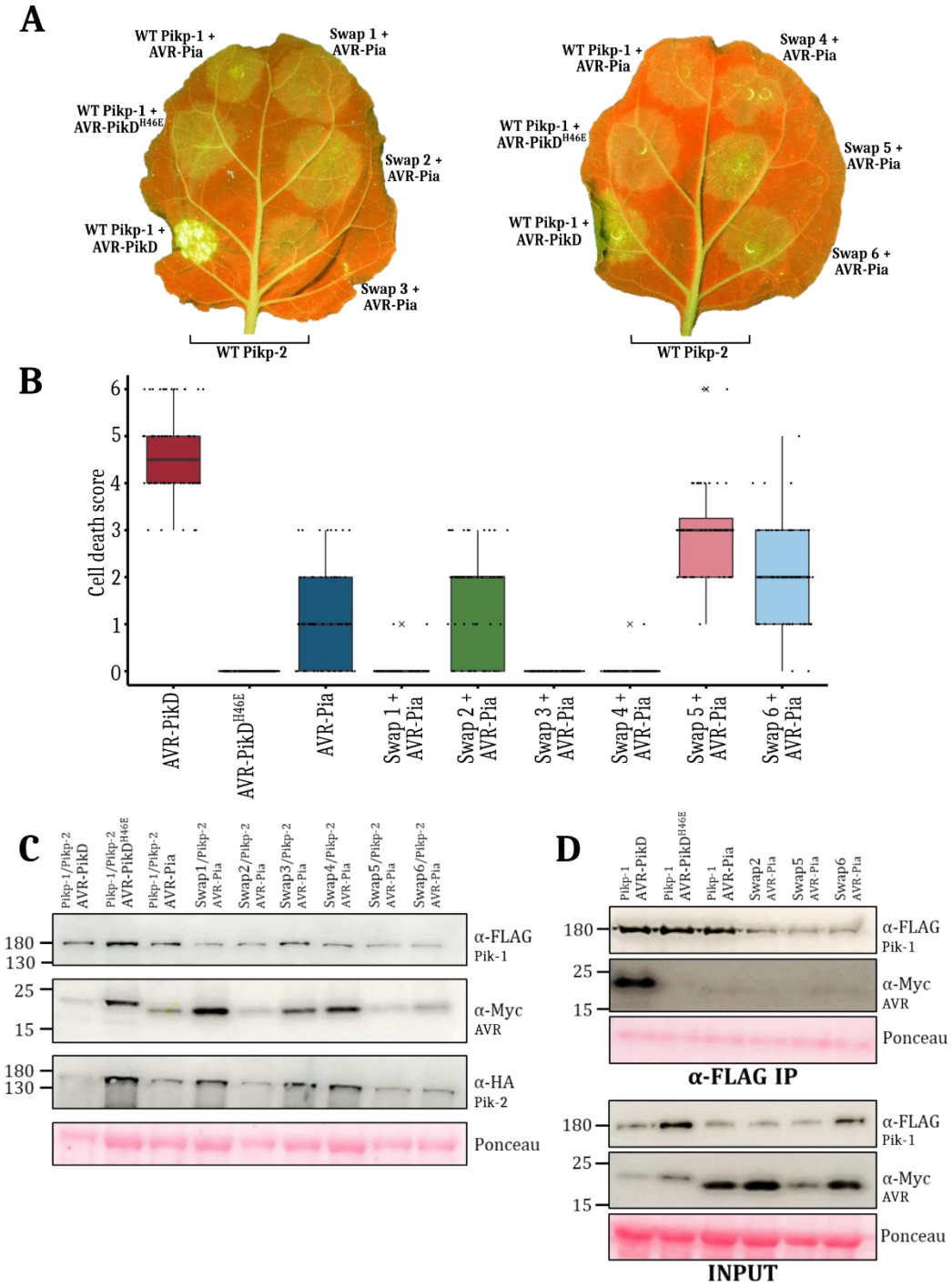


Figure 4.3-2: RGA5/Pikp-HMA swaps can enhance response to AVR-Pia.

A) Representative leaf images taken under UV light 5 dpi showing cell death progression at infiltration sites. **B)** Box plot showing cell death scores for 60 technical repeats over 3 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles. **C)** Western blot showing the expression of proteins at the infiltration sites. Blot shown is representative of 3 experiments showing similar results. **D)** Co-IP western blot showing protein input from plant extract (bottom) and outcome of immunoprecipitation of Pikp-1 or Pikp-1 mutants with FLAG magnetic beads (top). Blot is representative of 2 experiments showing similar results.

One notable observation is that the swaps showing a marked increase in cell death response (Swap5 and 6) are not part of the effector interface seen in the AVR-Pia/Pikp-HMA crystal structure, and in fact are on the opposite face of the HMA domain (see figure 4.3-1). The reasons for this are not immediately apparent, but it could imply that the increase in cell death results from a change of interaction with a different domain of Pikp-1, or a different immune signalling component, rather than a change in the interaction with AVR-Pia. If the mutations in the HMA domain are affecting intramolecular interactions, this could cause a change in how the immune response is triggered.

4.4 Combining swaps enhances the response

Following the positive changes seen in cell death response when elements of RGA5-HMA were transferred into Pikp-HMA, these investigations were extended by combining the successful swaps (Swap2, 5 and 6) to discover if this would enhance the response even further. These three swaps were combined in each possible combination, i.e. pairwise and then a triple swap (figure 4.4-1).

Figure 4.4-1: Design of RGA5/Pikp-HMA secondary structure swap combinations.

Diagram showing the location of each secondary structure swap combination mapped onto the AVR-Pia/Pikp-HMA structure. AVR-Pia is shown in blue, Pikp-HMA is shown in gold, and swapped sections from RGA5-HMA are shown in grey.

The cell death response for each of these combinations was tested as before, and it appeared that certain combinations of swaps had the potential to increase the immune response slightly further (figure 4.4-2). Indeed, for one of the swaps (Swap2+6) the cell

death response appeared almost equal to that of WT Pikp for AVR-PikD. Western blot analysis confirmed that each swap was expressed successfully in the leaf tissue (figure 4.4-2C). However, for co-IP experiments, the results were similar to those seen previously – i.e. there was inconsistency for the interaction with AVR-Pia. More repeats would be needed to obtain a clearer picture, but it appears overall that despite the increased cell death response, the interaction of the mutant NLRs with AVR-Pia is not stronger than for WT Pikp. In the data shown in figure 4.4-2D, none of Swap2+5, 2+6, 5+6, 2+5+6 or WT Pikp-1 were able to pull down AVR-Pia. This implies that even if a direct interaction is occurring in planta, it is too weak to be visualised by co-IP.

The HR-like cell death assays are not quantitative, so it is difficult to assess precisely which swaps give the biggest increase in response. However, from a visual inspection of the box plots (figures 4.3-2 and 4.4-2), the best swaps appear to be: Swap2+6, Swap5+6 and Swap5. These mutants were taken on for further investigation to try to understand the cause of the increased cell death response.

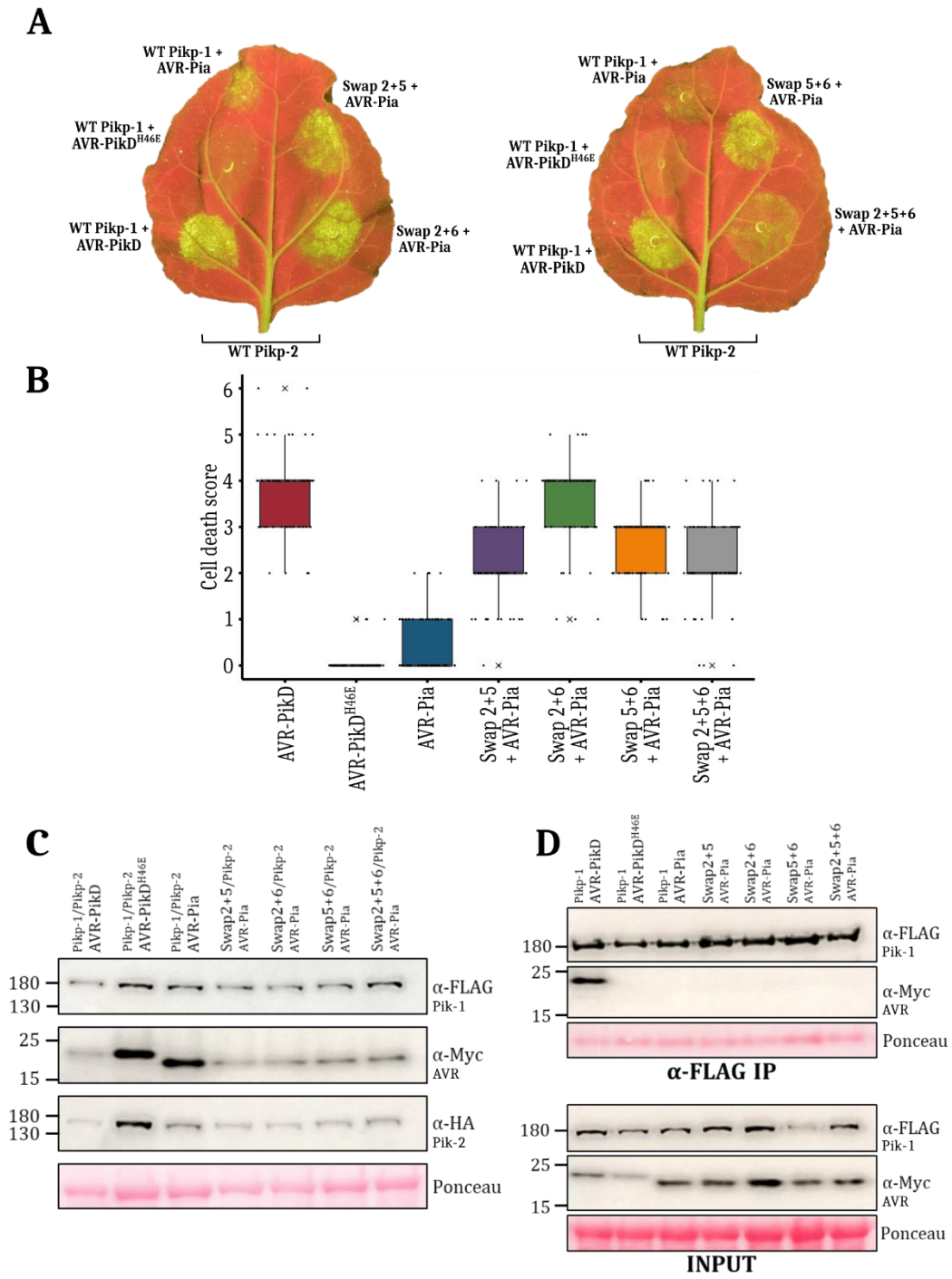


Figure 4.4-2: RGA5/Pikp-HMA swap combinations can enhance response to AVR-Pia.

A) Representative leaf images taken under UV light 5 dpi showing cell death progression at infiltration sites. **B)** Box plot showing cell death scores for 60 technical repeats over 3 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles. **C)** Western blot showing the expression of proteins at the infiltration sites. Blot shown is representative of 3 experiments showing similar results. **D)** Co-IP western blot showing protein input from plant extract (bottom) and outcome of immunoprecipitation of Pikp-1 or Pikp-1 mutants with FLAG magnetic beads (top).

4.5 RGA5-HMA in Pikp-1 background is autoactive

During testing of the RGA5/Pikp-HMA swaps, a further construct was generated that swapped the entire RGA5 HMA domain into the Pikp-1 background. The new construct was tested against AVR-Pia and an empty vector (EV) control and was found to be autoactive – i.e. there was a strong cell death response triggered by the NLR even in the absence of an effector (figure 4.5-1).

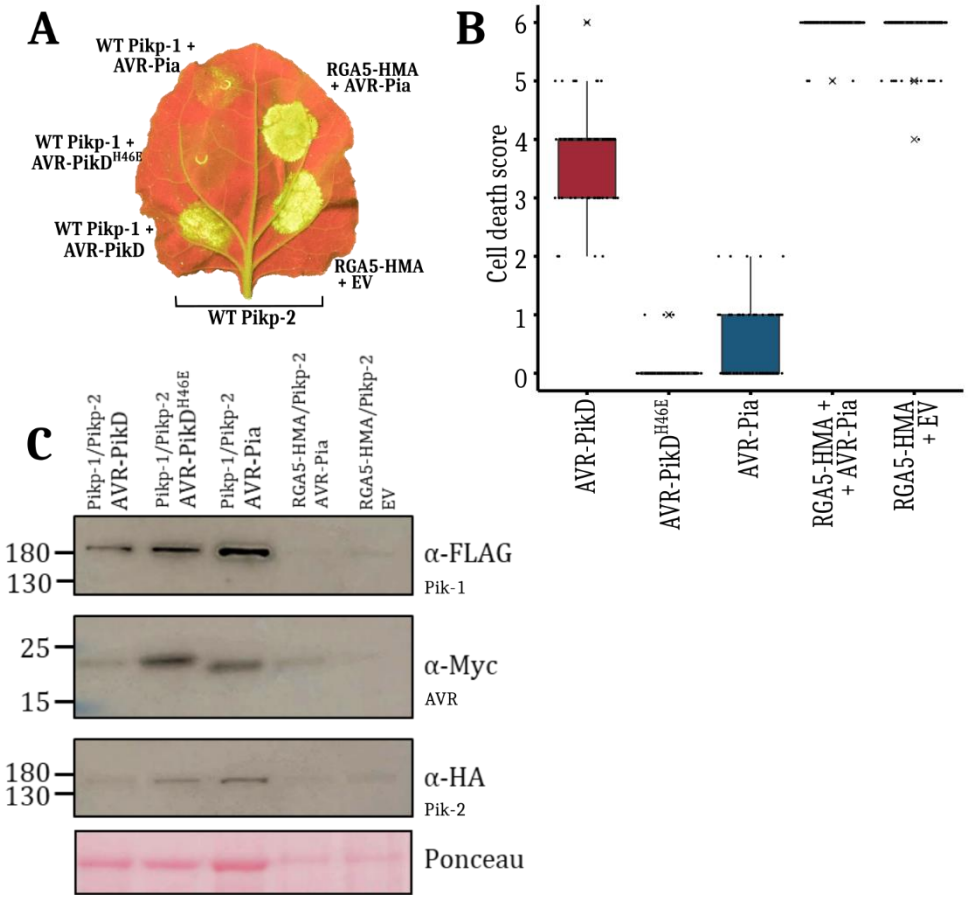


Figure 4.5-1: RGA5-HMA in Pikp-1 background triggers autoactive cell death in *N. benthamiana*.

A) Representative leaf images taken under UV light 5 dpi showing cell death progression at infiltration sites. **B)** Box plot showing cell death scores for 60 technical repeats over 3 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles. **C)** Western blot showing the expression of proteins at the infiltration sites. Blot shown is representative of 3 experiments showing similar results. **Note:** For brevity, 'RGA5-HMA' refers to the HMA domain of RGA5 inserted into the Pikp-1 background.

Knowledge of this autoactivity calls into question the increase in cell death seen when different RGA5/Pikp-HMA swaps were tested. It is possible that some of the swaps have

retained the features that lead to autoactivity, in which case the increase in cell death seen in these cases is actually unrelated to AVR-Pia.

To test whether the swaps that demonstrated a significant increase in cell death are actually autoactive, each of Swap2+6, Swap5+6 and Swap5 were re-tested against AVR-Pia, but were also tested against AVR-PikD (to investigate whether the swaps retained the ability to recognise AVR-PikD) and an empty vector control (to test for autoactivity). The results (figure 4.5-2) show (as discussed in previous sections) that each mutant can respond to AVR-Pia more strongly than WT Pikp, in the order Swap2+6 > Swap5+6 > Swap5. The data also indicates that each of the mutants retains the ability to respond to AVR-PikD, and in fact they all respond more strongly than WT Pikp. The most interesting observation is the response of each mutant to the empty vector control. For Swap2+6, the response to the empty vector is almost identical to the response to AVR-Pia, indicating that the increase in cell death response was due to autoactivity rather than enhanced recognition of the effector. However, for Swap5+6 the response to empty vector is much lower than the response to AVR-Pia and for Swap5 the response to EV is almost non-existent. This suggests that for both Swap5+6 and Swap5, the increased response to AVR-Pia is effector-dependant, rather than simply an autoactivity effect.

Figure 4.5-2C is a Western blot demonstrating that each construct has been successfully expressed in the leaf tissue. However, it should be noted that in the WT Pikp-1 and Swap5 samples, there is some contamination. When the blot was visualised with the α-Myc antibody there appeared to be some AVR-Pia contamination in the EV lanes. This contamination seems to have occurred after harvesting of the leaf tissue, because there was no cell death response in these samples on the leaves, which would have occurred if there had been AVR-Pia present at the infiltration site. The expression test was repeated with two more biological repeats, and AVR-Pia was occasionally present as a contaminant in the EV lanes, although not consistently in the same samples between repeats (data not shown). This is likely to have been an issue with processing the samples and running the Western blot, because there does not appear to have been any effect on consistency of the cell death scoring. This experiment needs repeating to obtain better quality data without contamination present.

As discussed for previous experiments, performing a co-IP indicates that none of WT Pikp, Swap2+6, Swap5+6 or Swap5 can consistently and irrefutably pull down AVR-Pia, although figure 4.5-2D shows some ambiguous bands on the blot that could possibly indicate some AVR-Pia interaction. However, each of the mutants tested (and WT Pikp-

1) were able to pull down AVR-PikD, indicating that the secondary structure swaps have not impeded this interaction in planta.

The autoactivity triggered by Swap2+6 is less strong than the autoactivity caused by the entire RGA5-HMA in the Pikp-1 background (figure 4.5-1), but the result could imply that α-helix 1 in the expected RGA5-HMA structure (Swap2) or β-strand 4 (Swap6) contributes heavily to the autoactivity. Given that the autoactivity of Swap5+6 is much lower than that of Swap2+6, it perhaps indicates that the greatest contribution to the autoactivity comes from Swap2, although it is not possible to confirm this through the data shown here. Indeed, neither the Swap2 nor Swap6 mutations show significant autoactivity alone, indicating that there could be an additive effect from the interplay of these two regions. Swap2 involves exchanging the non-functional metal binding loop in Pikp-HMA (Maqbool et al., 2015) for the equivalent region in RGA5-HMA. In Pikp-HMA, this loop is disordered in the crystal structure, and contains only one Cys residue in the region, rather than the required CxxC motif for metal-binding. In RGA5-HMA, the equivalent region is similarly missing the CxxC motif. It has been shown through additional work in the lab that this loop region has some involvement in autoactivity for related HMA proteins. sHMA1 is another (non-NLR) HMA domain-containing protein in rice and when the sHMA1 HMA domain was inserted into the Pikp-1 background, the resulting construct was strongly autoactive. Upon exchange of the metal-binding loop region of sHMA1 (namely '-SMPCEKS-') with the native loop from Pikp-1 ('-AMEGNNC'), the autoactivity was abolished (J. Maidment, unpublished). Although it was not possible within the timescale of the project, it would be interesting to find out whether mutating the metal binding loop in Swap2+6 would reduce the autoactivity, particularly given that the loop region in RGA5-HMA has greater sequence similarity to sHMA1 than to Pikp-1 (the sequence in this region is '-HMPCGKS'). Given that Swap5+6 has some low level of autoactivity, it is also possible that Swap6 contributes to the autoactive phenotype in a minor way, whereas Swap5 showed little or no autoactivity. However, more experiments would be needed to deconvolute the various intramolecular interactions taking place and determine the basis of the observed autoactivity, which may be caused by a combination of factors rather than one single region of the protein.

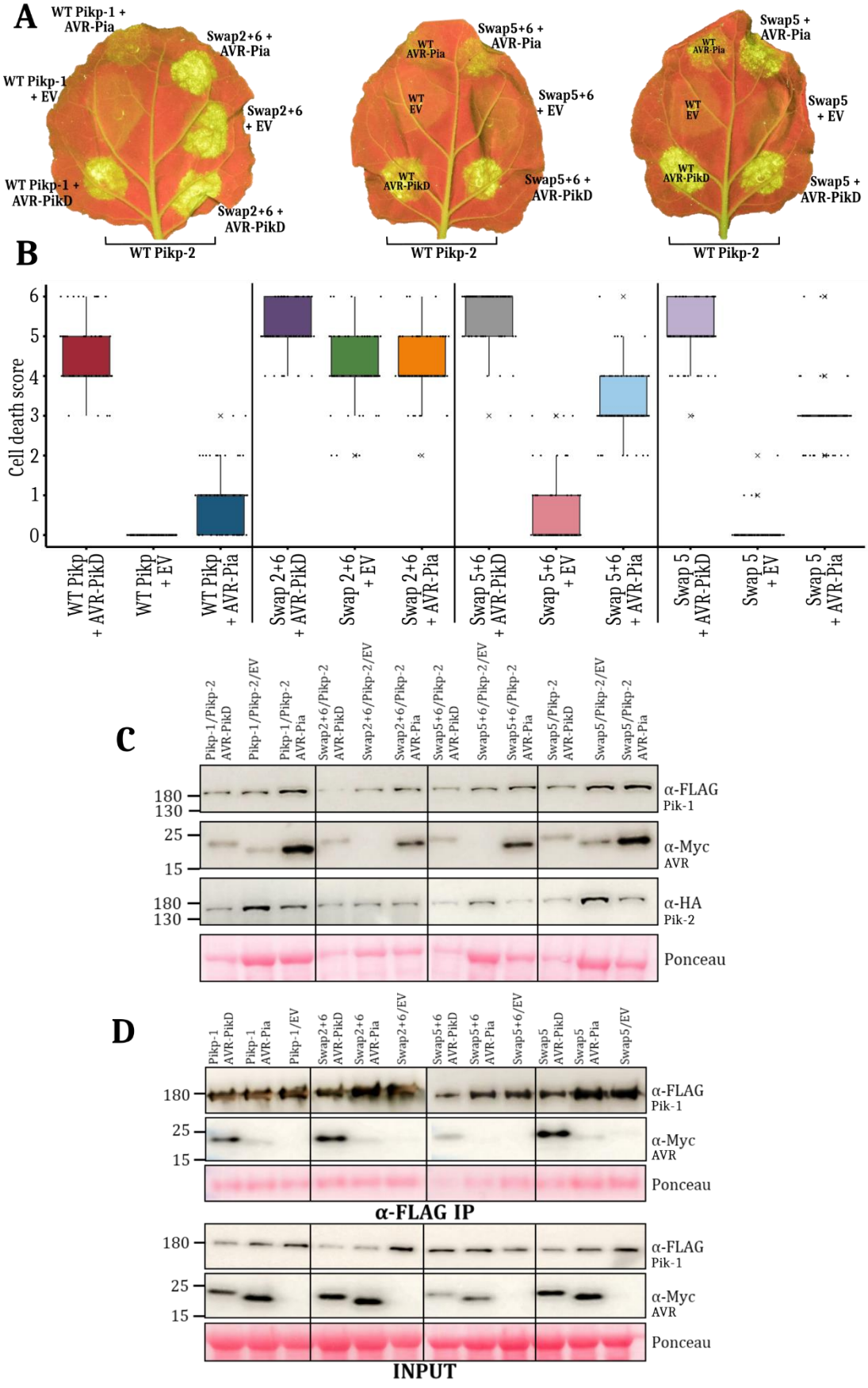

A
WT Pikp-1 + AVR-Pia
Swap2+6 + AVR-Pia
WT Pikp-1 + EV
Swap2+6 + EV
WT Pikp-1 + AVR-PikD
Swap2+6 + AVR-PikD
WT Pikp-2
WT AVR-Pia
Swap5+6 + AVR-Pia
WT EV
Swap5+6 + EV
WT AVR-PikD
Swap5+6 + AVR-PikD
WT Pikp-2
WT AVR-Pia
Swap5 + AVR-Pia
WT EV
Swap5 + EV
WT AVR-PikD
Swap5 + AVR-PikD
WT Pikp-2
B
Cell death score
0
1
2
3
4
5
6
WT Pikp + AVR-PikD
WT Pikp + EV
WT Pikp + AVR-Pia
Swap 2+6 + AVR-PikD
Swap 2+6 + EV
Swap 2+6 + AVR-Pia
Swap 5+6 + AVR-PikD
Swap 5+6 + EV
Swap 5+6 + AVR-Pia
Swap 5 + AVR-PikD
Swap 5 + EV
Swap 5 + AVR-Pia
C
Pikp-1/Pikp-2 AVR-PikD
Pikp-1/Pikp-2/EV
Pikp-1/Pikp-2 AVR-Pia
Swap2+6/Pikp-2 AVR-PikD
Swap2+6/Pikp-2/EV
Swap2+6/Pikp-2 AVR-Pia
Swap5+6/Pikp-2 AVR-PikD
Swap5+6/Pikp-2/EV
Swap5+6/Pikp-2 AVR-Pia
Swap5/Pikp-2 AVR-PikD
Swap5/Pikp-2/EV
Swap5/Pikp-2 AVR-Pia
180
130
α-FLAG Pik-1
25
15
α-Myc AVR
180
130
α-HA Pik-2
Ponceau
D
Pikp-1 AVR-PikD
Pikp-1 AVR-Pia
Pikp-1/EV
Swap2+6 AVR-PikD
Swap2+6 AVR-Pia
Swap2+6/EV
Swap5+6 AVR-PikD
Swap5+6 AVR-Pia
Swap5+6/EV
Swap5 AVR-PikD
Swap5 AVR-Pia
Swap5/EV
180
α-FLAG Pik-1
25
15
α-Myc AVR
Ponceau
α-FLAG IP
180
α-FLAG Pik-1
25
15
α-Myc AVR
Ponceau
INPUT

Figure 4.5-2: Some RGA5/Pikp-HMA swaps show an autoactive phenotype (see figure on previous page).

A) Representative leaf images taken under UV light 5 dpi showing cell death progression at infiltration sites. **B)** Box plot showing cell death scores for 58 technical repeats over 3 biological repeats scored at 5dpi, with all samples shown as dots and outliers indicated by 'x'. For each box, the centre line shows the median score and the edges of the box show the upper and lower quartiles. **C)** Western blot showing the expression of proteins at the infiltration sites. **D)** Co-IP western blot showing protein input from plant extract (bottom) and outcome of immunoprecipitation of Pikp-1 or Pikp-1 mutants with FLAG magnetic beads (top).

4.6 Purifying swaps in vitro

Alongside the testing of swaps for HR-like cell death in planta, the constructs that had conferred increased response to AVR-Pia were also cloned for heterologous expression and purification in vitro, with the aim to characterise them and discover whether the increased response in planta correlated with an increased binding in vitro. In total, Swap2, Swap5, Swap6, Swap2+5, Swap2+6, Swap5+6 and Swap2+5+6 were all cloned into the pOPINM vector and expressed and purified from *E. coli* SHuffle cells using the same method as for WT Pikp-HMA. All mutants were successfully expressed (figure 4.6-1), although the yields and stability of each protein varied quite considerably.

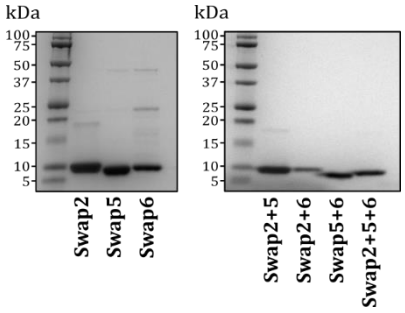


Figure 4.6-1: SDS-PAGE showing RGA5/Pikp-HMA swaps purified from *E. coli*.

SDS-PAGE showing purified protein for a selection of RGA5/Pikp-HMA swaps.

Intact mass spectrometry analysis for each mutant showed that full-length HMA was expressed in each case, but it was noted that for some of the mutants containing Swap5, the mass spectrometry analysis showed a peak at a molecular weight that was 2 Daltons lower than the expected full-length protein – this was true for Swap5 and Swap5+6 (see table 10).

Table 10: Comparison between the expected and actual molecular weights of purified swap proteins, measured using intact mass spectrometry.

Protein name	Expected mass (Da)	Actual mass (Da)	Difference (Exp - Act)
Swap2	8327.9	8327.6	0.3
Swap5	8392.8	8390.4	2.4
Swap6	8189.8	8189.7	0.1
Swap2+5	8358.9	8358.8	0.1
Swap2+6	8155.9	8155.8	0.1
Swap5+6	8220.8	8218.2	2.6
Swap2+5+6	8186.9	8186.8	0.1

A decrease of 2 Da shown in intact mass spectrometry can indicate the presence of a disulphide bond – when two cysteines become oxidised, the loss of two hydrogen atoms leads to a reduction of 2 Da in the molecular weight. WT Pikp-HMA only contains one cysteine residue, so is unable to form any intramolecular disulphide bonds, but replacement of Pikp-HMA α-2 with α-2 from RGA5-HMA (i.e. Swap5) inserts a new cysteine residue. Using Phyre2 to predict the structure of Swap5 (as described in section 4.3), it appears that the new cysteine residue has been positioned near to the cysteine present in α-1 (figure 4.6-2). Therefore, although it has not been proven, this might indicate that Swap5 can become stabilised in vitro due to the formation of an intramolecular disulphide bond. Mutants that contain Swap2 in addition to Swap5 (i.e. Swap2+5 and Swap2+5+6) do not show the 2 Da decrease in molecular weight, suggesting that there is no intramolecular disulphide bond formed. Although α-1 from RGA5-HMA (Swap2) also contains a cysteine residue, it is positioned in a slightly more N-terminal position in the sequence, so when Swap2 is inserted along with Swap5, the cysteine residues appear to be pushed further apart spatially in the structure (figure 4.6-2) and may thus be unable to form the disulphide bond. Note that because the structures shown in figure 4.6-2 are only models predicted using Phyre2, they cannot give a definitive picture of how the residue side chains would be positioned within the actual 3D structure, but can help support the hypothesis in the absence of real structural information.

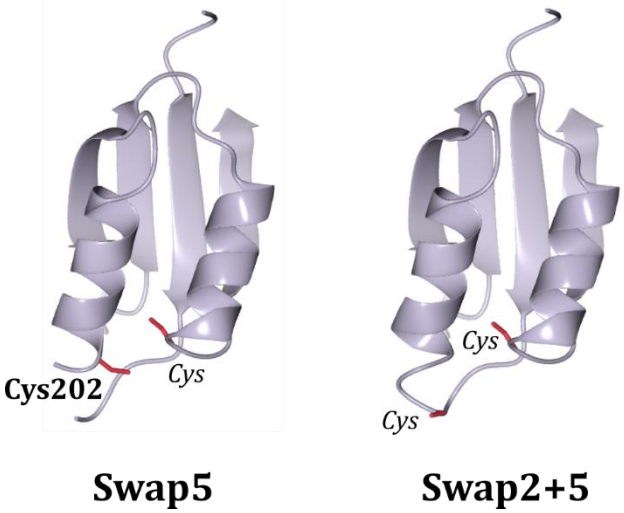


Figure 4.6-2: Phyre2 models of RGA5/Pikp-HMA Swap5 and Swap2+5.

Models of Swap5 and Swap2+5 generated by a one-to-one threading method using Phyre2, with the template structure as Pikp-HMA (PDB accession 5A6P). Cysteine side chains are shown in red, with the native Cys202 from Pikp-HMA shown in bold, and the cysteines that have been introduced by RGA5-HMA swaps shown in italics.

This preliminary characterisation of the swap proteins in vitro may help with understanding their intermolecular interactions, as more in planta and in vitro evidence of their behaviour is uncovered.

4.7 Swap2+6 has increased affinity for AVR-Pia

Having heterologously expressed and purified each swap that showed an increased cell death response to AVR-Pia, the affinities for each interaction could then be determined in vitro. To reduce the sample size, only the mutants with the greatest increase in response were tested – namely Swap2+6, Swap5+6 and Swap5. The affinities were tested using SPR, as described in Chapter 3. Given that, as described in section 4.5, it appeared that much of the response of Swap2+6 to AVR-Pia was caused by autoactivity rather than being effector-dependant, it was surprising to observe that the binding affinity of Swap2+6 for AVR-Pia had increased to 80 % of R_{max} (at 100 nM analyte concentration), compared to 43 % for WT Pikp-HMA (figure 4.7-1A). Perhaps even more intriguingly, Swap5+6 and Swap5, which both showed an enhanced response in planta coupled with lower autoactivity, were effectively unable to bind to AVR-Pia in vitro. This lack of correlation between in planta response and binding in vitro suggests that there are additional effects contributing to cell death in planta in these cases, aside from a simple

change in the interaction between effector and HMA domain. Further experiments would be needed to understand these effects.

Using a larger range of analyte (HMA) concentrations for analysis (figure 4.7-1B), it was possible to obtain a value for the equilibrium dissociation constant (K_D) of the Swap2+6/AVR-Pia interaction (figure 4.7-1). The value obtained was 10.7 nM, and while it was unfortunately not possible to quantify K_D for the WT Pikp-HMA/AVR-Pia interaction (see Chapter 3, section 3.3), this value is comparable to the affinity of WT Pikp-HMA for its 'matched' effector AVR-PikD (4.0 nM determined in this experiment, 5.9 nM determined in de la Concepcion, 2018). These SPR results show that the Swap2+6 mutant can interact more strongly with AVR-Pia than WT Pikp-HMA, and thus an increased affinity has been engineered through these mutations. However, the Swap2+6 autoactivity in planta must be addressed to make this mutation a viable construct for further exploration.

Further investigations have shown that it is possible to purify a Swap2+6/AVR-Pia complex, using the same technique as for the Pikp-HMA/AVR-Pia complex. This purified protein was used for crystallography trials, and some small crystals were obtained from a commercial screen. Diffraction of the crystals was poor (data not shown) but this provides a promising first step to obtaining the structure of the Swap2+6/AVR-Pia complex, and discovering how the interaction of the two proteins has been strengthened compared to WT Pikp-HMA.

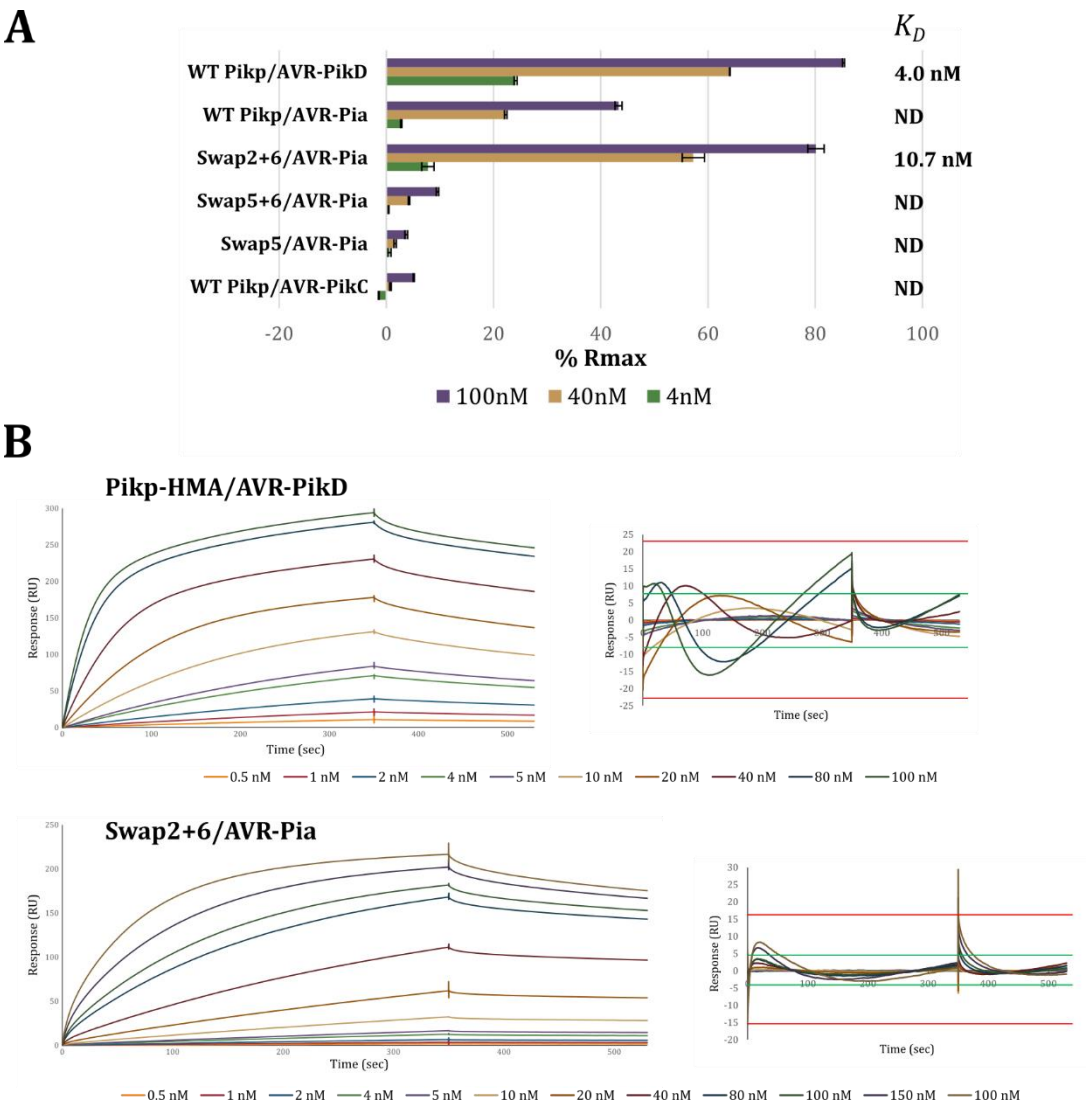


Figure 4.7-1: In vitro interactions of RGA5/Pikp-HMA swaps with AVR-Pia.

A) Surface plasmon resonance bar chart showing %Rmax, the percentage of theoretical maximum response for HMA binding to immobilised effector, at three different concentrations. Each measurement represents an average of three results, with error bars showing standard deviation. Chart shown is representative of three separate experiments. Where K_D values were obtainable, they are shown adjacent to the chart, 'ND' = Not Determined. **B)** Multi-cycle kinetics data for the interaction of Pikp-HMA with AVR-PikD (assuming a 2:1 fit model) and Swap2+6 with AVR-Pia (assuming a 1:1 fit model). Plots show the response units at each different analyte (HMA) concentration, with residuals plots alongside, showing difference between experimental and calculated data (green and red lines on residuals plots show data acceptability limits determined by the Biacore software). Data shown is representative of three separate experiments.

4.8 Discussion

The ability to engineer plant NLR proteins to respond more strongly to pathogen effectors is an exciting emerging technology. Armed with the knowledge of the AVR-Pia/Pikp-HMA crystal structure (Chapter 3), the work described here aimed to increase the response of Pikp to AVR-Pia in planta by engineering the HMA domain.

The first step was using the AVR-Pia/Pikp-HMA structure to design point mutations that might strengthen the binding. This proved challenging, as the proteins are held together predominantly by backbone interactions at the interface. The two mutants generated, Pikp-HMA^{A220E} and Pikp-HMA^{G214E}, were both unsuccessful at increasing the cell death response to AVR-Pia. The A220E mutation was made based on the knowledge that the residue in RGA5-HMA corresponding to the Pikp-HMA alanine was a dissimilar glutamate residue. However, the mutation appeared to have a negative impact on response in planta, indicating that this glutamate does not have a significant impact on the RGA5-HMA/AVR-Pia interaction, or that the impact is made in conjunction with other polymorphic residues in the region. The mutation of the small alanine residue to the bulky glutamate may in fact have introduced a steric clash that prevented a binding event from taking place. The second mutation, G214E had no effect on cell death, indicating that either the proposed new interaction at a loop region of the HMA was not formed, or did not have an impact on effector recognition.

In order to make a difference to the Pikp response to AVR-Pia, larger changes in the HMA domain were necessary. Swapping regions from the RGA5-HMA domain into Pikp-HMA proved a successful strategy for increasing the response in planta, and combining the successful swaps into new mutants enhanced this response even further. However, it was puzzling that the enhanced cell death response did not lead to an enhanced interaction in planta between the mutant Pikp proteins and AVR-Pia (as shown by co-IP). Equally puzzling was the fact that Swap5 and 6, which appeared to increase the immune response, were not at the expected interface with AVR-Pia. It was thought that perhaps the increase in cell death was caused by an intramolecular NLR interaction or an intermolecular interaction with another member of the signalling pathway.

Some light was shed on these puzzles when it was discovered that RGA5-HMA in the Pikp-1 background was strongly autoactive, triggering an immune response even in the absence of an effector. With this information, it was apparent that some of the increased 'immune response' may simply be due to autoactivity triggered by the insertion of part of RGA5-HMA. Upon investigation, it was discovered that the most 'successful' mutant, Swap2+6, fell into this category, and its entire cell death response in the presence of AVR-Pia could be accounted for by autoactivity. However, Swap5+6 and Swap5 showed a much weaker autoactive phenotype, indicating that only certain regions of RGA5-HMA are needed to trigger the autoactivity. With the experiments carried out so far, it is not possible to deconvolute these observations, and more work will be needed to understand which region is responsible for the autoactive effect. This would require testing all the

mutants (Swap1-6 and each swap combination) against the empty vector and investigating the levels of cell death for each. The interaction may in fact be more complicated than a single region, and could involve making further mutations to overcome the autoactivity. Given the similarity between the metal-binding loops of sHMA1 and RGA5-HMA, which are both known to cause autoactivity in the Pikp-1 background, the first logical test would be to mutate the metal-binding loop in Swap2+6 back to the native Pikp-1 loop, to see if that abolishes autoactivity, as seen for sHMA1 (J. Maidment, unpublished). For further investigation, it would be advantageous to repeat the autoactivity experiments using an empty vector in place of Pikp-2, to ensure that this autoactive phenotype is due to signalling transduced through the helper NLR, and not due to an additional effect.

Another complication encountered during the testing of the mutant Pikp proteins in planta was the difficulties in obtaining clear results for their interaction with AVR-Pia. Despite an enhanced cell death response, co-IP experiments did not show a gain-of-interaction for the mutants with AVR-Pia. However, as with WT Pikp and AVR-Pia (Chapter 3) it was difficult to get consistent results – for some biological repeats there appeared to be an interaction, whereas for other repeats no interaction could be seen. While it appears reasonable to conclude that any interaction between the proteins is weak, more repeats of the co-IP experiments would be needed to reach a proper consensus.

The three swap proteins that showed the greatest increase in cell death response to AVR-Pia were also tested for their interaction with AVR-Pia in vitro, via SPR. The results showed that there was little correlation between the in planta and in vitro observations. Despite Swap2+6, Swap5+6 and Swap5 all showing a strongly increased cell death response in planta, Swap5+6 and Swap5 were completely unable to bind AVR-Pia in SPR. Swap2+6, on the other hand, showed a strong increase in binding affinity – the K_D for the interaction was comparable to the binding affinity of the AVR-PikD/Pikp-HMA complex.

The disparity between in vitro and in planta studies observed with these swaps requires more investigation. It is possible that Swap5+6 and Swap5 can bind AVR-Pia in planta, but cannot bind the effector in vitro due to a factor such as the particular technique used or the different environment. Swap2+6 appears to be able to bind AVR-Pia strongly, so this remains the most promising candidate. If further investigations are able to prevent the autoactive phenotype, it is possible that Swap2+6 will still respond strongly to AVR-Pia. One priority for future investigations would be the improvement of Swap2+6/AVR-Pia complex crystals. If the structure of the complex can be obtained, it might be possible

to discover which residues are responsible for enhancing the binding affinity. If only a few residues are involved, these single mutations could be introduced back into the Pikp-HMA in the native Pikp-1 background. Hopefully this would lead to the same strength of binding, but without introducing the autoactivity from RGA5-HMA.

An additional consideration when studying these responses is that modifying the HMA domain of Pikp may not generate a truly 'RGA5-like' response to AVR-Pia. Although a superficially enhanced HR-like cell death response may be seen in planta, it is not clear how this is being generated or what the intermediate steps might be. One immediately apparent feature of RGA5 is that the HMA domain is located at the C-terminus, rather than centrally (after the coiled-coil domain) as seen in Pikp-1. Coupled with the fact that Pikp-HMA binds AVR-Pia at a different interface compared to AVR-PikD, it is conceivable that RGA5 and Pikp-1 are sensing the effectors in a different manner and perhaps generating different downstream signalling responses after perception. Indeed, given that the RGA5/RGA4 NLR pair appears to operate in a different manner to the Pik-1/Pik-2 pair i.e. by negative regulation rather than cooperation, it seems likely that the effector recognition and signalling are functioning in different ways. This could be one method by which the plant maintains a robust defence against *M. oryzae*. If two NLRs containing the same type of integrated domain can function through different downstream mechanisms, this might help to bypass any downstream signalling components that could be targeted by effectors. This is currently only speculation, and much more work would be needed to unravel the mechanisms of these NLR pairs. However, it should be noted that simply enhancing the strength of binding between HMA and effector may not be showing a full appreciation of the complexities of the signalling involved.

Nevertheless, with some modifications and further work, the Swap2+6 mutant could represent an example of an engineered integrated HMA domain with an enhanced capacity to respond to *M. oryzae* effectors.

PexRD24 targets a host phosphatase

Investigating the interaction of the *P. infestans* effector PexRD24 with its host target, PP1c

5.1 Introduction

In this chapter, the focus moves away from *M. oryzae* and MAX effectors to another important crop pathogen and another family of effectors – the RxLR effector family from *P. infestans*. RxLR effectors were discussed in Chapter 1 and differ from the MAX effectors in that they are defined not by a conserved structure, but by a conserved sequence. Nevertheless, as described in Chapter 1, many of the RxLR effectors are predicted to have conserved structural features, in the form of the WY domain fold (Chapter 1).

The RxLR effector that will be investigated here is PexRD24 (figure 5.1-1), also known as Pi04314 (PITG_04314) in some publications. This project was begun as a collaboration with the lab of Paul Birch (at The James Hutton Institute, Dundee), who have published their own work on this effector (discussed below). Before the commencement of the project, the effector was subjected to a yeast two-hybrid screen and a putative host interacting protein was identified as protein phosphatase 1 (Boevink, Wang, et al., 2016). Aside from the possible phosphatase host target, little was known about PexRD24 prior to the start of the project.

Phosphatase enzymes are vital in all kingdoms of life, as the regulation of phosphorylation state is important for correct post-translational modification or activation of many cellular components. In plants, phosphatases are known to be important in defence against both abiotic and biotic stress (Pais et al., 2009), and play a role in immune responses to some microbial pathogens. One recent example of the involvement of phosphatases in plant immunity is the enzyme PP2C38, which negatively regulates BIK1, an important signalling component for the PRRs FLS2 and EFR. Following perception of a possible pathogen, PP2C38 is phosphorylated, causing it to dissociate and allow full BIK1 activation (Couto et al., 2016). A number of other examples of phosphatase involvement in diverse aspects of plant immunity have been characterised in recent years (e.g. (Schweighofer et al., 2007; Widjaja et al., 2010)) and an effector from the bacterial pathogen *Pseudomonas syringae*, HopAO1, uses phosphatase activity directly to manipulate immune signalling (Macho et al., 2014). Exploitation of plant protein phosphorylation processes by pathogen effectors demonstrates the importance of phosphorylation to plant defences.

The serine/threonine phosphatase PP1 also controls diverse processes in plants (Lin et al., 1998), and regulates events such as stomatal opening (Takemiya et al., 2006). The catalytic subunit PP1c is controlled by a diverse range of regulatory subunits, which physically interact with the surface of the catalytic subunit through a conserved R-V-x-F

motif, where R is arginine, V is valine, x indicates any amino acid and F is phenylalanine (Cohen, 2002). Later research has suggested that the regulatory subunits can interact with the catalytic subunit through a range of different surface patches and motifs (Heroes et al., 2013). These regulatory subunits have wide-ranging functions including targeting the catalytic subunit to different subcellular localisations, such as the nucleus (Allen et al., 1998) and targeting to specific substrates (Bollen et al., 2010). The binding of a regulatory subunit may not cause significant global structural changes to the catalytic subunit but may cause alterations in the substrate binding groove or block access to certain binding sites (Verbinnen et al., 2017). This can enhance or inhibit enzyme activity towards a particular substrate.

PP1c has been extensively studied in mammals, and several crystal structures have been solved with the catalytic subunit in complex with inhibitors or regulatory subunits (for example (Choy et al., 2014; Hurley et al., 2007; Terrak et al., 2004)). The earliest structures of PP1c helped to elucidate its catalytic mechanism and shape an understanding of how the enzyme was regulated (Egloff et al., 1995; Goldberg et al., 1995). Metal ions were found bound to the catalytic subunit and have an important functional rather than structural role (Goldberg et al., 1995). Their identities are thought to be Zn^{2+} and Fe^{2+} in native PP1c (although Fe^{3+} and Mg^{2+} cannot be ruled out) but are substituted for Mn^{2+} when the enzyme is heterologously expressed from bacteria (Heroes et al., 2015). These structural studies have indicated that dephosphorylation is catalysed in a single step, and highlighted the importance of a particular histidine residue found at the catalytic site (Egloff et al., 1995). This histidine residue is the basis of the phosphatase-dead mutant described in later experiments.

PP1c purified from bacteria has been shown to exhibit different properties and activities compared to native PP1c (Egloff et al., 1995). PP1c produced in *E. coli* is often highly insoluble (Peti et al., 2013), perhaps one reason why the structure of an apo PP1c has not been solved (Ragusa et al., 2010). A number of different strategies, such as the use of solubility tags and molecular chaperones, have been shown to improve yields of soluble protein (Peti et al., 2013).

Since work began on this project, a collaborating group have published their findings on how the effector PexRD24 manipulates PP1c (Boevink, Wang, et al., 2016). PexRD24 contains the RVxF sequence motif of a PP1c regulatory subunit (figure 5.1-1) and was shown to be re-localising PP1c out of the host nucleolus in in planta studies. One hypothesis given was that the effector may be directing PP1c to an alternative substrate in the nucleoplasm (Boevink, Wang, et al., 2016). It was shown that virus-induced gene

silencing of PP1c isoforms in the host reduced the proliferation of *P. infestans* (Boevink, Wang, et al., 2016) and this observation led PP1c to be described as a susceptibility factor (Boevink, McLellan, et al., 2016); i.e. a host protein that aids pathogen infection (van Schie & Takken, 2014). Another recent example (Murphy et al., 2018) showed that a host kinase enzyme acts as a susceptibility factor that can be targeted by the RxLR effector Pi17316 from *P. infestans*.

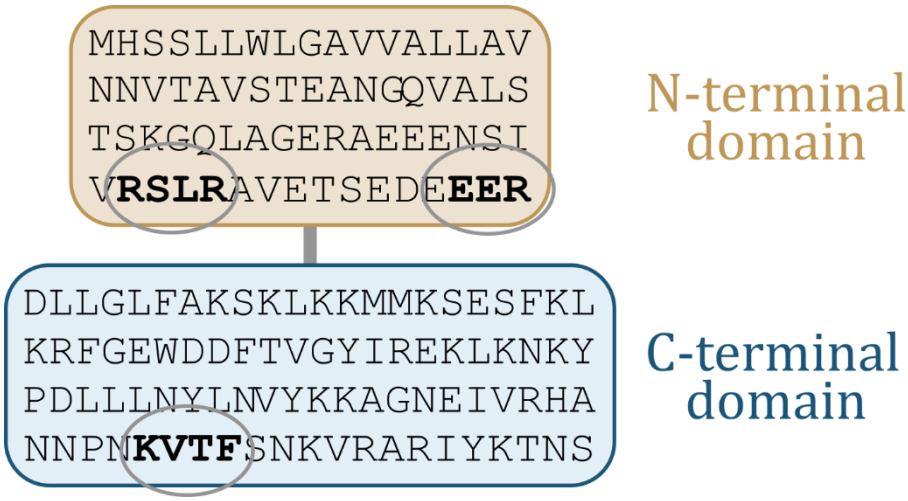


Figure 5.1-1: Diagram highlighting sequence features of PexRD24.

Diagram showing the amino acid sequence of the N-terminal (translocation) and C-terminal (effector) domains of PexRD24. The RxLR and EER motifs are circled in the N-terminal domain, and the RVxF motif (in this case KVTF) is shown in the C-terminal domain. Letters indicate the single amino acid code.

The work described in Boevink et al. has provided many clues about the purpose of PexRD24 as an effector but have not been able to elucidate its exact mode of action. In this chapter, investigations into obtaining stable, soluble PexRD24 and PP1c proteins in vitro are discussed, along with efforts to understand whether PexRD24 is affecting the enzyme activity of PP1c.

5.2 PP1c isoforms can be expressed from *E. coli*

Four different isoforms of PP1c were identified as possible PexRD24 interactors from a yeast-2-hybrid screen of a library made from infected potato cDNA (Boevink, Wang, et al., 2016), although only three were discussed in a later publication by Boevink et al. (Boevink, Wang, et al., 2016). As briefly discussed in section 5.1, working with phosphatase enzymes in vitro is notoriously difficult, so the first task was to discover whether the proteins could be purified in a stable and soluble manner for further study.

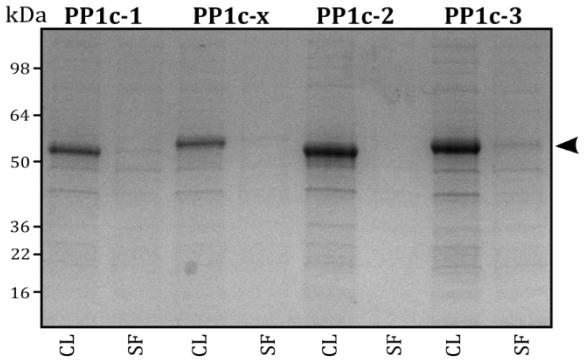
Each PP1c construct was a different variant of the catalytic subunit of protein phosphatase 1. For these in vitro studies, only the C-terminal effector domain was used, the construct did not include the N-terminal RxLR-containing domain, which is predicted to be disordered in RxLR effectors (Boutemy et al., 2011). The predominant differences between each PP1c isoform are found at the termini (figure 5.2-1A), and these differences could have an effect not just on the interaction with PexRD24, but also on the behaviour of the protein in solution, such as its solubility. Therefore, each construct was subjected to a small-scale expression test, in the expression vector pOPINS3C (yielding a cleavable N-terminal SUMO solubility tag on the protein of interest), BL21(DE3) *E. coli* and LB media. Each construct was expressed successfully, but there was little to none of this protein present in the soluble fraction (figure 5.2-1B). The only isoform that appeared to have a noticeable level of solubility was PP1c-3. After trials using two different vectors, two *E. coli* strains and three different types of expression media (figure 5.2-1C), it was not possible to increase protein solubility for the other isoforms. Therefore, PP1c-3 was scaled up in the original expression conditions (pOPINS3C, BL21(DE3), LB) (figure 5.2-1D).

However, despite the purification of soluble protein there were continuing difficulties. For example, the purified protein began to precipitate out of solution over time and intact mass spectrometry analysis indicated that there was significant degradation of the protein, with residues largely being degraded from the C-terminus. The expected molecular weight of PP1c-3 (following cleavage of the His-SUMO tag) was 36885.4 Da. Intact mass spectrometry analysis showed that there was indeed a peak at 36884.9 Da, corresponding to full-length protein, but there were other species identified from the analysis (figure 5.2-2) at 36189.4 Da, 35147.9 Da and 34492.0 Da, which correlated with PP1c-3 with residues removed from the C-terminus and, in the case of the 34492.0 Da species, from the N-terminus as well. These truncations will be discussed later in more detail.

A

```
PP1c-3  ----------------MDQNVLDDIITRLLEVK-GKPGKQVVLTEAEIKQLCLVAKETFLR
PP1c-2  ----------------MDPAAVDRIIEKLLEVRLSKPGKLVQLSESEIKQLCVSSRDIFVK
PP1c-1  --------MAQNGQGIEPAVLDDIINRLLEFRNARTVRQVQLSEAEIRSLCTAARAIFLQ
PP1c-x  MAQNEHQQQQQQQGLIEPAVLDDIINRLLEFRNARTVRQVQISEAEIRSLCSASREIFLQ
                  ::  .:* ** :***.: .:  : * ::*:**:.**  ::  *::

PP1c-3  QPNLLELEAPIKICGDIHGQYSDLLRLFEYGGLPPQSNYLFLGDYVDRGKQSLETICLLL
PP1c-2  QPNLLELEAPIKICGDIHGQYSDLLRLFEYGGFPPKANYLFLGDYVDRGKQSLETICLLL
PP1c-1  QPNLLELEAPIKICGDIHGQYGDLLRLFEYGGFPPQSNYLFLGDYVDRGKQSLETICLLL
PP1c-x  QPNLLELEAPIKICGDIHGQYGDLLRLFEYGGFPPEANYLFLGDYVDRGKQSLETICLLL
        **********************.**********:**::************************

PP1c-3  AYKIKYPENFFLLRGNHECASINRIYGFYDECKRRFNVRLWKIFTDCFNCLPVAALIDEK
PP1c-2  AYKIKYPENFFLLRGNHECASINRIYGFYDECKRRFNVKLWKSFTDCFNCLPVAALIDEK
PP1c-1  AYKIKYPENFFLLRGNHECASINRIYGFYDECKRRFNVRLWKVFTDCFNCLPVAALIDDK
PP1c-x  AYKIKYPENFFLLRGNHECASINRIYGFYDECKRRFNVKLWKCFTECFNCLPVAALIDEK
        ****************************************:*** **:************:*

PP1c-3  ILCMHGGLSPDLNHLDQIRGLQRPTDVPDAGLLCDLLWSDPSKEVQGWGMNDRGVSYTFG
PP1c-2  ILCMHGGLSPDLSSLDQIRNLPRPTAIPDTGLLCDLLWSDPGKDVKGWGMNDRGVSYTFG
PP1c-1  ILCMHGGLSPDLTDMDLIRNLPRPTDVPDSGLLCDLLWSDPSREVKGWGMNDRGVSYTFG
PP1c-x  ILCMHGGLSPDLTNLDQIRNLPRPTDVPDSGLLCDLLWSDPSREVKGWGMNDRGVSYTFG
        ************. :* ** * *** :**:***********.::*:***************

PP1c-3  ADKVTEFLEKHDLDLICRAHQVVEDGYEFFANRQLVTVFSAPNYCGEFDNAGAMMSVDET
PP1c-2  PDKVSEFLSKHDLDLVCRAHQVVEDGYEFFAERQLVTIFSAPNYCGEFDNAGAMMSVDEN
PP1c-1  PDKVAEFLMQQDMDLVCRAHQVVEDGYEFFAERQLVTVFSAPNYCGEFDNAGAMMSVDEN
PP1c-x  PDKVAEFLMQHDMDLVCRAHQVVEDGYEFFAERQLVTIFSAPNYCGEFDNAGAMMSVDES
         ***:*** ::*:**:****************:*****:**********************.

PP1c-3  LMCSFQILKPADKKSKFSFGSTTTAKPGSPTGMKSFFNSKA
PP1c-2  LMCSFQILKPAEKKNKFMM----------------------
PP1c-1  LMCSFQILKPADRKPRFL----------------------
PP1c-x  LMCSFQILKPTDRKPRFL----------------------
        **********:::* :*
```

B



C

	LB	AIM	EnPresso
BL21(DE3)	pOPINF pOPINS3C	pOPINS3C	pOPINS3C
SHuffle	pOPINS3C	pOPINS3C	

D

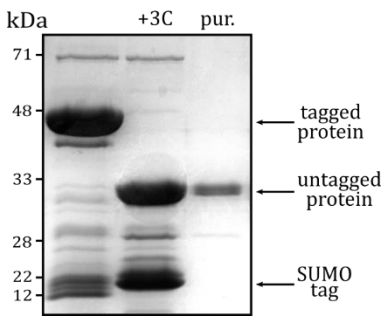


Figure 5.2-2: Comparison of PP1c isoforms (see figure on previous page).

A: Sequence alignment of the 4 PP1c isoforms. **B:** SDS-PAGE analysis of expression test for each PP1c isoform, tagged with an N-terminal SUMO tag. CL = crude Cell Lysate, SF = Soluble Fraction. Arrow indicates approximate expected location of each PP1c protein on the gel. **C:** Diagram indicating different expression conditions screened for PP1c isoforms, with different culture media in the top row, and different *E. coli* strains in the first column. Each cell shows the expression vector used alongside the corresponding media/ strain. AIM = Autoinduction media, LB = Lysogeny broth. **D:** SDS-PAGE analysis showing the purification of soluble PP1c-3 from pOPINS3C, BL21(DE3) and LB media, +3C = sample after addition of 3C protease, pur. = purified protein.

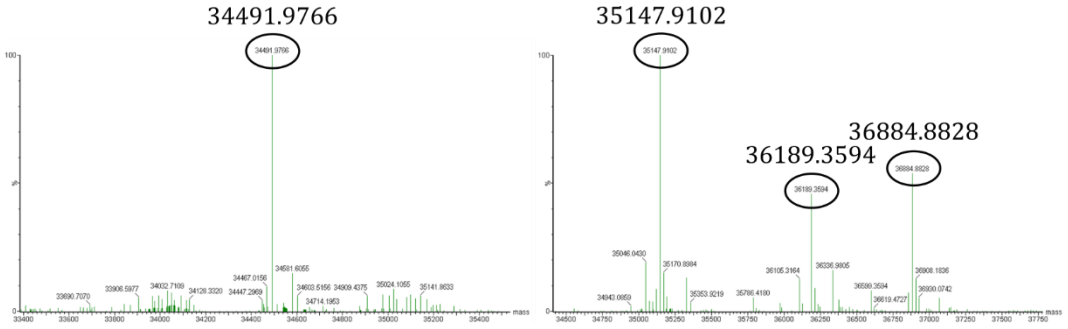


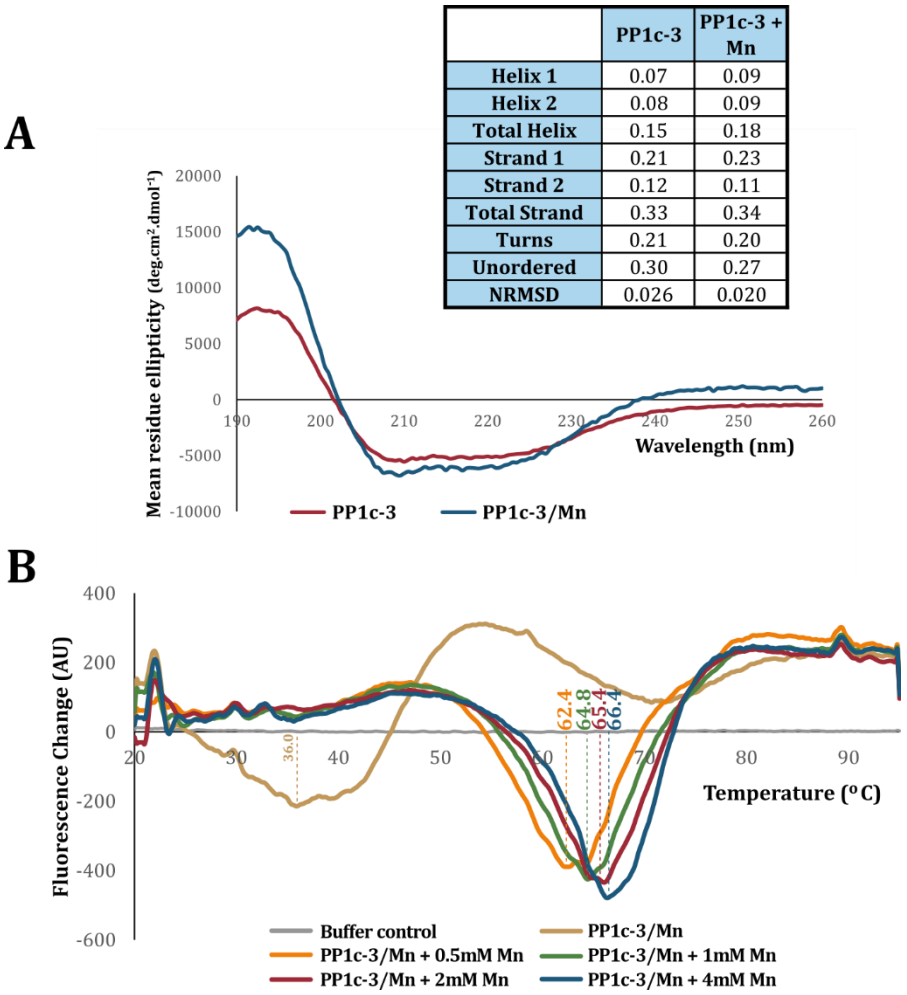
Figure 5.2-1: Intact mass spectrometry analysis of PP1c-3.

Spectra showing different species identified by peaks in intact mass spectrometry analysis. Peaks of interest are circled, with the exact mass (in Daltons) labelled above each for clarity.

Although other attempts were made to obtain soluble protein for the remaining PP1c isoforms, these continued to be unsuccessful and will not be discussed further.

As mentioned in section 5.1, PP1c purified from *E. coli* is a manganese-containing enzyme, so it was considered that the addition of Mn^{2+} ions during protein production may assist in the production of functional, well-folded protein. For an initial trial 1 mM Mn^{2+} (in the form of $MnCl_2$) was added to the *E. coli* growth media, prior to induction of PP1c expression. The purified protein behaved the same with or without the addition of Mn^{2+}, i.e. it was relatively unstable in solution. The secondary structure features of PP1c-3 purified with and without the additional manganese were assessed using circular dichroism (CD) to determine whether the metal ions had made a difference to the overall structure of the protein. When overlaid (figure 5.2-3A), the far-UV spectra for the proteins appear very similar, indicating that the manganese has not changed the global structure of the protein. Although there were no distinct features in the spectra, the analysis appears to show a typical CD signature for a folded protein. The online analysis DichroWeb tool (Whitmore & Wallace, 2004) was used to assign the proportions of different secondary structure elements (figure 5.2-3A), and the results show a mixture of α-helical and β-strand elements within the protein.

To further investigate the manganese requirements for PP1c-3, the purified protein (purified with additional Mn^{2+} in the media) was tested for thermal stability using differential scanning fluorimetry (DSF), also known as the thermofluor or thermal shift assay. This technique studies the stability of a protein by monitoring its temperature-induced denaturation using the fluorescent dye SYPRO® Orange. SYPRO® Orange binds to hydrophobic residues within the protein, and as the protein is denatured (unfolded) with increasing temperature, the measured fluorescence is enhanced as the hydrophobic residues are exposed. When the derivative of fluorescence change versus temperature is plotted, the lowest point on the curve indicates the melting temperature (T_m).



	PP1c-3	PP1c-3 + Mn
Helix 1	0.07	0.09
Helix 2	0.08	0.09
Total Helix	0.15	0.18
Strand 1	0.21	0.23
Strand 2	0.12	0.11
Total Strand	0.33	0.34
Turns	0.21	0.20
Unordered	0.30	0.27
NRMSD	0.026	0.020

Figure 5.2-3: Purified PP1c is stabilised by the addition of manganese.

A: Circular dichroism (CD) analysis comparing the secondary structure features of PP1c-3 expressed with and without Mn in the media, plot shows far-UV spectra for each. Numbers indicate the fraction of each type of secondary structure, where helix and strand assignments are divided into regular (helix/strand 1) and distorted (helix/strand 2) (Sreerama et al., 1999). NRMSD is the normalised root mean squared deviation. Secondary structure assignments were carried out using DichroWeb, solutions were obtained from the CDSSTR method using reference dataset 7 (Sreerama & Woody, 2000). **B:** Differential Scanning Fluorimetry (DSF) analysis showing the change in protein stability when different concentrations of Mn are added to the purified protein. PP1c protein had been purified with 1 mM Mn^{2+} in growth media. Protein unfolding is measured by an increase in fluorescence reading as temperature increases, and plots show the first derivative curves of thermal denaturation data. Numbers shown indicate the approximate melting temperature (T_m) of the proteins as an indication of stability.

Figure 5.2-3B shows that when PP1c-3 is purified with Mn^{2+} in the media, T_m is 36.0 °C. It should be noted that this protein had been stored at 4 °C for several days before running the assay, so it is likely that it was less stable than freshly purified protein. However, when additional Mn^{2+} ions are added to the assay buffer for this experiment, the melting temperature increases dramatically, and continues to increase with increasing concentrations of Mn^{2+} ions. When 0.5 mM Mn^{2+} is added to the assay buffer,

the T_m is 62.4 °C, and when 4 mM Mn^{2+} is added, T_m has further increased to 66.4 °C. This implies that the addition of Mn^{2+} ions to purified protein can have a dramatic effect on protein stability. Therefore, in subsequent purifications of PP1c-3, 2 mM Mn^{2+} was added to the protein following purification, prior to concentration and storage of protein.

As mentioned previously, PP1c-3 showed degradation when analysed using intact mass spectrometry. Several peaks seen in the analysis correlate to constructs with loss of residues from either the C-terminus or both the C- and N-terminus. Given that the purified protein appeared to be naturally degrading to these points, this suggested that these represented more stable constructs. Two of these constructs were cloned (figure 5.2-4A) and expression trials were carried out. The construct PP1c-3^{7-310}, which comprised a removal of 6 residues from the N-terminus alongside a truncation from the C-terminus (giving a species at 34492.0 Da), was completely insoluble. PP1c-3^{1-308}, which was generated with only a truncation from the C-terminus (35147.9 Da), was successfully expressed and purified from *E. coli* (figure 5.2-4B). This implies that the N-terminus of the protein might be important for solubility of PP1c-3. However, purified PP1c-3^{1-308} still precipitated upon storage, suggesting that there are still issues with stability.

In a publication regarding the stability of phosphatase enzymes from rabbit muscle, Zhang et al. (Zhang et al., 1994) concluded that the enzyme was stable when residues were truncated from the C-terminus, but not from the N-terminus, which agrees with the results shown here. Further, the authors remarked that loss of 33 residues (until the end of the conserved 'Q-I-L-K' residues) still allowed formation of a stable, active enzyme. Although the PP1c described in that paper is somewhat different to PP1c-3 used in this project (approximately 75 % sequence similarity) and therefore the conclusions may not be universally applicable for all PP1c enzymes, these findings could help to identify a more stable PP1c-3 construct. In addition, the corresponding cleavage of these same residues (truncation after the 'Q-I-L-K' conserved residues) from PP1c-3 yields a protein with greater similarity to the other PP1c isoforms, as those isoforms have a truncated C-terminus compared to PP1c-3 (figure 5.2-1A). In Boevink et al., (Boevink, Wang, et al., 2016) an amino acid alignment of the different PP1c isoforms from *Solanum tuberosum* trims the alignment to this same point, stating that these are the most well-conserved residues. Although these terminal residues in PP1c-3 could have an impact on its properties in vitro, it would be more informative for the sequence to closely resemble the other isoforms, to get a more representative picture of the behaviour of all PP1c

variants in vitro. Therefore, an additional truncation of PP1c-3 was made, PP1c-3$^{1-293}$, which removes these poorly conserved residues from the C-terminus (figure 5.2-4A).

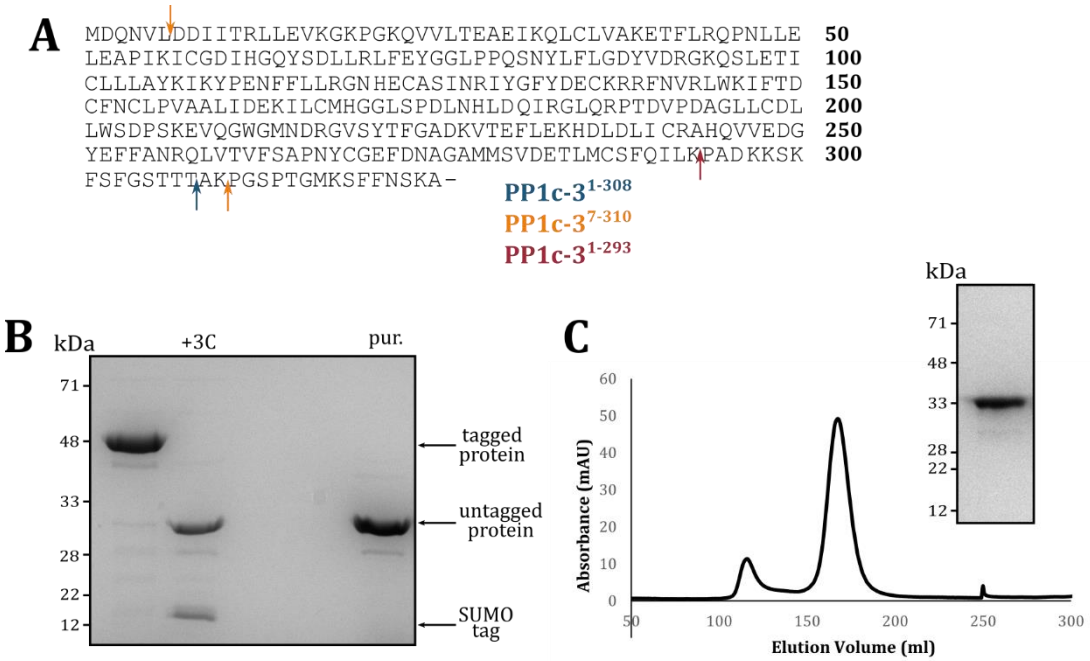


Figure 5.2-4: Generating truncated versions of PP1c-3.

A: Diagram showing the location of truncation sites for each version of PP1c-3. **B**: SDS-PAGE analysis showing PP1c-3$^{1-308}$ at each stage of purification from pOPINS3C, BL21(DE3) and LB media, +3C = sample after addition of 3C protease, pur. = purified protein. **C**: Superdex 75 26/600 gel filtration trace of PP1c-3$^{1-293}$ following cleavage of SUMO tag by 3C protease. Major peak on the trace at 166 ml elution volume. Accompanying SDS-PAGE shows purified PP1c-3$^{1-293}$ protein.

Figure 5.2-4C shows the purification of this construct, using the same expression conditions as WT full-length PP1c-3. Following successful purification of PP1c-3$^{1-293}$, intact mass spectrometry indicated that full-length PP1c-3$^{1-293}$ was present in the sample, with no indication of protein degradation. One interesting observation, which was seen consistently when this protein was purified, was an additional peak in the spectrum at a molecular weight 32 Da larger than the full-length protein (a peak at 33465.8 Da). The consistency in the appearance of this species in different protein preparations suggests that it is not simply a contaminant and could indicate some protein modification. In intact mass spectrometry analysis, such species can usually be identified through a database of common protein modifications on individual residues. However, no explanation for the source or identity of this species has yet been found. Given that the protein degradation issue appeared to have been solved by the generation

of this new construct, this truncation version of PP1c-3 (PP1c-3^{1-293}) was used for subsequent enzyme assays and for some of the later crystallisation trials.

Boevink et al. (Boevink, Wang, et al., 2016) described the production of a phosphatase-dead version of PP1c-1, which had completely attenuated enzymatic activity due to mutation of a key catalytic histidine. Therefore, the equivalent point mutation was generated in PP1c-3, namely a mutation of His129 to an alanine, using an overlapping PCR method. Having generated the mutant construct (in both the WT and PP1c-3^{1-293} backgrounds), expression trials were carried out using the same expression conditions as for WT PP1c-3. However, it appeared that mutation of this key catalytic residue also affected the solubility of the protein, and it was not possible to achieve protein solubility using WT expression conditions. PP1c-3^{H129A} was then tested in a medium-throughput expression screen, where two different expression vectors, four different *E. coli* strains, two different chemical chaperones and two different types of media were tested (figure 5.2-5A). None of these combinations were successful at producing soluble protein (results not shown). Following this screening, it was found that small amounts of soluble PP1c-3^{H129A} could be purified when co-expressed with the GroEL/GroES chaperones. The pGro7 plasmid (Takara Chaperone Plasmid Set #3340) was cotransformed with the expression construct and induced using L-arabinose, prior to induction of PP1c-3^{H129A}. With the assistance of the molecular chaperones during expression, full-length PP1c-3^{H129A} was expressed (figure 5.2-5B) and purified, although there appeared to be a large proportion of the protein forming aggregates. The largest peak marked with an 'x' in figure 5.2-5B was analysed by SDS-PAGE and appeared to contain the protein of interest (result not shown). However, the elution volume is only 115 ml, which indicates that the protein is present as an aggregated, rather than monomeric form. The smallest peak on the trace, at 167 ml also contained PP1c-3^{H129A}, and was at the correct elution volume to correlate with monomeric protein. Only the fractions corresponding to the monomeric protein were taken and used in further experiments. This phosphatase-dead version of PP1c-3 was later used as a negative control in the enzyme activity assays (section 5.5).

A

BL21 pOPINF	BL21 pOPINF DMSO	BL21 pOPINF EtOH	BL21 pOPINS3C	BL21 pOPINS3C DMSO	BL21 pOPINS3C EtOH
AI pOPINF	AI pOPINF DMSO	AI pOPINF EtOH	AI pOPINS3C	AI pOPINS3C DMSO	AI pOPINS3C EtOH
R2I pOPINF	R2I pOPINF DMSO	R2I pOPINF EtOH	R2I pOPINS3C	R2I pOPINS3C DMSO	R2I pOPINS3C EtOH
SoluBL21 pOPINF	SoluBL21 pOPINF DMSO	SoluBL21 pOPINF EtOH	SoluBL21 pOPINS3C	SoluBL21 pOPINS3C DMSO	SoluBL21 pOPINS3C EtOH

B

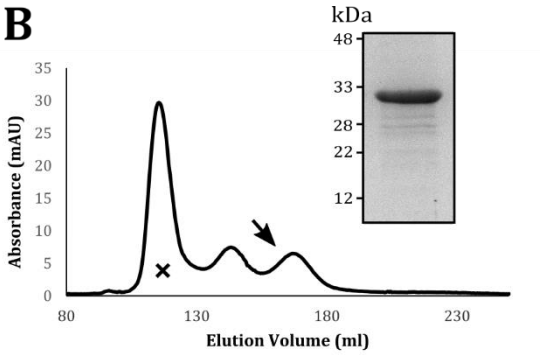


Figure 5.2-5: Generating a phosphatase-dead mutant of PP1c-3.

A: Grid showing different expression conditions screened for PP1c-3^{H129A}. Within each square, *E. coli* strain used is shown in the top row (BL21, AI = arabinose-inducible, R2I = Rosetta™ 2 pLacI, soluBL21), expression vector used in shown in the middle row (pOPINF or pOPINS3C), chemical chaperone is shown in the bottom row (no chaperone, DMSO or ethanol). This same grid plan was screened for both LB and PowerBroth™ media. **B**: Superdex 75 26/600 gel filtration trace of PP1c-3^{H129A} following cleavage of SUMO tag by 3C protease. 'X' indicates suspected elution region of aggregated protein. Arrow indicates peak on the trace where POI was found, at 167 ml elution volume. Accompanying SDS-PAGE shows purified PP1c-3^{H129A} protein.

5.3 PexRD24 can be expressed from *E. coli*

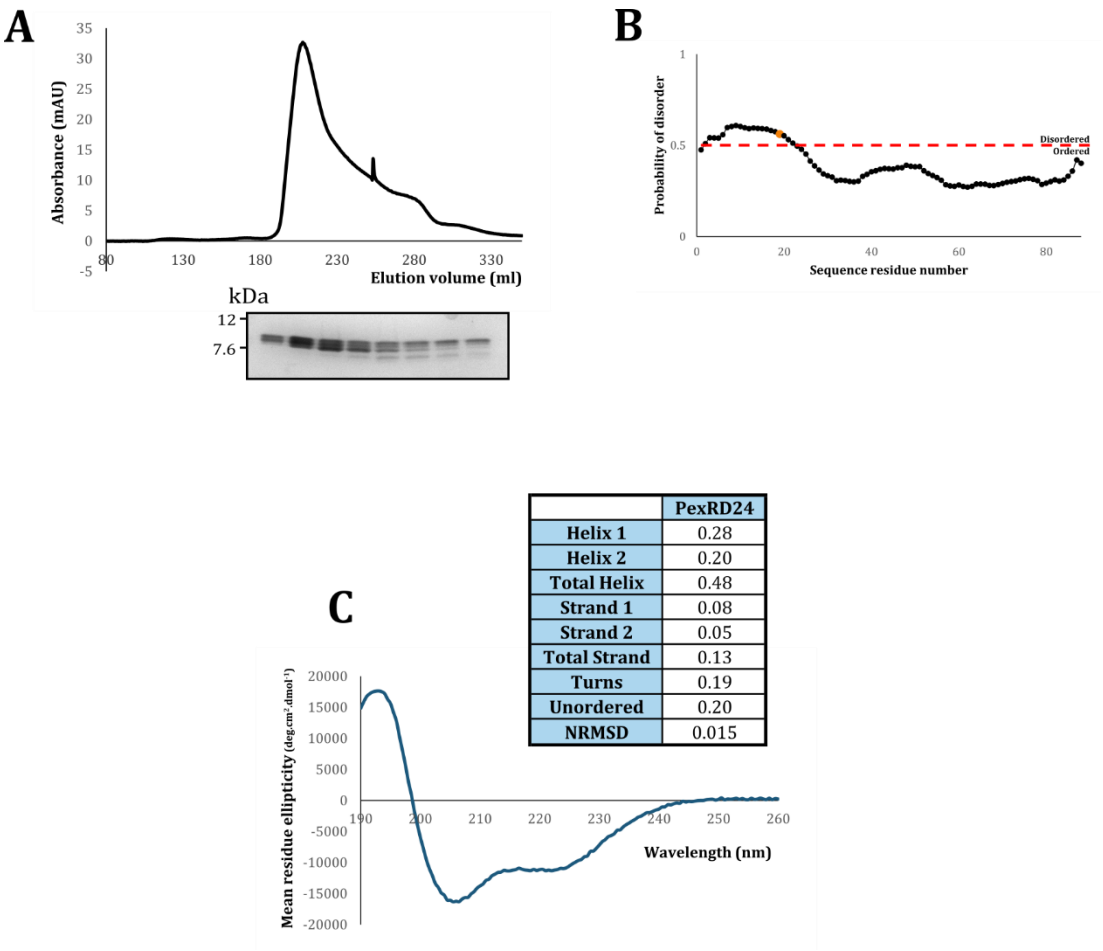
Expression of full-length PexRD24 was more straightforward than PP1c. Although several different expression conditions were tested during screening, the most successful were those using pOPINS3C, an arabinose-inducible *E. coli* strain and LB media. However, during the final gel filtration purification stage, the peak on the trace showed a tail effect (Figure 5.3-1A), which could indicate a number of issues. It is possible that the protein is interacting non-specifically with the resin in the gel filtration column, and therefore taking longer to elute from the column, or could indicate that there is some degradation of residues from the purified protein, leading to truncated species eluting from the gel filtration column. SDS-PAGE analysis of the purified protein shows multiple bands at the expected size (Figure 5.3-1A), supporting the degradation theory. Finally, intact mass spectrometry showed that although full-length PexRD24 was present in the sample at 10429.5 Da, there were several other peaks in the spectrum. One of these peaks, at 7631.3 Da, is slightly ambiguous as it appears to be a truncation of PexRD24, but it could correlate with either a truncation from the C-terminus (cleavage of 24 residues from the C-terminus gives a molecular weight of 7631.0 Da), or could be a truncation primarily from the N-terminus (cleavage of 20 residues from the N-terminus and only 5 residues from the C-terminus yields a species with a molecular weight of 7631.8 Da). Analysis using the RONN online disorder prediction tool (Yang et al., 2005) indicated that the N-terminal region of the protein might be disordered (figure 5.3-1B).

Disordered regions of a protein can be more prone to degradation, because they are not folded or held together and stabilised by intramolecular interactions. This might suggest that the major truncation seen in the intact mass spectrometry analysis corresponds to loss of N-terminal rather than C-terminal residues, although it has not been proven.

A truncated construct, corresponding to cleavage of this putative disordered N-terminus, was generated but proved difficult to express (results not shown). Therefore, despite these potential degradation issues, the full-length PexRD24 construct was used for all further experiments. Although multiple bands could be seen in SDS-PAGE analysis, the protein was soluble and could be stored at high concentrations without significant precipitation. An additional consideration is that it is not yet known how PexRD24 interacts with PP1c-3 (as discussed in section 5.1). PP1c interacting proteins frequently contain short amino acid motifs that 'dock' onto the surface of PP1c (Heroes et al., 2013), and, aside from the RVxF motif known in PexRD24, it is unknown which other areas of the effector might be important for interaction. Indeed, Heroes et al. (Heroes et al., 2013) discuss that PP1c-interacting motifs are often found in regions that are intrinsically disordered, but that become more ordered upon interaction with PP1c. Although it is not clear how far this would apply to a pathogen effector, it may be detrimental to cleave off putative disordered regions of PexRD24, in case they are important for interaction with PP1c-3.

Once purified, PexRD24 was analysed using circular dichroism. The spectrum (figure 5.3-1C) shows two distinct dips at 208 nm and 222 nm, which are characteristic of α-helical

proteins (Greenfield, 2006). DichroWeb analysis indicates that PexRD24 contains around 48 % helical structure (figure 5.3-1C).



	PexRD24
Helix 1	0.28
Helix 2	0.20
Total Helix	0.48
Strand 1	0.08
Strand 2	0.05
Total Strand	0.13
Turns	0.19
Unordered	0.20
NRMSD	0.015

Figure 5.3-1: Purification and characterisation of PexRD24.

A: Superdex 75 26/600 gel filtration trace of PexRD24 following cleavage of SUMO tag by 3C protease. Peak seen at 207 ml elution volume, with 'tailing'. Accompanying SDS-PAGE shows fractions taken from under the peak. **B**: RONN disorder prediction (Yang et al., 2005) showing the probability of disorder for each residue in the amino acid sequence. Residue labelled in orange is putative site of N-terminal cleavage identified by intact mass spectrometry. **C**: Circular dichroism (CD) far-UV spectrum for PexRD24. Numbers indicate the fraction of each type of secondary structure, where helix and strand assignments are divided into regular (helix/strand 1) and distorted (helix/strand 2). NRMSD is the normalised root mean squared deviation. Secondary structure assignments were carried out using DichroWeb, solutions were obtained from the CDSSTR method using reference dataset 7.

As discussed in section 5.1, PP1c regulatory subunits often contain an 'RVxF' motif, which is important for binding to the catalytic subunit. Boevink *et al*. had generated an RVxF mutant of PexRD24 (Pi04314mut, here called PexRD24mut), where each of these residues was mutated to an alanine, generating a KVTF > AAAA mutation. This construct was shown to no longer interact with PP1c isoforms in yeast or in planta, and could not re-localise PP1c from the nucleolus (Boevink, Wang, et al., 2016). Using the same

construct as in Boevink et al., this mutant was cloned into pOPINS3C and expression was trialled from *E. coli* under the same conditions as for WT PexRD24. PexRD24mut was expressed on a small scale and partially purified using a Ni-IMAC column (figure 5.3-2). A clear band in SDS-PAGE analysis appears to show expression of soluble, full-length protein (tagged with a His-SUMO tag at the N-terminus). Another band seen by SDS-PAGE is at the correct molecular weight (approximately 13 kDa) to be the His-SUMO tag alone (which would also bind to the Ni-IMAC column and elute along with the tagged protein). This indicates that the solubility tag might be detaching from the PexRD24mut protein after expression, although untagged PexRD24mut cannot be obviously seen on the gel. This preliminary trial indicates that PexRD24mut can be expressed in a soluble form from *E. coli*.

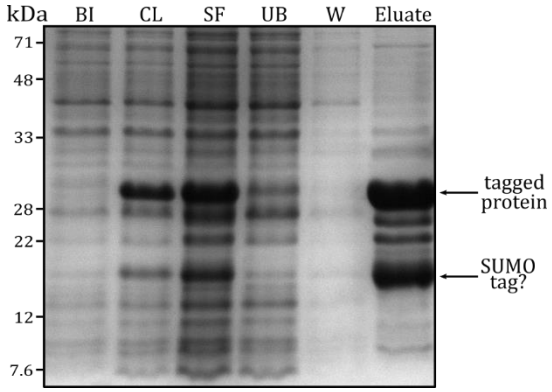


Figure 5.3-2: Expression and partial purification of PexRD24mut.

SDS-PAGE showing the expression and partial purification of the KVTF > AAAA PexRD24 mutant (PexRD24mut). BI = Before Induction of protein expression, CL = crude Cell Lysate, SF = Soluble Fraction, UB = Unbound, i.e. flow through from loading SF onto Ni-NTA column, W = Wash, i.e. flow through from washing Ni-NTA column with buffer, Eluate = SUMO-His tag protein eluted from the column and concentrated 4-fold.

It was hypothesised that this mutant would no longer be able to form a complex with PP1c, and a co-expression study was carried out using PexRD24mut and PP1c-3 (data not shown). The rationale behind the co-expression was that His-SUMO tagged PexRD24mut would not be able to co-purify with untagged PP1c-3, and therefore PP1c-3 would flow through a Ni-IMAC column without binding, whereas PexRD24mut would bind to the column and emerge when eluted with high concentrations of imidazole. Unfortunately, the experiment proved inconclusive. A superior experiment would have been to use analytical gel filtration or SPR to assess complex formation in vitro, but as described later in section 5.4, it was not possible at this stage. Therefore, no conclusions could be drawn about the ability of the KVTF > AAAA mutant to bind to PP1c-3 in vitro.

5.4 Interaction of PexRD24 and PP1c

Having shown that both PexRD24 and PP1c-3 could be expressed and purified in vitro, the next stage was to determine whether they could form a complex. Initial experiments used analytical gel filtration to assess complex formation between the two proteins. However, the results did not indicate that any complex was formed between the two partners (see later and figure 5.4-3B). This might indicate that the two proteins do not form a complex in vitro, but another explanation is that the instability of the proteins – i.e. the tendency of PP1c-3 to precipitate out of solution and the N-terminal degradation of PexRD24 – prevents formation of a complex in vitro. If the proteins were more stable, it is possible that they would be able to form a complex.

The next strategy was to co-express the two partners in an attempt to stabilise the proteins. For this co-expression strategy, a His-SUMO tagged construct of PexRD24 and an untagged construct of PP1c-3$^{1\text{-}293}$ were cotransformed into arabinose-inducible *E. coli* and expressed in LB media. During IMAC purification, the His-SUMO tagged PexRD24 should co-elute with the untagged PP1c-3$^{1\text{-}293}$ if the two proteins form a complex. By allowing complex formation before the proteins are purified, the partners should not have degraded or precipitated, giving a better chance for successful complex formation. Figure 5.4-1A shows that co-purification was successful, as SDS-PAGE analysis shows two major bands in the purified protein at the approximate correct size to be PexRD24 (10.4 kDa) and PP1c-3$^{1\text{-}293}$ (33.4 kDa). It was noted that the trace from the final purification gel filtration stage (figure 5.4-1B) showed a tailing effect on the peak, which could be an indication of degradation. However, intact mass spectrometry of the purified protein complex indicated that both full-length proteins were present in the sample. This suggests that co-expression has not only allowed complex formation, but has also stabilised the proteins. After four weeks of storage at 4 ^{o}C, a repeat intact mass spectrometry analysis indicated that there was some degradation of the effector, suggesting that there was not complete protein stabilisation, but increased stability compared to PexRD24 expressed alone. Additionally, there was still some precipitation of the complex at high protein concentrations, but this was to a lesser extent than PP1c-3$^{1\text{-}293}$ alone. Overall, it is possible to purify a PexRD24/PP1c-3 complex by co-expression, and the proteins appear to stabilise each other in vitro.

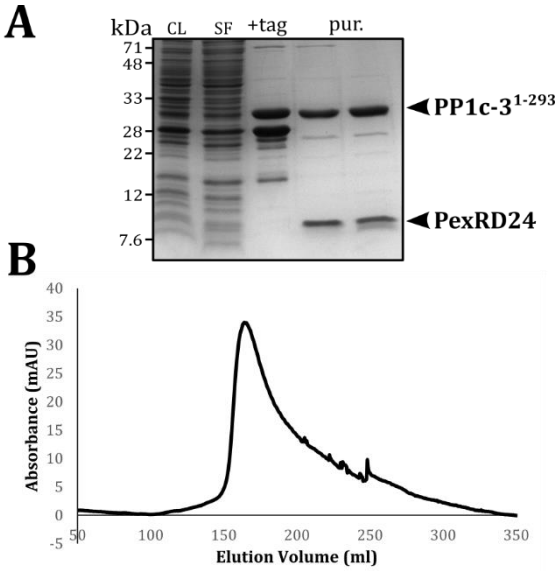


Figure 5.4-1: Purification of a PP1c-3$^{1-293}$/PexRD24 complex.

A: SDS-PAGE analysis showing the expression and purification of the PP1c-3$^{1-293}$/PexRD24 complex, where PP1c-3$^{1-293}$ is untagged, and PexRD24 is tagged with an N-terminal SUMO tag. CL = crude Cell Lysate, SF = Soluble Fraction, +tag = partially purified complex before removal of solubility tag, pur. = purified protein. **B:** Superdex 75 26/600 gel filtration trace of PP1c-3$^{1-293}$/PexRD24 complex following cleavage of SUMO tag by 3C protease. Peak seen at 166 ml elution volume, with 'tailing'.

Despite the progress with stabilising proteins through complex formation, the regular occurrence of protein precipitation leads to reduced yield and potential issues with measuring protein concentration accurately. Not only this, but protein that is unstable could give misleading results in experiments, such as false negatives in protein interaction assays. Therefore, a commercial buffer screen from Molecular Dimensions (RUBIC Buffer Screen MD1-96) was used to screen for more suitable buffer conditions that might enhance stability. Previously, all proteins had been purified in a standard gel filtration buffer containing 20 mM HEPES (pH 7.5) and 150 mM NaCl (with Mn^{2+} added to PP1c-3 after purification), but this screen tests a range of different buffer systems, salt concentrations and pH. The screen is performed using a DSF method as described in section 5.2. Samples of the protein (in the standard gel filtration buffer) were added to a 96-well screen containing different combinations of buffer conditions, and SYPRO® Orange dye was added. The protein was then subjected to incremental temperature increases, with fluorescence readings being taken at each temperature as the protein unfolded. Instead of plotting the derivative curves (as described in section 5.2), here the raw data was plotted, and an example curve is shown in figure 5.4-2A, indicating the features of each trace. As shown, T_m is defined as the temperature at which half maximum fluorescence is achieved. The traces for each buffer condition in the screen were analysed

visually, by plotting groups of traces together that allowed comparison of a single variable e.g. increasing imidazole concentration. The conditions that yielded the highest T_m from each group were then plotted together, and so on until the highest-ranking buffer conditions could be identified. Sample traces for some of the buffer conditions are shown in figure 5.4-2B, indicating how the changes in buffer can greatly affect the thermal stability of the proteins. For example, it is clear from figure 5.4-2B that buffer condition C2 is extremely destabilising, because the fluorescence intensity is high even at low temperatures – this indicates that the protein has become unfolded as soon as it encounters the buffer condition, before it has been heated. By contrast, the trace for buffer condition D12 shows low fluorescence intensity at low temperatures before a sharp increase with clearly defined T_m at approximately 60 ^oC. Broadly, this indicates that buffer condition D12 would be much more suitable for the PexRD24/PP1c-3 complex than buffer condition C2. Following visual analysis, the most stabilising buffer conditions were found to be sodium phosphate buffer with a high salt concentration at neutral pH. The highest T_m overall was conferred by buffer condition D4, which contained 100 mM sodium phosphate (pH 7.5) and 250 mM NaCl. These parameters were used to create a new gel filtration buffer. One difficulty to overcome was that $MnCl_2$ cannot be readily added to this new buffer because the Mn precipitates as manganese (II) phosphate. Therefore, it was not possible to add additional Mn^{2+} to the protein after purification, as had been standard practise. Instead, 1 mM $MnCl_2$ was added to the media prior to inducing protein expression, anticipating that the protein will take up Mn^{2+} as required during expression. The PexRD24/ PP1c-3^{1-293} complex was then purified using the new phosphate buffer, and it was clear in the final gel filtration stage that the elution peak was larger and more symmetrical than when the complex was purified using the standard HEPES-based gel filtration buffer (figure 5.4-2C). Following purification, the final yield was significantly increased, and the protein itself was more stable and less prone to precipitation. This indicates that the increased salt concentration and use of a phosphate buffer system is advantageous for protein stability. In addition, both partners of the complex were purified separately in the same phosphate buffer. PexRD24 appeared to be successfully purified, and the previous 'tailing' effect seen on the gel filtration trace (see figure 5.3-1A) is reduced by the higher ionic strength of the phosphate buffer. PP1c-3 could also be concentrated to much higher concentrations (above 10 mg/ml) without visible precipitation when purified in the phosphate buffer.

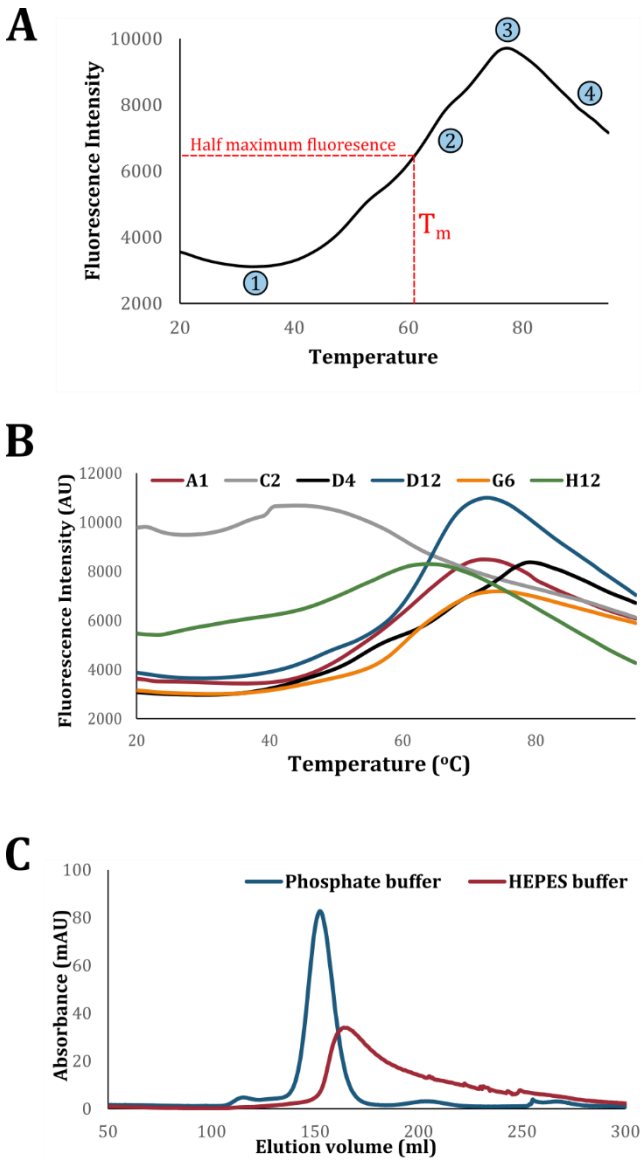


Figure 5.4-2: The PP1c-3$^{1-293}$/PexRD24 complex is more stable in a phosphate-based buffer.

A: Representative DSF melt curve from a buffer screen. The graph shows fluorescence intensity plotted against temperature. *1*: At low temperatures the protein is folded, and SYPRO® orange cannot bind. *2*: The gradient of the curve increases as the protein begins to unfold, allowing SYPRO® orange dye to bind to hydrophobic residues in the core, increasing fluorescence. *3*: The fluorescence intensity reaches a maximum as all the protein is unfolded. *4*: Dye begins to dissociate from the protein, reducing fluorescence levels. The point at which the fluorescence intensity is half its maximum value is where 50% of the protein is unfolded. The corresponding temperature at this point is known as T_m – the melting temperature of the protein. **B:** DSF analysis showing the change in PP1c-3$^{1-293}$/PexRD24 complex protein stability in different buffers. Plot shows fluorescence intensity plotted against temperature. Graph shows 6 example curves from the commercial buffer screen, labels indicate their position in the grid. **C:** Superdex 75 26/600 gel filtration traces of PP1c-3$^{1-293}$/PexRD24 complex following cleavage of SUMO tag by 3C protease, in either a HEPES or phosphate-based buffer.

Since the optimised buffer conditions appear to enhance the stability of PP1c-3$^{1-293}$ and its complex with PexRD24, it was considered that the stabilisation might allow the proteins to form a complex in vitro, rather than just through co-expression. This would open up a number of possibilities for different experiments, including biophysical analysis to quantify the interaction of the two proteins. Previous experiments showed that mixing purified PexRD24 and PP1c-3 and assessing complex formation by analytical gel filtration was unsuccessful, because no obvious peak shift was seen to indicate complex formation, and traces were of poor quality. A trace showing this experiment using proteins purified in the HEPES gel filtration buffer is shown in figure 5.4-3B. For this experiment, no complex formation is evident, as two small peaks in the mixed protein sample can be seen at similar elution volumes to each protein alone. Note that in this comparison experiment, full length WT PP1c-3 was used, rather than the truncated PP1c-3$^{1-293}$.

The same experiment was then repeated using proteins that had been purified in the optimised phosphate gel filtration buffer. The analytical gel filtration column was also equilibrated in this buffer. Both PP1c-3$^{1-293}$ and PexRD24 were run separately, and a PP1c-3$^{1-293}$/PexRD24 complex generated by co-expression was also assessed to give an expected elution volume. Finally, PP1c-3$^{1-293}$ and PexRD24 were mixed together and incubated for 7 hours on ice prior to running down the gel filtration column. Figure 5.4-3A shows that the traces for PP1c-3$^{1-293}$, PexRD24 and the complex generated by co-expression all show single, symmetrical peaks. There is a clear peak shift from 11.1 ml to 10.3 ml when PP1c-3$^{1-293}$ is co-expressed along with PexRD24. When the two proteins are mixed together in a 1:1 ratio in vitro, the peak on the trace is less clear. It appears to be a double peak, with the two maxima at 9.8 and 10.7 ml (figure 5.4-3A). It is not obvious why the peak should appear in this way – it is possible that there could be a different oligomerisation state being formed when the proteins are mixed in vitro, leading to slightly different elution volumes (for example, if the PexRD24:PP1c-3 complex was in a 2:1 ratio rather than a 1:1 ratio, this could account for the peak at 9.8 ml, which appears to indicate a larger molecular weight than the size of a 1:1 complex, but smaller than the size of a 2:2 complex). Alternatively, there could just be a deformity in the peak that gives a misleading indication of the elution volume. Nevertheless, there appears to be a definite peak shift when the two partners are mixed in vitro – if no complex was formed, peaks would be expected at the same elution volumes as the uncomplexed proteins (11.1 and 12.8 ml). This gives a promising indication that using the stabilising phosphate buffer for protein purification allows PP1c-3 and PexRD24 to form a complex when mixed in vitro. Further work is required to confirm this, with a repeat of the analytical

gel filtration experiment needed, alongside accompanying SDS-PAGE analysis of the fractions for confirmation of results.

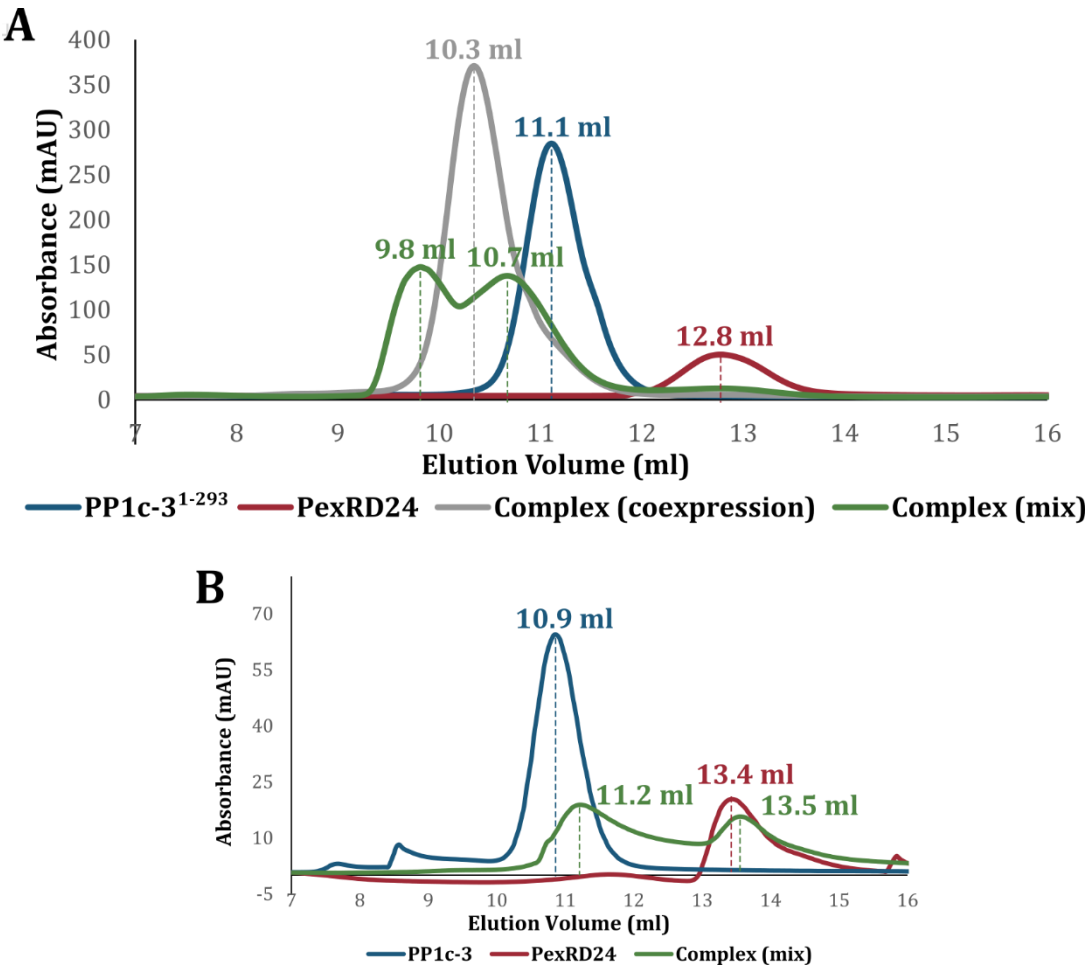


Figure 5.4-3: Stabilised proteins allow PP1c-3$^{1-293}$/PexRD24 complex formation in vitro**.**

A: Overlaid analytical gel filtration traces showing the elution of PP1c-3$^{1-293}$ (blue) and PexRD24 (red) alone, along with PP1c-3$^{1-293}$/PexRD24 both mixed together in vitro (green) and co-expressed as a complex (grey). All samples purified in phosphate gel filtration buffer, and experiment was run in the same buffer. Volumes indicate elution volumes. **B**: Comparison with experiment in **A**, using the standard HEPES gel filtration buffer for the experiment and protein purification. Overlaid analytical gel filtration traces showing the elution of (full-length) PP1c-3 (blue) and PexRD24 (red) alone, along with PP1c-3/PexRD24 mixed together in vitro (green). Volumes indicate elution volumes.

Alongside experiments to study the biochemistry of PP1c and PexRD24, a key breakthrough into learning how they interact would be to obtain the atomic structure of the two proteins in complex. Gaining structural insights into how PexRD24 binds to the PP1c host target might give some indication of what impact the effector has, if any, on the conformation and/or activity of the enzyme. To achieve this, numerous crystallisation screens were set up for the purified PP1c-3/PexRD24 complex. These screens were set up for both full-length PP1c-3 and for PP1c-3$^{1-293}$, as well as using

protein that had been purified in both the standard HEPES-based gel filtration buffer and the optimised phosphate buffer. Numerous protein crystals were obtained from commercial screens, but they were usually flat or needle-like and of insufficient quality to test for X-ray diffraction. Images of a sample of these crystals are shown in figure 5.4-4, along with descriptions of their crystallisation conditions. Crystal quality was improved upon using the more stabilised PP1c-3 construct (PP1c-$3^{1\text{-}293}$) and using the phosphate gel filtration buffer for protein purification, so it is possible that further screening and optimisation of the complex would yield diffraction-quality crystals in the future.

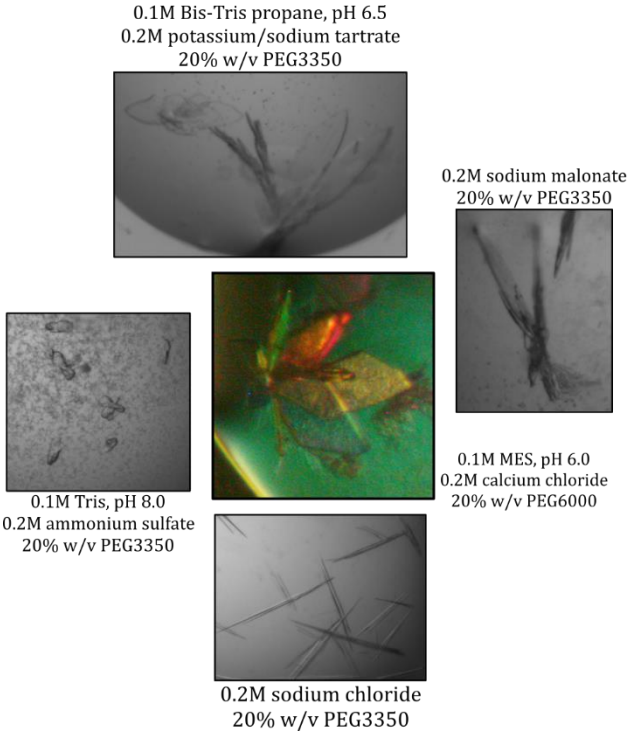


Figure 5.4-4: Images of sample PP1c-3/PexRD24 crystals.

Photographs of PP1c-3/PexRD24 crystals grown in different commercial crystallisation screens. Labels indicate the conditions of crystallisation. **Centre**: Colour photo of PP1c-$3^{1\text{-}293}$/PexRD24 complex crystals, purified in phosphate gel filtration buffer, from PACT™ screen (Molecular Dimensions). **Clockwise from top**: PP1c-3/PexRD24 complex crystals, purified in HEPES gel filtration buffer from PACT™ screen. PP1c-3/PexRD24 complex crystals, purified in HEPES gel filtration buffer from PACT™ screen. PP1c-$3^{1\text{-}293}$/PexRD24 complex crystals, purified in HEPES gel filtration buffer, from PEGs suite (Qiagen). PP1c-$3^{1\text{-}293}$/PexRD24 complex crystals, purified in HEPES gel filtration buffer, from custom-made KISS screen (Clare Stevenson).

5.5 Effect of PexRD24 on PP1c enzyme activity

Phosphatase enzymes are responsible for dephosphorylating numerous substrates in plant and animal cells (see section 5.1). One hypothesis for PexRD24 targeting PP1c is that the effector is manipulating the enzyme activity for the benefit of the pathogen. To test this hypothesis, assays were designed to confirm the activity of purified PP1c-3 in vitro, and test whether PexRD24 has any effect on the enzyme activity.

An important consideration throughout the phosphatase assays was the availability of manganese ions. For the enzymes to be active, it was found that Mn^{2+} was an essential component of either the assay buffer itself or the buffer containing the purified protein. Addition of Mn^{2+} during expression of the protein was not sufficient to generate an active enzyme. When PP1c-3 was purified in a HEPES-based gel filtration buffer, 2 mM $MnCl_2$ was added to the purified protein, but in the case of PP1c-3 purified in a phosphate-based buffer, $MnCl_2$ (1 mM) was added only to the assay buffer, to minimise precipitation of manganese phosphate.

A further issue encountered when designing the phosphatase activity assays was that PP1c-3^{1-293} and PexRD24 could not form a complex when mixed in vitro. Later experiments using an optimised purification buffer indicated that complex formation might be possible when the proteins are stabilised (see section 5.4), but these results were not available when designing the enzyme assays. Therefore, the only method available to obtain the two proteins in complex was through co-expression. By necessity, this means that in the enzyme assays, the PP1c-3/PexRD24 complex studied is from a different protein preparation obtained using a slightly different methodology (i.e. co-expression). The experiment would be better controlled if the same preparation of PP1c-3 could be directly compared with and without the presence of PexRD24, but given that this was not possible, this variable needs to be taken into consideration when interpreting the results. With the advancement of a buffer that promotes protein stability, future work on this project could focus on improving this aspect of the assay.

Two different methods were used to measure the phosphatase activity of PP1c-3. The first method uses a generic threonine phosphopeptide (K-R-pT-I-R-R) (Merck) as a substrate for the enzyme (figure 5.5-1A). When the phosphatase enzyme dephosphorylates the peptide, the reagent malachite green is used to detect free phosphate in the buffer, yielding a green colour generated by a complex of malachite green, molybdate and phosphate formed under acidic conditions (Feng et al., 2011), that absorbs at 620 nm. This assay was carried out using PP1c-3^{1-293}, the PexRD24/ PP1c-3^{1-}

^{293} complex formed by co-expression, and the phosphatase dead mutant PP1c-3^{H129A} (made in the PP1c-3$^{1-293}$ background). Initial tests were carried out to identify the optimum enzyme concentrations and reaction times to use in the assay. An observation from these tests was that the complex appeared much more efficient at dephosphorylation activity than the phosphatase enzyme alone. Low concentrations of complex were able to hydrolyse much more substrate in a given time period than the same concentration of phosphatase alone (data not shown). Indeed, in the results for the assay (figure 5.5-1A) the absorbance at 620 nm (directly related to the amount of free phosphate) is higher (up to two-fold increased) for the PexRD24/ PP1c-3$^{1-293}$ complex than for PP1c-3$^{1-293}$ alone. This suggests that the enzyme activity of PP1c-3 is enhanced by PexRD24. There could be a number of reasons for this effect. It is possible that the effector is promoting an enhanced phosphatase activity. An alternative explanation is that when the phosphatase is purified as a complex, it is more stable than phosphatase purified as an isolated catalytic domain (described elsewhere in this chapter), and the more stable enzyme might be expected to have a higher rate of activity. The negative control PP1c-3^{H129A} showed no activity at all, indicating that it has been successfully generated as a phosphatase-dead mutant. This assay was repeated three times and comparable results were obtained for each repeat. Figure 5.5-1A shows the results for this assay displayed as the raw data – i.e. absorbance at 620 nm. This raw data can be converted to the amount of phosphate released by generating a standard curve (using malachite green to detect known concentrations of phosphate in solution). Using the standard curve, the absorbance readings can be converted to picomoles of phosphate for the unknown concentrations of phosphate released by the enzyme. However, difficulties were encountered in obtaining a fully linear standard curve, and it was decided to plot raw data to avoid any over-interpretation of the results. Upon further optimisation of the assay, it would be possible to convert to a more standard representation of the phosphatase assay results.

Although the assay described above demonstrated the success of generating active PP1c-3 in vitro, it was unclear if the perceived increase in activity upon complexation with PexRD24 was simply due to enhanced stabilisation of the enzymes. Following development of the optimised buffer for purification of PP1c-3, the phosphatase assay was repeated. However, a crucial factor in stabilising PP1c-3 in solution is the addition of phosphate to the purification buffer, and an assay that relies on the detection of free phosphate in solution is obviously unsuitable in this situation. Therefore, a new assay needed to be identified that would not rely on the detection of free phosphate. In this case, the substrate *p*-nitrophenyl phosphate (pNPP) was used, which, upon cleavage by

phosphatase, is converted to *p*-nitrophenol. Under alkaline conditions, this product is converted to the *p*-nitrophenolate ion, which absorbs at 405 nm, and appears yellow (Lorenz, 2011). By detecting the presence of *p*-nitrophenolate, the phosphate in the assay solution deriving from the buffer composition should not affect the results. For this provisional experiment, the assay was not optimised for enzyme concentration or reaction time, and the same final enzyme concentration (0.4 μM) was used as described for the previous assay. The protocol for the assay was adapted from a phosphatase activity assay by Kelker et al. (Kelker et al., 2009). Substrate concentration was used according to this published protocol. As shown in figure 5.5-1B, the curves for PP1c-3$^{1-293}$ and the PP1c-3$^{1-293}$/PexRD24 complex are not saturating, indicating that the enzyme is fully turning over even 8 mM substrate in the 15 minute reaction time. To obtain better results, the concentration of enzyme may need to be lowered, or the concentration of substrate increased. As expected, the negative control showed no enzyme activity. A commercially-available purified phosphatase had also been purchased for use as a positive control. The recombinant human PPP1A protein purchased from Abcam (ab113150) showed significantly lower activity than PP1c-3. The proteins share around 82 % sequence identity and are originally sourced from different organisms (human and *S. tuberosum*). Both proteins were generated by recombinant expression from *E. coli*, but the commercially purchased enzyme includes an N-terminal His tag, which has been cleaved off in the case of PP1c-3. An additional consideration is that it was not possible to confirm the concentration of the enzyme supplied. The manufacturers data sheet shows the enzyme to be at 5 μM concentration, but the small quantity available meant that it was not practical to confirm this using equipment in the lab, as it would have used up a significant quantity of the sample. This means that it is not certain that the enzyme was actually used at 0.4 μM as presumed. Therefore, although the positive control was used as accurately as possible, and chosen to be as close as possible in sequence identity to the PP1c-3 enzyme, it was not a suitable control for use in this instance.

The assay using the stabilised proteins in phosphate buffer was only carried out once due to time constraints. However, it demonstrates that the PP1c-3 protein can cleave the pNPP substrate. Interestingly, despite the need to modify the concentrations of substrate and enzyme, the absorbance readings of PP1c-3$^{1-293}$ in complex with PexRD24 appears to be higher than the enzyme alone, as seen for the previously described assay. The consistency of this effect suggests that it is not related to a particular condition of either assay (such as buffer composition, or substrate used) but it is still not clear whether the effect is due to innate stabilisation cause by complex formation, or an effect that is

specific for the interaction of PexRD24. Further experiments would be needed to determine this.

Overall, the phosphatase assays carried out thus far have demonstrated that PP1c-3$^{1-293}$ is an active enzyme, that is capable of removing phosphate from two different synthetic substrates in vitro. It also appears that when PexRD24 is bound to PP1c-3$^{1-293}$, it does not inhibit the activity of the enzyme. The fact that the activity of the PP1c-3$^{1-293}$/PexRD24 complex appears greater than for the enzyme alone cannot be fully accounted for without further experimentation.

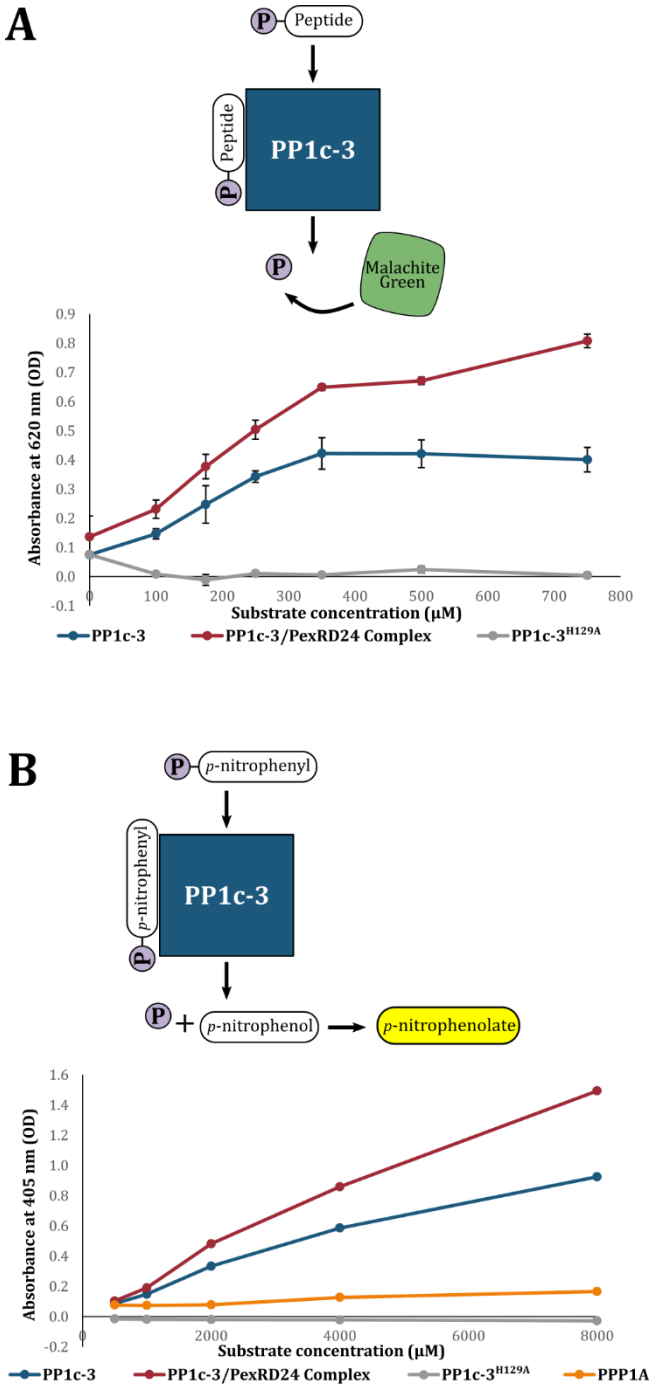


Figure 5.5-1: Enzyme assays suggest that PexRD24 does not inhibit PP1c-3 activity.

A: Diagram outlining the mechanism of the phosphatase assay involving a synthetic phosphopeptide substrate with detection by malachite green. Scatter plot shows absorbance at 620 nm for different concentrations of phosphopeptide substrate, turned over by PP1c-$3^{1\text{-}293}$, PP1c-$3^{1\text{-}293}$/PexRD24 complex or the PP1c-$3^{1\text{-}293H129A}$ negative control. Measurements were taken in triplicate and assay was carried out three times, error bars represent standard deviation. **B:** Diagram outlining the phosphatase assay mechanism using pNPP as the substrate. Scatter plot shows absorbance at 405 nm for different pNPP substrate concentrations when turned over by PP1c-$3^{1\text{-}293}$, PP1c-$3^{1\text{-}293}$/PexRD24 complex, the PP1c-$3^{1\text{-}293H129A}$ negative control, or the PPP1A control enzyme (Abcam). **Note**: PP1c-$3^{1\text{-}293}$ construct is used throughout, for all proteins used for both assays. For brevity, it is shown as 'PP1c-3' in the figure labels.

5.6 Discussion

The goal of this project was to use structural biology and biochemistry to understand how the *P. infestans* effector PexRD24 manipulates its host target protein phosphatase 1. It was hoped that obtaining structural information about how the proteins interact, and discovering if the effector altered enzyme activity, would complement the work by Paul Birch's lab in studying the localisation and interaction of the two proteins in planta.

Initial expression screens indicated that only one of the PP1c isoforms identified from the original yeast-2-hybrid screen, PP1c-3, could be expressed in a soluble manner from *E. coli*. The other isoforms remained insoluble, despite numerous trials. It should be noted that this includes PP1c-1, the isoform discussed at length in Boevink et al. (Boevink, Wang, et al., 2016). Unfortunately, it was not possible to characterise this particular isoform in vitro, but it was hoped that the conclusions made in this work about the behaviour and activity of PP1c-3 would also be applicable to PP1c-1.

As discussed in section 5.1, it is well documented in the literature that protein phosphatases can be unstable when expressed from *E. coli* and purified in vitro. Using molecular chaperones, altering buffer conditions and stabilisation with interacting proteins are all well documented techniques for obtaining soluble phosphatase proteins, and these methods were adopted for PP1c-3. Initial purification attempts focussed on the use of different *E. coli* strains and solubility tags, but this was later diversified. For the phosphatase dead mutant PP1c-3^{H129A}, soluble protein was particularly elusive, and only obtained via co-expression with the GroEL/GroES chaperones. Difficulties with obtaining the soluble protein led to slow progress in several areas of the project. Progress was made by truncating PP1c-3 to make a more stable construct, PP1c-3$^{1-293}$, which was less prone to degradation, but the protein would still readily precipitate out of solution when stored.

A further observation made when PP1c-3$^{1-293}$ was analysed by intact mass spectrometry was that there appeared to be an additional species, alongside the full-length protein, present in the sample that was 32 Da larger than the molecular weight of the full-length PP1c-3$^{1-293}$. This species was seen repeatedly when analysing PP1c-3$^{1-293}$, either alone or in complex with PexRD24. The identity of this species was not found, but could represent a post-translational modification of the protein, for example the oxidation of a methionine side chain (Gerhard Saalbach, personal communication). However, there was not enough information to identify the species exactly. Modification of amino acid

side chains can affect the stability and properties of a protein, but whether or not this putative modification had any impact on protein stability is not possible to determine.

One problematic result of the instability of PP1c-3 (and the degradation of PexRD24) was that it was not possible to obtain complex formation by mixing purified proteins in vitro. This limited the experiments that could be carried out, as biophysical techniques require measuring the actual process of complex formation to assess strength of binding.

One area of focus for the project was designing an assay to test the enzyme activity of PP1c-3, and how it was affected by PexRD24. Preliminary assays showed that both full length PP1c-3 and the PP1c-3^{1-293} truncation were active enzymes, but the phosphatase activity was dependent on manganese ions being provided to the purified protein. However, when studying the effect of PexRD24 on the enzyme activity, these assays relied on the effector being co-expressed and co-purified with the enzyme. This made it difficult to assess whether any alteration of enzyme activity was due to the stabilisation of the enzyme through co-expression, or whether it was due to a specific input from the effector. At high concentrations of substrate, the absorbance (as an indicator of enzyme activity) was enhanced about two-fold when the effector was present, and this effect could simply be due to the enzyme being better folded in its active state when stabilised by a partner. It is known that the PP1c catalytic subunit is heavily reliant on regulatory subunits for specifying localisation or substrate in vivo, so it might be expected that the catalytic subunit would be stabilised by a partner, as it does not function in isolation in vivo. To test this further, PP1c-3 could be co-expressed with another phosphatase-interacting protein, and the activity examined to see if the same effect occurs when PP1c is stabilised by a different partner. However, one conclusion that can be drawn from the assays is that the effector is not inhibiting the enzyme in vitro. This relates to the study by Boevink et al. because the authors tested the enzyme activity of GFP-tagged PP1c purified from plant tissue and found it to be active whether or not it was bound to PexRD24. In addition, a phosphatase-dead version of PP1c did not enhance pathogen colonisation to the same extent as active enzyme, indicating that phosphatase activity is required for PexRD24 function (Boevink, Wang, et al., 2016). It should be noted that testing enzyme activity in vitro using artificial substrates may not be giving an accurate picture of the enzyme activity in vivo. For example, this heterologously expressed PP1c is manganese-dependent, but the metal ions incorporated in vivo are thought to be Fe^{2+} and Zn^{2+} for eukaryotes. The addition of Mn^{2+} ions to the enzyme are considered to potentially alter the activity and substrate specificity of the enzyme (Verbinnen et al.,

2017). Nevertheless, without knowledge of the native PP1c substrates, this provides a useful starting point for understanding PexRD24 action in more detail.

Boevink et al. proposed that PexRD24 may be affecting salicylic and jasmonic-acid defence signalling pathways. Alongside a role in PP1c re-localisation, regulatory subunits can have a role in modifying the substrate specificity of the enzyme. For example, the regulatory subunit MYPT1 causes a significant change in the catalytic site of PP1c to enhance its activity towards a particular substrate, in this case myosin (Terrak et al., 2004). If the enhancement of the catalytic activity of PP1c observed here in the presence of PexRD24 is not simply due to stabilisation of the protein in vitro, it may mean that PexRD24 is in some way altering the catalytic site of the enzyme to have an enhanced activity towards the substrates. An alteration of catalytic site could mean that PexRD24 is not only re-localising the enzyme but altering its specificity for a particular substrate, such as a component of the salicylic or jasmonic-acid defence signalling pathway. However, in the absence of any structural information at present, this remains speculation.

Late on in the project, a major breakthrough occurred with the identification of a buffer composition, using phosphate and higher sodium chloride concentration, that stabilised PP1c-3 in vitro. The purified protein could be used at higher concentration without precipitation, and a preliminary analytical gel filtration experiment indicated that the stabilising buffer might now allow complex formation to occur in vitro. Although not feasible within the time scale of the project, it would now be possible to carry out, for example, SPR experiments to quantify the strength of binding between PP1c-3 and PexRD24. This could also be extended to testing the interaction between the RVxF mutant, PexRD24mut, and PP1c-3, to confirm that this interaction motif mutant no longer binds to PP1c-3. Similarly, it would be interesting to test whether the phosphatase-dead mutant, PP1c-3^{H129A}, still interacts with PexRD24, despite the mutation in its catalytic motif. Being able to achieve complex formation by mixing rather than co-expressing the proteins would also allow for better control in enzyme activity assays, as the same preparation of enzyme could be used when mixed with the effector and when tested in isolation.

Finally, the structural characterisation of the interaction of PP1c and PexRD24 was one aim of the project. Although the crystal structure has not been uncovered within the timescale of the project, crystals have been generated, which could be further optimised. Understanding how PexRD24 interacts with PP1c could give further clues as to how it

influences the enzyme, and what the function might be in re-localising it out of the nucleolus in planta.

Overall, the major outcome for this project has been the optimisation of techniques to generate stable PP1c and PexRD24 for studies in vitro. This required considerable effort, perhaps unsurprising given the known issues with purifying stable PP1c (Peti et al., 2013). Now that a stabilising buffer and optimal constructs have been developed, it might be possible in the future to understand in greater detail how PexRD24 manipulates its host target.

General Discussion

Recent insights into plant immunity have revealed a sensitive, complex system for detecting and responding to plant pathogens (Wu et al., 2018). Threats to global food security mean that studying devastating diseases of our food crops is more important than ever. Both the pathosystems explored in this PhD thesis are of huge importance to agriculture, and both *M. oryzae* and *P. infestans* represent model organisms for fungi and oomycetes respectively. The theme that runs throughout this thesis is using the study of molecular protein-protein interactions to gain new understanding of how host and pathogen proteins interact. It is of fundamental importance to analyse these interactions at the biochemical and structural level, because this enables us to understand how we can manipulate these interactions to benefit the plant. However, these two projects ultimately explore the intricacies of host/pathogen interaction from both sides of the battlefield. In the case of the effector PexRD24 interacting with a host phosphatase enzyme, it is possible to see the pathogen exploiting a host system, namely the use of additional subunits to target a generic enzyme, to enhance disease. On the opposing side, the ability of the host plant to turn a putative HMA domain-containing pathogen target into a detection system by integrating it into an immune receptor that can recognise sequence-divergent pathogen effectors, namely AVR-Pia and AVR-PikD, is a powerful tool in defence. In both cases, it would appear that we can witness the elegance and complexity of the plant/pathogen immunity arms race.

There are still many unanswered questions regarding recognition and signalling events in response to effector proteins. The recent explosion in research surrounding the integration of non-canonical domains into NLR proteins has provided new models that can shed light not just on how immune receptors are functioning, but also give clues about possible effector targets. Research into these integrated domains is therefore important for all areas of plant immunity. This project has expanded the field of research on integrated domains by showing how an integrated domain in the Pik-1 NLR can recognise an effector, AVR-Pia, that is also recognised by a different NLR pair containing an evolutionarily related but distinct integrated domain. Although the response to the 'mismatched' effector AVR-Pia is considerably weaker than for the 'matched' effector AVR-PikD, there are some interesting implications for this work, as it appears that Pik-1 is binding two sequence-unrelated effector proteins at different domain interfaces, and yet is still able to trigger an immune response. It would be fascinating to discover whether the differential binding also leads to different intramolecular and intermolecular interactions within the NLR pair.

Similar experiments could be carried out to investigate whether the same cross-reactivity would occur for another effector that is recognised by the RGA5/RGA4 pair of NLRs, AVR1-CO39. It would be interesting to draw parallels between AVR-Pia and AVR1-CO39, which are also two sequence-divergent MAX effectors, but are recognised by the same NLR. On the other hand, it would be valuable to reverse the experimental design and examine more closely whether the cross-reactivity exists for the RGA5/RGA4 NLRs responding to AVR-PikD. Given the clear differences in location of integrated domain and effector binding interfaces between the Pik-1/Pik-2 and RGA5/RGA4 pairs of NLRs, it would be interesting to uncover the level of plasticity in the system i.e. whether both pairs of NLRs are capable of recognising both effectors.

Although relatively little is currently known about the host targets of AVR-Pia and AVR-Pik effectors, it seems likely that the targets are highly similar or the same, given that they are recognised by such similar integrated HMA-containing domains. Therefore, it could be possible that the differential binding of the effectors by their immune receptors is related to how the effectors bind their target. It could be that the two effectors bind two different surfaces of a host protein, exerting different effects on it and thus different effects on the host plant; possibly emphasising an important role for the target protein in immunity or host cell function. This could also link to the different ways in which the HMA domain is integrated into RGA5 and Pik-1; possibly an attempt to expose the most important binding surface for each effector in the full-length context of the NLR. However, with little knowledge of the full-length NLR structure or inter-/intra-molecular interactions, this is only speculation. Further investigations into the function and nature of host target(s) will shed light on these theories. The concept of diverse pathogens producing multiple effectors that converge on certain conserved target proteins has been discussed previously, with the speculation that using sequence-divergent effectors in a redundant manner to target certain host proteins might help to alleviate the threat posed by plant immune receptors – if a plant is able to recognise one effector, another effector might be able to achieve the same pathogenicity role without recognition (Weßling et al., 2014). As a striking example, two sequence-divergent effectors from *Pseudomonas aeruginosa* show significant functional convergence, as both can play a role in pathogen invasion of both prokaryotic and eukaryotic cells (Jiang et al., 2014).

There has been considerable recent progress in identifying putative effectors using bioinformatic approaches and identifying host targets using screening approaches (Dalio et al., 2018). However, the importance of using a range of methods to validate predicted function is clear, to ensure that false predictions from screening procedures are

uncovered. Part of this project involved the careful testing of an interaction between a pathogen effector, PexRD24, and its host target, PP1c, showing that an observation made in a yeast two-hybrid screening procedure was maintained for purified proteins in vitro. Next, obtaining structural information for how PexRD24 is binding to its PP1c host target would be extremely useful, alongside the characterisation of the interaction using biophysical techniques. This may help provide new information regarding the changes that PexRD24 causes when it interacts with PP1c, whether it causes changes to the overall structure of the protein, manipulates the catalytic site or makes an allosteric change. Crucially, this could provide clues about the role of the effector in pathogen infection. This research contributes to a growing field of knowledge about the effector biology of *P. infestans*.

Both the effectors that formed the major focus of this research (AVR-Pia and PexRD24) are part of distinct effector families with conserved characteristics. AVR-Pia is part of the MAX family, and contains a conserved core fold, while PexRD24 is part of the RxLR family, which contains a conserved sequence motif, predicted to be involved in effector translocation. Greater knowledge of the members of these conserved effector families may help with an understanding of their similarities and differences. While it is clear that a conserved structure does not necessarily equal a conserved function within these groups, further research will help to cement knowledge of why pathogens evolve numerous effectors around the same framework. It seems likely that the basic core structure confers stability or ease of translocation within the host (de Guillen et al., 2015). Continuing research could not just help to identify new effector proteins within these families, but perhaps identify new effector families with conserved structural and sequence motifs.

Both projects discussed in this thesis contribute to the general understanding of immunity, but particularly for the work on AVR-Pia recognition, there is potential to build on this in the future for engineering rice, and other, NLR immune receptors. The ability to manipulate the rice NLR Pik-1 to recognise AVR-Pia more strongly in the model plant *N. benthamiana* has potential for future experimentation. Although additional work is required to remove the autoactivity of the engineered receptor, this work could lead onto testing of the receptor in rice plants. Challenging the engineered rice with *M. oryzae* strains containing both AVR-PikD and AVR-Pia would test the efficacy of the system. By studying effector families with conserved features, it might be possible to discover ways that NLRs could recognise additional members of a family that contains a particular motif.

Alongside implications for engineering in the native host plant rice, it is known that integrated domains are found in the immune receptors of many other important crop plants, including wheat, maize and soybean (Sarris et al., 2016). Recent examples have shown the importance of integrated domains in a variety of these crops. For example, in wheat, NLR proteins containing a zinc-finger BED domain are known to provide resistance to another fungal disease, the yellow stripe rust fungus (Marchal et al., 2018). Engineering of integrated domains in one plant species could have wider implications, partly because some features might be applicable across different types of domain, but also because the same type of integrated domain is often found in multiple plant species. For example, heavy metal-associated domains are also predicted to be integrated in some NLRs found in rapeseed (Sarris et al., 2016). The use of effector proteins to target host phosphatase enzymes is also found beyond oomycetes, and examples are known in bacterial pathogens, such as *Ralstonia solanacearum*, a soil-borne pathogen with a wide host range. A recent example showed that a Type III effector from *R. solanacearum* could interfere with host cellular homeostasis by indirectly targeting protein phosphatase 2A (Popa et al., 2016). Phosphatase enzymes are so important in all forms of life that targeting by pathogenic organisms appears to be conserved in different pathosystems. Expanding the field of research in this area could have wider implications in the future; by using shared knowledge, the opportunities to develop new and broader resistance strategies for a variety of crops are exciting.

Gaining an increased understanding of molecular plant-pathogen interactions through structure/function studies brings us a step closer to engineering the plant immune system to better recognise and respond to pathogenic organisms.

Appendix 1

Using a new Golden Gate vector suite for expression of proteins from the *E. coli* host

Introduction

The choice of vector system for expressing proteins is important and a key first step in successful protein production. The choice of vector can depend on a number of factors, including ease of cloning and compatibility with other systems. Crucial for heterologous protein production is the consideration of, for example, appropriate promoters and solubility tags to enable expression of soluble, stable, high yielding proteins. Many different vector systems and cloning strategies have been developed, each of which may have different uses and advantages for a particular situation (Celie et al., 2016).

One versatile system is the pOPIN vector suite (Berrow et al., 2007), designed by the Oxford Protein Production Facility (OPPF). This collection of vectors enables a range of constructs to be produced, giving scope for different solubility tags at either the N- or C-terminus of the protein. An additional advantage of this system is that the vectors are suitable for use in a range of expression hosts, including *E. coli*, insect and human cell lines. Finally, the In-Fusion® (Takara Bio) cloning procedure is rapid and single-step, enabling efficient insertion of constructs into the pOPIN vectors. The pOPIN vector suite is currently used universally in the Banfield group for heterologous protein expression from *E. coli*, allowing production of proteins for structural and biochemical work.

Despite the pOPIN vector suite being well-established and successful within the lab, it does have a small number of drawbacks. To test different solubility tags, the construct in question needs to be separately cloned into different pOPIN vectors using the In-Fusion® enzyme for each reaction. The In-Fusion® mix is expensive, and if many constructs need to be screened, the cost could be a limitation. In addition, the Banfield group frequently uses in planta model systems to correlate biological data with results obtained in vitro. Currently, this requires recloning all constructs from pOPIN vectors into the vectors used for in planta transient expression. This is time consuming and uses additional resources.

In 2008, a new cloning strategy was devised that was heralded as a 'One Pot, One Step' technique and was named 'Golden Gate' (Engler et al., 2008). Golden Gate cloning uses Type IIS restriction enzymes that cleave outside of their recognition sequence, leaving four base pair overhangs that can be of any identity. By selecting these overhangs rationally, it is possible to create seamless joins in the construct without making any vector assembly scars. Carefully and specifically chosen overhangs allow all the different parts of the vector (promoters, tags, terminators etc.) to be mixed together with the acceptor vector at the same time (i.e. 'One Pot, One Step'), because the specific overhangs mean that the parts will only ligate together in the defined order. By developing an

agreed system of 4 bp overhangs, known as the MoClo modular cloning strategy (Weber et al., 2011), resources can be readily shared between groups, because all parts of the same type have the same overhangs (Patron et al., 2015). This presents the opportunity to mix and match parts such as promoters and tags without the need for additional steps (e.g. designing primers or PCRs), allowing greater variety with only a few standard DNA parts. Golden Gate cloning and the MoClo system are now used as standard within the lab for generating constructs for agroinfiltration and in planta studies. Assembling various parts in the acceptor vector is cost effective, because the same DNA parts can be used to make multiple constructs, requiring only the readily-available Type IIS restriction enzymes BsaI and BpiI, and T4 DNA ligase, which is used to reassemble the parts. In general, Golden Gate cloning is very efficient, although if a large number of parts are being inserted into the vector, this can reduce the efficiency and make cloning slightly more problematic.

Golden Gate cloning is therefore more cost-effective and somewhat more versatile than the pOPIN vector suite. A proposed strategy to take advantage of this was to make the pOPIN vector backbone compatible with a Golden Gate cloning system, which would allow constructs to be cloned for both *E. coli* and in planta expression in the same manner, whilst keeping the benefits of the pOPIN vector for driving protein expression from the *E. coli* host. This chapter describes my participation in the project to develop such a system and test its viability.

For the following work, Mark Youles (TSL SynBio) was responsible for design and generation of all vectors. Cloning and expression of different constructs was carried out by me in each case.

Designing the pOPIN/GG vector system

To generate a hybrid of the pOPIN and Golden Gate systems, the principle is to make an expression system that has the features of the pOPIN vector, but can be cloned into using the standard Golden Gate overhangs and cloning techniques. This new hybrid vector will be known as 'pOPIN/GG'. As described above, this combines the power of pOPIN for soluble protein production with the Golden Gate system, which is both cheaper to use and also compatible with the lab's work using in planta transient expression (see Chapter 3 and 4).

In the pOPIN system, different vectors are used to achieve constructs with different tags – e.g. pOPINF encodes an N-terminal 6xHis tagged construct, while pOPINS3C encodes

an N-terminal SUMO tagged construct and pOPINE encodes a C-terminal 6xHis tag. In the Golden Gate system, only one type of acceptor plasmid is required for cloning, because the acceptor backbone does not contain any promoters, terminators or tags; these must all be inserted during the digestion-ligation reaction when the coding sequence is inserted. In the case of the pOPIN/GG system, the vector would be a hybrid between these two cases. The pOPIN backbone already contains the T7 promoter and terminator for inducible expression in *E. coli*, so these parts do not need to be added during cloning. However, for versatility, the vector needs to be capable of receiving different solubility and affinity tags at the N- and C-termini of the construct. To achieve this using the Golden Gate overhang system, it is necessary to design two different pOPIN/GG vectors – one for C-terminal tagging and one for N-terminal tagging. For simplicity, only the vectors for N-terminal tagging will be discussed – namely vectors F1, F3 and F5. (Vectors for C-terminal tagging are named F2, F4 and F6, but will not be discussed further).

To make a complete pOPIN/GG construct containing a coding sequence and N-terminal tag, three parts are needed.

1) **The acceptor vector**. This acceptor plasmid is made from the pOPIN backbone, with a RFP (red fluorescence protein) Golden Gate-compatible cassette inserted in place of the native lacZ selection gene and plasmid-encoded His tag. This Golden Gate cassette contains the correct Golden Gate MoClo assembly standard overhangs to allow insertion of an N-terminal tag and coding sequence in the correct order. Namely, it uses the 'CCAT' 5' overhang for joining to the start of the N-terminal tag and the 'GCTT' overhang for joining the 3' end of the coding sequence into the vector. BsaI sites flank these overhangs to allow digestion during cloning. This process works in the same way as normal Golden Gate cloning, except that in this case, only two parts need to be inserted because the promoter and terminator are already in the backbone.

2) **The N-terminal tag**. Any Golden Gate compatible N-terminal tag can be used in the pOPIN/GG system, meaning that it is possible to recapitulate many members of the pOPIN suite with a single acceptor vector e.g. using an N-terminal 6xHis tag makes an equivalent construct to POPINF, an N-terminal 6xHis-SUMO tag would recapitulate pOPINS3C etc. N-terminal tags for Golden Gate cloning are readily available from TSL SynBio, using the compatible overhangs and BsaI sites as usual.

3) **The coding sequence**. One advantage of the pOPIN/GG system is that any Golden Gate compatible level 0 coding sequence can be cloned directly into the pOPIN/GG level 1 acceptor plasmid. For a lab that works both in vitro and in planta, this removes the

need for additional PCR steps, which would normally be required if testing a construct in planta using the Golden Gate system and in vitro using the pOPIN suite. Any coding sequence with compatible overhangs and BsaI sites can be used in the pOPIN/GG vectors.

The description above shows that, in theory, using the hybrid system of pOPIN and Golden Gate should be relatively straightforward. However, the system required development and testing to ensure it was working as expected. Three rounds of correction were required, and the adaptations and tests are described below.

Vector development and testing

For the first generation of the pOPIN/GG vector, now called F1, the Golden Gate-compatible cassette was inserted into the pOPIN backbone using EcoRI and BamHI restriction. Within the cassette, BpiI sites were included so that the constructs made in the pOPIN/GG plasmid (in effect a Golden Gate level 1 vector) could be cut out and cloned into a Golden Gate level 2 vector, using a digestion-ligation reaction, if required. This acceptor vector was then used to clone an N-terminal 6xHis tag and the coding sequence of a test protein CpTIE-TED, here abbreviated to 'TED'. TED was used as a test protein for these constructs because it is known to be highly expressed from the pOPINF vector (Walden et al., 2015) so provides a useful way to directly compare the pOPIN suite and pOPIN/GG vector. The main features of this construct are shown in figure 0-1.

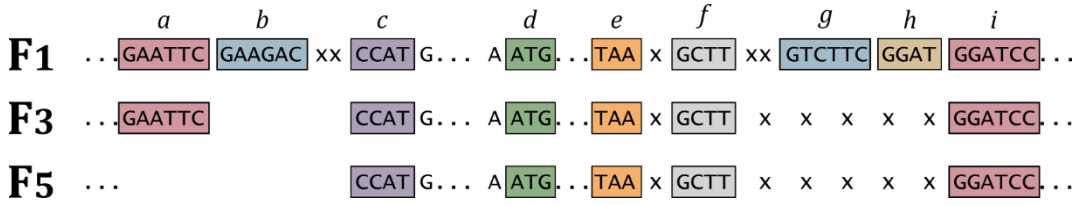


Figure 0-1: Features of the pOPIN/GG vectors.

Diagram showing the different features of the pOPIN/GG vector - versions F1, F3 and F5, used for generating an N-terminally tagged construct. '...' represents varying sequence of varying lengths, 'x' represents a single base of varying identity. Letters indicate different features:
a: EcoRI site
b: BpiI site
c: 'CCAT' Golden Gate vector overhang, incorporating the 'ATG' start codon for the N-terminal protein tag
d: 'AATG' Golden Gate overhang, incorporating the 'ATG' coding sequence start codon, which will present as a methionine residue scar in the expressed protein.
e: Stop codon for coding sequence
f: 'GCTT' Golden Gate vector overhang
g: BpiI site
h: Duplicated bases (redundant)
i: BanHI site

Figure 0-1 shows the restriction sites, BpiI sites, and overhangs used for joining the parts. One additional feature is four duplicated base pairs (labelled 'h') just inside the BamHI site. These bases were included through an error during cloning and perform no function. The stop codon at the end of the TED coding sequence is shown labelled 'e'. One important feature to note is that the 'AATG' overhang that joins the 3' end of the N-terminal tag to the 5' end of the TED coding sequence (labelled 'd') causes the translation of a methionine residue between the tag and the coding sequence, which remains as a scar in the final expressed protein.

Figure 0-2 shows the result of an expression trial from this F1 construct. TED was expressed from pOPINF, to act as a comparison. Both plasmids were transformed into an *E. coli* expression strain and a simple expression test was performed on a small scale. Cells were lysed to check for soluble expression, but the proteins were not purified. Figure 0-2 shows that for pOPINF, a clear band of overexpressed protein is visible by SDS-PAGE, around 28 kDa (the actual expected size of the construct is 23.3 kDa). This protein is present both in the lysed cells and in the soluble fraction, indicating that the TED protein is readily expressed from pOPINF, as previously published (Walden et al., 2015). However, for the pOPIN/GG vector, no overexpressed protein can be seen at the equivalent location on the gel. This test was repeated with other constructs (data not shown) and similar results were seen, indicating that expression from the pOPIN/GG vector was not working as expected.

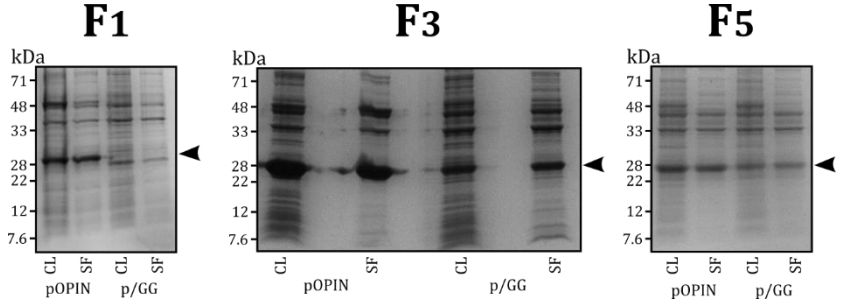


Figure 0-2: Testing expression from the pOPIN/GG vectors.

SDS-PAGE analysis of CpTIE-TED (TED) protein expression from *E. coli*, comparing the pOPINF vector and pOPIN/GG vector (versions F1, F3, F5). CL = crude Cell Lysate, SF = Soluble Fraction. Arrow indicates expected location of the TED protein on the gel. Note: For brevity, 'pOPIN/GG' is represented as 'p/GG'.

One explanation for the lack of expression is that the addition of the extra bases in the pOPIN/GG vector compared to pOPINF (principally the EcoRI/BamHI restriction sites for inserting the Golden Gate cassette, and the BpiI sites for putative onward cloning),

are causing issues with expression. The additional bases have been positioned between the T7 promoter and the transcription start site, which could lead to non-functional transcription. To help cut down the additional sequence between the promoter and the N-terminal tag, the non-essential BpiI sites were removed from the cassette. In addition, the four duplicated bases (labelled 'h' in figure 0-1 F1 diagram) were removed, as these are redundant. This adapted vector was called F3, and expression was tested as described for the F1 version. Figure 0-2 shows that in this case, there did appear to be soluble TED protein expression from the pOPIN/GG vector, as a band of overexpressed protein can be seen around 28 kDa in both the whole cell lysate and the soluble fraction. This suggests that reducing the number of additional bases between the promoter and expression construct have enabled transcription. However, the relative abundance of protein expressed from the pOPIN/GG system is much lower than from pOPINF. Repeated expression tests (data not shown) indicate that this is seen consistently.

Although expression was possible from the pOPIN/GG F3 vector, the lower levels of protein production could cause significant issues. TED proteins are highly soluble and stable, meaning that even with lower expression levels, there is still a clear protein band visible by SDS-PAGE. However, for a protein that is expressed at a lower level from the pOPIN vector suite, a depletion of around 50% from the pOPIN/GG system compared to pOPIN could mean that obtaining viable yields of soluble protein is not possible. Given that protein produced from *E. coli* is commonly used for biochemical assays or crystallography, using a vector that drives high levels of expression is very important. Therefore, the pOPIN/GG system is not a feasible replacement for the pOPIN suite unless it can produce comparable levels of protein. Further modifications were designed to make the pOPIN/GG vector almost identical to its pOPIN parent. To do this, the unwanted EcoRI site was cut out using an EcoRV/XbaI exchange, resulting in a plasmid that was free of the 5' EcoRI scarring. The pOPIN template vector was also found to naturally contain the required 'CCAT' 5' overhang, meaning that the fusion point between the pOPIN backbone and the Golden Gate N-terminal tag became completely seamless (figure 0-1, version F5). Having this portion of the vector completely identical to pOPINF should hopefully allow the promoter to drive expression at the same levels. Upon testing, it was found that this was indeed the case (figure 0-2) because the expression of soluble TED protein is now similar between pOPINF and pOPIN/GG F5 version. From the SDS-PAGE analysis shown in figure 0-2, it does seem that the amount of protein expressed from pOPIN/GG F5 may be slightly lower than from pOPINF, but the levels are comparable. The discrepancy may be due to some of the other, minor differences between the two vectors. The pOPIN vector backbone did require some domestication

before being used for Golden Gate cloning (all BsaI and BpiI restriction sites in the backbone must be removed, or the restriction enzymes will cleave the vector backbone during the digestion-ligation reaction). However, overall, the pOPIN/GG vector F5 version is functional after the modifications made.

Mep1 protein expression using pOPIN/GG

The 'TED' protein described previously, which was used to test the viability of the pOPIN/GG vector, has been extensively tested before in the lab (Walden et al., 2015), and is known to be stable and highly expressed. Next, the new pOPIN/GG system (the F5 vector) was used for expression of a protein that was unfamiliar. This protein, Mep1, is an effector protein from *Magnaporthe oryzae* (Yan & Talbot, 2016) and is largely uncharacterised (plasmid provided by the group of N. Talbot). Mep1 was cloned into pOPIN/GG F5 with an N-terminal 6xHis tag, and was expressed from BL21(DE3) *E. coli* in LB media. The effector was purified using an IMAC/GF technique, and figure 0-3 shows the final gel filtration step during purification, with accompanying SDS-PAGE analysis. Intact mass spectrometry analysis (data not shown) showed that full-length Mep1 was successfully purified, indicating that the pOPIN/GG system also works for a different, previously uncharacterised protein.

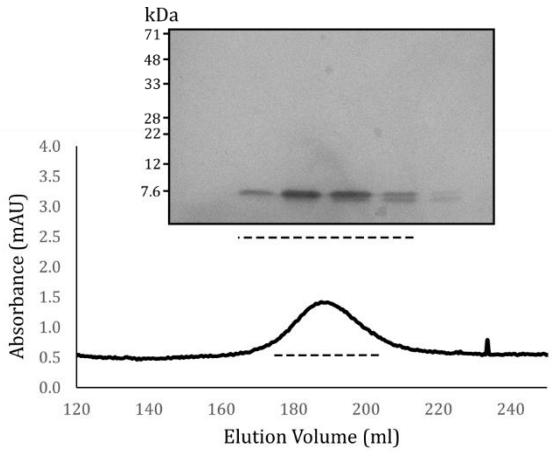


Figure 0-3: Expression of Mep1 from the pOPIN/GG vector F5.

Following cleavage of 6xHis tag by 3C protease, Mep1 was flowed down a Superdex 75 26/600 gel filtration column and showed a small peak on the trace at 189 ml elution volume (poor protein absorbance at 280 nm). Accompanying SDS-PAGE analysis shows fractions taken from under the peak.

Discussion

The aim of this project was the development of a vector that was suitable for expressing high yields of soluble protein in the *E. coli* heterologous expression system, whilst also being compatible with the Golden Gate cloning method used to produce constructs for in planta studies. Several iterations and rounds of testing were required to produce a functional vector. It was found that non-existent or low levels of transcription were caused by additional bases that were inserted into the pOPIN backbone to make it Golden Gate compatible. Removal of all additional bases between the promoter region and transcription start site enabled functionality to be restored. Cloning of constructs into the new vector is simple and efficient, as only two parts (the coding sequence and tag) need to be inserted into the acceptor vector. One point of note is that the pOPIN/GG cloning method is not completely scarless. Due to the nature of the overhangs in the MoClo assembly standard, a methionine residue is made in the expressed protein between the N-terminal tag and the protein of interest (POI). The result of this is that when the protein is cleaved from its purification tag using 3C protease, the remaining scar is three residues long 'Gly-Pro-Met-POI' as opposed to the two-residue scar on a protein expressed from a pOPIN vector ('Gly-Pro-POI'). It is hoped that this additional residue will not cause any issues for the purified protein. This chapter has described the successful use of pOPIN/GG to express two proteins with an N-terminal 6xHis tag, but the vector has the capacity to accept a range of other N-terminal tags. Future work and testing by others will explore the use of other tags, and their relative success compared to equivalent vectors from the pOPIN suite. As mentioned, equivalent versions of each pOPIN/GG vector were generated for making C-terminally tagged constructs. Initial tests have indicated that these vectors are working correctly, but further work is required.

Overall, the design and generation of the new pOPIN/GG vector system has been a success, and provides an alternative for protein expression from *E. coli* that has added Golden Gate compatibility.

Appendix 2

List of DNA Primers used throughout the project

List of DNA primers used throughout project		
Name	**Sequence**	**Use**
PP1c A1-1F	AGGAGATATACCATGGCTCAAAATGGGCAGGGGATAGAAC	Cloning PP1c into pOPINA
PP1c A1-1R	GTGGTGGTGGTGTTTTCACAAGAACCGAGGTTTTCTATCAGC	Cloning PP1c into pOPINA
PP1c A1-2F	AGGAGATATACCATGGCACAAAATGAGCATCAGCAGCAG	Cloning PP1c into pOPINA
PP1c A1-2R	GTGGTGGTGGTGTTTTCATAAAAACCGAGGTTTTCTATCTG	Cloning PP1c into pOPINA
PP1c A3-2F	AGGAGATATACCATGGACCCTGCAGCTGTCGATAGGATTATTG	Cloning PP1c into pOPINA
PP1c A3-2R	GTGGTGGTGGTGTTTTCACATCATAAACTTATTTTTCTTCTCTG	Cloning PP1c into pOPINA
PP1c D2-2F	AGGAGATATACCATGGACCAGAATGTGTTGGATGATATAATAAC	Cloning PP1c into pOPINA
PP1c D2-2R	GTGGTGGTGGTGTTTTCATGCTTTGGAATTAAAAAAGGACTTC	Cloning PP1c into pOPINA
PP1c A2-1truncF	AGGAGATATACCATGGGATTAATAGAGCCTGCTGTTCTTG	Cloning PP1c into pOPINA
PP1c A2-1truncR	GTGGTGGTGGTGTTTTCATGGCTTCAGTATCTGGAACGAACAC	Cloning PP1c into pOPINA
D2_35148_f_1059	AAGTTCTGTTTCAGGGCCCGGACCAGAATGTGTTGGATGATATAATAAC	Cloning truncated PP1c-3
D2_35148_r_1060	ATGGTCTAGAAAGCTTTAAGTAGTTGTACTTCCAAAGCTGAATTTGG	Cloning truncated PP1c-3
D2_34492_f_1061	AAGTTCTGTTTCAGGGCCCGGATGATATAATAACTAGGCTTCTTGAAG	Cloning truncated PP1c-3
D2_34492_r_1062	ATGGTCTAGAAAGCTTTATTTAGCAGTAGTTGTACTTCCAAAGCTG	Cloning truncated PP1c-3
4314_trunc_f_1088	AAGTTCTGTTTCAGGGCCCGAGCTTCAAGCTGAAGAGGTTTGGAG	Cloning truncated PexRD24
4314_8225_r_1089	ATGGTCTAGAAAGCTTTACGAGTTGGTTTTGTAGATACGAGCTCG	Cloning truncated PexRD24
4314_7631_r_1090	ATGGTCTAGAAAGCTTTAGATACGAGCTCGGACTTTGTTCGA	Cloning truncated PexRD24
PP1cD2-1FF_982c	AAGTTCTGTTTCAGGGCCCGATGGACCAGAATGTGTTGGATGATATAATAAC	Cloning PP1c-3into pOPINS3C/F
PP1cD2-1FR_983c	ATGGTCTAGAAAGCTTTATGCTTTGGAATTAAAAAAGGACTTC	Cloning PP1c-3into pOPINS3C/F
Oligo 04314_298FF	AAGTTCTGTTTCAGGGCCCGGATTTGCTTGGGCTTTTTGCCAAGAGCAAGC	Cloning PexRD24 into pOPINS3C/F
Oligo 04314_299FR	ATGGTCTAGAAAGCTTTACGAGTTGGTTTTGTAGATACGAGCTGGGAC	Cloning PexRD24 into pOPINS3C/F

D2QILK_r_1167	ATGGTCTAGAAAGCTTTATTTTAATATTTGAAAGGAGCACATTAATG	Cloning truncated PP1c-3
1275_Mep1_fwd	AAGGTCTCTAATGCATGAGGCCAGCGCCGGTACC	Cloning Mep1 into pOPIN/GG
1276_Mep1_rvs	AAGGTCTCTAAGCTTAGATACCATTGTCTTTATGGTCGGGCCT	Cloning Mep1 into pOPIN/GG
1303_PiaPikDN_S3C_fwd	AAG TTC TGT TTC AGG GCC CGG AAA CGG GCA ACA AAT ATA TAG	Cloning AVR-PiaNAVR-PikD into pOPINS3C
1304_PiaPikDN_S3C_rvs	ATG GTC TAG AAA GCT TTA GTA AGG CTC GGC AGC AAG CC	Cloning AVR-PiaNAVR-PikD into pOPINS3C
1316_TEDGG_fwd	AAGGTCTCTAATGTCTCCGACGGGTGGTCTGCAAAC	Cloning TED into pOPIN/GG
1317_TEDGG_rvs	AAGGTCTCTAAGCTTACGGCGTCACGAATTTGGTGCCC	Cloning TED into pOPIN/GG
1374_AVRPikD-N_S3C_fwd	AAGTTCTGTTTCAGGGCCCGGAATGCTTCTGGTTCATGTTC	Cloning truncated AVR-PikD into pOPINS3C
704-AVR-Pik-Rev	ATGGTCTAGAAAGCTTTAGAAGCCCGGACGTTTTTTACC	Cloning truncated AVR-PikD into pOPINS3C
1375_AVRPikD-N_A_fwd	AGGAGATATACCATGGAATGCTTCTGGTTCATGTTCAAAAAC	Cloning truncated AVR-PikD into pOPINA
959RevAVRFLpOPINA	GTGGTGGTGGTGTTTTTAGAAGCCCGGACGTTTTTTAC	Cloning truncated AVR-PikD into pOPINA
PikD-For-pOPINE-962	AGGAGATATACCATGGGGAGCGATAGCGAAGTGAAC	Cloning AVR-Pik into pOPINE
1385_AVRPia_pOPINE_rvs	GTGATGGTGATGTTTGTAAGGCTCGGCAGCAAGCC	Cloning AVR-Pia into pOPINE
PikD-Rev-pOPINE-963	GTGATGGTGATGTTTGAAGCCCGGACGTTTTTTAC	Cloning AVR-Pik into pOPINE
1467_AVRPikD-NE1_S3C_fwd	AAGTTCTGTTTCAGGGCCCGGGCATCCCGGTCCCGGAATGC	Cloning truncated AVR-PikD into pOPINS3C
1468_AVRPikD-NE1_A_fwd	AGGAGATATACCATGGGCATCCCGGTCCCGGAATGC	Cloning truncated AVR-PikD into pOPINA
1469_AVRPikD-NE2_S3C_fwd	AAGTTCTGTTTCAGGGCCCGGACCATCCGGGCATCCCGG	Cloning truncated AVR-PikD into pOPINS3C
1470_AVRPikD-NE2_A_fwd	AGGAGATATACCATGGACCATCCGGGCATCCCGGTCCC	Cloning truncated AVR-PikD into pOPINA
1497_AVRPiaPikDN_GGN_fwd	AAGGTCTCTAATGGAAACGGGCAACAAATATATAGAAAAACGCGC	Cloning AVR-PiaNAVR-PikD into lvl1 GG
1498_AVRPia_GGN_fwd	AAGGTCTCTAATGGCGCCAGCTAGATTTTGCGTCTATTACG	Cloning AVR-Pia into lvl1 GG
1499_AVRPia_GGN_rvs	AAGGTCTCTAAGCCTAGTAAGGCTCGGCAGCAAGCCAATCCGG	Cloning AVR-Pia into lvl1 GG
1510_PikDN_GGN_fwd	AAGGTCTCTAATGGAATGTTTTTGGTTTATGTTTAAAAACAACGTACG	Cloning truncated AVR-PikD into lvl1 GG
1658_RGA5SWAP6_M_rvs	ATGGTCTAGAAAGCTTTACTCAACTAACTCCACCATCAAGAGCTCGGCG	Cloning RGA5/Pikp-HMA Swap6 into pOPINM

718-Pik-p1-For	AAGTTCTGTTTCAGGGCCCGGGTCTGAAACAAAAAATCGT	Cloning Pikp-HMA into pOPINM
1164_Pikp+5aa_rev	ATGGTCTAGAAAGCTTTAATCTTTATTTGCTTGCGAGACTTGCAGCAGTTC	Cloning Pikp-HMA into pOPINM
1719_RGA5fullswap_fwd	TTGGTCTCACAGAAAGATAGTTGTTAAGGTGCACATGCCATGCGG	Cloning RGA5-HMA into pCR8
1720_RGA5fullswap_rvs	AAGGTCTCACGTCTTTCTCAACTAACTCCACCATCAAGAGCTCGGCG	Cloning RGA5-HMA into pCR8
GW1F	GTTGCAACAAATTGATGAGCAATGC	Sequencing pCR8
GW2R	GTTGCAACAAATTGATGAGCAATTA	Sequencing pCR8
M13F	TGTAAAACGACGGCCAGT	Sequencing pCR8
M13R	CAGGAAACAGCTATGACC	Sequencing pCR8
T7 promoter	TAATACGACTCACTATAGGG	Sequencing pOPIN
T7 terminator	TAGTTATTGCTCAGCGGTGG	Sequencing pOPIN
T7 terminator (pOPINRev)	CACCACCTTCTGATAGGCAG	Sequencing pOPIN
GG_level0_F (1180)	CGTTATCCCCTGATTCTGTGGATAAC	Sequencing GG
GG_level0_R (1181)	GTCTCATGAGCGGATACATATTTGAATG	Sequencing GG
level_1_R (1182)	GAACCCTGTGGTTGGCATGCACATAC	Sequencing GG
level_1_F (1183)	CTGGTGGCAGGATATATTGTGGTG	Sequencing GG
pCambiaF	GGCTCGTATGTTGTGTGG	Sequencing pCambia
1076_PikP1 int 1	TGGAGGCAAAACAACGG	Sequencing Pikp-1
1077_PikP1 int 2	GGTCGTCTGGTCAGGAGG	Sequencing Pikp-1
1078_PikP1 int 3	ATTTGGAGATTTTGTATGTGG	Sequencing Pikp-1
1081-PikP1 int seq	GTTGCGAGCACTGGAGG	Sequencing Pikp-1
1097 Pikp1 int seq rev	GAAGTGCGTTCCCAAGG	Sequencing Pikp-1
1248_AllHMAmutants_p1_Fwd	AAGGTCTCACAGAAAAACAGGGCTAAAGCAAAAAATCGTG	Sequencing Pikp-1

References

Abreu-Neto, J. B., Turchetto-Zolet, A. C., Oliveira, L. F. V., Bodanese Zanettini, M. H., & Margis-Pinheiro, M. (2013). Heavy metal-associated isoprenylated plant protein (HIPP): characterization of a family of proteins exclusive to plants. *The FEBS Journal, 280*(7), 1604-1616.

Allen, P. B., Kwon, Y. G., Nairn, A. C., & Greengard, P. (1998). Isolation and characterization of PNUTS, a putative protein phosphatase 1 nuclear targeting subunit. *Journal of Biological Chemistry, 273*(7), 4089-4095.

Andolfo, G., & Ercolano, M. R. (2015). Plant innate immunity multicomponent model. *Frontiers in Plant Science, 6*, 987.

Ariya-anandech, K., Chaipanya, C., Teerasan, W., Kate-Ngam, S., & Jantasuriyarat, C. (2018). Detection and allele identification of rice blast resistance gene, Pik, in Thai rice germplasm. *Agriculture and Natural Resources, 52*(6), 525-535.

Ashikawa, I., Hayashi, N., Yamane, H., Kanamori, H., Wu, J., Matsumoto, T., Ono, K., & Yano, M. (2008). Two adjacent Nucleotide-Binding Site–Leucine-Rich Repeat class genes are required to confer Pikm-specific rice blast resistance. *Genetics, 180*(4), 2267-2276.

Askari, N., Correa, R. G., Zhai, D., & Reed, J. C. (2012). Expression, purification, and characterization of recombinant NOD1 (NLRC1), a NLR family member. *Journal of Biotechnology, 157*(1), 75-81.

Baggs, E., Dagdas, G., & Krasileva, K. V. (2017). NLR diversity, helpers and integrated domains: making sense of the NLR IDentity. *Current Opinion in Plant Biology, 38*, 59-67.

Bai, S., Liu, J., Chang, C., Zhang, L., Maekawa, T., Wang, Q., Xiao, W., Liu, Y., Chai, J., Takken, F. L. W., Schulze-Lefert, P., & Shen, Q.-H. (2012). Structure-function analysis of barley NLR immune receptor MLA10 reveals its cell compartment specific activity in cell death and disease resistance. *PLoS Pathogens, 8*(6), e1002752.

Bebber, D. P. (2015). Range-expanding pests and pathogens in a warming world. *Annual Review of Phytopathology, 53*(1), 335-356.

Bent, A. F., & Mackey, D. (2007). Elicitors, effectors, and R genes: The new paradigm and a lifetime supply of questions. *Annual Review of Phytopathology, 45*(1), 399-436.

Bentham, A., Burdett, H., Anderson, P. A., Williams, S. J., & Kobe, B. (2017). Animal NLRs provide structural insights into plant NLR function. *Annals of Botany, 119*(5), 698-702.

Bernoux, M., Burdett, H., Williams, S. J., Zhang, X., Chen, C., Newell, K., Lawrence, G. J., Kobe, B., Ellis, J. G., Anderson, P. A., & Dodds, P. N. (2016). Comparative analysis of the flax immune receptors L6 and L7 suggests an equilibrium-based switch activation model. *The Plant Cell, 28*(1), 146-159.

Bernoux, M., Ellis, J. G., & Dodds, P. N. (2011). New insights in plant immunity signaling activation. *Current Opinion in Plant Biology, 14*(5), 512-518.

Bernoux, M., Ve, T., Williams, S., Warren, C., Hatters, D., Valkov, E., Zhang, X., Ellis, J. G., Kobe, B., & Dodds, P. N. (2011). Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. *Cell Host Microbe, 9*(3), 200-211.

Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I., & Owens, R. J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research, 35*(6), e45-e45.

Bialas, A., Zess, E. K., De la Concepcion, J. C., Franceschetti, M., Pennington, H. G., Yoshida, K., Upson, J. L., Chanclud, E., Wu, C. H., Langner, T., Maqbool, A., Varden, F. A., Derevnina, L., Belhaj, K., Fujisaki, K., Saitoh, H., Terauchi, R., Banfield, M. J., & Kamoun, S. (2018). Lessons in effector and NLR biology of plant-microbe systems. *Molecular Plant-Microbe Interactions, 31*(1), 34-45.

Birch, P. R., Rehmany, A. P., Pritchard, L., Kamoun, S., & Beynon, J. L. (2006). Trafficking arms: oomycete effectors enter host plant cells. *Trends in Microbiology, 14*(1), 8-11.

Block, A., Toruño, T. Y., Elowsky, C. G., Zhang, C., Steinbrenner, J., Beynon, J. L., & Alfano, J. R. (2014). The *Pseudomonas syringae* type III effector HopD1 suppresses effector-triggered immunity, localizes to the endoplasmic reticulum, and targets the *Arabidopsis* transcription factor NTL9. *New Phytologist, 201*(4), 1358-1370.

Boevink, P. C., McLellan, H., Gilroy, E. M., Naqvi, S., He, Q., Yang, L., Wang, X., Turnbull, D., Armstrong, M. R., Tian, Z., & Birch, P. R. J. (2016). Oomycetes seek help from the plant: *Phytophthora infestans* effectors target host susceptibility factors. *Molecular Plant, 9*(5), 636-638.

Boevink, P. C., Wang, X., McLellan, H., He, Q., Naqvi, S., Armstrong, M. R., Zhang, W., Hein, I., Gilroy, E. M., Tian, Z., & Birch, P. R. (2016). A *Phytophthora infestans* RXLR effector targets plant PP1c isoforms that promote late blight disease. *Nature Communications, 7*, 10311.

Bollen, M., Peti, W., Ragusa, M. J., & Beullens, M. (2010). The extended PP1 toolkit: designed to create specificity. *Trends in Biochemical Sciences, 35*(8), 450-458.

Boller, T., & Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual Review of Plant Biology, 60*, 379-406.

Bonardi, V., Tang, S., Stallmann, A., Roberts, M., Cherkis, K., & Dangl, J. L. (2011). Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proceedings of the National Academy of Sciences, 108*(39), 16463-16468.

Bos, J. I. B., Armstrong, M. R., Gilroy, E. M., Boevink, P. C., Hein, I., Taylor, R. M., Zhendong, T., Engelhardt, S., Vetukuri, R. R., Harrower, B., Dixelius, C., Bryan, G., Sadanandom, A., Whisson, S. C., Kamoun, S., & Birch, P. R. J. (2010). *Phytophthora infestans* effector AVR3a is essential for virulence and manipulates plant immunity by stabilizing host E3 ligase CMPG1. *Proceedings of the National Academy of Sciences, 107*(21), 9909-9914.

Boutemy, L. S., King, S. R. F., Win, J., Hughes, R. K., Clarke, T. A., Blumenschein, T. M. A., Kamoun, S., & Banfield, M. J. (2011). Structures of *Phytophthora* RXLR effector proteins: A conserved but adaptable fold underpins functional diversity. *Journal of Biological Chemistry, 286*(41), 35834-35842.

Bozkurt, T. O., Schornack, S., Win, J., Shindo, T., Ilyas, M., Oliva, R., Cano, L. M., Jones, A. M. E., Huitema, E., van der Hoorn, R. A. L., & Kamoun, S. (2011). *Phytophthora infestans* effector AVRblb2 prevents secretion of a plant immune protease at the haustorial interface. *Proceedings of the National Academy of Sciences, 108*(51), 20832-20837.

Breen, S., Williams, S. J., Winterberg, B., Kobe, B., & Solomon, P. S. (2016). Wheat PR-1 proteins are targeted by necrotrophic pathogen effector proteins. *The Plant Journal, 88*(1), 13-25.

Casey, L. W., Lavrencic, P., Bentham, A. R., Cesari, S., Ericsson, D. J., Croll, T., Turk, D., Anderson, P. A., Mark, A. E., Dodds, P. N., Mobli, M., Kobe, B., & Williams, S. J. (2016). The CC domain structure from the wheat stem rust resistance protein Sr33 challenges paradigms for dimerization in plant NLR proteins. *Proceedings of the National Academy of Sciences, 113*(45), 12856-12861.

Celie, P. H., Parret, A. H., & Perrakis, A. (2016). Recombinant cloning strategies for protein expression. *Current Opinion in Structural Biology, 38*, 145-154.

Cesari, S. (2018). Multiple strategies for pathogen perception by plant immune receptors. *New Phytologist, 219*(1), 17-24.

Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T., & Dodds, P. N. (2014). A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. *Frontiers in Plant Science, 5*, 606.

Cesari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y., Shimamoto, K., Dodds, P., Terauchi, R., & Kroj, T. (2014). The NB-LRR proteins RGA4 and RGA5

interact functionally and physically to confer disease resistance. *EMBO Journal, 33*(17), 1941-1959.

Cesari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A., Rivas, S., Alaux, L., Kanzaki, H., Okuyama, Y., Morel, J.-B., Fournier, E., Tharreau, D., Terauchi, R., & Kroj, T. (2013). The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. *The Plant Cell, 25*(4), 1463-1481.

Chang, J. H., Urbach, J. M., Law, T. F., Arnold, L. W., Hu, A., Gombar, S., Grant, S. R., Ausubel, F. M., & Dangl, J. L. (2005). A high-throughput, near-saturating screen for type III effector genes from *Pseudomonas syringae*. *Proceedings of the National Academy of Sciences, 102*(7), 2549-2554.

Chaparro-Garcia, A., Schwizer, S., Sklenar, J., Yoshida, K., Petre, B., Bos, J. I. B., Schornack, S., Jones, A. M. E., Bozkurt, T. O., & Kamoun, S. (2015). *Phytophthora infestans* RXLR-WY effector AVR3a associates with Dynamin-Related Protein 2 required for endocytosis of the Plant Pattern Recognition Receptor FLS2. *PLoS One, 10*(9), e0137071.

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 1), 12-21.

Choi, H. W., & Klessig, D. F. (2016). DAMPs, MAMPs, and NAMPs in plant innate immunity. *BMC Plant Biology, 16*(1), 232.

Choy, M. S., Hieke, M., Kumar, G. S., Lewis, G. R., Gonzalez-DeWhitt, K. R., Kessler, R. P., Stein, B. J., Hessenberger, M., Nairn, A. C., Peti, W., & Page, R. (2014). Understanding the antagonism of retinoblastoma protein dephosphorylation by PNUTS provides insights into the PP1 regulatory code. *Proceedings of the National Academy of Sciences, 111*(11), 4097-4102.

Cohen, P. T. (2002). Protein phosphatase 1--targeted in many directions. *Journal of Cell Science, 115*(Pt 2), 241-256.

Cook, D. E., Mesarich, C. H., & Thomma, B. P. H. J. (2015). Understanding plant immunity as a surveillance system to detect invasion. *Annual Review of Phytopathology, 53*(1), 541-563.

Costanzo, S., & Jia, Y. (2010). Sequence variation at the rice blast resistance gene Pi-km locus: Implications for the development of allele specific markers. *Plant Science, 178*(6), 523-530.

Couto, D., Niebergall, R., Liang, X., Bücherl, C. A., Sklenar, J., Macho, A. P., Ntoukakis, V., Derbyshire, P., Altenbach, D., Maclean, D., Robatzek, S., Uhrig, J., Menke, F., Zhou, J.-M., & Zipfel, C. (2016). The *Arabidopsis* protein phosphatase PP2C38 negatively regulates the central immune kinase BIK1. *PLoS Pathogens, 12*(8), e1005811.

Cui, H., Tsuda, K., & Parker, J. E. (2015). Effector-triggered immunity: From pathogen perception to robust defense. *Annual Review of Plant Biology, 66*(1), 487-511.

Dalio, R. J. D., Herlihy, J., Oliveira, T. S., McDowell, J. M., & Machado, M. (2018). Effector biology in focus: A primer for computational prediction and functional characterization. *Molecular Plant-Microbe Interactions, 31*(1), 22-33.

Dangl, J. L., & Jones, J. D. G. (2001). Plant pathogens and integrated defence responses to infection. *Nature, 411*, 826.

de Guillen, K., Ortiz-Vallejo, D., Gracy, J., Fournier, E., Kroj, T., & Padilla, A. (2015). Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi. *PLoS Pathogens, 11*(10), e1005228.

de la Concepcion, J. C., Franceschetti, M., Maqbool, A., Saitoh, H., Terauchi, R., Kamoun, S., & Banfield, M. J. (2018). Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen. *Nature Plants, 4*(8), 576-585.

Deslandes, L., & Rivas, S. (2012). Catch me if you can: bacterial effectors and plant targets. *Trends in Plant Science, 17*(11), 644-655.

Dodds, P. N., & Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews: Genetics, 11*, 539-548.

Dong, S., Yin, W., Kong, G., Yang, X., Qutob, D., Chen, Q., Kale, S. D., Sui, Y., Zhang, Z., Dou, D., Zheng, X., Gijzen, M., M. Tyler, B., & Wang, Y. (2011). *Phytophthora sojae* avirulence effector Avr3b is a secreted NADH and ADP-ribose pyrophosphorylase that modulates plant immunity. *PLoS Pathogens, 7*(11), e1002353.

dos Reis, S. P., Lima, A. M., & de Souza, C. R. B. (2012). Recent molecular advances on downstream plant responses to abiotic stress. *International Journal of Molecular Sciences, 13*(7), 8628.

Du, Y., Berg, J., Govers, F., & Bouwmeester, K. (2015). Immune activation mediated by the late blight resistance protein R1 requires nuclear localization of R1 and the effector AVR1. *New Phytologist, 207*(3), 735-747.

Egloff, M. P., Cohen, P. T., Reinemer, P., & Barford, D. (1995). Crystal structure of the catalytic subunit of human protein phosphatase 1 and its complex with tungstate. *Journal of Molecular Biology, 254*(5), 942-959.

Eitas, T. K., & Dangl, J. L. (2010). NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Current Opinion in Plant Biology, 13*(4), 472-477.

El Kasmi, F., Chung, E.-H., Anderson, R. G., Li, J., Wan, L., Eitas, T. K., Gao, Z., & Dangl, J. L. (2017). Signaling from the plasma-membrane localized plant immune receptor RPM1 requires self-association of the full-length protein. *Proceedings of the National Academy of Sciences, 114*(35), E7385-E7394.

El Kasmi, F., & Nishimura, M. T. (2016). Structural insights into plant NLR immune receptor function. *Proceedings of the National Academy of Sciences, 113*(45), 12619-12621.

Ellis, J. G. (2016). Integrated decoys and effector traps: how to catch a plant pathogen. *BMC Biology, 14*(1), 13.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 4), 486-501.

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS One, 3*(11), e3647.

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr, 62*(Pt 1), 72-82.

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr, 69*(Pt 7), 1204-1214.

Fan, G., Yang, Y., Li, T., Lu, W., Du, Y., Qiang, X., Wen, Q., & Shan, W. (2018). A *Phytophthora capsici* RXLR effector targets and inhibits a plant PPIase to suppress endoplasmic reticulum-mediated immunity. *Molecular Plant, 11*(8), 1067-1083.

Feng, J., Chen, Y., Pu, J., Yang, X., Zhang, C., Zhu, S., Zhao, Y., Yuan, Y., Yuan, H., & Liao, F. (2011). An improved malachite green assay of phosphate: mechanism and application. *Analytical Biochemistry, 409*(1), 144-149.

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., & Gurr, S. J. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature, 484*, 186.

Flor, H. H. (1971). Current status of the Gene-for-Gene concept. *Annual Review of Phytopathology, 9*(1), 275-296.

Franceschetti, M., Maqbool, A., Jiménez-Dalmaroni, M. J., Pennington, H. G., Kamoun, S., & Banfield, M. J. (2017). Effectors of filamentous plant pathogens: Commonalities amid diversity. *Microbiology and Molecular Biology Reviews, 81*(2).

Frescatada-Rosa, M., Robatzek, S., & Kuhn, H. (2015). Should I stay or should I go? Traffic control for plant pattern recognition receptors. *Current Opinion in Plant Biology, 28*, 23-29.

Fry, W. (2008). *Phytophthora infestans*: the plant (and R gene) destroyer. *Molecular Plant Pathology, 9*(3), 385-402.

Fukao, Y. (2012). Protein–protein interactions in plants. *Plant and Cell Physiology, 53*(4), 617-625.

Fukuoka, S., Saka, N., Koga, H., Ono, K., Shimizu, T., Ebana, K., Hayashi, N., Takahashi, A., Hirochika, H., Okuno, K., & Yano, M. (2009). Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science, 325*(5943), 998-1001.

Garman, E. F., & Owen, R. L. (2006). Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr, 62*(Pt 1), 32-47.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. e., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In J. M. Walker (Ed.), *The Proteomics Protocols Handbook* (pp. 571-607). Totowa, NJ: Humana Press.

Gerland, P., Raftery, A. E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B. K., Chunn, J., Lalic, N., Bay, G., Buettner, T., Heilig, G. K., & Wilmoth, J. (2014). World population stabilization unlikely this century. *Science, 346*(6206), 234-237.

Giraldo, M. C., Dagdas, Y. F., Gupta, Y. K., Mentlak, T. A., Yi, M., Martinez-Rocha, A. L., Saitoh, H., Terauchi, R., Talbot, N. J., & Valent, B. (2013). Two distinct secretion systems facilitate tissue invasion by the rice blast fungus *Magnaporthe oryzae*. *Nature Communications, 4*, 1996.

Goldberg, J., Huang, H. B., Kwon, Y. G., Greengard, P., Nairn, A. C., & Kuriyan, J. (1995). Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature, 376*(6543), 745-753.

Gómez-Gómez, L., & Boller, T. (2000). FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Molecular Cell, 5*(6), 1003-1011.

Goss, E. M., Tabima, J. F., Cooke, D. E. L., Restrepo, S., Fry, W. E., Forbes, G. A., Fieland, V. J., Cardenas, M., & Grünwald, N. J. (2014). The Irish potato famine pathogen *Phytophthora infestans* originated in central Mexico rather than the Andes. *Proceedings of the National Academy of Sciences, 111*(24), 8791-8796.

Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols, 1*(6), 2876-2890.

Grouffaud, S., Whisson, S. C., Birch, P. R. J., & van West, P. (2010). Towards an understanding on how RxLR-effector proteins are translocated from oomycetes into host cells. *Fungal Biology Reviews, 24*(1), 27-36.

Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M. V., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I. B., Bulone, V., Cai, G., Cakir, C., Carrington, J. C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M. A., Fugelstad, J., Gilroy, E. M., Gnerre, S., Green, P. J., Grenville-Briggs, L. J., Griffith, J.,

Grünwald, N. J., Horn, K., Horner, N. R., Hu, C.-H., Huitema, E., Jeong, D.-H., Jones, A. M. E., Jones, J. D. G., Jones, R. W., Karlsson, E. K., Kunjeti, S. G., Lamour, K., Liu, Z., Ma, L., MacLean, D., Chibucos, M. C., McDonald, H., McWalters, J., Meijer, H. J. G., Morgan, W., Morris, P. F., Munro, C. A., O'Neill, K., Ospina-Giraldo, M., Pinzón, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D. C., Schumann, U. D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., van de Vondervoort, P. J. I., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B. C., van West, P., Ristaino, J., Govers, F., Birch, P. R. J., Whisson, S. C., Judelson, H. S., & Nusbaum, C. (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature, 461*, 393.

Hajdukiewicz, P., Svab, Z., & Maliga, P. (1994). The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. *Plant Molecular Biology, 25*(6), 989-994.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology, 166*(4), 557-580.

Heath, M. C. (2000). Nonhost resistance and nonspecific plant defenses. *Current Opinion in Plant Biology, 3*(4), 315-319.

Henry, E., Toruño, T. Y., Jauneau, A., Deslandes, L., & Coaker, G. L. (2017). Direct and indirect visualization of bacterial effector delivery into diverse plant cell types during infection. *The Plant Cell*.

Heroes, E., Lesage, B., Gornemann, J., Beullens, M., Van Meervelt, L., & Bollen, M. (2013). The PP1 binding code: a molecular-lego strategy that governs specificity. *The FEBS Journal, 280*(2), 584-595.

Heroes, E., Rip, J., Beullens, M., Van Meervelt, L., De Gendt, S., & Bollen, M. (2015). Metals in the active site of native protein phosphatase-1. *Journal of Inorganic Biochemistry, 149*, 1-5.

Hu, Z., Yan, C., Liu, P., Huang, Z., Ma, R., Zhang, C., Wang, R., Zhang, Y., Martinon, F., Miao, D., Deng, H., Wang, J., Chang, J., & Chai, J. (2013). Crystal structure of NLRC4 reveals its autoinhibition mechanism. *Science, 341*(6142), 172-175.

Hurley, T. D., Yang, J., Zhang, L., Goodwin, K. D., Zou, Q., Cortese, M., Dunker, A. K., & DePaoli-Roach, A. A. (2007). Structural basis for regulation of protein phosphatase 1 by Inhibitor-2. *Journal of Biological Chemistry, 282*(39), 28874-28883.

Jiang, F., Waterfield, Nicholas R., Yang, J., Yang, G., & Jin, Q. (2014). A *Pseudomonas aeruginosa* Type VI secretion phospholipase D effector targets both prokaryotic and eukaryotic cells. *Cell Host Microbe, 15*(5), 600-610.

Jiang, R. H. Y., Tripathy, S., Govers, F., & Tyler, B. M. (2008). RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with

more than 700 members. *Proceedings of the National Academy of Sciences, 105*(12), 4874-4879.

Jones, D. A., Bertazzoni, S., Turo, C. J., Syme, R. A., & Hane, J. K. (2018). Bioinformatic prediction of plant-pathogenicity effector proteins of fungi. *Current Opinion in Microbiology, 46*, 43-49.

Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature, 444*, 323.

Jones, J. D. G., Witek, K., Verweij, W., Jupe, F., Cooke, D., Dorling, S., Tomlinson, L., Smoker, M., Perkins, S., & Foster, S. (2014). Elevating crop disease resistance with cloned genes. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1639), 20130087.

Kamoun, S., Furzer, O., Jones, J. D. G., Judelson, H. S., Ali, G. S., Dalio, R. J. D., Roy, S. G., Schena, L., Zambounis, A., Panabières, F., Cahill, D., Ruocco, M., Figueiredo, A., Chen, X.-R., Hulvey, J., Stam, R., Lamour, K., Gijzen, M., Tyler, B. M., Grünwald, N. J., Mukhtar, M. S., Tomé, D. F. A., Tör, M., Van Den Ackerveken, G., McDowell, J., Daayf, F., Fry, W. E., Lindqvist-Kreuze, H., Meijer, H. J. G., Petre, B., Ristaino, J., Yoshida, K., Birch, P. R. J., & Govers, F. (2015). The Top 10 oomycete pathogens in molecular plant pathology. *Molecular Plant Pathology, 16*(4), 413-434.

Kanzaki, H., Yoshida, K., Saitoh, H., Fujisaki, K., Hirabuchi, A., Alaux, L., Fournier, E., Tharreau, D., & Terauchi, R. (2012). Arms race co-evolution of *Magnaporthe oryzae* AVR-Pik and rice Pik genes driven by their physical interactions. *The Plant Journal, 72*(6), 894-907.

Kelker, M. S., Page, R., & Peti, W. (2009). Crystal structures of protein phosphatase-1 bound to nodularin-R and tautomycin: a novel scaffold for structure-based drug design of serine/threonine phosphatase inhibitors. *Journal of Molecular Biology, 385*(1), 11-21.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols, 10*(6), 845-858.

Kerppola, T. K. (2006). Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. *Nature Protocols, 1*(3), 1278-1286.

Kim, M., Ahn, J. W., Jin, U. H., Choi, D., Paek, K. H., & Pai, H. S. (2003). Activation of the programmed cell death pathway by inhibition of proteasome function in plants. *Journal of Biological Chemistry, 278*(21), 19406-19415.

Kim, S. H., Qi, D., Ashfield, T., Helm, M., & Innes, R. W. (2016). Using decoys to expand the recognition specificity of a plant disease resistance protein. *Science, 351*(6274), 684-687.

King, S. R., McLellan, H., Boevink, P. C., Armstrong, M. R., Bukharova, T., Sukarta, O., Win, J., Kamoun, S., Birch, P. R., & Banfield, M. J. (2014). *Phytophthora infestans* RXLR effector PexRD2 interacts with host MAPKKK epsilon to suppress plant immune signaling. *The Plant Cell, 26*(3), 1345-1359.

Krasileva, K. V., Dahlbeck, D., & Staskawicz, B. J. (2010). Activation of an *Arabidopsis* resistance protein is specified by the in planta association of its Leucine-Rich Repeat domain with the cognate oomycete effector. *The Plant Cell, 22*(7), 2444-2458.

Krissinel, E. (2015). Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Research, 43*(W1), W314-319.

Kroj, T., Chanclud, E., Michel-Romiti, C., Grand, X., & Morel, J. B. (2016). Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytologist, 210*(2), 618-626.

Larebeke, N. V., Engler, G., Holsters, M., Den Elsacker, S. V., Zaenen, I., Schilperoort, R. A., & Schell, J. (1974). Large plasmid in *Agrobacterium tumefaciens* essential for crown gall-inducing ability. *Nature, 252*, 169.

Le Roux, C., Huet, G., Jauneau, A., Camborde, L., Trémousaygue, D., Kraut, A., Zhou, B., Levaillant, M., Adachi, H., Yoshioka, H., Raffaele, S., Berthomé, R., Couté, Y., Parker, Jane E., & Deslandes, L. (2015). A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell, 161*(5), 1074-1088.

Lin, Q., Li, J., Smith, R. D., & Walker, J. C. (1998). Molecular cloning and chromosomal mapping of type one serine/threonine protein phosphatases in *Arabidopsis thaliana*. *Plant Molecular Biology, 37*(3), 471-481.

Liu, G., Greenshields, D. L., Sammynaiken, R., Hirji, R. N., Selvaraj, G., & Wei, Y. (2007). Targeted alterations in iron homeostasis underlie plant defense responses. *Journal of Cell Science, 120*(4), 596-605.

Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., & Berkmen, M. (2012). SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories, 11*, 56.

Lorenz, U. (2011). Protein tyrosine phosphatase assays. *Current Protocols in Immunology, CHAPTER*, Unit-11.17.

Lu, S. (2012). Use of the Yeast Two-Hybrid system to identify targets of fungal effectors. In M. D. Bolton & B. P. H. J. Thomma (Eds.), *Plant Fungal Pathogens: Methods and Protocols* (pp. 165-189). Totowa, NJ: Humana Press.

Lukasik, E., & Takken, F. L. W. (2009). STANDing strong, resistance proteins instigators of plant defence. *Current Opinion in Plant Biology, 12*(4), 427-436.

Macho, A. P., Schwessinger, B., Ntoukakis, V., Brutus, A., Segonzac, C., Roy, S., Kadota, Y., Oh, M. H., Sklenar, J., Derbyshire, P., Lozano-Duran, R., Malinovsky, F. G., Monaghan, J., Menke, F. L., Huber, S. C., He, S. Y., & Zipfel, C. (2014). A bacterial tyrosine phosphatase inhibits plant pattern recognition receptor activation. *Science, 343*(6178), 1509-1512.

Maqbool, A., Saitoh, H., Franceschetti, M., Stevenson, C. E., Uemura, A., Kanzaki, H., Kamoun, S., Terauchi, R., & Banfield, M. J. (2015). Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife, 4*.

Marchal, C., Zhang, J., Zhang, P., Fenwick, P., Steuernagel, B., Adamski, N. M., Boyd, L., McIntosh, R., Wulff, B. B. H., Berry, S., Lagudah, E., & Uauy, C. (2018). BED-domain containing immune receptors confer diverse resistance spectra to yellow rust. *bioRxiv*.

McCann, H. C., Nahal, H., Thakur, S., & Guttman, D. S. (2012). Identification of innate immunity elicitors using molecular signatures of natural selection. *Proceedings of the National Academy of Sciences, 109*(11), 4215-4220.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography, 40*(4), 658-674.

McDermott, J. E., Corrigan, A., Peterson, E., Oehmen, C., Niemann, G., Cambronne, E. D., Sharp, D., Adkins, J. N., Samudrala, R., & Heffron, F. (2011). Computational prediction of Type III and IV secreted effectors in gram-negative bacteria. *Infection and Immunity, 79*(1), 23-32.

McLellan, H., Boevink, P. C., Armstrong, M. R., Pritchard, L., Gomez, S., Morales, J., Whisson, S. C., Beynon, J. L., & Birch, P. R. J. (2013). An RxLR effector from *Phytophthora infestans* prevents ee-localisation of two plant NAC transcription factors from the endoplasmic reticulum to the nucleus. *PLoS Pathogens, 9*(10), e1003670.

Murphy, F., He, Q., Armstrong, M., Giuliani, L. M., Boevink, P. C., Zhang, W., Tian, Z., Birch, P. R. J., & Gilroy, E. M. (2018). The potato MAP3K StVIK is required for the *Phytophthora infestans* RXLR effector Pi17316 to promote disease. *Plant Physiology, 177*(1), 398-410.

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D, 67*(4), 355-367.

Newman, M.-A., Sundelin, T., Nielsen, J., & Erbs, G. (2013). MAMP (microbe-associated molecular pattern) triggered immunity in plants. *Frontiers in Plant Science, 4*(139).

Nyarko, A., Singarapu, K. K., Figueroa, M., Manning, V. A., Pandelova, I., Wolpert, T. J., Ciuffetti, L. M., & Barbar, E. (2014). Solution NMR structures of *Pyrenophora*

tritici-repentis ToxB and its inactive homolog reveal potential determinants of toxin activity. *Journal of Biological Chemistry, 289*(37), 25946-25956.

Okuyama, Y., Kanzaki, H., Abe, A., Yoshida, K., Tamiru, M., Saitoh, H., Fujibe, T., Matsumura, H., Shenton, M., Galam, D. C., Undan, J., Ito, A., Sone, T., & Terauchi, R. (2011). A multifaceted genomics approach allows the isolation of the rice Pia-blast resistance gene consisting of two adjacent NBS-LRR protein genes. *The Plant Journal, 66*(3), 467-479.

Ortiz, D., de Guillen, K., Cesari, S., Chalvon, V., Gracy, J., Padilla, A., & Kroj, T. (2017). Recognition of the *Magnaporthe oryzae* effector AVR-Pia by the decoy domain of the rice NLR immune receptor RGA5. *The Plant Cell, 29*(1), 156-168.

Ose, T., Oikawa, A., Nakamura, Y., Maenaka, K., Higuchi, Y., Satoh, Y., Fujiwara, S., Demura, M., Sone, T., & Kamiya, M. (2015). Solution structure of an avirulence protein, AVR-Pia, from *Magnaporthe oryzae*. *Journal of Biomolecular NMR, 63*(2), 229-235.

Padmanabhan, M., Cournoyer, P., & Dinesh-Kumar, S. P. (2009). The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. *Cellular Microbiology, 11*(2), 191-198.

Pais, S. M., Gonzalez, M. A., Tellez-Inon, M. T., & Capiati, D. A. (2009). Characterization of potato (*Solanum tuberosum*) and tomato (*Solanum lycopersicum*) protein phosphatases type 2A catalytic subunits and their involvement in stress responses. *Planta, 230*(1), 13-25.

Park, C. H., Shirsekar, G., Bellizzi, M., Chen, S., Songkumarn, P., Xie, X., Shi, X., Ning, Y., Zhou, B., Suttiviriya, P., Wang, M., Umemura, K., & Wang, G.-L. (2016). The E3 ligase APIP10 connects the effector AvrPiz-t to the NLR receptor Piz-t in rice. *PLoS Pathogens, 12*(3), e1005529.

Patron, N. J., Orzaez, D., Marillonnet, S., Warzecha, H., Matthewman, C., Youles, M., Raitskin, O., Leveau, A., Farre, G., Rogers, C., Smith, A., Hibberd, J., Webb, A. A., Locke, J., Schornack, S., Ajioka, J., Baulcombe, D. C., Zipfel, C., Kamoun, S., Jones, J. D., Kuhn, H., Robatzek, S., Van Esse, H. P., Sanders, D., Oldroyd, G., Martin, C., Field, R., O'Connor, S., Fox, S., Wulff, B., Miller, B., Breakspear, A., Radhakrishnan, G., Delaux, P. M., Loque, D., Granell, A., Tissier, A., Shih, P., Brutnell, T. P., Quick, W. P., Rischer, H., Fraser, P. D., Aharoni, A., Raines, C., South, P. F., Ane, J. M., Hamberger, B. R., Langdale, J., Stougaard, J., Bouwmeester, H., Udvardi, M., Murray, J. A., Ntoukakis, V., Schafer, P., Denby, K., Edwards, K. J., Osbourn, A., & Haseloff, J. (2015). Standards for plant synthetic biology: a common syntax for exchange of DNA parts. *New Phytologist, 208*(1), 13-19.

Paulus, J. K., & van der Hoorn, R. A. L. (2018). Tricked or trapped—Two decoy mechanisms in host–pathogen interactions. *PLoS Pathogens, 14*(2), e1006761.

Peti, W., Nairn, A. C., & Page, R. (2013). Structural basis for protein phosphatase 1 regulation and specificity. *The FEBS Journal, 280*(2), 596-611.

Petre, B., & Kamoun, S. (2014). How do filamentous pathogens deliver effector proteins into plant cells? *PLOS Biology, 12*(2), e1001801.

Petre, B., Saunders, D. G., Sklenar, J., Lorrain, C., Win, J., Duplessis, S., & Kamoun, S. (2015). Candidate effector proteins of the rust pathogen *Melampsora larici-populina* target diverse plant cell compartments. *Molecular Plant-Microbe Interactions, 28*(6), 689-700.

Pieterse, C. M. J., Does, D. V. d., Zamioudis, C., Leon-Reyes, A., & Wees, S. C. M. V. (2012). Hormonal modulation of plant immunity. *Annual Review of Cell and Developmental Biology, 28*(1), 489-521.

Popa, C., Li, L., Gil, S., Tatjer, L., Hashii, K., Tabuchi, M., Coll, N. S., Ariño, J., & Valls, M. (2016). The effector AWR5 from the plant pathogen *Ralstonia solanacearum* is an inhibitor of the TOR signalling pathway. *Scientific Reports, 6*, 27058.

Pritchard, L., & Birch, P. R. J. (2014). The zigzag model of plant–microbe interactions: is it time to move on? *Molecular Plant Pathology, 15*(9), 865-870.

Ragusa, M. J., Dancheck, B., Critton, D. A., Nairn, A. C., Page, R., & Peti, W. (2010). Spinophilin directs protein phosphatase 1 specificity by blocking substrate binding sites. *Nature Structural & Molecular Biology, 17*(4), 459-464.

Ryder, L. S., & Talbot, N. J. (2015). Regulation of appressorium development in pathogenic fungi. *Current Opinion in Plant Biology, 26*, 8-13.

Sarma, G. N., Manning, V. A., Ciuffetti, L. M., & Karplus, P. A. (2005). Structure of Ptr ToxA: An RGD-containing host-selective toxin from *Pyrenophora tritici-repentis*. *The Plant Cell, 17*(11), 3190-3202.

Sarris, P. F., Cevik, V., Dagdas, G., Jones, J. D. G., & Krasileva, K. V. (2016). Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biology, 14*(1), 8.

Schornack, S., van Damme, M., Bozkurt, T. O., Cano, L. M., Smoker, M., Thines, M., Gaulin, E., Kamoun, S., & Huitema, E. (2010). Ancient class of translocated oomycete effectors targets the host nucleus. *Proceedings of the National Academy of Sciences, 107*(40), 17421-17426.

Schweighofer, A., Kazanaviciute, V., Scheikl, E., Teige, M., Doczi, R., Hirt, H., Schwanninger, M., Kant, M., Schuurink, R., Mauch, F., Buchala, A., Cardinale, F., & Meskiene, I. (2007). The PP2C-type phosphatase AP2C1, which negatively regulates MPK4 and MPK6, modulates innate immunity, jasmonic acid, and ethylene levels in *Arabidopsis*. *The Plant Cell, 19*(7), 2213-2224.

Segretin, M. E., Pais, M., Franceschetti, M., Chaparro-Garcia, A., Bos, J. I. B., Banfield, M. J., & Kamoun, S. (2014). Single amino acid mutations in the potato immune receptor R3a expand response to *Phytophthora* effectors. *Molecular Plant-Microbe Interactions, 27*(7), 624-637.

Seybold, H., Trempel, F., Ranf, S., Scheel, D., Romeis, T., & Lee, J. (2014). Ca^{2+} signalling in plant immune response: from pattern recognition receptors to Ca^{2+} decoding mechanisms. *New Phytologist, 204*(4), 782-790.

Shao, Z.-Q., Xue, J.-Y., Wu, P., Zhang, Y.-M., Wu, Y., Hang, Y.-Y., Wang, B., & Chen, J.-Q. (2016). Large-scale analyses of angiosperm Nucleotide-Binding Site-Leucine-Rich Repeat genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant Physiology, 170*(4), 2095-2109.

Sonah, H., Deshmukh, R. K., & Bélanger, R. R. (2016). Computational prediction of effector proteins in fungi: Opportunities and challenges. *Frontiers in Plant Science, 7*, 126.

Sperschneider, J., Catanzariti, A. M., DeBoer, K., Petre, B., Gardiner, D. M., Singh, K. B., Dodds, P. N., & Taylor, J. M. (2017). LOCALIZER: subcellular localization prediction of both plant and effector proteins in the plant cell. *Scientific Reports, 7*, 44598.

Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., & Taylor, J. M. (2018). Improved prediction of fungal effector proteins from secretomes with EffectorP 2.0. *Molecular Plant Pathology, 19*(9), 2094-2110.

Sperschneider, J., Dodds, P. N., Singh, K. B., & Taylor, J. M. (2018). ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. *New Phytologist, 217*(4), 1764-1778.

Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J. M., & Taylor, J. M. (2016). EffectorP: predicting fungal effector proteins from secretomes using machine learning. *New Phytologist, 210*(2), 743-761.

Spoel, S. H., & Dong, X. (2012). How do plants achieve immunity? Defence without specialized immune cells. *Nature Reviews: Immunology, 12*(2), 89-100.

Sreerama, N., Venyaminov, S. Y., & Woody, R. W. (1999). Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. *Protein Science : A Publication of the Protein Society, 8*(2), 370-380.

Sreerama, N., & Woody, R. W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical Biochemistry, 287*(2), 252-260.

Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification, 41*(1), 207-234.

Sukarta, O. C. A., Slootweg, E. J., & Goverse, A. (2016). Structure-informed insights for NLR functioning in plant immunity. *Seminars in Cell & Developmental Biology, 56*, 134-149.

Sun, Y., Li, L., Macho, A. P., Han, Z., Hu, Z., Zipfel, C., Zhou, J.-M., & Chai, J. (2013). Structural basis for flg22-induced activation of the *Arabidopsis* FLS2-BAK1 immune complex. *Science, 342*(6158), 624-628.

Swiderski, M. R., Birker, D., & Jones, J. D. G. (2009). The TIR domain of TIR-NB-LRR resistance proteins is a signaling domain involved in cell death induction. *Molecular Plant-Microbe Interactions, 22*(2), 157-165.

Takemiya, A., Kinoshita, T., Asanuma, M., & Shimazaki, K. (2006). Protein phosphatase 1 positively regulates stomatal opening in response to blue light in *Vicia faba*. *Proceedings of the National Academy of Sciences, 103*(36), 13549-13554.

Team, R. D. C. (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Terrak, M., Kerff, F., Langsetmo, K., Tao, T., & Dominguez, R. (2004). Structural basis of protein phosphatase 1 regulation. *Nature, 429*(6993), 780-784.

Thomma, B. P., Nurnberger, T., & Joosten, M. H. (2011). Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *The Plant Cell, 23*(1), 4-15.

Thordal-Christensen, H., Birch, P. R. J., Spanu, P. D., & Panstruga, R. (2018). Why did filamentous plant pathogens evolve the potential to secrete hundreds of effectors to enable disease? *Molecular Plant Pathology, 19*(4), 781-785.

Thulasi Devendrakumar, K., Li, X., & Zhang, Y. (2018). MAP kinase signalling: interplays between plant PAMP- and effector-triggered immunity. *Cellular and Molecular Life Sciences, 75*(16), 2981-2989.

Trdá, L., Boutrot, F., Claverie, J., Brulé, D., Dorey, S., & Poinssot, B. (2015). Perception of pathogenic or beneficial bacteria and their evasion of host immunity: pattern recognition receptors in the frontline. *Frontiers in Plant Science, 6*, 219.

van der Biezen, E. A., & Jones, J. D. G. (1998). Plant disease-resistance proteins and the gene-for-gene concept. *Trends in Biochemical Sciences, 23*(12), 454-456.

van der Hoorn, R. A. L., & Kamoun, S. (2008). From Guard to Decoy: A new model for perception of plant pathogen effectors. *The Plant Cell, 20*(8), 2009-2017.

van Schie, C. C., & Takken, F. L. (2014). Susceptibility genes 101: how to be a good host. *Annual Review of Phytopathology, 52*, 551-581.

Verbinnen, I., Ferreira, M., & Bollen, M. (2017). Biogenesis and activity regulation of protein phosphatase 1. *Biochemical Society Transactions, 45*(1), 89-99.

Vetukuri, R. R., Whisson, S. C., & Grenville-Briggs, L. J. (2017). *Phytophthora infestans* effector Pi14054 is a novel candidate suppressor of host silencing mechanisms. *European Journal of Plant Pathology, 149*(3), 771-777.

Walden, M., Edwards, J. M., Dziewulska, A. M., Bergmann, R., Saalbach, G., Kan, S.-Y., Miller, O. K., Weckener, M., Jackson, R. J., Shirran, S. L., Botting, C. H., Florence, G. J., Rohde, M., Banfield, M. J., & Schwarz-Linek, U. (2015). An internal thioester in a pathogen surface protein mediates covalent host binding. *eLife, 4*, e06638.

Wang, C.-I. A., Gunčar, G., Forwood, J. K., Teh, T., Catanzariti, A.-M., Lawrence, G. J., Loughlin, F. E., Mackay, J. P., Schirra, H. J., Anderson, P. A., Ellis, J. G., Dodds, P. N., & Kobe, B. (2007). Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *The Plant Cell, 19*(9), 2898-2912.

Wang, S., Boevink, P. C., Welsh, L., Zhang, R., Whisson, S. C., & Birch, P. R. J. (2017). Delivery of cytoplasmic and apoplastic effectors from *Phytophthora infestans* haustoria by distinct secretion pathways. *New Phytologist, 216*(1), 205-215.

Wang, X., Boevink, P., McLellan, H., Armstrong, M., Bukharova, T., Qin, Z., & Birch, Paul R. (2015). A host KH RNA-binding protein is a susceptibility factor targeted by an RXLR effector to promote late blight disease. *Molecular Plant, 8*(9), 1385-1395.

Wawra, S., Trusch, F., Matena, A., Apostolakis, K., Linne, U., Zhukov, I., Stanek, J., Koźmiński, W., Davidson, I., Secombes, C. J., Bayer, P., & van West, P. (2017). The RxLR motif of the host targeting effector AVR3a of *Phytophthora infestans* is cleaved before secretion. *The Plant Cell*.

Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS One, 6*(2), e16765.

Weßling, R., Epple, P., Altmann, S., He, Y., Yang, L., Henz, S. R., McDonald, N., Wiley, K., Bader, K. C., Gläßer, C., Mukhtar, M. S., Haigis, S., Ghamsari, L., Stephens, A. E., Ecker, J. R., Vidal, M., Jones, J. D. G., Mayer, K. F. X., Ver Loren van Themaat, E., Weigel, D., Schulze-Lefert, P., Dangl, J. L., Panstruga, R., & Braun, P. (2014). Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe, 16*(3), 364-375.

Whisson, S. C., Avrova, A. O., Boevink, P. C., Armstrong, M. R., Seman, Z. A., Hein, I., & Birch, P. R. J. (2011). Exploiting knowledge of pathogen effectors to enhance late blight resistance in potato. *Potato Research, 54*(4), 325-340.

Whisson, S. C., Boevink, P. C., Moleleki, L., Avrova, A. O., Morales, J. G., Gilroy, E. M., Armstrong, M. R., Grouffaud, S., van West, P., Chapman, S., Hein, I., Toth, I. K., Pritchard, L., & Birch, P. R. J. (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature, 450*, 115.

Whitmore, L., & Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research, 32*(Web Server issue), W668-673.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Widjaja, I., Lassowskat, I., Bethke, G., Eschen-Lippold, L., Long, H. H., Naumann, K., Dangl, J. L., Scheel, D., & Lee, J. (2010). A protein phosphatase 2C, responsive to the bacterial effector AvrRpm1 but not to the AvrB effector, regulates defense responses in *Arabidopsis*. *The Plant Journal, 61*(2), 249-258.

Williams, S. J., Sohn, K. H., Wan, L., Bernoux, M., Sarris, P. F., Segonzac, C., Ve, T., Ma, Y., Saucet, S. B., Ericsson, D. J., Casey, L. W., Lonhienne, T., Winzor, D. J., Zhang, X., Coerdt, A., Parker, J. E., Dodds, P. N., Kobe, B., & Jones, J. D. G. (2014). Structural basis for assembly and function of a heterodimeric plant immune receptor. *Science, 344*(6181), 299-303.

Wilson, R. A., & Talbot, N. J. (2009). Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nature Reviews Microbiology, 7*(3), 185-195.

Win, J., Kamoun, S., & Jones, A. M. E. (2011). Purification of effector–target protein complexes via transient expression in *Nicotiana benthamiana*. In J. M. McDowell (Ed.), *Plant Immunity: Methods and Protocols* (pp. 181-194). Totowa, NJ: Humana Press.

Win, J., Krasileva, K. V., Kamoun, S., Shirasu, K., Staskawicz, B. J., & Banfield, M. J. (2012). Sequence divergent RXLR effectors share a structural fold conserved across plant pathogenic oomycete species. *PLoS Pathogens, 8*(1), e1002400.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., & Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D, 67*(4), 235-242.

Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography, 43*(1), 186-190.

Wróblewski, T., Spiridon, L., Martin, E. C., Petrescu, A.-J., Cavanaugh, K., Truco, M. J., Xu, H., Gozdowski, D., Pawłowski, K., Michelmore, R. W., & Takken, F. L. W. (2018). Genome-wide functional analyses of plant coiled–coil NLR-type pathogen receptors reveal essential roles of their N-terminal domain in oligomerization, networking, and immunity. *PLOS Biology, 16*(12), e2005821.

Wu, C.-H., Abd-El-Haliem, A., Bozkurt, T. O., Belhaj, K., Terauchi, R., Vossen, J. H., & Kamoun, S. (2017). NLR network mediates immunity to diverse plant pathogens. *Proceedings of the National Academy of Sciences, 114*(30), 8113-8118.

Wu, C.-H., Derevnina, L., & Kamoun, S. (2018). Receptor networks underpin plant immunity. *Science, 360*(6395), 1300-1301.

Wu, C.-H., Krasileva, K., Banfield, M., Terauchi, R., & Kamoun, S. (2015). The "sensor domains" of plant NLR proteins: more than decoys? *Frontiers in Plant Science, 6*(134).

Xiong, Q., Ye, W., Choi, D., Wong, J., Qiao, Y., Tao, K., Wang, Y., & Ma, W. (2014). *Phytophthora* Suppressor of RNA Silencing 2 is a conserved RxLR effector that promotes infection in soybean and *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions, 27*(12), 1379-1389.

Xue, M., Yang, J., Li, Z., Hu, S., Yao, N., Dean, R. A., Zhao, W., Shen, M., Zhang, H., Li, C., Liu, L., Cao, L., Xu, X., Xing, Y., Hsiang, T., Zhang, Z., Xu, J.-R., & Peng, Y.-L. (2012). Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*. *PLOS Genetics, 8*(8), e1002869.

Yan, X., & Talbot, N. J. (2016). Investigating the cell biology of plant infection by the rice blast fungus *Magnaporthe oryzae*. *Current Opinion in Microbiology, 34*, 147-153.

Yang, Z. R., Thomson, R., McNeil, P., & Esnouf, R. M. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics, 21*(16), 3369-3376.

Yoshida, K., Saitoh, H., Fujisawa, S., Kanzaki, H., Matsumura, H., Yoshida, K., Tosa, Y., Chuma, I., Takano, Y., Win, J., Kamoun, S., & Terauchi, R. (2009). Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *The Plant Cell, 21*(5), 1573-1591.

Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., Thines, M., Weigel, D., & Burbano, H. A. (2013). The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLife, 2*, e00731.

Yuan, B., Zhai, C., Wang, W., Zeng, X., Xu, X., Hu, H., Lin, F., Wang, L., & Pan, Q. (2011). The Pik-p resistance to *Magnaporthe oryzae* in rice is mediated by a pair of closely linked CC-NBS-LRR genes. *Theoretical and Applied Genetics, 122*(5), 1017-1028.

Zhai, C., Lin, F., Dong, Z., He, X., Yuan, B., Zeng, X., Wang, L., & Pan, Q. (2011). The isolation and characterization of Pik, a rice blast resistance gene which emerged after rice domestication. *New Phytologist, 189*(1), 321-334.

Zhang, M., & Coaker, G. (2017). Harnessing effector-triggered immunity for durable disease resistance. *Phytopathology, 107*(8), 912-919.

Zhang, X., Bernoux, M., Bentham, A. R., Newman, T. E., Ve, T., Casey, L. W., Raaymakers, T. M., Hu, J., Croll, T. I., Schreiber, K. J., Staskawicz, B. J., Anderson, P. A., Sohn, K. H., Williams, S. J., Dodds, P. N., & Kobe, B. (2017). Multiple functional self-association interfaces in plant TIR domains. *Proceedings of the National Academy of Sciences, 114*(10), E2046-E2052.

Zhang, Y., Dorey, S., Swiderski, M., & Jones, J. D. G. (2004). Expression of RPS4 in tobacco induces an AvrRps4-independent HR that requires EDS1, SGT1 and HSP90. *The Plant Journal, 40*(2), 213-224.

Zhang, Y., Zhao, J., Li, Y., Yuan, Z., He, H., Yang, H., Qu, H., Ma, C., & Qu, S. (2016). Transcriptome analysis highlights defense and signaling pathways mediated by rice pi21 gene with partial resistance to *Magnaporthe oryzae*. *Frontiers in Plant Science, 7*(1834).

Zhang, Z., Zhao, S., Bai, G., & Lee, E. Y. (1994). Characterization of deletion mutants of the catalytic subunit of protein phosphatase-1. *Journal of Biological Chemistry, 269*(19), 13766-13770.

Zhang, Z. M., Zhang, X., Zhou, Z. R., Hu, H. Y., Liu, M., Zhou, B., & Zhou, J. (2013). Solution structure of the *Magnaporthe oryzae* avirulence protein AvrPiz-t. *Journal of Biomolecular NMR, 55*(2), 219-223.

Zheng, X., McLellan, H., Fraiture, M., Liu, X., Boevink, P. C., Gilroy, E. M., Chen, Y., Kandel, K., Sessa, G., Birch, P. R. J., & Brunner, F. (2014). Functionally redundant RXLR effectors from *Phytophthora infestans* act at different steps to suppress early flg22-triggered immunity. *PLoS Pathogens, 10*(4), e1004057.

Zipfel, C. (2008). Pattern-recognition receptors in plant innate immunity. *Current Opinion in Immunology, 20*(1), 10-16.

Zipfel, C. (2014). Plant pattern-recognition receptors. *Trends in Immunology, 35*(7), 345-351.

Zipfel, C., & Rathjen, J. P. (2008). Plant immunity: AvrPto targets the frontline. *Current Biology, 18*(5), R218-R220.