USING A RICE BLAST EFFECTOR TARGET TO ENGINEER NLR IMMUNE RECEPTORS WITH NOVEL RECOGNITION SPECIFICITIES

Josephine Harriet Robertshaw Maidment

Thesis submitted to the University of East Anglia

for the degree of Doctor of Philosophy

February 2020

This work is dedicated to Iwan Griffiths.

Abstract

Arable farmers across the world are engaged in a constant battle with plant pathogens. The rice blast fungus *Magnaporthe oryzae* is a threat to global rice production. During infection, *M. oryzae* secretes effector proteins which interact with host targets to promote pathogen virulence. Specific effectors can also be recognised by intracellular nucleotide-binding, leucine-rich repeat (NLR) immune receptor proteins, activating host defences.

Recognition of the rice blast effector AVR-Pik is mediated by the paired rice NLR proteins Pik-1/Pik-2. An integrated heavy metal-associated (HMA) domain was previously identified in Pik-1, to which certain variants of AVR-Pik directly bind to trigger resistance. Multiple Pik-1 alleles have evolved in rice and exhibit differential responses to different AVR-Pik effector variants. AVR-PikC and AVR-PikF do not interact with any known Pik-1 alleles and thus evade plant defences.

In this work, I used biochemical and structural techniques to characterise the interaction between AVR-Pik and the HMA domain of its putative virulence target, OsHIPP19. I demonstrated that a second rice blast effector, AVR-Pia, also interacts with OsHIPP19-HMA. By modifying the HMA domain of Pik-1 to resemble OsHIPP19, I engineered two Pik-1 variants (Pikp-1$^{SNK\text{-}EKE}$ and Pikp-1^{HIPP19_mbl7}) which trigger HR-like cell death in response to AVR-PikC/AVR-PikF in a model system. Pikp-1$^{SNK\text{-}EKE}$ was engineered through targeted mutagenesis of Pik-1 guided by the crystal structure of AVR-PikF in complex with OsHIPP19-HMA. Exchanging the HMA domain of Pikp-1 for that of OsHIPP19 resulted in an autoactive chimera, however subsequent modifications produced a second Pik-1 variant, Pikp-1^{HIPP19_mbl7}, which was not autoactive and could respond to AVR-PikC/AVR-PikF in a model plant.

The engineered Pik-1 NLR proteins are being incorporated into transgenic rice and barley to investigate whether they confer resistance to *Magnaporthe oryzae*. Rationally engineered immune receptors offer opportunities to control emerging and rapidly evolving pathogens, contributing to global food security.

Table of Contents

Table of figures

Table of tables

Acknowledgements

I would like to begin by thanking my supervisor, Professor Mark Banfield, for giving me the opportunity to conduct this research and for guiding my work over the last four years. His calm manner and support have been invaluable. My thanks also go to the other members of my supervisory committee, Sophien Kamoun, Marina Franceschetti and Abbas Maqbool, for the insightful comments and productive discussions which have shaped this thesis.

I am very grateful to all members of the Banfield group, past and present, who have been a source of support, advice and experience over the last four years. I'd particularly like to thank Marina Franceschetti for not only being a fantastic mentor, but also for the time and effort she spent establishing many of the tools and methods that provided a foundation for this project. I was fortunate to supervise three excellent students during my PhD, and would like to thank Thomas Gate, Vera Sham and Lauren Grubb for their scientific contributions and for being so lovely to work with.

I'd like to express my gratitude to Dave Lawson, Clare Stevenson and Julia Mundy (JIC protein crystallography and biophysical analysis platforms) and to Gerhard Saalbach and Carlo de Oliveria Martins (JIC proteomics platform) for their contributions to this work. I have also benefited from the considerable barley expertise of Matthew Moscou's group in the Sainsbury Laboratory, and am very grateful to Phon Green and Helen Brabham for all their help and advice. I would also like to thank the support staff at JIC, especially the media kitchen and horticultural services teams, for all the work they do to facilitate the research that goes on at the John Innes Centre.

This project began with experiments carried out by members of Ryohei Terauchi's group at Iwate Biotechnology Research Center in Japan, and I am very grateful for their scientific input and also for their hospitality during my visit to Iwate. In particular, my thanks go to Motoki Shimizu for his ongoing work with the transgenic rice plants.

As part of the BBSRC doctoral training programme, I was fortunate enough to spend three months working with the capacity building team at the BecA-ILRI Hub in Nairobi, Kenya. The experiences I gained through this placement have really shaped my approach to research, and I am so grateful to everybody I worked with during that time for sharing their knowledge and enthusiasm.

I owe a great many thanks to Hannah, Helen, Beth, Freya, Jenny and Sarah for providing encouragement, laughter, perspective and friendship over the last four years. I'd also like to thank all the women at Lakenham Hewett rugby club for picking me up (literally and figuratively) when I've been down; your support means more than I can say.

I'd like to thank my brilliant parents for encouraging me to follow my interests, and for always supporting and valuing my education. My sisters, Alex and Evie, continue to inspire me with their determination and resilience, and I am so grateful to them for their faith in me during this PhD. To my friend-who-is-basically-a-sister Katy, thank you for always being just a phone call away.

Abbreviations

This thesis uses standard abbreviations for nucleic acids (one letter code, ATGCU) and amino acids (one and three letter codes, e.g. R or Arg).

Standard SI units are used, unless stated otherwise.

A_{280}	absorbance at 280nm
ADP	adenosine diphosphate
AIM	auto-induction medium
ATP	adenosine triphosphate
AU	absorbance units
AVR	avirulence protein
BIC	biotrophic interfacial complex
bp	base pair
CC	coiled coil
CD	circular dichroism
cDNA	complementary DNA
CDS	coding sequence
CRN	crinkling and necrosis
cv.	cultivar
Da (kDa)	Dalton (kiloDalton)
DAMP	damage associated molecular pattern
dATP	deoxyadenosine triphosphate
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dpi	days post-infiltration
DTT	dithiothreitol

EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
EIHM	extrainvasive hyphal membrane
EPPO	European and Mediterranean Plant Protection Organization
EV	empty vector
FC	flow cell
FOG	Ficoll and Orange G
HA	haemagglutinin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	HellFire (His:Flag)
HIPP	heavy metal associated isoprenylated plant protein
HMA	heavy metal associated
HPP	heavy metal associated plant protein
HR	hypersensitive response
HRP	horseradish peroxidase
ID	integrated domain
IMAC	immobilised metal affinity chromatography
IP	immunoprecipitation
IPTG	isopropyl β-D-1-thiogalactopyranoside
IRRI	International Rice Research Institute
ITC	isothermal titration calorimetry
JCSG	Joint Centre for Structural Genomics
KA	Kanto (rice cultivar)
LB	lysogeny broth (Lucia-Bertani)
LDS	lithium dodecyl sulphate
LRR	leucine rich repeat
LysM	lysine motif
MAMP	microbe associated molecular pattern
MAS	marker assisted selection
MAX	Magnaporthe AVRs and ToxB-like

MBP	maltose binding protein
MW	molecular weight
NB-ARC	nucleotide binding Apaf-1, R protein, CED4 shared
NLR	nucleotide-binding leucine rich repeat
NMR	nuclear magnetic resonance
NTA	nitrilotriacetic acid
OD_{600}	optical density measured at a wavelength of 600nm
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PISA	protein interfaces, surfaces and assemblies
PR	pathogenesis-related
PRR	pattern recognition receptor
PTI	pattern triggered immunity
PVDF	polyvinylidene difluoride
R	(disease) resistance
RenSeq	resistance gene enrichment sequencing
RLK	receptor-like kinase
RLP	receptor-like protein
RMSD	root-mean-square deviation of atomic positions
RNA	ribonucleic acid
RU	response units
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SOC	super optimal broth with catabolite repression
SPR	surface plasmon resonance
STAND	signal transduction ATPase with numerous domains
SUMO	small ubiquitin-like modifier
TAE	tris-acetate-EDTA
TBS	tris-buffered saline
TCEP	tris(2-carboxyethyl)phosphine

Tfb	transformation buffer
TIR	TOLL/interleukin-1 receptor
TLS	translation-libration-screw
T_m	melting temperature
TTSS	type III secretion system
UTR	untranslated region
UV	ultraviolet
v/v	volume/volume
w/v	weight/volume
WT	wild type

1

General introduction

1

General introduction

1.1 The rice blast pathogen threatens global rice production

1.1.1 Rice is a vital staple food crop

Rice (*Oryza sativa*) is a staple food for more than half of the world's population. In 2011, rice accounted for 19% of total calorie intake worldwide and over 50% of total calorie intake in Bangladesh, Cambodia, Laos, and Vietnam (IRRI, 2018). Asia produces and consumes 90% of world's rice. While some high-to-middle income Asian countries have seen consumption decline slightly in recent years as increased wealth has promoted diversification in diet, rice consumption per capita continues to increase in many other Asian countries and in the region overall (Seck et al., 2012, IRRI et al., 2010). Over recent decades, Africa has also increased its reliance on rice for calorie intake (Elert, 2014), largely due to shifting consumer preferences linked to urbanisation (Seck et al., 2012). The continued consumption of rice as a staple food, combined with projected population growth (particularly in Asia and Africa), is likely to increase demand for rice into the future.

1.1.2 The rice blast fungus *Magnaporthe oryzae* threatens rice production

Global rice production is limited by diverse pathogens and pests, which can reduce both overall yield and the nutritional quality of the crop (Wang et al., 2014, Pennisi, 2010). One of the major threats to rice production is the filamentous ascomycete fungus *Magnaporthe oryzae*, which causes rice blast disease. Typical symptoms include the formation of diamond-shaped lesions with a white-grey centre surrounded by a dark

Figure 1.1 Symptoms of rice blast disease

Typical symptoms of rice blast disease. Photographs are from the International Rice Research Institute. ***A.*** *Neck blast causing collapse of the panicle.* ***B.*** *Brown necrotic lesions symptomatic of collar blast.* ***C.*** *and* ***D.*** *Pale grey diamond-shaped lesions on infected leaves.*

brown border, on infected leaves, and black to grey-brown lesions on the neck or nodes (figure 1.1). Both neck and node infection can lead to girdling or collapse of the stem.

Rice blast disease has been reported in 88 countries spanning all major rice-growing regions around the world (EPPO, 2019). Yield losses are variable (Pennisi, 2010) and typically depend on a range of factors including the developmental stage of the plant, the part of the plant infected, environmental conditions and the production ecosystem (Mew et al., 2018). For example, infection during the seeding or early tillering stages can completely kill the plant, while infection at the maturation stage affects grain filling, lowering yield. In addition to reducing yield, blast infection can also reduce the quality of any grain harvested, as the disease often results in chalkier grains with lower market value (Mew et al., 2018). The rice production ecosystem affects the likelihood of a blast epidemic; rice grown in tropical lowlands is least prone to severe blast infection, while subtropical and temperate lowland production systems are more susceptible to devastating epidemics (Mew et al., 2018).

It has been estimated that between 1975 and 1990, rice blast destroyed sufficient rice to feed around 60 million people per year (Wang et al., 2014). Based on world rice harvest statistics for 2009/2010 and yield losses of 10-35%, Fisher et al. estimated that rice blast disease annually destroys crops that would otherwise be capable of feeding 212-742 million people for a year (Fisher et al., 2012). A recent study on the global burden of crop pests and pathogens provides a reduced estimate of the yield loss caused by rice blast disease, at 4.3 % of rice production worldwide (Savary et al., 2019). While estimates of the yield losses caused by rice blast disease vary, it is evident that this fungus poses a significant threat to global rice production.

1.1.3 Infection biology of *Magnaporthe oryzae*

The infection cycle of *M. oryzae* begins with the attachment of a teardrop-shaped, 3-cell conidium (asexual spore produced by ascomycete fungi) to the rice leaf cuticle (Wilson and Talbot, 2009). An adhesive, released from an apical compartment in the spore tip, enables the conidium to stick to the cuticle. A single polarised germ tube emerges from the conidium and grows across the surface of the leaf, before differentiation into a dome-shaped appressorium is induced by contact with hydrophobic cutin and lipid monomers.

Differentiated appressorium cells possess chitin-rich cell walls and a layer of melanin between the cell wall and the cell membrane. As the appressorium matures, high concentrations of compatible solutes such as glycerol accumulate, and the resulting influx of water enables the generation of turgor pressures of up to 8MPa (de Jong et al., 1997). The melanin layer has a key role in this process, as it prevents the pressure-driven exit of solutes from the cell and thereby facilitates turgor generation. Analysis of mutants lacking this melanin layer found them to be non-pathogenic (Chumley and Valent, 1990). The turgor pressure translates into mechanical force, which drives the penetration peg through the leaf cuticle.

Invasive hyphae extend from the initial infection site to form a branched hyphal network. The hyphae are encapsulated by the extra-invasive hyphal membrane, which is formed by invagination of the plant cell membrane and remains continuous with it. During early infection, *M. oryzae* forms bulging, branched cells which are well adapted for the acquisition of nutrients from the host plant. The fungus can colonise adjacent cells via plasmodesmata in a process that requires hyphal constriction (Sakulkoo et al., 2018). As a hemibiotrophic fungus, *M. oryzae* may exist biotrophically for a prolonged period. Eventually, rapid branching of hyphae causes disease lesions, and conidiophores (specialised fungal hyphae capable of producing conidia) develop. Sporulation occurs under conditions of high humidity, and conidia are transferred to neighbouring plants by the wind or via splash inoculation (Wilson and Talbot, 2009).

1.1.4 Current strategies for control of rice blast disease

At present, methods for control of rice blast disease can be divided into cultural control, chemical control, and resistant varieties. Of these, resistant varieties are widely considered the most effective and sustainable method of controlling the disease, though an integrated approach to rice blast management is required to successfully combat the pathogen (van Esse et al., 2019).

1.1.4.1 Cultural control

Good cultural control practices can limit the incidence and impact of rice blast disease but are unlikely to prevent it entirely. Methods of cultural control include management

of potential sources of inoculum, reducing susceptibility through appropriate application of fertilisers, and limiting plant density to reduce spread of the disease.

Infected rice crop residue, such as rice straw, can harbour the fungus and be a key source of inoculum, causing disease outbreaks in subsequent plantings (Mew et al., 2018). Rice straw is often stored around a farm for use as animal feed, compost or thatching. Appropriate disposal of infected rice straw can eliminate a key source of inoculum and reduce the incidence of blast disease.

Excessive use of nitrogen fertiliser increases susceptibility to rice blast fungus. The genetic and mechanistic basis by which this occurs is not well understood, but it has been shown that high nitrogen levels can partially break down resistance triggered by known resistance genes (Ballini et al., 2013). By contrast, application of silicon fertiliser can reduce the incidence of rice blast disease (Hayasaka et al., 2005). Increased levels of silica in the leaf epidermis is reported to provide an increased physical barrier to fungal penetration (Hayasaka et al., 2008). Additionally, accumulation of silicon has been shown to induce physiological changes in rice plants which may contribute to reduced susceptibility to *M. oryzae* (Sun et al., 2010). Appropriate application of fertilisers can therefore reduce the risk of rice blast disease.

1.1.4.2 Chemical control

The extent of the use of chemical control methods has varied across different production systems in different countries. For example, in Japan, rice production is highly intensive and has involved substantial application of fungicides. However, in less economically prosperous rice-growing regions, effective fungicides can be physically inaccessible and/or prohibitively expensive. Fungicides can be effective at controlling blast disease, however, extensive use of fungicides has numerous deleterious consequences. They can cause environmental pollution, have negative effects on other organisms in the rice ecosystem, and can induce chemical resistance, preventing long-term efficacy in control of the fungus.

The widely used fungicide probenazole (under the brand name Oryzemate™) differs from most available fungicides by activating host plant defences, rather than directly targeting the fungus (Yoshioka et al., 2001). Application of probenazole has been shown

to induce production of antifungal compounds and defence-related enzymes and promote reinforcement of cell walls through lignification (Iwata, 2001). Fungicides that function in this way are less likely to disrupt other organisms in the environment or induce chemical resistance in the fungus and therefore represent a promising option for future chemical control of rice blast disease.

1.1.4.3 Genetic resistance

Developing resistant rice varieties is arguably the most economical and effective method of preventing rice blast disease. However, the production of varieties with durable resistance is complicated by the genetic diversity within and between *M. oryzae* populations, and the capacity of the fungus to evolve in response to widespread deployment and cultivation of resistant varieties.

Classical breeding programmes to develop resistant varieties involve identification of resistance in a crop variety or wild relative and crossing the resistant plant with an elite cultivar. Resistant progeny are selected, and backcrossed to the elite parent. Multiple cycles of selection and backcrossing are required to deliver a resistant plant with the desirable qualities of the elite cultivar.

Marker-assisted selection (MAS) can be exploited by breeders to avoid laborious and costly phenotypic screening and allow selection of desirable plant genotypes at the seedling stage (Collard and Mackill, 2008). This approach requires reliable markers close to the gene-of-interest, and a high-throughput, cost-effective screening process. AgRenSeq combines association genetics and resistance gene enrichment sequencing (RenSeq) to rapidly identify and clone novel resistance genes (Arora et al., 2019). Cloned resistance genes can be incorporated into breeding programmes, but also be used as precise molecular markers for marker-assisted selection.

Deployment of single resistance genes is not advisable due to the potential for a pathogen to overcome the resistance. Stacking, or pyramiding, multiple resistance genes is more likely to confer durable disease resistance (Djian-Caporalino et al., 2014). While it can be difficult to identify plants containing multiple resistance genes through phenotypic screening alone, marker-assisted pyramiding can be used to select plants containing the desired genes. This approach has been successfully used to incorporate blast resistance

genes into elite varieties (Pinta et al., 2013, Hittalmani et al., 2000). However, even when using molecular markers to expediate selection, the number of backcrosses required to avoid significant linkage drag means that this incorporating multiple resistance genes through conventional breeding approaches can take years, if not decades.

Genetic modification and gene editing technologies provide a wealth of opportunities for generating resistant varieties. These technologies can enable transfer of resistance genes from non-sexually compatible species, conferring novel recognition specificity. For example, transfer of a pigeonpea resistance gene into soybean conferred resistance to Asian soybean rust (Kawashima et al., 2016). Genetic modification also facilitates rapid incorporation of multiple resistance genes into an elite variety; three resistance genes against late blight from wild potato species were incorporated into popular African highland potato varieties and shown to confer resistance in the field (Ghislain et al., 2019). Incorporating these resistance genes via conventional breeding approaches would have required decades of backcrossing and selection to eliminate undesirable genetic material from the wild species. Finally, genetic modification opens avenues for rational engineering of novel resistance genes, based on a fundamental understanding of their structure and function (De la Concepcion et al., 2019, Kim et al., 2016, Helm et al., 2019). However, legislative barriers and public mistrust of genetic modification approaches can prevent breeders and farmers from benefiting from these approaches.

1.1.5 Wheat blast disease is an emerging threat

Different lineages of *Magnaporthe oryzae* can cause blast disease on different host species (Valent et al., 2019). These include the commercially important crops wheat (*Triticum aestivum*) and barley (*Hordeum vulgare*), and the subsistence crops finger millet (*Eleusine coracana*) and foxtail millet (*Setaria italica*) which are widely cultivated by smallholder farmers in Africa and Asia.

At present, the spread of wheat blast disease is a serious cause for concern. Wheat blast disease was first reported in Brazil in the mid-1980s, and was subsequently found in Bolivia, Paraguay and Argentina. It is thought to have emerged following a host jump mediated by widespread cultivation of a wheat cultivar lacking a key resistance gene (Inoue et al., 2017). The disease spread to Bangladesh from South America in 2016, where

it affected over 15,000 hectares of wheat, causing yield losses of up to 90 % (Islam et al., 2016). Wheat blast has since been reported in India (Islam et al., 2019), and there are concerns that the disease could spread through Asia and into Europe, with devastating consequences for wheat production (Cruz and Valent, 2017, Mottaleb et al., 2018).

While there are likely to be key differences between *M. oryzae* infection of wheat and rice, it is hoped that the extensive research on the rice pathosystem can be utilised to develop strategies to control wheat blast disease.

1.2 Most plants are resistant to most pathogens

Plants are resistant to the majority of the pathogens they encounter in their natural environment. This resistance is mediated by diverse constitutive and inducible defence mechanisms, sometimes referred to as nonhost resistance. Successful infection of a plant by a pathogen requires specific adaptations to circumvent these defences.

1.2.1 Constitutive barriers to pathogen invasion

Plants present a physical barrier to infection in the form of the cuticle and cell wall. The cuticle consists predominantly of the waxy polymer cutin and other hydrophobic waxes, and covers the aerial epidermis. It provides protection against both abiotic and biotic stresses. The cell wall, primarily composed of cellulose microfibrils in a matrix of hemicelluloses and pectin, represents a second physical barrier. The cell wall is not just a passive barrier to infection, it can be actively remodelled and reinforced in response to pathogen attack (Underwood, 2012).

Many biotrophic and hemibiotrophic fungal pathogens such as *Magnaporthe oryzae* have evolved mechanisms to directly penetrate both the cuticle and cell wall in order to establish infection (de Jong et al., 1997, Chumley and Valent, 1990). By contrast, bacterial pathogens typically exploit natural openings in the plant cuticle, such as stomata, or wounding sites caused by pest attack or environmental damage. Necrotrophic pathogens, which kill plant cells and survive on the dead tissue, often secrete hydrolytic enzymes to degrade the plant cell walls.

In addition to these physical barriers, many plants possess a chemical barrier against attack in the form of constitutively produced antimicrobial compounds (Osbourn, 1996). While some compounds are diffuse and prevent microbial growth on the surface of the plant, many are stored in the vacuole or other organelles and released upon cell damage. Biotrophic and hemibiotrophic pathogens may avoid the effects of these compounds, as they do not destroy the plant cells during infection.

1.2.2 Induction of defences by recognition of conserved "patterns"

Plants can induce additional defence responses through recognition of conserved molecular patterns that indicate the presence of an invading microbe. These molecular patterns can broadly be divided into "non-self" microbe-associated molecular patterns (MAMPs), or "modified-self" damage-associated molecular patterns (DAMPs). MAMPs are conserved molecules found in a broad group of microorganisms and are generally either integral to the structure or function of the microbe or are produced by a core metabolic activity (Boller and Felix, 2009). DAMPs are host-derived molecules whose presence indicates microbial attack (Hou et al., 2019).

MAMPs and DAMPs are recognised by transmembrane pattern recognition receptors (PRRs), which can be classified as either receptor-like kinases (RLKs) or receptor-like proteins (RLPs). Receptor-like kinases consist of an intracellular kinase domain, a single-pass transmembrane domain and an extracellular "sensor" domain, which may incorporate leucine-rich repeats (LRRs), carbohydrate-binding lysine motifs (LysMs), lectin motifs or epidermal growth factor (EGF)-like domains (Macho and Zipfel, 2014). Receptor-like proteins share a similar architecture but lack the kinase domain and require interaction with other proteins to initiate downstream signalling pathways. While the conserved structure of RLKs and RLPs has enabled the identification of an abundance of PRRs in plant genomes, for the most part the microbial targets of these receptors remain unknown.

Fungal MAMPs identified to date include chitin oligosaccharides released during infection as a consequence of fungal cell wall degradation by host enzymes (Stacey and Shibuya, 1997, Wan et al., 2008), and a 22kDa fungal protein, ethylene-inducing xylanase (EIX) (Bailey et al., 1990, Ron and Avni, 2004). In rice, chitin oligosaccharides

are recognised by OsCEBiP (chitin oligosaccharide elicitor binding protein), an RLP with two extracellular LysM domains (Hayafune et al., 2014, Kaku et al., 2006, Gubaeva et al., 2018). As an RLP, OsCEBiP lacks a kinase signalling domain and requires interaction with OsCERK1 (Hayafune et al., 2014, Shimizu et al., 2010, Kaku et al., 2006). Recognition of EIX involves LeEIX1 and LeEIX2, receptor-like proteins (RLPs) identified in tomato (*Solanum lycopersicon*, formerly *Lycopersicon esculentum*). They consist of an extracellular leucine-rich repeat domain, transmembrane domain and cytoplasmic domain with a mammalian endocytosis signal, and share 85% sequence similarity (Ron and Avni, 2004). The LRR-RLK BAK1 has been shown to be involved in signalling responses to a number of MAMPs including bacterial flg22 and elf26 (Chinchilla et al., 2007, Heese et al., 2007) and the DAMPs Pep1 and Pep2.

Events downstream of PRRs include ion fluxes across the plasma membrane, production of reactive oxygen species, cell wall reinforcement, callose deposition, closure of plasmodesmata and stomata, and changes in gene expression, such as upregulation of pathogenesis-related (PR) proteins (reviewed in (Boller and Felix, 2009)). This array of defensive responses is collectively referred to as pattern-triggered immunity, or PTI, and is typically mediated by cell-surface receptors.

1.3 Effectors suppress plant defences and promote infection

In order to cause disease, pathogens must overcome both constitutive and inducible defences. In many cases, this is achieved through the production and secretion of proteins known as effectors. Effectors are structurally and functionally diverse, but ultimately promote pathogen virulence through suppressing plant defences and manipulating host physiology. Traditional views of the plant immune system drew lines between molecular patterns and effectors; as our understanding of plant immunity has developed, it has become apparent that these lines are blurred. Some effectors are widely distributed across different species and can themselves be considered as molecular patterns (Stergiopoulos et al., 2010). Similarly, certain MAMPs may have a function in suppressing plant defences, similar to effectors.

Certain effectors are recognised by host intracellular immune receptors (discussed in detail in section 1.4), activating plant defence responses and preventing infection. For this reason, many characterised effectors have been somewhat counterintuitively termed avirulence (AVR) proteins, as their presence can be detected by a cognate immune receptor, triggering plant defences and preventing the pathogen causing disease. However, in host plants lacking the corresponding immune receptor, the effector contributes to pathogen virulence.

1.3.1 Identification of effector proteins

Gram-negative bacterial pathogens deliver effector proteins into the plant cell cytoplasm using the type III secretion system (TTSS). Translocation of bacterial effectors requires an export signal in the N-terminal region of the protein; the export signal does not include a specific amino acid sequence but can be reliably identified by certain patterns and compositions of amino acids (Wang et al., 2011). Importantly, the TTSS export signal has facilitated the identification of effector repertoires in many important bacterial phytopathogens.

For oomycetes, the discovery of conserved motifs required for effector translocation has enabled identification of two large effector families. RXLR effectors are defined by a N-terminal signal peptide, followed by an RXLR motif (where X represents any amino acid) and an acidic region commonly ending in the amino acids EER. Mutation of these motifs has been shown to prevent translocation of the RXLR effector AVR3a (Whisson et al., 2007). The CRN (Crinkling and Necrosis) effector family share a conserved N-terminal domain containing a LXLFLAK motif, which has been shown to be involved in effector translocation (Schornack et al., 2010). The C-terminal domains of RXLR and CRN effectors are highly diverse (Win et al., 2007), and the functions of many of these effectors are yet to be elucidated.

Identification of fungal effectors has proved considerably more challenging due to the apparent lack of conserved motifs required for translocation into the host cell. Fungal pathogens exhibit considerable diversity in their modes of infection and appear to deliver effectors through different mechanisms. For example, *Ustilago maydis* secretes effectors from the invading hyphal tip (Bielska et al., 2014), while secretion of cytoplasmic

effectors by *Magnaporthe oryzae* involves a specialised structure known as the biotrophic interfacial complex (discussed further in section 1.3.3)(Giraldo et al., 2013). Approaches to predict effectors from fungal secretomes have typically relied upon characteristics shared by many experimentally validated effectors, such as small size, high cysteine content, genomic location and evidence for diversifying selection, however these are not universal traits of known fungal effector proteins. More recently, the machine-learning method EffectorP has been developed to identify fungal effectors (Sperschneider et al., 2016, Sperschneider et al., 2018). However, despite the availability of genomic data for multiple fungal plant pathogens, reliable identification of fungal effectors remains a challenge.

1.3.2 Effectors interact with a range of host targets

Effector proteins are highly diverse, and in many cases show little to no sequence similarity to proteins of known function. Of the effectors that have been functionally characterized, many have been found to have roles in suppression of plant immune signalling, manipulation of host metabolism, and transcriptional reprogramming (Khan et al., 2018).

Strains of *Pseudomonas syringae* are reported to secrete between 9 and 39 effectors into the host cell cytoplasm (Baltrus et al., 2011), while analysis of the effector complement of *Ralstonia solanacearum* has shown that strains typically contain 60 to 75 distinct effectors (Deslandes and Genin, 2014). By contrast, the genomes of filamentous plant pathogens are predicted to contain several hundreds of effector proteins (Lo Presti et al., 2015, Saunders et al., 2012, Nemri et al., 2014). The reduced effector repertoire of bacterial plant proteins may reflect their smaller overall genome size. Alternatively, the extended effector complement of filamentous plant pathogen may be a consequence of a more complex infection process (Büttner, 2016). Infection by most filamentous plant pathogens involves multiple discrete stages, each involving the formation and maintenance of complex structures. Filamentous pathogens typically form extensive interaction interfaces with the host, and effectors are likely to be required to prevent initiation of plant defences. Furthermore, biotrophic pathogens rely upon the living host for survival, and remodel host metabolism to obtain essential nutrients. This extensive metabolic reprogramming is likely to require contributions from multiple effectors.

While the size and content of the effector repertoire is variable between pathogen strains and species, certain effectors are consistently found in multiple strains or across species. These effectors have been termed "core" effectors and are hypothesised to have key functions in pathogen virulence that have prevented their loss during pathogen diversification. Functional characterisation of these "core" effectors is likely to identify key processes targeted by plant pathogens which could in turn create opportunities to engineer the plant to evade manipulation by the pathogen.

An emerging theme in studies of effector function is that many effectors, even within a pathogen species, target similar host proteins. This redundancy may offer a buffer, should an effector become recognised by the plant immune system. Alternatively, the expression of the effectors may be differentially regulated, and they may be produced at different stages of infection.

Conversely, individual effectors have been shown to target multiple host proteins. A meta-analysis of effector-target interactions found that 68% of bacterial effectors target more than one host protein, with 32% targeting multiple host proteins from different molecular classes (Khan et al., 2018). For example, the *Pseudomonas syringae* effector HopZ1a is an acetyltransferase which can acetylate various host targets to suppress plant immunity. These targets include tubulin, interfering with transport of immunity-related proteins, and JAZ proteins, inducing their degradation and indirectly downregulating salicylic acid-induced immune signalling pathways (Jiang et al., 2013, Lee et al., 2012). Effector-target interactions have been less widely characterised in fungal plant pathogens, largely due to challenges in identification of effectors as discussed previously. However, the *Magnaporthe oryzae* effector AVR-Piz-t has been reported to interact with at least four different host targets, specifically include two RING E3 ubiquitin ligases, APIP6 (Park et al., 2012) and APIP10 (Park et al., 2016), a bZip-type transcription factor, APIP5 (Wang et al., 2016), and a protein with homology to the nucleoporin protein Nup98, APIP12 (Tang et al., 2017).

As more fungal effectors are functionally characterised, it will be interesting to see if there is a similar trend for a single effector interacting with multiple host targets. It is tempting to speculate that the smaller genome size of bacterial pathogens may have selected for multi-functional effector proteins.

The capacity of effector proteins to interact with several different host proteins is likely to be closely linked to their structure. Some effectors may display a multidomain architecture, with different domains interacting with different host targets. In other cases, flexible or disordered regions may support interaction with multiple targets. Obtaining structural information about how effectors bind their targets is likely to advance our understanding of how effectors can interact with multiple host proteins.

1.3.3 *Magnaporthe oryzae* secretes apoplastic and cytoplasmic effectors

The effectors produced by *Magnaporthe oryzae* can be divided into two classes based on their localisation to either the plant cell cytoplasm or the apoplastic space between the fungal cell wall and the plant-derived extra-invasive hyphal membrane (EIHM) (Giraldo et al., 2013). Apoplastic effectors are dispersed throughout the EIHM compartment, while cytoplasmic effectors are preferentially accumulated in the biotrophic interfacial complex (BIC), a plant-derived, membrane-rich structure which is initially located at the tip of primary invasive hyphae but which becomes subapical as the invasive hyphae develop. While secretion of apoplastic effectors is via the conventional ER-to-Golgi secretory pathway that has been described in other fungi, secretion of cytoplasmic effectors involves a different mechanism requiring the exocyst complex components Exo70 and Sec5, and the t-SNARE Sso1 (Giraldo et al., 2013).

1.3.4 *Magnaporthe oryzae* effectors

While a considerable number of *Magnaporthe oryzae* effector candidates have now been identified and cloned (Petit-Houdenot et al., 2019), only a few have been functionally characterised. These include AVR-Pita, AVR-Pii, AVR-Piz-t and ACE1.

AVR-Pita is a 223aa protein which shows similarity to neutral zinc metalloproteases from *Aspergillus* sp (Orbach et al., 2000, Jia et al., 2000). It most closely resembles NpII from *Aspergillus oryzae*, and shared residues include a 10aa sequence containing two histidine residues thought to chelate zinc (Orbach et al., 2000). However, the putative metal binding properties of AVR-Pita have not been confirmed experimentally, and the function of this effector remains unknown.

AVR-Pii is a 70aa secreted protein which interacts with OsExo70-F2 and OsExo70-F3, members of the OsExo70 family of rice proteins involved in polarised exocytosis as part of the exocyst complex (Fujisaki et al., 2015). The purpose of this interaction remains uncertain, though it is necessary for Pii-dependent resistance (Fujisaki et al., 2015).

AVR-Piz-t has been shown to interact with numerous host targets, including two RING E3 ubiquitin ligases, APIP6 (Park et al., 2012) and APIP10 (Park et al., 2016), a bZip-type transcription factor, APIP5 (Wang et al., 2016), and a protein with homology to the nucleoporin protein Nup98, APIP12 (Tang et al., 2017). AVR-Piz-t suppresses the ubiquitination activity of both APIP6 and APIP10, and is in turn ubiquitinated and targeted for degradation via the 26S proteasome pathway. Both APIP6 and APIP10 appear to have a role in PTI, as knockdown lines show compromised immune responses (Park et al., 2012, Park et al., 2016). AVR-Piz-t also suppresses the transcriptional activity and protein accumulation of APIP5. Interestingly, APIP5 appears to function as a negative regulator of cell death, and the activity of AVR-Piz-t promotes necrosis (Wang et al., 2016). It is somewhat surprising that AVR-Piz-t can interact with such diverse targets with seemingly contradictory outcomes; the structural basis and regulation of these interactions remains to be determined.

ACE1 is a polyketide synthase-nonribosomal peptide synthetase (PKS-NRPS) hybrid protein (Bohnert et al., 2004). ACE1 is located in a cluster with 15 other genes, 14 of which are thought to be involved in secondary metabolism and all of which are upregulated during the penetration stage of the infection process. It is thought that ACE1 acts as part of a biosynthetic pathway which produces a secondary metabolite that is then detected by Pi33, the cognate NLR protein for ACE1 (Collemare et al., 2008).

Chapters 3 and 4 of this thesis focus on the interaction between two additional cytoplasmic *M. oryzae* effectors, AVR-Pik and AVR-Pia, and a putative virulence target. Understanding how *M. oryzae* effectors perturb the plant immune system can offer new insights into the requirements for successful infection. This in turn can inform strategies to engineer resistance to the fungus.

1.3.5 The MAX effectors: a structurally conserved effector family

Structural studies of *Magnaporthe oryzae* effector proteins has led to the identification of a sequence-diverse but structurally conserved family of effector proteins named the MAX (Magnaporthe AVRs and ToxB-like) effectors (de Guillen et al., 2015). The family includes at least four *Magnaporthe oryzae* effectors (AVR-Piz-t, AVR-Pia, AVR1-CO39 and AVR-Pik) and ToxB, a proteinaceous toxin from *Pyrenophora tritici-repentis*.

The MAX effectors share a conserved six-stranded β-sandwich fold, which is generally stabilised by a disulphide bond between cysteine residues in β1 and at the start of, or immediately prior to, β5. β1, β2 and β6 typically form one β-sheet, with β3, β4 and β5 forming a second. The length and orientation of the β-strands vary, as do the lengths and positions of the loops connecting adjacent β-strands. Notably, AVR-PikD has a 22 amino acid extension at the N-terminus.

Interestingly, despite sharing a common fold, the *Magnaporthe oryzae* effectors appear to target different host proteins. As described earlier, AVR-Piz-t interacts with diverse host targets to modulate the plant ubiquitination system and manipulate gene expression (Wang et al., 2016, Tang et al., 2017, Park et al., 2012, Park et al., 2016). By contrast, AVR-Pia, AVR1-CO39 and AVR-Pik interact with proteins containing a heavy metal associated domain (discussed in section 1.6). ToxB causes chlorosis in host plants, but while the N-terminal region of the protein plays a crucial role in the activity of ToxB (Figueroa Betts et al., 2011), the precise mechanism of action of the toxin remains unknown.

This is not the first example of a conserved structural fold among sequence diverse effector proteins. The crystal structures of AVR3a11 from *Phytopthora capsici* and PexRD2 from *P. infestans* revealed that, despite sharing less than 20% sequence identity, the two proteins share a core α-helical fold consisting of W and Y motifs, known as the WY-fold (Boutemy et al., 2011). A screen of the RXLR and non-RXLR proteome of *P. infestans* and *Hyaloperonospora arabidopsidis* indicated that the WY-fold is present in around 44% of *P. infestans* and 26% of *H. arabidopsidis* RXLR effectors. The loops between α-helices exhibit variation in both length and composition, and the α-helices themselves can vary at the amino acid level as long as the hydrophobic core of the helical

bundle remains conserved (Boutemy et al., 2011). Further structural studies of effector proteins may identify novel effector families with conserved structures.

1.4 Intracellular NLR immune receptors detect specific effectors to trigger immunity

Certain cytoplasmic effectors can be recognised by intracellular immune receptors of the nucleotide binding, leucine rich repeat (NLR) protein family. Many of the resistance (R) genes that have been identified and incorporated into breeding programmes to date encode NLR proteins.

1.4.1 NLR protein architecture

NLR proteins form part of the innate immune system in both plants and animals, though are thought to have independent origins and have evolved convergently (Yue et al., 2012). While NLR proteins in both systems play key roles in intracellular surveillance and activation of immune responses, animal NLR proteins detect conserved non-self and modified-self molecular patterns (similar to plant PRRs) rather than the pathogen/strain-specific effectors recognised by plant NLR proteins.

NLR proteins are members of the STAND (Signal Transduction ATPase with Numerous Domains) protein family. They are multidomain proteins, consisting of a C-terminal leucine-rich repeat domain, a central NB-ARC (nucleotide-binding Apaf-1, R protein, CED4-shared) domain, and a variable N-terminal domain. In plants, the N-terminus typically consists of either a TOLL/interleukin-1 receptor (TIR) domain or a coiled coil (CC) domain, and this distinction has been used to classify NLRs into two subfamilies (Pan et al., 2000). In animals, N-terminal domains are more variable, though prevalent N-terminal domains include a caspase recruitment domain (CARD), pyrin domain (PYD), or a baculovirus inhibitor of apoptosis protein repeat (BIR) domain.

1.4.2 Recognition of effector proteins

NLR proteins detect the presence of their cognate effector either through directly interacting with the effector, or by sensing an effector-mediated change in an intermediate protein (figure 1.2). This intermediate protein may be the virulence target

of the effector, which is "guarded" by the NLR protein (the guard model) (Jones and Dangl, 2006), or a non-functional mimic of the virulence target (the decoy model) (van der Hoorn and Kamoun, 2008). A well-characterised example of indirect effector sensing is the recognition of the *Pseudomonas syringae* effector AVR-Pto by the tomato NLR protein Prf. Prf interacts with the serine/threonine kinase Pto in the absence of the effector, and a direct interaction between AVR-Pto and Pto instigates Prf-dependent immune signalling (Oh and Martin, 2011). However, whether this example typifies the guard or decoy model is undetermined; Pto has not been implicated in basal defence, and it has been suggested that the true virulence targets of AVR-Pto are receptor-like kinases involved in PTI and that Pto therefore functions as a decoy (van der Hoorn and Kamoun, 2008).

Several examples of a direct interaction between an effector and cognate NLR protein have been reported. The *Magnaporthe oryzae* effector AVR-Pita was shown to directly interact with the LRR domain of the rice NLR protein Pi-ta in yeast two-hybrid and far western analyses (Jia et al., 2000). Yeast two-hybrid experiments also demonstrated a direct interaction between *Melampsora lini* AVR567 effector alleles and the cognate flax (*Linum usitatissimum*) L567 NLR protein alleles (Dodds et al., 2006). Ideally, direct interactions between effectors and NLR proteins would be validated *in vitro* using purified recombinant proteins, however ongoing challenges in producing full-length NLR proteins in heterologous systems have so far prevented this.

A recent advance in the field of NLR biology has been the identification of unconventional protein domains in many NLR proteins. The integrated decoy hypothesis (Cesari et al., 2014a) proposes that these domains share a common origin with the host target of the effector recognised by the NLR protein, and function as effector sensor domains. While these domains were initially referred to as "integrated decoys", the term "integrated domains" is generally preferred (Wu et al., 2015), as it makes no assumptions about the function, or lack of function, of these domains.

Studies have estimated that 3.5-5.2% of higher plant NLR proteins contain integrated domains (Kroj et al., 2016, Sarris et al., 2016). The identities of these protein domains are diverse, and include DNA-binding BED and WRKY domains, protein kinases, thioredoxins and heavy-metal associated domains. The integration sites are also variable;

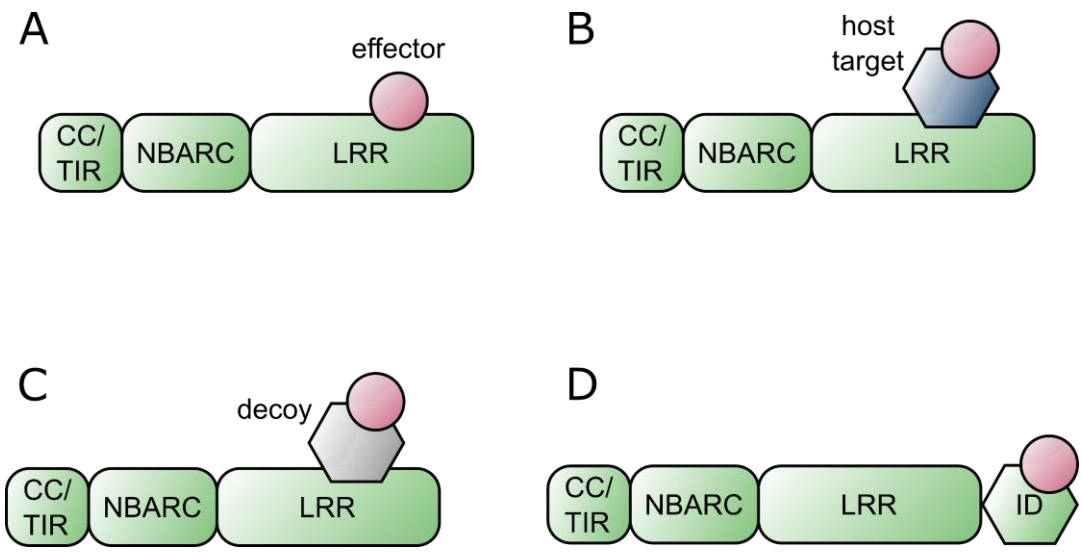


Figure 1.2 NLR proteins detect the presence of their cognate effector in different ways.

Schematic representations of the different models of effector perception by NLR proteins (shown in green). The effector is represented as a pink circle. The host target is represented as a purple hexagon, and a non-functional mimic of the host target as a grey hexagon. Abbreviations used for NLR protein domains are CC = coiled coil, TIR = TOLL/interleukin-1 receptor, NBARC = nucleotide-binding Apaf-1, R protein, CED4-shared, LRR = leucine rich repeat, ID = integrated domain. ***A.*** *Recognition via a direct interaction between the effector and a canonical domain of the NLR protein.* ***B.*** *Indirect recognition of an effector by "guarding" and detecting a change in the target of the effector (the guard model).* ***C.*** *Indirect recognition of an effector by detecting a change in a non-functional mimic of the target of the effector (the decoy model).* ***D.*** *Recognition via a direct interaction between the effector and an integrated domain resembling the effector target.*

integrated domains have been identified at both termini of NLR proteins, as well as between canonical NLR protein domains. Integration of non-canonical protein domains is proposed to have occurred independently on multiple occasions. A study of NLR proteins containing integrated domains in grasses identified an amino acid motif upstream of the integration which may have made the protein more amenable to integration of a non-canonical domain (Bailey et al., 2018). The mechanism by which unusual domains are integrated into NLR proteins is an area for future study.

A direct interaction between an effector and an integrated domain has been demonstrated in several cases. The *Arabidopsis thaliana* NLR protein RRS1 contains an integrated WRKY domain towards the C-terminus of the protein, and the effectors AvrRps4 (from *Pseudomonas syringae*) and PopP2 (from *Ralstonia solanacearum*) directly interact with the WRKY domain to trigger immune signalling (Sarris et al., 2015). The rice NLR proteins Pik-1 and RGA5 each contain an integrated heavy metal associated domain, to which their cognate *Magnaporthe oryzae* effectors AVR-Pik and AVR-Pia/AVR1-CO39 (De la Concepcion et al., 2018, Maqbool et al., 2015, Guo et al., 2018, Ortiz et al., 2017) bind to activate plant defences.

In light of the discovery of integrated domains, and the increased availability and power of bioinformatic tools for prediction of protein domains, it will be interesting to revisit cases where an effector has been reported to directly interact with canonical NLR protein domains. AVR-PikD was initially reported to directly interact with the CC domain of the rice NLR Pik-1 (Kanzaki et al., 2012), however it is now known that Pik-1 contains an integrated domain adjacent to the CC domain with which AVR-PikD interacts.

1.4.3 Singletons, pairs, networks

Harold Flor's gene-for-gene model proposes that a single R gene mediates a response to a specific effector protein. In some instances this appears to be the case, with a single NLR protein capable of triggering effector-dependent immunity. Examples of these so-called "singleton" NLR proteins include the MLA (mildew locus A) proteins, which directly recognise the powdery mildew effector AVRa. Different MLA alleles respond to different AVRa alleles, however transient expression of a single MLA allele with its cognate effector in a heterologous system is sufficient to trigger HR-like cell death (Saur

et al., 2019). Similarly, the *Arabidopsis thaliana* NLR protein RPS5 and RPP13 trigger HR-like cell death in a heterologous system when transiently expressed alongside their cognate effectors AvrPphB and ATR13, respectively (Ade et al., 2007, Leonelli et al., 2011). These "singleton" NLR proteins are capable of both sensing the presence of a specific effector and activating downstream immune signalling pathways. The structural basis of this activation has recently been elucidated for the *Arabidopsis thaliana* "singleton" NLR protein ZAR1 (discussed in detail in section 1.4.4).

However, some NLR proteins have specialised as either "sensor" NLR proteins, responsible for effector detection, or "helper" NLR proteins, responsible for activating immune signalling. Sensor and helper NLR proteins can either function exclusively with a specific partner, or form part of a promiscuous immune receptor network, where multiple sensor NLR proteins can interact with one or more helper NLR proteins.

Several sensor/helper pairs have been identified, where the two partners function together to active immune signalling in response to a particular effector. In many cases, these NLR protein pairs are genetically linked, with transcription controlled by a shared promoter. Interestingly, these NLR protein pairs do not share a common mechanism of action. In the case of the paired NLR proteins RRS1/RPS4 and RGA4/RGA5, the "helper" NLR protein (RPS4 or RGA4) is autoactive, and constitutively triggers HR-like cell death. In the absence of an effector, the "sensor" NLR protein (RRS1 or RGA5) represses the autoactivity of the "helper". When the effector is present, this repression is relieved, and the "helper" NLR protein triggers immune signalling. By contrast, neither Pik-1 (sensor NLR) nor Pik-2 (helper NLR) is autoactive, and both proteins are required for effector-triggered immunity. This indicates that paired NLR proteins can deliver immunity through at least two distinct mechanisms; negative regulation of the helper NLR protein by the sensor, and cooperation of the helper and sensor.

A study of solanaceous NLR proteins revealed an extensive network of CC-NLR immune receptors (Wu et al., 2017). The network involves a wide array of sensor NLR proteins, capable of detecting effector proteins from diverse bacterial, fungal, oomycete, nematode and viral pathogens, which interact with a smaller number of helper NLR proteins known as NRC (NLR required for cell death) proteins. A single NRC protein can be required for the activity of multiple different sensor NLR proteins. Unlike the pairs

discussed earlier, the sensor and helper NLR proteins comprising this network do not appear to be clustered together in the genome (Wu et al., 2017). Networks involving TIR-NLR proteins have also been identified; multiple TIR-NLR proteins require members of the RPW8-NLR protein NRG1 subclade to activate immune signalling (Castel et al., 2019).

1.4.4 Activation of NLR immune receptors

The NB-ARC domain contains three ATP binding motifs, namely kinase-1a (also known as p-loop or Walker A), kinase-2a (Walker B) and kinase-3a, and is capable of binding and hydrolysing ATP (Tameling et al., 2002). This ATPase activity has long been thought to be critical for the signalling function of NLR proteins, with ATP association and hydrolysis mediating transition between OFF and ON states (Takken et al., 2006, Takken and Goverse, 2012).

Two seminal papers have provided significant structural insights into NLR protein activation (Wang et al., 2019b, Wang et al., 2019a). HOPZ-ACTIVATED RESISTANCE 1 (ZAR1) is an *Arabidopsis thaliana* CC-NLR protein which recognises at least three bacterial effectors through association with different pseudokinases. The *Xanthomonas campestris* pv. *campestris* effector AvrAC is recognised by ZAR1. In the absence of the effector, a molecule of inactive, ADP-bound ZAR1 forms a heterodimer with the pseudokinase RKS1. AvrAC uridylates a receptor-like cytoplasmic kinase, PBS1-like protein 2 (PBL2), and the modified PBL2^{UMP} binds to the ZAR1-RKS1 complex. Cryo-EM structures of the ZAR1-RKS1 complex with and without PBL2^{UMP} revealed that binding of PBL2^{UMP} causes conformational changes in the nucleotide-binding domain of ZAR1 which lead to the release of bound ADP from the NLR protein, priming it for activation (Wang et al., 2019b). Subsequent binding of ATP induces oligomerisation of the ZAR1-RKS1-PBL2^{UMP} complex to form a pentameric wheel-like structure that has been termed the resistosome (Wang et al., 2019a). The cryo-EM structure of the resistosome identified a key conformational change in α1 of the ZAR1 CC domains; the α-helices project outwards and form a 5-helix funnel-like structure. It has been proposed that this structure could cause a hypersensitive response by inserting into the plasma membrane to form pores, analogous to certain cell-death causing toxins (Wang et al., 2019a).

The structural characterisation of a full-length NLR protein at different stages of activation represents a significant advance in our understanding of plant NLR protein biology. The ZAR1 resistosome is reminiscent of the oligomeric inflammasome structure formed following activation of the animal NLR proteins NAIP2 and NLRC4 (Zhang et al., 2015a). Oligomerisation of NLR proteins may represent a conserved feature of activation. ZAR1 is a singleton NLR protein, however as described in section 1.4.3 many NLR proteins are specialised as either sensors or helpers and function in pairs or as part of complex networks. It remains to be seen how the ZAR1 activation model may apply to paired NLR proteins. One possibility is that an activated NLR protein pair recruits additional copies of one or both of the NLR proteins into an oligomeric structure. For example, the NAIP2/NLRC4 inflammasome consists of a single molecule of NAIP2 assembled with multiple copies of NLRC4. Another area for future study is how integrated domains can be incorporated into the ZAR1 model of NLR activation. The pseudokinase RKS1 is constitutively associated with ZAR1 and is responsible for binding PBL2^{UMP}. RKS1 functions similarly to an integrated domain, and integrated domains at the C-terminus of an NLR protein may function in a similar manner. However, integrated domains have been identified both at the N-termini of NLR proteins and between canonical protein domains (Bailey et al., 2018, Sarris et al., 2016, Kroj et al., 2016), and as of yet it is unknown how these integrated domains align with the ZAR1 activation model.

1.4.5 The hypersensitive response

Recognition of an effector by its cognate NLR protein activates signalling pathways which culminate in a suite of events collectively termed the hypersensitive response (HR). Little is known about the signalling processes downstream of CC-NLR proteins which culminate in a hypersensitive response. By contrast, immune responses triggered by TIR-NLR proteins require the nucleocytoplasmic lipase-like protein EDS1 (Aarts et al., 1998), which forms a heterodimeric complex with either SAG101 or PAD4 (Wagner et al., 2013).

A characteristic feature of the HR is localised necrotic cell death, which is thought to prevent colonisation of adjacent cells, restricting the spread of the pathogen. While it was originally thought that this cell death alone conditioned resistance to a pathogen

following effector perception, it has since been shown that cell death and resistance can be uncoupled. Resistance can be activated in the absence of cell death; the *Pseudomonas syringae* effector HopZ5 triggers resistance in *Arabidopsis thaliana* Col-0 plants without induction of cell death (Jayaraman et al., 2017). Conversely, cell death is not always sufficient for resistance. An *A. thaliana* mutant could induce cell death in response to certain *P. syringae* strains but was still susceptible to infection (Century et al., 1995). The interplay between cell death and resistance is an area of ongoing research in the plant pathology field.

1.4.6 NLR proteins as targets for engineering

As most R genes characterised to date encode NLR proteins, these immune receptors are attractive targets for protein engineering. Even though approaches such as AgRenSeq are accelerating the discovery of novel NLR proteins in crops and their wild relatives, they can only identify what is already present in nature. Pathogens evolve rapidly and can modify their effector repertoire to evade the plant immune system. Using synthetic biology to engineer NLR proteins capable of recognising recently evolved and/or currently unrecognised effectors could provide an extended suite of NLR proteins collectively conferring durable resistance to various pests and pathogens.

The discovery of integrated domains in NLR proteins has opened up new avenues for engineering. A particularly attractive goal is to exchange the integrated domain of a characterised NLR protein for an alternative protein domain with the aim of generating an NLR protein capable of recognising novel effectors.

This idea is exemplified by ongoing work to engineer NLR protein-mediated recognition of proteases from various pathogen species. The *Pseudomonas syringae* effector protease AVRPphB cleaves the protein kinase PBS1. This proteolytic activity is detected by the *Arabidopsis* NLR protein RPS5, which activates plant defences (Shao et al., 2003, Ade et al., 2007). Exchanging the AVRPphB cleavage site in PBS1 for the cleavage site of different effector proteases altered the recognition specificity of RPS5 in the model system *Arabidopsis thaliana* (Kim et al., 2016). A PBS1 variant containing the cleavage site of the soybean mosaic virus NIa protease could be cleaved by the protease, and proteolytic cleavage triggered cell death in soybean protoplasts (Helm et al., 2019).

Additionally, it has been shown that AVRPphB can induce an immune response in both barley and wheat. In both systems, cleavage of PBS1 orthologs by AVRPphB is detected by a NLR protein (PBR1) (Carter et al., 2019). Engineering PBS1 variants with diverse protease cleavage sites is therefore a highly promising strategy to develop novel resistances in important crop species.

1.5 The paired NLR proteins Pik-1 and Pik-2 recognise the rice blast effector AVR-Pik via an integrated heavy metal associated domain in Pik-1

1.5.1 Pik-1 and Pik-2

A pair of rice NLR proteins, Pik-1 and Pik-2, cooperatively activate plant defences in response to the *Magnaporthe oryzae* effector AVR-Pik. The Pik locus is located on the long arm of chromosome 11. The Pik-1 and Pik-2 genes are genetically linked; the genes are arranged in a head-to-head orientation with a shared promoter (figure 1.3).

While Pik-2 has a typical CC-NLR structure, Pik-1 contains a heavy metal associated (HMA) domain (pfam PF00403) integrated between the canonical CC and NB-ARC domains (figure 1.4). The HMA domain is conserved among prokaryotes and eukaryotes and is found in proteins involved in transport or detoxification of a range of heavy metals.

The HMA domain was first characterised in the copper-transporting P-type ATPases MNK and WND, which are involved in Menkes and Wilson disease, respectively (Bull and Cox, 1994, Bull et al., 1993). The domain includes an MXCXXC metal-binding motif (where X represents any amino acid); the two cysteine residues can coordinate a bound metal ion (Arnesano et al., 2001). The solution structure of the fourth HMA domain of MNK shows it is comprised of a four-stranded antiparallel β-sheet and two α-helices packed in a αβ-sandwich fold (Gitschier et al., 1998). This structure is often referred to as a "ferredoxin-like" fold.

Interestingly, Pik-1 is not the only rice NLR protein known to contain an integrated HMA domain. The paired rice NLR proteins RGA5 and RGA4 trigger plant defences in response to the *M. oryzae* effectors AVR-Pia and AVR1-CO39 (Cesari et al., 2013). There

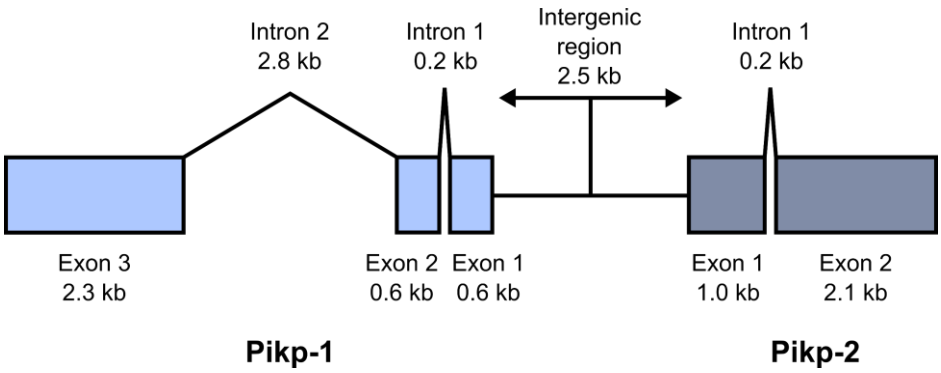


Figure 1.3 Map of the Pik locus.

Exons are represented as rectangles and introns as connecting diagonal lines. The intergenic region, containing the promoter, is represented as a straight line with arrows indicating the direction of transcription.

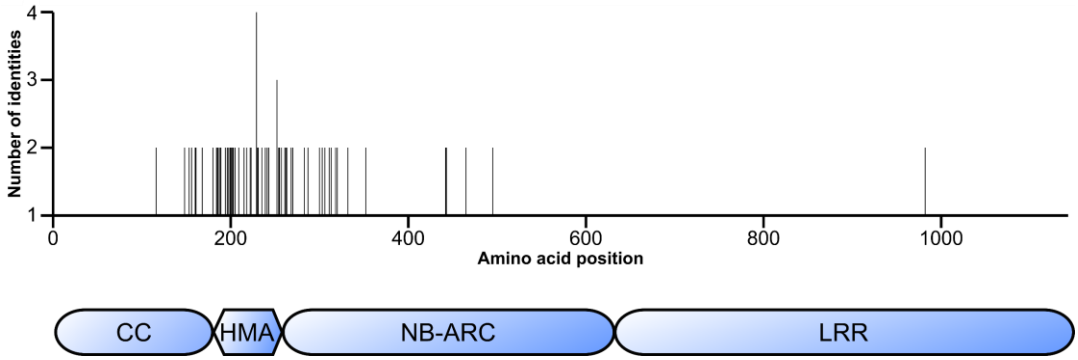


Figure 1.4 Amino acid polymorphisms between Pik-1 alleles cluster in and around the HMA domain.

Amino acid sequences were obtained for each of the Pik-1 alleles in table 1.1 from Genbank or directly from the referenced paper. Sequences were aligned in Clustal Omega. The number of different amino acid identities at each position in the Pik-1 protein was determined and plotted in R using the ggplot2 package (R Core Development Team, 2018, Wickham, 2016). A schematic representation of the Pik-1 domains is shown below the plot for reference.

is a HMA domain (initially described as a RATX1, related to ATX1, domain) at the C-terminus of RGA5, to which AVR-Pia and AVR1-CO39 bind (Cesari et al., 2013, Ortiz et al., 2017, Guo et al., 2018).

To date, seven Pik-1/Pik-2 alleles have been described at this locus (table 1) and they share between 95 % and 99 % amino acid sequence similarity. Interestingly, most of the polymorphic residues between Pik-1 alleles are clustered around the HMA domain (figure 1.4).

1.5.2 AVR-Pik

The *Magnaporthe oryzae* effector AVR-Pik was cloned using an association genetics approach (Yoshida et al., 2009). This approach identified 5 alleles of AVR-Pik (designated A-E), differing in between 1 and 4 amino acid positions (figure 1.5). A subsequent analysis of 39 *M. oryzae* isolates collected from Asia, Africa, Europe and America found that isolates containing AVR-PikD, AVR-PikE, AVR-PikA and AVR-PikC were present worldwide, however AVR-PikB was not identified in any of the 39 isolates (Kanzaki et al., 2012). From this, the authors concluded that AVR-PikB is a comparatively rare allele.

Recently, a novel AVR-Pik allele was identified in Thai *M. oryzae* isolates and designated AVR-PikF (figure 1.5) (Longya et al., 2019). Longya et al. investigated the presence or absence of AVR-Pik in 58 Thai isolates and found that AVR-PikD was by far the most prevalent, with AVR-PikA, AVR-PikE and AVR-PikF also identified at lower frequencies. AVR-PikF has since been reported in Chinese *M. oryzae* isolates (Li et al., 2019a).

The Pik alleles have different recognition specificities for the different AVR-Pik alleles. For example, the Pikp allele, from the K60 rice variety, recognises only the AVR-PikD allele, while the Pikm allele, from the Tsuyuake rice variety, recognises AVR-PikD, AVR-PikE and AVR- PikA. The recognition specificities of the different Pik alleles, where known, are summarised in table 1.2. Crucially, none of the Pik alleles recognise AVR-PikC or AVR-PikF. *Magnaporthe oryzae* isolates carrying either of these alleles alone are virulent on rice containing any of the characterised Pik alleles.

Interestingly, Longya et al. found that over half of *M. oryzae* isolates containing AVR-

Table 1.1 Summary of the Pik alleles that have been described to date.

Pik allele	Rice variety	Reference	Genbank accession
Pik*	Kusabue	(Zhai et al., 2011)	HM048900
Pik*-KA	Kanto51	(Ashikawa et al., 2012)	AB616658
Pikp	K60	(Yuan et al., 2011)	HM035360
Pikm	Tsuyuake	(Ashikawa et al., 2008)	AB462256
Pikh	K3	(Zhai et al., 2014)	HQ662330
Piks	Shin 2	-	AET36547
Pike	Xiangzao143	(Chen et al., 2015)	-
Pi1	LAC23	(Hua et al., 2012)	HQ606329

Pik carry two AVR-Pik alleles, specifically AVR-PikD and AVR-PikF (Longya et al., 2019). Sequence data from Li et al. also hinted at the presence of two AVR-Pik alleles in some isolates (Li et al., 2019a), however they did not investigate this further. The presence of multiple AVR-Pik alleles may represent an important transition stage in the evolution of virulent *M. oryzae* strains carrying novel AVR-Pik alleles. *M. oryzae* isolates containing both AVR-PikD and AVR-PikF are avirulent on K60 ($Pikp^{+}$) rice plants, however future loss of AVR-PikD could lead to a highly virulent *M. oryzae* population capable of overcoming Pik-mediated resistance in the field (Longya et al., 2019).

A consistent finding of all studies that have investigated the incidence of AVR-Pik in rice blast isolates is that AVR-Pik is present in a majority of *M. oryzae* isolates surveyed and exhibits high allelic diversity (Longya et al., 2019, Huang et al., 2014, Kanzaki et al., 2012, Li et al., 2019a). This apparent selection for mutation rather than loss hints at an important function of AVR-Pik in blast infection.

1.5.3 Interaction of AVR-Pik with the integrated HMA domain of Pik-1 is necessary for recognition

A combination of *in vitro* and *in vivo* approaches previously demonstrated that AVR-PikD directly interacts with the integrated HMA domain of Pikp-1 (Maqbool et al., 2015). Mutations in AVR-PikD that prevent the interaction with Pikp-HMA were

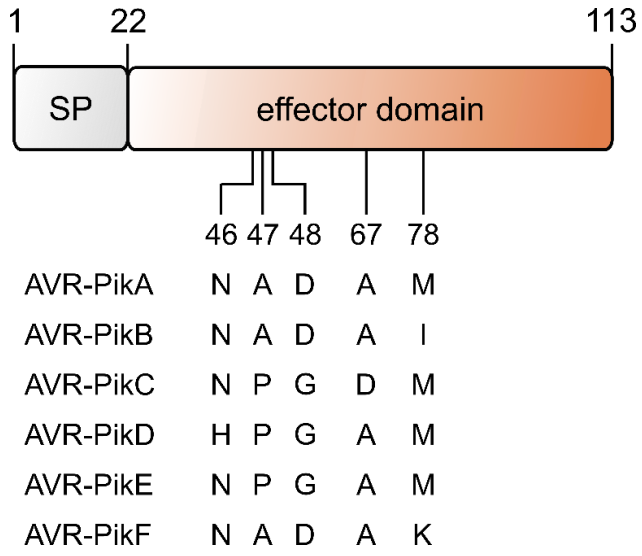


Figure 1.5 Alleles of the AVR-Pik effector protein.

Schematic representation of the AVR-Pik effector protein indicating polymorphic amino acid residues which vary between the six different AVR-Pik alleles identified to date (Yoshida et al., 2009, Kanzaki et al., 2012, Longya et al., 2019). Single letter codes are used to denote amino acid identities. SP = signal peptide.

Table 1.2 Different Pik alleles have different recognition specificities for AVR-Pik alleles.

R indicates that rice containing a particular Pik allele is resistant to M. oryzae *carrying the corresponding AVR-Pik allele, S indicates that the rice containing the Pik allele is susceptible. The table was constructed using data from (Kanzaki et al., 2012, Longya et al., 2019, Li et al., 2019a).*

		Pik allele				
		Pikp	**Piks**	**Pik*-KA**	**Pikm**	**Pikh**
AVR-Pik allele	**AVR-PikA**	S	S	S	R	R
	AVR-PikC	S	S	S	S	S
	AVR-PikD	R	R	R	R	R
	AVR-PikE	S	S	R	R	R
	AVR-PikF	S	S	S	S	S

identified, and transgenic *M. oryzae* carrying these mutated AVR-PikD alleles were fully susceptible on K60 ($Pikp^+$) rice lines (Maqbool et al., 2015), indicating that the interaction between the effector and integrated HMA domain is necessary for activation of Pik-mediated resistance.

As described in section 1.5.2, while the Pikp allele only recognises AVR-PikD, the Pikm allele recognises AVR-PikD, AVR-PikE and AVR-PikA. The difference in recognition specificity between these two alleles can be attributed to differences in binding of the effector alleles to the HMA domains of Pikp-1 and Pikm-1. Pikm-HMA binds AVR-PikD, AVR-PikE and AVR-PikA with high affinity, while Pikp-HMA binds AVR-PikD with high affinity, and weakly interacts with AVR-PikE (De la Concepcion et al., 2018). By solving the crystal structures of Pikm-HMA in complex with AVR-PikD, AVR-PikE and AVR-PikA, and comparing these to the structure of Pikp-HMA in complex with AVR-PikD, structural differences at the binding interfaces were identified which explain the contrasting effector interaction profiles of the two HMA domains (De la Concepcion et al., 2018).

Additionally, the crystal structures of the complexes between AVR-Pik alleles and the HMA domains of Pik-1 alleles were used to identify two amino acids in the HMA domain (Lys262 and Glu263) which are important for the extended recognition specificity of Pikm (De la Concepcion et al., 2019). The corresponding amino acids in Pikp-1 (Asn and Lys, respectively) were replaced with these two amino acids from Pikm-1, to generate the double mutant $Pikp\text{-}1^{NK\text{-}KE}$. These two mutations in Pikp-1 were sufficient to extend high affinity binding to AVR-PikE and AVR-PikA *in vitro* (De la Concepcion et al., 2019).

1.6 AVR-Pik interacts with members of a family of proteins containing a heavy metal associated domain

1.6.1 AVR-Pik interacts with HMA domain-containing proteins

Collaborators at Iwate Biotechnology Research Center (Japan) carried out a yeast two-hybrid screen with a rice cDNA library to identify putative host targets of AVR-PikD. AVR-PikD was cloned into a bait vector, and the cDNA library was cloned into prey

vectors. The screen identified four proteins as interacting with AVR-PikD, and the interactions were subsequently confirmed by pairwise yeast two-hybrid experiments (Ryohei Terauchi, personal communication).

Each of the proteins interacting with AVR-PikD contained a heavy metal associated domain, and, based on their size, they were referred to as small heavy metal associated domain containing (sHMA) proteins. BLAST searches revealed nearly 100 sHMA proteins in the rice genome. Additional pairwise yeast two-hybrid experiments were performed with a selection of these sHMA proteins. The interaction with AVR-PikD was found to be specific to a subset of sHMA proteins.

1.6.2 HIPPs and HPPs – two families of HMA-domain containing proteins in vascular plants

Over 100 sHMA proteins have now been identified in rice, and these can be divided into two broad families: the heavy metal-associated isoprenylated plant proteins (HIPPs), which contain a C-terminal CααX (where α represents an aliphatic amino acid, and X represents any amino acid) isoprenylation motif in addition to one or more HMA domains, and the heavy metal associated plant proteins (HPPs), which lack the isoprenylation motif (de Abreu-Neto et al., 2013). The subset of sHMA proteins with which AVR-PikD interacts includes members of both families.

1.6.3 HIPPs and HPPs are involved in plant responses to abiotic and biotic stresses

HIPPs and HPPs have been implicated in heavy metal homeostasis (de Abreu-Neto et al., 2013, Tehseen et al., 2010) and responses to various abiotic stresses including cold stress and drought stress (de Abreu-Neto et al., 2013, Barth et al., 2009). Furthermore, several HIPPs have been shown to have a role in plant responses to biotic stresses. AtHIPP3 has been identified as a zinc-binding, nuclear-localised negative regulator of the salicylate-dependent plant immunity pathway (Zschiesche et al., 2015). OsHIPP05 (Pi21) is a rice blast susceptibility gene (Fukuoka et al., 2009). A deletion from a proline-rich region of the protein, including the consensus PXXPXXP motif which may have a role in protein-protein interactions, results in a loss-of-function allele which confers resistance to rice blast. OsHIPP05 therefore appears to be a susceptibility factor in rice

blast disease. Other HIPPs have also been shown to be susceptibility factors for various pests and pathogens. AtHIPP27 is a susceptibility gene for beet cyst nematode, with loss of function mutants displaying reduced susceptibility to the pest (Radakovic et al., 2018). NbHIPP26 is targeted by the potato mop-top virus movement protein to facilitate the spread of the virus and knock down of NbHIPP26 inhibited long-distance movement of the virus (Cowan et al., 2018).

The significance of the interaction between AVR-Pik and HIPPs/HPPs in rice blast disease remains unknown. Work by our collaborators to characterise the mechanism of action of AVR-Pik in rice is ongoing.

1.7 Aims and objectives

Magnaporthe oryzae poses a significant threat to rice farmers across the world, and there is a clear need to develop new solutions against rice blast disease. Engineering NLR proteins to recognise new effectors could offer new sources of resistance to *M. oryzae*. Due to their similarity to effector targets, integrated domains have considerable potential as foci for engineering efforts. Broadly, this work aims to understand how *M. oryzae* effector proteins interact with a putative host target, which shares similarity with an integrated domain in a characterised NLR protein, and then utilise this information to engineer rice with enhanced blast resistance.

In Chapter 3, I characterise the interaction between AVR-Pik and the HMA domain of OsHIPP19 using biochemical and structural approaches and compare this with the previously studied interactions between AVR-Pik and the integrated HMA domains of Pik-1. Chapter 4 explores the interaction between a second *Magnaporthe oryzae* MAX effector, AVR-Pia, and the HMA domain of OsHIPP19. In chapters 5 and 6, I describe two novel approaches to engineer Pik-1 variants capable of triggering plant defences in response to the currently unrecognised effector alleles AVR-PikC and AVR-PikF. Both engineering strategies were informed by the characterisation of the interaction between AVR-Pik and OsHIPP19-HMA. Finally, chapter 7 outlines the progress that has been made towards stable transgenic rice and barley plants expressing the engineered Pik-1 variants.

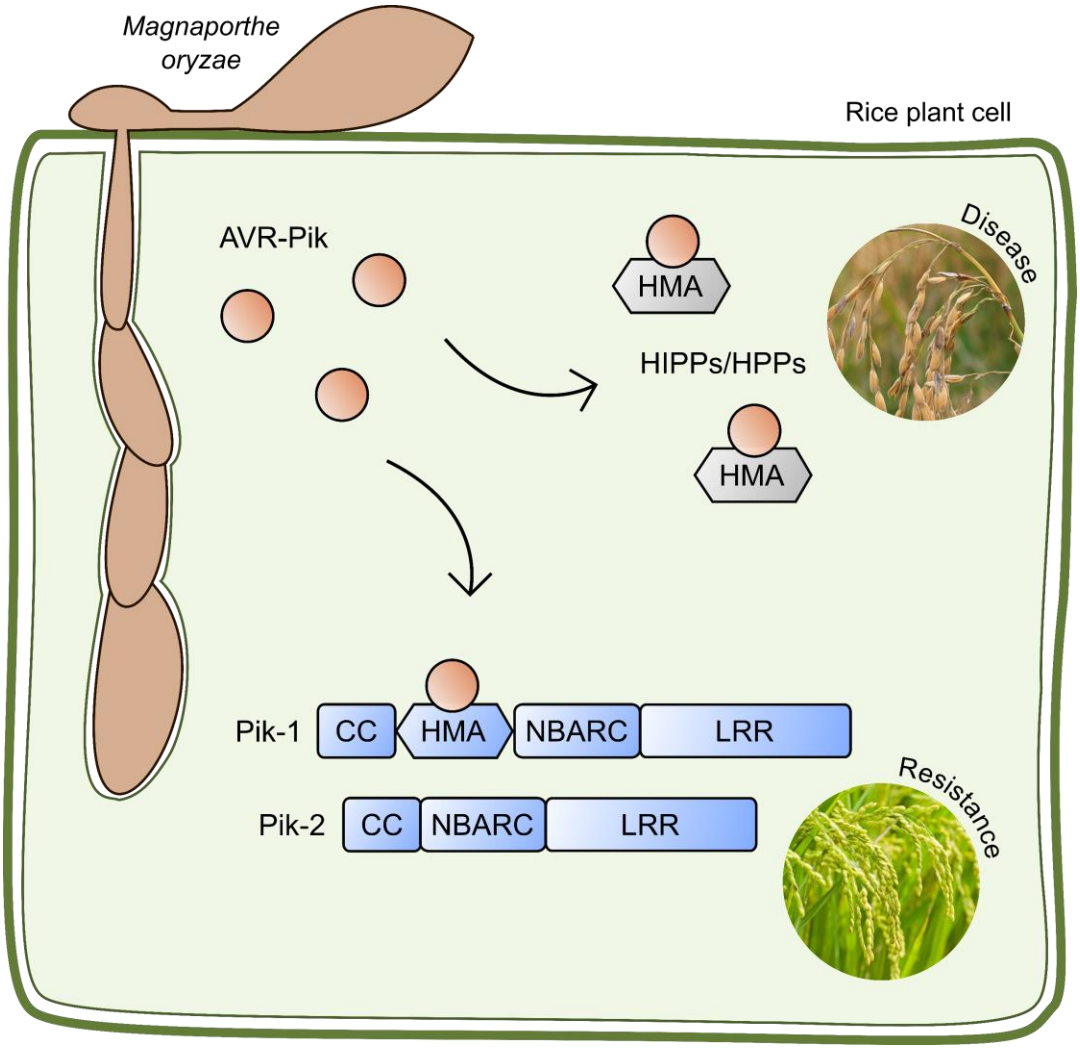


Figure 1.6 Overview of the Pik/AVR-Pik/HIPP/HPP system.

Magnaporthe oryzae *secretes the effector AVR-Pik (orange circle) into the host cell cytoplasm. AVR-Pik interacts with the HMA domain of OsHIPP19 (grey) to presumably promote virulence of the pathogen, contributing to disease. The paired NLR proteins Pik-1 and Pik-2 (blue) mediate recognition of AVR-Pik; direct binding of the effector to the HMA domain of Pik-1 activates plant immunity.*

2

Materials and methods

2

Materials and methods

2.1 General chemicals, reagents and buffers

2.1.1 Chemicals

Chemicals were supplied by Sigma-Aldrich (now Merck), Melford, ForMedium™, Thermo Fisher Scientific or VWR (Avantor) unless specified otherwise.

2.1.2 Antibiotics

1000X stocks of carbenicillin, spectinomycin, kanamycin and gentamycin were prepared by dissolving into deionised water. A 1000X stock of rifampicin was prepared by dissolving into 100% DMSO, and a 1000X stock of chloramphenicol was prepared by dissolving into ethanol. All solutions were filter-sterilised using Minisart® 0.22µm filters (Sartorius). Carbenicillin, kanamycin and gentamycin stocks were stored at 4°C and spectinomycin, chloramphenicol and rifampicin stocks at -20°C.

The 1000X stocks were used to supplement both liquid and solid media to the following final concentrations; carbenicillin (100 µg/ml), kanamycin (30 µg/ml), gentamycin (50 µg/ml), rifampicin (50 µg/ml), tetracycline (5 µg/ml), spectinomycin (100 µg/ml), chloramphenicol (34 µg/ml).

2.2 Bacterial methods

2.2.1 Bacterial strains

Table 2.1 describes the bacterial strains used in this project. The *Escherichia coli* strains DH5α, STELLAR™ and OneShot™ Top10 were used for cloning. *E. coli* SHuffle® T7 Express was used for production of recombinant proteins. *Agrobacterium tumefaciens*

Table 2.1 Bacterial strains used in this research project.

	Strain	Genotype	Source / Reference
Escherichia coli	DH5α	F- Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 λ-	ThermoFisher Scientific
	STELLAR™	F–, ara,Δ(lac-proAB) [Φ80d lacZΔM15], rpsL(str), thi, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, dam, Dcm	Clontech
	OneShot™ Top10	F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG	ThermoFisher Scientific
	SHuffle® T7 Express	fhuA2 lacZ::T7 gene1 [lon] ompT ahpC gal λatt::pNEB3-r1-cDsbC (SpecR, lacIq) ΔtrxB sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10 --TetS) endA1 Δgor Δ(mcrC-mrr)114::IS10	New England Biolabs (Lobstein et al., 2012)
Agrobacterium tumefaciens	GV3101	C58 (rif R) Ti pMP90 (pTiC58DT-DNA) (gent R) Nopaline (pSoup tet R)	(Van Larebeke et al., 1974)
	AGL1	C58 RecA (rif R/carbR) Ti pTiBo542DT-DNA Succinamopine	(Jin et al., 1987)

GV3101 was used for transient expression of genes of interest in *Nicotiana benthamiana*, while *A. tumefaciens* AGL1 was used for stable transformation of rice and barley plants.

2.2.2 Media for bacterial growth

All media were prepared using deionised water and sterilised by autoclaving. Following sterilisation, liquid media were cooled to room temperature, and media containing agar to ~50 °C, before the addition of temperature-sensitive supplements. All supplements were filter-sterilised using Minisart® 0.22µm filters (Sartorius).

2.2.2.1 Lysogeny broth

Lysogeny broth (LB, also known as Luria-Bertani) medium consisted of 10 g/L tryptone, 5 g/L yeast extract and 10 g/L sodium chloride. The final pH of the medium at 25 °C was 7 ± 0.2.

2.2.2.2 Auto-induction medium

Auto-induction medium (AIM) (Studier, 2005) consisted of 10 g/L tryptone, 5 g/L yeast extract, 3.3 g/L ammonium sulphate, 6.8 g/L monobasic potassium phosphate, 7.1 g/L disodium phosphate, 0.5 g/L glucose, 2 g/L α-lactose, 0.15 g/L magnesium sulphate and 0.03 g/L trace elements.

2.2.2.3 Super Optimal broth with Catabolite repression (SOC)

Super Optimal broth with Catabolite repression (SOC) consisted of 20 g/L tryptone, 5 g/L yeast extract, 0.58 g/L sodium chloride, 0.186 g/L potassium chloride, 2.03 g/L magnesium chloride, 2.46 g/L magnesium sulphate and 3.6 g/L glucose.

2.2.2.4 Lysogeny broth agar plates

Lysogeny broth (prepared as described in section 2.2.2.1) was supplemented with 1.1% (w/v) agar prior to autoclaving. For blue-white selection, isopropyl β-D-1-thiogalactopyranoside (IPTG; final concentration of 1mM) and X-gal (final concentration of 200 μg/ml) were added to the LB agar.

2.2.3 Preparation of chemically competent *E. coli*

Chemically competent *E. coli* cells were prepared using a modified version of the protocol described by Hanahan (Hanahan, 1983). A glycerol stock of competent *E. coli* cells was used to inoculate 10 ml LB medium. The culture was incubated at 37 °C for ≥ 12 hours with agitation (200-250 rpm), and subsequently used to inoculate 100 ml LB medium. The culture was incubated as before until the OD_{600} (optical density measured at a wavelength of 600 nm) of the culture reached 0.3. The cells were incubated on ice for 5 minutes and pelleted by centrifugation (7 minutes at 2000 x *g*). Pelleted cells were resuspended in 40 ml filter-sterilised TfbI buffer (30 mM potassium acetate, 10 mM calcium chloride, 100 mM rubidium chloride, 50 mM manganese chloride, 15% (v/v)

glycerol, pH 5.8) to a homogeneous suspension, incubated on ice for 5 minutes, then pelleted by centrifugation (7 minutes at 2000 x *g*). Pelleted cells were resuspended to a homogeneous suspension in 4 ml filter-sterilised TfbII buffer (10 mM MOPS, 75 mM calcium chloride, 10 mM rubidium chloride, 15% (v/v) glycerol, pH 6.5). Resuspended cells were incubated on ice for ≥ 15 minutes, and then aliquoted into prechilled, sterile Eppendorf tubes. These were flash-frozen in liquid nitrogen and stored at -70 °C.

2.2.4 Transformation of chemically competent *E. coli*

For transformation, chemically competent *E. coli* cells were thawed on ice. Immediately upon thawing, plasmid DNA was added and the cells incubated on ice for a further 30 minutes. The cells were then subjected to a heat shock (1 minute at 42 °C) before being returned to ice for 2 minutes. 500 µl ice-cold SOC medium was added to the cells, which were then recovered by incubation at 37 °C with agitation (200-250 rpm) for 1 hour. Cells were then spread on LB agar plates containing appropriate supplements. Plates were transferred to a static incubator at 37°C for ≥16 hours, until single colonies could be observed on the plate.

2.2.5 Preparation of electrocompetent *A. tumefaciens*

For preparation of electrocompetent *A. tumefaciens*, a glycerol stock of competent cells was used to inoculate 5 ml LB medium supplemented with tetracycline. The culture was incubated at 28 °C for 24 hours with agitation (200-250 rpm) and diluted (1:200) into 200 ml LB medium supplemented with tetracycline. The culture was incubated as before until the OD_{600} reached 0.5, transferred to prechilled centrifuge tubes and incubated on ice for >30 minutes. Cells were pelleted by centrifugation (15 minutes at 2000 x *g* at 4 °C). Pelleted cells were resuspended in 50 ml ice-cold 10 % glycerol, and centrifuged (15 minutes at 2000 x *g* at 4 °C). Pelleted cells were resuspended in 200 µl ice-cold 10 % glycerol and aliquoted into prechilled, sterile Eppendorf tubes. These were flash-frozen in liquid nitrogen and stored at -70 °C.

2.2.6 Transformation of electrocompetent *A. tumefaciens*

For transformation, electrocompetent *A. tumefaciens* cells were thawed on ice. Immediately upon thawing, plasmid DNA was added, and the mixture transferred to a

prechilled 2 mm electroporation cuvette (Geneflow / Cell Projects). A 2500 V pulse was supplied using an Eppendorf® Electroporator 2510. 500 µl ice-cold SOC medium was added to the electroporated cells, which were then recovered by incubation at 28 °C with agitation (200-250 rpm) for 1 hour. Cells were spread on LB agar plates containing appropriate antibiotics. Plates were transferred to a static incubator at 28 °C for ≥36 hours until single colonies were observed on the plate.

2.2.7 Preparation of glycerol stocks

Glycerol stocks were prepared by combining 750 µl bacterial LB culture (grown for ≥16 hours at 37 °C for *E. coli*, or for ≥36 hours at 28 °C for *A. tumefaciens*) with 750 µl sterile 100 % glycerol (final concentration of 50 % (v/v) glycerol) in a sterile 2 ml tube. Glycerol stocks were stored at -70 °C.

2.3 DNA methods

2.3.1 Isolation of plasmid DNA from *E. coli*

Following transformation, a single colony was used to inoculate 10 ml LB medium supplemented with the appropriate antibiotic. After a minimum of 8 hours at 37 °C with agitation (200-250 rpm), cells were pelleted by centrifugation (4,000 x *g* for 10 minutes). Plasmid DNA was extracted from the pelleted cells using either the ISOLATE II Plasmid Mini kit (Bioline), or the Nucleospin® Plasmid kit (Macherey-Nagel), according to the manufacturer's instructions. Plasmid DNA was eluted from the spin column using 2.5 mM Tris-HCl, pH 8.0, which was pre-heated to 50 °C prior to application to the column.

Golden Gate assembly of large constructs or constructs containing multiple components requires higher quality plasmid DNA than can be obtained with the ISOLATE II Plasmid Mini or Nucleospin® Plasmid kits. A single colony was used to inoculate 50 ml LB medium supplemented with the appropriate antibiotic. Cells were pelleted after a minimum of 16 hours at 37 °C with agitation (200 rpm). Plasmid DNA was extracted using the PureYield™ Plasmid Midiprep System (ProMega) according to the manufacturer's instructions.

The concentration of plasmid DNA was determined using a NanoVue (GE Healthcare). The A_{260}/A_{280} ratio was measured to assess the purity of the sample (an A_{260}/A_{280} ratio of ≥1.8 is generally considered to be acceptable). Following isolation, plasmid DNA was stored at -20 °C.

2.3.2 Polymerase chain reaction

2.3.2.1 Synthesis of DNA oligonucleotide primers

DNA oligonucleotide primers were synthesised by Integrated DNA Technologies. Oligonucleotides were supplied in a lyophilised form and resuspended in dH_2O to a concentration of 100 µM. 10 µM working stocks were prepared by 10X dilution in dH_2O. Both 10 µM and 100 µM stocks were stored at -20 °C. All oligonucleotide primers used in this work are listed in the appendix.

2.3.2.2 Amplification of DNA fragments for cloning

VELOCITY DNA polymerase (Bioline) was used to amplify DNA fragments for subsequent use in cloning reactions. VELOCITY is a high-fidelity polymerase with 3' – 5' proofreading exonuclease activity. The components and conditions of the reaction are described in table 2.2 and table 2.3, respectively. PCR was carried out in a T3000 Thermocycler (Biometra).

Mutations in a template sequence were either made by incorporating the mutation in one of the primers or generated by 3-primer PCR. This method incorporates an internal primer (15-25 bases) carrying the desired mutation. 3 µl of the third 10 µM primer was included in the 50 µl reaction mixture, and the PCR carried out as described above using VELOCITY DNA polymerase.

2.3.2.3 Colony PCR

Colony PCR was used to investigate the presence of a construct of interest in bacterial colonies following transformation. Each single, isolated colony was lightly touched with a sterile pipette tip and transferred to a PCR tube containing the reaction mixture.

MyTaq™ DNA polymerase (Bioline) was used for colony PCR as the enzyme has a rapid amplification rate and is supplied with a premixed buffer containing dNTPs, $MgCl_2$, and

Table 2.2 Reaction components for VELOCITY PCR.

Component	Volume (µl)
5X HiFi buffer	10
dNTPs (10 mM)	1
Forward primer (10 µM)	3
Reverse primer (10 µM)	3
VELOCITY DNA polymerase	1
Template DNA	Variable
dH_2O	31
Total reaction volume	**50**

Table 2.3 Thermocycling conditions for VELOCITY PCR.

PCR stage	Temperature (°C)	Time (s)	
Initial denaturation	98	120	
Denaturation	98	30	35 cycles
Annealing	(Primer T_a) - 5	30	
Extension	72	30	
Final extension	72	600	

a gel loading dye. The reaction components and conditions are described in tables 2.4 and 2.5, respectively. PCR was carried out in a T3000 Thermocycler (Biometra).

2.3.2.4 PCR to confirm successful transformation of barley plants

Following extraction of genomic DNA from transgenic barley plants, PCR was carried out with primers specific to the hygromycin resistance cassette to test for successful transformation of the plants. MyTaq™ polymerase (Bioline) was used due to the large number of samples being tested. The reaction mixture and conditions were as described above.

Table 2.4 Reaction components for MyTaq PCR.

Component	Volume (µl)
5x MyTaq reaction buffer	4
Forward primer (10µM)	3
Reverse primer (10µM)	3
MyTaq DNA polymerase	0.5
Template	Variable
dH_2O	9.5
Total reaction volume	**20**

Table 2.5 Thermocycling conditions for MyTaq PCR.

PCR stage	Temperature (°C)	Time (s)	
Initial denaturation	95	60	
Denaturation	95	15	30 cycles
Annealing	(Primer T_a) - 5	15	
Extension	72	30*	

* 90 s for DNA fragments > 1 kb

2.3.3 Agarose gel electrophoresis

Agarose gels were prepared by dissolving 1 % (w/v) molecular biology grade agarose in TAE buffer (40 mM Tris-acetate pH 8.0, 1.0 mM EDTA) by pulse-heating in a microwave. The agarose solution was cooled to approximately 50 °C and either ethidium bromide (final concentration of 0.5 µg/ml) or Midori^{Green} Advance DNA stain (Nippon Genetics Europe; 1 µl stain per 25 ml 1 % agarose) was added to enable visualisation of DNA under UV light. The molten agarose was poured into moulds containing well-combs and allowed to cool to room temperature to set the gel.

DNA samples were combined with 4X FOG loading dye (12 % (w/v) Ficoll 400 and 0.25 % (w/v) Orange G). The MyTaq Red PCR buffer contains an agarose gel loading dye;

no FOG loading dye was added to DNA samples in this buffer. Samples were loaded onto the gel alongside a molecular weight marker. The molecular weight markers used in this project are shown in table 2.6.

Gel electrophoresis was carried out at a constant voltage of 80-100 V in TAE buffer. DNA bands were visualised using an ultraviolet (UV) transilluminator.

2.3.4 Purification of DNA from an agarose gel

Following electrophoresis, DNA bands of interest were excised from the agarose gel using a razor blade. Purification of DNA from the gel was carried out using a NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel) according to the manufacturer's instructions.

2.3.5 Synthesis of gene fragments

The HMA domain of OsHIPP19 containing seven amino acids from Pikp-1 (described in chapter 6) was commercially synthesised by Integrated DNA Technologies as a gBlocks® Gene Fragment. The fragment was designed to enable direct cloning into the pCR8 vector to produce a level 0 plasmid that could subsequently be used to generate full-length Pik-1 in a level 1 vector. Synthesised DNA was supplied in a lyophilised form and was resuspended in 2.5 mM Tris-HCl pH 8.5 to a final concentration of 20 ng/µl.

Table 2.6 DNA molecular weight markers used in this project.

Molecular weight marker	Manufacturer
1 kb Plus ladder	Invitrogen
1 kb Plus ladder (formerly called the 2-log DNA ladder)	New England Biolabs
Penn State 1 kb ladder	(Henrici et al., 2017)
Penn State 100 bp ladder	(Henrici et al., 2017)

2.3.6 Cloning

2.3.6.1 In-Fusion™ cloning

In-Fusion™ cloning was used to clone a gene of interest into a pOPIN vector (Berrow et al., 2007). Developed by the Oxford Protein Production Facility (OPPF), the suite of pOPIN expression vectors enable proteins to be produced in various expression hosts, with a range of different affinity and solubility tags.

Prior to cloning, the pOPIN vector is linearised by double-digestion with specific restriction enzymes (table 2.7). A PCR product, containing the sequence of interest flanked by 15 base pair adaptor sequences, is generated and purified. For In-Fusion cloning, 1 µl 5X InFusion HD enzyme premix was combined with 50-100 ng of linearised pOPIN vector and 10-50 ng of purified PCR fragment. The reaction mixture was made up to a final volume of 5 µl with dH_2O and incubated at 42 °C for 30 minutes. Following incubation, the reaction mixture was used to transform *E. coli* STELLAR cells as described in section 2.4.4.

2.3.6.2 TOPO® cloning

The pCR™8/GW/TOPO® vector was used as a level 0 acceptor for Pik-1 domain constructs. While *BsaI* digestion of standard level 0 acceptors produces overhangs specific for the part type, using the pCR™8 vector allows production of custom

Table 2.7 pOPIN expression vectors used to produce recombinant proteins in E. coli*.*

pOPIN vector	Restriction enzymes for linearisation	Antibiotic resistance	Copy number	Protein product
pOPIN-M	KpnI / HindIII	Carbenicillin	High	6xHis:MBP^3C^POI
pOPIN-S3C	KpnI / HindIII	Carbenicillin	High	6xHis:SUMO^3C^POI
pOPIN-E	NcoI / PmeI	Carbenicillin	High	POI:6xHis*
pOPIN-A	NcoI / DraI	Kanamycin	Low	POI*

*Solubility tags can be added to the N-terminus by subcloning from pOPIN-M/pOPIN-S3C.

overhangs. In the case of the Pik-1 domain constructs, this allows multiple segments of the CDS to be assembled into a level 1 vector. Additionally, the backbone of the pCR™8 vector does not contain any *BsaI* restriction sites. It should be noted that the subsequent development of the universal level 0 acceptor pUAP1 allows production of custom overhangs within the standard Golden Gate modular cloning system.

DNA molecules were first modified by A-tailing, which adds an adenine to the 3' end of blunt-ended, double-stranded DNA. 6.8 µl DNA (20-50 ng/µl) was combined with 1 µl 10x ThermoPol® buffer (NEB), 2 µl 1mM dATP (final concentration of 0.2mM dATP) and 0.2 µl Taq DNA polymerase (NEB). The reaction mixture was incubated at 72 °C for 20 minutes.

4.5 µl A-tailed DNA was combined with 0.5 µl pCR™8 vector and 1 µl salt solution (1.2 M NaCl, 0.06 M $MgCl_2$). The pCR™8 vector and salt solution were supplied as part of the pCR™8/GW/TOPO® TA Cloning Kit (Thermo Fisher Scientific). The reaction mixture was incubated at room temperature for 30 minutes, and used to transform competent OneShot™ TOP10 *E. coli* cells as described in section 2.2.4.

2.3.6.3 Golden Gate

Golden Gate cloning was used to generate constructs for transient expression in *Nicotiana benthamiana* and for stable transformation of rice and barley. This cloning method allows directional assembly of multiple DNA fragments in a single reaction, combining digestion by type IIS restriction enzymes with subsequent ligation by a T4 DNA ligase (Engler et al., 2014, Weber et al., 2011, Engler et al., 2008). Type IIS restriction enzymes recognise asymmetric DNA sequences and cleave outside their recognition site. Cleavage by the enzymes *BsaI* or *BpiI* results in a 4 base pair single-stranded overhang. By designing parts to have specific overhangs following digestion, DNA fragments can be assembled in a particular order through complementary overhangs.

Golden Gate cloning has been used by members of the laboratory to generate assorted constructs for experiments *in planta* (Maqbool et al., 2015, De la Concepcion et al., 2018), some of which were used in the experiments in chapter 5 and chapter 6. Members of the laboratory have also double-domesticated (removed endogenous *BpiI* or *BsaI* sites)

template DNA for Pikp-1, Pikp-2 and AVR-Pik, enabling mutations to be made with comparative ease.

The Golden Gate MoClo assembly standards (Weber et al., 2011, Engler et al., 2014) define Golden Gate plasmids as either level 0, level 1 or level 2 based on their contents. Level 0 modules contain genetic elements such as promoters, terminators, epitope tags or coding sequences. Level 1 constructs contain complete transcriptional units; level 1 acceptor vectors are binary vectors, so these constructs can be used for *A. tumefaciens*-mediated transient gene expression *in planta*. Level 2 constructs contain multiple transcriptional units, and can therefore be used for stable transformation, where a separate selectable marker cassette is required alongside the gene of interest. To assemble fragments into a level 0 or level 2 acceptor, *BpiI* is used. To assemble fragments into a level 1 acceptor, *BsaI* is used.

The MoClo assembly standard syntax enables efficient assembly of compatible parts and convenient exchange of parts between scientists. For example, following *BsaI* cleavage of any level 0 N-terminal tag module, the exposed overhangs are CCAT (5') and AATG (3'). For more complex assemblies, such as the assembly of Pikp-1 from multiple level 0 modules each encoding a Pikp-1 domain, custom overhangs were designed. Table 2.8 summarises the level 0 modules used in this project. For efficient and accurate assembly, overhangs should not share more than two matching bases.

When assembling multigene level 2 constructs, the order in which the level 1 transcriptional units are incorporated into the level 2 construct is determined by the level 1 acceptor. *BpiI* digestion of the level 1 acceptors exposes different overhangs which determine their position in the level 2 construct.

Table 2.9 lists all the acceptor vectors used in this project. pCR™8 is included, as it was used to generate level 0 modules where custom overhangs were required following *BsaI* digestion, however as described in section 2.3.6.2, these modules were generated by TOPO® with Gateway® cloning. Apart from pCR™8, all acceptor vectors were obtained from TSL SynBio.

Table 2.8 Level 0 modules used in Golden Gate cloning.

Level 0 part		5' overhang	3' overhang	Source
Promoter + 5' UTR (no N-terminal tag)	Mas	GGAG	AATG	TSL SynBio (pICH85281)
	35S	GGAG	AATG	TSL SynBio (pICH51277)
Promoter + 5' UTR (N-terminal tag)	Ubi10	GGAG	CCAT	TSL SynBio (pICSL13005)
N-terminal tag	4xMyc	CCAT	AATG	TSL SynBio (pICSL30009)
Coding sequence	Pikp-1_CC	AATG	CAGA	Marina Franceschetti
	Pikp-1_HMA*	CAGA	GACG	Marina Franceschetti**
	OsHIPP19-HMA	CAGA	GACG	-
	Pikp-1_NBARC*	GACG	CTTT	Marina Franceschetti
	Pikp-1_LRR	CTTT	TTCG	Marina Franceschetti
	Pikp-2_CC			-
	Pikp-2_NBARC*			-
	Pikp-2_LRR			-
	AVR-PikF	AATG	GCTT	-
C-terminal tag	HellFire (6xHis: 3xFLAG)	TTCG	GCTT	TSL SynBio (pICSL50001)
	6xHA	TTCG	GCTT	TSL SynBio (pICSL50009)
3' UTR + terminator	Mas	GCTT	CGCT	TSL SynBio (pICSL51277)
	35S	GCTT	CGCT	TSL SynBio (pICH41414)
Miscellaneous	Pikp-1_35Ster*			-
	Pikp-2_35Ster			-
	Native Pik promoter			Marina Franceschetti

*versions of these domains containing mutations used the same overhangs.

**Marina Franceschetti generated the level 0 module containing WT Pikp-HMA. I generated additional level 0 modules containing mutated versions of Pikp-HMA.

Digestion-ligation reaction mixtures were prepared as described in table 2.10 in a final volume of 20 µl. 100 ng acceptor plasmid was used for assembling level 1 constructs, while 200 ng acceptor plasmid was used for assembling level 2 constructs. High-quality plasmid DNA isolated from *E. coli* cells by midiprep was used for all level 2 assemblies and for most level 1 assemblies. Thermocycling conditions for the digestion-ligation reaction are shown in table 2.11.

Following a digestion-ligation reaction, 5 µl of the reaction mixture was used to transform chemically competent *E. coli* STELLAR™ cells as described in section 2.2.4.

2.3.7 DNA sequencing

Sequencing reactions were performed by LightRun™ sequencing (Eurofins Genomics / GATC Biotech). 400-500 ng of purified plasmid were combined with 2.5 µl of a 10 µM working stock of an appropriate sequencing primer and brought to a total reaction volume of 10 µl with dH_2O. Sequencing results were analysed with Serial Cloner and Chromas.

2.4 Recombinant protein production and purification

2.4.1 Protein production in *E. coli*

A glycerol stock of *E. coli* SHuffle cells transformed with the construct of interest was used to inoculate 100 ml LB medium supplemented with appropriate antibiotics. The culture was incubated overnight (16-20 hours) at 30 °C with agitation (200 rpm), and then subcultured into 4-8x 1 L of autoinduction media (AIM) (Studier, 2005), supplemented with appropriate antibiotics, in a 2 L Erlenmeyer flask. The volume of culture added was adjusted so that the initial OD_{600} of the 1 L AIM culture was 0.07.

To produce OsHIPP19-HMA, the 1 L cultures were incubated at 30 °C with agitation (200 rpm) for 18-22 hours. To produce all other proteins, the 1 L cultures were incubated at 30 °C with agitation (200 rpm) until an OD_{600} of 0.4-0.8 was reached. The cultures were then incubated for a further 14-16 hours at 18 °C with agitation (200 rpm).

The culture was transferred into Nalgene® 1000 ml centrifuge bottles, and the cells pelleted by centrifugation at 5,410 x *g* for 7 minutes. The cell pellets were resuspended

Table 2.9 Acceptor vectors used in Golden Gate cloning.

	Features	Cloning selection	Antibiotic resistance
pCR™8	Used for level 0 modules requiring custom overhangs	-	Spectinomycin
pUAP1	Universal level 0 acceptor	RFP	Chloramphenicol
pICH47732	Level 1 acceptor (position 1)	Blue-white	Carbenicillin
pICH47742	Level 1 acceptor (position 2)	Blue-white	Carbenicillin
pICH47751	Level 1 acceptor (position 3)	Blue-white	Carbenicillin
pICSL4723	Level 2 acceptor	Orange-white	Kanamycin

Table 2.10 Reaction components for a digestion-ligation reaction.

Component	Quantity
Acceptor plasmid	100-200 ng
Plasmids containing each module/part to be inserted	2:1 molar ratio of insert:acceptor
T4 DNA ligase buffer (NEB)	1.5 µl
Bovine Serum Albumin (1 mg/ml)	1.5µl
T4 DNA ligase (NEB)	200 units
BsaI or *BpiI* (NEB)	10 units

Table 2.11 Thermocycling conditions for a digestion-ligation reaction

Temperature (°C)	Time (s)	
37	20	
37	180	x 27 cycles
16	240	
50	300	
80	300	

by pipetting in ice-cold lysis buffer (50 mM Tris-HCl pH 8.0, 50 mM glycine, 5 % glycerol, 500 mM NaCl and 20 mM imidazole, supplemented with cOmplete™ EDTA-free Protease Inhibitor Cocktail). Resuspended cells were either stored at -70 °C or immediately lysed for protein extraction.

2.4.2 Cell lysis and isolation of soluble fraction

Resuspended cells were lysed by sonication using a Vibra-Cell™ sonicator (SONICS) with a single 13 mm probe. The sonicator was used at 40 % amplitude, with a 1 s pulses delivered at 4 s intervals for a total pulse time of 4-8 minutes, depending on the cell density of the resuspension. The resuspended cells were kept on ice for the duration of the sonication process. 1 µl of the resulting whole cell lysate was diluted in 9 µl lysis buffer and combined with 2.5 µl SDS loading buffer for analysis by SDS-PAGE.

The total cell lysate was transferred to pre-chilled 50 ml Nalgene® Oak Ridge centrifuge tubes, and clarified by centrifugation at 36250 x *g* for 30 minutes at 4 °C. The soluble fraction of the cell lysate (the supernatant) was transferred to a clean tube. 1 µl of soluble fraction was diluted in 9 µl lysis buffer and combined with 2.5 µl SDS loading buffer for analysis by SDS-PAGE.

2.4.3 Purification of tagged protein from soluble fraction

The tagged protein was purified from the soluble fraction by immobilised metal affinity chromatography (IMAC) and gel filtration, carried out in tandem with an ÄKTAxpress system at 4 °C. The soluble fraction of the cell lysate was injected onto a 5 ml HisTrap™ HP NTA column (GE Healthcare), pre-activated with Ni^{2+}. The 6xHis affinity tag binds to Ni^{2+}, enabling purification of 6xHis-tagged proteins. The column was then washed with 100 ml of binding/wash buffer (50 mM Tris-HCl pH 8.0, 50 mM glycine, 5 % glycerol, 500 mM NaCl and 20 mM imidazole). The small amount of imidazole in the binding/wash buffer reduces non-specific interactions between the Ni^{2+} and endogenous proteins containing adjacent histidine residues. Bound proteins were then eluted with 25 ml of elution buffer (50 mM Tris-HCl pH 8.0, 50 mM glycine, 500 mM NaCl, 500 mM imidazole, 5% (v/v) glycerol). The imidazole outcompetes the 6xHis tag for Ni^{2+} binding, and the displaced protein is thus eluted from the column. The eluate was directly applied to a Superdex™ 75 HiLoad™ 26/600 gel filtration column (GE Healthcare) pre-

equilibrated in gel filtration buffer (20 mM HEPES pH 7.5 and 150 mM NaCl). For purification of OsHIPP19-HMA, 1 mM TCEP was included in the gel filtration buffer. Gel filtration was carried out at a flow rate of 3.2 ml/min, and the eluate collected in 8 ml fractions. The fractions were analysed by SDS-PAGE to confirm which contained the protein of interest.

2.4.4 Cleavage of tag with 3C protease

Fractions from the gel filtration containing the tagged protein were pooled. To cleave the affinity/solubility tags from the protein of interest, recombinant human rhinovirus (HRV) 3C protease was added (10 µg 3C protease per mg of tagged protein) and incubated at 4 °C overnight.

2.4.5 Separation of solubility tag from protein of interest

The untagged protein of interest was separated from the tag by affinity chromatography. The HMA domains of OsHIPP19, Pikm-1, Pikp-1 and Pikp-1 variants were each expressed from pOPIN-M and therefore produced with a N-terminal 6xHis:MBP tag, which was cleaved by 3C protease. To separate the HMA domain from the 6xHis:MBP tag, and remove the (6xHis-tagged) 3C protease, the protein was manually injected onto a 5 ml HisTrap™ HP Ni^{2+}- NTA column (GE Healthcare) and a 5 ml MBPTrap™ HP dextrin sepharose column (GE Healthcare) connected in tandem and pre-equilibrated in binding/wash buffer (see section 2.4.3).

The first 8 ml of flow-through was described as flow-through 1, and represented the "dead" volume of the columns. The remaining flow-through was collected separately and referred to as flow-through 2. The columns were washed with 30 ml binding/wash buffer, and the wash-through collected in two 15 ml fractions (wash-through 1 and 2). The HisTrap™ and MBPTrap™ columns were then separated, and bound protein (the 6xHis:MBP tag, and 6xHis-tagged 3C protease) eluted from each of them with elution buffer (see section 2.4.3) and 0.5 M NaOH, respectively. Fractions were analysed by SDS-PAGE to confirm the presence of the untagged HMA protein in the flow-through 2 and wash-through fractions.

The AVR-Pik effector alleles and AVR-Pia were expressed from pOPIN-E and produced with a N-terminal SUMO tag, which was cleaved by addition of 3C protease, and a C-terminal non-cleavable 6xHis tag. To separate the SUMO tag from the 6xHis-tagged effector, the protein was manually injected onto a 5 ml HisTrap™ HP Ni^{2+}- NTA column (GE Healthcare) pre-equilibrated in binding/wash buffer. The column was washed with 15 ml binding/wash buffer, and the 6xHis-tagged effector (and 6xHis-tagged 3C protease) eluted with 15 ml of elution buffer. Fractions were analysed by SDS-PAGE to confirm the presence of the 6xHis-tagged effector in the eluate.

To further purify the protein of interest, and separate the 6xHis-tagged 3C protease from 6xHis-tagged effector protein, a second gel filtration stage was carried out with an ÄKTAxpress system at 4 °C. The protein was injected onto a Superdex™ 75 HiLoad™ 26/600 gel filtration column (GE Healthcare) pre-equilibrated in gel filtration buffer (20 mM HEPES pH 7.5 and 150 mM NaCl, with 1 mM TCEP for purification of OsHIPP19-HMA). As before, gel filtration was carried out at a flow rate of 3.2 ml/min, and the eluate collected in 8 ml fractions. Fractions were analysed with SDS-PAGE to confirm the presence of the protein of interest. Fractions containing the purified protein were pooled and concentrated by centrifugation (7500 x *g* at 4 °C) using 20 ml and 2 ml VivaSpin® centrifugal concentrators (Sartorius) with a molecular weight cut-off appropriate for the protein. The protein concentration was determined (see section 2.4.6), and if the purified protein was not immediately required, it was stored at -70 °C.

2.4.6 Measuring protein concentration

Protein concentrations were determined using a Direct Detect® Infrared Spectrometer (Merck). This technique detects amide bonds (Strug et al., 2014), and unlike other techniques for measuring protein concentration, does not depend on the presence of specific amino acids in the protein of interest.

The protein was diluted to fall within the measurable range of 0.2-1 mg/ml. The Direct Detect® Assay-free card contains four membrane positions; 2 µl buffer was applied to the first position, and 2 µl of protein was applied to each of the three remaining positions. The card was left at room temperature until the membrane had completely dried before

the protein concentration was measured. The protein sample was measured in triplicate, and the measurements averaged.

2.4.7 Intact mass spectrometry

Intact mass spectrometry was performed by Dr Gerhard Saalbach and Dr Carlo de Oliveira Martins at the John Innes Centre Proteomics Platform. Spectra were obtained using a Synapt G2-Si mass spectrometer coupled to an Acquity UPLC system (Waters).

2.4.8 Production of protein complexes for X-ray crystallography

2.4.8.1 OsHIPP19-HMA and AVR-PikF

Chemically competent *E. coli* SHuffle cells were co-transformed with MBP:OsHIPP19-HMA in pOPIN-A and AVR-PikF in pOPIN-S3C as described in section 2.2.4. These constructs produce MBP-tagged OsHIPP19-HMA, and 6xHis:SUMO-tagged AVR-PikF. A single, well-isolated colony from the LB agar plate was used to inoculate 100 ml LB medium supplemented with appropriate antibiotics (both carbenicillin and kanamycin, to select for cells containing both plasmids). The culture was incubated overnight (16-20 hours) at 30 °C with agitation (200 rpm), and then subcultured into 8x 1 L of autoinduction media (AIM) (Studier, 2005), as described in section 2.4.1. The 1 L cultures were incubated at 30 °C with agitation (200 rpm) for 18-22 hours.

Cells were pelleted by centrifugation, resuspended in lysis buffer and lysed as described in section 2.4.2. The soluble fraction was obtained from the total cell lysate by centrifugation at 36250 x *g* for 30 minutes at 4 °C. The tagged protein complex was purified from the soluble fraction by immobilised metal affinity chromatography (IMAC) and gel filtration, carried out as described in section 2.4.3. As MBP:OsHIPP19-HMA is not 6xHis-tagged, this protein will only bind to the HisTrap™ HP column if it is in complex with 6xHis:SUMO-tagged AVR-PikF. Cleavage of the affinity and solubility tags was carried out by addition of 3C protease, as described in section 2.4.4.

To separate the 6xHis:SUMO and MBP tags from the untagged protein complex, the sample was manually injected onto a 5 ml HisTrap™ HP Ni^{2+}- NTA column (GE Healthcare) and a 5 ml MBPTrap™ HP dextrin sepharose column (GE Healthcare) connected in tandem and pre-equilibrated in binding/wash buffer (see section 2.4.3).

Flow-through, wash-through and eluate were collected as described for the purification of HMA domains in section 2.4.5. The flow-through and wash-through fractions containing the protein complex were further purified by gel filtration, carried out with an ÄKTAxpress system at 4 °C. The protein was injected onto a Superdex™ 75 HiLoad™ 26/600 gel filtration column (GE Healthcare) pre-equilibrated in gel filtration buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP). Fractions were analysed with SDS-PAGE to confirm the presence of the proteins of interest. As OsHIPP19-HMA and AVR-PikF are too similar in size to separate by SDS-PAGE, the sample was sent for intact mass spectrometry (section 2.4.7) to confirm the presence of both proteins in the sample. Fractions containing the purified protein were pooled and concentrated by centrifugation (see section 2.4.5), then immediately used to set up crystallisation screens.

2.4.8.2 Pikp-HMA$^{SNK-EKE}$ and AVR-PikC

Chemically competent *E. coli* SHuffle cells were co-transformed with Pikp-HMA$^{SNK-EKE}$ in pOPIN-M and AVR-PikC in pOPIN-A as described in section 2.2.4. These constructs produce 6xHis:MBP-tagged Pikp-HMA$^{SNK-EKE}$, and untagged AVR-PikC. A single, well-isolated colony from the LB agar plate was used to inoculate 100 ml LB medium supplemented with appropriate antibiotics (both carbenicillin and kanamycin, to select for cells containing both plasmids). The culture was incubated overnight (16-20 hours) at 30 °C with agitation (200 rpm), and then subcultured into 8x 1 L of autoinduction media (AIM) (Studier, 2005), as described in section 2.4.1. The 1 L cultures were incubated at 30 °C with agitation (200 rpm) until an OD_{600} of 0.4-0.8 was reached, then for a further 14-16 hours at 18 °C with agitation (200 rpm).

Cells were pelleted by centrifugation, resuspended in lysis buffer and lysed as described in section 2.4.2. The soluble fraction was obtained and the tagged protein complex was purified from the soluble fraction by IMAC and gel filtration as described in section 2.4.3. As AVR-PikC is untagged, it can only bind to the HisTrap™ HP column if in complex with 6xHis:MBP:Pikp-HMA$^{SNK-EKE}$. Cleavage of the 6xHis:MBP tag from Pikp-HMA$^{SNK-EKE}$ was carried out by addition of 3C protease, as described in section 2.4.4.

The 6xHis:MBP tag was separated from the protein complex using tandem HisTrap™ HP Ni^{2+}- NTA column (GE Healthcare) and MBPTrap™ HP dextrin sepharose column (GE

Healthcare) as described for the purification of HMA domains in section 2.4.5. Fractions containing the protein complex were further purified by gel filtration, using a Superdex™ 75 HiLoad™ 26/600 gel filtration column (GE Healthcare) pre-equilibrated in gel filtration buffer (20 mM HEPES pH 7.5, 150 mM NaCl). Fractions containing the purified protein were pooled and concentrated by centrifugation (see section 2.4.5), then immediately used to set up crystallisation screens.

2.4.8.3 OsHIPP19-HMA and AVR-Pia

A co-expression strategy was initially attempted to produce a complex between OsHIPP19-HMA and AVR-Pia. Chemically competent *E. coli* SHuffle cells were co-transformed with MBP:OsHIPP19-HMA in pOPIN-A and AVR-Pia in pOPIN-S3C as described in section 2.2.4. However, when 1 L cultures of co-transformed *E. coli* SHuffle cells were incubated at 30 °C overnight, the yield of soluble 6xHis:SUMO:AVR-Pia was very low.

The two proteins were therefore produced and purified separately, as described in sections 2.4.1 to 2.4.5. Prior to the final gel filtration, the untagged proteins were combined and incubated on ice for >2 hours to allow complex formation. The complex was then purified by gel filtration carried out with an ÄKTAxpress system at 4 °C. The protein sample was injected onto a Superdex™ 75 HiLoad™ 26/600 gel filtration column (GE Healthcare) pre-equilibrated in gel filtration buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP). Fractions containing the purified protein were pooled and concentrated by centrifugation (see section 2.4.5), then immediately used to set up crystallisation screens.

2.5 SDS-PAGE

Two different methods were used for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Unless specified in the relevant figure legend, method 1 was used.

2.5.1 Method 1

Pre-cast RunBlue™ TEO-Tricine SDS Mini gels (Expedeon) were fitted into a compatible RunBlue™ gel running tank. 12 % or 16 % single-percentage gels and 4-20 % gradient gels were used according to the molecular weight(s) of the protein(s) to be resolved. The tank was filled with 1X RunBlue™ TEO-Tricine SDS running buffer. Protein samples were combined with RunBlue™ 4X LDS Sample Buffer supplemented with 25 mM DTT and heated at 95 °C for 5 minutes (unless specified otherwise). Denatured samples were loaded onto the gel alongside a molecular weight protein ladder (either RunBlue™ Prestained Molecular Weight marker (Expedeon), Precision Plus Protein™ Dual Xtra Standards (Bio-Rad) or PageRuler™ Plus prestained protein ladder (Thermo Fisher)). Gel electrophoresis was carried out at a constant voltage of 160-180 V (or 100 V if proteins were to be subsequently transferred for Western blotting) until the dye front was within 5 mm of the end of the gel. Proteins were either transferred to a PVDF membrane for Western blotting (see section 2.12) or visualised by incubation with InstantBlue™ Coomassie Protein Stain (Expedeon) at room temperature for >60 minutes. Stained gels were rinsed in deionised water and imaged using a GeneSys G:Box imaging system (Syngene).

2.5.2 Method 2

17 % single-percentage SDS-PAGE gels were made in the laboratory and stored at 4 °C for up to two weeks until needed. The resolving gel was prepared by combining 17 % (w/v) polyacrylamide in 375 mM Tris-HCl pH 8.8 with 0.1 % (w/v) SDS, with 0.1 % (w/v) ammonium persulphate (AMPS) and 0.04 % (v/v) N,N,N',N'-tetramethylethylenediamine (TEMED) added immediately prior to pouring into Mini-PROTEAN (Bio-Rad) 1.0 mm glass gel casting plates assembled in a Mini-PROTEAN Tetra System Casting Frame (Bio-Rad). To produce a level boundary between the resolving and stacking gels, a thin layer of water-saturated butanol was applied on top of the resolving gel during polymerisation. When the resolving gel had set, the butanol was removed. The stacking gel was prepared by combining 5 % w/v polyacrylamide, 63 mM Tris-HCl, pH 6.8 with 0.1 % w/v SDS, with 0.1 % w/v ammonium persulfate and 0.1 % v/v N,N,N',N'tetramethylethylenediamine added immediately prior to pouring on top of the resolving gel. Well combs were inserted into the stacking gel. Upon setting, the gels

were fitted into a Mini-PROTEAN Tetra Cell electrophoresis tank (Bio-Rad) and the combs removed. The tank was filled with 1X Tris-Tricine SDS running buffer (25 mM Tris-HCl, 250 mM NaCl, 0.1 % (w/v) SDS). Samples were combined with 4X 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) and heated at 95 °C for 5 minutes. Denatured samples were loaded onto the gel alongside the RunBlue™ Prestained Molecular Weight marker (Expedeon). Gel electrophoresis was carried out at a constant voltage of 160-180 V until the dye front was within 5 mm of the end of the gel. Proteins were visualised by incubation with InstantBlue™ Coomassie Protein Stain (Expedeon) at room temperature for >60 minutes. Stained gels were rinsed in deionised water and imaged using a GeneSys G:Box imaging system (Syngene).

2.6 Circular dichroism

2.6.1 Preparation of OsHIPP19-HMA sample for circular dichroism

Following purification, the OsHIPP19-HMA protein was in the buffer used for the final gel filtration stage (20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP). As both HEPES and chloride ions absorb strongly in the far UV (<200 nm), the purified OsHIPP19-HMA protein was exchanged into a phosphate buffer (20 mM potassium phosphate, pH 7.2) by repeated cycles of concentration and dilution. The protein was concentrated to a volume of approximately 200 µl using a VivaSpin® centrifugal concentrator (Sartorius), diluted to a volume of 2 ml in phosphate buffer, and the process repeated a minimum of 3 times. The sample was diluted in phosphate buffer to a final concentration of 0.2 mg/ml.

2.6.2 Obtaining circular dichroism spectra

Circular dichroism was carried out with a Chirascan™ Plus CD Spectrometer (Applied Photophysics). Prior to use, the machine was completely purged of oxygen using nitrogen (N_2) gas. Spectra were collected at wavelengths between 185 nm and 260 nm, in steps of 0.5 nm. The temperature of the sample compartment was set to 20 °C. Spectra were initially collected with an empty cuvette to check machine background. Spectra were then acquired for buffer only, and for the OsHIPP19-HMA protein sample. In each case, 3 spectra were acquired and averaged. The final trace for the protein sample was

obtained by subtracting the trace obtained for buffer only. The Chirascan™ software was used to convert the units from millidegrees (machine units) to mean residue molar ellipticity (MRME), which corrects the measurements to account for protein concentration, using the equation:

$$MRME = \frac{millidegrees \times MRW}{10 \times L \times C}$$

where:

$$MRW = \frac{molar\ mass}{(number\ of\ amino\ acids) - 1} = \frac{8322.8}{(77) - 1} = 109.5$$

$$C = protein\ concentration\ (mg.ml^{-1})$$

$$L = path\ length\ (cm)$$

2.6.3 Assignment of secondary structure features

The DichroWeb server (Whitmore and Wallace, 2004, Whitmore and Wallace, 2008) was used to assign secondary structure features using the CDSSTR analysis method. Four of the available reference protein sets (3, 4, 6 and 7) were compatible with spectra collected at wavelengths between 185 nm and 260 nm (Sreerama and Woody, 2000). Secondary structure features were assigned based on each of these four reference sets.

2.7 Analytical gel filtration

Experiments were conducted at 4 °C using a Superdex 75 10/300 GL column (GE Healthcare) equilibrated in running buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP). To investigate whether the effector and HMA domain form a complex, the two proteins were combined in a 1:1 molar ratio and incubated on ice for 1 hour prior to analysis. Each protein was also analysed alone, at a concentration equivalent to that present in the complex. For each experiment, 100 µl protein was injected at a flow rate of 0.5 ml/min, and 500 µl fractions were collected for analysis with SDS-PAGE.

2.8 Surface plasmon resonance

Surface plasmon resonance experiments were carried out using a Biacore™ T100 system (GE Healthcare Life Sciences). The experimental buffer was 20 mM HEPES pH 7.5, 860 mM NaCl, 0.1 % Tween®20. All experiments were conducted at an analysis temperature of 25 °C. Reagents and proteins were flowed over the sensor chip at a constant flow rate of 30 µl/min.

Flow cell (FC) 2 of a nitrilotriacetic acid (NTA) sensor chip (GE Healthcare Life Sciences) was activated with 30 µl of 0.5 mM $NiCl_2$. Following activation, 30 µl of 6xHis-tagged AVR-Pik effector was immobilised on FC2. The concentration of the immobilised effector was adjusted to achieve effector capture of ~250 response units (RU).

The HMA domain was flowed across both FC1 (the reference cell) and FC2 for an association time of 360 seconds. Experimental buffer was then flowed across both flow cells for a further 180 seconds to monitor dissociation of the HMA domain. To regenerate the chip, 30 µl 0.35 M ethylenediaminetetraacetic acid (EDTA) pH 8.0 was applied to both flow cells. To remove any residual EDTA, 15 µl buffer was flowed across both cells. The NTA sensor chip was then re-activated with $NiCl_2$ for the next experimental cycle. The background response from FC1 (non-specific binding of the HMA domain to the chip) was subtracted from the response from FC2.

For $\%R_{max}$ experiments, the binding response (R_{obs}) was measured immediately after the end of the injection of the analyte and expressed as a percentage of the theoretical maximum response (R_{max}), calculated as follows:

$$R_{max}\,(RU) = \frac{M_W(analyte)}{M_W(ligand)} \times \text{stoichiometry} \times \text{ligand capture (RU)}$$

For OsHIPP19-HMA and Pikm-HMA, the stoichiometry of HMA:effector binding was 1:1. For Pikp-HMA (and Pikp-HMA mutants) the binding stoichiometry was 2:1. For experiments involving OsHIPP19-HMA, the HMA domains were tested at 2 nM, 5 nM and 20 nM; otherwise the HMA domains were tested at 4 nM, 40 nM and 100 nM. 3 technical replicates were conducted for each concentration, and the experiment repeated 3 times. Boxplots were generated using the ggplot2 package (Wickham, 2016) implemented in R (R Core Development Team, 2018).

For analysis of the kinetics of the interaction between OsHIPP19-HMA and AVR-Pik effectors, the HMA domain was flowed over the chip at a wider range of concentrations between 0.25nM and 25nM. In addition to subtracting the background response from FC1, the trace obtained when buffer was flowed over the chip instead of the HMA domain was subtracted from each experimental trace. The Biacore™ evaluation software was used to fit a 1:1 binding model to the data. The R_{max} was fitted locally to reflect the regeneration of the chip and subsequent recapture of Ni^{2+} and effector between cycles. The resulting model estimates for the k_a and k_d values were used to calculate the reported K_D:

$$K_D(\mathrm{M}) = \frac{k_d\ (s^{-1})}{k_a\ (M^{-1}s^{-1})}$$

The experiment was repeated three times.

2.9 Crystallographic methods

2.9.1 Screening crystallisation conditions

In order to conduct X-ray diffraction studies, protein crystals, consisting of multiple, symmetry-related molecules, are required. While some proteins crystallise readily, some will only form crystals under very specific conditions, and others, particularly those with highly flexible structures, may not crystallise at all. Identifying suitable conditions for protein crystallisation can necessitate extensive screening, altering buffer system, pH, precipitant(s), protein concentration, and temperature, amongst other variables. Initial investigation of crystallisation conditions was carried out using commercial screens. Each screen consists of 96 different crystallisation conditions, designed to sample a range of crystallisation space. The commercial screens used in this project are listed in table 2.12.

Crystallisation experiments were conducted using the sitting drop vapour diffusion method (figure 2.1). A small drop containing purified protein and precipitant(s) is sealed in a chamber with a larger reservoir containing the precipitant at a higher concentration. Gradual diffusion of water vapour from the protein drop to the reservoir increases the concentration of precipitant and protein in the drop. Under ideal conditions, this

facilitates the formation of crystalline protein nuclei, which then grow to form large protein crystals suitable for X-ray diffraction studies.

Sitting drop vapour diffusion experiments were set up in MRC 2-drop 96-well crystallisation plates (Molecular Dimensions). A Rainin Liquidator™ 96 pipetting system (Mettler Toledo) was used to transfer 40 µl screen solution to the reservoir. Using either an Oryx Nano or Oryx8 crystallisation robot (Douglas Instruments), 0.3 µl purified protein was combined with 0.3 µl reservoir solution and dispensed as a sitting drop. The crystallisation plate has 2 wells for each reservoir, allowing 2 different protein concentrations to be trialled per screen. The wells were immediately sealed with a ClearVue (UV light-permeable) adhesive seal (Molecular Dimensions). The crystallisation plates were incubated at 20 °C, and periodically imaged under daylight and UV light by an automated Minstrel Crystallisation Imager (Rigaku).

Where crystals in commercial screens were not considered to be optimal for X-ray diffraction studies, optimisation screens were set up to sample crystallisation space around the condition(s) that gave the initial hits. Optimisation screens were designed using XSTEP software (Douglas Instruments), and the reservoir solutions prepared and dispensed using an Oryx8 crystallisation robot (Douglas Instruments). The sitting drops were then dispensed as for the commercial screens.

Table 2.12 Crystallisation screens used in this project.

Screen	Supplier
Morpheus	Molecular Dimensions
JCSG	Molecular Dimensions
Structure	Molecular Dimensions
KISS	Molecular Dimensions (designed by Clare Stevenson)
PEG	Qiagen
MIDAS	Molecular Dimensions
PACT	Molecular Dimensions

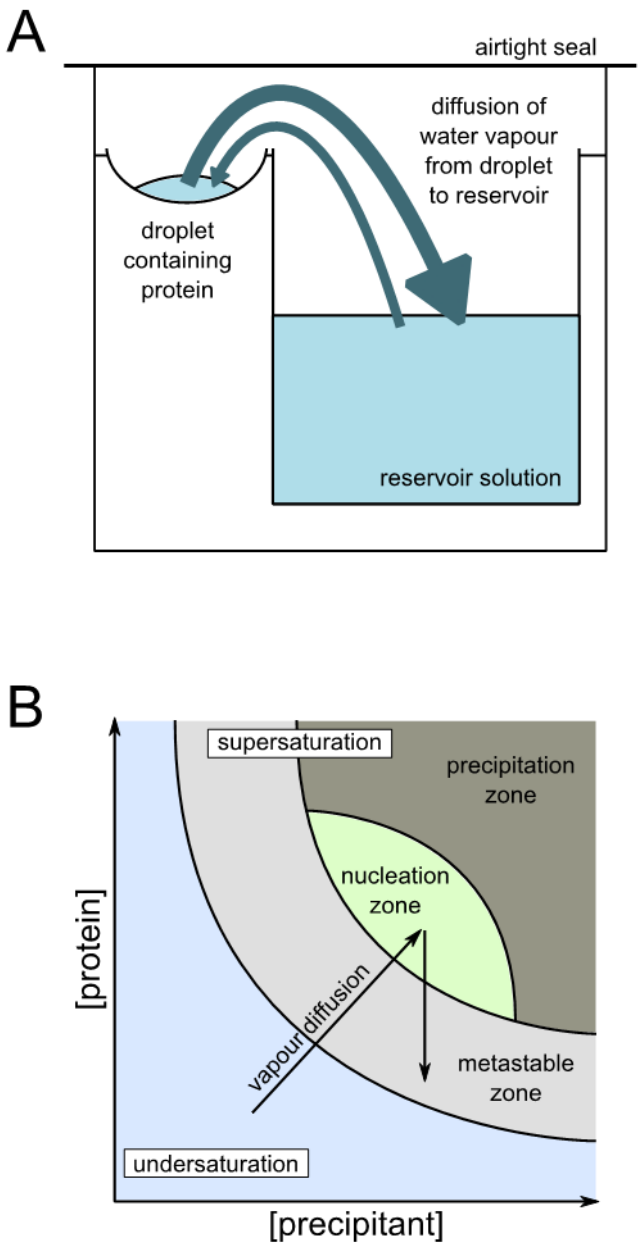


Figure 2.1 Protein crystallisation by vapour diffusion.

A. *Diagram illustrating the experimental setup for protein crystallisation by sitting drop vapour diffusion.* ***B.*** *Protein solubility phase diagram summarising crystallisation by sitting drop vapour diffusion under ideal conditions.*

2.9.2 Seeding

Seeding involves the addition of small protein crystal fragments (seeds) into the sitting drop, which can promote ordered growth of larger crystals during vapour diffusion experiments. Initial attempts to crystallise the complex between OsHIPP19-HMA and AVR-PikF resulted in fine, needle-like crystals in multiple conditions in the commercial Morpheus™ screen (see chapter 3). The crystals from condition H4 were used for seeding. A seed stock was prepared by adding 2 µl of reservoir solution to the drop containing the crystals and transferring the solution and crystals to an Eppendorf tube containing an additional 48 µl of reservoir solution. The tube was vortexed to fragment the crystals.

2.9.3 Crystal mounting and cryoprotection

X-ray diffraction data collection requires the protein crystal to be exposed to ionising X-rays, which leads to the formation of free radicals. These can damage the crystal lattice and distort the structure of the protein. To minimise diffusion of free radicals and subsequent radiation damage to the protein crystal, X-ray diffraction experiments are performed at cryogenic temperatures (Garman and Owen, 2006). The crystals are flash-frozen in liquid nitrogen immediately after removal from the sitting drop, and data collection is conducted with the crystal positioned in a stream of liquid nitrogen. However, this process can lead to formation of crystalline ice both on the surface of the protein crystal, and within solvent channels. Expansion of crystalline ice in solvent channels can disrupt the internal structure of the protein, and the presence of crystalline ice, particularly on the surface of the protein crystal, causes characteristic “ice rings” in X-ray diffraction images which can obscure the diffraction of the protein crystal. To reduce the formation of crystalline ice, a cryoprotectant is added to the protein crystal prior to flash-freezing. Cryoprotectants such as ethylene glycol or glycerol replace water molecules within the solvent channels of the protein.

All crystallisation conditions in the commercial Morpheus® screen contain sufficient concentrations of glycerol, ethylene glycol or polyethylene glycol (PEG) to effectively cryoprotect any protein crystals that form. For crystals that formed in conditions that did not contain sufficient cryoprotectant, the crystal was mounted in a loop and

transferred to a small drop of cryoprotectant solution, before being flash-frozen in liquid nitrogen. The cryoprotectant solution used was the corresponding reservoir solution supplemented with 20 % ethylene glycol. Using a cryoprotectant similar to the reservoir solution reduces the risk of the crystal dissolving or becoming stressed during the cryoprotection process.

2.9.4 Data collection

X-ray diffraction data collection was carried out on beamlines I03 and I04 at the Diamond Light Source synchrotron facility (Oxfordshire, UK). Between 3 and 5 test images with an oscillation angle of 45° were collected to determine optimal data collection strategies using EDNA (Incardona et al., 2009) and visually assess the likely resolution of the X-ray diffraction. Datasets consisting of 3600 images were then collected. For the oscillation angle and wavelength used for a dataset, see the relevant results section.

2.9.5 Data processing

Data reduction was carried out using either the Xia2 or AutoPROC pipelines (Vonrhein et al., 2011, Sauter et al., 2004, Grosse-Kunstleve et al., 2002, Evans, 2006) implemented in CCP4i2 (Winn et al., 2011). These pipelines carry out indexing, integration, and scaling of the diffraction dataset. Indexing involves the identification of diffraction spots (in 2D), which are then mapped to scattering vectors which correspond to 3D coordinates in the reciprocal lattice (hkl). This provides an approximation of the unit cell dimensions and an estimate for the Bravais lattice. These parameters are refined and used for data integration, where the positions of the diffraction spots are predicted and the intensity of each spot (I_{hkl}) is measured. Systematic absences, where expected diffraction spots are not present, can help to predict the space group of the crystal. Scaling is then carried out to put the intensities of symmetrically equivalent reflections onto a common scale.

The scaled, but unmerged, data file was the passed to AIMLESS (implemented in CCP4i2) for merging (Evans and Murshudov, 2013). During merging, symmetrically equivalent reflections are combined into a single value. Initially, no resolution limit was set, and the data was processed to the highest resolution supplied in the data file. Subsequent

inspection of merging statistics was used to determine an appropriate resolution. Additionally, plots representing R_{merge} as a function of image number were used to identify if the crystal was significantly damaged by radiation in the later stages of data collection. If a significant increase in R_{merge} was detected in later images, these images were excluded from the final merging.

2.9.6 Phase determination

Information about both the amplitudes and phases of diffracted X-rays are required to generate an electron density map. While the amplitudes of diffracted X-rays can be determined from the intensities of the diffraction spots, information about the phases is lost during the experiment. This is known as the "phase problem" of protein crystallography.

Phase information can be obtained by experimental phasing or molecular replacement. For experimental phasing, heavy atoms are introduced into the protein either during production (for example, selenomethionine can be incorporated in the place of methionine) or by soaking protein crystals in a solution of heavy atoms prior to flash-freezing. The anomalous scattering of X-rays by these heavy atoms can be used to obtain the phases. Alternatively, molecular replacement, where the phases from a known protein structure (the model) with similarity to the structure in question are "borrowed" and applied to the dataset. All structures presented in this thesis were solved by molecular replacement using PHASER (McCoy et al., 2007).

2.9.7 Model refinement and validation

Crystallographic refinement aims to improve the fit between the experimental data and the protein model. The goodness-of-fit can be assessed using the crystallographic R factor and free R factor (R_{free}). The R factor can be inappropriately reduced by overmodelling the data, as the model is built based on the experimental data to which it is then compared. By contrast, the free R factor uses a random subset of experimental data, which is excluded from the refinement process, for cross-validation. This measure is generally preferred as a goodness-of-fit indicator. A perfect agreement between the model and the experimental data would be indicated by R and R_{free} values of 0, while refining a model against a randomly generated dataset would give R and R_{free} values close

to 0.6. Ideally, R and R_{free} values should not differ by >0.05; greater deviation between the two statistics can indicate overmodelling of the data. R and R_{free} generally increase with resolution; for a 2 Å structure, values for R and R_{free} of < 0.2 and 0.25, respectively, would support a good fit.

COOT (Emsley et al., 2010) was used to visualise the electron density map and built model and carry out manual rebuilding and adjustment to improve the fit of the model within the electron density. REFMAC (Murshudov et al., 2011), as implemented in CCP4i2, was used for refinement. 50 refinement cycles were carried out following each rebuilding step. In the later stages of refinement, 5 cycles of translation-libration-screw (TLS) refinement (Winn et al., 2003) were included prior to the main refinement. TLS refinement allows some consideration of anisotropy in the data without necessitating the large number of parameters used by full anisotropic refinement. Where the automatic weighting given to the experimental data relative to protein geometry by REFMAC was considered too high or too low, the weighting was manually specified. The final models were produced by iterative cycles of refinement using REFMAC and rebuilding with COOT, until a good agreement was reached between the model and experimental data as determined by the values of R and R_{free}. Model validation was carried out using the MolProbity web server. All crystal structure figures presented in this thesis were produced with the CCP4 molecular graphics program (CCP4mg) (McNicholas et al., 2011).

2.9.8 Interface analysis

Interface analysis was carried out with QtPISA (Krissinel and Henrick, 2007, Krissinel, 2010) implemented in CCP4i2. Prior to analysis, water molecules were removed from the PDB file, and where multiple copies of the HMA domain or effector were present in the final model, one copy of the HMA/effector complex was selected for analysis, and additional copies were removed from the PDB file. Where comparisons have been drawn between the interfaces of different HMA/effector complexes, equivalent amino acids have been included in the analysis, as shown in table 2.13.

The radar plots generated by QtPISA provide a visual representation of the interface parameters determined by the software. The radar beam for each parameter represents

Table 2.13 Amino acids included in QtPISA analysis of HMA/effector protein complexes.

Structure	HMA	AVR-Pik
Pikp-HMA^{NK-KE}/AVR-PikE Pikp-HMA^{NK-KE}/AVR-PikC Pikp-HMA$^{SNK-EKE}$/AVR-PikC	Leu187-Glu262, excluding Glu198-Asn201 inclusive	Ala32-Phe113
Pikm-HMA/AVR-PikA	Met189-Glu263, excluding Asp201-Lys202 inclusive	Ile33-Phe113
OsHIPP19-HMA/AVR-PikF	Lys2-Glu76, excluding Glu14-Lys15 inclusive	Ile33-Phe113

the probability that the value of that parameter indicates a biologically relevant assemblage in the PDB, with a probability of zero located at the centre of the radar. All radar plots shown are generated using reference parameter "Interface Type", so that probabilities are calculated with reference to PDB entries of type "protein-protein".

2.10 *Nicotiana benthamiana* cell death assays

2.10.1 Growth of *Nicotiana benthamiana* plants

Nicotiana benthamiana plants were grown in a controlled environment room at 22°C constant temperature and 80% relative humidity, with a 16 hour photoperiod. 4-week old plants were used for cell death assays.

2.10.2 Preparation of *Agrobacterium tumefaciens*

Glycerol stocks of *A. tumefaciens* GV3101 cells transformed with the relevant T-DNA constructs were used to inoculate 10 ml LB medium supplemented with appropriate antibiotics (rifampicin, gentamycin and either carbenicillin, for Golden Gate level 1 constructs, or kanamycin, for p19 in pCAMBIA). Cultures were incubated at 28 °C with agitation (180 rpm) for 40-48 hours. Cells were pelleted by centrifugation (3500 x *g* for 15 minutes) and resuspended in 1 ml MMA buffer (10 mM MES pH 5.6, 10 mM $MgCl_2$,

150 mM acetosyringone). The OD_{600} of a 40x dilution of the resuspended cells was measured using a BioMate™ spectrophotometer and used to calculate the OD_{600} of the resuspended cells. The resuspended cells were then diluted in MMA buffer to give stock solutions of defined OD_{600}.

Stock solutions of resuspended *A. tumefaciens* cells were combined to enable co-infiltration of different combinations of NLR proteins and effectors. The viral silencing suppressor p19 is reported to increase protein production levels in *N. benthamiana* and was included in all experiments. The final concentrations (OD_{600}) of *A. tumefaciens* transformed with T-DNA constructs for NLR proteins, effectors and p19 were 0.4, 0.6 and 0.1, respectively, giving a total OD_{600} of 1.5 (2 x (NLR proteins = 0.4) + 1 x (effector = 0.6) + 1 x (p19 = 0.1)). Where one or more component was omitted, the sample was made up to an OD_{600} of 1.5 with *A. tumefaciens* transformed with empty vector.

2.10.3 Agroinfiltration for cell death assays

The different combinations of NLR proteins and effectors were spot-infiltrated on *N. benthamiana* leaves. For each experiment, all NLR/effector combinations were infiltrated on each leaf, and 2 leaves were infiltrated per plant. Depending on the experiment, 10-20 leaves were infiltrated in each experiment. The position of each sample was rotated around the leaf to account for within-leaf variation (Bashandy et al., 2015). Following infiltration, plants were left in the controlled environment room under the same conditions described in section 2.10.1.

2.10.4 Leaf imaging and cell death scoring

Leaves were harvested 5 days post-infiltration (dpi) and photographed under daylight (adaxial surface) and UV light (abaxial surface). Leaf photographs were used to assign a cell death score to each spot, using the seven-point (0-6) scale (figure 2.2) developed by Maqbool et al (Maqbool et al., 2015). Scores from three independent experiments (30-60 leaves in total) were combined. The data were plotted as dot plots using the ggplot2 package (Wickham, 2016) implemented in R (R Core Development Team, 2018).

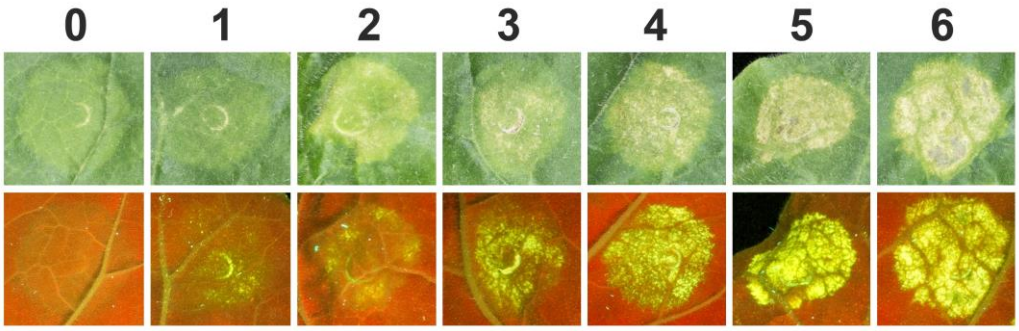


Figure 2.2 Scoring scale used for cell death assays.

The seven-point scale used to score cell death assays as previously published (Maqbool et al., 2015). The top row shows the infiltrated site under daylight, and the bottom row shows the infiltrated site under ultraviolet light.

2.10.5 Protein extraction from *N. benthamiana* leaves to confirm protein production

To confirm the production of each of the proteins in *N. benthamiana*, 3 leaf discs (1 cm diameter) were taken from infiltrated leaves 2 dpi, transferred to Eppendorf tubes and flash-frozen in liquid nitrogen. A micropestle (pre-cooled in liquid nitrogen) was used to grind the leaf tissue. 300 µl of plant protein extraction buffer (GTEN (25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 10 % v/v glycerol), 10 mM DTT, 2 % (w/v) PVPP, 0.1 % Tween®-20, 1x plant protease inhibitor cocktail (Sigma)) was added to each Eppendorf tube of ground leaf tissue. The tube was briefly vortexed to mix, and transferred to a pre-cooled microfuge. Cell debris was pelleted by centrifugation at 20000 x *g* for 2 minutes at 4 °C. The supernatant was transferred to a clean, pre-cooled Eppendorf tube, and centrifuged for a further 2 minutes at 20000 x *g* at 4 °C. 40 µl clarified supernatant was transferred to an Eppendorf tube containing 10 µl SDS-PAGE loading dye RunBlue™ 4X LDS Sample Buffer (Expedeon) and boiled at 95 °C for 5 minutes. SDS-PAGE and Western blotting were carried out as described in sections 2.5 and 2.12 to confirm the presence of the proteins of interest in the leaf tissue.

2.11 Co-immunoprecipitation

2.11.1 Growth of *Nicotiana benthamiana* plants

Nicotiana benthamiana plants were grown in a glasshouse with variable temperature and light levels depending on the local climate. 4-week old plants were infiltrated for co-immunoprecipitation experiments.

2.11.2 Preparation of *Agrobacterium tumefaciens*

30 µl of a (thawed) glycerol stock of *A. tumefaciens* GV3101 cells transformed with the relevant T-DNA constructs was spread evenly on an LB agar plate supplemented with appropriate antibiotics (rifampicin, gentamycin and either carbenicillin, for Golden Gate level 1 constructs, or kanamycin, for p19 in pCAMBIA). Plates were incubated in a static incubator at 28 °C for 40-48 hours. Cells were then scraped from the surface of the LB agar plate using a sterile loop and resuspended in 1 ml MMA buffer (10 mM MES pH 5.6, 10 mM $MgCl_2$, 150 mM acetosyringone). The OD_{600} of a 40x dilution of the resuspended cells was measured using a BioMate™ spectrophotometer and used to calculate the OD_{600} of the resuspended cells. The resuspended cells were then diluted in MMA buffer to give stock solutions of defined OD_{600}.

Stock solutions of resuspended *A. tumefaciens* cells were combined to enable co-infiltration of different combinations of NLR proteins and effectors. As for the cell death assays, the viral silencing suppressor p19 was included in all experiments. The final concentrations (OD_{600}) of *A. tumefaciens* transformed with T-DNA constructs for NLR proteins, effectors and p19 were 0.4, 0.6 and 0.1, respectively, giving a total OD_{600} of 1.1 (1 x (NLR protein = 0.4) + 1 x (effector = 0.6) + 1 x (p19 = 0.1)). Pik-2 was not included, to prevent cell death following infiltration. Where one component was omitted, the sample was made up to an OD_{600} of 1.1 with *A. tumefaciens* transformed with empty vector.

2.11.3 Agroinfiltration for co-immunoprecipitation analysis

To produce proteins for co-immunoprecipitation analysis, whole-leaf infiltrations were carried out. 3 leaves, on 3 different plants, were infiltrated with each NLR/effector

combination. 2 leaves were infiltrated on each plant (with 2 different combinations). Whole leaves were harvested 3 dpi and the leaf midrib removed before leaves infiltrated with the same NLR/effector combination were pooled and flash-frozen in liquid nitrogen.

2.11.4 Extraction of protein from *Nicotiana benthamiana* leaves

Leaf tissue was ground to a fine powder in liquid nitrogen using a pre-chilled pestle and mortar. The leaf powder was then weighed and thoroughly resuspended in 2X leaf tissue mass of ice-cold plant protein extraction buffer (see 2.10.5). Plant cell debris was pelleted by centrifugation at 4,200 x *g* for 30 minutes at 4 °C. From this point until protein elution from the magnetic beads, all steps were carried out at 4 °C using pre-chilled materials. The supernatant was removed from the sample and filtered using a 0.45 µm Minisart® syringe filter (Sartorius). 20 µl filtered extract (the co-immunoprecipitation input) was combined with 5 µl SDS-PAGE loading dye.

2.11.5 Co-immunoprecipitation

α-FLAG® M2 magnetic beads (Merck, formerly Sigma-Aldrich) were supplied as a 50% suspension in 50 % glycerol with 10mM sodium phosphate, 150mM sodium chloride, pH 7.4 and 0.02 % (w/v) sodium azide (PBA/A). Prior to use, the beads were equilibrated in immunoprecipitation (IP) buffer (GTEN + 0.1 % Tween®-20) by repeated separation of the beads from the supernatant using a magnetic rack, removal of the supernatant, and resuspension of the beads in 5X bead volume IP buffer, before being finally resuspended in 2X the original volume of IP buffer (a 25 % suspension).

40 µl of the resuspended α-FLAG® M2 magnetic beads were combined with 1 ml filtered plant protein extract and mixed thoroughly by turning end-over-end for 1 hour at 4 °C. The beads were washed 5X by separating the beads from the supernatant using a magnetic rack, removing the supernatant, and resuspending the beads in 1 ml ice-cold IP buffer.

After the final wash, the beads were resuspended in 30 µl SDS loading buffer. Bound proteins were eluted by boiling at 70 °C for 10 minutes. The magnetic beads were

removed from the eluate prior to analysis of both the input and eluate by SDS-PAGE and Western blotting as described in sections 2.5 and 2.12.

2.12 Western blotting

Following SDS-PAGE of plant protein extracts from co-immunoprecipitation experiments or cell death assays, proteins were visualised by Western blotting. Proteins were transferred from the SDS-PAGE gel to a polyvinylidene difluoride (PVDF) membrane using a Trans-Blot® Turbo™ transfer system (Bio-Rad) according to the manufacturer's guidelines. The pre-programmed "High Molecular Weight" protocol was used for all transfers (10 minutes, 1.3 A for a single gel or 2.5 A for simultaneous protein transfer from two gels, up to 25V).

Following protein transfer, the PVDF membrane was blocked in 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 >60 minutes at 4 °C with gentle agitation. The blocked membrane was then incubated in TBS-T + 5 % (w/v) dried milk powder containing the primary (1°) antibody overnight (>12 hours) at 4 °C with gentle agitation. For the concentrations of each of the antibodies used, see table 2.14. The membrane was then washed 4 times with TBS-T, with a 15 minute incubation in TBS-T with gentle agitation at room temperature between each wash.

Excess TBS-T was removed from the membrane, and the epitope-tagged proteins visualised using the LumiBlue ECL Extreme reagents (Expedeon). 150 µl of each of the two ECL Extreme reagents (peroxide solution and luminol/enhancer solution) were

Table 2.14 Primary antibodies used in this project. All antibodies were conjugated to horseradish peroxidase (HRP).

1° Antibody	Supplier	Working dilution
α-FLAG-HRP	Generon (CPA9020)	1:10000
α-HA-HRP	ThermoFisher Scientific (26183-HRP)	1:3000
α-myc-HRP	Santa Cruz Biotechnology (9E10)	1:5000

mixed and immediately applied to the membrane, ensuring complete coverage. Chemiluminescence was detected using the ImageQuant LAS 500 spectrophotometer (GE Healthcare).

To enable the membrane to be re-probed with a different antibody, the membrane was incubated with 15 ml Restore™ Western Blot Stripping Buffer (ThermoFisher Scientific) for 30 minutes at room temperature with gentle agitation to remove the previous antibody. The membrane was then washed 3 times with TBS-T, and following the final wash, the membrane was incubated in TBS-T for 20 minutes at room temperature with gentle agitation. The restored membrane was blocked, incubated with 1° antibody, washed and visualised as described earlier.

After all proteins of interest had been visualised, the membrane was incubated in Ponceau S stain (0.1 % w/v Ponceau S in 5 % v/v acetic acid) for >15 minutes at room temperature with gentle agitation to observe total protein loading. The membrane was then incubated in dH_2O for 10-20 minutes to remove background stain, and imaged using a document scanner.

2.13 Barley methods

2.13.1 Barley transformation

Agrobacterium tumefaciens mediated transformation of barley (*Hordeum vulgare* cv Golden Promise) embryos was carried out by Matthew Smoker and colleagues at the Sainsbury Laboratory Plant Transformation Platform using a protocol adapted from (Hensel et al., 2008, Harwood et al., 2009). Transformed plantlets were supplied in jiffy containers for potting into soil.

2.13.2 Growth of barley plants

Transformed plantlets were transferred from jiffy containers into 1 L pots containing John Innes cereal mix (40 % medium grade peat, 40 % sterilised soil, 20 % horticultural grit, 1.3 kg/m^3 PG Mix 14-16-18 + Te base fertiliser, 1 kg/m^3 Osmocote Mini 16-8-11 2 mg + Te 0.02% B, wetting agent, 3kg/m^3 Maglime, 300g/m^3 Exemptor (insecticide)). Plants were grown in either a controlled environment room (75 % humidity, 16 hour

photoperiod, 100 % light level, 18 °C day temperature, 12 °C night temperature) or a glasshouse, depending on available space. Plants transferred to the glasshouse were covered with a plastic bag for 3 days to give moderately high humidity during acclimatisation.

2.13.3 Genomic DNA extraction from barley leaf tissue

This protocol was developed by members of the 2Blades group (The Sainsbury Laboratory) for a 96-well plate format, and adapted by members of the Moscou group (The Sainsbury Laboratory) for Eppendorf tubes.

Leaves were sampled 5-10 days after the plants had been transferred to pots. A 4 cm piece of the youngest leaf on the plant was harvested, and carefully rolled around forceps before being transferred to an Eppendorf tube. The rolling process avoids creasing the leaf; wounds caused by creasing can lead to oxidation of isolated DNA. Eppendorf tubes were kept on ice while further samples were collected. If the genomic DNA extraction was not carried out immediately after sampling, leaf tissue samples were stored at -70 °C.

Two stainless steel balls (5/32" diameter, OPM Diagnostics) were added to each Eppendorf tube. The samples were lyophilised for 36-48 hours, and the lyophilised leaf tissue was pulverised using a Geno/Grinder® (SPEX® Sample Prep) for 1 minute at 1500 rpm. The tubes were centrifuged for 30 seconds at 2300 x *g* to return leaf powder on the lid to the base of the tube.

From this point onwards, all steps were carried out in a fume hood. 500 µl CTAB extraction buffer (1 % (w/v) CTAB, 100 mM Tris pH 8.0, 700 mM NaCl, 10 mM EDTA, 140mM β-mercaptoethanol) was added to each Eppendorf tube, taking care to avoid cross-contamination of samples. Tubes were inverted several times to ensure all tissue was in contact with the buffer, then incubated for 45 minutes in a 65 °C water bath, mixing by inversion every 15 minutes.

Tubes were centrifuged for 1 minute at 2300 x *g* to remove sample from the lid of the tube. 200 µl ice cold 5 M potassium acetate was added to each sample, and the tubes inverted several times to mix. Samples were incubated on ice for >20 minutes. Following the incubation, 250 µl chloroform: isoamyl alcohol (24:1) was added, and the samples mixed by agitation for 5 minutes. The samples were then centrifuged for 10 minutes at

2300 x *g*. Taking care not to disturb the interface between the resulting two layers, 500 µl of the upper aqueous layer was added to an Eppendorf tube containing 220 µl isopropanol and inverted several times to mix. Tubes were centrifuged for 15 minutes at 2300 x *g* to pellet the DNA/RNA.

The supernatant was carefully poured away, and residual supernatant allowed to evaporate. The pellet was resuspended in 200 µl TE buffer (10 mM Tris pH 8.0, 1mM EDTA) supplemented with 0.2 mg/ml RNase A, and incubated for 10 minutes at 65 °C. Tubes were centrifuged for 1 minute at 2300 x *g* to remove sample from the lid of the tube. 300 µl isopropanol: 4.4 M NH_4Ac (7:1) was added to each tube and mixed by inversion. Tubes were centrifuged for 10 minutes at 2300 x *g* to pellet the DNA. The supernatant was carefully decanted, and residual supernatant allowed to evaporate. 250 µl of 70 % ethanol was added to rinse the DNA pellet. The tubes were centrifuged for 3 minutes at 2300 x *g*, and the 70 % ethanol poured off. Tubes were left open at room temperature for 10 minutes to evaporate any remaining ethanol.

DNA pellets were resuspended in 100 µl TE buffer and stored at 4 °C overnight before being used for PCR to confirm the presence of the transgene (see section 2.3.2.4).

3

The biochemical and structural basis of the interaction between AVR-Pik and OsHIPP19

3

The biochemical and structural basis of the interaction between AVR-Pik and OsHIPP19

3.1 Introduction

Four HIPPs/HPPs were identified as interactors of AVR-Pik in the yeast two-hybrid screen carried out by our collaborators (Ryohei Terauchi, personal communication). This chapter focuses on one of these proteins, OsHIPP19 (*Oryza sativa* HMA domain-containing isoprenylated plant protein 19). OsHIPP19 is a 122 amino acid protein with a N-terminal HMA domain and a C-terminal isoprenylation motif (-CSIM; (Capell et al., 2005, Clarke, 1992)). The HMA domain and isoprenylation motif are separated by a flexible segment containing proline-rich motifs typically associated with protein-protein interactions. The canonical MXCXXC metal-binding motif is degenerate in OsHIPP19 (MPCEKS), with the second cysteine replaced by a serine. The HMA domain of OsHIPP19 shares 51% identity at the amino acid level with the HMA domains of both Pikp-1 and Pikm-1.

This chapter aims to biochemically and structurally characterise the interaction between AVR-Pik and the HMA domain of OsHIPP19 (OsHIPP19-HMA). I first sought to confirm the interaction between AVR-PikD and the HMA domain of OsHIPP19 using purified proteins *in vitro*, and then to investigate the affinity of the interaction using biophysical techniques. I then aimed to extend this analysis to other AVR-Pik effector alleles, hypothesising that AVR-Pik alleles which avoid interacting with the integrated HMA domain of Pik-1 will retain the ability to interact with the HMA domain of OsHIPP19. Finally, I attempted to obtain structural information about the interaction

between AVR-Pik and OsHIPP19-HMA. I hypothesised that AVR-Pik will bind to the HMA domain of OsHIPP19 at a comparable interface to that of the HMA domains of Pikp-1 and Pikm-1.

3.2 Results

3.2.1 Production and purification of the HMA domain of OsHIPP19 from *E. coli*

Work by Professor Hiromasa Saitoh (Iwate Biotechnology Research Center / Tokyo University of Agriculture), Dr Abbas Maqbool (John Innes Centre), and Dr Marina Franceschetti (John Innes Centre) defined the HMA domain of OsHIPP19 (OsHIPP19 $^{2-77}$). Dr Abbas Maqbool cloned the domain into the expression vector pOPIN-M (pOPIN-M:OsHIPP19 $^{2-77}$) for production of recombinant OsHIPP19 $^{2-77}$ in E. coli. *Intact mass spectrometry of the purified OsHIPP19-HMA protein was performed by Dr Gerhard Saalbach (John Innes Centre).*

The HMA domain of Pikp-1 was originally defined as residues Gly186-Ser258 based on a secondary structure modelling approach. This protein could be produced in *E. coli* and was used in the initial characterisation of the interaction between Pikp-HMA and AVR-PikD (Maqbool et al., 2015). However, in the crystal structure of Pikp-HMA, the final β-strand (β_4) appears slightly truncated, suggesting that the complete domain involves additional residues at the C-terminus.

Professor Hiromasa Saitoh and Dr Abbas Maqbool attempted to produce full length OsHIPP19 in *E. coli*. They observed that the protein broke down into two main products, which were identified by mass spectrometry to be OsHIPP19$^{1-77}$ and OsHIPP19$^{1-102}$. An initial construct for the HMA domain of Pikm-1 based on sequence alignment with the original construct for Pikp-HMA (Gly187-Ser259) did not result in soluble protein. However, addition of five amino acids to the C-terminus (Gly187-Asp264) to generate a product comparable to OsHIPP19$^{1-77}$ enabled production of soluble Pikm-HMA. These five amino acids were hypothesised to complete the Pik-HMA domain. Subsequent crystal structures of Pikp-HMA and Pikm-HMA with these additional amino acids demonstrates that these additional amino acids form part of the HMA domain (De la

Concepcion et al., 2018). From here on, OsHIPP19$^{2-77}$ is referred to as OsHIPP19-HMA (figure 3.1a).

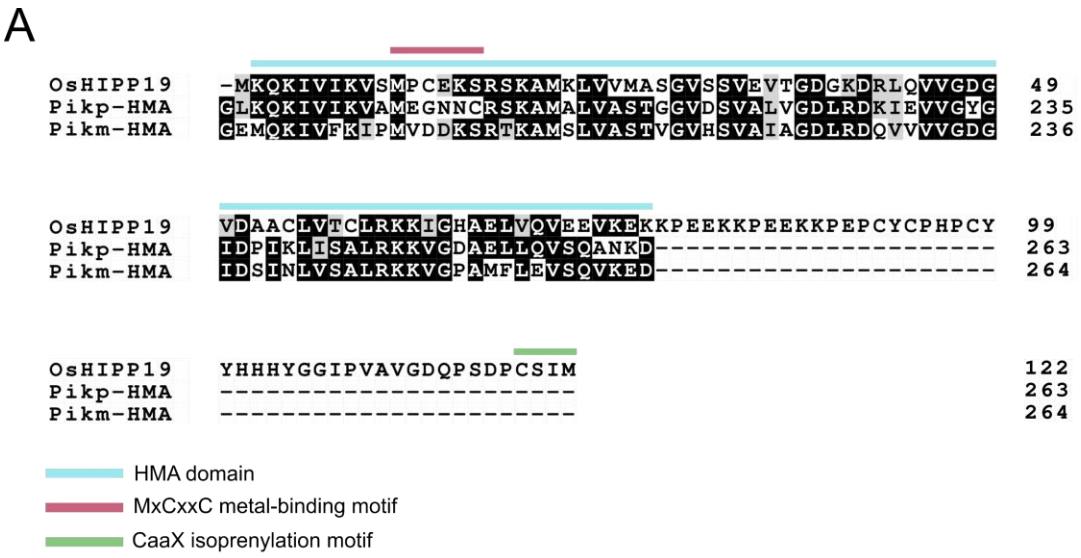
The nucleotide sequence encoding these amino acids was amplified from a synthetic construct (produced by GenScript) containing the OsHIPP19 sequence codon-optimised for gene expression in *E. coli*. Figure 3.1b shows the native and codon-optimised nucleotide sequence of full-length OsHIPP19. The amplicon was cloned into the *E. coli* expression vector pOPIN-M (Berrow et al., 2007) by InFusion cloning (see Materials and Methods).

The pOPIN-M::OsHIPP19-HMA construct produces a fusion protein with a N-terminal hexa-histidine (6xHis) tag adjacent to a maltose binding protein (MBP) tag. The 6xHis tag facilitates purification by immobilised metal affinity chromatography (IMAC), and the MBP tag promotes protein solubility and can also be used in affinity chromatography. A human rhinovirus (HRV) 3C protease cleavage site (LEVLFQGP) is located between the MBP tag and OsHIPP19-HMA. Both tags can therefore be removed from OsHIPP19-HMA by addition of 3C protease. As the protease cleaves between the glutamine (Q) and glycine (G) in the cleavage site, a glycine-proline "scar" remains at the N-terminus of purified OsHIPP19-HMA.

The *E. coli* SHuffle cell line (Lobstein et al., 2012) has been successfully used by other members of the lab to produce the HMA domains of Pikp-1 and Pikm-1. Competent *E. coli* SHuffle cells were transformed with pOPIN-M::OsHIPP19-HMA, and cultured in auto-induction media (Studier, 2005) to induce protein production. Cells were harvested by centrifugation and lysed by sonication. The soluble fraction was obtained from the total cell lysate by centrifugation.

Analysis of the total cell lysate and soluble fraction by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) showed a strong band close to the 50 kDa marker (the 6xHis:MBP:OsHIPP19-HMA fusion protein has an expected molecular weight of 50.9 kDa) in both the total cell lysate and soluble fraction (figure 3.2a), demonstrating that the protein is produced in *E. coli,* and is soluble.

The fusion protein was purified from the soluble fraction by immobilised metal affinity chromatography (IMAC) and gel filtration carried out in tandem using an ÄKTAxpress



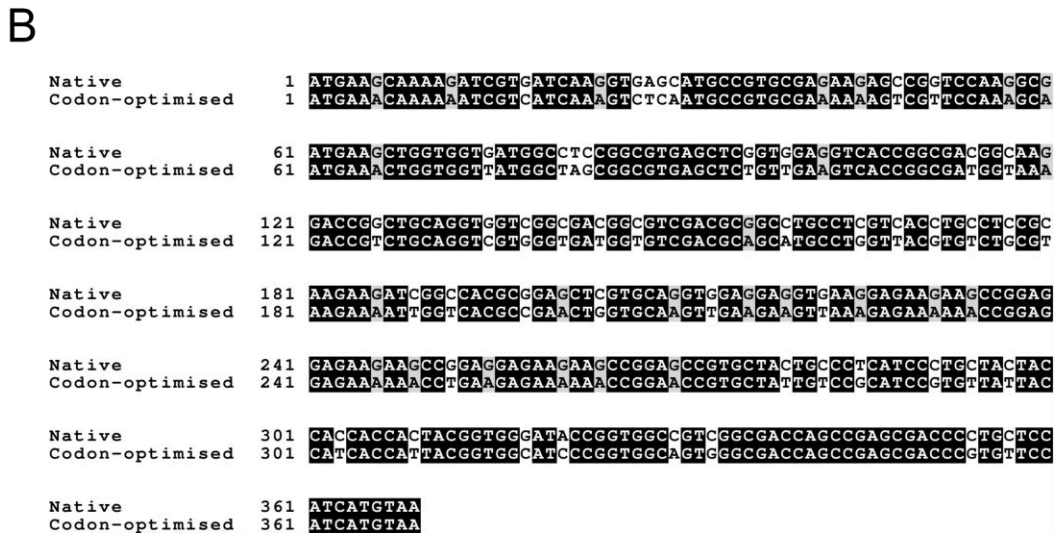


Figure 3.1 Amino acid and nucleotide sequences of OsHIPP19.

A. *Multiple amino acid sequence alignment of OsHIPP19 with HMA domains of Pikp-1 and Pikm-1. The HMA domain of OsHIPP19 is shown by the blue line, the canonical MxCxxC motif is highlighted by the red line, and the CaaX isoprenylation motif is highlighted by the green line.* ***B.*** *Nucleotide sequence alignment of OsHIPP19 from rice (native) and codon-optimised for expression in* E. coli. *Alignments were performed with Clustal Omega and coloured using BoxShade.*

as described in Materials and Methods. The trace from the gel filtration (figure 3.2b) shows two adjacent peaks in the absorbance at 280 nm (A_{280}; peak maxima at 144 ml and 190 ml), with no baseline separation between peaks. SDS-PAGE analysis of fractions corresponding to both peaks showed a band at the expected size for the 6xHis:MBP:OsHIPP19-HMA fusion protein (figure 3.2b). In the later fractions, one or two additional bands could be observed below the band of the expected size for the fusion protein. These may be the result of partial degradation of the protein, or premature separation of the 6xHis:MBP tag. Fractions from the gel filtration containing the 6xHis:MBP:OsHIPP19-HMA fusion protein, as identified by SDS-PAGE, were pooled. To remove the 6xHis:MBP tag for the final purification of OsHIPP19-HMA, the fusion protein was incubated with 3C protease at 4 °C overnight. SDS-PAGE analysis of the protein sample before and after 3C protease treatment confirmed that the 50.9 kDa fusion protein was successfully cleaved into two products (figure 3.2c). The band slightly above the 37 kDa marker corresponds to the 42.6 kDa 6xHis:MBP tag. The band between the 10 kDa and 15 kDa markers, although slightly higher than would be expected, corresponds to OsHIPP19-HMA.

To remove the 6xHis:MBP tag and 6xHis-tagged 3C protease, the sample was passed through tandem HisTrap and MBPTrap columns. OsHIPP19-HMA was present in the flow-through and wash-through from the columns, while the 6xHis:MBP tag, uncleaved 6xHis:MBP:OsHIPP19-HMA, and 6xHis:3C protease were retained until specifically eluted (figure 3.2d).

The flow-through and wash-through fractions containing OsHIPP19-HMA were pooled and a final gel filtration step was carried out. To prevent aggregation of OsHIPP19-HMA induced by cysteine oxidation, 1 mM TCEP was added to the gel filtration buffer. OsHIPP19-HMA contains no tyrosine and tryptophan residues, and therefore absorbs UV light at 280 nm very poorly (the molar extinction coefficient of OsHIPP19-HMA is just 360 $cm^{-1}M^{-1}$). Consequently, the trace from the final gel filtration step is virtually a flat line (figure 3.3a). By analysing the fractions by SDS-PAGE, however, it can be observed that OsHIPP19-HMA eluted from the column between 226 ml and 258 ml. Furthermore, SDS-PAGE shows that the protein is present as a single band, with no observable contaminants (figure 3.3a).

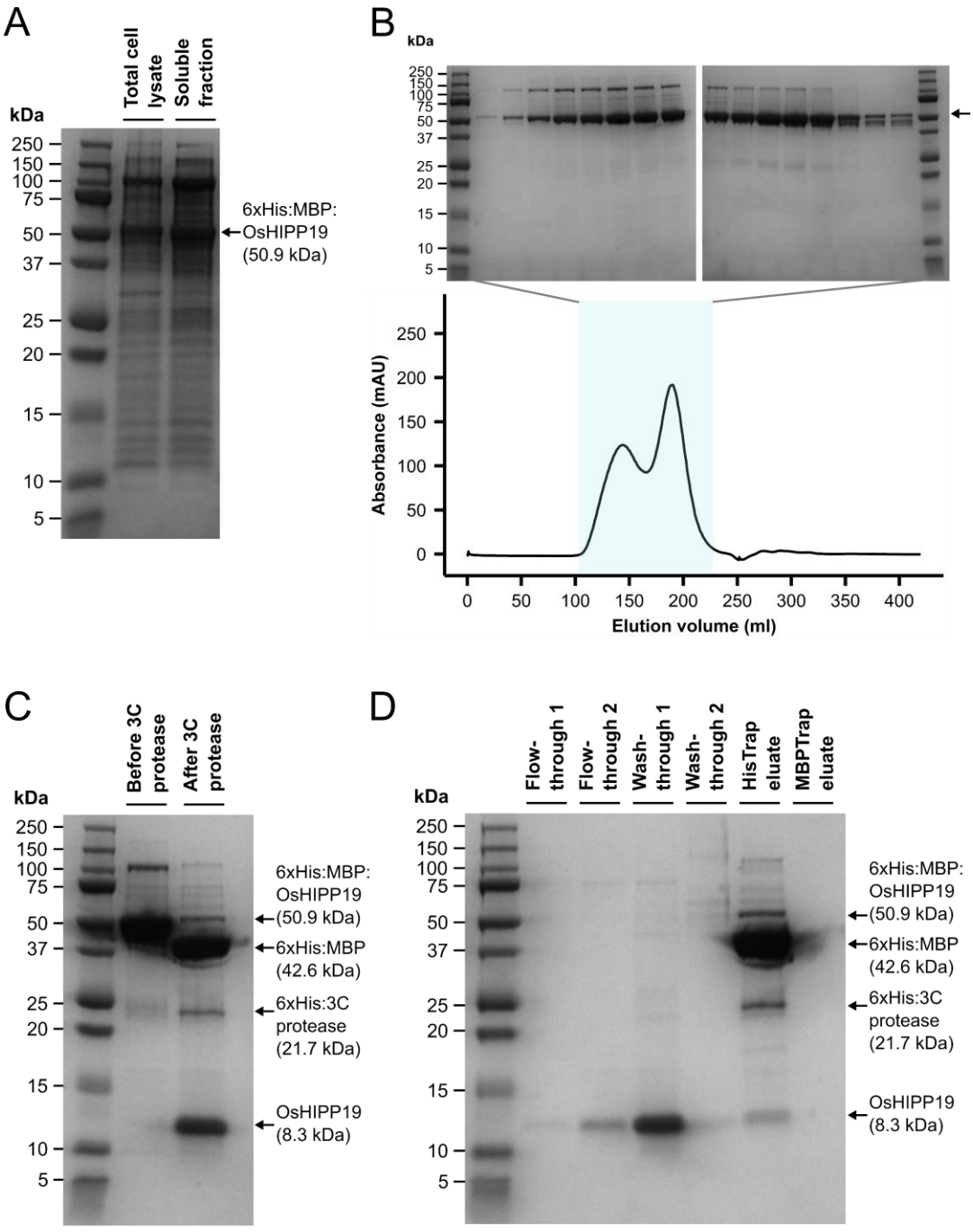


Figure 3.2 Stages in the purification of OsHIPP19-HMA from E. coli ***SHuffle.***

A. *SDS-PAGE gel showing the total cell lysate and soluble fraction from* E. coli *cells expressing pOPIN-M::OsHIPP19-HMA.* ***B.*** *Trace from the gel filtration of 6xHis:MBP:OsHIPP19-HMA. SDS-PAGE gel shows fractions corresponding to the shaded area of the trace (114-218 ml).* ***C.*** *SDS-PAGE gel showing the sample before and after 3C protease cleavage.* ***D.*** *SDS-PAGE gel of fractions from the tandem HisTrap and MBPTrap columns. Arrows indicate bands of interest.*

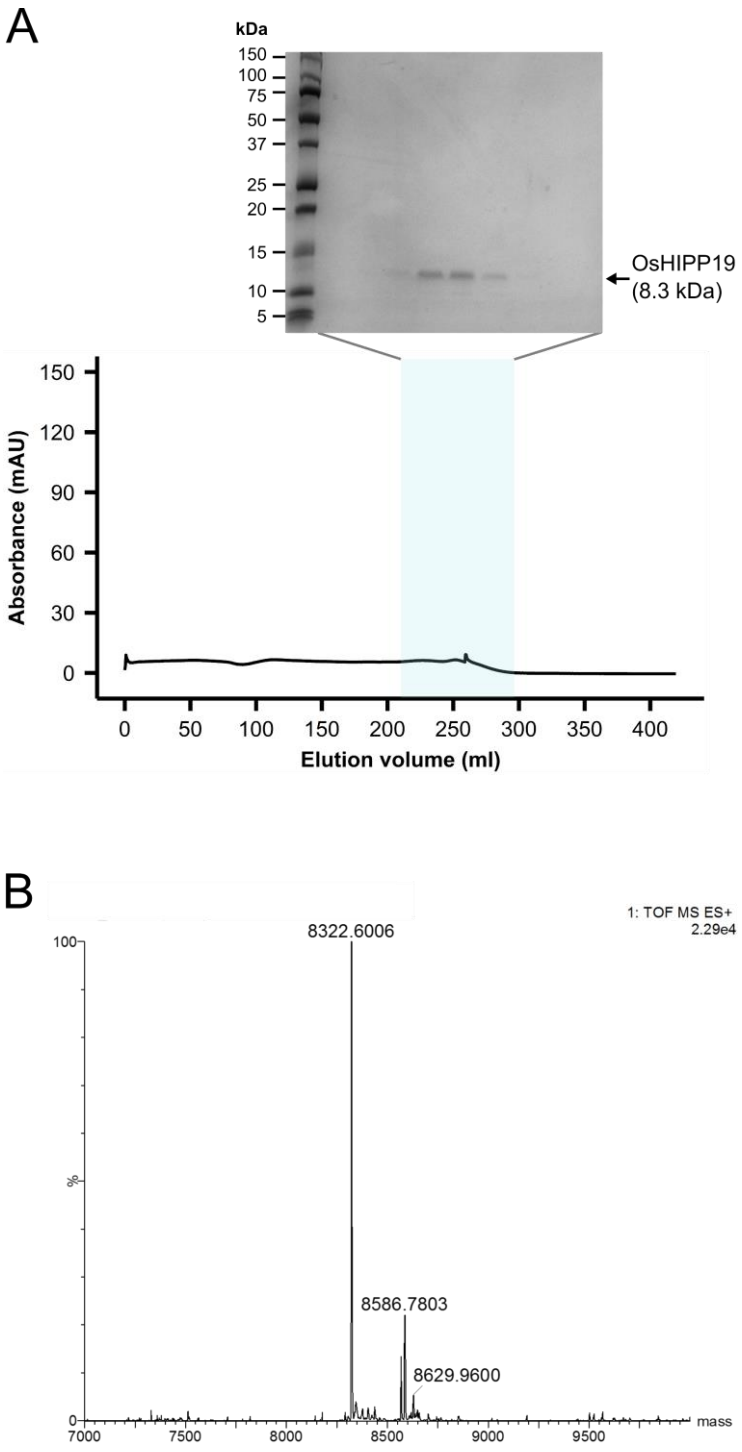


Figure 3.3 Final gel filtration and intact mass spectrometry analysis of OsHIPP19-HMA.

A. *Trace from the final gel filtration of OsHIPP19-HMA. SDS-PAGE gel shows fractions corresponding to the shaded area of the trace (210-298 ml).* ***B.*** *Intact mass spectrometry analysis of the purified sample. The major peak at 8323 Da exactly matches the expected mass of OsHIPP19-HMA.*

To confirm the identity of the protein, the sample was given to the JIC proteomics team for in-tact mass spectrometry. The results showed a major peak at 8323 Da (figure 3.3b), exactly matching the expected mass of the protein.

3.2.2 Circular dichroism spectroscopy confirms that purified OsHIPP19-HMA folds into secondary structure elements

When producing a protein in a heterologous system for biochemical and biophysical analysis, it is important to verify that the protein can fold into appropriate secondary structures and maintains these structures during the purification process.

Circular dichroism is a spectroscopic technique used to investigate structural features of optically active chiral molecules, such as proteins. Optically active chiral molecules differentially absorb left and right circularly polarised light. Measuring this differential absorption at specific wavelengths in the UV-far UV range can be used to draw conclusions about protein secondary structure. Specifically, α-helices show two negative peaks at 222 nm and 208 nm, and a larger positive peak at 192 nm, while β-sheets show a negative peak between 215 nm and 219 nm, and a larger positive peak between 195 nm and 202 nm.

As both HEPES and chloride ions absorb strongly in the far UV (<200 nm), the purified OsHIPP19-HMA protein was exchanged into a phosphate buffer (20 mM potassium phosphate, pH 7.2) by repeated cycles of concentration and dilution. Spectra were obtained in triplicate for wavelengths from 185 nm to 260 nm with a step size of 0.5 nm. The spectra were averaged, and after subtracting spectra obtained when only buffer was present, the units converted from millidegrees to mean residue molar ellipticity (MRME) to correct for protein concentration.

The trace for OsHIPP19-HMA showed features characteristic of both α-helices and β-sheets (figure 3.4). Negative peaks are present at both 222 nm and 208 nm, with a larger positive peak at around 190 nm, indicating α-helices are present. Furthermore, a negative peak is observed between 215 nm and 219 nm, and the MRME is positive between 185 nm and 197 nm, showing a β-sheet contribution. To further investigate the relative composition of secondary structure elements, the averaged, blank subtracted spectrum was submitted to the DichroWeb server (Whitmore and Wallace, 2004,

Whitmore and Wallace, 2008). Using the CDSSTR analysis method with each of the four available reference protein sets (3, 4, 6 and 7) (Sreerama and Woody, 2000), the proportion of different secondary structural elements present in the protein were calculated (table 3.1). The results are similar regardless of the reference protein set used, and support the visual analysis of the spectrum, indicating that the protein forms both α-helices and β-sheets.

These findings are consistent with the expectation that the protein would adopt the previously characterised HMA domain fold; the fold comprises a four-stranded antiparallel β-sheet and two α-helices arranged in an α-β sandwich. Importantly, this demonstrates that OsHIPP19-HMA, when produced in a heterologous system (*E. coli*) and purified, can fold to form secondary structure elements. The ability to produce the HMA domain of OsHIPP19 in *E. coli* and purify it to homogeneity provides the opportunity to study this protein *in vitro*.

3.2.3 Production and purification of AVR-PikD from *E. coli*

Dr Abbas Maqbool provided the pOPIN-E::SUMO:AVR-PikD construct. Dr Abbas Maqbool and Professor Hiromasa Saitoh originally produced and purified AVR-PikD:6xHis. Intact mass spectrometry of the purified protein was performed by Dr Gerhard Saalbach.

The pOPIN-E::SUMO:AVR-PikD construct results in a fusion protein consisting of AVR-PikD (residues Glu22 to Phe93, lacking the N-terminal signal peptide) with a N-terminal small ubiquitin-like modifier (SUMO) solubility tag and a C-terminal 6xHis affinity tag. A 3C protease cleavage site between the SUMO tag and AVR-PikD enables cleavage of the solubility tag during the later stages of purification. The 6xHis tag is retained as it is required for surface plasmon resonance experiments (discussed later in this chapter).

Protocols for the production and purification of AVR-Pik effectors from *E. coli* SHuffle have been developed and optimised by lab members and are published (De la Concepcion et al., 2018, Maqbool et al., 2015). Briefly, competent *E. coli* SHuffle cells were transformed with pOPIN-E::SUMO:AVR-PikD, and cultured in auto-induction media to induce protein production. Cell lysis, clarification of lysate, IMAC and gel filtration were

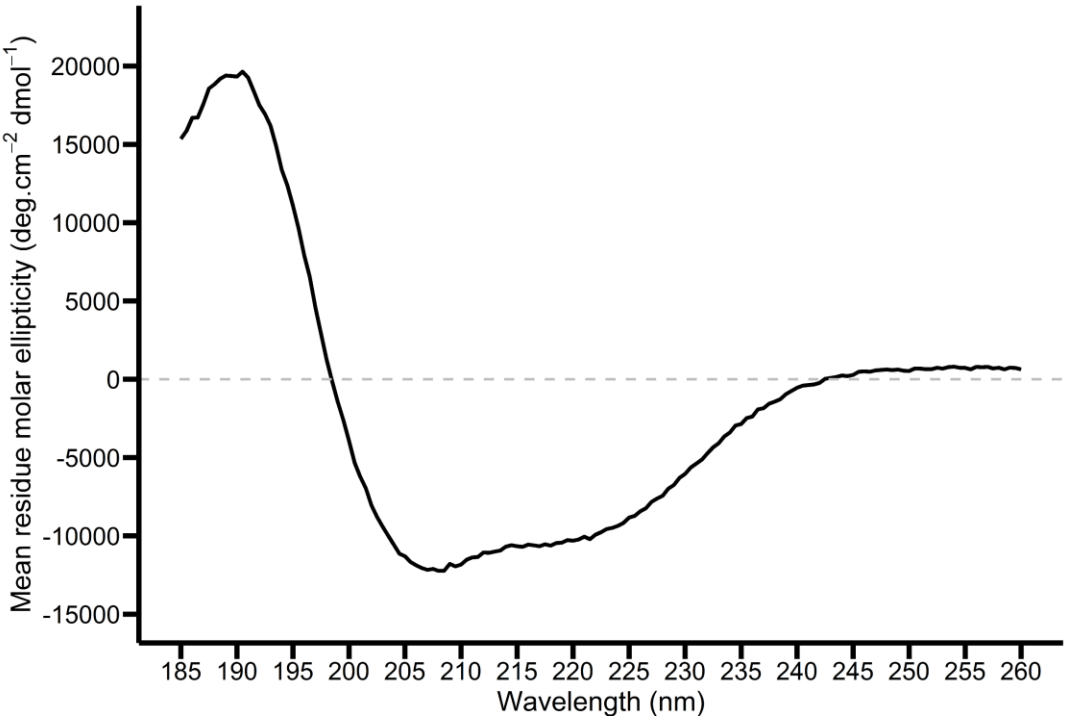


Figure 3.4 Circular dichroism spectrum of OsHIPP19-HMA.

The circular dichroism spectrum shown is the average of three spectra obtained for wavelengths from 185 nm to 260 nm with a step size of 0.5 nm following buffer subtraction.

Table 3.1 Secondary structure assignment for OsHIPP19-HMA from circular dichroism spectra.

Secondary structure assignments were calculated by the DichroWeb server (Whitmore and Wallace, 2004, Whitmore and Wallace, 2008) using the CDSSTR analysis method with each of the four reference sets (Sreerama and Woody, 2000). Helix and strand components are subdivided into regular (1) or distorted (2). NMRSD (normalised root mean squared deviation) is a measure of how well the assigned composition matches the experimental data.

Reference set	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	NMRSD
3	0.20	0.14	0.10	0.08	0.19	0.29	0.011
4	0.17	0.12	0.13	0.09	0.20	0.29	0.009
6	0.20	0.14	0.09	0.07	0.17	0.33	0.011
7	0.19	0.13	0.09	0.07	0.18	0.34	0.010

carried out as described for 6xHis:MBP:OsHIPP19-HMA in section 3.2.1. Following cleavage of the SUMO tag by addition of 3C protease, the sample was passed through a HisTrap column. As the protein remains His-tagged, it was retained in the column until specifically eluted, while the SUMO tag was present in the flow-through and wash-through. The His-tagged 3C protease could not be separated from the His-tagged AVR-PikD in this stage. However, the final gel filtration stage was sufficient to separate the two proteins, as demonstrated by the single band on the SDS-PAGE at the size expected for AVR-PikD:6xHis. As is apparent from the strong peak at 280 nm in the trace for the final gel filtration of AVR-PikD:6xHis (figure 3.5a), AVR-PikD contains multiple tyrosine and tryptophan residues and thus absorbs well at 280 nm, with a molar extinction coefficient of 25,680 $cm^{-1}M^{-1}$. Fractions containing the purified protein were pooled and concentrated. Intact mass spectrometry analysis of the final sample confirmed the identity and purity of the protein, with the peak at 11,784 Da exactly matching the expected mass of AVR-PikD accounting for the formation of a single disulphide bond (minus ~2 Da) as previously reported (figure 3.5b).

3.2.4 AVR-PikD interacts with the HMA domain of OsHIPP19 with nanomolar affinity

The first objective of this project was to validate the interaction between AVR-PikD and OsHIPP19 *in vitro* using purified proteins. To achieve this, analytical gel filtration was used to qualitatively test for an interaction. Analytical gel filtration, also known as size exclusion chromatography, is a technique which separates proteins according to their size and shape. The sample of interest is injected onto a column containing the stationary phase; spherical beads composed of a composite cross-linked agarose and dextran matrix. The matrix creates pores of various sizes in the beads. Larger proteins, or protein complexes, cannot enter as many of these pores as smaller proteins. Consequently, larger proteins progress through the column more rapidly, and elute from the column earlier than smaller proteins, which can occupy more of the pores in the stationary phase and are therefore retained in the column for a longer period.

This technique can be used for investigating protein complex formation; if two proteins interact, the resulting complex will be unable to enter as many of the pores in the stationary phase as either of the two proteins alone and will therefore elute from the

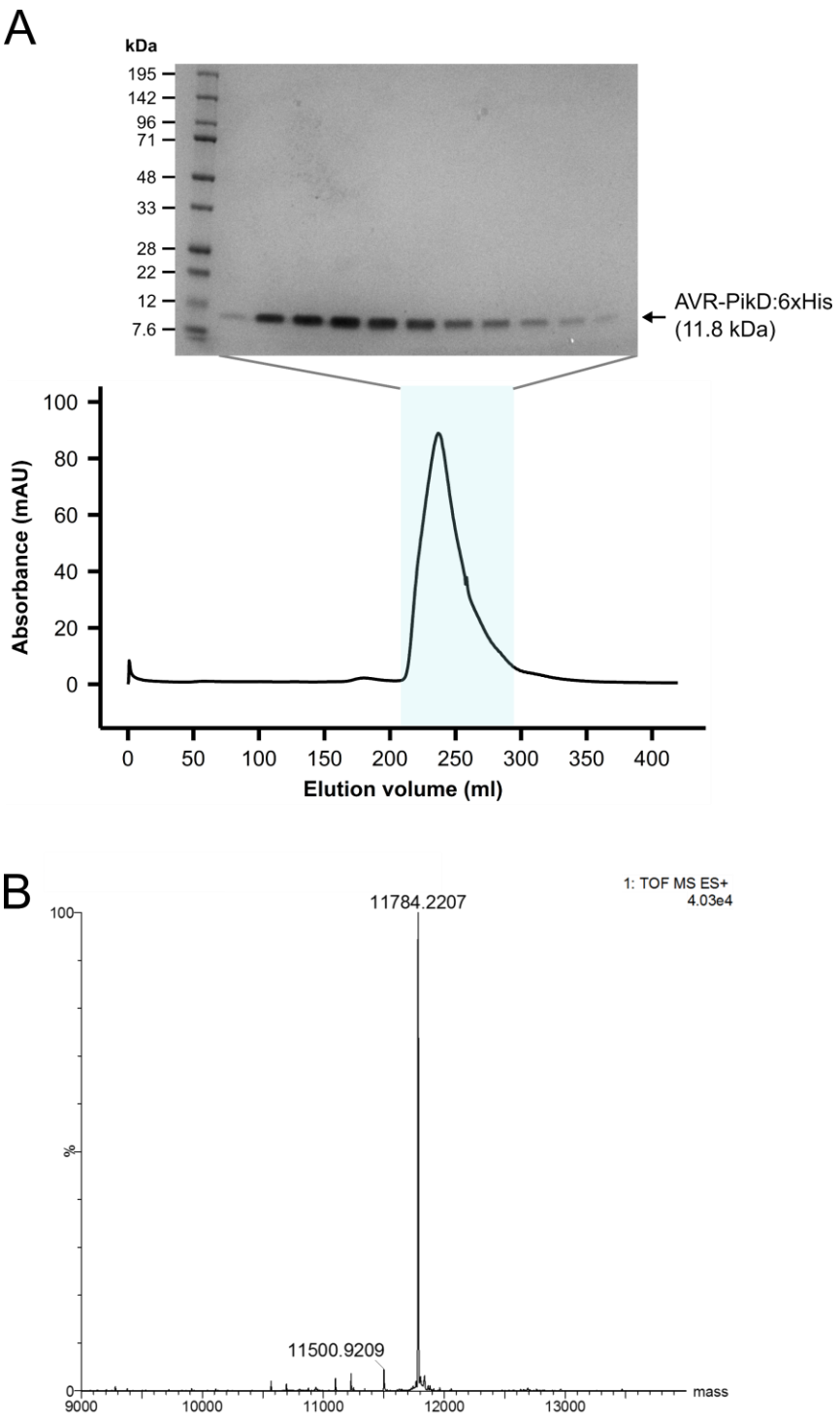


Figure 3.5 Final gel filtration and intact mass spectrometry analysis of AVR-PikD:6xHis.

A. *Trace from the final gel filtration of AVR-PikD:6xHis. SDS-PAGE gel shows fractions corresponding to the shaded area of the trace (210-295 ml).* ***B.*** *Intact mass spectrometry analysis of the purified sample. The major peak at 11784 Da exactly matches the expected mass of the protein, accounting for the formation of a single disulphide bond (11786 Da minus ~2 Da).*

column earlier than the individual proteins. The elution of proteins from the gel filtration column can typically be observed by monitoring the absorbance at 280nm. However, as discussed in section 3.2.1, OsHIPP19-HMA lacks tryptophan and tyrosine residues, and consequently absorbs very weakly at 280nm.

OsHIPP19-HMA was shown by SDS-PAGE of gel filtration fractions to elute from the column at approximately 13 ml. AVR-PikD eluted from the gel filtration column at a volume of 15.8ml, as indicated by the peak in absorbance at 280nm and confirmed by SDS-PAGE of the fraction corresponding to the peak. When AVR-PikD and OsHIPP19-HMA were combined, a shift in peak absorbance to an earlier elution volume of 12.3ml was observed. SDS-PAGE of fractions corresponding to the peak confirmed that AVR-PikD and OsHIPP19-HMA co-elute (figure 3.6a). These results demonstrate that AVR-PikD and OsHIPP19-HMA form a complex *in vitro*.

While analytical gel filtration can be used to qualitatively ascertain whether two proteins interact, biophysical techniques such as surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) can quantify the affinity and kinetics of protein-protein interactions. SPR had previously been used successfully by other members of the group to investigate the affinity of interactions between AVR-Pik effectors and integrated HMA domains (Maqbool et al., 2015, De la Concepcion et al., 2018), and was therefore chosen to study the affinity of the interaction between AVR-PikD and OsHIPP19-HMA.

Surface plasmon resonance measures the change in refractive index caused by the binding of molecules to a sensor chip. To investigate protein-protein interactions by SPR, one protein (the ligand) must be immobilised on the surface of the sensor chip, while a second protein (the analyte) is flowed across the chip. AVR-PikD was immobilised on a Ni^{2+}-NTA sensor chip via a non-cleavable 6xHis tag at the C-terminus of the protein. Using a multicycle kinetics approach, OsHIPP19-HMA, at concentrations ranging from 0.25nM to 25nM, was flowed over the chip for 360 seconds and the association (in response units, RU) recorded. Buffer was then flowed over the chip for a further 180 seconds to record protein dissociation. A 1:1 binding model was fitted to the data. Figure 3.6b shows the observed responses (coloured lines) and the fitted model (black lines), with the residuals (observed response – response predicted by model) plotted below. The residuals are small, indicating good agreement between the model and the experimental

data. If the residuals are entirely within the green acceptance thresholds determined by the Biacore T100 evaluation software (figure 3.6b), the model fits well to the experimental data. If the residuals fall outside the red acceptance thresholds, the model should be considered a poor fit for the data. The model was used to determine the association constant, k_a, and dissociation constant, k_d, for the interaction, from which the equilibrium dissociation constant, K_D, could be calculated according to the equation:

$$K_D = \frac{k_d}{k_a}$$

The calculated K_D value for the interaction between AVR-PikD and OsHIPP19 for the replicate shown is 0.7nM. The experiment was carried out three times, with similar results. The rate constants and standard errors associated with each constant for all three replicates are shown in table 3.2. These results demonstrate that AVR-PikD binds to OsHIPP19 with nanomolar affinity.

3.2.5 AVR-PikD interacts with OsHIPP19-HMA with higher affinity than with the integrated HMA domains of Pik-1

The HMA domain of OsHIPP19 shares 51% amino acid sequence identity with the integrated HMA domains of both Pikp-1 and Pikm-1 (see alignment in figure 3.1a). Previous work has shown that AVR-PikD binds to the integrated HMA domains of Pikp-1 and Pikm-1 with nanomolar affinity (De la Concepcion et al., 2018, Maqbool et al., 2015). To compare the binding of AVR-PikD to the integrated Pik-HMA domains with its binding to OsHIPP19-HMA, surface plasmon resonance (SPR) was used.

As described in 3.2.4, AVR-PikD was immobilised on a Ni^{2+}-NTA chip. Three different concentrations of each HMA domain (2 nM, 5 nM and 20 nM) were flowed over the chip, and the binding stability (R_{obs}, measured in response units (RU)) recorded. Mutating glutamate-230 of Pikp-HMA to arginine has previously been shown to disrupt the interaction of the HMA domain with AVR-PikD (Marina Franceschetti, personal communication); $Pikp^{E230R}$-HMA was therefore used as a negative control.

The R_{obs} for each HMA was then expressed as a percentage of the maximum theoretical response (R_{max}) that would be obtained if each immobilised molecule of AVR-PikD was

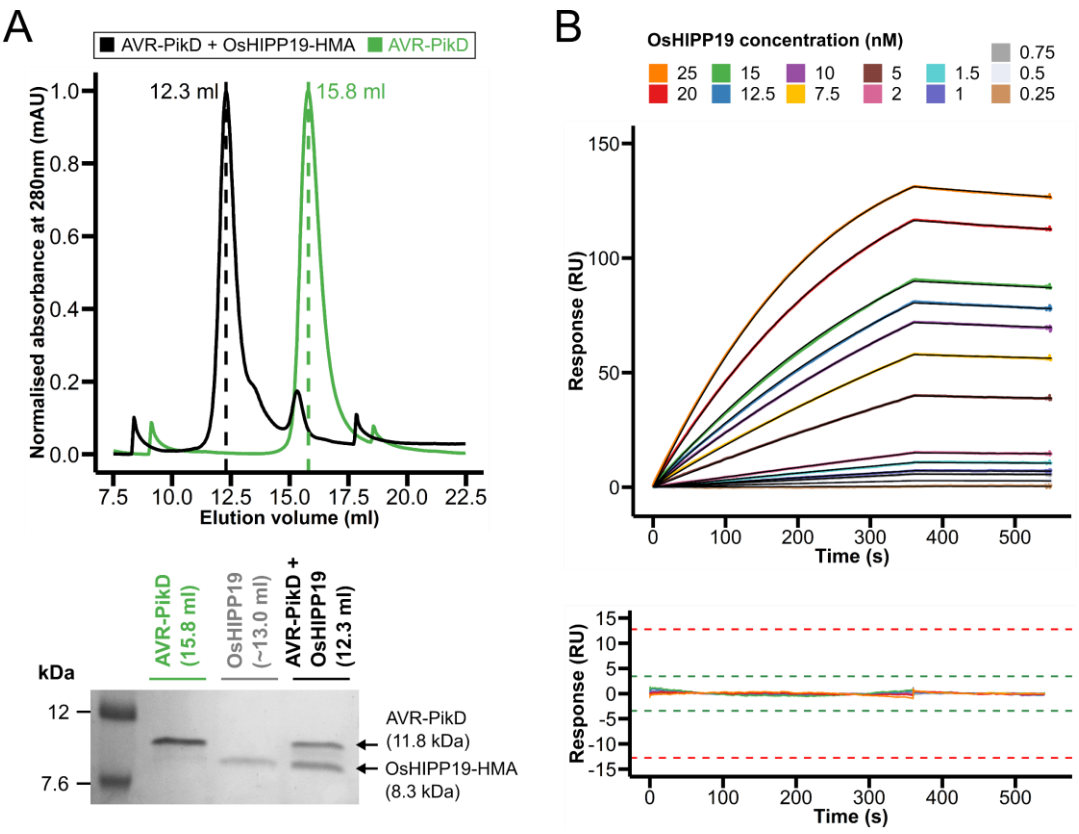


Figure 3.6 AVR-PikD interacts with OsHIPP19-HMA in vitro.

A. *Analytical gel filtration traces for AVR-PikD alone (green) and AVR-PikD with OsHIPP19-HMA (black). Sharp spikes at 8-9 ml and 18-19 ml are caused by a mechanical fault and occur at regular 10 ml intervals, but do not otherwise interfere with the experiment. The SDS-PAGE gel shows fractions from the peak elution volumes of each sample, demonstrating that AVR-PikD and OsHIPP19 co-elute. SDS-PAGE was carried out with method 2 (see Materials and Methods).* ***B.*** *Multicycle kinetics data from surface plasmon resonance (coloured lines) and 1-to-1 binding model fitted to the data (black lines) with the residuals plotted below. Data shown were used to calculate the rate constants described in the table below (replicate 1). Green and red acceptance thresholds are determined by the Biacore T100 evaluation software.*

Table 3.2 Rate constants for the interaction between AVR-PikD and OsHIPP19-HMA.

Replicate	k_a (10^5 M^{-1} s^{-1})	SE	k_d (10^{-4} s^{-1})	SE	K_D (nM)
1	2.610	0.005	1.914	0.004	0.7
2	3.523	0.004	2.186	0.004	0.6
3	2.387	0.003	2.314	0.002	0.9

occupied by the HMA domains. The theoretical R_{max} is calculated by the following equation:

$$R_{max} = \frac{M_W(analyte)}{M_W(ligand)} \times R_{ligand} \times binding\ stoichiometry$$

For Pikm-HMA and OsHIPP19-HMA, the binding stoichiometry is 1:1, while for Pikp-HMA and Pikp^{E230R}-HMA, the binding stoichiometry is 2:1 (De la Concepcion et al., 2018, Maqbool et al., 2015).

Consistent with previous data, AVR-PikD interacted with the HMA domains of Pikp-1 and Pikm-1 with similar apparent affinity. Interestingly, AVR-PikD bound to OsHIPP19-HMA with higher apparent affinity (larger $\%R_{max}$) than to the HMA domains of either Pikp-1 and Pikm-1 at each of the three concentrations tested (figure 3.7).

3.2.6 AVR-Pik alleles which do not interact with integrated Pik-1 HMA domains interact with the HMA domain of OsHIPP19 with nanomolar affinity

Dr Abbas Maqbool provided SUMO:AVR-PikA, SUMO:AVR-PikC and SUMO:AVR-PikE in pOPIN-E. Dr Marina Franceschetti and Professor Chatchawan Jantasuriyarat (Kasetsart University, Thailand) provided SUMO:AVR-PikF in pOPIN-E (not codon-optimised for E. coli *expression), and conducted an initial analytical gel filtration experiment to test for an interaction between AVR-PikF and OsHIPP19-HMA.*

To date, six alleles of AVR-Pik have been defined and designated A-F (Kanzaki et al., 2012, Longya et al., 2019). AVR-PikB was characterised in a Japanese blast isolate (isolate 9505-3; (Yoshida et al., 2009)) and has not been identified in other populations (Kanzaki et al., 2012). Consequently, this allele was excluded from previous studies involving the AVR-Pik alleles (Kanzaki et al., 2012, Maqbool et al., 2015, De la Concepcion et al., 2018) and is similarly not included here. The five remaining alleles differ in just five amino acid positions but show differential binding to the integrated Pikp-HMA and Pikm-HMA domains (see table 1.2). Crucially, neither AVR-PikC nor AVR-PikF interact with either Pikp-HMA or Pikm-HMA, and rice plants containing Pikp-1/Pikp-2 or Pikm-1/Pikm-2 are susceptible to *Magnaporthe oryzae* isolates carrying AVR-PikC or AVR-PikF.

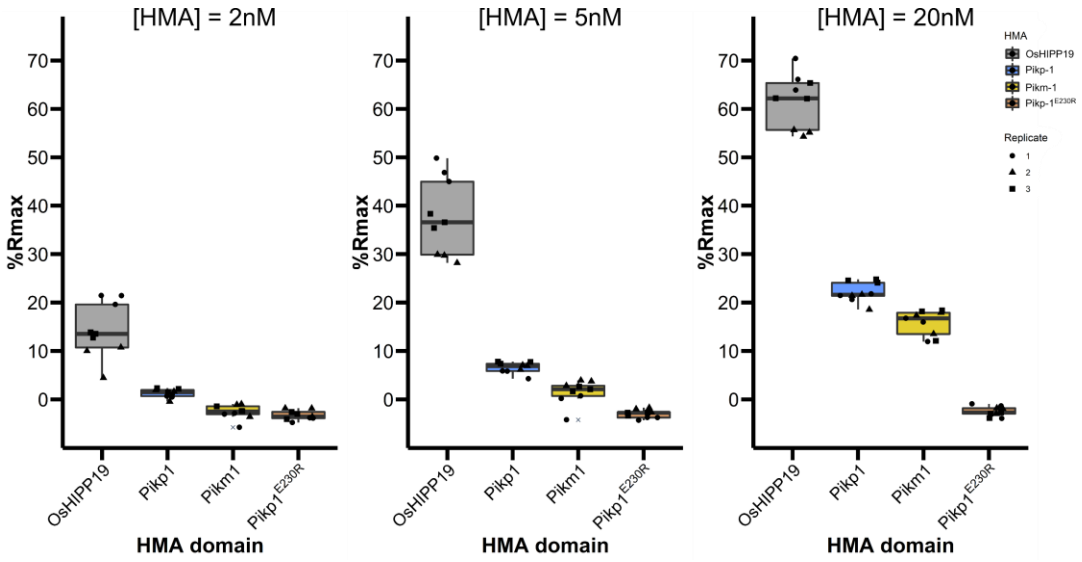


Figure 3.7 AVR-PikD binds more tightly to the HMA domain of OsHIPP19 than to the integrated HMA domains of Pikp-1 and Pikm-1.

$\%R_{max}$ is the percentage of the theoretical maximum response, assuming a 1:1 HMA:effector binding model for OsHIPP19-HMA and Pikm-HMA, and a 2:1 binding model for Pikp-HMA and $Pikp^{E230R}$-HMA (De la Concepcion et al., 2018, Maqbool et al., 2015). The centre line of the box represents the median and the box limits are the upper and lower quartiles. The whiskers extend to the smallest value within Q_1 - 1.5x the interquartile range (IQR) and the largest value within Q_3 + 1.5x IQR. Individual data points are represented as black shapes. The experiment was repeated three times, with each experiment containing three technical replicates. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).

The amino acid polymorphisms that distinguish AVR-PikC and AVR-PikF from other AVR-Pik alleles are adaptive and prevent these alleles from interacting with the HMA domains of the Pik NLR proteins. I hypothesised that, despite carrying these disruptive polymorphisms, these alleles will still interact with the putative virulence target OsHIPP19. To investigate whether each of the AVR-Pik alleles interacts with the HMA domain of OsHIPP19, the four additional AVR-Pik alleles (AVR-PikA, AVR-PikC, AVR-PikE and AVR-PikF) were produced in *E. coli* and purified as described for AVR-PikD in section 3.2.3. Analytical gel filtration demonstrated that each of the AVR-Pik alleles, including AVR-PikC and AVR-PikF, form a complex with OsHIPP19-HMA (figure 3.8).

To test whether the amino acid polymorphisms that prevent AVR-PikC and AVR-PikF from interacting with Pik HMA domains reduce their affinity for OsHIPP19-HMA relative to AVR-PikD, surface plasmon resonance (SPR) was used to determine the equilibrium dissociation constant (K_D) for their interaction with OsHIPP19-HMA as described for AVR-PikD in section 3.2.3. As for AVR-PikD, a 1:1 binding model was fitted to the experimental data for AVR-PikC (figure 3.9a) and AVR-PikF (figure 3.9b) and the residuals calculated and plotted. In both cases, the residuals are small, demonstrating a good fit between the experimental data and fitted model. The rate constants k_a and k_d, associated standard errors, and calculated equilibrium dissociation constants (K_D) are shown in table 3.3. The K_D for the interaction between AVR-PikF and OsHIPP19-HMA is 0.9nM, and the K_D for the interaction between AVR-PikC and OsHIPP19-HMA is 1.5nM. As for AVR-PikD, the experiment was repeated three times, with similar results. This demonstrates that AVR-PikC and AVR-PikF also bind OsHIPP19 with high affinity.

3.2.7 Mutations at interface 2, which disrupt the interaction of AVR-PikD with Pikp-HMA, do not affect the interaction with OsHIPP19-HMA

Dr Abbas Maqbool provided pOPIN-E::SUMO:AVR-PikD^{H46E} and pOPIN-E::SUMO:AVR-PikD^{D66R} constructs for protein production.

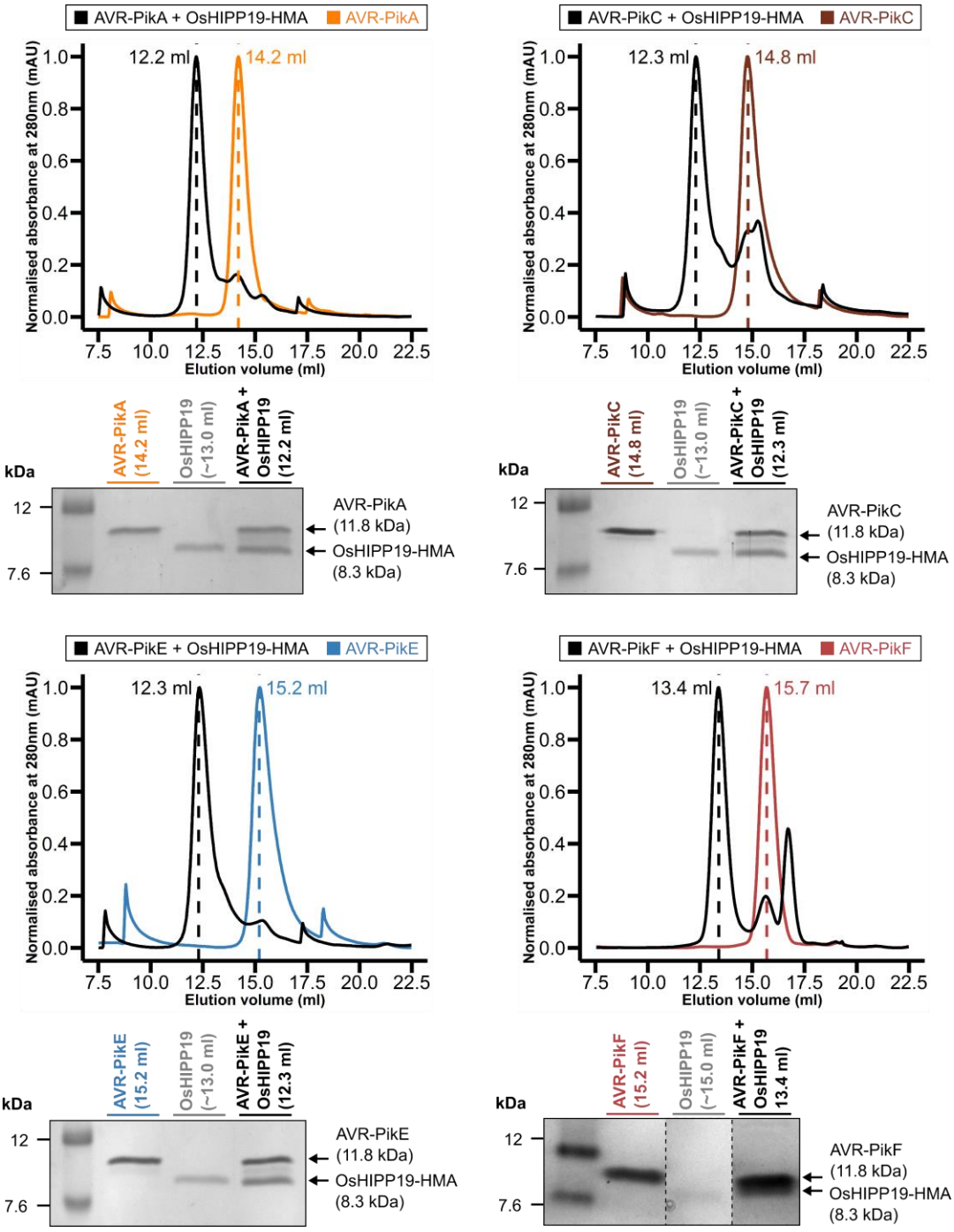


Figure 3.8 AVR-PikA, AVR-PikC, AVR-PikE and AVR-PikF interact with OsHIPP19-HMA.

Analytical gel filtration traces for each effector alone (coloured) and with OsHIPP19-HMA (black). The peak shift observed when each effector is combined with OsHIPP19-HMA indicates complex formation. SDS-PAGE gels show fractions from the peak elution volumes of each sample. The same OsHIPP19-HMA fraction was used for the SDS-PAGE gels for AVR-PikA, AVR-PikC and AVR-PikE. SDS-PAGE gels for AVR-PikA, AVR-PikC and AVR-PikE were carried out with method 2 (see Materials and Methods).

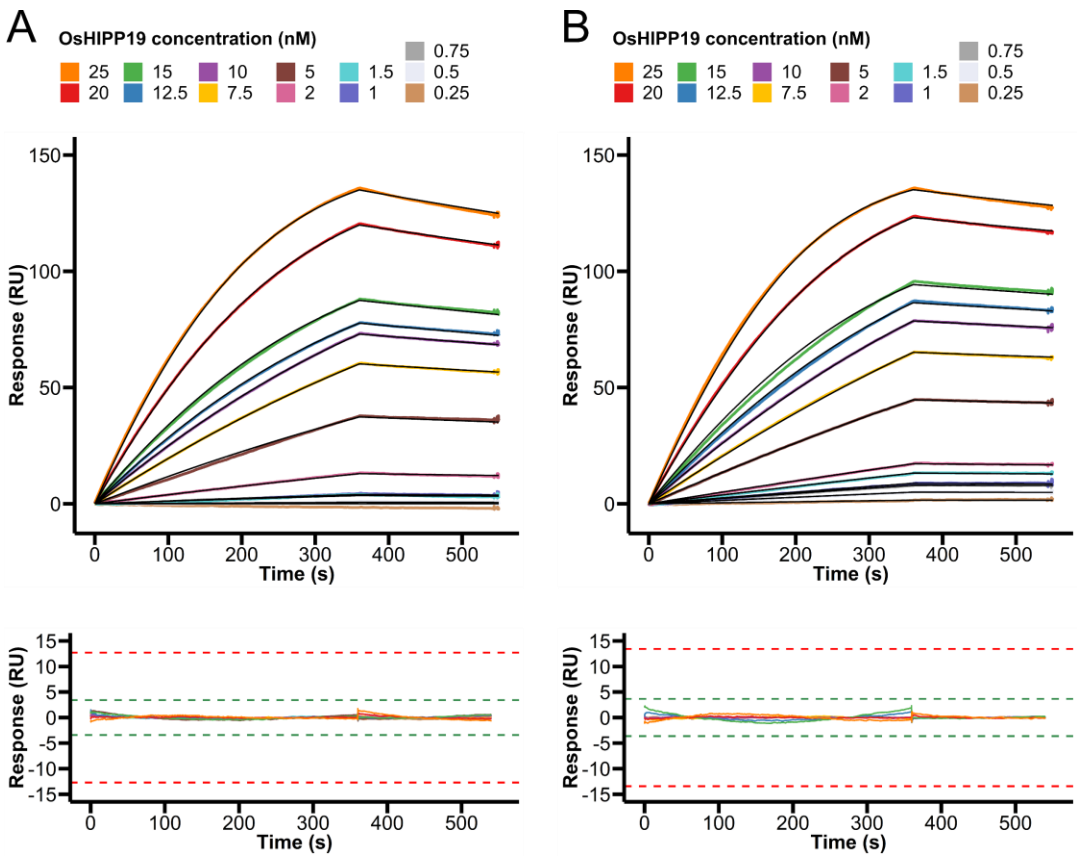


Figure 3.9 AVR-PikC and AVR-PikF bind OsHIPP19-HMA with nanomolar affinity.

Multicycle kinetics data for the interaction with ***A.*** *AVR-Pik or* ***B.*** *AVR-PikF with OsHIPP19-HMA (coloured lines), and 1-to-1 binding model fitted to the data (black lines) with the residuals plotted below. Data shown were used to calculate the rate constants in the table below (replicate 1). Green and red acceptance thresholds for residuals are determined by the Biacore T100 evaluation software.*

Table 3.3 Rate constants for the interactions between AVR-PikC/F and OsHIPP19-HMA.

	Replicate	k_a (10^5 M^{-1} s^{-1})	SE	k_d (10^{-4} s^{-1})	SE	K_D (nM)
AVR-PikC	1	2.884	0.004	4.318	0.004	1.5
	2	4.254	0.010	4.507	0.010	1.1
	3	2.950	0.024	5.609	0.027	1.9
AVR-PikF	1	3.989	0.006	3.485	0.006	0.9
	2	4.731	0.005	4.008	0.005	0.8
	3	3.690	0.004	3.709	0.004	1.0

Based on the structure of Pikp-HMA in complex with AVR-PikD, Maqbool et al. (Maqbool et al., 2015) designed a series of AVR-PikD mutants, each with a single amino acid change at the binding interface, which were hypothesised to disrupt the interaction with Pikp-HMA. Two of these mutants, AVR-PikD^{H46E} and AVR-PikD^{D66R}, failed to interact with Pikp-HMA in both yeast-two-hybrid and surface plasmon resonance (SPR) experiments (Maqbool et al., 2015). The side chain of AVR-PikD^{Asp66} forms hydrogen bonds with the side chain of Pikp-HMA^{Lys195} and the main chain amide group of Pikp-HMA^{Asp224} (figure 3.10a). AVR-PikD^{His46} is positioned in a pocket on the surface of Pikp-HMA, and forms hydrogen bonds with the side chains of Pikp-HMA^{Ser218} and Pikp-HMA^{Glu230} (figure 3.10b). It should be noted that while these experiments were performed with the original Pikp-HMA construct (Gly186-Ser258) lacking the C-terminal five amino acids of the HMA domain, subsequent work has shown that the Pikp-HMA construct with the complete HMA domain (Gly186-Asp263) does not interact with AVR-PikD^{H46E} in yeast-two-hybrid and SPR experiments (De la Concepcion et al., 2018)).To test whether these mutations disrupt the interaction with OsHIPP19-HMA, the AVR-PikD mutants were produced in *E. coli* and purified as described for AVR-PikD in section 3.2.3, and analytical gel filtration was used to test for complex formation *in vitro* as described in section 3.2.4. AVR-PikD^{D66R} and AVR-PikD^{H46E} each co-eluted with OsHIPP19-HMA at an earlier elution volume than observed for the effector or OsHIPP19-HMA alone (figure 3.10c and 3.10d). This demonstrates that AVR-PikD^{H46E} and AVR-PikD^{D66R} can each form a complex with OsHIPP19-HMA.

3.2.8 Purification of the AVR-PikF/OsHIPP19-HMA protein complex

Dr Marina Franceschetti and Professor Chatchawan Jantasuriyarat provided pOPIN-S3C::AVR-PikF.

Following the observation that all AVR-Pik effectors bind OsHIPP19-HMA with high affinity, I aimed to determine the structural basis of the interaction and establish at the atomic level how the binding interface(s) of the AVR-Pik/OsHIPP19 complex differs from the previously characterised interfaces between AVR-Pik and the integrated HMA domains. As neither AVR-PikC nor AVR-PikF interact with the integrated HMA

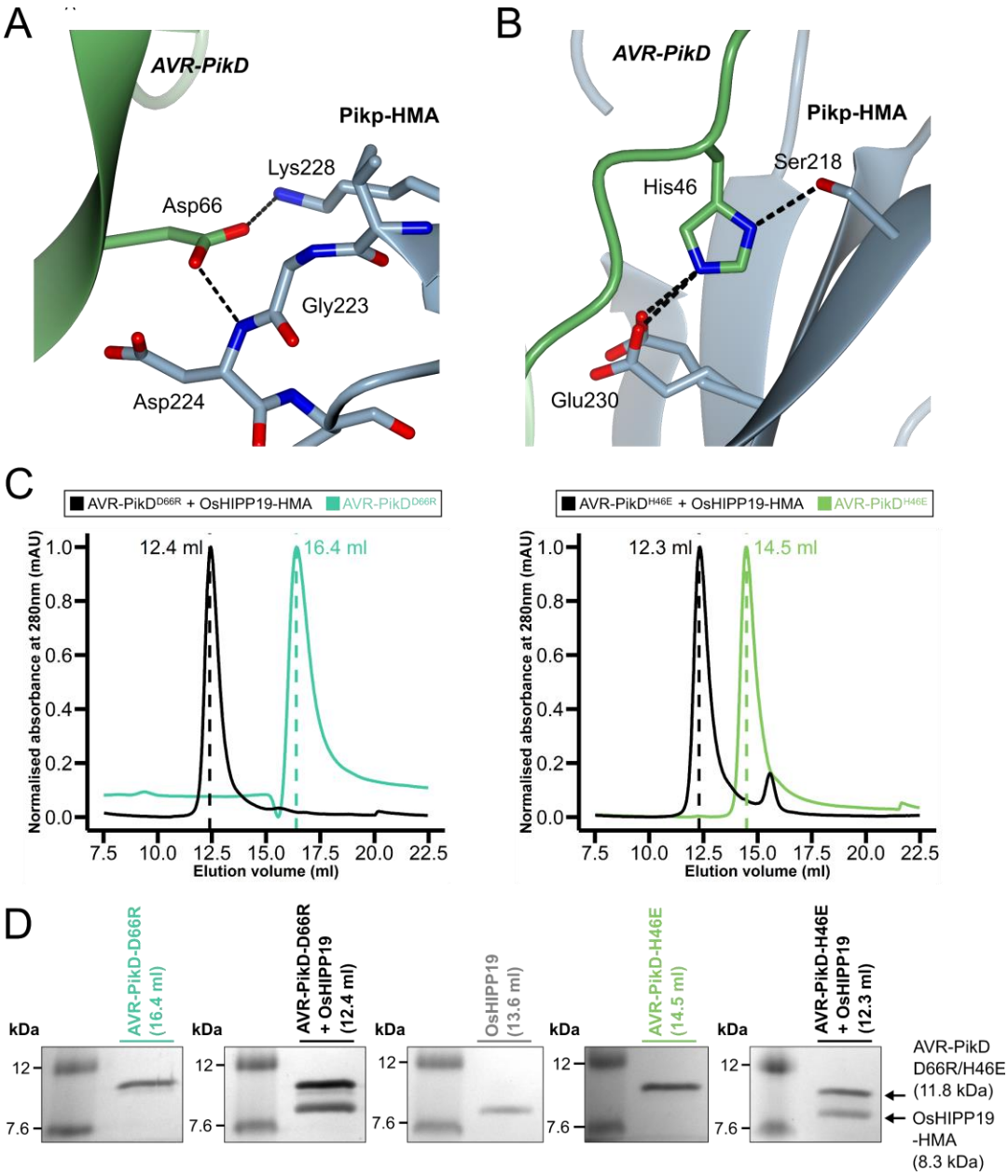


***Figure 3.10 AVR-PikD^{D66R} and AVR-PikD^{H46E} interact with OsHIPP19-HMA* in vitro.**

A. *AVR-PikD^{Asp66} forms hydrogen bonds with Pikp-HMA^{Lys195} and the main chain of Pikp-HMA^{Asp224}.* ***B.*** *AVR-PikD^{His46} is positioned in a pocket on the surface of Pikp-HMA. For both A. and B., the effector is coloured in green and Pikp-HMA in blue. The main chains of the protein are represented as ribbons, with key residues shown as cylinders. The accession code for the PDB file is 5A6W (Maqbool et al., 2015). The N-terminal arm of AVR-PikD has been hidden from the foreground in panel A for clarity.* ***C.*** *Analytical gel filtration traces for each effector alone (coloured) and with OsHIPP19-HMA (black). D. SDS-PAGE gels (method 2; see Materials and Methods) show fractions from the peak elution volumes of each sample.*

domains of Pikp-1 or Pikm-1, their interaction with OsHIPP19-HMA was of particular interest.

The nucleotide sequence encoding MBP:OsHIPP19-HMA (without the 6xHis tag) was amplified from pOPIN-M::OsHIPP19-HMA and the amplicon was cloned into the *E. coli* expression vector pOPIN-A by InFusion cloning. The pOPIN-A::MBP:OsHIPP19-HMA construct produces OsHIPP19-HMA with a N-terminal maltose binding protein (MBP) tag (which can be cleaved by addition of 3C protease), but no 6xHis affinity tag. pOPIN-S3C::AVR-PikF produces AVR-PikF with an N-terminal 6xHis affinity tag and small ubiquitin-like modifier (SUMO) solubility tag, both of which can be cleaved by addition of 3C protease. Crucially, as only the AVR-PikF protein is 6xHis-tagged, the presence of MBP:OsHIPP19-HMA in the eluate from the initial IMAC step is indicative of complex formation.

Competent *E. coli* SHuffle cells were co-transformed with pOPIN-A::MBP:OsHIPP19-HMA and pOPIN-S3C::AVR-PikF. The expression and purification pipeline was similar to that described for OsHIPP19-HMA in section 3.1.1 (see Materials and Methods for full details). While it was not possible to resolve the two proteins on an SDS-PAGE gel following cleavage of the solubility tags due to their similar size (figure 3.11a), intact mass spectrometry confirmed the presence of both proteins in the final sample (figure 3.11b).

3.2.9 Crystallisation of the complex between AVR-PikF and OsHIPP19-HMA

Protein crystals were harvested by Dr Clare Stevenson (John Innes Centre).

Initial crystallisation trials were conducted using the sitting drop vapour diffusion method with the purified complex of AVR-PikF and OsHIPP19-HMA at 22 mg/ml. Two commercial screens, Morpheus™ and JCSG™ (Molecular Dimensions), were trialled. Initial hits were obtained in multiple conditions in the Morpheus™ screen. The Morpheus™ screen uses three buffer systems at different pHs, combined with four precipitant mixes and eight additive mixes, giving a total of 96 distinct conditions. The layout of the screen, and the conditions in which crystals were obtained, are shown in figure 3.12a.

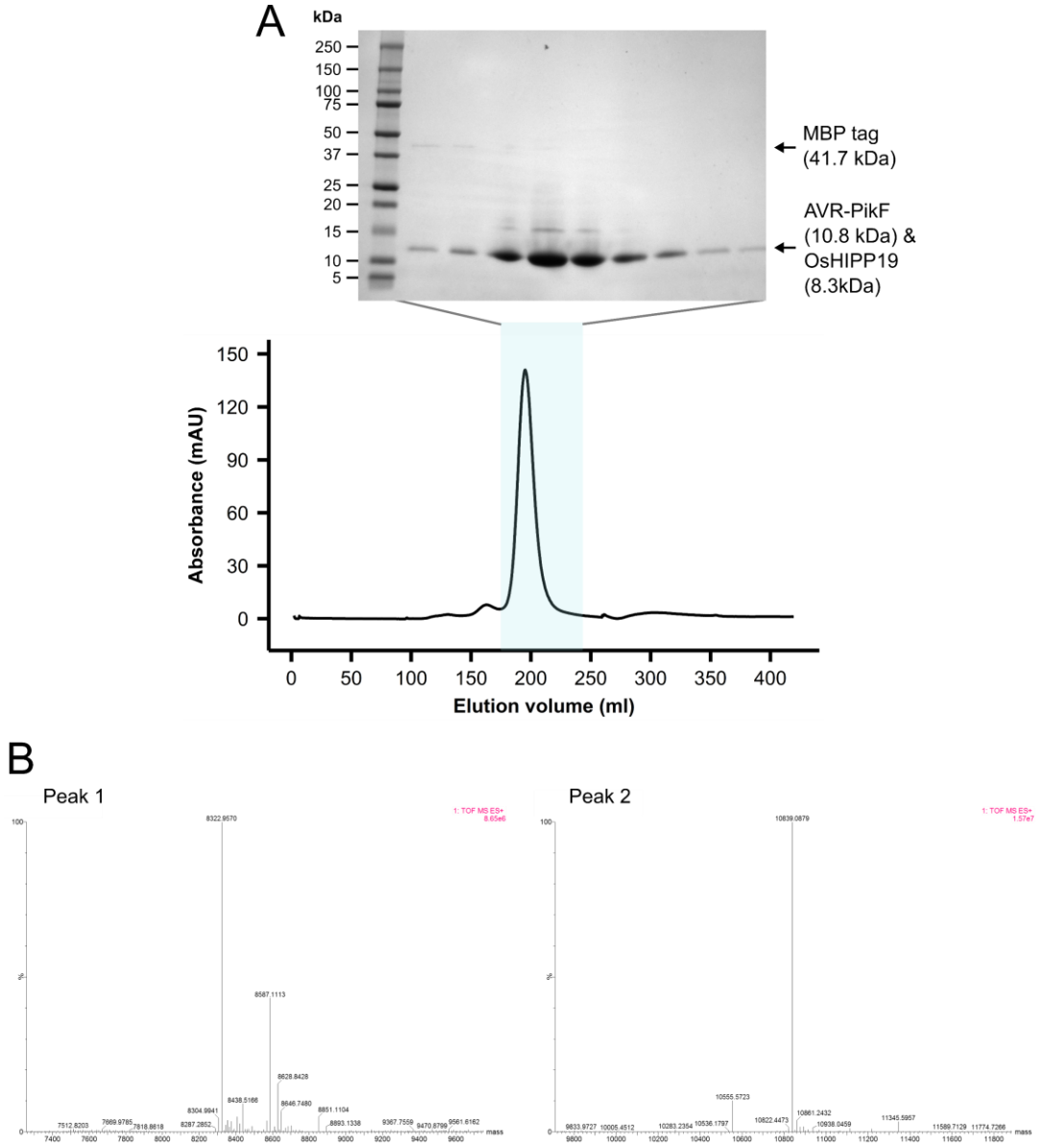


Figure 3.11 Final gel filtration of OsHIPP19-HMA in complex with AVR-PikF and intact mass spectrometry analysis of the purified complex.

A. *Trace from the final gel filtration of OsHIPP19-HMA in complex with AVR-PikF. SDS-PAGE gel shows fractions corresponding to the shaded area of the trace (175-243 ml). The gel shows some (likely MBP tag) contamination in the earlier fractions; these were not pooled for the final sample.* ***B.*** *Intact mass spectrometry analysis of the purified sample. Peak 1 exactly matches the expected mass of OsHIPP19-HMA (8323 Da) and peak 2 exactly matches the expected mass of AVR-PikF (10839 Da), accounting for the formation of a single disulphide bond (minus ~2 Da). While a faint band at approximately 15kDa was visible on the SDS-PAGE gel, no peak was detected at this mass, suggesting it may be a gel artefact.*

All initial hits consisted of showers of thin, needle-like crystals which were not suitable for X-ray diffraction studies. Examples of initial hits are shown in figure 3.12b. These fine needles are typically a result of excessive nucleation, with little protein remaining in the drop for subsequent growth of individual crystals. To optimise crystallisation in order to obtain crystals suitable for X-ray diffraction studies, three approaches were taken. First, reducing the initial concentration of protein can reduce nucleation; therefore, the Morpheus™ screen was repeated with the purified AVR-PikF/OsHIPP19-HMA protein at 11 mg/ml. Second, seeding can be used to control nucleation. Crystals from condition H4 of the initial Morpheus™ screen were used for a seed stock. The Morpheus™ screen was repeated again with the purified AVR-PikF/OsHIPP19-HMA protein at 11 mg/ml with the seed stock. Finally, each of the initial conditions in which crystals were observed contained Morpheus™ Precipitant Mix 4, which comprises 25% v/v 2-methyl-2,4-pentanediol, 25% w/v PEG 1000 and 25% w/v PEG 3350. As this mix is largely PEG-based, varying the molecular weights and concentrations of PEG could result in improved crystals. The commercially available PEGs Suite screen (Qiagen) includes PEG at a variety of molecular weights, concentrations and pH values, and trials were set up using this screen with purified AVR-PikF/OsHIPP19-HMA at 11 mg/ml.

Each of these approaches resulted in the formation of crystals in multiple conditions. In several conditions, the crystals were larger and more suitable for X-ray diffraction studies. Six crystals from six different conditions (table 3.4) were selected for harvesting. All conditions in the Morpheus™ screen are cryoprotected; for crystals from the PEGs Suite screen, 20% ethylene glycol in mother liquor was used as a cryoprotectant. Crystals were individually mounted in a loop, flash frozen in liquid N_2 and transferred to the Diamond Light Source synchrotron facility.

Crystallisation trials were also conducted using the sitting drop vapour diffusion method with the purified complex of AVR-PikC and OsHIPP19-HMA. Seven different commercial screens were trialled, using purified complex at various concentrations. Hits were obtained in just two conditions in the JCSG *Plus*™ (Molecular Dimensions) screen with protein at 15 mg/ml. However, this batch of the screen was contaminated with a fungus, and, despite repeated attempts, the crystals could not be reproduced.

A

Additive mix		1	2	3	4	5	6	7	8	9	10	11	12
Divalents	A	(✓)			✓								✓
Halogens	B				✓				✓				✓
Nitrate Phosphate Sulfate	C				✓				✓				✓
Alcohols	D												
Ethylene glycols	E								✓				
Monosaccharides	F				✓				✓				
Carboxylic acids	G	(✓)			✓				✓				✓
Amino acids	H				✓*				✓				✓

Buffer system	1 (pH 6.5)				2 (pH 7.5)				3 (pH 8.5)			
Precipitant mix	1	2	3	4	1	2	3	4	1	2	3	4

✓ = needle-like crystals (✓) = amorphous crystalline ✓* = used for seed stock

B

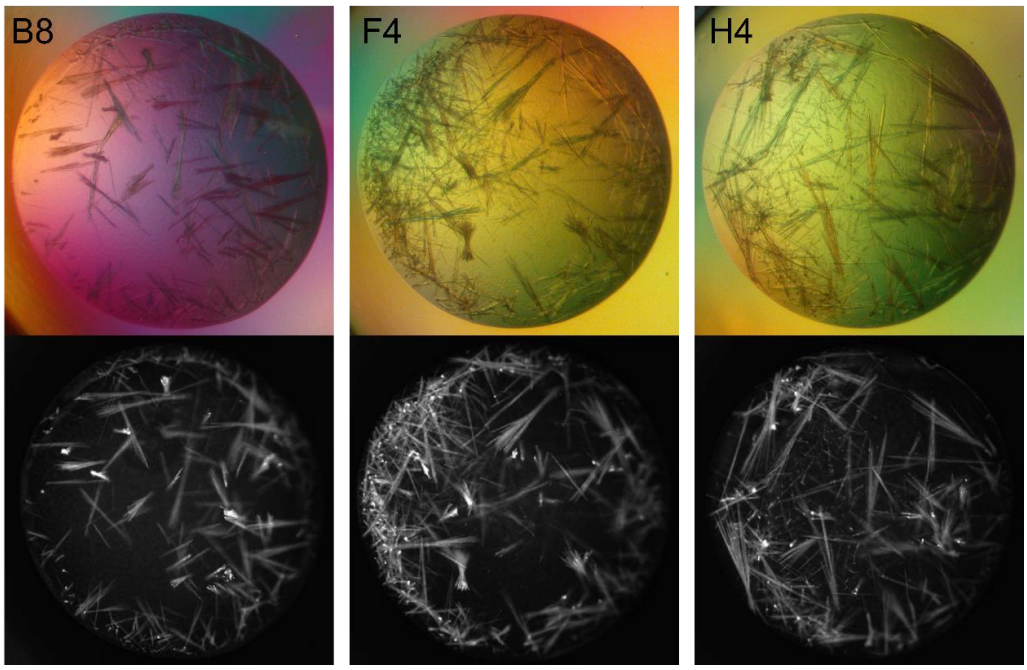


Figure 3.12 Initial hits in crystallisation of AVR-PikF in complex with OsHIPP19-HMA.

A. *Schematic representation of the commercial Morpheus™ crystallisation screen, indicating the conditions in which needle-like crystals were obtained.* ***B.*** *Examples of crystals obtained in the Morpheus™ screen. The top image shows the crystals through a polarising lens, and the bottom image shows the crystals under UV light.*

Consequently, the decision was made to focus on the complex between AVR-PikF and OsHIPP19.

3.2.10 Data collection, processing, and phase determination

Data collection was carried out by Professor David Lawson (John Innes Centre), Dr Clare Stevenson and Dr Marina Franceschetti.

X-ray diffraction data was collected at beamline I03 of the Diamond Light Source synchrotron facility. Following preliminary screening, a diffraction dataset was collected for the crystal harvested from condition D9 of the Morpheus™ screen at a wavelength of 0.98Å. Figure 3.13 shows the crystal in the sitting drop (panel A) and mounted in the loop prior to data collection (panel B). The dataset consisted of 3600 images, with an oscillation angle of 0.1°.

Data reduction was carried out using the autoPROC pipeline (Vonrhein et al., 2011), with the scaled but unmerged data file passed to AIMLESS (as implemented in CCP4i2) (Evans and Murshudov, 2013, Winn et al., 2011). Unit cell parameters were determined to be a= 29.78, b= 53.78, c= 98.03, α= β= γ= 90, and the data was processed in the space group $P\,2_1\,2_1\,2_1$. Based on various merging statistics, the (scaled, unmerged) data was reprocessed by AIMLESS to a maximum resolution of 1.9Å, with an overall R_{merge} of 12.9%. Data processing statistics are shown in table 3.5.

The phases were solved by molecular replacement using PHASER (McCoy et al., 2007), as implemented in CCP4i2. The crystal structure of a monomer of Pikp-HMA in complex with AVR-PikD (PDB accession number 5A6W, chains B and C) was used as a model. A unique solution was found with a translation function (TF) Z-score of 12.3. A TF Z-score of >8 is generally considered to indicate successful phasing.

3.2.11 Model building and refinement

The phased data and amino acid sequences of OsHIPP19-HMA and AVR-PikF were passed to BUCCANEER (Cowtan, 2006) for automated protein model building. 162 residues were successfully built into two fragments, and all were assigned to a chain. Of the two chains supplied, 94.2% of the residues were successfully built. The refinement R-factor (R) was 27%, and the free-R factor (R_{free}) was 34%.

Table 3.4 Summary of crystallisation conditions which gave rise to the crystals that were most suitable for X-ray diffraction data collection.

Screen	Condition	Details
Morpheus™, with seed	C1	Buffer system 1, precipitant mix 1, NPS additive mix
Morpheus™, without seed	D9	Buffer system 3, precipitant mix 1, alcohols additive mix
PEGs Suite, with seed	A1	0.1M sodium acetate pH 4.6, 40% (v/v) PEG 200
	A5	0.1M sodium acetate pH 4.6, 25% (w/v) PEG 1000
	E6	0.2M sodium chloride, 20% (w/v) PEG 3350
	G11	0.2M ammonium sulphate, 20% (w/v) PEG 3350

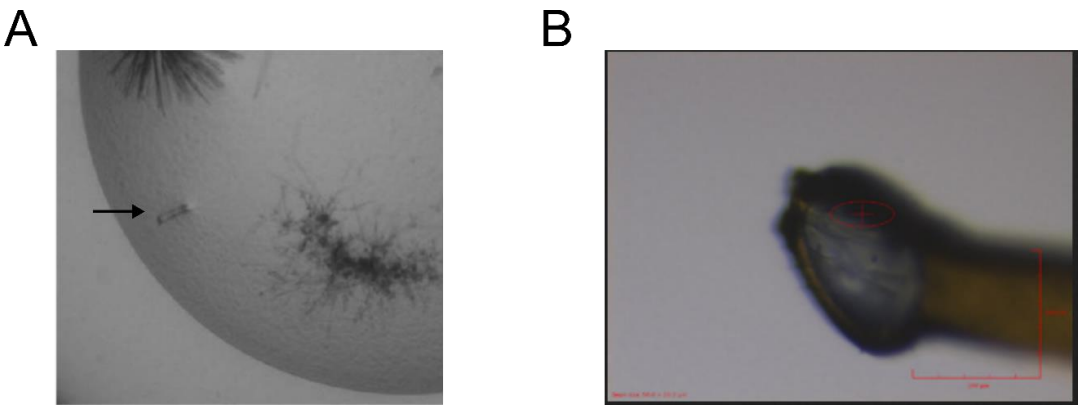


Figure 3.13 Protein crystal used for X-ray diffraction data collection.

A. *The protein crystal from condition D9 of the commercial Morpheus™ screen, without seeding, with protein at a concentration of 11 mg/ml, prior to harvesting from the sitting drop.* ***B.*** *The same crystal mounted in a loop prior to data collection.*

Table 3.5 Data collection and refinement statistics for the AVR-PikF/OsHIPP19-HMA complex.

Data collection statistics	
Wavelength (Å)	0.98
Space group	$P2_12_12_1$
Cell dimensions *a*, *b*, *c* (Å)	29.78, 53.78, 98.03
Resolution (Å)*	98.03-1.90 (1.94-1.90)
R_{merge} (%)	12.9 (112.9)
$I/\sigma I$	12.4 (1.9)
Completeness (%)	100 (100)
Unique reflections	13077 (811)
Redundancy	12.6 (11.1)
CC(1/2) (%)	99.9 (91.0)
Refinement and model statistics	
Resolution (Å)	98.03-1.90 (1.95-1.90)
R_{work}/R_{free} (%)	19.6/23.3 (27.5/27.1)
No. atoms	
Protein	2416
Water	67
B-factors	
Protein	38.1
Water	37.4
R.m.s deviations	
Bond lengths (Å)	0.015
Bond angles (°)	1.661
Ramachandran plot (%)**	
Favoured	97.99
Allowed	2.01
Outliers	0
MolProbity Score	1.39 (98th percentile)

* The highest resolution shell is shown in parenthesis.
** As calculated by MolProbity

Iterative rounds of manual rebuilding, refinement and validation were carried out using COOT (Emsley et al., 2010) and REFMAC (Murshudov et al., 2011). In the later stages of refinement, translation-libration-screw (TLS) parameters were included, using automatically defined TLS groups, to reflect anisotropic displacement of a rigid body (the defined groups of atoms) (Winn et al., 2003). The TLS parameters were refined over 5 cycles prior to the main refinement. In addition, the weighting of the geometric restraints against the experimental data was manually set to 0.1.

Following refinement, the R-factor for the model was 19.6 %, and the R_{free} was 23.3 %. The final model comprised amino acids Ile-33 to Phe-113 of AVR-PikF, and Pro-1 (part of the glycine-proline scar produced by 3C protease cleavage of the 6xHis:MBP tag; see section 3.2.8) to Glu-76 of OsHIPP19-HMA. OsHIPP19^{Glu14} and OsHIPP19^{Lys15} were excluded from the final model as the electron density was not sufficiently well-defined to position them. 67 water molecules were present in the final model.

Final validation of the model was carried out with MolProbity (Chen et al., 2010), which assesses model quality based on protein geometry and steric clashes. A Ramachandran plot is a graphical representation of the phi (φ) and psi (ψ) dihedral angles of each of the amino acids that make up a protein. Certain combinations of φ and ψ are favoured, while others are disallowed due to steric hindrance. The Ramachandran plots for the refined model of AVR-PikF/OsHIPP19-HMA showed that all residues were within the allowed regions, with 98 % (146 residues) within the favoured regions (figure 3.14).

Rotamer analysis revealed that 126 residues were favoured rotamers, however 5 residues were classified as outliers. Each of these residues was inspected, and more favourable rotamers were not supported by the electron density. The clashscore for the model, reflecting the number of steric overlaps per 1000 atoms, is 2.03. This is excellent, placing the model in the 100th percentile among structures of comparable resolution.

MolProbity calculates a summary score reflecting the quality of the model based on clashes, residues in not-favoured areas of the Ramachandran plot and rotamer outliers (Chen et al., 2010). The MolProbity score for the structure of AVR-PikF/OsHIPP19-HMA is 1.39, which places it in the 98th percentile of structures of a similar resolution.

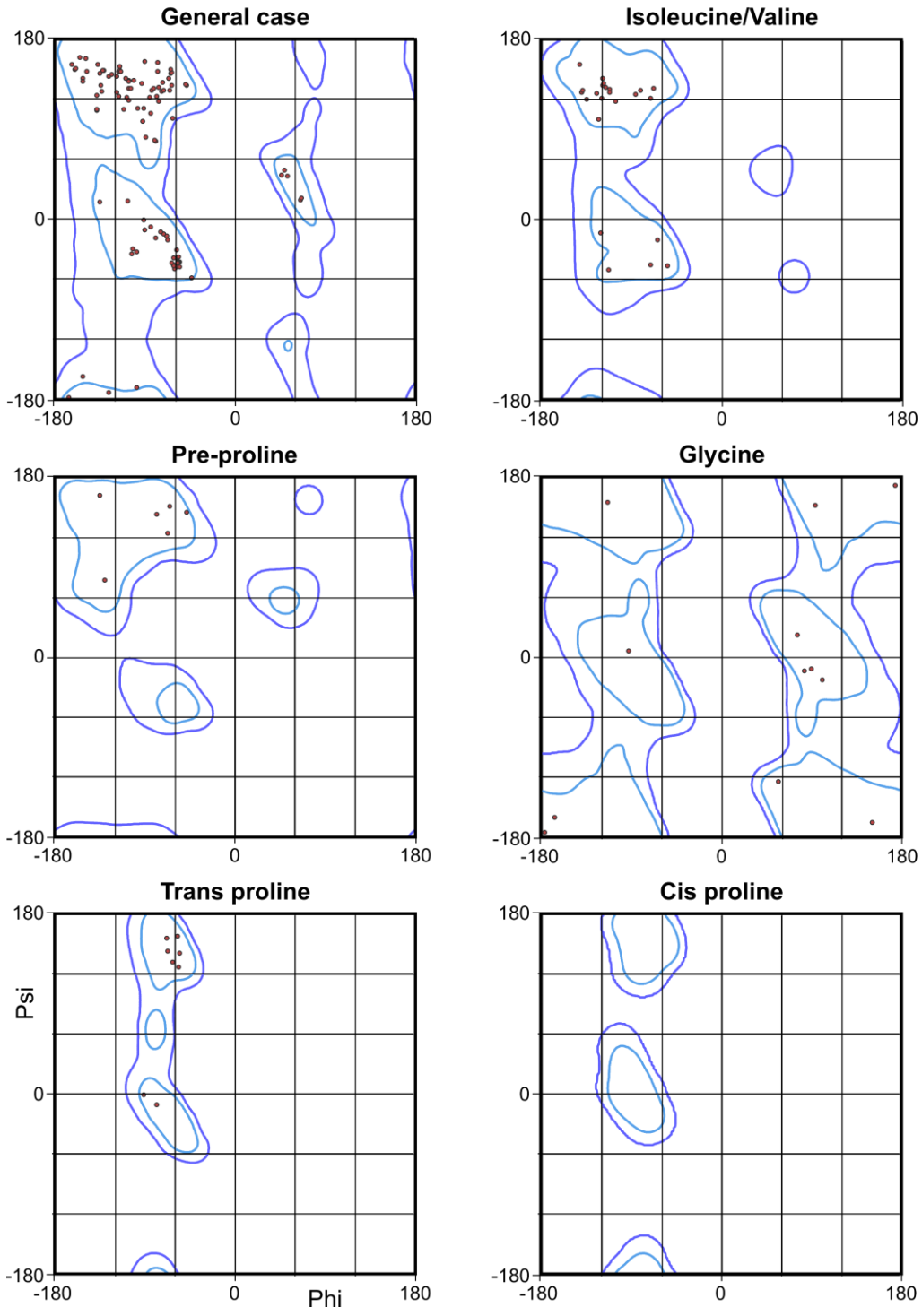


Figure 3.14 Ramachandran plots for the structure of AVR-PikF in complex with OsHIPP19-HMA.

100% of residues were in allowed regions of the plots (within the purple lines), with 97.99% of residues lying within favoured regions (the blue lines). Ramachandran plots were produced with MolProbity.

3.2.12 The crystal structure of AVR-PikF in complex with OsHIPP19-HMA

As anticipated, the HMA domain of OsHIPP19 adopts the well-characterised HMA fold (Pfam: PF00403) consisting of a four-stranded antiparallel β-sheet and two α-helices arranged in an α-β sandwich. The loop between β1 and α1 containing the degenerate metal-binding motif (MPCEKS) is poorly defined in the electron density, and OsHIPP19$^{Glu14\text{-}Lys15}$ could not be positioned.

The structure of AVR-PikF does not differ significantly from the previously determined structures of other AVR-Pik alleles. The effector consists of a core six-stranded β-sandwich structure, conserved among the MAX effectors, with an N-terminal extension comprising AVR-PikF$^{Arg31\text{-}Pro52}$. A disulphide bond between AVR-PikF^{Cys54} and AVR-PikF^{Cys70} stabilises the β-sandwich structure.

OsHIPP19-HMA and AVR-PikF form a 1:1 complex. The position of AVR-PikF relative to the HMA domain of OsHIPP19 is similar to the previously determined structures of other AVR-Pik effector alleles in complex with the integrated Pik-HMA domains (figure 3.15) (De la Concepcion et al., 2018, Maqbool et al., 2015). AVR-PikF only differs from AVR-PikA by a single polymorphism at position 78 (Longya et al., 2019), and the structures of the OsHIPP19-HMA/AVR-PikF complex and Pikm-HMA/AVR-PikA complex (PDB accession code 6FUD) are similar. The root-mean-square deviation of atomic positions (RMSD), as calculated in Coot (Emsley et al., 2010) using secondary structure matching, between the HMA domains is 0.83Å using 74 of 78 residues. The RMSD between the AVR-Pik effectors is 0.44Å using 81 of 81 residues. The overall RMSD between the two complexes is 0.71Å using 155 of 158 residues.

Analysis of the binding interface using QtPISA (Krissinel, 2010, Krissinel and Henrick, 2007) reveals that the interface between OsHIPP19-HMA and AVR-PikF is extensive, burying 23.1% and 19.5% of the total accessible surface area of the HMA domain (1068.5 Å^2) and effector (1022.0 Å^2) respectively. The total interface area (sum of the buried surface area of each component divided by two) for the OsHIPP19-HMA/AVR-PikF complex is 1045.3 Å^2, larger than the total interface area of the Pikm-HMA/AVR-PikA complex (918.3 Å^2) (De la Concepcion et al., 2018), and indeed any of the Pik-HMA/AVR-Pik complexes determined to date. The radar plots (figure 3.16) generated

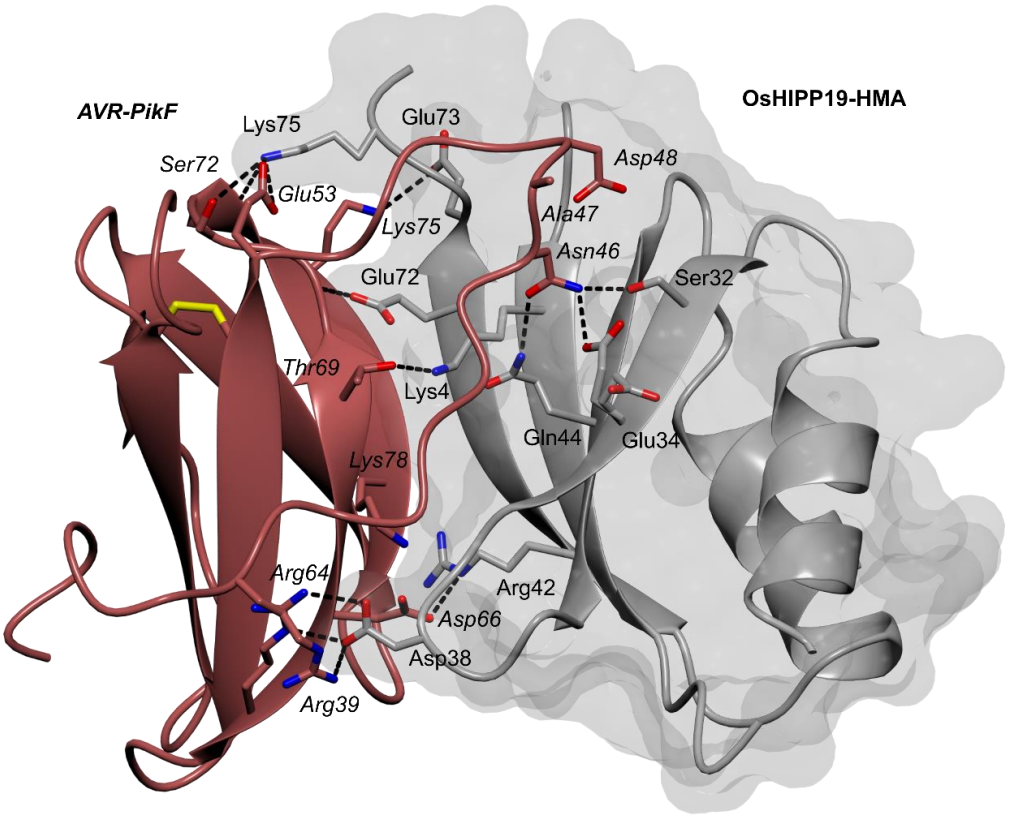


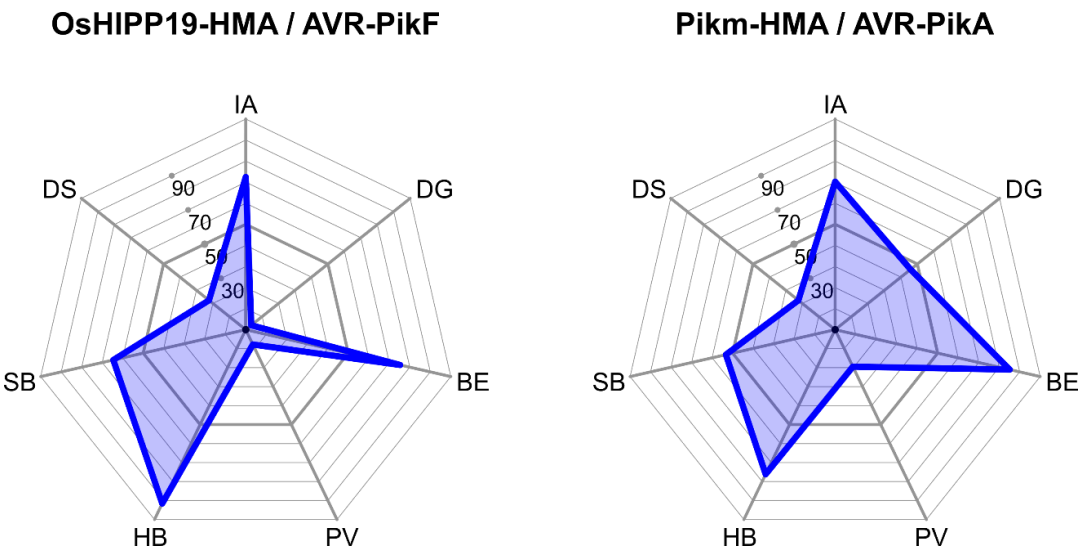
Figure 3.15 The crystal structure of AVR-PikF in complex with OsHIPP19-HMA.

The structures of AVR-PikF and OsHIPP19 are represented as red and grey ribbons respectively, with the molecular surface of OsHIPP19 also displayed. The side chains of amino acids of interest are displayed as cylinders. Hydrogen bonds are represented by dashed lines.

Table 3.6 Summary of interface analysis carried out by QtPISA for the structure of OsHIPP19-HMA/AVR-PikF and the previously published structure of Pikm-HMA/AVR-PikA (De la Concepcion et al., 2018) (PDB accession code 6FUD).

		OsHIPP19 / AVR-PikF	Pikm / AVR-PikA
AVR-Pik	B.S.A. (Å)	1022.0	891.0
	% B.S.A. of total	19.5	16.6
HMA	B.S.A. (Å)	1068.5	945.3
	% B.S.A. of total	23.5	20.7
Total interface area* (Å)		1045.3	918.2
Number of hydrogen bonds		17	12
Number of salt bridges		10	7

* Total interface area is the total B.S.A. (Buried Surface Area) of each component divided by two.



	OsHIPP19 / AVR-PikF	Pikm / AVR-PikA
IA: Interface area	1045.3 Å	918.3 Å
DG: Solvation energy	0.1 kcal/mol	-5.6 kcal/mol
BE: Binding energy	-11.2 kcal/mol	-13.5 kcal/mol
PV: Hydrophobic P-value	0.6917	0.4405
HB: Hydrogen bonds	17	12
SB: Salt bridges	10	7
DB: Disulphide bonds	0	0

Figure 3.16 Comparison of the binding interfaces of OsHIPP19-HMA/AVR-PikF and Pikm-HMA/AVR-PikA using QtPISA.

Radar plots produced by QtPISA provide a visual representation of the binding interfaces of the OsHIPP19-HMA/AVR-PikF and Pikm-HMA/AVR-PikA complexes, based on the seven parameters outlined above. The radar beam for each parameter represents the probability that the value of that parameter indicates a biologically relevant assemblage in the PDB, with a probability of zero located at the centre of the radar.

by QtPISA provide a graphical representation of the binding interface according to seven interface parameters.

3.2.13 Differences at all three interfaces, but particularly at interface 3, likely underpin the higher affinity of AVR-Pik for OsHIPP19-HMA relative to the integrated Pikm-1 HMA domain

Previous analysis of the interface between AVR-PikA and Pikm-HMA revealed three main regions, numbered interfaces 1 to 3, which contribute to the interaction between the twoproteins. Similarly, three distinct regions can be identified in the structure of OsHIPP19-HMA/AVR-PikF. Differences between OsHIPP19-HMA and Pikm-HMA at each of these three interfaces contribute to the differences in specificity and affinity of the interactions between the HMA domains and AVR-Pik alleles.

Interface 1 of AVR-PikA/Pikm-HMA was characterised by a weak (3.5Å) hydrogen bond between the side chain of Pikm-1^{Lys191} and the main-chain carbonyl group of AVR-PikA^{Thr69}, and a hydrophobic interface contributed by Pikm-1^{Met189} (De la Concepcion et al., 2018). Pikm-1^{Lys191} is conserved in OsHIPP19 (OsHIPP19^{Lys4}), and the side chain of OsHIPP19^{Lys4} forms a similar hydrogen bond (3.3Å) with the main-chain carbonyl of AVR-PikF^{Thr69}. However, OsHIPP19^{Lys4} also forms a second, stronger (2.9Å) hydrogen bond with the side chain of AVR-PikF^{Thr69}. The hydrophobic interface contributed by Pikm-1^{Met189} is absent in OsHIPP19, however the N-terminal methionine of OsHIPP19 has been replaced with a glycine-proline cloning artefact, which could alter the interaction of the effector with the extreme N-terminus of OsHIPP19.

In both complexes, interface 2 involves residues from β2 and β3 of the HMA domain (Pikm-1$^{Ser219\text{-}Val233}$ and OsHIPP19$^{Ser31\text{-}Val46}$), which interact with residues in β2 and the N-terminal extension (including the polymorphic residues 46, 47 and 48) of AVR-PikF. AVR-PikA and AVR-PikF share the asparagine-alanine-aspartate (NAD) triad in the polymorphic positions 46, 47 and 48. The interactions between the residues in these positions and residues in β2 and β3 of the HMA domain underpin the differential recognition of AVR-PikD, AVR-PikE, and AVR-PikA by Pikm (De la Concepcion et al., 2018). The side chain of AVR-PikA^{Asn46} forms a single hydrogen bond (2.9Å) with Pikm-1^{Ser219}. By contrast, the side chain of AVR-PikF^{Asn46} is rotated relative to AVR-PikA^{Asn46},

and forms two hydrogen bonds with the side chains of OsHIPP19^{Ser32} (2.5Å) and OsHIPP19^{Gln44} (2.9Å) (figure 3.17a). Pikm-1^{Lys195}, located in β1 of the HMA, also contributes to interface 2; the side chain of Pikm-1^{Lys195} forms a hydrogen bond with the side chain of AVR-PikA^{Asp66}. While this lysine is conserved in OsHIPP19 (OsHIPP19^{Lys8}), β1 is shifted away from the interface with the effector, and no hydrogen bond is formed between OsHIPP19^{Lys8} and AVR-PikF^{Asp66}. Instead, AVR-PikF^{Asp66} forms hydrogen bonds with the main-chain amide group of OsHIPP19^{Asp38} (2.8Å) and side chain of OsHIPP^{Arg42} (2.9Å). The side chain of Pikm-1^{Asp225} forms two hydrogen bonds with the side chain of AVR-PikA^{Arg64}; the aspartate is conserved in OsHIPP19 (OsHIPP19^{Asp38}) and forms similar interactions with AVR-PikF^{Arg64}. The side chain of OsHIPP19^{Asp38} contributes an additional, weak, hydrogen bond (3.8Å) to the side chain of AVR-PikF^{Arg39}.

Interface 3 comprises residues from β4 of the HMA domain extending to the C-terminus of the protein (Pikm-1$^{Met254\text{-}Asp264}$ and OsHIPP19$^{Glu67\text{-}Glu76}$). In the complex between Pikm-HMA/AVR-PikA, this interface is defined by main-chain hydrogen bonding between β4 of the HMA domain and β3 of AVR-PikA, and, notably, the positioning of Pikm-1^{Lys262} into a pocket on the surface of the effector. The conserved OsHIPP19^{Lys75} binds into a similar pocket on the surface of AVR-PikF, formed by AVR-PikF^{Glu53}, AVR-PikF^{Tyr71}, AVR-PikF^{Ser72} and AVR-PikF^{Trp74}, forming hydrogen bonds with the main-chain carbonyl group and side chain of AVR-PikF^{Ser72} (2.7Å and 3.4Å, respectively) and the side chain of AVR-PikF^{Glu53} (2.8Å and 3.4Å).

In addition to main-chain hydrogen bonding between β4 of OsHIPP19-HMA and β3 of AVR-PikF, additional hydrogen bonds are contributed by OsHIPP19^{Glu72} and OsHIPP19^{Glu73}. Strikingly, the side chain of OsHIPP19^{Glu72} (the residue in the corresponding position in Pikm-HMA is serine) forms a hydrogen bond (2.8Å) with the main-chain amide group of AVR-PikF^{Tyr71} (figure 3.17b). In addition, the side chain of OsHIPP19^{Glu73} (a glutamine in the corresponding position in Pikm-HMA) forms a salt bridge interaction with the side chain of AVR-PikF^{Lys75}. Overall, the hydrogen bonding between the HMA domain and effector at interface 3 is more extensive in the OsHIPP19-HMA/AVR-PikF complex than in the Pikm-HMA/AVR-PikA complex, which is likely to contribute to the difference in binding affinity.

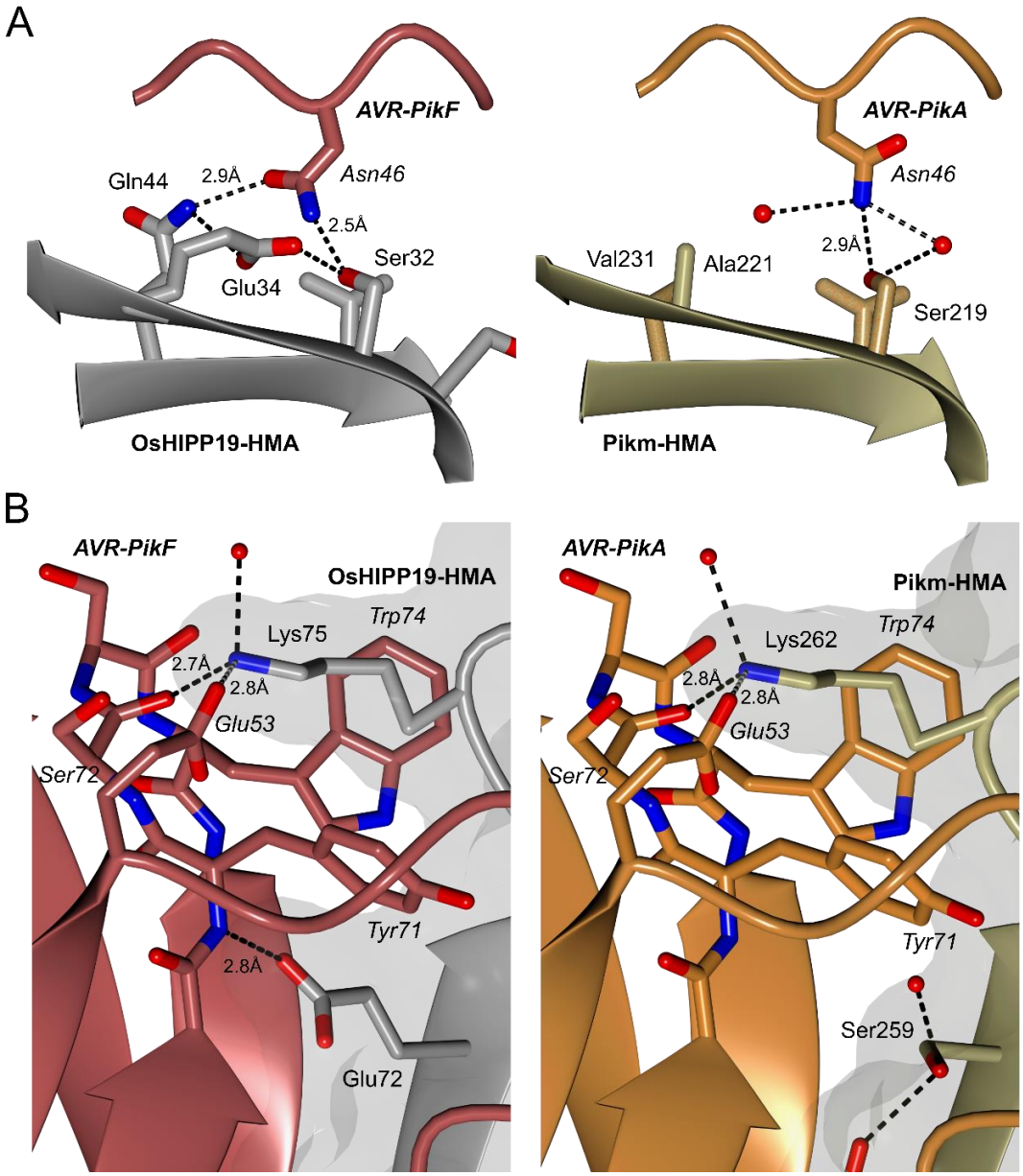


Figure 3.17 Close-up views comparing interfaces 2 and 3 in the structures of OsHIPP19-HMA/AVR-PikF and Pikm-HMA/AVR-PikA.

The structures are represented as grey (OsHIPP19-HMA), red (AVR-PikF), gold (Pikm-HMA) and orange (AVR-PikA) ribbons, with key residues shown as cylinders. Hydrogen bonds are represented as dashed lines. Bond lengths are shown for intermolecular hydrogen bonds. ***A.*** *AVR-PikF^{Asn46} forms two hydrogen bonds with OsHIPP19-HMA compared to the one formed between AVR-PikA^{Asn46} and Pikm-HMA. The side chain of OsHIPP19^{Glu34} exists in two alternate conformations, both supported by the electron density. For clarity, only the relevant conformation is shown.* ***B.*** *OsHIPP19^{Glu72} forms an additional hydrogen bond with the main chain of AVR-PikF. The molecular surface of the HMA domains are shown in grey (OsHIPP19-HMA) and gold (Pikm-HMA).*

3.2.14 AVR-PikF^{Lys78} is located at the binding interface with OsHIPP19 and is likely to be adaptive

The only difference between AVR-PikA, which binds to Pikm-1 to trigger plant immunity, and AVR-PikF, which does not, is the residue at position 78 (methionine in AVR-PikA and lysine in AVR-PikF). The polymorphic AVR-PikF^{Lys78} is positioned at the binding interface with OsHIPP19 (figure 3.18) and is well defined in the electron density. I hypothesise that while this residue is sufficient to disrupt the interaction between AVR-PikF and the integrated HMA domains, increased intermolecular interactions between AVR-PikF and OsHIPP19, particularly at interface 3, are sufficient to compensate for the disruptive influence of the AVR-PikF^{Lys78} side chain and maintain the interaction between the two proteins.

3.3 Discussion

In this chapter, I used biochemical and structural techniques to characterise the interaction between AVR-Pik and the HMA domain of rice heavy-metal associated isoprenylated plant protein 19 (OsHIPP19). I first confirmed the interaction between AVR-PikD and OsHIPP19-HMA which was observed in yeast two-hybrid assays by our collaborators, and then used biophysical techniques to demonstrate that AVR-PikD interacts with OsHIPP19-HMA with nanomolar affinity. Furthermore, I observed that AVR-PikD interacts with OsHIPP19-HMA with higher apparent affinity than with the integrated Pik-HMA domains. Interestingly, I found that all AVR-Pik alleles, including AVR-PikC and AVR-PikF, interact with OsHIPP19-HMA with nanomolar affinity. By solving the crystal structure of AVR-PikF bound to OsHIPP19-HMA and comparing it to published structures of AVR-Pik effectors in complex with either Pikp-HMA or Pikm-HMA, I identified differences between the binding interfaces which may explain the increased affinity and broader specificity of the AVR-Pik/OsHIPP19-HMA interaction.

The identification of diverse protein domains integrated into the typical structure of an NLR protein has offered new insights into the molecular mechanisms through which NLR proteins sense the presence of effectors. The integrated domain hypothesis proposes that these domains have their origins in the host virulence targets of the effector, and that the similarity of the integrated domain to the host target results in the effector

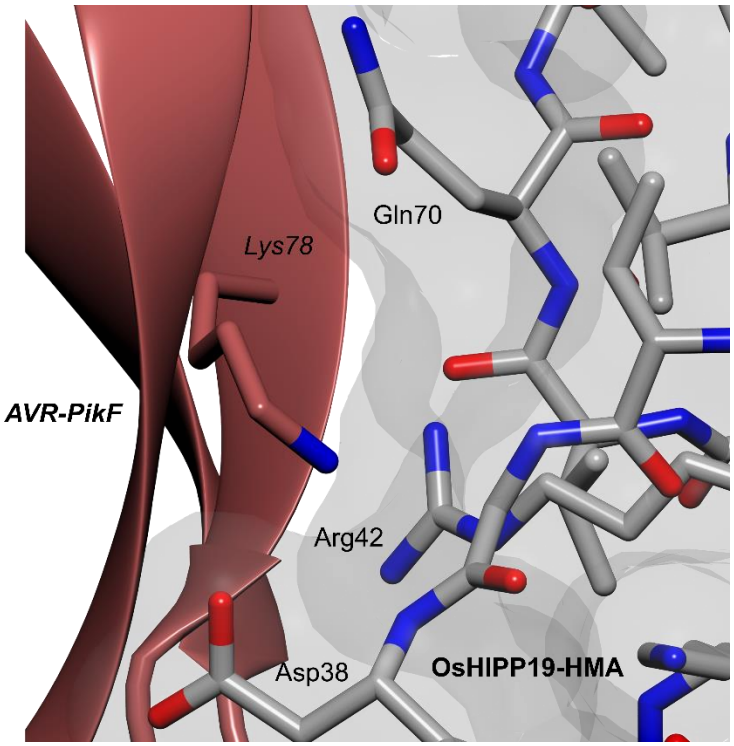


Figure 3.18 AVR-PikF^{Lys78} is at the binding interface.

The structure of AVR-PikF is represented as a red ribbon with AVR-PikF^{Lys78} shown as a cylinder. The structure of OsHIPP19 is represented as grey cylinders with the molecular surface of OsHIPP19 also displayed.

interacting with the NLR protein and triggering immune signalling. It might therefore be expected that the integrated HMA domains of Pikp-1 and Pikm-1 would have evolved to bind the effector more tightly than the HMA domain of the putative virulence target OsHIPP19. Somewhat surprisingly, our results show that AVR-PikD interacts with OsHIPP19-HMA with higher apparent affinity than with either Pikp-HMA or Pikm-HMA.

It is firstly important to acknowledge that the HMA domains are being studied in isolation, without the biological context of the full-length protein. Other Pik-1 domains may provide additional contacts with the effector which increase the affinity of the receptor for AVR-PikD. The rice NLR protein RGA5 also has an integrated HMA domain through which it recognises the *M. oryzae* effector AVR-Pia, however other domains also contribute to AVR-Pia binding (Ortiz et al., 2017). As the molecular details of the mechanism by which Pik-1 and Pik-2 cooperate to initiate immune signalling are unknown, Pik-2 may also influence the interaction between AVR-PikD and Pik-1.

While AVR-PikD interacts with Pikp-HMA and Pikm-HMA with lower apparent affinity than with OsHIPP19-HMA, the binding affinity is clearly sufficient to activate the NLR protein and trigger immune signalling. Therefore, higher affinity binding of the integrated HMA domain to the effector may not provide any additional evolutionary benefit. Furthermore, given that HIPPs/HPPs have been identified as susceptibility factors for other pathogens (Cowan et al., 2018, Fukuoka et al., 2009, Radakovic et al., 2018) it is not unreasonable to hypothesise that other effectors may also target HIPPs/HPPs. The Pik-1/Pik-2 NLR pair may recognise additional effectors outside the AVR-Pik allelic series, trading higher affinity binding to AVR-Pik for breadth of binding to other effectors.

The integrated domain hypothesis proposes that these domains have their origins in the host virulence targets of the effector (Cesari et al., 2013). It would therefore be expected that the effector binds in a similar manner to both its virulence target and to the integrated domain of the NLR protein. Consistent with this, I observed that the global binding interface between AVR-PikF and the HMA domain of the putative virulence target OsHIPP19 is virtually identical to that between AVR-Pik and the integrated HMA domains of Pikp-1 and Pikm-1. The ability to compare the interaction of an effector with

a putative host target and the integrated domain of a cognate NLR protein may make the Pik/AVR-Pik/HIPP system a good model to further probe the evolution and function of integrated domains in NLR proteins in the future.

The Ala67Asp mutation which distinguishes AVR-PikC from AVR-PikE, and the Met78Lys mutation which distinguishes AVR-PikF from AVR-PikA, are both located at the interface of the effector with the HMA domain. These mutations appear to be adaptive (Longya et al., 2019, De la Concepcion et al., 2018) and sufficient to prevent interaction with the integrated HMA domains of the various Pik-1 alleles. However, I found that these mutations do not disrupt the interaction between the effector and the HMA domain of the putative virulence target OsHIPP19, and that both AVR-PikC and AVR-PikF bind OsHIPP19-HMA with nanomolar affinity.

By solving the crystal structure of AVR-PikF bound to OsHIPP19-HMA, I have shown that the total interface area between the two proteins is larger than between any of the previously studied AVR-Pik/Pik-HMA complexes. Additionally, there are additional intermolecular hydrogen bonds and salt bridges between AVR-PikF and OsHIPP19-HMA at all three of the interfaces compared to AVR-PikA and Pikm-HMA. The crystal structure of the complex between OsHIPP19/AVR-PikF reveals that the side chain of AVR-PikF^{Lys78} adopts an unusual conformation, potentially due to steric clash of more favourable rotamers with OsHIPP19-HMA. This suggests that the interaction is maintained by additional intermolecular contacts between the proteins, particularly at interface 3, rather than compensatory mutations in the HMA domain to accommodate the AVR-PikF^{Lys78} side chain.

The emergence of an allelic series of AVR-Pik effectors is likely to have been driven by the deployment of rice varieties carrying Pik alleles (Kanzaki et al., 2012, Li et al., 2019a, Longya et al., 2019) AVR-PikC and AVR-PikF avoid binding to the integrated HMA domain of Pik-1, and therefore *M. oryzae* isolates carrying AVR-PikC and AVR-PikF overcome Pik-mediated blast resistance in the field. The work described in this chapter characterising the interaction between AVR-Pik alleles and the HMA domain of OsHIPP19 could be used to guide future efforts to engineer a Pik-1 variant with an HMA domain resembling that of OsHIPP19, to which these stealthy alleles bind. Such a variant may be capable of delivering resistance to *M. oryzae* carrying AVR-PikC and AVR-PikF.

4

Two *Magnaporthe oryzae* effectors bind OsHIPP19 at different interfaces

4

Two *Magnaporthe oryzae* effectors bind OsHIPP19 at different interfaces

4.1 Introduction

The *Magnaporthe oryzae* effector, AVR-Pia, was identified and cloned in the same association genetics study that identified AVR-Pik (Yoshida et al., 2009). The solution structure of the effector has been solved by two independent groups using NMR spectroscopy and found to consist of a six-stranded β-sandwich structure, leading to its classification as a member of the MAX effector family (Ose et al., 2015, Ortiz et al., 2017). Despite sharing just 17 % amino acid sequence identity, AVR-Pia and AVR-Pik adopt a similar core structure, though AVR-Pia lacks the N-terminal extension of AVR-PikD formed by amino acids Arg31 to Pro52 (figure 4.1).

Recognition of AVR-Pia is mediated by a pair of rice CC-NLR proteins, RGA4 and RGA5, which also recognise the sequence-unrelated MAX effector AVR1-CO39. Like Pik-1 and Pik-2, RGA4 and RGA5 are genetically linked; the genes are found in a head-to-head orientation under the control of a shared promoter on chromosome 11 of the rice genome (Cesari et al., 2013). RGA5 functions as a sensor NLR protein and has an integrated HMA domain (previously referred to as a related to ATX1, or RATX1, domain) at the C-terminus of the protein (figure 4.2a) (Cesari et al., 2013, Cesari et al., 2014b). The HMA domain shares 51 % amino acid identity with the integrated HMA domain of Pikp-1, and the domains appear to have been integrated into the two NLR proteins in two separate evolutionary events. RGA4 has the typical structure of a CC-NLR protein, with no integrated domain. In the absence of the effector, RGA5 represses the otherwise-active

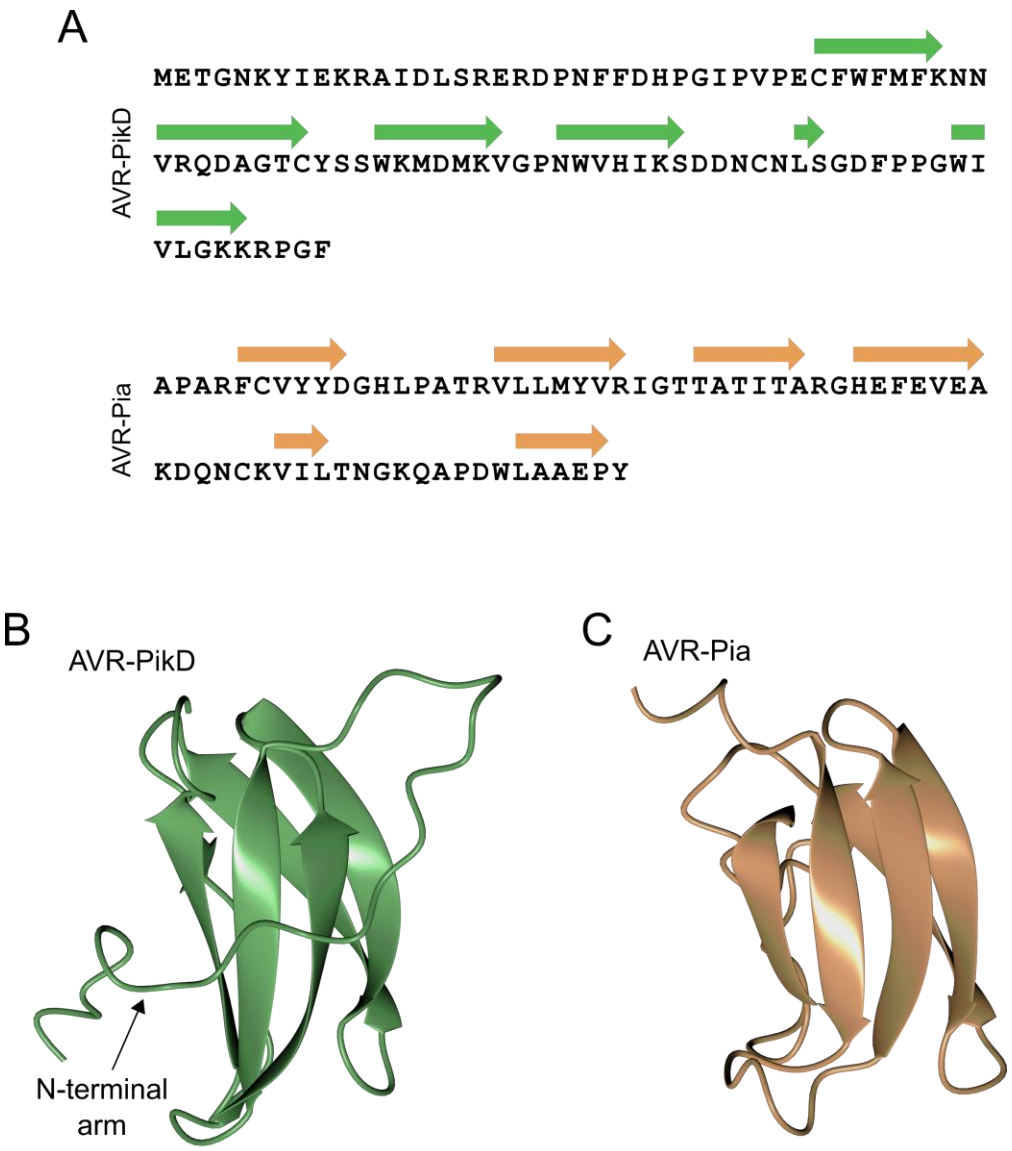


Figure 4.1 AVR-PikD and AVR-Pia share a common structural fold.

A. *Amino acid sequences of AVR-PikD (top) and AVR-Pia (bottom). β-strands are represented by coloured arrows above the amino acid sequence.* ***B.*** *Crystal structure of AVR-PikD (from PDB accession number 6G10, (De la Concepcion et al., 2019) with β-strands represented as green ribbons.* ***C.*** *Crystal structure of AVR-Pia (from PDB accession number 6Q76, (Varden et al., 2019)) with β-strands represented as orange ribbons.*

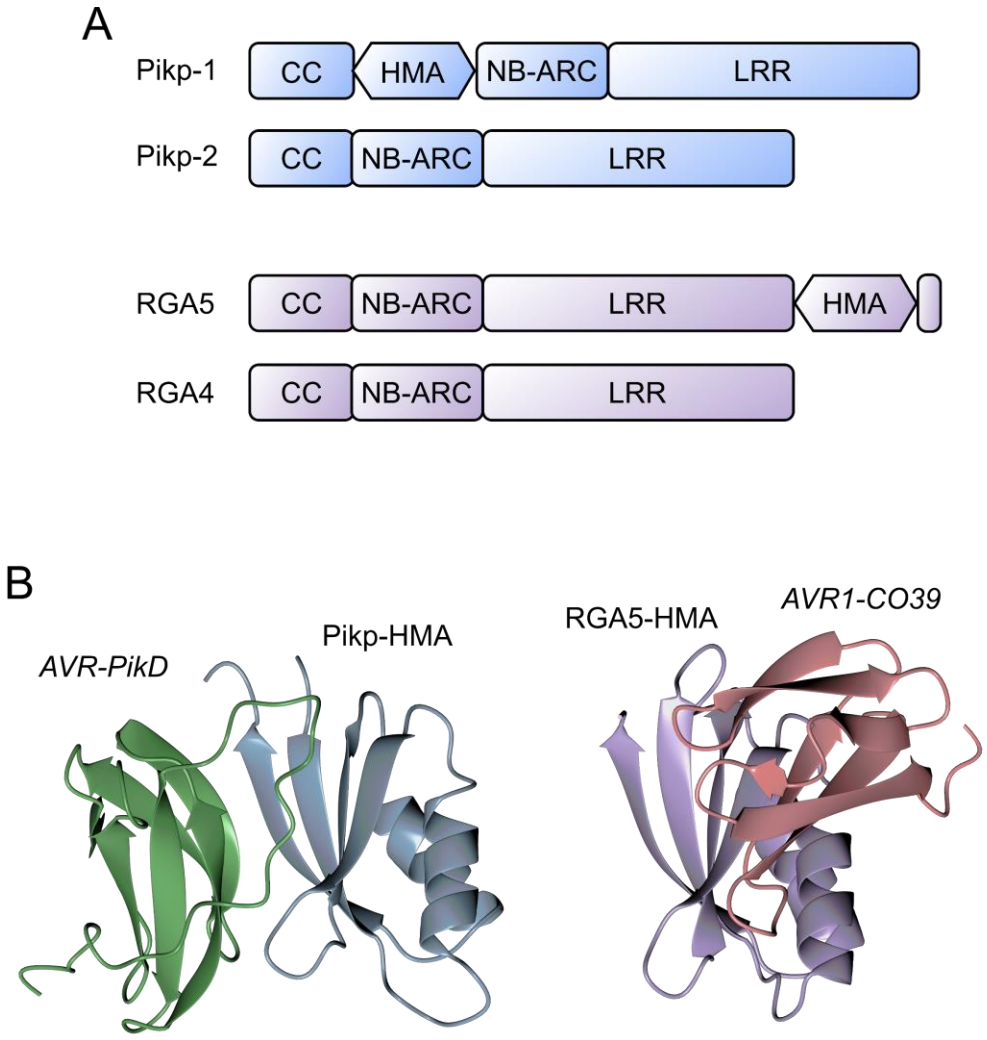


Figure 4.2 The integrated HMA domains of Pikp-1 and RGA5 bind their corresponding effectors at different interfaces.

A. *Schematic representations of the paired NLR proteins Pikp-1/Pikp-2 and RGA4/RGA5. CC = coiled coil, NB-ARC = nucleotide-binding Apaf-1, R protein, CED4-shared, LRR = leucine-rich repeat, HMA = heavy metal associated.* ***B.*** *Crystal structures of Pikp-HMA in complex with AVR-PikD (PDB accession code 6G10, (De la Concepcion et al., 2019)) and RGA5-HMA in complex with AVR1-CO39 (PDB accession code 5ZNG, (Guo et al., 2018)). Structures are represented as ribbons. Pikp-HMA and RGA5-HMA are shown in the same orientation to enable comparison of the effector binding interfaces.*

RGA4. AVR-Pia (or AVR1-CO39) directly interacts with the integrated HMA domain of RGA5, and this interaction is required to relieve the negative regulation by RGA5, enabling RGA4 to activate defence signalling (Cesari et al., 2014b).

The crystal structure of RGA5-HMA in complex with AVR1-CO39 (PDB accession code 5ZNG) revealed that, despite the structural similarity between the integrated HMA domains and between AVR1-CO39 and AVR-Pik, the two effectors bind their respective HMA domains at different interfaces (figure 4.2b). In the structure of AVR1-CO39 and RGA5-HMA, an antiparallel β-sheet forms between the four-stranded β-sheet of RGA5-HMA and β1, β2 and β6 of AVR1-CO39, mediated by the antiparallel alignment of β2 from RGA5-HMA and β2 from AVR1-CO39 (Guo et al., 2018). By contrast, in the structures of AVR-Pik in complex with integrated HMA domains, an antiparallel β-sheet forms between the β-sheet of the HMA domain and β3, β4 and β5 of the effector (Maqbool et al., 2015, De la Concepcion et al., 2018). Additionally, the interface between AVR1-CO39 and RGA5-HMA is mostly supported by main-chain hydrogen bonding, with very few side chain interactions (Guo et al., 2018).

Interestingly, Pikp-1 and Pikp-2 confer moderate resistance to *M. oryzae* isolates carrying AVR-Pia (Varden et al., 2019). AVR-Pia weakly interacts with the HMA domain of Pikp-1 *in vitro*, and the crystal structure of the complex (PDB accession code 6Q76) demonstrated that AVR-Pia interacts with Pikp-HMA at a similar interface to that observed for AVR1-CO39 with RGA5-HMA (Varden et al., 2019). This interface is, again, mostly comprised of backbone interactions, and the total interface area between AVR-Pia and Pikp-HMA is approximately half that between AVR-PikD and Pikp-HMA.

While there is currently no structure available for the complex between AVR-Pia and RGA5-HMA, it is reasonable to hypothesise that the complex is likely to involve a similar binding interface to the characterised interfaces between AVR-Pia/Pikp-HMA and AVR1-CO39/RGA5. This hypothesis is further supported by the NMR titration experiments performed to identify the HMA-interacting surface of AVR-Pia. Amino acids with recorded chemical shift changes from the unbound to RGA5-HMA bound state mostly localised to β2 and β3 of AVR-Pia (Ortiz et al., 2017), consistent with the interface involved in the interaction with Pikp-HMA.

AVR-Pia and AVR1-CO39 appear to interact more weakly with RGA5-HMA than AVR-PikD with Pikp-HMA. Isothermal titration calorimetry (ITC) experiments determined the equilibrium dissociation constants (K_D) for the interactions between AVR-Pia or AVR1-CO39 and RGA5-HMA as 1.8 μM and 5.4 μM, respectively (Ortiz et al., 2017), while the K_D for the interaction between AVR-PikD and Pikp-HMA has been determined (by surface plasmon resonance) to be less than 10 nM (Maqbool et al., 2015, De la Concepcion et al., 2018). It should be acknowledged that there are substantial differences between ITC and SPR; a key example being that for ITC, both proteins are in solution, while for SPR, one protein is tethered to a sensor chip. However, these results suggest that the interaction between AVR-Pia/AVR1-CO39 and RGA5-HMA is several orders of magnitude weaker than that between AVR-PikD with Pikp-HMA. AVR-Pia has been shown to associate with regions of RGA5 outside the HMA domain, as the effector immunoprecipitates with a truncated version of RGA5 lacking the C-terminal HMA domain (RGA5$_{\Delta HMA}$) (Ortiz et al., 2017). However, the association with RGA5$_{\Delta HMA}$ was not sufficient to relieve the RGA5-mediated repression of RGA4 and activate immune signalling, indicating that the interaction with the HMA domain, albeit weak, is crucial for recognition of AVR-Pia (Ortiz et al., 2017).

The integrated domain hypothesis proposes that integrated domains in NLR proteins have their origins in the host targets of the recognised effector. I hypothesised that AVR-Pia and AVR1-CO39 interact with rice proteins containing one or more HMA domains. Given that AVR-Pik interacts with OsHIPP19, I was interested to explore whether multiple *M. oryzae* effectors could interact with the same potential target.

In this chapter, I aimed to investigate whether AVR-Pia, like AVR-Pik, interacts with the HMA domain of OsHIPP19. Upon identifying that AVR-Pia interacts with OsHIPP19-HMA in analytical gel filtration experiments, I compared the relative binding affinity of the interaction with that observed between AVR-PikD and OsHIPP19-HMA. I hypothesised that AVR-Pia would interact with OsHIPP19-HMA at a comparable interface to that observed in the crystal structure of AVR-Pia in complex with Pikp-HMA and endeavoured to use X-ray crystallography to obtain structural information about the interaction between AVR-Pia and OsHIPP19-HMA.

4.2 Results

4.2.1 AVR-Pia interacts with OsHIPP19-HMA in analytical gel filtration

Dr Freya Varden (John Innes Centre) provided AVR-Pia (without the signal peptide) in pOPIN-E, and purified protein used for some of the experiments described in this chapter.

AVR-Pia was produced in *E. coli* and purified as described for AVR-PikD in chapter 3 (section 3.2.3). To test whether AVR-Pia interacts with the purified HMA domain of OsHIPP19, analytical gel filtration was used. When analysed alone, AVR-Pia eluted in a single peak at a volume of 15.0 ml. By contrast, when analysed following incubation with OsHIPP19-HMA in an equimolar ratio, two peaks were observed at 12.8 ml and 15.0 ml (figure 4.3a). The additional peak at an earlier elution volume indicates that the two proteins interact, as OsHIPP19-HMA absorbs light very poorly at 280 nm (discussed in section 3.2.1). SDS-PAGE confirmed that fractions corresponding to the peak at 12.8 ml contain both AVR-Pia and OsHIPP19-HMA (figure 4.3a). However, the additional peak at 15.0 ml indicates that the sample contains some uncomplexed AVR-Pia. Analytical gel filtration is a qualitative technique, however a significant peak for the uncomplexed sample can hint at a weak interaction between the two proteins.

4.2.2 The interaction between AVR-Pia and OsHIPP19-HMA is weaker than that between AVR-PikD and OsHIPP19-HMA

To quantitatively assess whether AVR-Pia binds to OsHIPP19-HMA with lower affinity than AVR-PikD, surface plasmon resonance was used. AVR-Pii, another *M. oryzae* effector which targets Exo70 proteins (Fujisaki et al., 2015), was used as a negative control. The effectors were produced with a C-terminal non-cleavable 6xHis affinity tag. As described in chapter 3 (section 3.2.4), each effector was immobilised on a Ni^{2+}-NTA chip via the 6xHis tag. Three different concentrations of OsHIPP19-HMA (2nM, 5nM and 20nM) were flowed over the chip, and the binding stability (R_{obs}, measured in response units (RU)) recorded and expressed as a percentage of the maximum theoretical response observable (%R_{max}). The %$R_{max,}$ was calculated using an assumed binding stoichiometry of 1:1 for each effector:OsHIPP19-HMA pair. Experiments were

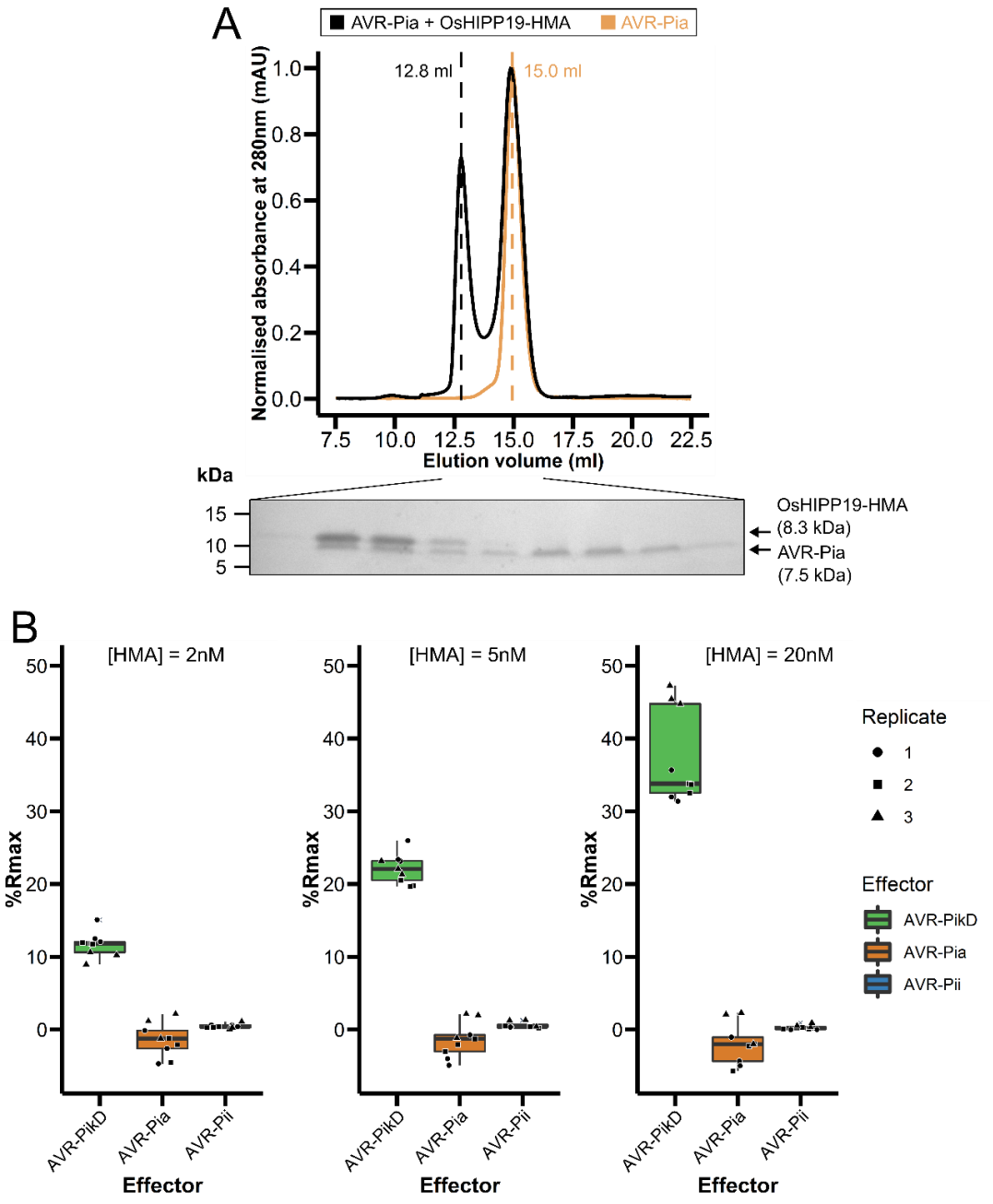


Figure 4.3 AVR-Pia interacts weakly with the HMA domain of OsHIPP19-HMA.

A. *Analytical gel filtration traces for AVR-Pia alone (orange) and AVR-Pia with OsHIPP19-HMA (black). The SDS-PAGE gel shows fractions collected from the gel filtration of AVR-Pia and OsHIPP19-HMA across the two peaks.* ***B.*** *OsHIPP19-HMA does not bind AVR-Pia at OsHIPP19-HMA concentrations ≤20 nM. %R_{max} is the percentage of the theoretical maximum response. The experiment was repeated three times, with each experiment containing three technical replicates. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).*

conducted in triplicate, and each of the three experiments consisted of three technical replicates. The $\%R_{max}$ observed at each of the three concentrations tested for the interaction between AVR-PikD and OsHIPP19-HMA was comparable to the values observed in previous experiments. As anticipated, no binding was detectable between AVR-Pii and OsHIPP19-HMA. Additionally, no binding was evident between AVR-Pia and OsHIPP19-HMA at any of the three concentrations investigated (figure 4.3b). Therefore, it can be concluded that the interaction between OsHIPP19-HMA and AVR-Pia is weaker than that between OsHIPP19-HMA and AVR-PikD.

4.2.3 The preliminary crystal structure of AVR-Pia/OsHIPP19-HMA indicates that AVR-Pia binds to OsHIPP19-HMA at a different interface to AVR-PikD

I aimed to use X-ray crystallography to determine the structural basis of the interaction between AVR-Pia and OsHIPP19-HMA, and investigate whether AVR-Pia binds to OsHIPP19-HMA at a comparable interface to that observed for the complex between AVR-Pia and Pikp-HMA (Varden et al., 2019).

4.2.3.1 Purification of the AVR-Pia/OsHIPP19-HMA protein complex

Dr Freya Varden provided pOPIN-S3C::AVR-Pia. Intact mass spectrometry was performed by Dr Gerhard Saalbach.

Co-expression of OsHIPP19-HMA and AVR-PikF was used to successfully produce and purify the complex between the two proteins. However, for the complex between OsHIPP19-HMA and AVR-Pia, different conditions were required for the successful production of each of the proteins in 1 L *E. coli* SHuffle cells (as detailed in section 2.4.8.3). As a result, the two proteins were produced and purified from *E. coli* cell lysate separately. *E. coli* SHuffle were transformed with either AVR-Pia in pOPIN-S3C, which produces AVR-Pia with a N-terminal 6xHis affinity tag and SUMO solubility tag, or OsHIPP19-HMA in pOPIN-M, which produces OsHIPP19-HMA with a N-terminal 6xHis affinity tag and MBP solubility/affinity tag. In both cases, the tags are cleavable by 3C protease. Following cleavage of the tagged protein, the tags were separated from the untagged proteins by affinity chromatography. The untagged proteins were then combined and incubated on ice to promote complex formation, after which the complex

was purified by gel filtration. The presence of both proteins in the final sample was confirmed by intact mass spectrometry.

4.2.3.2 Crystallisation of the protein complex

Crystals were harvested by Dr Clare Stevenson.

Initial crystallisation trials were carried out using the sitting drop vapour diffusion method with the purified complex of AVR-Pia and OsHIPP19-HMA at 10 mg/ml. Two commercial screens, Morpheus™ and JCSG *Plus*™ (Molecular Dimensions), were trialled. Crystals appeared in 5 different conditions in the JCSG *Plus*™ screen, and 3 in the Morpheus™ screen (table 4.1). Many of the crystals were fairly small, and while some looked suitable for X-ray diffraction studies, I aimed to improve the quality of the crystals by optimising the crystallisation conditions.

The complexity of the individual conditions in the Morpheus™ screen makes optimising crystals obtained in this screen challenging. However, the conditions that gave rise to crystals in the JCSG *Plus*™ screen contained fewer components and shared common features. All 5 conditions contained 20-25 % PEG 3350, 4 of the 5 included an ammonium salt (0.2 M), and 2 of the 5 used a 0.1 M BIS-TRIS buffer at pH 5.5. A 96-well optimisation screen was set up based around these parameters. Reservoir solutions were prepared containing PEG 3350 at a concentration between 17 % and 27 %, and 0.1 M BIS-TRIS buffer adjusted to a pH between 5.5 and 6.5. The reservoir solution for each condition contained 0.2 M ammonium chloride. The screen was set up with two different concentrations of the protein complex, 5 mg/ml and 10 mg/ml. The layout of the screen is shown in figure 4.4.

Improved crystals were obtained in multiple conditions in the optimisation screen (figure 4.4). Interestingly, many of the crystals were hexagonal. Four crystals were selected for harvesting from the optimisation screen set up with the protein complex at a concentration of 10 mg/ml. The crystals in the sitting drop are shown in figure 4.5a. Crystals 1 and 2 were taken from condition D12 (0.1 M BIS-TRIS pH 5.93, 27.1 % (w/v) PEG3350, 0.2 M ammonium chloride), while crystals 3 and 4 were harvested from condition A11 (0.1 M BIS-TRIS pH 5.5, 26.2 % (w/v) PEG 3350, 0.2 M ammonium chloride). 20% ethylene glycol in mother liquor was used as a cryoprotectant. Crystals

Table 4.1 Conditions in the commercial Morpheus™ and JCSG-plus™ screens which resulted in crystals of OsHIPP19-HMA/AVR-Pia.

Screen	Condition	Details
Morpheus™	D4	0.12M alcohols, 50% precipitant mix 4, 0.1M buffer system 1 pH 6.5
	H1	0.1M amino acids, 50% precipitant mix 1, 0.1M buffer system 1 pH 6.5
	H4	0.1M amino acids, 50% precipitant mix 4, 0.1M buffer system 1 pH 6.5
JCSG-*plus*™	A8	20% PEG 3350, 200mM ammonium formate
	A9	20% PEG 3350, 200mM ammonium chloride
	C3	20% PEG 3350, 200mM ammonium nitrate
	H3	25% PEG 3350, 0.1M BIS-TRIS pH 5.5
	H10	25% PEG 3350, 200mM ammonium acetate, 100mM BIS-TRIS pH 5.5

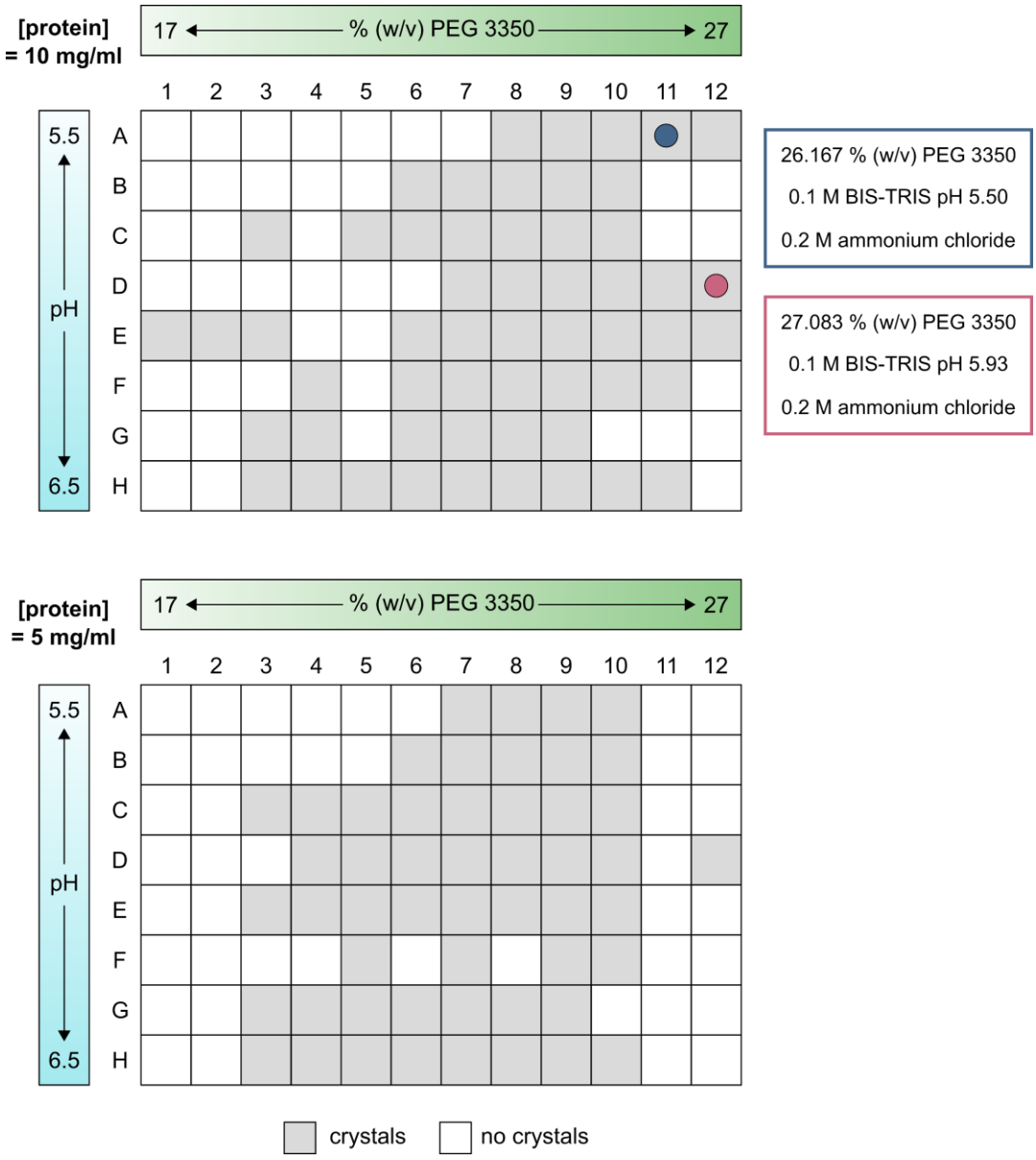


Figure 4.4 Optimisation of crystallisation conditions for OsHIPP19-HMA in complex with AVR-Pia.

Schematic representation of the 96-well screen set up to optimise crystallisation conditions for the OsHIPP19-HMA/AVR-Pia complex by varying the pH of a 0.1 M BIS-TRIS buffer and the concentration of PEG 3350. 0.1 M ammonium chloride was included in all conditions. Grey squares indicate conditions where crystals formed, white squares signify conditions with no crystals. The purple and pink circles indicate the conditions (detailed in the boxes with matching coloured outlines) which gave rise to the crystals selected for X-ray diffraction studies.

were individually mounted in a loop, flash frozen in liquid N_2 and transferred to the Diamond Light Source synchrotron facility for X-ray diffraction studies.

4.2.3.3 Data collection and processing

X-ray diffraction data was collected with guidance from Dr Clare Stevenson and Professor David Lawson. Professor David Lawson gave significant help and advice with data processing, as highlighted throughout the following section.

X-ray diffraction data was collected at beamline I04 of the Diamond Light Source synchrotron facility. Four datasets from three crystals (see table 4.2) were collected at a wavelength of 0.98 Å, each comprising 7200 images with an oscillation angle of 0.05°. Figure 4.5b shows each of the crystals mounted in a loop immediately prior to data collection.

The ISpyB data collection user interface for the Diamond Light Source passes datasets to several autoprocessing pipelines for automatic data reduction. The different pipelines use subtly different algorithms, and depending on the dataset, some pipelines may fail to appropriately process the data. Typically, the pipelines will provide similar estimates for the unit cell parameters and space group. However, for each of the four datasets, the outputs differed considerably between pipelines (table 4.2).

Professor David Lawson measured distances between reflections in the same plane to estimate two of the unit cell lengths as approximately 30 Å and 420 Å. In certain images, only every 6th reflection was present in the plane corresponding to the 420 Å unit cell axis. These systematic absences are indicative of a 6_1 or 6_5 screw axis. Two of the autoprocessed solutions estimated the space group as $P6_1\,2\,2$ or $P6\,2\,2$, with two different unit cell sizes ($a = 20$, $b = 20$, $c = 420$, or $a = 35$, $b = 35$, $c = 420$). Calculation of the solvent content with a single copy of a 1:1 complex between AVR-Pia and OsHIPP19-HMA (total molecular weight of 15.8 kDa) in the asymmetric unit allows the exclusion of the first unit cell ($a = b = 20$) as the expected solvent content for $P6_1\,2\,2$ / $P6_5\,2\,2$ is negative, and only 18% for $P6_1/P6_5$. The solvent content of protein crystals is typically around 43 % but can range between 27 % and 65 % (Matthews, 1968). For the second unit cell ($a = b = 35$), one copy of the complex in the asymmetric unit gives a solvent content of 44 %

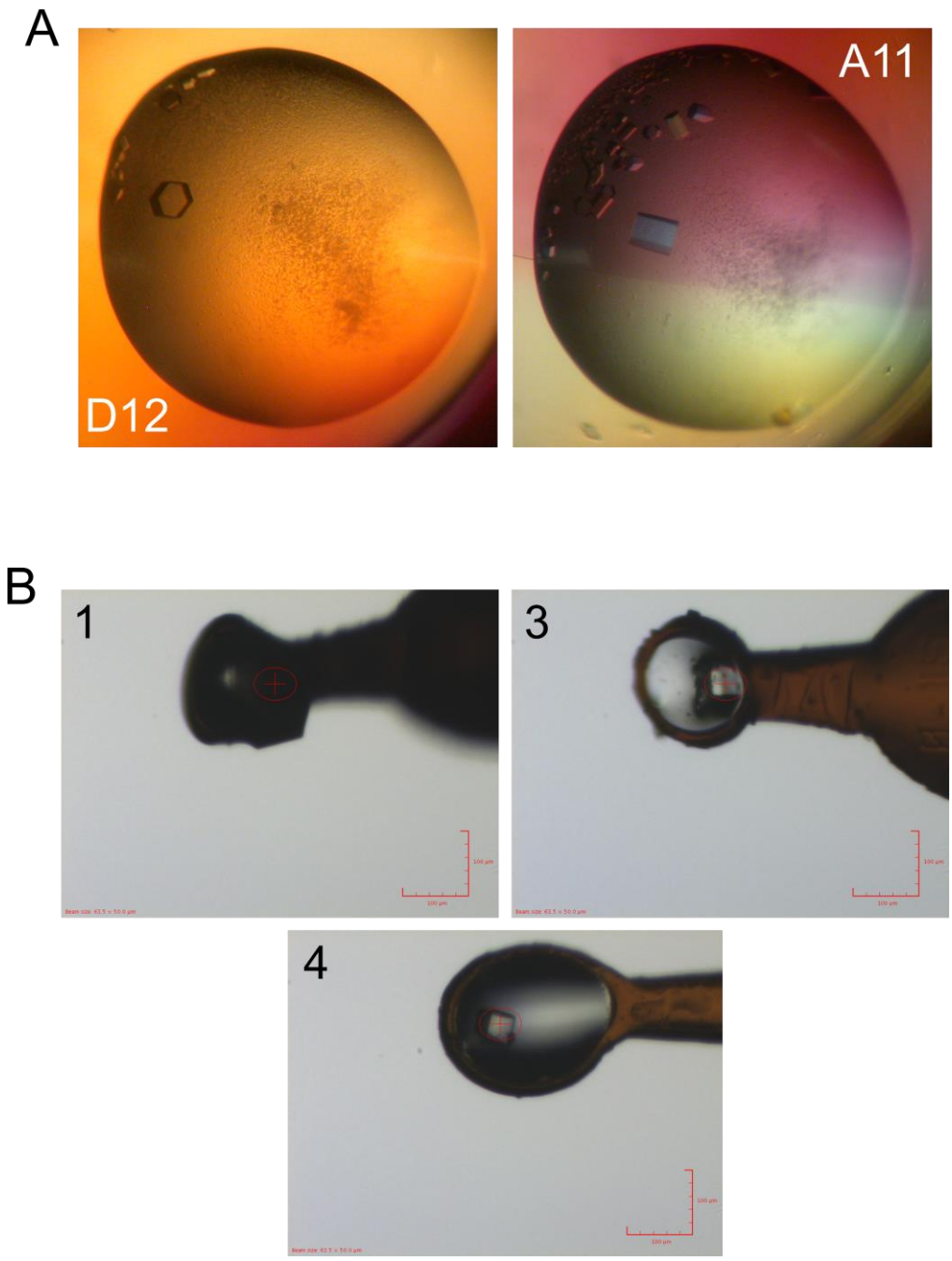


Figure 4.5 Crystals of OsHIPP19-HMA in complex with AVR-Pia.

A. *Crystals in the sitting drops from conditions D12 and A11 of the 96-well optimisation screen.* ***B.*** *Crystals 1, 3 and 4 mounted in loops prior to X-ray diffraction data collection. Crystal 1 is the large hexagonal crystal shown in sitting drop D12. Crystals 3 and 4 are smaller crystals from sitting drop A11. Crystal numbers correspond to those in table 4.2. No dataset was collected from crystal 2 (also from sitting drop D12), so it is not shown in the loop.*

for the space group $P6_1\,2\,2$ / $P6_5\,2\,2$. This supported using these unit cell and space group estimates for further processing.

The dataset from crystal 3 was reprocessed with xia2 using the DIALS pipeline, specifying the space group as $P\,6\,2\,2$ and the unit cell parameters as $a = b = 35$, $c = 420$, $\alpha = \beta = 90$, $\gamma = 120$. The scaled but unmerged data file was passed to AIMLESS (as implemented in CCP4i2) for merging. Based on the merging statistics, the maximum resolution was limited to 2.2 Å, giving an overall R_{merge} of 13.2 %. Data processing statistics are shown in table 4.3.

The phases were solved by molecular replacement using PHASER (McCoy et al., 2007), as implemented in CCP4i2. The crystal structures of AVR-Pia (PDB accession number 6Q76) and OsHIPP19-HMA (from the structure of OsHIPP19-HMA in complex with AVR-PikF solved in chapter 3) were supplied as models. A unique solution was found in the space group $P\,6_5\,2\,2$. Iterative rounds of manual rebuilding and refinement were carried out using COOT (Emsley et al., 2010) and REFMAC (Murshudov et al., 2011). TLS parameters were included for automatically defined TLS groups and refined over 5 cycles prior to the main refinement. However, following multiple cycles of refinement, certain regions of the electron density map remained highly fragmented, particularly on the surfaces away from the binding interface. It was not possible to accurately position parts of both protein chains into the density, and the R-factor and free R-factor could not be reduced below 30 %.

The data was reprocessed with xia2 using the DIALS pipeline, selecting the space group as $P6$ and the unit cell parameters as $a = b = 35$, $c = 420$, $\alpha = \beta = 90$, $\gamma = 120$. Merging and molecular replacement were carried out as described for the data processed with the space group $P\,6\,2\,2$. Data processing statistics are shown in table 4.3. A unique solution was found, with two copies of the complex in the asymmetric unit. However, once again, the electron density map was fragmented, and the R-factor and free R-factor stalled above 30 %.

Professor David Lawson identified an issue with the spot finding and indexing of the reflections during data reduction with DIALS. Approximately half of reflections were not indexed by the program. The dataset was passed to Dr Graeme Winter (Diamond

Table 4.2 Estimated space group (SG) and unit cell dimensions (UCD) by different autoprocessing pipelines for the four diffraction datasets generated.

Crystal	Dataset	Autoprocessing pipeline							
		AutoProc		FastDP		Xia2 3dii		Xia2 DIALS	
		SG	UCD	SG	UCD	SG	UCD	SG	UCD
1	1	*F*2 2 2	a = 35 b = 60 c = 421	*F*2 2 2	a = 35 b = 60 c = 421	$P2_1\ 2_1\ 2_1$	a = 35 b = 60 c = 421	*C*1 2 1	a = 35 b = 60 c = 420
	2	*C*1 2 1	a = 60 b = 35 c = 66	*H*3 2	a = 35 b = 35 c = 421	*F*2 2 2	a = 35 b = 60 c = 421	$P6_1\ 2\ 2$	a = 20 b = 20 c = 421
3	1	*F*2 2 2	a = 35 b = 60 c = 421	*C*1 2 1	a = 35 b = 60 c = 103	$P2_1\ 2_1\ 2_1$	a = 35 b = 60 c = 419	*C*1 2 1	a = 60 b = 35 c = 418
4	1	*F*2 2 2	a = 35 b = 60 c = 421	*P*6 2 2	a = 35 b = 35 c = 420	*F*2 2 2	a = 35 b = 60 c = 431	$P2\ 2\ 2_1$	a = 17 b = 30 c = 421

Table 4.3 Data processing statistics for the dataset collected from crystal 3 with each of the three space groups/unit cell dimensions trialled.

Data processing statistics			
Space group	$P6_5\,2\,2$	$P6$	$C\,1\,2\,1$
Cell dimensions *a*, *b*, *c* (Å)	34.87, 34.87, 420.83	34.87, 34.87, 420.83	60.21, 34.77, 418.14
Resolution (Å)*	35.07-2.20 (2.27-2.20)	35.07-2.20 (2.27-2.20)	34.82-2.00 (2.05-2.00)
R_{merge} (%)	13.2 (82.5)	12.8 (78.3)	11.9 (48.5)
$I/\sigma I$	12.8 (3.0)	5.6 (2.8)	6.5 (2.3)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
Unique reflections	8973 (702)	14712 (1231)	59711 (4470)
Redundancy	33.0 (35.2)	20.1 (20.1)	6.4 (6.6)
CC(1/2) (%)	99.7 (99.0)	99.7 (98.7)	99.3 (95.5)

* The highest resolution shell is shown in parenthesis.

Light Source), who developed the DIALS pipeline for data reduction. He was able to process the dataset more satisfactorily in the space group *C*1 2 1 with unit cell parameters a = 60, b = 35, c = 420, α = 90, β = 92, γ = 90.

The scaled but unmerged dataset provided by Dr Graeme Winter was supplied to AIMLESS as implemented in CCP4i2 for merging. Merging statistics supported processing the data to a maximum resolution of 2.0 Å. Molecular replacement was carried out as described earlier, and a unique solution was found with six copies of the complex in the asymmetric unit in a linear arrangement (figure 4.6). Interestingly, the electron density map for the copy of the complex at one end of the asymmetric unit is very good, and the majority of the residues can be satisfactorily modelled in the density. By contrast, the electron density for the copy of the complex at the other end of the asymmetric unit is highly fragmented, and no residues can be positioned with confidence (figure 4.6). This suggests that the space group and/or unit cell parameters may still be inaccurate.

4.2.3.4 Preliminary structure of AVR-Pia in complex with OsHIPP19-HMA.

While the quality of the model is not sufficient to form conclusions about the positions of individual residues or the hydrogen bonds that may exist between them, I can conclude that AVR-Pia binds to OsHIPP19-HMA at a similar interface to that observed for AVR-Pia to Pikp-HMA and AVR1-CO39 to RGA5-HMA. The effector binds to OsHIPP19-HMA at an interface formed by β2 and α1 of the HMA domain, with β2 of the effector aligned in an antiparallel direction to β2 of the HMA domain (figure 4.6). The interface between Pikp-HMA/AVR-Pia and RGA5-HMA/AVR1-CO39 is predominantly composed of main chain hydrogen bonds between β2 of the effector and β2 of the HMA domain; while it is not possible to reliably determine the hydrogen bonding pattern between OsHIPP19-HMA and AVR-Pia, it seems likely that similar hydrogen bonds between β2 of AVR-Pia and β2 of OsHIPP19-HMA contribute to the interface. This would result in a continuous β-sheet being formed by the four β-strands of the HMA domain and β1, β2 and β6 of AVR-Pia.

Importantly, AVR-Pia binds to OsHIPP19-HMA at a different interface to that targeted by AVR-Pik (figure 4.7). AVR-Pik binds to OsHIPP19-HMA at the side of the HMA

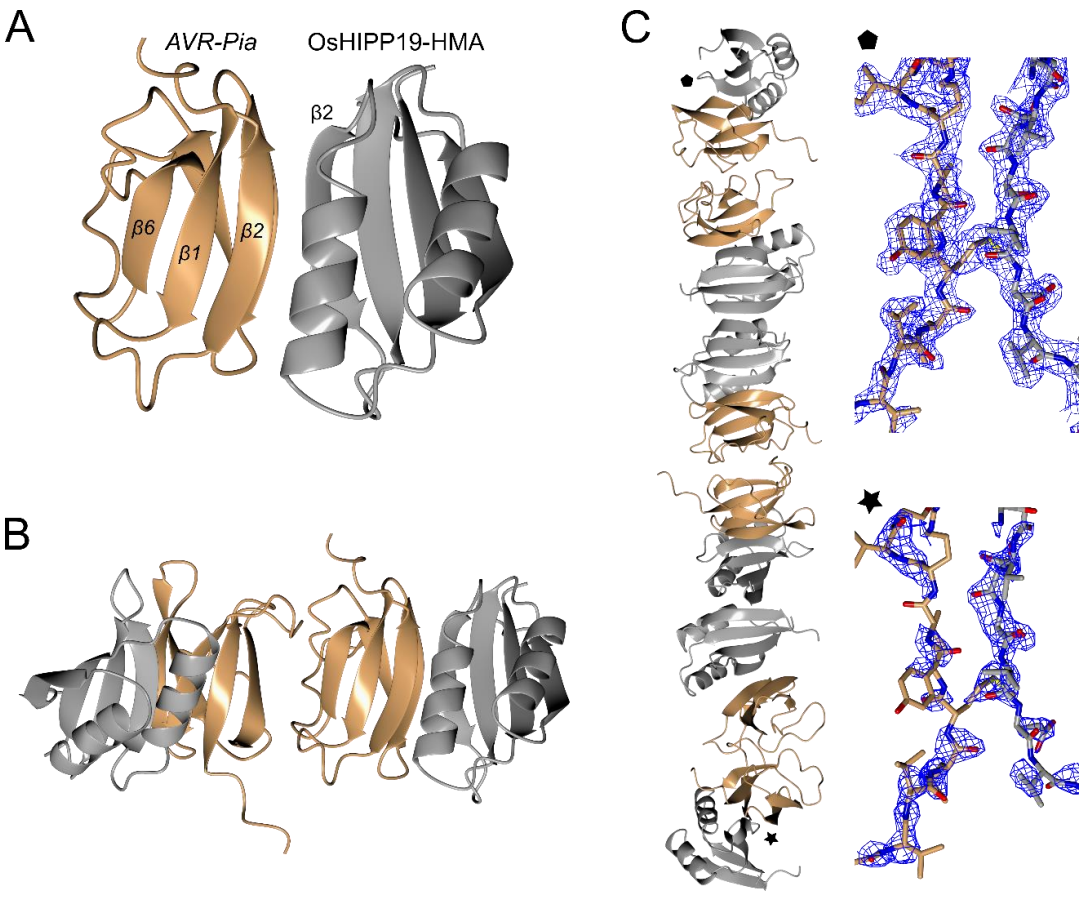


Figure 4.6 Preliminary models of OsHIPP19-HMA in complex with AVR-Pia.

OsHIPP19-HMA is represented as a grey ribbon, and AVR-Pia as an orange ribbon. ***A.*** *Model obtained from diffraction data processed in the space group* P6_5 *2 2 (one copy of the complex in the asymmetric unit). Relevant secondary structure features are indicated.* ***B.*** *Model obtained from diffraction data processed in the space group* P*6 (two copies of the complex in the asymmetric unit).* ***C.*** *Model obtained from diffraction data processed in the space group* C *1 2 1 with six copies of the complex arranged linearly in the asymmetric unit. The interface between AVR-Pia and OsHIPP19-HMA and the corresponding electron density are shown for the copies at the ends of the asymmetric unit. The structures are shown as cylinders, and the electron density is represented as blue mesh.*

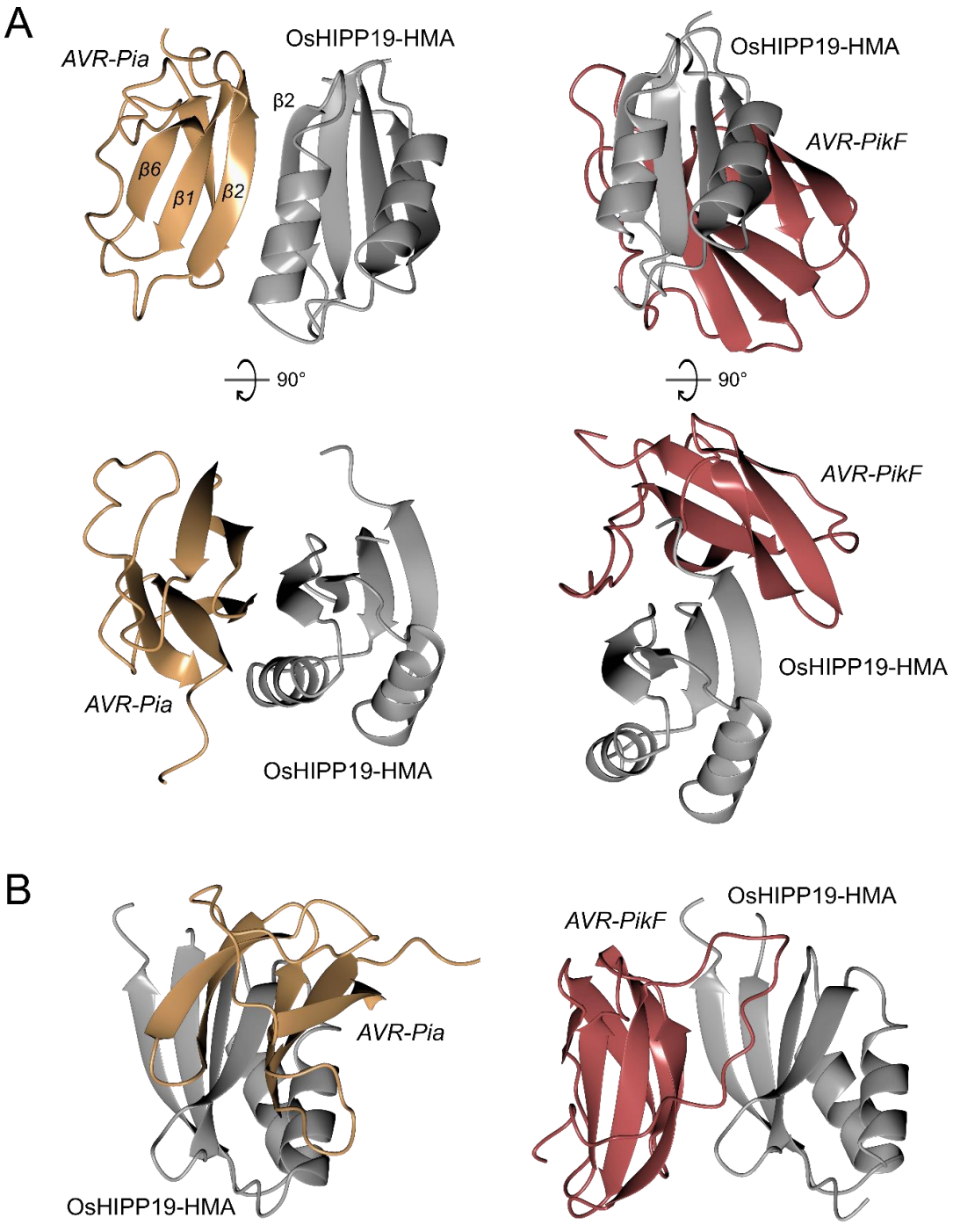


Figure 4.7 AVR-Pia and AVR-PikF bind to OsHIPP19-HMA at different interfaces.

Comparison between the structures of OsHIPP19-HMA in complex with AVR-PikF and in complex with AVR-Pia. OsHIPP19-HMA is represented as a grey ribbon, AVR-PikF as a red ribbon and AVR-Pia as an orange ribbon. ***A.*** *The crystal structures in two orientations, with a 90° rotation between views.* ***B.*** *The position of AVR-Pia relative to the "classical" view of the AVR-Pik/HMA domain complex.*

domain characterised by the four-stranded antiparallel β-sheet, with residues from each of the four β-strands contributing to the interaction with the effector. Figure 4.7 illustrates the different binding sites of the two effectors; OsHIPP19-HMA is shown in the same orientation for each side-by-side comparison.

4.3 Discussion

In this chapter, I have demonstrated that AVR-Pia interacts with OsHIPP19-HMA, albeit weakly. While I was able to obtain X-ray diffraction datasets from crystals of the protein complex, challenges in data processing have so far prevented satisfactory refinement of the model. However, the available electron density maps support the hypothesis that the interface between AVR-Pia and OsHIPP19-HMA is similar to the interface between AVR-Pia and Pikp-HMA, and markedly different to that between AVR-PikF and OsHIPP19-HMA.

Each of the three electron density maps contained well-defined regions, where individual amino acid side chains could be clearly positioned. However, the maps also contained regions of fragmented electron density where the protein main chain cannot be positioned with confidence. It seems likely that the space group has not been correctly identified. Pseudosymmetry, where non-crystallographic symmetry is mistaken for crystallographic symmetry, can lead to uncertainty about the true space group of a crystal. Additionally, the dataset shows evidence of diffraction anisotropy. While the REFMAC refinement pipeline includes anisotropic scaling algorithms which can address moderate anisotropy, severe anisotropy can limit the quality of the electron density map and cause the R factor and free R factor to stall. There were also substantial issues with finding and indexing reflections during data reduction; collecting data with the detector positioned further back from the crystal should enable the diffraction spots to be better resolved, which would improve indexing. This is particularly important along the *c*-axis where reflections are close together. An alternative approach would be to screen different crystallisation conditions with the aim of obtaining a different crystal form, as a crystal with different unit cell parameters and/or space group may improve downstream processing. Improving the quality of the model would allow a more detailed analysis of the interface between OsHIPP19-HMA and AVR-Pia.

The interaction between AVR-Pia and OsHIPP19-HMA appears to be weaker than that between AVR-Pik and OsHIPP19-HMA; while the two proteins form a complex when incubated together on ice, no binding was observed between AVR-Pia and OsHIPP19-HMA in surface plasmon resonance experiments. One limitation of surface plasmon resonance is that the technique requires one protein to be tethered to the sensor chip. For this experiment, AVR-Pia, AVR-Pik and AVR-Pii were each produced with a C-terminal 6xHis affinity tag, via which they were immobilised on a Ni^{2+}-NTA chip. This process can affect the surface available for binding of the analyte. SPR has been used extensively to study the interactions between AVR-Pik effectors and HMA domains, and the tagging and immobilisation of AVR-Pik does not compromise the interaction of the effector with the HMA domain. However, while AVR-Pia interacts with OsHIPP19-HMA at a different interface, previous work has found an interaction between AVR-Pia and Pikp-HMA in surface plasmon resonance experiments (Varden et al., 2019). As the interface between AVR-Pia and Pikp-HMA is comparable to that between AVR-Pia and OsHIPP19-HMA, it is reasonable to conclude that the region of AVR-Pia responsible for the interaction with HMA domains is available for binding during SPR experiments.

Another possibility is that the concentrations of OsHIPP19-HMA were too low to observe binding. The reported K_D for the interaction between AVR-Pia and RGA5-HMA, as determined by isothermal titration calorimetry, was 1.8 μM. If binding to OsHIPP19-HMA is in a similar, micromolar, range, then 20nM OsHIPP19-HMA is unlikely to be sufficient to observe an interaction by SPR. Unfortunately, at concentrations >25 nM, OsHIPP19-HMA binds to the reference flow cell, interfering with measurement of the association and dissociation of the HMA domain and the bound effector. An alternative technique, such as isothermal titration calorimetry, may represent a more viable method to quantify the binding affinity between OsHIPP19-HMA and AVR-Pia.

OsHIPP19 consists of a N-terminal HMA domain and a C-terminal proline-rich region, which is predicted to be disordered. Proline-rich regions are often associated with protein-protein or protein-ligand interactions. The C-terminal portion of OsHIPP19 may have a role in stabilising the interaction between AVR-Pia and the HMA domain and/or increasing the binding affinity. Intriguingly, a proline-rich region of comparable

length is present at the C-terminus of RGA5, immediately following the integrated HMA domain. It should be noted that the C-terminus of OsHIPP19-HMA is at the opposite side of the protein to the AVR-Pia binding interface, though without structural information about the proline-rich region it is not possible to draw conclusions about its position relative to the effector. Unfortunately, attempts to produce and purify full-length OsHIPP19, or OsHIPP19 lacking the putative isoprenylation motif at the extreme C-terminus of the protein, from *E. coli* have so far been unsuccessful.

Over 100 HIPPs and HPPs have been identified in rice to date (de Abreu-Neto et al., 2013) (Ryohei Terauchi, personal communication). AVR-Pia may target a HIPP/HPP other than OsHIPP19 and binds to the HMA domain of that HIPP/HPP with higher affinity than to OsHIPP19-HMA. The interaction between AVR-Pia and OsHIPP19-HMA may be non-specific and not biologically relevant. Alternatively, a low affinity interaction with OsHIPP19-HMA may be sufficient for the effector to fulfil its virulence function. The roles of AVR-Pik and AVR-Pia in promoting pathogen infection remain unknown, and work is ongoing to identify their mechanism of action.

Given the limited number of effectors for whom interacting partners have been identified, the finding that three *M. oryzae* effectors (AVR-Pik, AVR-Pia and AVR1-CO39) each interact with small HMA domain-containing proteins is quite striking. Several HIPPs have been identified as susceptibility factors for infection; NbHIPP26 is required for long-distance movement of the potato mop-top virus (Cowan et al., 2018), AtHIPP27 is a susceptibility gene for beet cyst nematode infection (Radakovic et al., 2018), and OsHIPP05 (Pi21) is a susceptibility gene for rice blast disease (Fukuoka et al., 2009). It is tempting to speculate that HIPPs are targeted by effectors from multiple pathogens. Understanding the function of OsHIPP19, and the consequence of interaction with AVR-Pik and AVR-Pia during rice blast disease, could shed light on the role of HIPPs in disease and immunity.

5

Structure-guided engineering of a novel Pik-1 variant with extended recognition capability

5

Structure-guided engineering of a novel Pik-1 variant with extended recognition capability

5.1 Introduction

The Pik NLR proteins confer resistance to *Magnaporthe oryzae* carrying certain AVR-Pik effector alleles. However, the stealthy effector alleles AVR-PikC and AVR-PikF are not recognised by any of the Pik alleles that have been identified and functionally characterised to date. Previous work has demonstrated that AVR-PikC and AVR-PikF avoid binding to the integrated HMA domain of Pikp-1 and Pikm-1 and are thus able to evade Pikp- or Pikm-triggered immunity (De la Concepcion et al., 2018, Longya et al., 2019). In chapter 3 of this thesis, I demonstrated that all AVR-Pik alleles, including AVR-PikC and AVR-PikF, interact with the HMA domain of the putative host target OsHIPP19. I hypothesised that if I can modify the HMA domain of Pikp-1 to resemble the HMA domain of OsHIPP19, I could engineer an NLR protein which binds to both AVR-PikC and AVR-PikF and activates immunity in response to *M. oryzae* isolates carrying either of these effector alleles.

Structural information can inform efforts to engineer proteins with specific properties. In the context of plant-pathogen interactions, this may include receptors with potential to recognise a more diverse range of effector proteins, or virulence targets capable of evading manipulation by effectors. This concept is exemplified by the engineering of a modified jasmonate receptor protein with reduced sensitivity to coronatine, a jasmonic acid (JA)-mimicking phytotoxin produced by *P. syringae*. The mutated COI1 protein

remains capable of maintaining endogenous jasmonate signalling through its interaction with JA (Zhang et al., 2015b). The crystal structure of the COI1-JAZ co-receptor was used to identify residues in the JA-Ile/coronatine binding pocket, and an alanine-to-valine mutation at position 384 in COI1 was predicted to prevent coronatine binding without affecting JA binding. COI1^{A384V} transgenic *A. thaliana* plants showed increased resistance to coronatine-producing strains of *Pseudomonas syringae*, consistent with the expected reduction in coronatine binding of COI1^{A384V}. Crucially, they also possessed a level of defence to chewing insects comparable with transgenic WT COI1 and Col-0 plants, demonstrated JA-induced root growth inhibition, and retained male fertility; these responses indicate the maintenance of JA signalling (Zhang et al., 2015b). The study demonstrates that structural information can inform targeted mutagenesis to produce modified proteins capable of evading pathogen effectors.

Previous work has demonstrated that structure-guided engineering can expand the recognition profile of Pikp-1. While Pikp only recognises the AVR-PikD effector allele, Pikm recognises AVR-PikD, AVR-PikE and AVR-PikA. Comparisons between the crystal structures of Pikp-HMA and Pikm-HMA in complex with the different effector alleles identified key differences at the binding interfaces which determine the recognition specificity of the NLR protein (De la Concepcion et al., 2018). Replacing two amino acids in Pikp-1 with the corresponding amino acids of Pikm-1 (Asn2612Lys and Lys262Glu, giving Pikp-1^{NK-KE}) was sufficient to extend the recognition specificity of Pikp to AVR-PikE and AVR-PikA (De la Concepcion et al., 2019). Pikp-1^{NK-KE} essentially recapitulates the recognition profile of Pikm, but provides an important proof-of-concept that structure-guided engineering can extend the recognition specificity of an NLR protein, highlighting the potential to engineer recognition to effectors not currently detected in nature.

In this chapter, I used a structure-guided approach to engineer a Pik-1 variant which triggers cell death in response to the AVR-PikC and AVR-PikF in a model system. Exchanging the HMA domain of Pikp-1 for the HMA domain of OsHIPP19 resulted in an autoactive NLR protein which triggered cell death in the absence of an effector. However, by making a targeted point mutation, informed by the crystal structure of OsHIPP19-HMA in complex with AVR-PikF, in the Pikp-1^{NK-KE} background, I

engineered a Pik-1 variant which is not autoactive and delivers an AVR-PikC/AVR-PikF-dependent cell death response in the model system *Nicotiana benthamiana*.

5.2 Results

5.2.1 AVR-PikF does not trigger Pikp/Pikm-mediated cell death in *Nicotiana benthamiana*.

This section includes my contribution to (Longya et al., 2019).

Constructs for in planta *expression (Pikp-1:HF, Pikp-2:6xHA, Pikm-1:HF, Pikm-2:6xHA and Myc:AVR-PikD) were provided by Dr Marina Franceschetti, Hannah Langlands and Juan Carlos De la Concepcion (John Innes Centre). The Western blot to check protein production in* Nicotiana benthamiana *was carried out by Apinya Longya (John Innes Centre / Kasetsart University).*

Nicotiana benthamiana is well-established as a model system to investigate the response of NLR proteins to specific effectors. Proteins of interest are transiently expressed in *N. benthamiana* by infiltrating a suspension of *Agrobacterium tumefaciens* cells carrying the relevant genes. If NLR protein(s) recognise the effector, programmed cell death (characteristic of a hypersensitive response) occurs at the infiltration site, with necrotic tissue typically visible 5 days post-infiltration. In addition, a characteristic feature of HR-like cell death is the accumulation of phenolic compounds, which auto-fluoresce upon exposure to ultra-violet (UV) light.

Advantages of this model system are that *N. benthamiana* plants can be grown comparatively cheaply and easily, and that protocols for transient expression of genes by agroinfiltration in *N. benthamiana* are well established. Furthermore, this system enables screening of large numbers of different combinations of NLR proteins and effectors with reasonable efficiency.

To date, resistance of rice cultivars to *M. oryzae* isolates has correlated with observed HR-like cell death following transient expression of NLR/effector combinations in *N. benthamiana* (De la Concepcion et al., 2018, Maqbool et al., 2015). For example, a rice cultivar containing Pikm is resistant to *M. oryzae* isolates carrying either AVR-PikD,

AVR-PikE or AVR-PikA. Correspondingly, HR-like cell death is observed when Pikm-1/Pikm-2 are transiently co-expressed with AVR-PikD, AVR-PikE or AVR-PikA, in *N. benthamiana* (De la Concepcion et al., 2018).

Rice cultivars containing Pikp (IRBLKP-K60) or Pikm (IRBLKM-TS) are susceptible to *Magnaporthe oryzae* isolates carrying AVR-PikF (Longya et al., 2019). To investigate whether Pikp-1/Pikp-2 and Pikm-1/Pikm-2 trigger cell death when co-expressed with AVR-PikF in *N. benthamiana,* AVR-PikF was cloned with an N-terminal 4xmyc epitope tag into a level 1 Golden Gate binary vector (Weber et al., 2011, Engler et al., 2014). The paired NLR proteins (Pikp-1:HF/Pikp-2:HA or Pikm-1:HF/Pikm-2:HA) were transiently co-expressed with either myc:AVR-PikD or myc:AVR-PikF by agroinfiltration in *N. benthamiana*. The assay was repeated with 20 leaves in each of three independent experiments. The leaves were photographed 5 days post-infiltration (dpi), and the extent of cell death scored on a seven-point scale (see figure 2.2).

Neither Pikp-1/Pikp-2 nor Pikm-1/Pikm-2 triggered cell death when transiently co-expressed with AVR-PikF (figure 5.1). As anticipated, based on previous results from the lab (De la Concepcion et al., 2018, Maqbool et al., 2015), both Pikp-1/Pikp-2 and Pikm-1/Pikm-2 triggered HR-like cell death when co-expressed with AVR-PikD. This is evidenced by the presence of necrotic tissue and strong autofluorescence under UV light in the infiltrated region. To confirm that the lack of cell death when myc:AVR-PikF was co-expressed with the Pik NLR proteins was not due to poor production of the myc:AVR-PikF protein in *N. benthamiana*, the presence of each of the proteins in the infiltrated tissue was confirmed by Western blot (figure 5.1c).

Crucially, these results demonstrate a continued correlation between blast resistance in rice and HR-like cell death in *N. benthamiana*. This correlation supports the use of the cell death assay in *N. benthamiana* for screening engineered NLR proteins for extended response to AVR-PikC and AVR-PikF, as engineered NLR proteins which trigger cell death when co-expressed with these effector alleles can be considered good candidates for delivering blast resistance in rice.

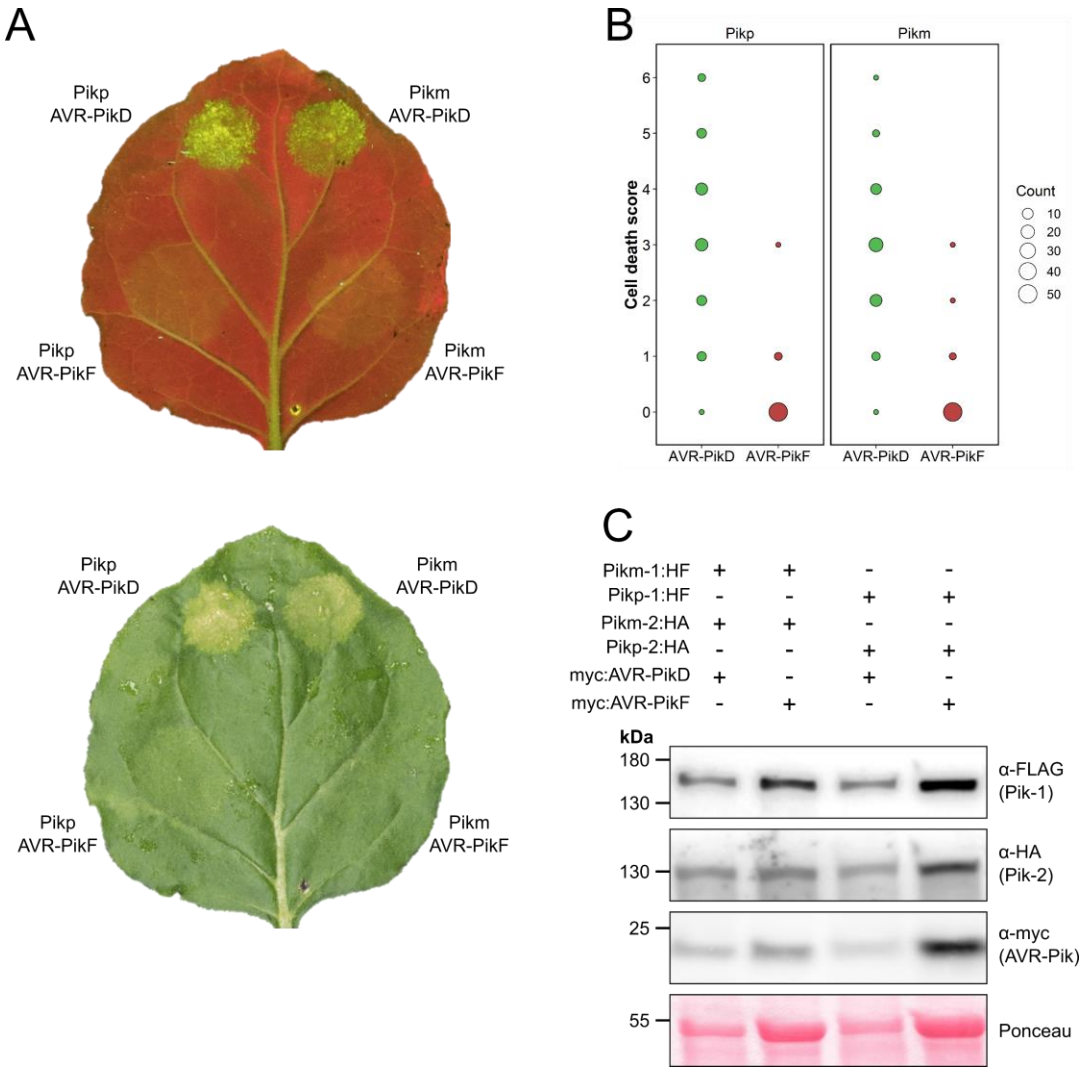


Figure 5.1 Neither Pikp nor Pikm respond to AVR-PikF in the* N. benthamiana *cell death assay.

A. *Representative leaf images showing cell death response under UV light (top) and daylight (bottom) showing the extent of cell death 5 dpi. The image under daylight has been horizontally flipped to provide the same leaf orientation for both images.* ***B.*** *Dot plot summarising the results obtained from 55 leaves in 3 independent experiments. Leaves were scored according to the scale in figure 2.2. The size of the dot at each cell death value is proportional to the number of leaves receiving that score. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).* ***C.*** *Western blots confirming production of the proteins in* N. benthamiana.

5.2.2 A Pikp-1^{HIPP19} chimera is autoactive in a Pikp-2-dependent manner

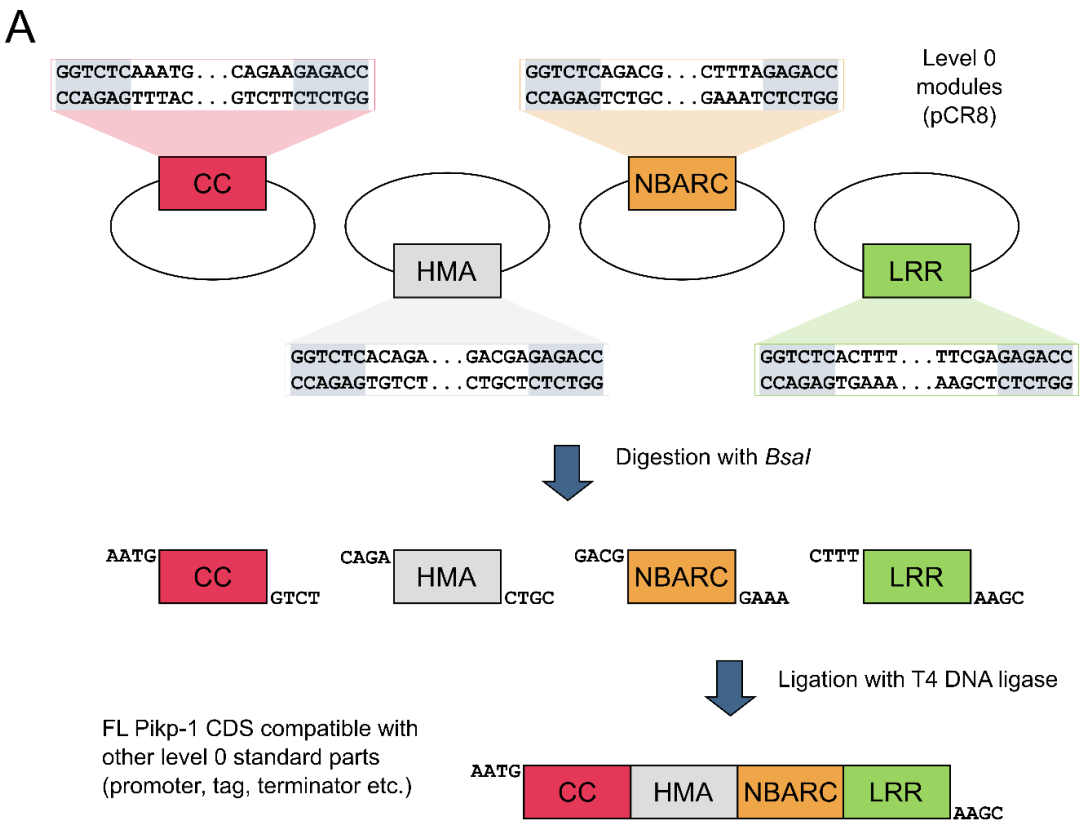
Hannah Langlands double-domesticated Pikp-1. Dr Marina Franceschetti designed and generated the level 0 modules encoding the WT Pikp-1 domains.

The results from Chapter 3 demonstrate that AVR-PikC and AVR-PikF each interact with the HMA domain of OsHIPP19 with nanomolar affinity. Previous work from the laboratory has demonstrated that a direct interaction between the effector and the HMA domain of Pik-1 is necessary for a cell death response in *N. benthamiana*, and that the strength of binding *in vitro* correlates with the extent of the cell death response. I therefore hypothesised that replacing the HMA domain of Pikp-1 with that of OsHIPP19 would result in an NLR capable of binding to AVR-PikC and AVR-PikF and delivering a cell death response *in planta*.

The Golden Gate modular cloning system was used to exchange amino acids 188-263 (inclusive) of Pikp-1 with amino acids 2-77 (inclusive) of OsHIPP19. Pikp-1 had previously been double-domesticated and the CDS divided into four level 0 modules, each encoding one of the NLR protein domains. Digestion of the level 0 constructs with the type IIS restriction enzyme *BsaI* exposes custom overhangs which promote assembly of the full-length Pikp-1 CDS in a level 1 acceptor (figure 5.2a). This system facilitates the construction of chimeric Pik-1 proteins and allows mutagenesis of individual domains prior to incorporation into the full length NLR protein. A compatible level 0 module containing the CDS for the HMA domain of OsHIPP19 (amino acids 2-77) was generated (figure 5.2b) and used in place of Pikp-HMA in the full-length assembly. The resulting chimera is described as Pikp-1^{HIPP19} from here on.

To investigate the activity of Pikp-1^{HIPP19}, the chimera was tested in the *N. benthamiana* cell death assay. Pikp-1^{HIPP19}:HF was transiently co-expressed with myc:AVR-PikD or empty vector, with and without Pikp-2:HA. Pikp-1:HF/Pikp-2:HA/myc:AVR-PikD were co-expressed as a positive control, and Pikp-1:HF/Pikp-2:HA/empty vector as a negative control.

In the presence of Pikp-2:HA, Pikp-1^{HIPP19}:HF elicits strong, effector-independent HR-like cell death (figure 5.3). This autoactive phenotype is dependent on the presence of



B

```
GGTCTCACAGAAAAACAGGGCTAAAACAAAAAATCGTCATCAAAGTCTCAATGCCGTGCGAAAAAA
   R  K  T  G  L  K  Q  K  I  V  I  K  V  S  M  P  C  E  K  S

GTCGTTCCAAAGCAATGAAACTGGTGGTTATGGCTAGCGGCGTGAGCTCTGTTGAAGTCACCGGCG
  R  S  K  A  M  K  L  V  V  M  A  S  G  V  S  S  V  E  V  T  G  D

ATGGTAAAGACCGTCTGCAGGTCGTGGGTGATGGTGTCGACGCAGCATGCCTGGTTACGTGTCTGC
  G  K  D  R  L  Q  V  V  G  D  G  V  D  A  A  C  L  V  T  C  L  R

GTAAGAAAATTGGTCACGCCGAACTGGTGCAAGTTGAAGAAGTTAAAGAGAAAGACGTGAGACC
  K  K  I  G  H  A  E  L  V  Q  V  E  E  V  K  E  K  D
```

Figure 5.2 Golden Gate assembly of a Pikp-1^{HIPP19} chimera

A. *Schematic representation of the assembly of full-length Pikp-1 coding sequence (CDS) from level 0 modules each encoding the CDS for a Pikp-1 domain, flanked by* BsaI *restriction sites. Digestion with* BsaI *reveals complementary overhangs, facilitating assembly of the full length Pikp-1 CDS. This can be combined with level 0 modules encoding promoter, terminator and tag sequences for assembly of a level 1 construct.* ***B.*** *Nucleotide sequence (top) encoding the HMA domain of OsHIPP19, flanked by the required overhangs and* BsaI *restriction sites. This sequence was cloned into pCR8 to generate a level 0 construct. The amino acid sequence is shown beneath the nucleotide sequence.*

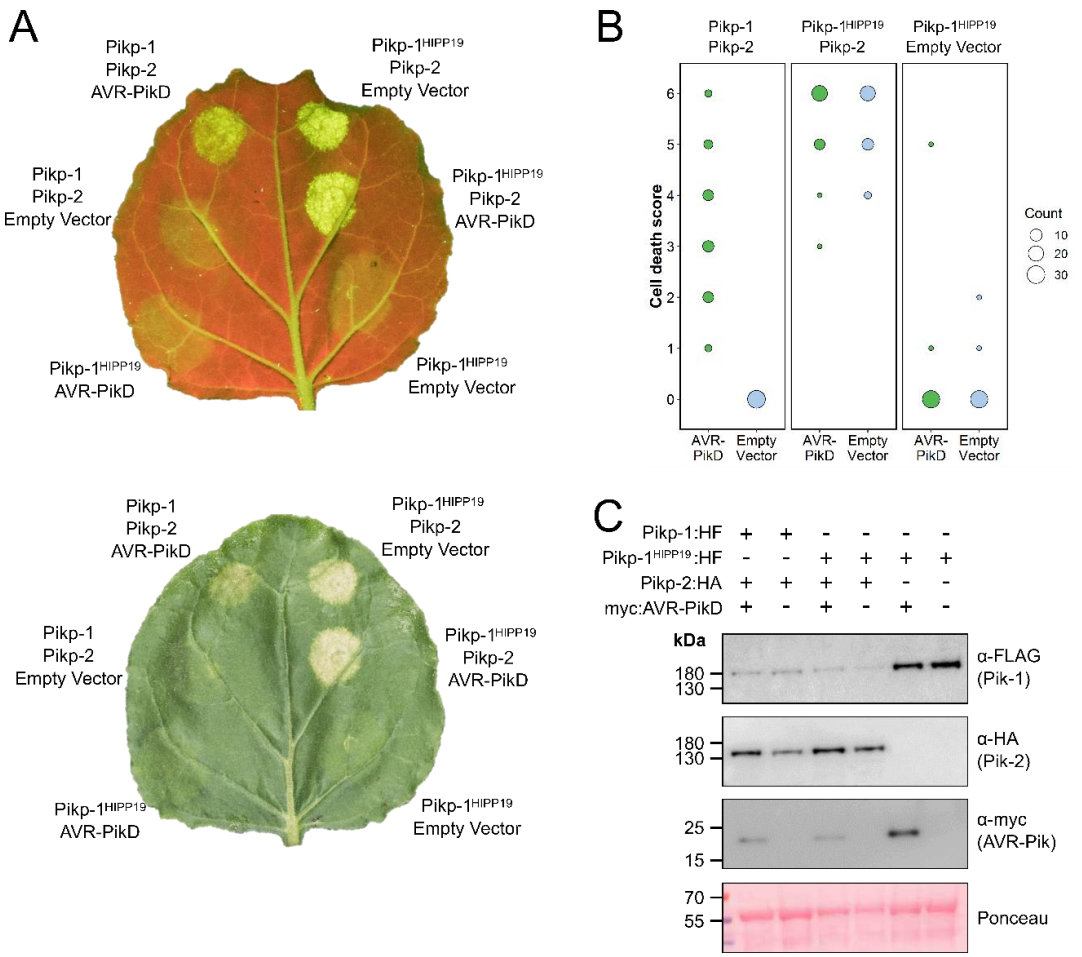


Figure 5.3 Pikp-1^{HIPP19} is autoactive in a Pikp-2-dependent manner.

A. *Representative leaf images showing cell death response under UV light (top) and daylight (bottom) showing the extent of cell death 5 dpi. The image under daylight has been horizontally flipped to provide the same leaf orientation for both images.* ***B.*** *Dot plot summarising the results obtained from 30 leaves in 3 independent experiments. Leaves were scored according to the scale in figure 2.2. The size of the dot at each cell death value is proportional to the number of leaves receiving that score. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).* ***C.*** *Western blots confirming production of the proteins in* N. benthamiana.

Pikp-2; in its absence, no HR-like cell death was observed. The autoactivity of Pikp-1^{HIPP19} prevents its use as a functional NLR protein for recognition of AVR-PikC or AVR-PikF.

Given that Pikp-1^{HIPP19} is autoactive, two approaches could be taken to engineer a functional NLR capable of responding to AVR-PikC and AVR-PikF.

1. Generate a limited number of structure-guided mutations at the effector-binding interface of the HMA domain of Pikp-1, to extend binding to AVR-PikC and AVR-PikF without making substantial changes which could lead to autoactivation.
2. Identify the amino acids in the HMA domain of OsHIPP19 that are responsible for the autoactivity of Pikp-1^{HIPP19} and replace these with the corresponding amino acids in Pikp-1 to alleviate autoactivity.

Work relating to the second strategy is described and discussed in chapter 6. This chapter now focuses on the first strategy; structure-guided mutagenesis of Pikp-1 to extend its response to AVR-PikC and AVR-PikF.

5.2.3 Pikp-1^{N261K} (Pikh-1) does not trigger cell death in response to AVR-PikC in *N. benthamiana*

Constructs for in planta *expression of myc:AVR-PikD, myc:AVR-PikE, myc:AVR-PikA and myc:AVR-PikC were provided by Dr Marina Franceschetti and Juan Carlos De la Concepcion.*

In the crystal structures of OsHIPP19-HMA and Pikm-HMA in complex with AVR-Pik, the side chain of OsHIPP19^{Lys75}/ Pikm-1^{Lys262} is positioned in a pocket of negative charge on the surface of the effector. In Pikp-HMA, the side chain of Pikp-1^{Lys262} is present in this binding pocket, however this lysine is shifted one position towards the C-terminus relative to OsHIPP19^{Lys75} and Pikm-1^{Lys262}. Consequently, the backbone of Pikp-HMA (residues Ser258-Asn261) loops out away from the effector, and is hypothesised to reduce the strength of binding at that interface (De la Concepcion et al., 2018).

Pikh-1 differs from Pikp-1 by a single asparagine-to-lysine polymorphism at position 261. Therefore, Pikh-1 has a lysine in the same amino acid position as Pikm-1. If Pikh-

1^{Lys261}, rather than Pikh-1^{Lys262}, occupies the binding pocket on the surface of the effector, Pikh-1$^{Ser258-Ala260}$ may be brought into closer proximity to the effector, stabilising the interaction.

There is conflicting data over the recognition profile of Pikh-1. Based on results from *Magnaporthe oryzae* spot inoculation experiments, Kanzaki et al. reported that the monogenic K3 rice line containing Pikh is resistant to *M. oryzae* strains carrying AVR-PikD, AVR-PikE and AVR-PikA, but not AVR-PikC (Kanzaki et al., 2012). By contrast, it has been recently reported that rice cultivars containing Pikh are resistant to *M. oryzae* strains carrying AVR-PikC in spray inoculation assays.

The *N. benthamiana* cell death assay was used to investigate whether transient co-expression of Pikh-1, Pikh-2 and AVR-PikC triggers HR-like cell death. Pikh-2 is identical to Pikp-2; for clarity, the protein will be referred to as Pikp-2 in this work. A level 1 construct expressing Pikh-1:HF was assembled from level 0 Pikp-1 domain modules and a level 0 Pikp-HMA domain module carrying the Asn261Lys mutation that distinguishes Pikh-1 from Pikp-1. Pikp-1:HF, Pikp-2:HA and myc:AVR-PikD were co-infiltrated as a positive control, and Pikh-1:HF, Pikp-2:HA and empty vector as a negative control.

Transient co-expression of Pikh-1:HF, Pikp-2:HA and myc:AVR-PikD yielded a strong cell death response (stronger than that observed following co-expression of Pikp-1:HF, Pikp-2:HA and myc:AVR-PikD), however only a weak response was observed following co-expression of Pikh-1:HF, Pikp-2:HA and myc:AVR-PikE, and co-expression with either myc:AVR-PikA or myc:AVR-PikC did not result in significant cell death (figure 5.4). The presence of all proteins was confirmed by Western blot.

5.2.4 Pikp-1^{NK-KE} triggers a stronger cell death response than Pikh-1 when co-expressed with AVR-Pik effectors

The construct for in planta *expression of Pikp-1^{NK-KE}:HF was provided by Marina Franceschetti and Juan Carlos de la Concepcion.*

Independent work by Dr Marina Franceschetti and Juan Carlos de la Concepcion identified that two amino acid changes in Pikp-1 to the corresponding amino acids in Pikm-1 could extend its response to AVR-PikE and AVR-PikA (though not to AVR-

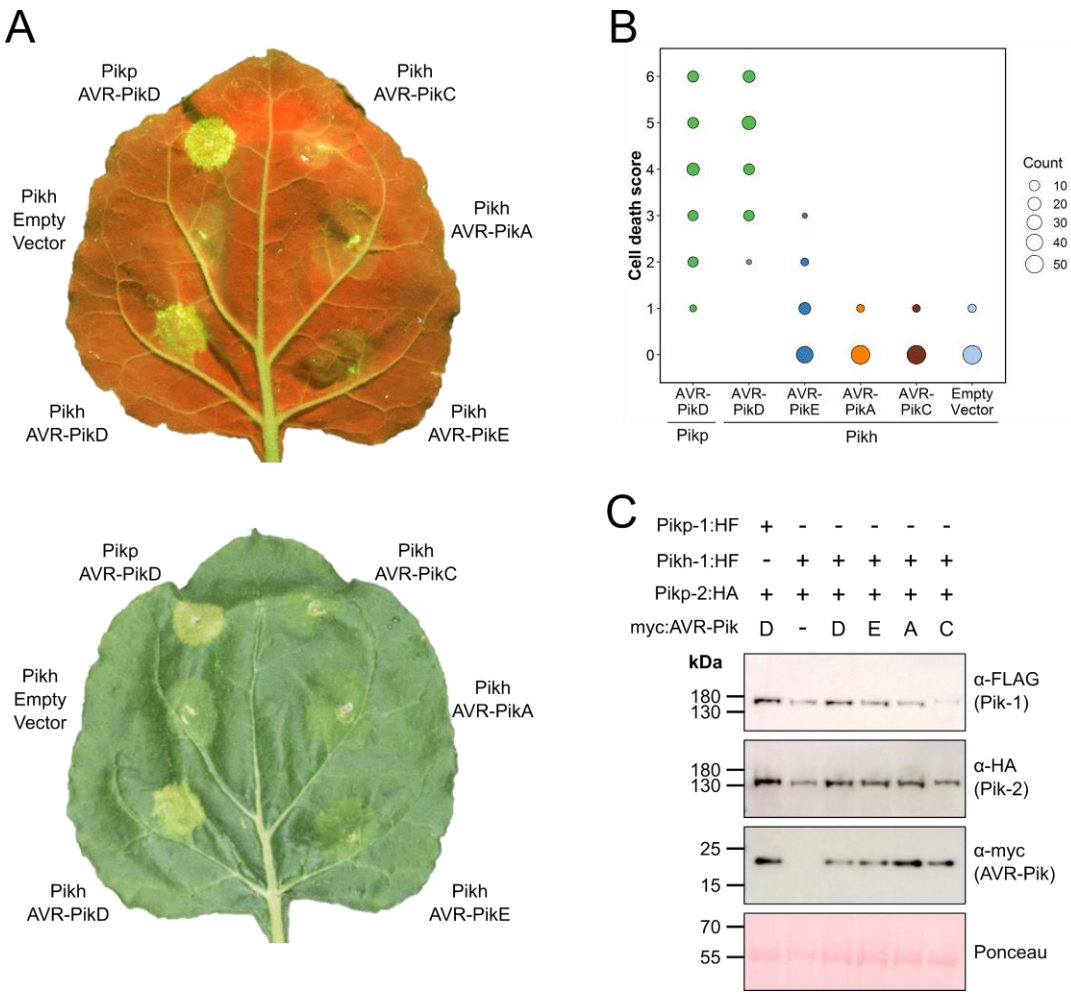


Figure 5.4 Pikh does not trigger cell death in response to AVR-PikC.

A. *Representative leaf images showing cell death response under UV light (top) and daylight (bottom) showing the extent of cell death 5 dpi. The image under daylight has been horizontally flipped to provide the same leaf orientation for both images.* ***B.*** *Dot plot summarising the results obtained from 60 leaves in 3 independent experiments. Leaves were scored according to the scale in figure 2.2. The size of the dot at each cell death value is proportional to the number of leaves receiving that score. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).* ***C.*** *Western blots confirming production of the proteins in* N. benthamiana.

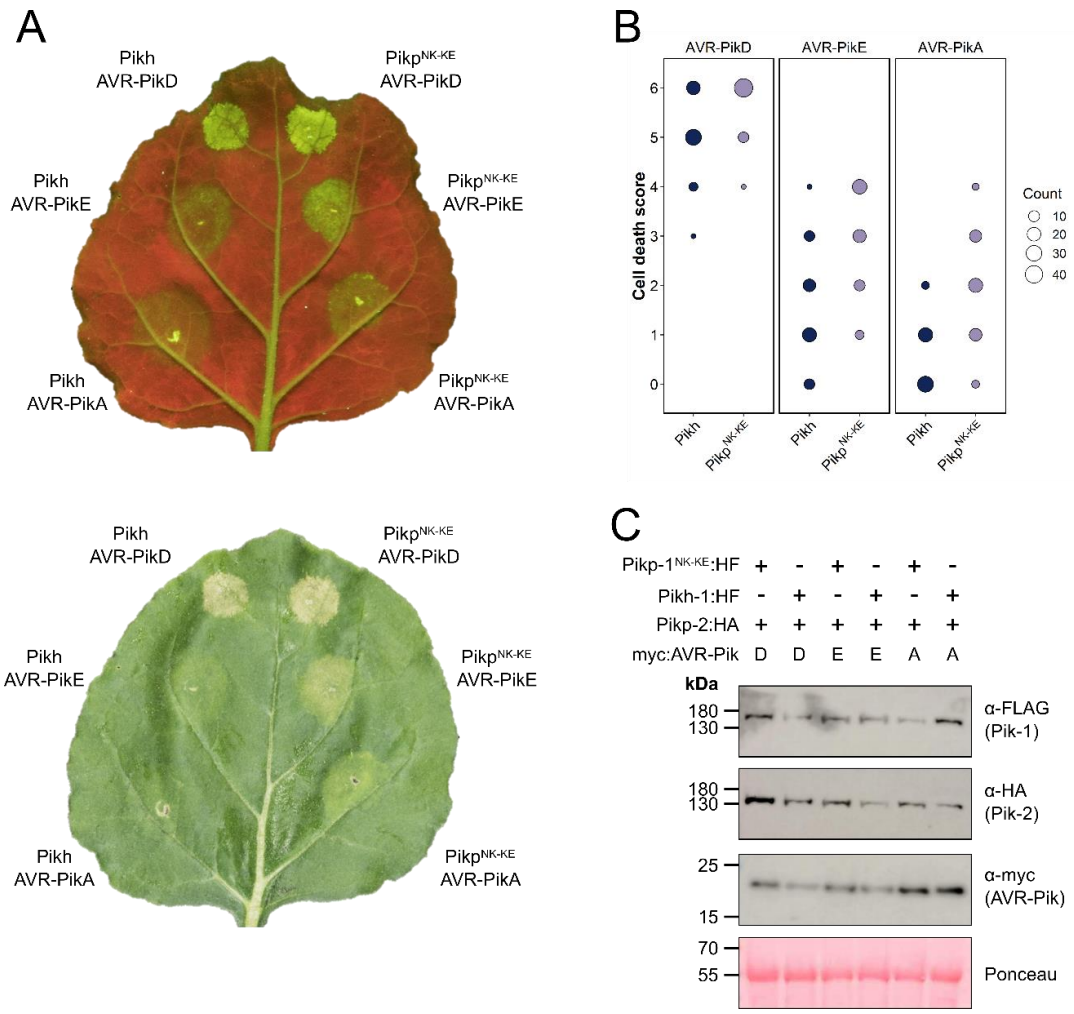


Figure 5.5 Pikp^{NK-KE} triggers a stronger cell death response than Pikh to AVR-Pik effectors.

A. *Representative leaf images showing cell death response under UV light (top) and daylight (bottom) showing the extent of cell death 5 dpi. The image under daylight has been horizontally flipped to provide the same leaf orientation for both images.* ***B.*** *Dot plot summarising the results obtained from 57 leaves in 3 independent experiments. Leaves were scored according to the scale in figure 2.2. The size of the dot at each cell death value is proportional to the number of leaves receiving that score. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).* ***C.*** *Western blots confirming production of the proteins in* N. benthamiana*.*

PikC), effectively recapitulating the phenotype of Pikm-1 (De la Concepcion et al., 2019). One of these changes was the same asparagine-261 to lysine substitution present in Pikh-1; the second was the mutation of the adjacent lysine-262 to glutamate. The resulting Pikp-1 mutant with these two amino acid substitutions is described as Pikp-1$^{NK\text{-}KE}$. A stronger cell death response was observed when Pikp-1$^{NK\text{-}KE}$:HF was transiently co-expressed with Pikp-2:HA and myc:AVR-PikD than that observed when Pikp-1:HF, Pikp-2:HA and myc:AVR-PikD were co-expressed.

In the crystal structures, Pikm-HMA^{Glu262} does not appear to play a role in effector binding. To investigate the effect, if any, of the additional lysine-262 to glutamate substitution in the extent of the cell death response, Pikh-1:HF and Pikp-1$^{NK\text{-}KE}$:HF were each transiently co-expressed with Pikp-2:HA and either myc:AVR-PikD, myc:AVR-PikE or myc:AVR-PikA in *N. benthamiana*. Pikp-1$^{NK\text{-}KE}$:HF/Pikp-2:HA consistently triggered a stronger cell death response than Pikh-1:HF/Pikp-2:HA for each of the three effector alleles (figure 5.5).

5.2.5 The crystal structure of the complex between Pikp-HMA$^{NK\text{-}KE}$ and AVR-PikC

Juan Carlos De la Concepcion carried out co-immunoprecipitation experiments with Pikp-1$^{NK\text{-}KE}$ and AVR-PikC (De la Concepcion et al., 2019), purified the complex from E. coli *and conducted crystallisation trials. Crystals were harvested by Dr Clare Stevenson. X-ray diffraction data collection was carried out with Juan Carlos De la Concepcion, Dr Clare Stevenson and Professor David Lawson.*

Juan Carlos De la Concepcion observed that AVR-PikC weakly associated with Pikp-1$^{NK\text{-}KE}$ in co-immunoprecipitation experiments (De la Concepcion et al., 2019). Subsequently, he purified the complex of Pikp-HMA$^{NK\text{-}KE}$/AVR-PikC from *E. coli* and obtained protein crystals in multiple conditions of the commercial Morpheus™ crystallisation screen with the protein complex at 13 mg/ml. A subset of these were each mounted in a loop, flash-frozen in liquid nitrogen and transferred to the Diamond Light Source synchrotron facility.

5.2.5.1 Data collection and processing

X-ray diffraction data was collected at beamline I04 of the Diamond Light Source synchrotron facility. A dataset was collected from a crystal harvested from condition F8 (buffer system 2 pH 7.5, precipitant mix 4, monosaccharides additive mix) of the commercial Morpheus™ screen. Figure 5.6 shows the crystal in the sitting drop (panel A) and mounted in the loop prior to data collection (panel B). The dataset consisted of 3600 images with a between-image oscillation angle of 0.1 °, and was collected at a wavelength of 0.98 Å. Data reduction was carried out using the AutoProc pipeline (Vonrhein et al., 2011), with the scaled but unmerged data file passed to AIMLESS (implemented in CCP4i2) (Evans and Murshudov, 2013, Winn et al., 2011). Unit cell parameters were determined to be a = 66.78, b = 80.21, c = 105.68, α = β = γ = 90. The data was processed satisfactorily in the space group $P\,2_1\,2_1\,2_1$. Guided by merging statistics, the scaled, unmerged dataset was reprocessed by AIMLESS to a maximum resolution of 2.15 Å, resulting in an overall R_{merge} of 5.3%. Data collection and processing statistics are shown in table 5.1.

The phases were solved by molecular replacement with PHASER (McCoy et al., 2007), as implemented in CCP4i2, using the structure of a dimer of Pikp-HMA in complex with AVR-PikD (PDB accession code 5A6W) as a model. A unique solution was found with a TF Z-score of 13.6, indicating successful structure solution. The asymmetric unit contains 2 copies of a dimer of Pikp-HMA$^{NK\text{-}KE}$ and a monomer of AVR-PikC.

5.2.5.2 Model refinement and validation

With only a few amino acid differences between the structures of Pikp-HMA/AVR-PikD and Pikp-1$^{NK\text{-}KE}$/AVR-PikC, these substitutions were made in COOT (Emsley et al., 2010) prior to initial refinement with REFMAC (Murshudov et al., 2011). Following this first refinement, R was 25.9 % and R_{free} was 29.8 %. Iterative rounds of manual rebuilding, refinement and validation were carried out using COOT and REFMAC.

The refined model had an R-factor of 22.2 %, with an R_{free} of 27.1 %. Table 5.2 summarises the N-terminal and C-terminal amino acid in each of the six protein chains, as well as amino acids which were excluded from the final model as the electron density

Table 5.1 Data collection and refinement statistics for the Pikp-HMA^{NK-KE}/AVR-PikC complex.

Data collection statistics	
Wavelength (Å)	
Space group	$P2_12_12_1$
Cell dimensions *a*, *b*, *c* (Å)	66.78, 80.21, 105.68
Resolution (Å)*	46.17-2.15 (2.22-2.15)
R_{merge} (%)	5.3 (99.6)
$I/\sigma I$	23.9 (2.3)
Completeness (%)	99.9 (99.8)
Unique reflections	31604 (2694)
Redundancy	13.2 (13.7)
CC(1/2) (%)	100.0 (92.5)
Refinement and model statistics	
Resolution (Å)	44.16-2.15 (2.21-2.15)
R_{work}/R_{free} (%)	22.2/27.1 (36.1/34.1)
No. atoms	
Protein	6959
Water	78
B-factors	
Protein	64.0
Water	57.4
R.m.s deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.479
Ramachandran plot (%)**	
Favoured	97.21
Allowed	2.79
Outliers	0
MolProbity Score	1.87 (89th percentile)

* The highest resolution shell is shown in parenthesis.
** As calculated by MolProbity

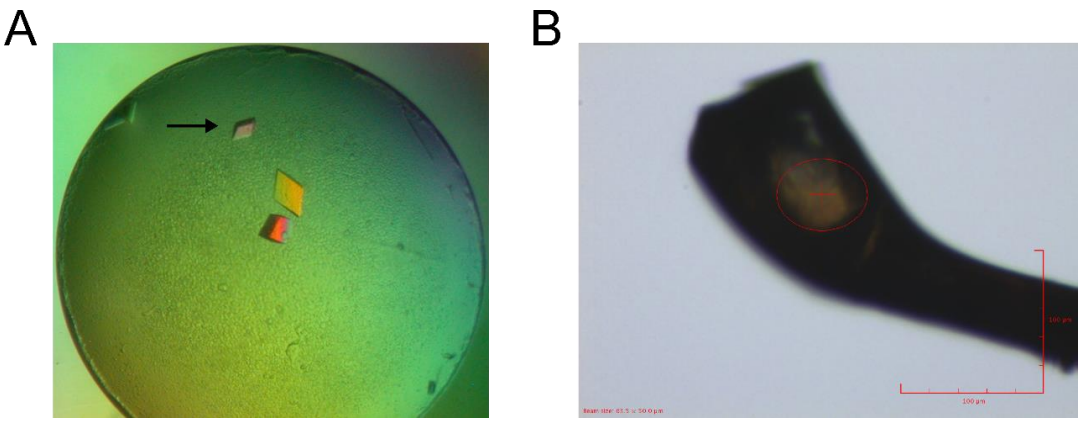


Figure 5.6 Protein crystal used for X-ray diffraction data collection.

A. *The protein crystal used for data collection (indicated by a black arrow) from condition F8 of the commercial Morpheus™ screen prior to harvesting from the sitting drop.* ***B.*** *The same crystal mounted in a loop prior to data collection.*

Table 5.2 Summary of the amino acids included in the final model for each protein chain.

Chain	Protein	N-terminal residue	C-terminal residue	Excluded residues
A	Pikp-HMA^{NK-KE}	Gly186	Gln259	Gly199 – Asn200
B	*Pikp-HMA^{NK-KE}	Gly186	Glu262	Glu198 – Asn200
C	AVR-PikC	Ile33	Pro111	-
D	Pikp-HMA^{NK-KE}	Leu187	Gln259	Gly199 – Asn201
E**	*Pikp-HMA^{NK-KE}	Gly186	Glu262	Glu198 – Asn201
F**	AVR-PikC	Ala32	Phe113	-

* indicates that this copy of Pikp-HMA^{NK-KE} interfaces with AVR-PikC.
** indicates that these chains were used for interface analysis.

was not sufficiently well-defined to position them. 78 water molecules were included in the refined model.

Final validation of the model was carried out with MolProbity (Chen et al., 2010). The Ramachandran plots for the refined model showed that all residues are in the allowed regions, with 97.21 % in the favoured regions (figure 5.7). Rotamer analysis classified 13 rotamers as outliers, however inspection of the electron density found that more favourable rotamers were not supported in each case. The clash score is 5.82, placing it in the 97th percentile of similar structures. The overall MolProbity score of 1.87 places the structure in the 89th percentile of structures of comparable resolution. Refinement and validation statistics are shown in table 5.1. The overall structure of AVR-PikC and the interfacing HMA domain of Pikp-1^{NK-KE} is shown in figure 5.8 and at the global level resembles the structures of previously solved AVR-Pik/HMA domain complexes. AVR-PikC differs from AVR-PikE by a single amino acid polymorphism at position 67 (alanine in AVR-PikE, aspartate in AVR-PikC). Comparison of the structure of Pikp-1^{NK-KE}/AVR-PikC with the structure of Pikp-1^{NK-KE}/AVR-PikE (PDB accession code 6R8M; (De la Concepcion et al., 2019)) shows that the two structures are very similar, with an overall RMSD of 0.58 Å across 154 amino acids.

Interface analysis using QtPISA revealed that, for the structure of Pikp-HMA^{NK-KE}/AVR-PikC, the binding interface buries 21.5 % and 18.0 % of the total accessible surface area of the HMA domain (971.5 Å^2) and effector (1018.2 Å^2) respectively. This is very similar to the interface between Pikp-HMA^{NK-KE}/AVR-PikE, which buries 22.0 % and 18.0 % of the surface area of the HMA domain (964.3 Å^2) and effector (993.9 Å^2) respectively. Furthermore, the number of hydrogen bonds and salt bridges at the interface are similar; 12 hydrogen bonds and 8 salt bridges form between Pikp-HMA^{NK-KE} and AVR-PikC, while 11 hydrogen bonds and 10 salt bridges form between Pikp-HMA^{NK-KE} and AVR-PikE. Key interface parameters for both complexes are shown in table 5.3 and visualised as radar plots in figure 5.9.

The crystal structure of Pikp-HMA^{NK-KE}/AVR-PikC shows that AVR-PikC^{Asp67} is located at the interface with Pikp-HMA^{NK-KE}. In the structure of Pikp-HMA^{NK-KE} /AVR-PikE, the side chain of Pikp-HMA$^{NK-KE_Asp224}$ forms two hydrogen bonds with the side chain of AVR-PikE^{Arg64} (figure 5.10). However, in the structure of Pikp-HMA^{NK-KE}/AVR-PikC, the

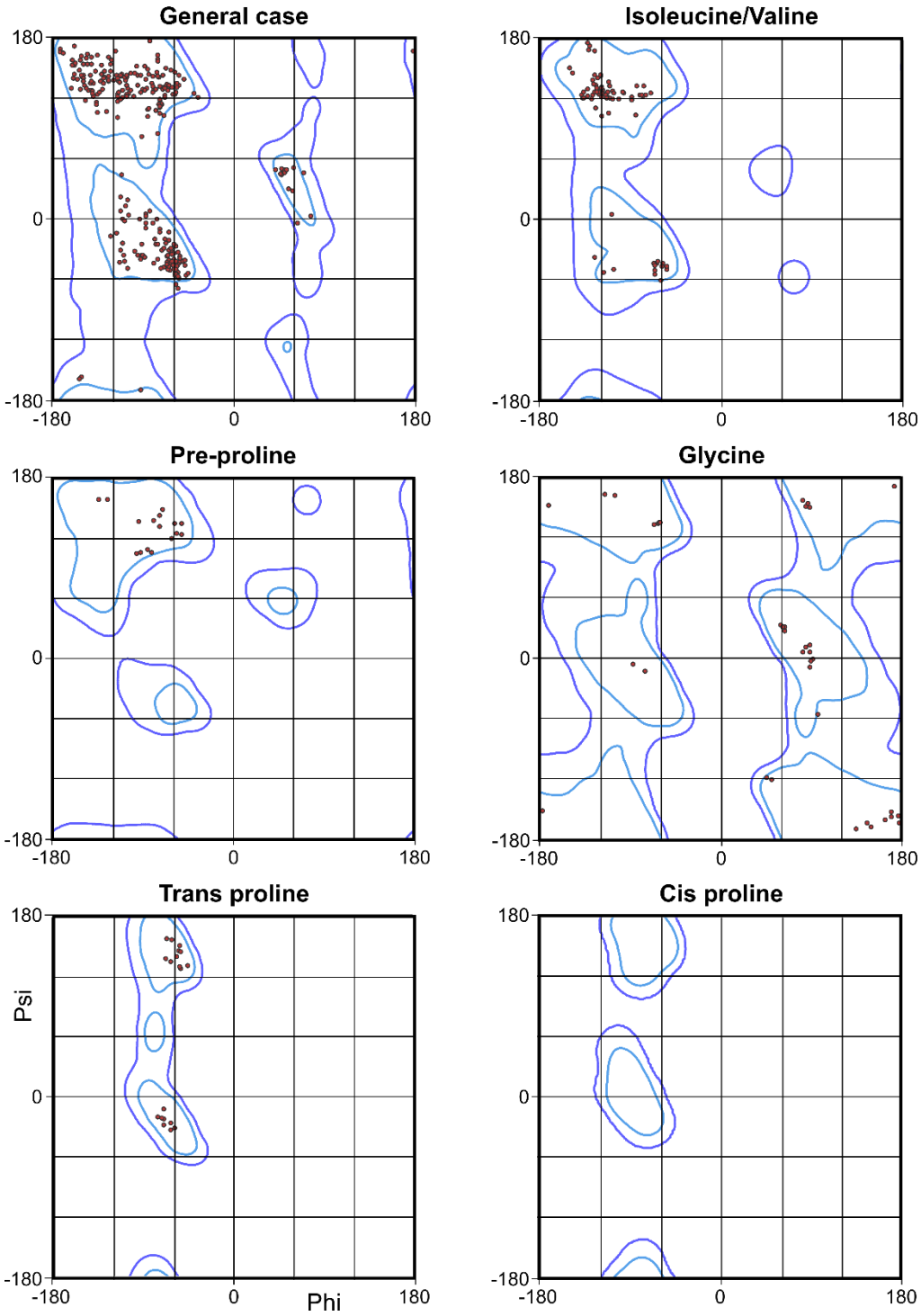


Figure 5.7 Ramachandran plots for the structure of AVR-PikC in complex with Pikp-HMA^{NK-KE}.

100 % of residues were in allowed regions of the plots (within the purple lines), with 97.21 % of residues lying within favoured regions (the blue lines). Ramachandran plots were produced with MolProbity.

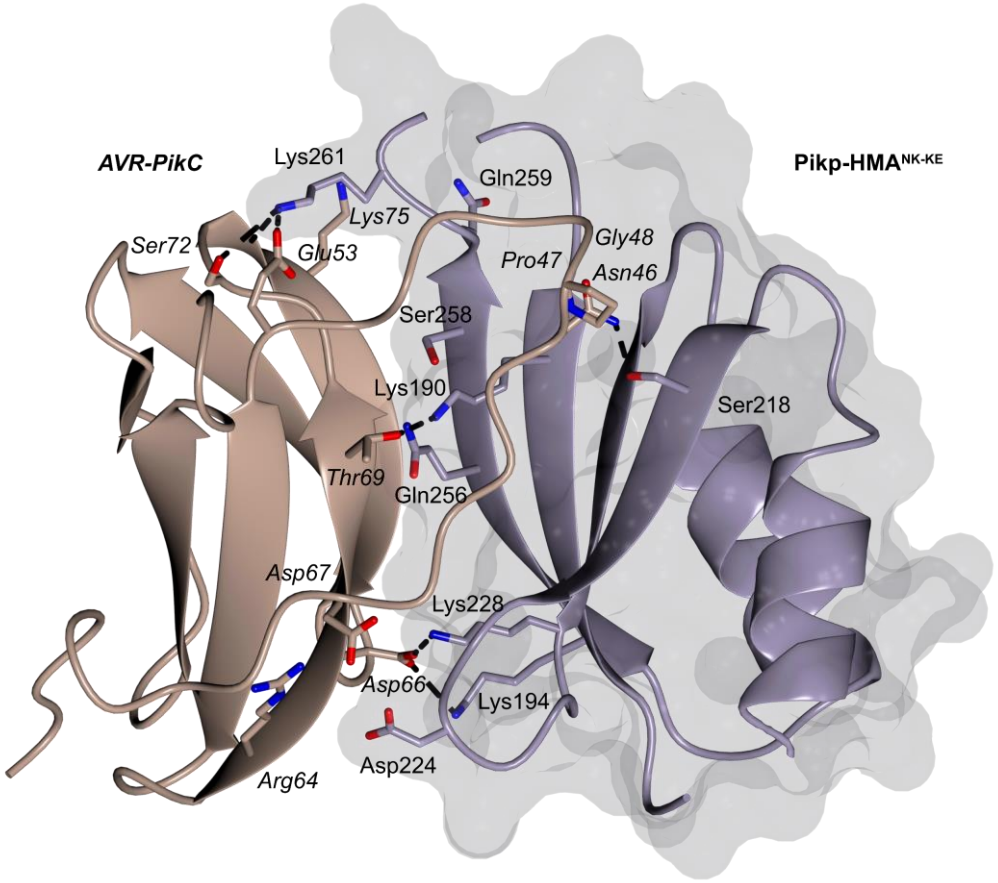


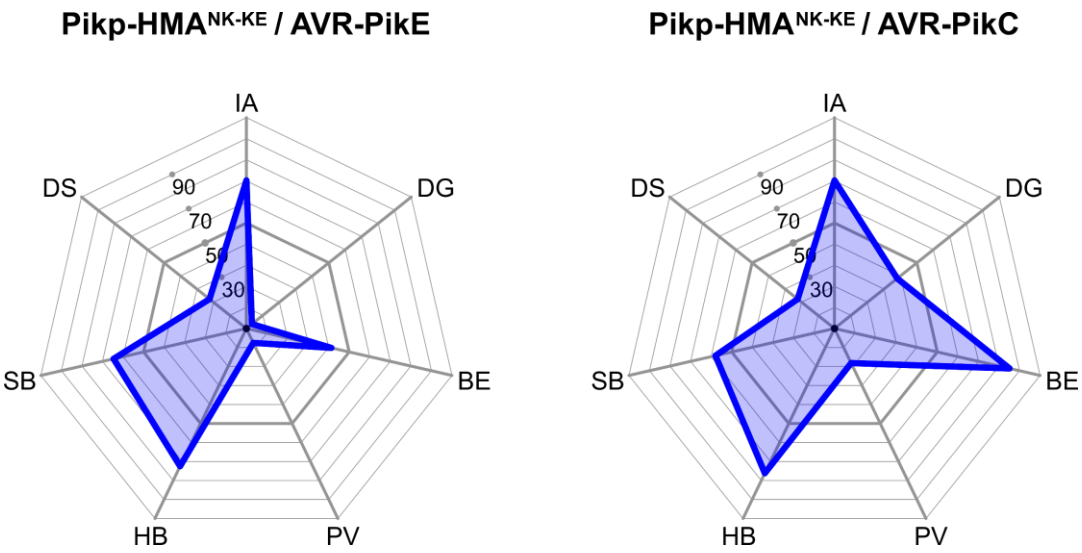
Figure 5.8 The crystal structure of AVR-PikC in complex with Pikp-HMA^{NK-KE}.

The structures of AVR-PikC and Pikp-HMA^{NK-KE} are represented as brown and lilac ribbons respectively, with the molecular surface of Pikp-HMA^{NK-KE} also displayed. Side chains of amino acids of interest are displayed as cylinders. Hydrogen bonds are represented by dashed lines.

Table 5.3 Summary of interface analysis by QtPISA for Pikp-HMA^{NK-KE}/AVR-PikC and the published structure of Pikp-HMA^{NK-KE}/AVR-PikE (PDB accession code 6R8M,(De la Concepcion et al., 2019)).

		Pikp^{NK-KE} / AVR-PikE	Pikp^{NK-KE} / AVR-PikC
AVR-Pik	B.S.A. (Å)	964.3	971.5
	% B.S.A. of total	18.0	18.0
HMA	B.S.A. (Å)	993.9	1018.2
	% B.S.A. of total	22.0	21.5
Total interface area* (Å)		979.1	994.9
Number of hydrogen bonds		11	12
Number of salt bridges		10	8

* Total interface area is the total B.S.A. (Buried Surface Area) of each component divided by two.



	Pikp^{NK-KE} / AVR-PikE	Pikp^{NK-KE} / AVR-PikC
IA: Interface area	979.1 Å	994.8 Å
DG: Solvation energy	0.1 kcal/mol	-5.1 kcal/mol
BE: Binding energy	-8.5 kcal/mol	-13.4 kcal/mol
PV: Hydrophobic P-value	0.7780	0.4997
HB: Hydrogen bonds	11	12
SB: Salt bridges	10	8
DB: Disulphide bonds	0	0

Figure 5.9 Comparison of the binding interfaces of Pikp-HMA^{NK-KE}/AVR-PikE and Pikp-HMA^{NK-KE}/AVR-PikC using QtPISA.

Radar plots produced by QtPISA provide a visual representation of the binding interfaces of the Pikp-HMA^{NK-KE}/AVR-PikE and Pikp-HMA^{NK-KE}/AVR-PikC complexes, based on the seven parameters outlined above. The radar beam for each parameter represents the probability that the value of that parameter indicates a biologically relevant assemblage in the PDB, with a probability of zero located at the centre of the radar.

loop containing Pikp-HMA$^{NK\text{-}KE_Asp224}$ is shifted away from the effector relative to its position in the Pikp-HMA$^{NK\text{-}KE}$ /AVR-PikE structure, likely due to a combination of steric clash and repulsion due to the matching charges of the Pikp-HMA$^{NK\text{-}KE_Asp224}$ and AVR-PikC^{Asp67} side chains. Consequently, Pikp-HMA$^{NK\text{-}KE_Asp224}$ does not form hydrogen bonds with AVR-PikC^{Arg64}; instead, the side chain of AVR-PikC^{Arg64} forms an intramolecular hydrogen bond with the side chain of AVR-PikC^{Asp67} (figure 5.10).

5.2.6 Mutating Pikp-1$^{NK\text{-}KE_Asp224}$ does not extend the cell death response of Pikp-1$^{NK\text{-}KE}$ to AVR-PikC

I hypothesised that replacing Pikp-HMA$^{NK\text{-}KE_Asp224}$ with either a lysine or an alanine could compensate for the disruptive influence of AVR-PikC^{Asp67} and facilitate recognition of AVR-PikC by Pikp-1$^{NK\text{-}KE}$. Exchanging Pikp-HMA$^{NK\text{-}KE_Asp224}$ for a lysine could support the formation of a new hydrogen bond between the oppositely charged side chains, while replacing Pikp-HMA$^{NK\text{-}KE_Asp224}$ with an alanine could limit the disruptive influence of AVR-PikC^{Asp67} and allow the Pikp-HMA$^{NK\text{-}KE}$ loop containing residue 224 to shift back towards the effector.

The Asp224Lys and Asp224Ala mutations were each made in the Pikp-1$^{NK\text{-}KE}$ background, resulting in Pikp-1$^{DNK\text{-}KKE}$ and Pikp-1$^{DNK\text{-}AKE}$, respectively. Pikp-1$^{DNK\text{-}KKE}$:HF and Pikp-1$^{DNK\text{-}AKE}$:HF were then tested in *N. benthamiana* cell death assays with either myc:AVR-PikD or myc:AVR-PikC. Pikp-1$^{NK\text{-}KE}$:HF was transiently co-expressed with Pikp-2:HA and either myc:AVR-PikD, as a positive control, or myc:AVR-PikC, as a negative control. Neither of the Asp224 mutations extended the response of the NLR to AVR-PikC, and both mutations reduced the cell death response of Pikp-1$^{NK\text{-}KE}$ to AVR-PikD (figure 5.11). The latter result is perhaps unsurprising, as these mutations would also prevent the two hydrogen bonds being formed between Pikp-1$^{NK\text{-}KE_Asp224}$ and AVR-PikD^{Arg64}.

5.2.7 Pikp-1$^{SNK\text{-}EKE}$ triggers cell death in response to AVR-PikC or AVR-PikF

As mutations to accommodate polymorphic residues were unsuccessful, I hypothesised that it may be possible to engineer a Pik-1 variant that interacts with AVR-PikC/AVR-

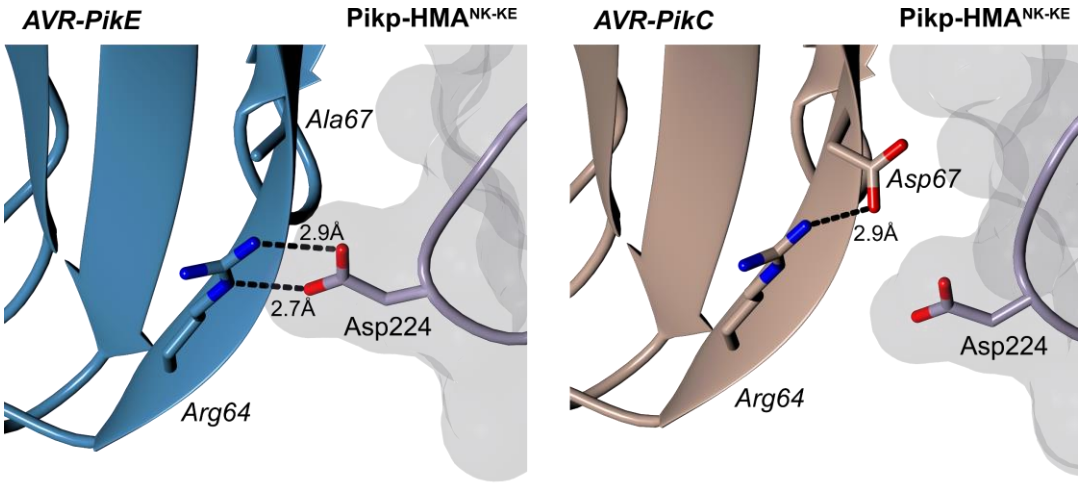


Figure 5.10 AVR-PikC^{Asp67} disrupts the hydrogen bonding between AVR-Pik^{Arg64} and Pikp-HMA$^{NK\text{-}KE_Asp224}$.

The structures are represented as lilac (Pikp-HMA$^{NK\text{-}KE}$), blue (AVR-PikE) and brown (AVR-PikC) ribbons, with key residues shown as cylinders. Hydrogen bonds are represented as dashed lines. Bond lengths are shown for relevant hydrogen bonds. The molecular surface of the HMA domains are shown in lilac. The left image shows two hydrogen bonds formed between AVR-PikE^{Arg64} and Pikp-HMA$^{NK\text{-}KE_Asp224}$ (PDB accession code 6R8M, (De la Concepcion et al., 2019)). The right image shows that the polymorphic AVR-PikC^{Asp67} prevents the formation of these intermolecular hydrogen bonds. For clarity, the N-terminal "arm" of AVR-Pik (residues Ala32-Asp45) is hidden from the foreground of both images.

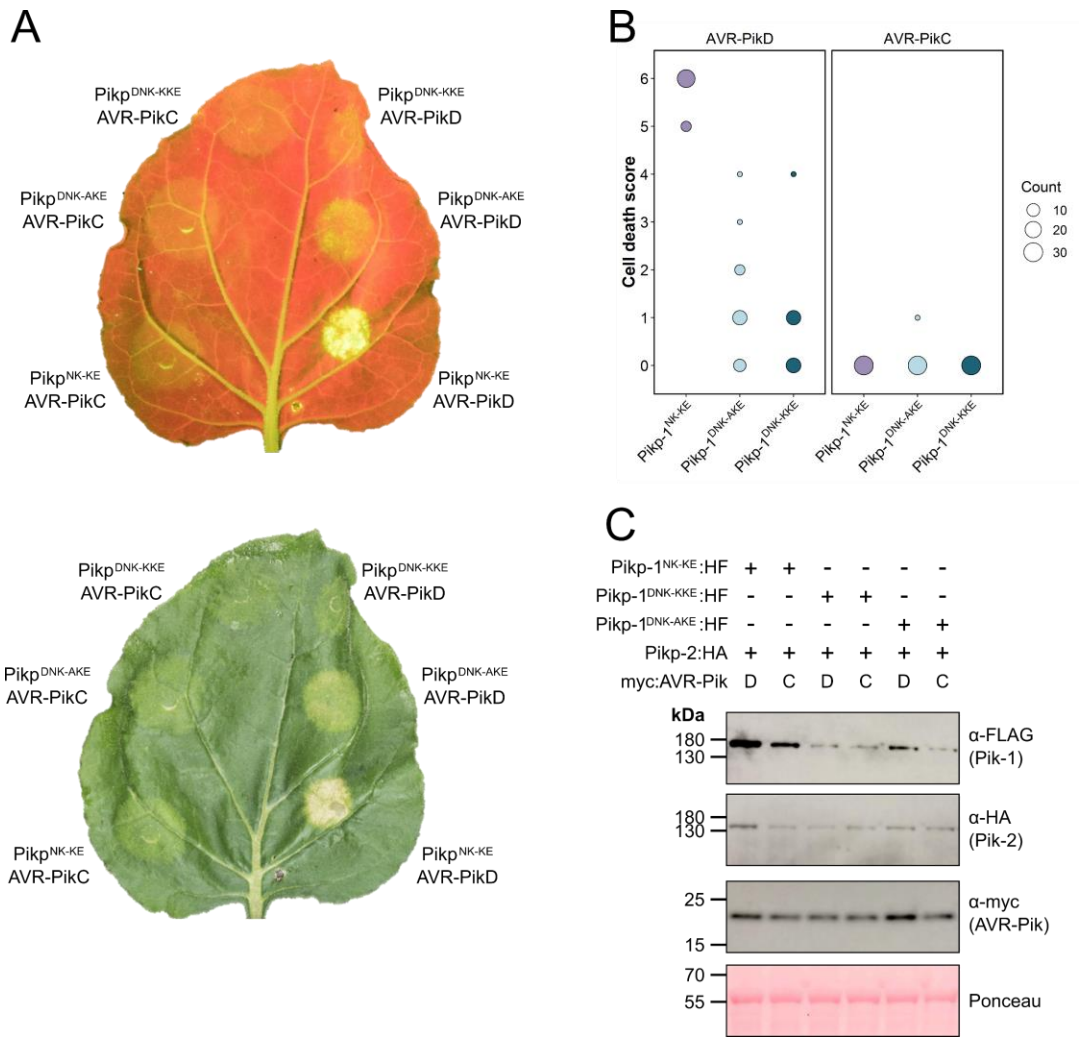


Figure 5.11 Mutating Asp224 of Pikp-1^{NK-KE} does not extend the response of the NLR to AVR-PikC.

A. *Representative leaf images showing cell death response under UV light (top) and daylight (bottom) showing the extent of cell death 5 dpi. The image under daylight has been horizontally flipped to provide the same leaf orientation for both images.* ***B.*** *Dot plot summarising the results obtained from 30 leaves in 3 independent experiments. Leaves were scored according to the scale in figure 2.2. The size of the dot at each cell death value is proportional to the number of leaves receiving that score. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).* ***C.*** *Western blots confirming production of the proteins in* N. benthamiana.

PikF by modifying other interfaces in the HMA domain, which may then compensate for disruption at the site of the polymorphic residue.

The crystal structure of OsHIPP19-HMA in complex with AVR-PikF revealed additional hydrogen bonds at interface 3 relative to the structures of integrated HMAs in complex with AVR-Pik. The side chain of OsHIPP19^{Glu72} was particularly striking; the corresponding residue in both Pikp-HMA and Pikm-HMA is serine, and while the hydroxyl group of the serine side chain only forms an intramolecular hydrogen bond within the HMA domain, the bulkier OsHIPP19^{Glu72} side chain extends into the space between the effector and HMA domain and forms a direct hydrogen bond with the effector backbone. I hypothesised that this residue may help to rigidify and stabilise the interactions at this interface to increase the binding affinity of the two proteins.

The Ser258Glu mutation was made in the Pikp-1^{NK-KE} background, and the triple mutant (Pikp-1$^{SNK-EKE}$ hereafter) was tested in *N. benthamiana* cell death assays as described previously. Strong HR-like cell death was observed when Pikp-1^{NK-KE}:HF or Pikp-1$^{SNK-EKE}$:HF was co-infiltrated with Pikp-2:HA and myc:AVR-PikD, and when myc:AVR-PikD was replaced by empty vector, no cell death was detected. Critically, infiltration of Pikp-1$^{SNK-EKE}$:HF, but not Pikp-1^{NK-KE}:HF, with Pikp-2:HA and myc:AVR-PikC or myc:AVR-PikF, resulted in cell death (figure 5.12 and figure 5.13).

To investigate whether the Ser258Glu mutation alone is sufficient to extend the cell death response to AVR-PikC, Pikp-1^{S258E}:HF was generated and screened in the cell death assay. When Pikp-1^{S258E}:HF was co-infiltrated with Pikp-2:HA and either myc:AVR-PikC or myc:AVR-PikF, no cell death was observed (figure 5.14). This indicates that all three mutations are required for HR-like cell death in response to AVR-PikC and AVR-PikF.

5.2.8 Pikp-1$^{SNK-EKE}$ interacts with AVR-PikC and AVR-PikF *in vitro* and *in vivo*

Pikp-HMA^{NK-KE} in pOPIN-M, for production of recombinant Pikp-HMA^{NK-KE} in E. coli*, was provided by Dr Marina Franceschetti and Juan Carlos De la Concepcion.*

Previous work has correlated the extent of the Pik/AVR-Pik-dependent cell death

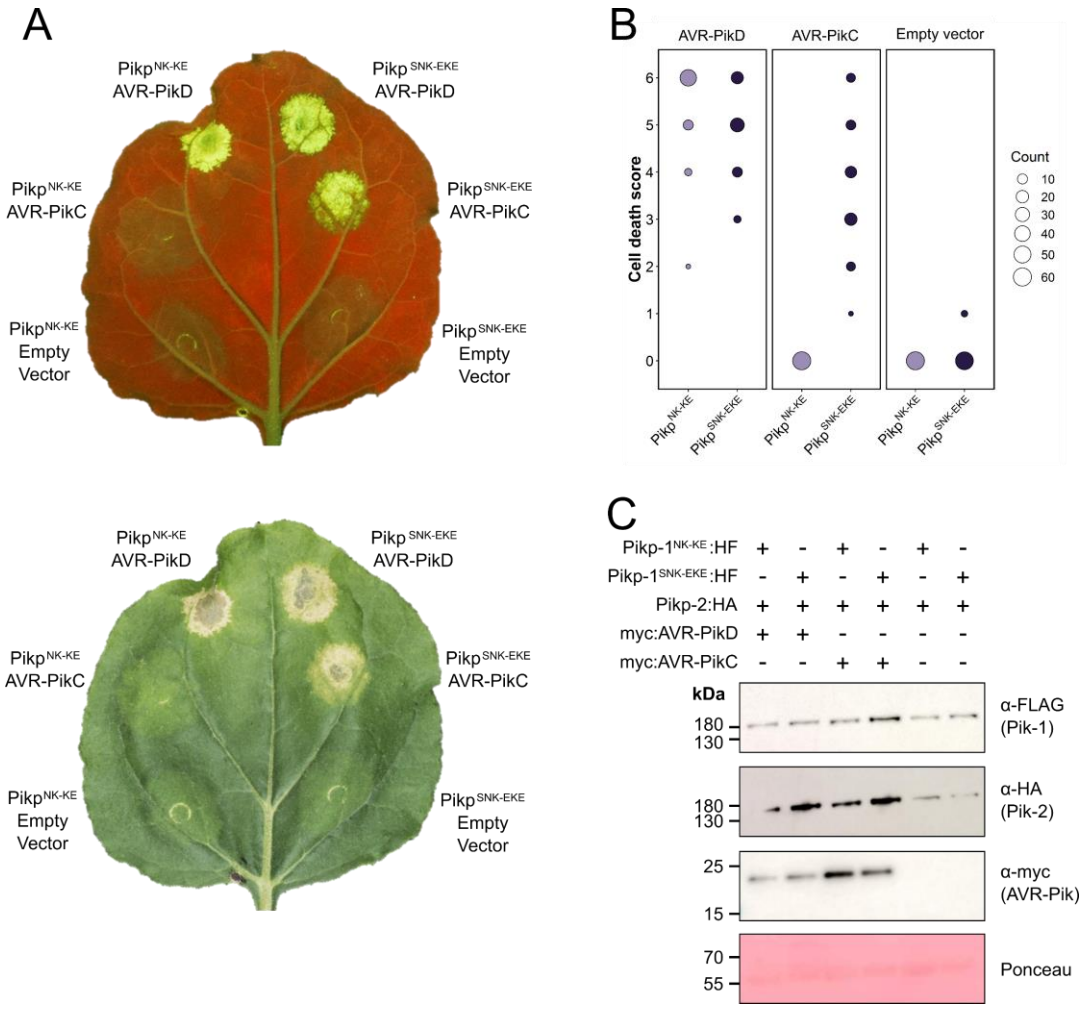


Figure 5.12 Pikp$^{SNK-EKE}$ triggers cell death in response to AVR-PikC.

A. *Representative leaf images showing cell death response under UV light (top) and daylight (bottom) showing the extent of cell death 5 dpi. The image under daylight has been horizontally flipped to provide the same leaf orientation for both images.* ***B.*** *Dot plot summarising the results obtained from 60 leaves in 3 independent experiments. Leaves were scored according to the scale in figure 2.2. The size of the dot at each cell death value is proportional to the number of leaves receiving that score. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).* ***C.*** *Western blots confirming production of the proteins in* N. benthamiana.

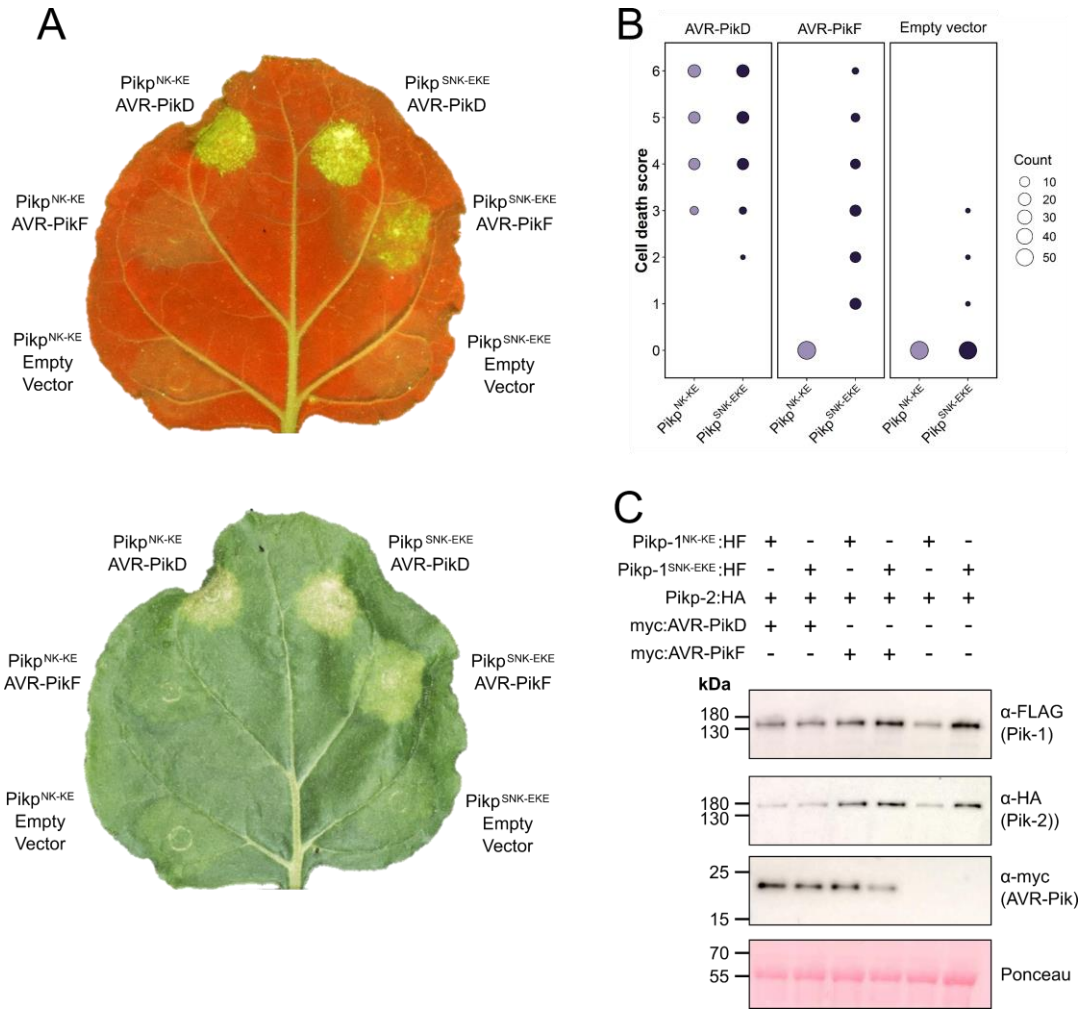


Figure 5.13 Pikp$^{SNK-EKE}$ triggers cell death in response to AVR-PikF

A. *Representative leaf images showing cell death response under UV light (top) and daylight (bottom) showing the extent of cell death 5 dpi. The image under daylight has been horizontally flipped to provide the same leaf orientation for both images.* ***B.*** *Dot plot summarising the results obtained from 57 leaves in 3 independent experiments. Leaves were scored according to the scale in figure 2.2. The size of the dot at each cell death value is proportional to the number of leaves receiving that score. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).* ***C.*** *Western blots confirming production of the proteins in* N. benthamiana.

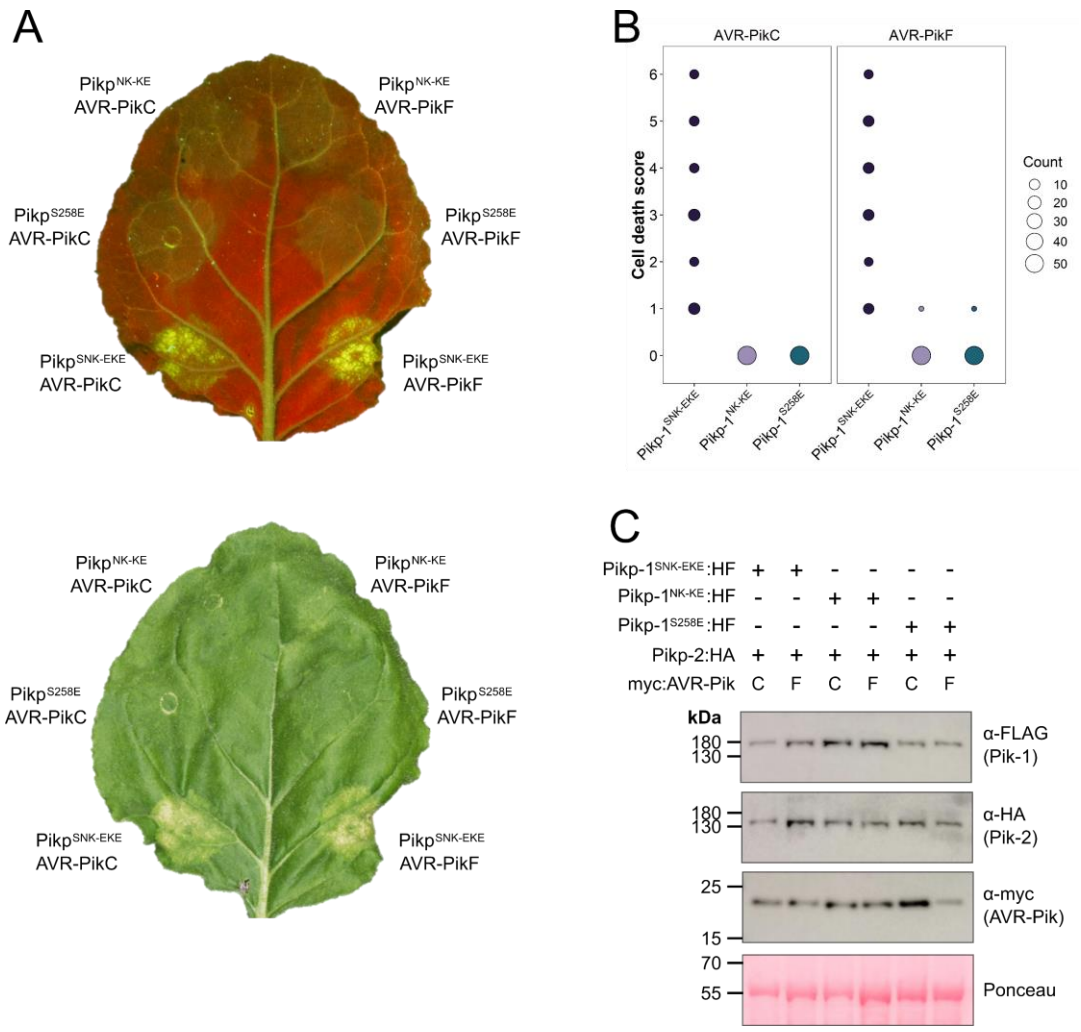


Figure 5.14 Pikp-1^{S258E} cannot trigger cell death in response to AVR-PikC or AVR-PikF.

A. *Representative leaf images showing cell death response under UV light (top) and daylight (bottom) showing the extent of cell death 5 dpi. The image under daylight has been horizontally flipped to provide the same leaf orientation for both images.* ***B.*** *Dot plot summarising the results obtained from 54 leaves in 3 independent experiments. Leaves were scored according to the scale in figure 2.2. The size of the dot at each cell death value is proportional to the number of leaves receiving that score. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).* ***C.*** *Western blots confirming production of the proteins in* N. benthamiana.

response in *N. benthamiana* with binding affinity *in vitro* and *in planta* (De la Concepcion et al., 2018, Maqbool et al., 2015). I hypothesised that the cell death response observed when Pikp-1$^{SNK-EKE}$ was transiently co-expressed with AVR-PikC or AVR-PikF can be attributed to increased binding affinity of the modified HMA domain for these effector alleles.

To investigate whether the Ser258Glu mutation increases the binding affinity of the HMA domain for AVR-PikC and AVR-PikF, surface plasmon resonance was used. Pikp$^{SNK-EKE}$-HMA was cloned into pOPIN-M, produced in *E. coli* and purified using the pipeline described previously for OsHIPP19-HMA.

AVR-PikC or AVR-PikF was immobilised on a Ni^{2+}-NTA sensor chip via a C-terminal non-cleavable 6xHis tag. Pikp-HMA (which has previously been shown not to bind AVR-PikC in SPR experiments), Pikp^{NK-KE}-HMA (for which weak binding to AVR-PikC has been observed (De la Concepcion et al., 2019)), or Pikp$^{SNK-EKE}$-HMA was flowed over the surface of the chip. 3 different concentrations of each HMA domain (4nM, 40nM and 100nM) were used. In each case, the HMA domain was presumed to be dimer, as previous work has shown that a dimer of Pikp-HMA interacts with a monomer of the effector. The binding stability (R_{obs}, measured in response units (RU)) was recorded, and expressed as a percentage of the maximum theoretical response (%R_{max}) as described in section 3.2.5.

As anticipated, Pikp-HMA did not bind to either AVR-PikC or AVR-PikF, evidenced by a %R_{max} of approximately zero at each of the 3 concentrations (figure 5.15). Weak binding was observed for Pikp^{NK-KE}-HMA to both AVR-PikC and AVR-PikF. By contrast, the Pikp$^{SNK-EKE}$-HMA bound to both effectors with higher apparent affinity (larger %R_{max}) than Pikp^{NK-KE}-HMA. This result was consistent across the 3 concentrations investigated, though binding (and %R_{max}) was low for all three HMA domains at 4nM (figure 5.15).

While working with purified proteins *in vitro* enables quantification of the binding affinity of the two proteins of interest, it is not possible to produce sufficient quantities of full-length Pik-1 in a heterologous system, therefore experiments *in vitro* use the HMA domains alone. Consequently, these experiments lack both the context of the full-length protein, and the biological context of the cellular environment.

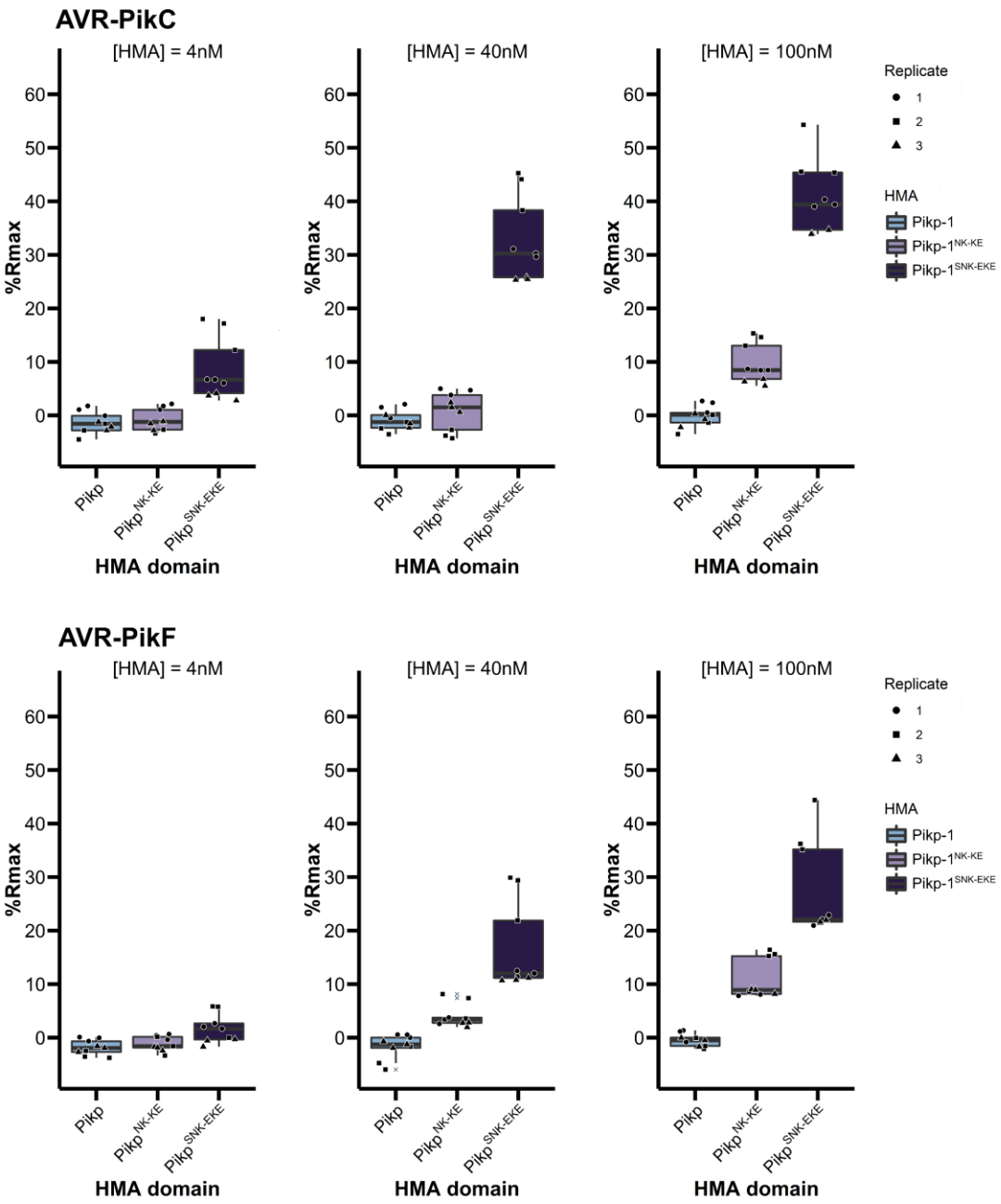


Figure 5.15 Pikp-HMA$^{SNK\text{-}EKE}$ has higher apparent affinity for AVR-PikC and AVR-PikF than Pikp-HMA$^{NK\text{-}KE}$.

%R_{max} is the percentage of the theoretical maximum response, assuming a 2:1 HMA:effector binding model. The centre line of the box represents the median and the box limits are the upper and lower quartiles. Whiskers extend to the smallest value within Q_1 - 1.5x the interquartile range (IQR) and the largest value within Q_3 + 1.5x IQR. The experiment was repeated 3 times, with each experiment containing 3 technical replicates. Individual data points are represented as black shapes. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).

Co-immunoprecipitation is a technique to test for association between two proteins of interest following their production in, and extraction from, a plant cell. In brief, both proteins of interest are transiently co-expressed in *N. benthamiana*, each with a different epitope tag. In the experiments described here, Pik-1 has a C-terminal HellFire (His:FLAG, HF) tag and AVR-Pik has a N-terminal myc tag. The plant extract is incubated with magnetic beads bound to the α-FLAG antibody, to immunoprecipitate Pik-1:HF and associated proteins. Following multiple wash steps to remove unbound protein, the immunoprecipitated proteins are eluted from the beads. The presence of the proteins of interest in the original plant extract (input) and eluate is determined by Western blotting. If the myc-tagged effector is present in the eluate, the effector associates with the FLAG-tagged NLR protein. It should be noted that co-immunoprecipitation experiments alone cannot demonstrate a direct interaction between proteins of interest, as intermediate proteins present in the plant extract may indirectly mediate association.

Pikp-1:HF, Pikp-1$^{NK\text{-}KE}$:HF and Pikp-1$^{SNK\text{-}EKE}$:HF were each transiently co-expressed with either myc:AVR-PikD, myc:AVR-PikC or myc:AVR-PikF. Pikp-2 was not included in order to prevent the onset of HR-like cell death, which reduces protein levels in the plant cell extract. Furthermore, previous work has shown that AVR-Pik can associate with Pik-1 in the absence of Pik-2 (De la Concepcion et al., 2018).

Co-expression of myc:AVR-PikD with empty vector was used as a negative control, to show that the effector does not bind non-specifically to the exposed anti-FLAG resin during the immunoprecipitation or remain in the eluate following immunoprecipitation due to, for example, insufficient washing of the resin. Pikp-1:HF was co-expressed with myc:AVR-PikD as a positive control. A band of the expected size for myc-tagged AVR-PikC and AVR-PikF was detected in the eluate following immunoprecipitation of Pikp-1$^{SNK\text{-}EKE}$:HF (figure 5.16), although the band for AVR-PikF was quite faint. No band corresponding to myc:AVR-PikC or myc:AVR-PikF was detected in the eluate following immunoprecipitation of Pikp-1:HF or Pikp-1$^{NK\text{-}KE}$:HF. This demonstrates that AVR-PikC and AVR-PikF associate with Pikp-1$^{SNK\text{-}EKE}$, but not with Pikp-1 or Pikp-1$^{NK\text{-}KE}$. This is consistent with the *in vitro* data obtained using surface plasmon resonance. Taken together, the results from these complementary approaches indicate that the Ser258Glu

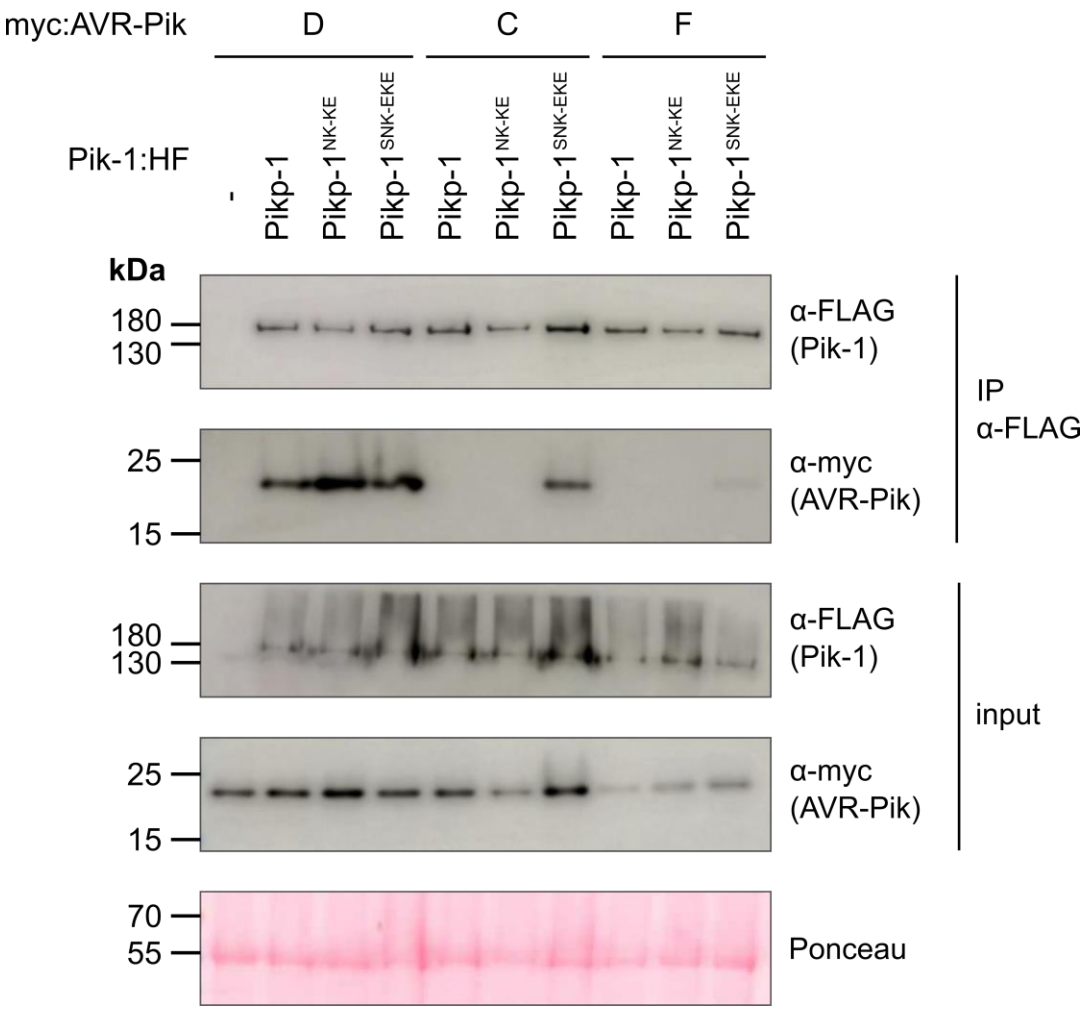


Figure 5.16 AVR-PikC and AVR-PikF co-immunoprecipitate with Pikp-1$^{SNK-EKE}$ but not with either Pikp-1^{NK-KE} or Pikp-1.

Western blots showing proteins present in the eluate following α-FLAG immunoprecipitation (IP) and proteins present in the plant extract prior to immunoprecipitation (input). Western blots are representative of 3 independent experiments.

mutation increases the binding affinity of Pikp-1^{NK-KE} for AVR-PikC and AVR-PikF to a sufficient level to trigger cell death *in planta*.

5.2.9 The crystal structure of the complex between Pikp-HMA$^{SNK-EKE}$ and AVR-PikC

To investigate whether the side chain of Pikp-HMA$^{SNK-EKE_Glu258}$ forms a similar hydrogen bond to that seen for OsHIPP19-HMA^{Glu72} in the crystal structure of OsHIPP19-HMA/AVR-PikF, and whether this mutation leads to any additional changes in the binding interface, I aimed to solve the crystal structure of Pikp$^{SNK-EKE}$-HMA in complex with AVR-PikC.

5.2.9.1 Purification and crystallisation of the Pikp-HMA$^{SNK-EKE}$/AVR-PikC complex

Intact mass spectrometry was performed by Dr Gerhard Saalbach.

Competent *E. coli* SHuffle cells were co-transformed with pOPIN-M::Pikp-HMA$^{SNK-EKE}$ and pOPIN-A::AVR-PikC. The pOPIN-M::Pikp-HMA$^{SNK-EKE}$ construct produces Pikp-HMA$^{SNK-EKE}$ with a N-terminal 6xHis:MBP tag (cleavable by addition of 3C protease), and the pOPIN-A::AVR-PikC construct produces untagged AVR-PikC. The complex was purified from *E. coli* cell lysate using a similar pipeline to that described in section 3.1.1 for OsHIPP19-HMA (see Materials and Methods for full details). The presence of both proteins in the final sample was confirmed by intact mass spectrometry.

Crystallisation trials were conducted using the sitting drop vapour diffusion method with the purified complex of AVR-PikC and Pikp-HMA$^{SNK-EKE}$ at either 10 mg/ml or 20 mg/ml in the commercial Morpheus™ screen. Large crystals, suitable for X-ray diffraction studies, were obtained in multiple conditions in the Morpheus™ screen at both protein concentrations. Five crystals were selected for harvesting (see figure 5.17 and table 5.4). Crystals were individually mounted in a loop, flash frozen in liquid N_2 and transferred to the Diamond Light Source synchrotron facility. All conditions in the commercial Morpheus™ screen are cryoprotected, so no additional cryoprotectant was necessary.

5.2.9.2 Data collection and structure solution

X-ray diffraction data was collected at beamline I03 of the Diamond Light Source synchrotron facility. Following preliminary screening of the five crystals, a diffraction dataset was collected for crystal 4 (see table 5.4) at a wavelength of 0.98 Å. The dataset comprised 3600 images, with an oscillation angle of 0.1° between images. Panel B of figure 5.17 shows the crystal mounted in the loop prior to data collection.

Data reduction was carried out using the Xia2 3dii pipeline (Grosse-Kunstleve et al., 2002, Sauter et al., 2004, Evans, 2006), and the unit cell parameters determined as a = 66.35, b = 83.13, c = 107.04, α = β = γ = 90. The scaled, but unmerged, data file was passed to AIMLESS (implemented in CCP4i2) (Evans and Murshudov, 2013, Winn et al., 2011), and, based on various merging statistics, the data processed to a maximum resolution of 2.05 Å with an overall R_{merge} of 8.1 %. Images 2701-3600 were excluded, as an increase in the value of R_{merge} towards the end of the run indicated radiation damage to the crystal. The data was processed in the space group $P\,2_1\,2_1\,2_1$. Data processing statistics are shown in table 5.5.

The phases were solved by molecular replacement using PHASER (McCoy et al., 2007), as implemented in CCP4i2. The partially refined structure of Pikp-$HMA^{NK\text{-}KE}$ in complex with AVR-PikC was used as a search model. Both copies of the Pikp-$HMA^{NK\text{-}KE}$ dimer/AVR-PikC monomer were supplied. A unique solution was found with a TF Z-score of 8.9. The asymmetric unit contains 2 copies of a dimer of Pikp-$HMA^{SNK\text{-}EKE}$ and a monomer of AVR-PikC.

5.2.9.3 Model refinement and validation

As the only difference in sequence between the model supplied for molecular replacement and the phased model is the Ser258Glu mutation, this residue was manually changed in COOT (Emsley et al., 2010) prior to initial refinement with REFMAC (implemented in CCP4i2) (Murshudov et al., 2011). Following this initial refinement, the R-factor was 24.9 % and the free R-factor was 28.8 %. Iterative rounds of manual rebuilding, refinement and validation were carried out using COOT and REFMAC. Refinement and validation statistics are shown in table 5.5.

Table 5.4 Crystallisation conditions that resulted in crystals of Pikp-HMA$^{SNK-EKE}$/AVR-PikC suitable for X-ray diffraction studies.

Crystal	Well	Condition details	[Protein]	Harvested by
1	A1	Buffer system 1, precipitant mix 1, divalents additive mix	20 mg/ml	CEMS
2	D3	Buffer system 1, precipitant mix 3, alcohols additive mix	10 mg/ml	CEMS
3			20 mg/ml	JM
4	D7	Buffer system 2, precipitant mix 3, alcohols additive mix	10 mg/ml	JM
5			10 mg/ml	CEMS

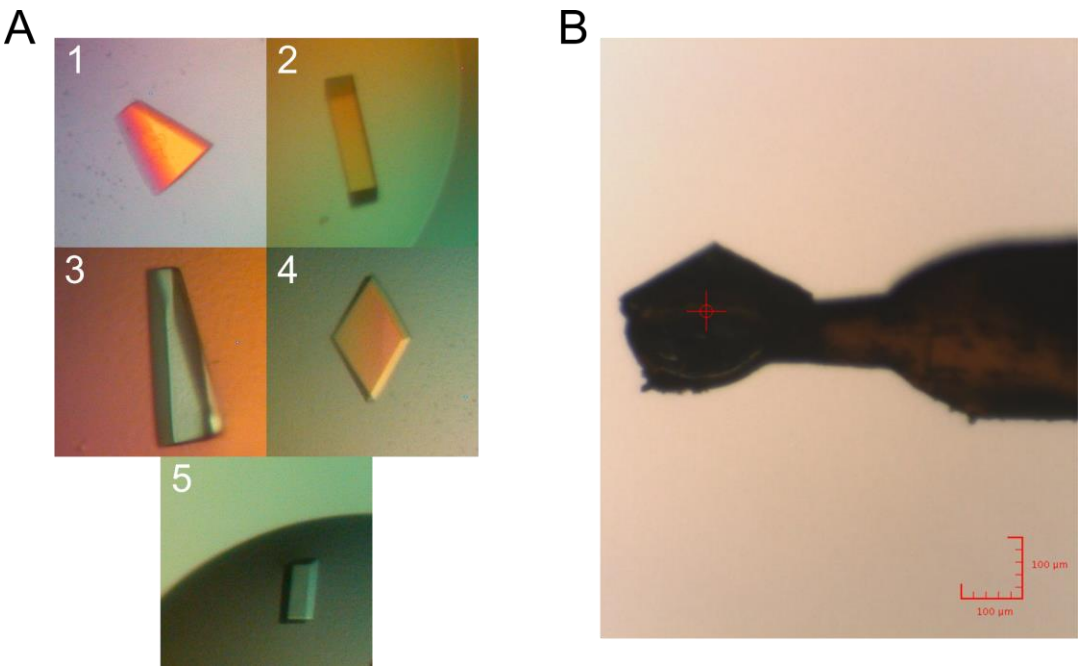


Figure 5.17 Protein crystals of Pikp-HMA$^{SNK-EKE}$ in complex with AVR-PikC.

A. *The 5 crystals selected to be sent to the Diamond Light Source synchrotron facility for X-ray diffraction studies. Numbers correspond to those in table 5.4.* ***B.*** *Crystal 4 mounted in a loop prior to X-ray diffraction data collection.*

Table 5.5 Data collection and refinement statistics for the Pikp-HMA$^{SNK-EKE}$/AVR-PikC complex.

Data collection statistics	
Wavelength (Å)	
Space group	P 2_1 2_1 2_1
Cell dimensions *a*, *b*, *c* (Å)	66.35, 83.13, 107.04
Resolution (Å)*	46.67-2.05 (2.11-2.05)
R_{merge} (%)	8.1 (104.8)
$I/\sigma I$	15.2 (2.0)
Completeness (%)	99.8 (99.4)
Unique reflections	37795 (2886)
Redundancy	9.9 (10.1)
CC(1/2) (%)	99.9 (83.8)
Refinement and model statistics	
Resolution (Å)	46.71-2.05 (2.10-2.05)
R_{work}/R_{free} (%)	20.4/25.9 (28.2/31.0)
No. atoms	
Protein	7016
Water	205
B-factors	
Protein	46.7
Water	46.8
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.589
Ramachandran plot (%)**	
Favoured	97.90
Allowed	2.10
Outliers	0
MolProbity Score	1.65 (93rd percentile)

* The highest resolution shell is shown in parenthesis.
** As calculated by MolProbity

The final model has an R-factor of 20.4 % and a free R-factor of 25.8 %. The N-terminal and C-terminal amino acids for each of the six protein chains are shown in table 5.6. As has been the case for previous structures of HMA domains in complex with AVR-Pik effector alleles, some residues in the loop between β1 and α1 of the HMA domain were poorly defined in the electron density and could not be positioned; excluded amino acids are listed in table 5.6. 205 water molecules were positioned in the final model.

Final validation of the model was carried out with MolProbity (Chen et al., 2010). The Ramachandran plots for the refined model show that all residues lie in allowed regions of the plots, with 97.90 % in favoured regions (figure 5.18). Rotamer analysis revealed 13 poor rotamers, however manual inspection in COOT confirmed that more favourable rotamers are not supported by the electron density. The clash score (number of steric overlaps per 1000 atoms) is 4.65, which is considered to be very good. The overall MolProbity score is 1.65, placing the model in the 93rd percentile of structures of comparable resolution.

5.2.10 The Ser258Glu mutation results in additional hydrogen bonds at interface 3

As would be anticipated, the global structures of Pikp-HMA$^{NK\text{-}KE}$/AVR-PikC and Pikp-HMA$^{SNK\text{-}EKE}$/AVR-PikC are very similar (figure 5.19). The overall RMSD, as calculated in COOT using secondary structure matching, is 0.38 Å using 154 residues. For the HMA domains only, the RMSD is 0.31 Å (using 72 residues) and for the AVR-PikC molecules, the RMSD is 0.20 Å (using 82 residues).

Interface analysis with QtPISA showed that, for the structure of Pikp-HMA$^{SNK\text{-}EKE}$/AVR-PikC, the binding interface buries 22.3 % and 18.2 % of the total accessible surface area of the HMA domain (978.5 Å^2) and effector (1042.6 Å^2) respectively. The interface between Pikp-HMA$^{SNK\text{-}EKE}$ and AVR-PikC is therefore fractionally larger than that of Pikp-HMA$^{NK\text{-}KE}$ and AVR-PikC (table 5.7). Key interface parameters are summarised in table 5.7 and represented in graphical form as radar plots in figure 5.20.

Strikingly, 15 hydrogen bonds and 9 salt bridges are formed between Pikp-HMA$^{SNK\text{-}EKE}$ and AVR-PikC, contrasting with the 12 hydrogen bonds and 8 salt bridges between Pikp-

Table 5.6 Summary of the amino acids included in the final model for each protein chain.

Chain	Protein	N-terminal residue	C-terminal residue	Excluded residues
A	Pikp-HMA$^{SNK-EKE}$	Leu187	Gln259	Gly198
B	* Pikp-HMA$^{SNK-EKE}$	Leu187	Glu262	Glu198 – Asn200
C	AVR-PikC	Asp34	Phe113	-
D	Pikp-HMA$^{SNK-EKE}$	Leu187	Ala260	Glu198 – Asn201
E**	* Pikp-HMA$^{SNK-EKE}$	Leu187	Glu262	Glu198 – Asn201
F**	AVR-PikC	Ala32	Phe113	-

* indicates that this copy of Pikp-HMA^{NK-KE} interfaces with AVR-PikC.
** indicates that these chains were used for interface analysis.

Table 5.7 Summary of interface analysis by QtPISA for Pikp-HMA$^{SNK-EKE}$/AVR-PikC.

		Pikp$^{SNK-EKE}$ / AVR-PikC	Pikp^{NK-KE} / AVR-PikC
AVR-Pik	B.S.A. (Å)	978.5	971.5
	% B.S.A. of total	18.2	18.0
HMA	B.S.A. (Å)	1042.6	1018.2
	% B.S.A. of total	22.3	21.5
Total interface area* (Å)		1010.6	994.9
Number of hydrogen bonds		15	12
Number of salt bridges		9	8

* Total interface area is the total B.S.A. (Buried Surface Area) of each component divided by two.

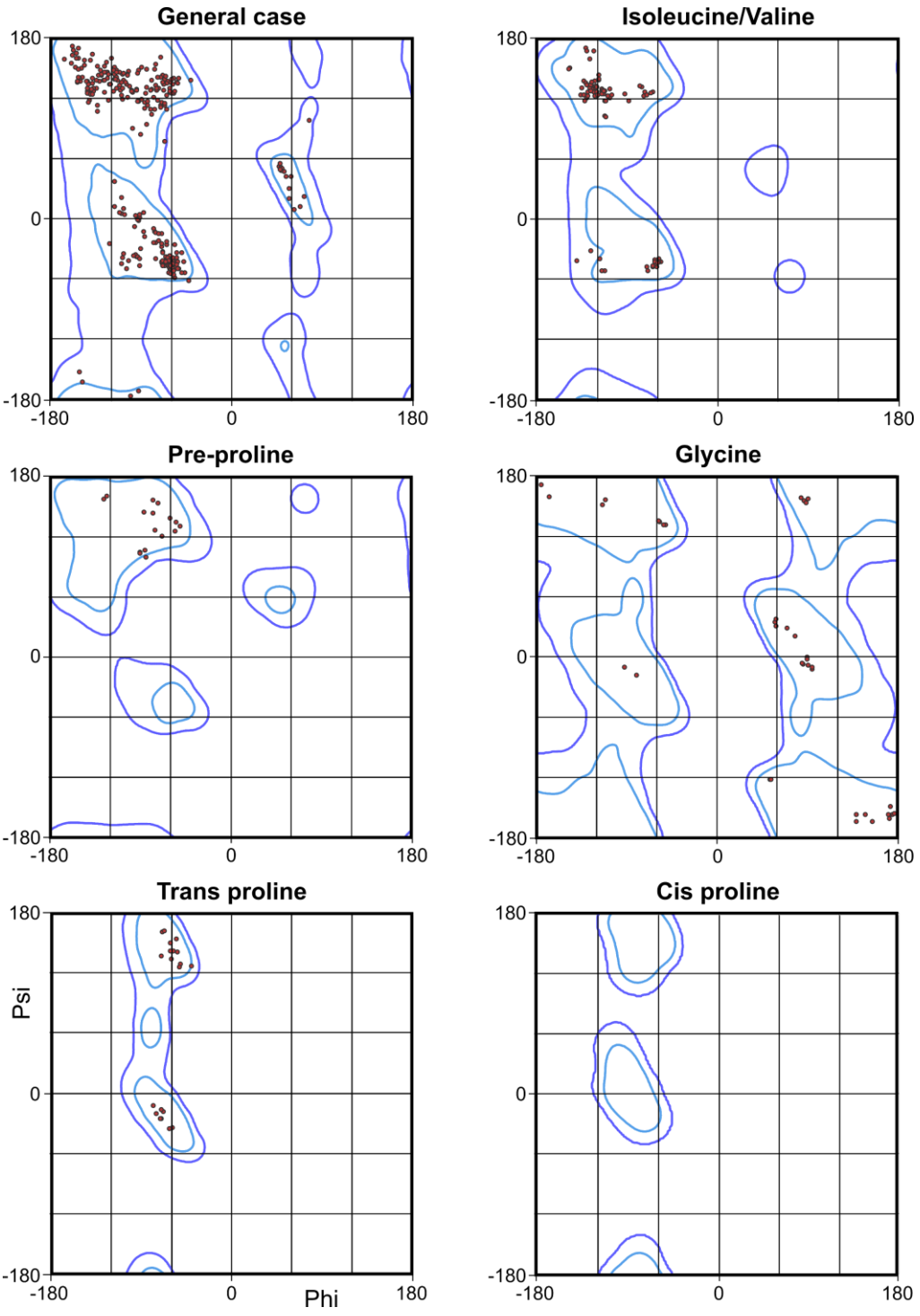


Figure 5.18 Ramachandran plots for the structure of AVR-PikC in complex with Pikp-HMA$^{SNK-EKE}$.

100% of residues were in allowed regions of the plots (within the purple lines), with 97.90% of residues lying within favoured regions (the blue lines). Ramachandran plots were produced with MolProbity.

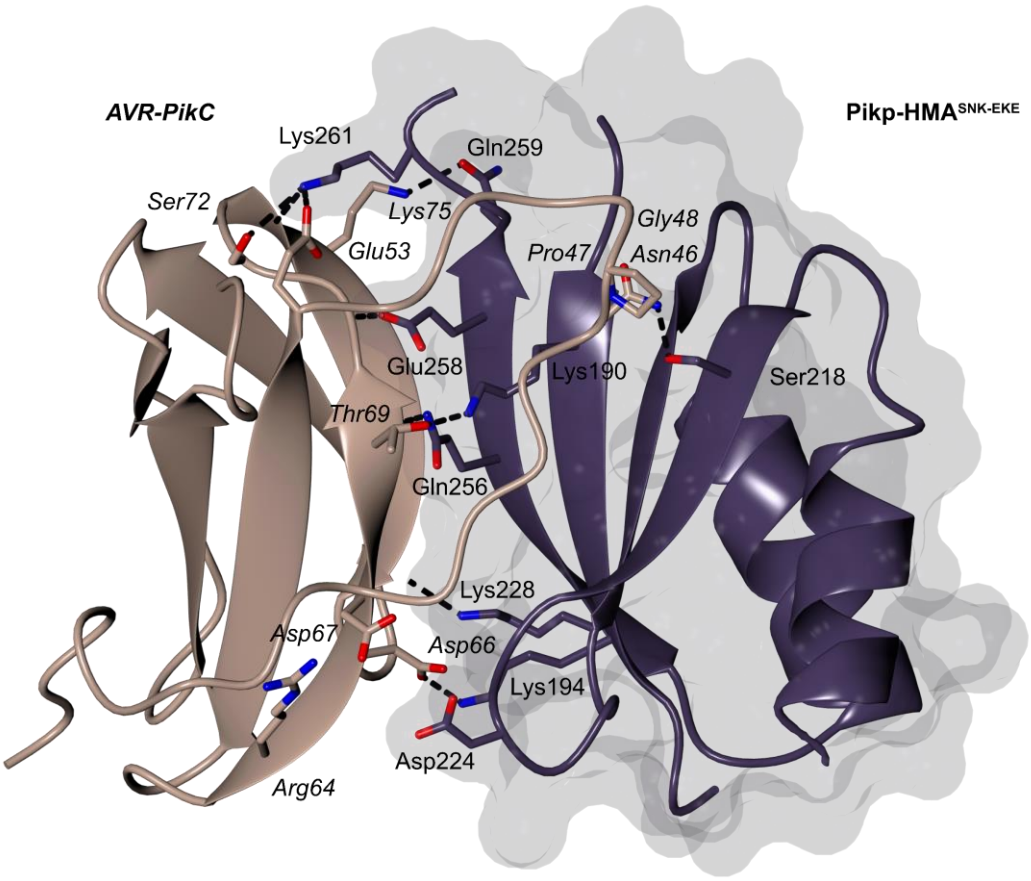
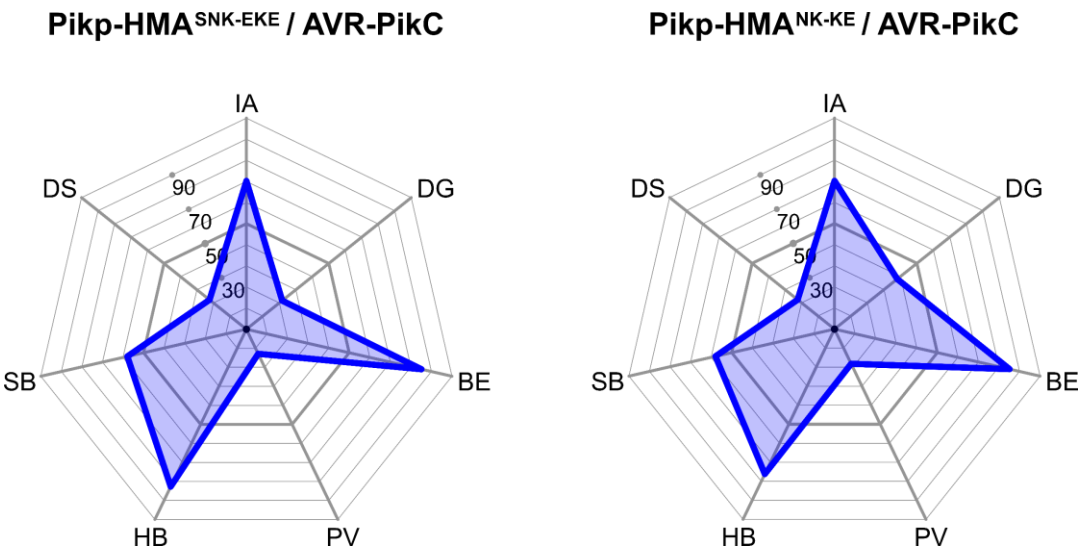


Figure 5.19 The crystal structure of Pikp-HMA$^{SNK-EKE}$ in complex with AVR-PikC.

The structures of AVR-PikC and Pikp-HMA$^{SNK-EKE}$ are represented as brown and purple ribbons respectively, with the molecular surface of Pikp-HMA$^{SNK-EKE}$ also displayed. The side chains of amino acids of interest are displayed as cylinders. Hydrogen bonds are represented by dashed lines.



	Pikp$^{SNK-EKE}$ / AVR-PikC	Pikp^{NK-KE} / AVR-PikC
IA: Interface area	1010.5 Å	994.8 Å
DG: Solvation energy	-3.2 kcal/mol	-5.1 kcal/mol
BE: Binding energy	-13.2 kcal/mol	-13.4 kcal/mol
PV: Hydrophobic P-value	0.5535	0.4997
HB: Hydrogen bonds	15	12
SB: Salt bridges	9	8
DB: Disulphide bonds	0	0

Figure 5.20 Comparison of the binding interfaces of Pikp-HMASNK-EKE/AVR-PikC and Pikp-HMANK-KE/AVR-PikC using QtPISA.

Radar plots produced by QtPISA provide a visual representation of the binding interfaces of the Pikp-HMA$^{SNK-EKE}$/AVR-PikC and Pikp-HMA^{NK-KE}/AVR-PikC complexes, based on the seven parameters outlined above. The radar beam for each parameter represents the probability that the value of that parameter indicates a biologically relevant assemblage in the PDB, with a probability of zero located at the centre of the radar.

HMA^{NK-KE} and AVR-PikC. The Ser258Glu point mutation therefore appears to have had a significant impact on hydrogen bonding at the binding interface.

Inspection of Pikp-HMA$^{SNK-EKE}$ revealed that, as anticipated, the side chain of Glu258 extends into the space between the effector and the HMA domain, and forms a direct hydrogen bond (2.9 Å) with the backbone of AVR-PikC (figure 5.21). Interestingly, however, this single mutation also supports two additional hydrogen bonds between AVR-PikC and residues comprising β4 of the HMA domain. The side chain of Pikp-HMA$^{SNK-EKE_Gln259}$ forms a hydrogen bond (3.0 Å) with the side chain of AVR-PikC^{Lys75}, and the side chain of Pikp-HMA$^{SNK-EKE_Gln256}$ forms a hydrogen bond (also 3.0 Å) with the main-chain carbonyl group of AVR-PikC^{Thr69} (figure 5.21). By contrast, in the structure of Pikp-HMA^{NK-KE}/AVR-PikC, the corresponding atoms are 3.7 Å apart. The Ser258Glu mutation appears to have shifted β4 of the HMA domain towards the effector, enabling the formation of additional hydrogen bonds at this interface.

As is the case for all structures between HMA domains and AVR-Pik effectors obtained to date, the interface between Pikp-HMA$^{SNK-EKE}$ and AVR-PikC is hydrated, with a network of hydrogen-bonded water molecules bridging the two proteins. Interestingly, the water molecules at the interface between Pikp-HMA$^{SNK-EKE}$ and AVR-PikC appear better defined in the electron density than those at the interface between Pikp-HMA^{NK-KE} and AVR-PikC. A tempting hypothesis is that the Ser258Glu mutation rigidifies the network of water molecules at the effector-HMA interface, further contributing to the stability of the protein complex.

5.3 Discussion

At present, no Pik alleles have been identified which recognise the AVR-PikC or AVR-PikF effector alleles. In this chapter, I have successfully engineered a Pik-1 variant (Pikp-1$^{SNK-EKE}$) which triggers cell death in response to the AVR-PikC and AVR-PikF in a model system. This engineering has been guided by the crystal structure of AVR-PikF in complex with OsHIPP19-HMA, and exploits the similarities and differences between the integrated HMA domain of Pik-1 and the HMA domain of the putative virulence target OsHIPP19.

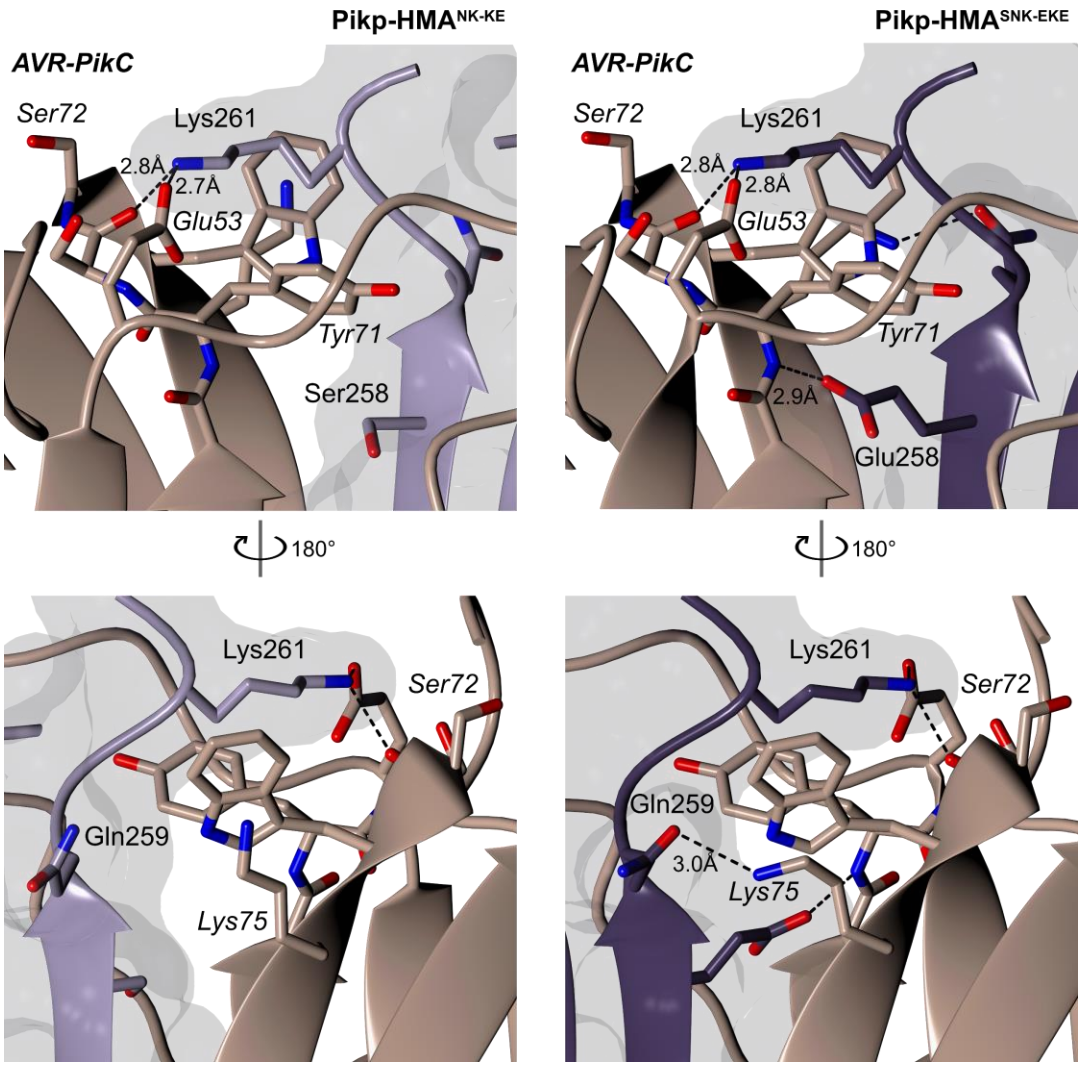


Figure 5.21 The Ser258Glu mutation leads to additional hydrogen bonds at interface 3 between Pikp-HMA$^{SNK-EKE}$ and AVR-PikC.

The structures are represented as lilac (Pikp-HMA^{NK-KE}), purple (Pikp-HMA$^{SNK-EKE}$) and brown (AVR-PikC) ribbons, with key residues shown as cylinders. Hydrogen bonds are represented as dashed lines. Bond lengths are shown for intermolecular hydrogen bonds. The molecular surface of the HMA domains are shown in lilac or purple, respectively. The bottom images are rotated 180° about the y-axis relative to the top images.

As discussed in chapter 1, the hypersensitive response activated by NLR-mediated effector recognition can be divided into cell death and pathogen resistance, and these two phenomena can be uncoupled during a pathogen response. Our assays in the model system *N. benthamiana* use cell death as a proxy for resistance in rice. These assays facilitate moderately fast and efficient screening of modified effector and NLR proteins; it would not be practical to conduct these assays using transgenic rice and transformed *M. oryzae*. Thus far, there is a tight correlation between cell death in *N. benthamiana* following agrobacterium-mediated transient expression of AVR-Pik alleles and Pik-1/Pik-2 NLR proteins, and resistance in rice containing the native Pik NLR proteins to *M. oryzae* isolates carrying AVR-Pik effector alleles. However, when drawing conclusions from these assays, it is important to keep in mind that we are using cell death in a model system as a proxy for resistance in a host.

The cell death assays are scored by using both the UV and daylight images to assign a score between 0 and 6 to each infiltrated spot (see scoring scale in figure 2.2). The leaves are scored blind by the same person across an experiment for consistency. This scoring system is inherently subjective; the possibility of developing an automated image pipeline was explored, but the method by which the leaves are illuminated (using hand-held UV lights) during imaging presented a barrier to this. The seven-point scoring scale has an advantage over a "cell death or no cell death" binary scoring, as it allows variation in the extent of cell death both within and between samples to be recorded and presented. However, a quantitative technique, such as an ion leakage assay, could enable more robust comparisons to be made between samples. Ion leakage assays quantify the movement of electrolytes out of plant tissues as a result of the loss of the integrity of the cell membrane which typically occurs during cell death (Hatsugai and Katagiri, 2018). This approach would be particularly valuable for the experiment investigating the difference in cell death response between Pikh-1 and Pikp-1$^{NK\text{-}KE}$ to the different AVR-Pik alleles (section 5.2.4). While the cell death assay indicates that Pikp-1$^{NK\text{-}KE}$ responds more strongly than Pikh-1 to each of the three AVR-Pik alleles, ion leakage assays would enable quantification of the extent of cell death in each case. However, when screening Pikp-1 variants for an extended response to AVR-PikC and AVR-PikF, the observation of cell death was sufficient to select candidates to take forward for *in vitro* studies and, importantly, into experiments in the rice/*M. oryzae* pathosystem.

The HMA domains of Pikp-1 and OsHIPP19 share 51% amino acid sequence similarity and are structurally very alike. However, when the HMA domain of Pikp-1 was exchanged for that of OsHIPP19, the resulting chimera (Pikp-1^{HIPP19}) was autoactive, triggering Pikp-2-dependent cell death in the absence of the effector. While the binding interface between the integrated HMA domain and AVR-Pik effector is well-characterised, the mechanism by which binding of the effector activates immune signalling by the NLR pair is unknown. It could be speculated that effector binding triggers a conformational change in Pik-1 and/or Pik-2, exposing surfaces for interaction with downstream signalling components. Exchanging the integrated Pikp-HMA domain for that of OsHIPP19 could cause similar conformational changes to allow activation of immune signalling in the absence of the effector. The cryo-EM structures of the *A. thaliana* NLR protein ZAR1 have shown that activation of this NLR protein involves oligomerisation to form a pentameric resistosome, with conformational changes in the coiled-coil domain exposing the N-terminal α-helix (Wang et al., 2019a). The HMA domain of Pik-1 is immediately adjacent to the coiled-coil domain, and if activation of Pik-1 and Pik-2 also requires rearrangement of their coiled-coil domains, introducing the HMA domain of OsHIPP19 could trigger conformational changes similar to those that occur during activation. Perhaps unsurprisingly, it appears that following the integration of the HMA domain into Pik-1, there has been close co-evolution of the HMA domain with the other domains of Pik-1 and Pik-2, potentially involving the acquisition of compensatory mutations in both the HMA domain and other Pik-1/Pik-2 domains, to enable the NLR proteins to detect AVR-Pik and instigate an effector-dependent immune response. While it would be highly desirable to be able to exchange integrated domains for other, unrelated, protein domains to generate NLR proteins with custom recognition specificities, the finding that exchanging Pikp-HMA for one with 51 % shared sequence identity results in autoactivity suggests that at present this is still a distant possibility. It should be noted that while Pikp-1^{HIPP19} is constitutively active in *N. benthamiana*, the activity of this protein in rice is unknown. I hypothesise that Pikp-1^{HIPP19} would also be autoactive in rice; the dependency of the autoactive phenotype on Pikp-2 suggests that the cell death is due to the specific activation of Pikp-1^{HIPP19} and Pikp-2, rather than the overexpression of a potentially misfolding or aggregating chimeric protein. If Pikp-1^{HIPP19} is similarly autoactive in rice, constitutive expression of

the gene could be lethal. Generating stable transgenic rice with Pikp-1^{HIPP19} and Pikp-2 under the control of an inducible promoter would allow the activity of Pikp-1^{HIPP19} to be assessed.

The AVR-Pik recognition profile of Pikh remains an area of ongoing research. Previously published research suggests that Pikh does not recognise or respond to AVR-PikC (Kanzaki et al., 2012), however recent experiments observed that near-isogenic rice lines (NILs) containing Pikh are resistant to *Magnaporthe oryzae* carrying AVR-PikC. To confirm that another resistance gene in the NIL is not responsible for the observed resistance to *Magnaporthe oryzae* carrying AVR-PikC, Pikh should be knocked out. If the knockout line loses resistance to *M. oryzae* carrying AVR-PikC, it can be concluded that Pikh is responsible for conferring the observed resistance. It is also possible that the resistance profile of Pikh depends on environmental conditions, the developmental stage of the plant during infection, or the method of inoculation, and that differences in any of these could contribute to the different observations regarding the resistance of Pikh.

Pikp^{NK-KE} gives a qualitatively stronger cell death response to AVR-PikD, AVR-PikE or AVR-PikA than Pikh. In the structures of Pikp-1^{NK-KE}, Pikp-1$^{NK-KE_Glu262}$ is positioned away from the binding interface and does not contribute to the interaction with the effector. It could be considered surprising that this mutation appears to increase the cell death response of the NLR protein to the effector. However, in Pikh-HMA, it is possible that the side chains of the two adjacent lysine residues (Lys261 and Lys262) compete for binding into the pocket of negative surface charge formed on the surface of the AVR-Pik effector. Binding of Pikh-1^{Lys262} into the pocket of negative surface charge would likely result in the looping out of the Pikh-1 main chain away from the effector as was observed in the structure of Pikp-HMA (De la Concepcion et al., 2018), weakening the interface. By contrast, binding of Pikh-1^{Lys261} into the pocket would potentially bring the main-chain in towards the effector, as seen in the structure of Pikm-HMA (De la Concepcion et al., 2018). The additional Lys262Glu mutation in Pikp-1^{NK-KE} could remove the competition between the two lysine side chains in Pikh-1 and ensure the binding of the side chain of Lys261 into the pocket.

The crystal structure of AVR-PikC in complex with Pikp-HMA^{NK-KE} has revealed how the single Ala67Asp polymorphism that distinguishes AVR-PikC from AVR-PikE allows

AVR-PikC to avoid recognition by Pik-1. AVR-PikC^{Asp67} lies at the interface between the effector and Pikp-HMA. The side chain of AVR-PikC^{Asp67} extends towards the HMA domain and prevents the formation of two intermolecular hydrogen bonds between AVR-PikC^{Arg64} and Pikp-HMA^{Asp224}. Additionally, a combination of steric clash and repulsion caused by the matching charges of the AVR-PikC^{Asp67} and Pikp-HMA^{Asp224} side chains shifts the loop between β2 and β3 containing Pikp-HMA^{Asp224} away from the effector.

Initial attempts at engineering a Pik-1 variant that responds to AVR-PikC focused on accommodating the disruptive AVR-PikC^{Asp67} side chain by mutating Pikp-1$^{NK\text{-}KE_Asp224}$. I hypothesised that mutating the aspartate to alanine would alleviate the repulsion and steric clash between the two aspartate side chains, bringing the loop between β2 and β3 back towards the effector. I also generated Pikp-1$^{DNK\text{-}KKE}$, hypothesising that mutating the aspartate to a lysine might result in the formation of alternative hydrogen bonds. However, neither of these mutations extended the response of Pikp-1$^{NK\text{-}KE}$ to AVR-PikC. As both mutations would abolish the formation of the two hydrogen bonds between Pikp-HMA$^{NK\text{-}KE_Asp224}$ and AVR-PikD^{Arg64}, unsurprisingly both mutants exhibited a reduced cell death response to AVR-PikD.

The two residues that differentiate Pikp-1$^{NK\text{-}KE}$ from Pikp-1 are both located at interface 3, and the Asn261Lys and Lys262Glu mutations extended the recognition of Pikp-1 to AVR-PikE and AVR-PikA (De la Concepcion et al., 2019). However, the polymorphic amino acids that distinguish AVR-PikE and AVR-PikA from AVR-PikD are located at interface 2; strengthening the interaction at interface 3 compensated for the disruptive effect of the adaptive polymorphisms in AVR-PikE and AVR-PikA.

By making an additional point mutation (Ser258Glu) in Pikp-1$^{NK\text{-}KE}$, I was able to increase its binding affinity for AVR-PikC and AVR-PikF. The crystal structure of Pikp-1$^{SNK\text{-}EKE}$ revealed that, in addition to the predicted additional hydrogen bond between the side chain of Pikp-1$^{SNK\text{-}EKE_Glu258}$ and the backbone of AVR-PikC, two additional intermolecular hydrogen bonds form between other amino acids at the interface. I hypothesise that this is due to β4 of the HMA domain being brought slightly closer to the effector, allowing hydrogen bonds to form between residues that were previously too distant for intermolecular hydrogen bonding to occur.

In all structures between AVR-Pik effectors and HMA domains solved to date, the interface between the two proteins is highly hydrated, with an extensive network of water molecules bridging the gap between the two proteins. Interestingly, the network of water molecules at the Pik-HMA/AVR-PikC interface appears rigidified in the crystal structure of Pikp-1$^{SNK\text{-}EKE}$ / AVR-PikC compared to the structure of Pikp-1$^{NK\text{-}KE}$ / AVR-PikC. At present, available programmes for analysis of macromolecular interfaces, such as QtPISA, do not consider water molecules. Recent advances in both data collection hardware and data processing programmes have increased the quality of electron density maps and enabled water molecules to be modelled with greater accuracy. A challenge for the field of macromolecular structural biology is to develop algorithms that consider the likely effects of water molecules at protein-protein interfaces when determining parameters such as the binding energy or solvation energy of the interface. This could be used to inform engineering efforts to further increase the affinity of Pik-HMA for the AVR-Pik effectors; replacing specific interface residues with hydrophobic residues to exclude water molecules from the interface could improve the affinity of the complex.

I have used a structure-guided engineering approach to generate a Pik-1 variant that triggers a cell death response in the model system *Nicotiana benthamiana* when co-expressed with Pikp-2 and AVR-PikC or AVR-PikF. While the correlation between cell death in *N. benthamiana* and blast resistance in rice makes this a promising candidate for delivering resistance to *M. oryzae* strains carrying AVR-PikC or AVR-PikF, it is necessary to generate stable transgenic rice lines carrying this engineered NLR protein and challenge these with *M. oryzae* strains expressing these effectors.

6

Engineering a Pikp-1:OsHIPP19 chimera with extended recognition capability

6

Engineering a Pikp-1:OsHIPP19 chimera with extended recognition capability

6.1 Introduction

In chapter 5, I demonstrated that exchanging the HMA domain of Pikp-1 for that of OsHIPP19 results in an autoactive chimera (Pikp-1^{HIPP19}) which triggers HR-like cell death in the absence of an effector. Such autoactive NLR proteins are considered to be of limited use for conferring pathogen resistance in the field; stable transgenic plants containing autoactive NLR proteins have been shown to exhibit severe stunting, fail to set seed (Zhou et al., 2019, Li et al., 2019b), or display spontaneous HR-like necrotic lesions across leaves and stems (Hu et al., 1996). While there is some evidence to suggest that low expression of an autoactive NLR protein can "prime" a plant for disease resistance (Li et al., 2019b), a comprehensive study on associated fitness costs in terms of seed set and crop yield is so far lacking.

The principal aim of the work described in this chapter was to modify Pikp-1^{HIPP19} to prevent autoactivity, without preventing an effector-dependent response. I hypothesised that certain amino acids in the HMA domain of Pikp-1 are required for the appropriate activation of the NLR protein, and that exchanging these amino acids for the corresponding residues in OsHIPP19-HMA leads to autoactivation. If these amino acids could be identified, it could be possible to engineer a modified chimera which is not constitutively active but still responds to AVR-PikC and AVR-PikF.

A secondary objective is to use the autoactive Pikp-1^{HIPP19} chimera to probe the mechanism underlying cell death signaling by the Pik-1/Pik-2 NLR proteins. While the

molecular mechanisms underlying the detection of AVR-Pik by Pik-1/Pik-2 have been well-characterised (Maqbool et al., 2015, De la Concepcion et al., 2018), how this translates into downstream signalling for activation of defence responses is less well understood.

As referenced in section 1.4.4, nucleotide exchange mediated by the NB-ARC domain is an integral part of NLR protein activation. Located in the NB-ARC domain, the P-loop (also known as the Walker A motif) has the consensus sequence GxxxxGKS/T. A functional P-loop is required for nucleotide binding, with the highly conserved lysine binding the β- and γ-phosphates of ATP. Mutation of the lysine to arginine has been reported to reduce ATP binding, and thus prevent the ATPase activity of NLR necessary for signalling (Tameling et al., 2002, Takken and Goverse, 2012, Takken et al., 2006, Wu et al., 2017, Deyrup et al., 1998). Also located in the NB-ARC domain, the MHD motif facilitates ADP binding. Mutation of the highly conserved aspartate to a valine is thought to destabilise ADP interaction leading to an increase in the active ATP-bound form of the NLR. Correspondingly, this mutation causes autoactivity in several NLR proteins (Williams et al., 2011, Bendahmane et al., 2002, Tameling et al., 2006, Van Ooijen et al., 2008).

Previous work has demonstrated that the functional P-loops of both Pikp-1 and Pikp-2 are required for effector-triggered HR-like cell death in the model system *Nicotiana benthamiana*, as mutation of the conserved lysine in the P-loop of either Pikp-1 or Pikp-2 to arginine prevents the cell death phenotype (Rafal Zdzralek and Professor Hiromasa Saitoh, personal communication). This contrasts with observations for the paired NLR proteins RGA4/RGA5 and RRS1/RPS4, where the P-loop of the sensor NLR protein (RGA5 and RRS1, respectively) is dispensable for immune signalling (Sohn et al., 2014, Cesari et al., 2013, Cesari et al., 2014b). Furthermore, it was found that mutation of the conserved aspartate of the MHD motif of Pikp-2 to valine does not cause autoactivation as has been reported for other NLR proteins, but instead prevents the Pik-mediated cell death. The MHD motif of Pikp-1 is degenerate (IHP), however mutating each of these three residues to alanine prevents effector-triggered cell death. Figure 6.1 shows the amino acid sequences of Pikp-1 and Pikp-2 with the P-loop and MHD motifs highlighted.

P-loop MHD motif

Pikp-1

```
   1  MEAAAMAVTAATGALAPVLVKLAALLDDGECNLLEGSRSDAEFIRSELEAVHSLLTPNILGRMGD
  66  DDAACKDGLIAEVRELSYDLDDAVDDFLELNFEQRRSASPFGELKARVEEHVSNRFSDWKLPAAS
 131  LPPSSVHRRAGLPPPDAELVGMDKRMEELTKLLEQGSNDASRWRKRKPHFPLRKTGLKQKIVIKV
 196  AMEGNNCRSKAMALVASTGGVDSVALVGDLRDKIEVVGYGIDPIKLISALRKKVGDAELLQVSQA
 261  NKDVKETTPMLAPVKSICEFHKVKTVCILGLPGGGKTTVARELYDALGTHFPCRVFVSVSPSSSP
 326  SPNLTKTLADIFAQAQLGVTDTLSTPYGGSGTGRALQQHLIDNISAFLLNKKYLIVIDDIWHWEE
 391  WEVIRKSIPKNDLGGRIIMTTRLNSIAEKCHTDDNDVFVYEVGDLDNNDALSLSWGIATKSGAGN
 456  RIGTGEDNPCYDIVNMCYGMPLALIWLSSALVGEIEELGGAEVKKCRDLRHIEDGILDIPSLQPL
 521  AESLCLGYNHLPLYLRTLLLYCSAYHWSNRIERGRLVRRWIAEGFVSEEKEAEGYFGELINRGWI
 586  TQHGDNNSYNYYEIHPVMLAFLRCKSKEYNFLTCLGLGSDTSTSASSPRLIRRLSLQGGYPVDCL
 651  SSMSMDVSHTCSLVVLGDVARPKGIPFYMFKRLRVLDLEDNKDIQDSHLQGICEQLSLRVRYLGL
 716  KGTRIRKLPQEMRKLKHLEILYVGSTRISELPQEIGELKHLRILDVRNTDITELPLQIRELQHLH
 781  TLDVRNTPISELPPQVGKLQNLKIMCVRSTGVRELPKEIGELNHLQTLDVRNTRVRELPWQAGQI
 846  SQSLRVLAGDSGDGVRLPEGVCEALINGIPGATRAKCREVLSIAIIDRFGPPLVGIFKVPGSHMR
 911  IPKMIKDHFRVLSCLDIRLCHKLEDDDQKFLAEMPNLQTLVLRFEALPRQPITINGTGFQMLESF
 976  RVDSRVPRIAFHEDAMPNLKLLEFKFYAGPASNDAIGITNLKSLQKVVFRCSPWYKSDAPGISAT
1041  IDVVKKEAEEHPNRPITLLINAGYKEISTESHGSSENIAGSSGIDTEPAQAQHDNLPAVRDDYKG
1106  KGILLDGRCPTCGRATKIEEETQDRVADIEIQTETTS
```

Pikp-2

```
   1  MELVVGASEATMKSLLGKLGNLLAQEYALISGIRGDIQYINDELASMQAFLRDLSNVPEGHSHGH
  61  RMKDWMKQIRDIAYDVEDCIDDFAHRLPQDSISDAKWSFLLTKIYELWTWWPRRVIASNIAQLKV
 131  RAQQIADRRSRYGVNNPEHLDSSSSARTRAVNYEIAEYQVTSPQIIGIKEPVGMKTVMEELEVWL
 196  TNPQAENGQAVLSIVGFGGVGKTTIATALYRKVSDKFQCRASVAVSQNYDQGKVLNSILSQVSNQ
 261  EQGSSTTISEKKNLTSGAKSMLKTALSLLRGNCICQPENDGNPDNTPIRLQETTDDDQNPRKLEQ
 326  LLAEKSYILLIDDIWSAETWESIRSILPKNNKGGRIIVTTRFQAVGSTCSPLETDRLHTVDFLTD
 391  DESQNLFNTSICESKIRKDSNKVDEQVPEEIWKICGGLPLAIVTMAGLVACNPRKACCDWSKLCK
 456  SLFPEQETPLTLDGVTRILDCCYNDLPADLKTCLLYLSIFPKGWKISRKRLSRRWIAEGFANEKQ
 521  GLTQERVAEAYFNQLTRRNLVRPMEHGSNGKVKTFQVHDMVLEYIMSKSIEENFITVVGGHWQMT
 586  APSNKVRRLSMQSSGSNRGSSTKGLNLAQVRSLTVFGNLNHMPFHSFNYGIIQVLDLEDWKGLKE
 651  RHMTEICQMLLLKYLSIRRTEISKIPSKIQKLEYLETLDIRETYVRDLPKSIVQLKRIISILGGN
 716  KNTRKGLRLPQEKSKKPIKNPSPQGKTKEPAKKGFLSQEKGKGAMKALRVLSGIEIVEESSEVAA
 781  GLHQLTGLRKLAIYKLNITKGGDTFKQLQSSIEYLGSCGLQTLAINDENSEFINSLGDMPAPPRY
 846  LVALELSGKLEKLPKWITSITTLNKLTISVTVLRTETLEILHILPSLFSLTFAFSLSAAKQDQDI
 911  IKDILENNKLDSDGEIVIPAEGFKSLKLLRFFAPLVPKLSFLDKNAMPALEIIEMRFKDFEGLFG
 976 IEILENLREVHLKVSDGAEAITKFLVNDLKVNTEKPKVFVDGIVTA
```

Figure 6.1 The P-loop and MHD motif of Pikp-1 and Pikp-2.

Amino acid sequences of Pikp-1 and Pikp-2. The P-loop motif (GxxxxGKT/S) is highlighted in gold, and the MHD motif is highlighted in blue.

6.2 Results

6.2.1 Autoactivation of the Pikp-1^{HIPP19} chimera requires the P-loop and MHD motif of Pikp-2

Constructs for transient expression of Pikp-2^{K217R}:HA and Pikp-2^{D559V}:HA were cloned with Rafal Zdrzalek (John Innes Centre) and Professor Hiromasa Saitoh.

While the autoactive Pikp-1^{HIPP19} chimera does not, as hoped, function as an immune receptor with extended response to AVR-PikC and AVR-PikF, it can nevertheless be used as a tool to probe the mechanism by which Pikp-1 and Pikp-2 co-operate to deliver effector-dependent immunity. As previous work has shown that intact P-loops and MHD motifs in both Pikp-1 and Pikp-2 are required for effector-triggered cell death, I aimed to investigate whether these are similarly required for the effector-independent cell death observed when Pikp-1^{HIPP19} and Pikp-2 are co-expressed.

To investigate whether the Pikp-2-dependent autoactivity of Pikp-1^{HIPP19} requires the P-loop of Pikp-1^{HIPP19}, the P-loop of Pikp-2 and the MHD motif of Pikp-2, level 1 constructs were generated for transient expression of Pikp-1^{HIPP19_K296R}:HF, Pikp-2^{K217R}:HA and Pikp-2^{D559V}:HA in *N. benthamiana*. Pikp-1^{HIPP19_K296R}:HF was co-expressed with WT Pikp-2:HA to investigate the role of the P-loop of Pikp-1^{HIPP19} in autoactivity. Pikp-1^{HIPP19}:HF was co-expressed with Pikp-2^{K217R}:HA or Pikp-2^{D559V}:HA to test whether the P-loop and MHD motif, respectively, are required for Pikp-1^{HIPP19} autoactivity. Pikp-1^{HIPP19}:HF was co-expressed with Pikp-2:HA as a positive control, and with empty vector as a negative control.

No HR-like cell death was observed when Pikp-1^{HIPP19}:HF was co-expressed with either Pikp-2^{K217R}:HA or Pikp-2^{D559V}:HA, indicating that both the P-loop and MHD motif of Pikp-2 are required for autoactivity of Pikp-1^{HIPP19}. Intriguingly, the cell death response was attenuated, but not abolished, when Pikp-1^{HIPP19_K217R} was co-expressed with Pikp-2 (figure 6.2). Production of proteins in *N. benthamiana* was confirmed by Western blot. This suggests that the P-loop of Pikp-2 is sufficient to partially activate downstream signalling, but that an intact P-loop of Pikp-1 is required for a full cell death response.

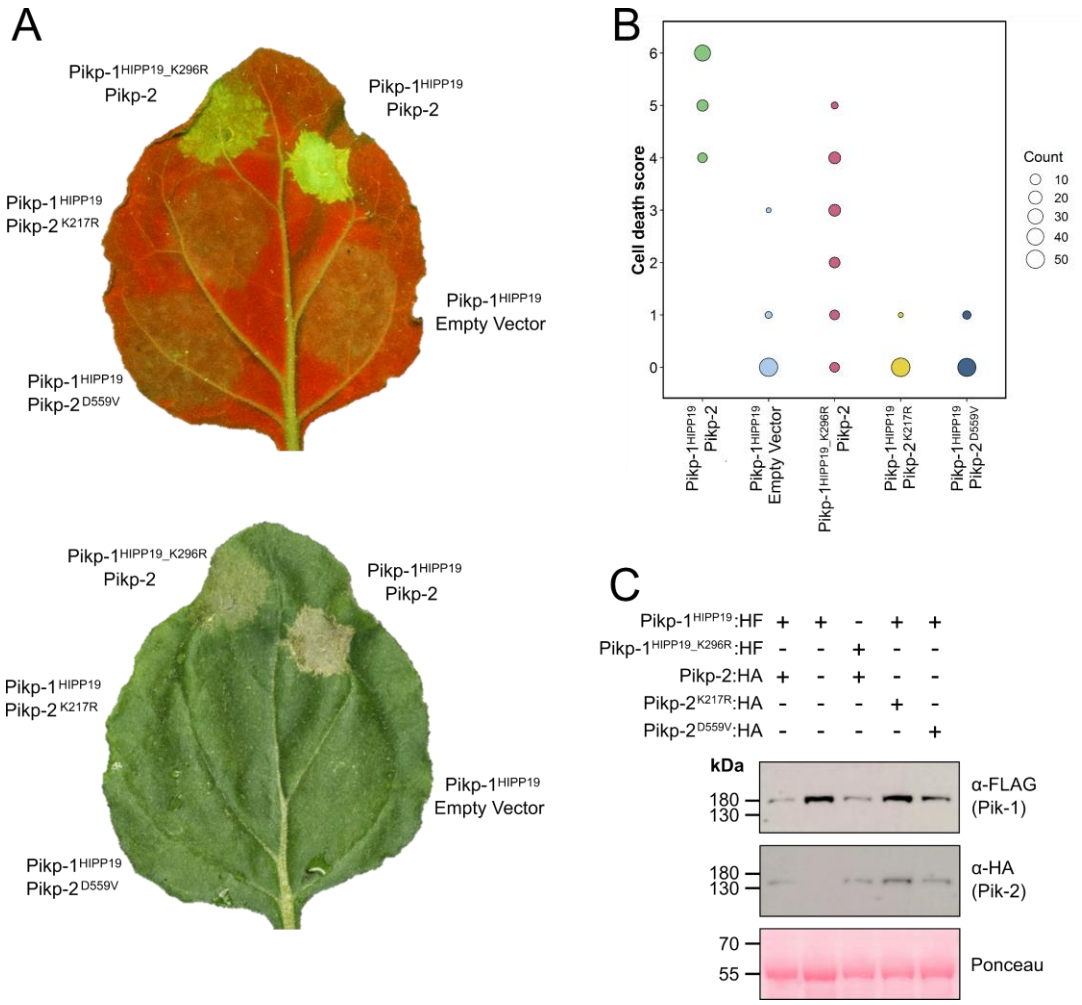


Figure 6.2 The autoactivity of Pikp-1^{HIPP19} requires the P-loop and MHD motif of Pikp-2.

A. *Representative leaf images showing cell death response under UV light (top) and daylight (bottom) showing the extent of cell death 5 dpi. The image under daylight has been horizontally flipped to provide the same leaf orientation for both images.* ***B.*** *Dot plot summarising the results obtained from 54 leaves in 3 independent experiments. Leaves were scored according to the scale in figure 2.2. The size of the dot at each cell death value is proportional to the number of leaves receiving that score. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).* ***C.*** *Western blots confirming production of the proteins in* N. benthamiana.

6.2.2 All AVR-Pik effector alleles associate with Pikp-1^{HIPP19} in co-immunoprecipitation experiments

Two of the three replicates of the co-immunoprecipitation experiment were conducted by Vera Sham (John Innes Centre) under my supervision.

While Pikp-1^{HIPP19} is autoactive and therefore not a functional immune receptor, I wanted to test whether the integration of the HMA domain of OsHIPP19 facilitates an interaction between AVR-PikC/AVR-PikF and the chimeric Pikp-1^{HIPP19} NLR protein. If Pikp-1^{HIPP19} interacts with AVR-PikC and AVR-PikF, there is the potential to modify the chimera to prevent autoactivity but retain a response to AVR-PikC and AVR-PikF.

Co-immunoprecipitation experiments were used to test for an association between each of the AVR-Pik effectors and Pikp-1^{HIPP19}. Each of the AVR-Pik effectors, with an N-terminal myc-tag, was transiently co-expressed with Pikp-1^{HIPP19}:HF in *N. benthamiana*. Pikp-2 was not included to prevent the strong cell death response that is triggered when Pikp-1^{HIPP19} and Pikp-2 are co-expressed. As a positive control, Pikp-1:HF was co-expressed with myc:AVR-PikD. As negative controls, Pikp-1:HF was co-expressed with myc:AVR-PikC, and Pikp-1^{HIPP19} was co-expressed with empty vector. Plant cell extracts were incubated with α-FLAG magnetic beads to immunoprecipitate Pikp-1^{HIPP19}:HF. Following multiple wash steps, bound proteins were eluted from the beads by boiling. The presence of the proteins in the input and eluate was determined by Western blotting. Each of the myc-tagged AVR-Pik effector alleles, including AVR-PikC and AVR-PikF, co-immunoprecipitated with Pikp-1^{HIPP19}, as evidenced by the bands of the appropriate size present in the eluate following immunoblotting with α-myc (figure 6.3).

From this, I can conclude that the integration of the HMA domain of OsHIPP19 into Pikp-1 extends binding of the NLR protein to other AVR-Pik effector alleles, including AVR-PikC and AVR-PikF. If the region of the HMA domain responsible for the autoactive phenotype of Pikp-1^{HIPP19} could be identified and modified to prevent autoactivation, the resulting NLR protein could be capable of triggering a cell death response to AVR-PikC and AVR-PikF.

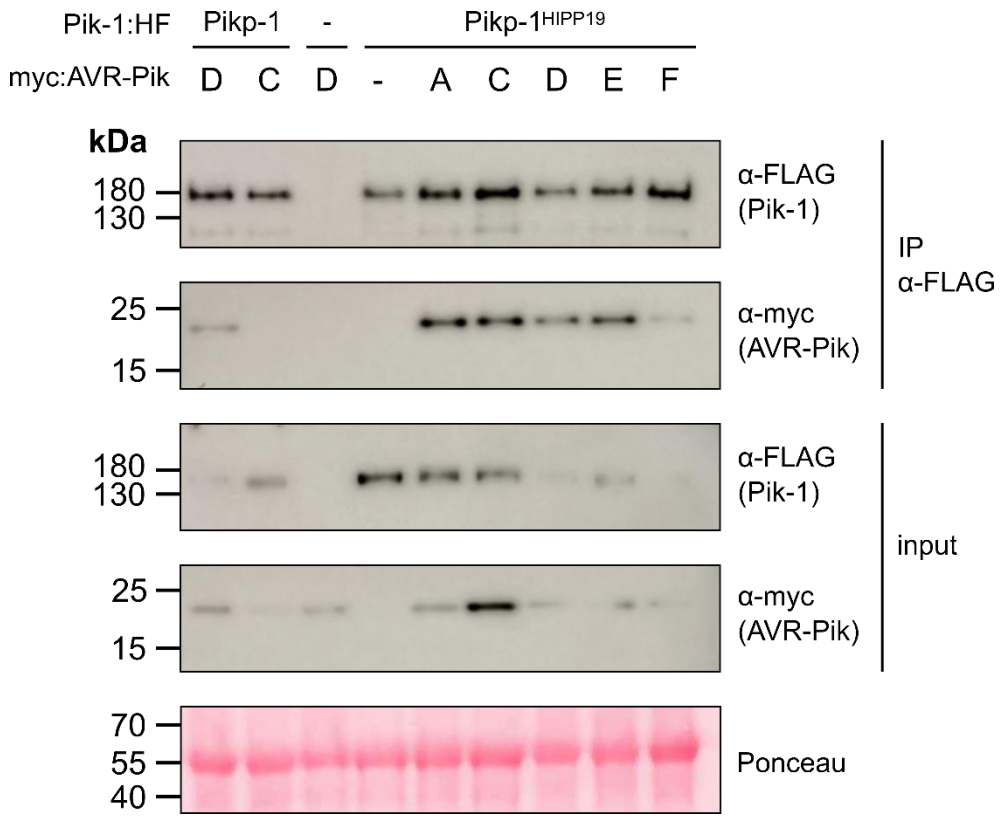


Figure 6.3 All AVR-Pik alleles co-immunoprecipitate with Pikp-1^{HIPP19}.

Western blots showing proteins present in the eluate following α-FLAG immunoprecipitation (IP) and proteins present in the plant extract prior to immunoprecipitation (input). Western blots are representative of 3 independent experiments.

6.2.3 Exchanging seven amino acids in Pikp-1^{HIPP19} for the corresponding amino acids in Pikp-1 removes autoactivity and the resulting NLR (Pikp-1$^{HIPP19\text{-}mbl7}$) responds to AVR-PikC and AVR-PikF

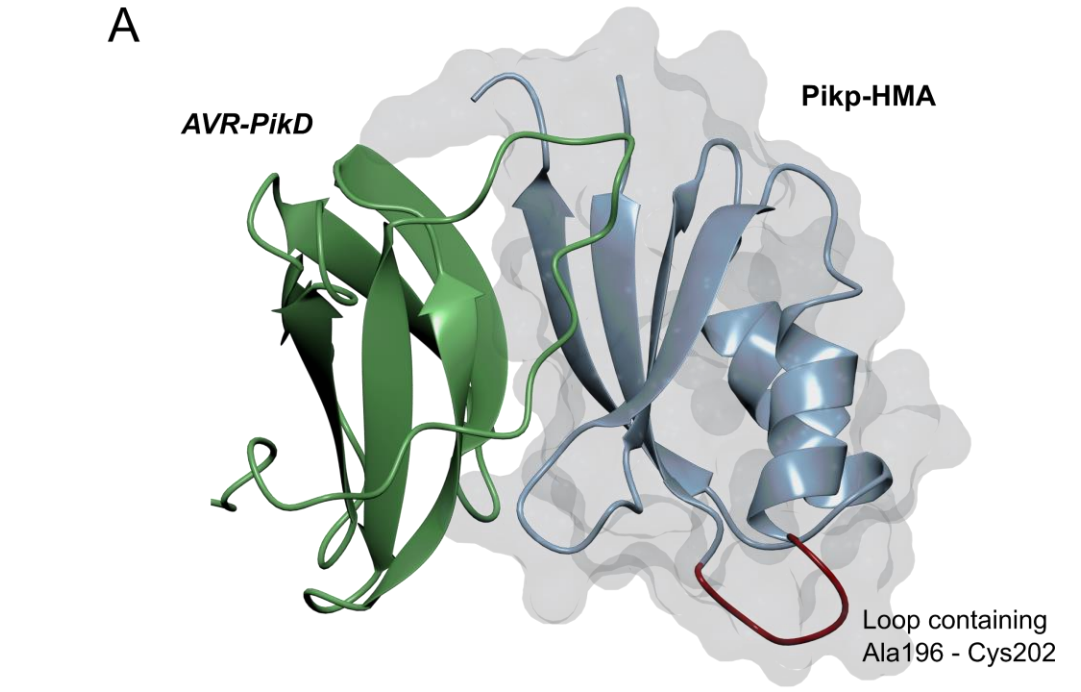
Aleksandra Ola Bialas (The Sainsbury Laboratory) identified seven amino acids in the HMA domain of a Pikp-1 chimera that determine autoactivity.

As part of her research on the evolution of Pikp-1, Aleksandra Bialas replaced the HMA domain of Pikp-1 with alternative HMA domains. As was the case for Pikp-1^{HIPP19}, the chimeric Pik-1 proteins she generated were autoactive. By systematically replacing regions of the integrated HMA domain with the corresponding region of Pikp-HMA, she identified seven amino acids (Ala196-Cys202, inclusive) that are involved in autoactivation. Replacing these seven amino acids in her integrated HMA domains with the corresponding amino acids (AMEGNNC) from Pikp-1 prevented autoactivity.

In Pikp-1, these seven amino acids are located in a loop between β1 and α1 of the HMA domain, away from the effector-binding interface (figure 6.4a). In most of the crystal structures of HMA domain/AVR-Pik complexes solved to date, several of the residues in this loop are excluded from the final models as the electron density is not sufficiently well defined to position them satisfactorily. The lack of defined electron density suggests that this loop is disordered.

A level 0 module encoding the HMA domain of OsHIPP19 with residues Ser10 to Ser16 (inclusive) replaced with residues Ala196 to Cys202 from Pikp-1 (referred to as OsHIPP19-HMA^{mbl7}, for the seven amino acids from the metal-binding loop) was generated from a commercially synthesised DNA fragment (figure 6.4b). This level 0 module was used to assemble full-length Pikp-1$^{HIPP19\text{-}mbl7}$.

The *N. benthamiana* cell death assay was used investigate the activity of this modified chimera. Pikp-1$^{HIPP19\text{-}mbl7}$:HF was transiently co-expressed with Pikp-2:HA and either myc:AVR-PikD, myc:AVR-PikC, myc:AVR-PikF or empty vector. Pikp-1^{HIPP19}:HF was transiently co-expressed with or without Pikp-2:HA as a positive or negative control, respectively. No cell death was observed when Pikp-1$^{HIPP19\text{-}mbl7}$:HF was co-expressed with Pikp-2:HA and empty vector, demonstrating that exchanging these seven amino acids



B

```
          Pikp-1   188   KQKIVIKVAMEGNNCRSKAMALVASTGGVDSVALVGDL
       OsHIPP19      2   KQKIVIKVSMPCEKSRSKAMKLVVMASGVSSVEVTGDG
OsHIPP19-mbl7        2   KQKIVIKVAMEGNNCRSKAMKLVVMASGVSSVEVTGDG

          Pikp-1   226   RDKIEVVGYGIDPIKLISALRKKVGDAELLQVSQANKD
       OsHIPP19     40   KDRLQVVGDGVDAACLVTCLRKKIGHAELVQVEEVKEK
OsHIPP19-mbl7       40   KDRLQVVGDGVDAACLVTCLRKKIGHAELVQVEEVKEK
```

Figure 6.4 Seven amino acids in Pikp-HMA determine autoactivity.

A. *The crystal structure of Pikp-HMA in complex with AVR-PikD (De la Concepcion et al., 2018) (PDB accession code 6FU9). Pikp-HMA and AVR-PikD are represented as blue and green ribbons respectively, with the molecular surface of Pikp-HMA also displayed. The position of the seven amino acids in a loop between β1 and α1 of the HMA domain is highlighted in red.* ***B.*** *The amino acid sequences of the HMA domains of Pikp-1, OsHIPP19, and OsHIPP19^{mbl7}. The seven amino acids are indicated by the red bar.*

from OsHIPP19-HMA with the corresponding amino acids in Pikp-HMA was sufficient to prevent the autoactivity of Pikp-1^{HIPP19}. Strong HR-like cell death was observed when Pikp-1$^{HIPP19\text{-}mbl7}$ was co-expressed with Pikp-2 and AVR-PikD, demonstrating that this modified chimera retains the ability to trigger cell death in response to AVR-Pik effectors. Crucially, Pikp-1$^{HIPP19\text{-}mbl7}$ triggered a strong cell death response when co-expressed with Pikp-2 and either AVR-PikC (figure 6.5) or AVR-PikF (figure 6.6).

6.2.4 All AVR-Pik effector alleles associate with Pikp-1$^{HIPP19\text{-}mbl7}$ in co-immunoprecipitation experiments

Two of the three replicates of the co-immunoprecipitation experiment were conducted by Vera Sham under my supervision.

As the seven amino acids involved in autoactivity are located away from the AVR-Pik binding interface of the HMA domain, I hypothesised that they would not affect the interaction between the HMA domain and effector. Co-immunoprecipitation experiments were used to test for an association between each of the AVR-Pik effectors and the modified chimera Pikp-1$^{HIPP19\text{-}mbl7}$.

Each myc-tagged effector was transiently co-expressed with Pikp-1$^{HIPP19\text{-}mbl7}$:HF in *N. benthamiana.* As in other co-immunoprecipitation experiments, Pikp-2 was not included to avoid HR-like cell death following co-infiltration. As a positive control, Pikp-1:HF and myc:AVR-PikD were co-expressed. As negative controls, Pikp-1:HF was co-expressed with myc:AVR-PikC, and Pikp-1$^{HIPP19\text{-}mbl7}$ was co-expressed with empty vector. Pikp-1$^{HIPP19\text{-}mbl7}$:HF was immunoprecipitated by incubation of the plant extract with α-FLAG magnetic beads. Following multiple wash steps, the beads were boiled to elute bound proteins, and the presence of the proteins in the input and eluate was determined by Western blotting. Each of the myc-tagged AVR-Pik alleles co-immunoprecipitated with Pikp-1$^{HIPP19\text{-}mbl7}$:HF, as bands of the appropriate size are present in the eluate following immunoblotting with α-myc (figure 6.7), indicating that the modified chimera interacts with each of the AVR-Pik alleles.

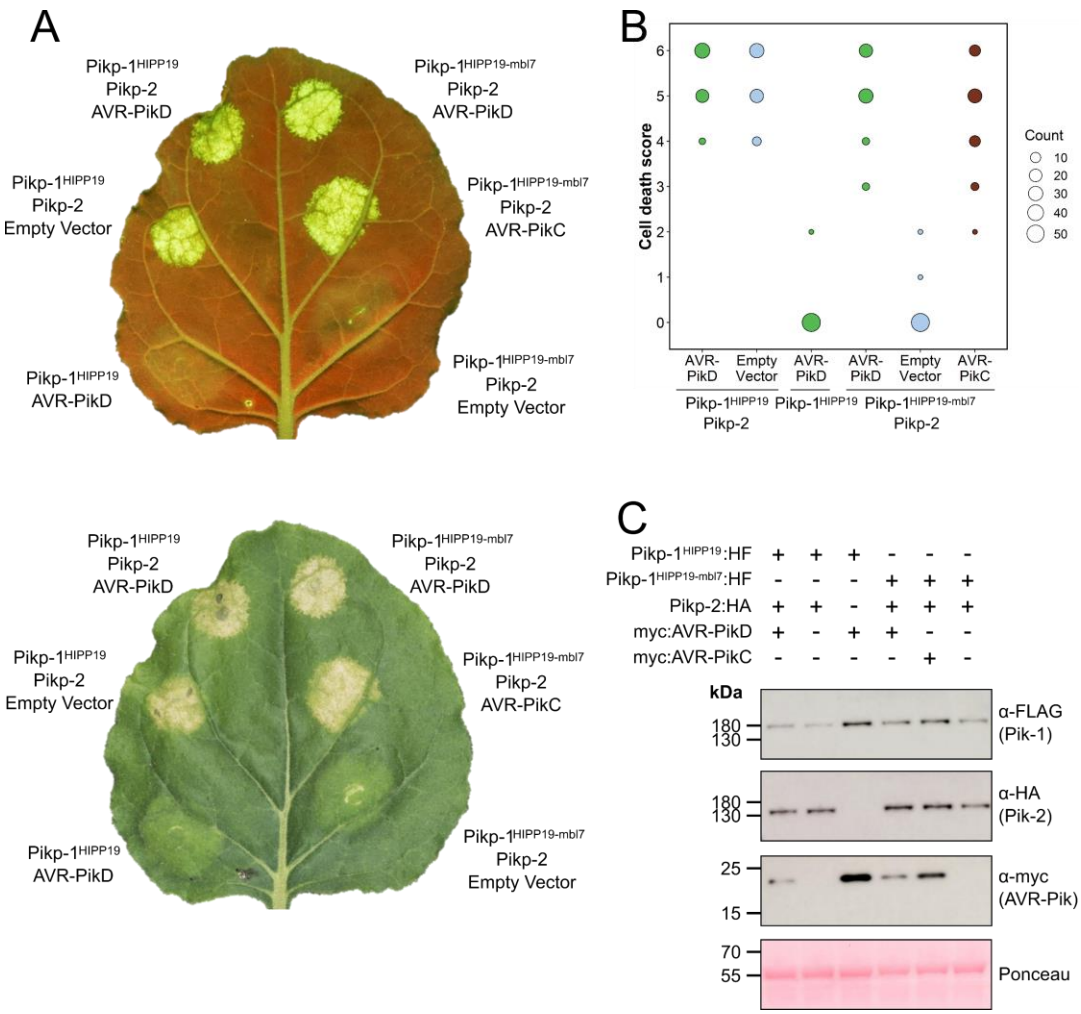


Figure 6.5 Pikp-1$^{HIPP19\text{-}mbl7}$ is not autoactive and triggers cell death in response to AVR-PikC.

A. *Representative leaf images showing cell death response under UV light (top) and daylight (bottom) showing the extent of cell death 5 dpi. The image under daylight has been horizontally flipped to provide the same leaf orientation for both images.* ***B.*** *Dot plot summarising the results obtained from 60 leaves in 3 independent experiments. Leaves were scored according to the scale in figure 2.2. The size of the dot at each cell death value is proportional to the number of leaves receiving that score. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).* ***C.*** *Western blots confirming production of the proteins in* N. benthamiana.

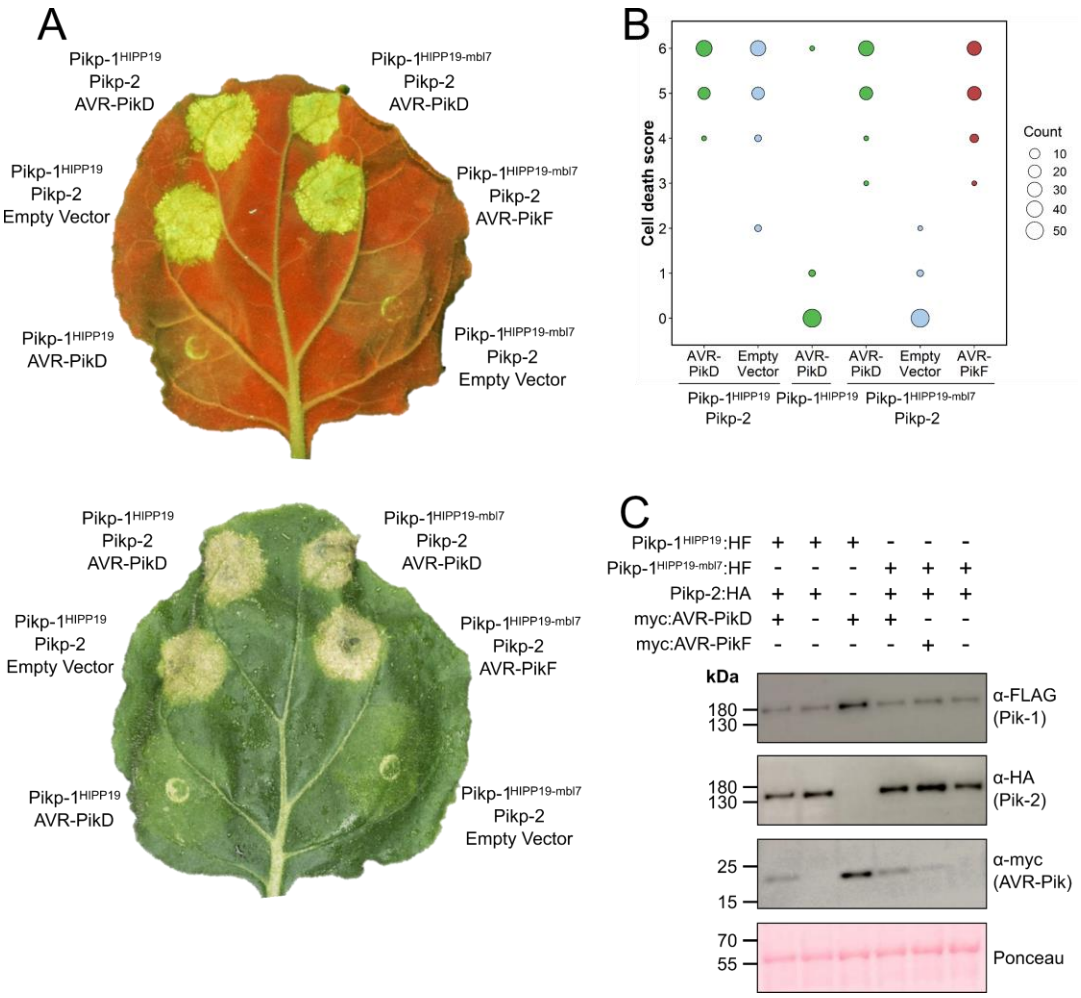


Figure 6.6 Pikp-1HIPP19-mbl7 is not autoactive and triggers cell death in response to AVR-PikF.

A. *Representative leaf images showing cell death response under UV light (top) and daylight (bottom) showing the extent of cell death 5 dpi. The image under daylight has been horizontally flipped to provide the same leaf orientation for both images.* ***B.*** *Dot plot summarising the results obtained from 58 leaves in 3 independent experiments. Leaves were scored according to the scale in figure 2.2. The size of the dot at each cell death value is proportional to the number of leaves receiving that score. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).* ***C.*** *Western blots confirming production of the proteins in* N. benthamiana*.*

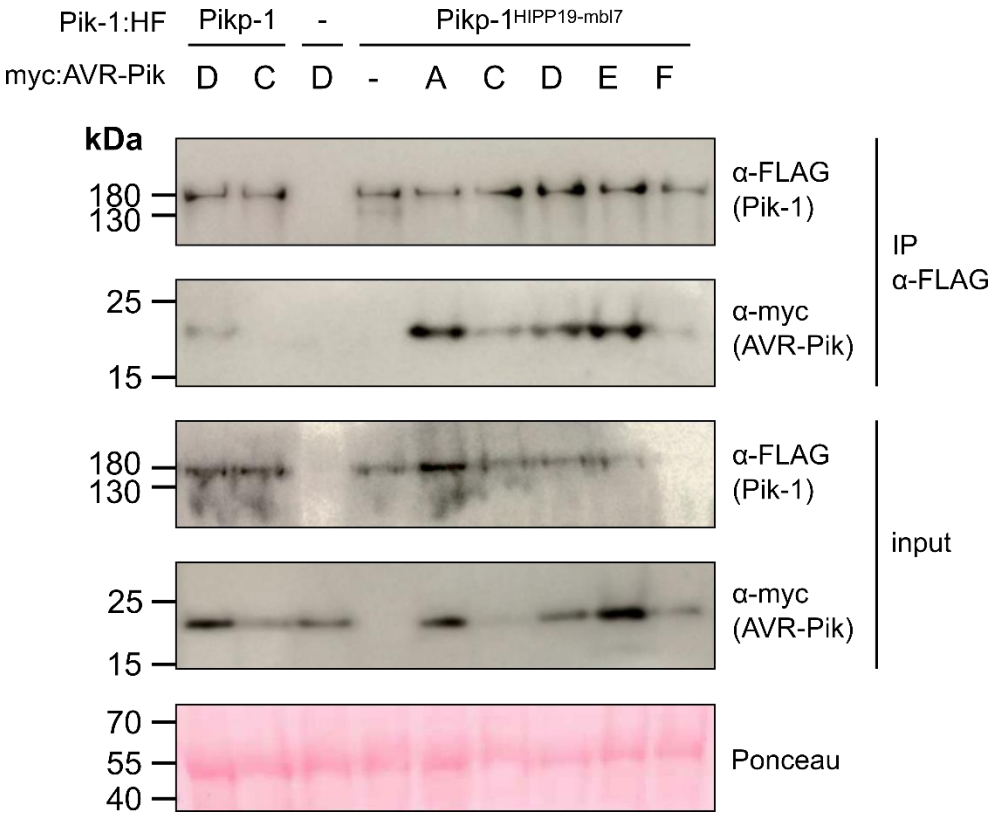


Figure 6.7 All AVR-Pik alleles co-immunoprecipitate with Pikp-1$^{HIPP19\text{-}mbI7}$.

Western blots showing proteins present in the eluate following α-FLAG immunoprecipitation (IP) and proteins present in the plant extract prior to immunoprecipitation (input). Western blots are representative of 3 independent experiments.

6.2.5 Pikp-1$^{HIPP19\text{-}mbl7}$ does not respond to AVR-Pia in cell death assays

The construct for transient expression of myc:AVR-Pia was provided by Dr Freya Varden.

The work described in Chapter 4 demonstrates that AVR-Pia interacts with the HMA domain of OsHIPP19, albeit with lower affinity than AVR-Pik. I hypothesised that, as the HMA domain of Pikp-1$^{HIPP19\text{-}mbl7}$ is almost identical to that of OsHIPP19, this chimeric NLR protein may interact with AVR-Pia and trigger cell death in *N. benthamiana*.

To investigate this, Pikp-1$^{HIPP19\text{-}mbl7}$:HF was transiently co-expressed with Pikp-2:HA and myc:AVR-Pia in *N. benthamiana*. As a positive control, Pikp-1$^{HIPP19\text{-}mbl7}$:HF was co-expressed with Pikp-2:HA and myc:AVR-PikD, and as a negative control, Pikp-1$^{HIPP19\text{-}mbl7}$:HF was co-expressed with Pikp-2:HA and empty vector. Previous work by Dr Freya Varden had found a very weak cell death response when Pikp-1:HF, Pikp-2:HA and myc:AVR-Pia were co-expressed (Varden et al., 2019); this combination was included for reference.

For one of the three replicates, the cell death observed for the positive control (Pikp-1$^{HIPP19\text{-}mbl7}$, Pikp-2, AVR-PikD) was lower than in, previous experiments and the other two replicates. However, even when only considering the two replicates where the positive control elicited strong cell death, no cell death was observed when AVR-Pia was co-expressed with Pikp-1$^{HIPP19\text{-}mbl7}$ and Pikp-2, nor when AVR-Pia was co-expressed with Pikp-1 and Pikp-2 (figure 6.8). Expression of all proteins, including AVR-Pia, was confirmed by Western blotting.

6.3 Discussion

Exchanging the HMA domain of Pikp-1 for the HMA domain of OsHIPP19 produced a chimeric NLR protein (Pikp-1^{HIPP19}) which activated a Pikp-2 dependent cell death response in the absence of the effector. In this chapter, I demonstrated that Pikp-1^{HIPP19} associates with all AVR-Pik alleles in co-immunoprecipitation experiments, including AVR-PikC and AVR-PikF. By exchanging seven amino acids in the HMA domain of Pikp-1^{HIPP19} for those in Pikp-1, I prevented autoactivation of the NLR protein, and the

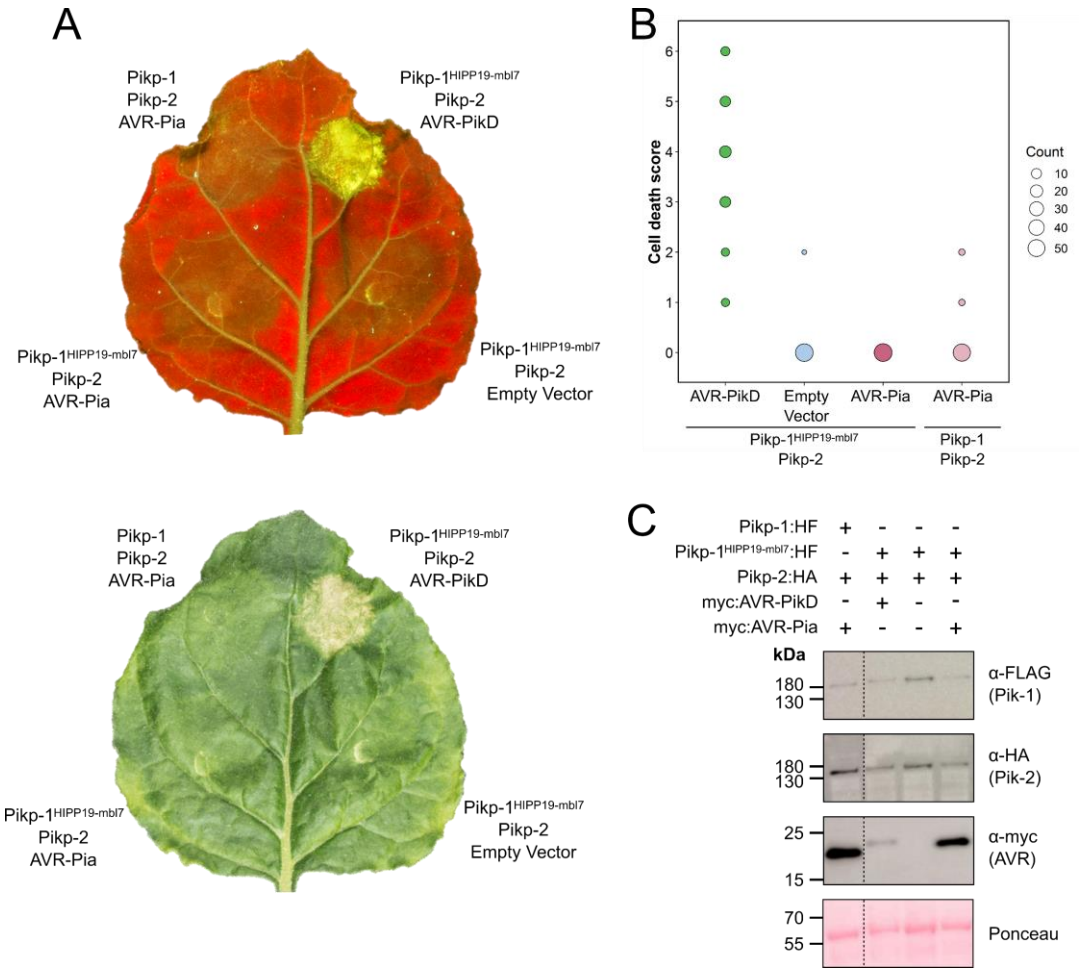


Figure 6.8 Pikp-1$^{HIPP19\text{-}mbl7}$ does not trigger cell death in response to AVR-Pia.

A. *Representative leaf images showing cell death response under UV light (top) and daylight (bottom) showing the extent of cell death 5 dpi. The image under daylight has been horizontally flipped to provide the same leaf orientation for both images.* ***B.*** *Dot plot summarising the results obtained from 59 leaves in 3 independent experiments. Leaves were scored according to the scale in figure 2.2. The size of the dot at each cell death value is proportional to the number of leaves receiving that score. Plots were produced using the ggplot2 package (Wickham, 2016) in R (R Core Development Team, 2018).* ***C.*** *Western blots confirming production of the proteins in* N. benthamiana. *Dashed line indicates that some lanes in the blot (containing samples not relevant to this experiment) have been omitted in this figure.*

resulting chimera triggered cell death in *Nicotiana benthamiana* in response to AVR-PikC or AVR-PikF.

In chapter 5, I showed that the autoactivity of Pikp-1^{HIPP19} was dependent on Pikp-2. In this chapter, I demonstrated that Pikp-1^{HIPP19} autoactivity requires an intact P-loop and MHD motif in Pikp-2. This is consistent with the work of Professor Hiromasa Saitoh and Rafal Zdrzalek, who demonstrated that the P-loop and MHD motif of Pikp-2 are required for the effector-dependent response of Pikp-1/Pikp-2 to AVR-PikD. Interestingly, mutating the conserved lysine in the P-loop of Pikp-1^{HIPP19} to arginine partially, but not completely, attenuated the cell death response. While the P-loop appears required for effector-dependent cell death, it is at least partially dispensable for effector-independent cell death caused by integration of a different HMA domain. While qualitative scoring of the *N. benthamiana* cell death assay indicates a slight reduction in the extent of cell death delivered when Pikp-1^{HIPP19_K296R} instead of Pikp-1^{HIPP19} is co-expressed with Pikp-2, use of a quantitative method such as ion leakage assays would enable more robust conclusions to be drawn on this topic. Additionally, assaying the ATPase activity of Pikp-1, Pikp-2, Pikp-1^{HIPP19} and the respective P-loop mutants would provide a more direct evaluation of the effect of the P-loop on activation of the Pik NLR proteins.

It is striking that just seven amino acids in the HMA domain prevent a Pik-1 chimera from being autoactive. Interestingly, these seven amino acids include the MXCXXC metal-binding motif of the HMA domain. This motif is degenerate in both Pikp-1 (MEGNNC) and OsHIPP19 (MPCEKS). Thus far, there is no evidence that either Pikp-HMA or OsHIPP19-HMA binds metal ions; no metal ions have been present in any of the solved crystal structures of Pik-1 HMA domains or OsHIPP19-HMA. The seven amino acids are located in the loop between β1 and α1 of the HMA domain. In the crystal structures of AVR-Pik/HMA domain complexes obtained to date, the electron density for this loop is typically poor compared to the density for the rest of the structure. This suggests that this loop is flexible relative to other regions of the protein. A tempting hypothesis is that this loop is involved in intramolecular interactions between domains in the Pik-1 NLR protein, keeping the protein in an inactive state in the absence of effector binding. Mutating these amino acids could prevent these intramolecular interactions, keeping Pik-1 in an active state and causing the autoactive phenotype.

Future work could investigate the interaction between the HMA domains of Pikp-1, OsHIPP19 and OsHIPP19^{mbl7} and the other domains of Pikp-1 and Pikp-2.

The principal outcome of this chapter is a second engineered Pik-1 variant which triggers HR-like cell death in response to the currently unrecognised effectors AVR-PikC and AVR-PikF in the model system *Nicotiana benthamiana*. The effector-binding interface of the HMA domain of Pikp-1^{HIPP19_mbl7} is virtually identical to that of OsHIPP19; it is unlikely that any mutation could arise in the effector that prevents it from interacting with Pikp-1^{HIPP19_mbl7} while retaining the ability to interact with OsHIPP19. However, AVR-PikD interacts with several HIPPs/HPPs in addition to OsHIPP19, and it is not inconceivable that the effector could mutate to lose interaction with OsHIPP19, but retain interaction with other HIPPs/HPPs, which could be sufficient for it to retain its virulence function.

The work presented in Chapter 4 demonstrates that AVR-Pia interacts weakly with the HMA domain of OsHIPP19 at a different binding interface to that targeted by AVR-Pik. However, while AVR-Pia interacts with OsHIPP19-HMA, the Pikp-1^{HIPP19_mbl7} chimera did not trigger cell death when co-expressed with Pikp-2 and AVR-Pia.

One explanation for this is that the binding affinity of AVR-Pia for OsHIPP19-HMA is not sufficiently high for the Pikp-1$^{HIPP19\text{-}mbl7}$ chimera to trigger a cell death response. Previous work with the wild-type Pikp-1 and Pikm-1 NLR proteins suggests a threshold binding affinity must be reached to activate immune signalling. For example, it was possible to observe a weak interaction between Pikp-HMA and AVR-PikE *in vitro*, co-purify and crystallise the complex and obtain the crystal structure (Maqbool et al., 2015, De la Concepcion et al., 2018). However, Pikp-1 does not trigger HR-like cell death in response to AVR-PikE in *Nicotiana benthamiana* cell death assays, and K60 (Pikp^{+}) rice plants are susceptible to *M. oryzae* carrying AVR-PikE (Maqbool et al., 2015, De la Concepcion et al., 2018). By contrast, Pikm-HMA binds AVR-PikE with higher affinity *in vitro*, and Pikm triggers HR-like cell death in *N. benthamiana* in response to AVR-PikE and confers resistance to AVR-PikE in a host pathosystem (De la Concepcion et al., 2018). A key next experiment would be to investigate whether AVR-Pia associates with Pikp-1$^{HIPP19\text{-}mbl7}$ in co-immunoprecipitation experiments; this could not be completed due to time constraints.

Another possible explanation for the lack of response of Pikp-1$^{HIPP19\text{-}mbl7}$ to AVR-Pia is that the HMA domain interface with which AVR-Pia interacts may not be compatible with complete activation of the Pik NLR proteins. However, K60 (Pikp^{+}) rice lines have shown moderate resistance when challenged with *M. oryzae* carrying AVR-Pia, suggesting that some degree of recognition/activation is possible.

As discussed in chapter 4, AVR-Pia may target other HIPPs or HPPs, and bind to these with higher affinity than to OsHIPP19. Integrating the HMA domain of one of these other HIPPs or HPPs into Pikp-1, with the appropriate mutations to prevent autoactivity, could produce a chimeric Pik-1 NLR protein capable of responding to AVR-Pia. However, the overall utility of such an NLR would be limited, as RGA5 and RGA4 already confer resistance to *M. oryzae* strains expressing AVR-Pia.

Of the limited number of *M. oryzae* effectors with known interactors, three (AVR-Pik, AVR-Pia and AVR1-CO39) have been shown to interact with HIPPs and/or HPPs. While the purpose of this interaction remains unknown, it seems reasonable to hypothesise that other effectors, both in *M. oryzae* and in other pathogen species, may also target HIPPs and HPPs. For example, the potato mop-top virus movement protein has been shown to interact with NbHIPP26 (Cowan et al., 2018). By identifying novel HIPP/HPP-interacting effectors and incorporating the HMA domain of the corresponding HIPP/HPP into Pik-1, it may be possible to engineer Pik-1 variants responding to diverse pathogens.

However, incorporating alternative HMA domains into Pikp-1 to recognise HIPP/HPP-interacting effectors may prove challenging. As illustrated by the autoactivity of Pikp-1^{HIPP19}, the integrated HMA domain has co-evolved with the other protein domains in Pik-1 and Pik-2 to deliver an appropriate effector-dependent immune response. An effector binding at a different interface, or with an alternative protein structure, may not activate the Pik NLR proteins. Furthermore, certain parts of the HMA domain may be occluded by other Pik-1 or Pik-2 domains, rendering them inaccessible for effector binding. The data presented in this chapter suggest that the loop between β1 and α1 containing the seven amino acids modified in the Pikp-1$^{HIPP19\text{-}mbl7}$ chimera is important for appropriate activation of Pikp-1. An effector interacting with the HMA domain at an

interface that blocks this loop, or interferes with other conformational changes that may occur during activation, could prevent immune signalling.

Chapters 5 and 6 present two complementary approaches for engineering NLR proteins that trigger HR-like cell death in a model system in response to currently unrecognised effectors. However, while results in model system *N. benthamiana* are promising, the true test of these engineered immune receptors is whether they confer resistance in the host pathosystem.

7

Towards transgenic rice and barley producing the engineered Pik-1 NLR proteins

7

Towards transgenic rice and barley producing the engineered Pik-1 NLR proteins

7.1 Introduction

The work described in chapters 5 and 6 demonstrates that Pikp-1$^{SNK\text{-}EKE}$ and Pikp-1$^{HIPP19\text{-}mbl7}$ can each trigger HR-like cell death when co-expressed with Pikp-2 and either AVR-PikC or AVR-PikF in the model system *Nicotiana benthamiana*. While the extent of cell death following transient expression of NLR protein/effector combinations in *N. benthamiana* has so far correlated with resistance of rice cultivars to *M. oryzae* isolates, it is important to test empirically whether these engineered NLR proteins can confer host resistance to *M. oryzae* isolates expressing AVR-PikC and AVR-PikF. This chapter outlines the progress that has been made towards generating stable transgenic rice and barley lines carrying either WT Pikp-1 (control), Pikp-1$^{SNK\text{-}EKE}$ or Pikp-1$^{HIPP19\text{-}mbl7}$, under the control of either the native rice Pik promoter, or the cauliflower mosaic virus 35S constitutive promoter.

Magnaporthe oryzae is best known for causing rice blast disease, however different strains of *M. oryzae* can cause disease on several other important grass species, including wheat, ryegrass, finger millet, foxtail millet and barley. Wheat blast is a significant emerging threat, as its recent spread into South East Asia from South America places large areas of wheat production at risk of the disease. Barley is a susceptible host for *M. oryzae* strains responsible for wheat blast and rice blast disease. Generating transgenic barley plants expressing the engineered NLR proteins would enable us to test the engineered variants against wheat blast isolates as well as rice blast isolates.

Following the identification of the two mutations (Asn261Lys and Lys262Glu) in Pikp-1 that extend the recognition profile of Pikp-1 to AVR-PikE and AVR-PikA, attempts were made to generate stable transgenic rice lines producing Pikp-1^{NK-KE}, with the aim of investigating whether the extended binding/recognition observed *in vitro* and in the model system *N. benthamiana* translated into blast resistance in a host pathosystem. Constructs were generated containing either WT Pikp-1:HF or Pikp-1^{NK-KE}:HF and Pikp-2:HA under the control of the native Pik promoter and used to transform rice (*Oryza sativa* cv. Nipponbare). Unfortunately, while the rice plants were successfully transformed with the constructs, and the incorporated genes were expressed (as confirmed by RT-PCR), no protein could be detected by Western blot and the transformed plants were susceptible to *M. oryzae* expressing AVR-PikD (the positive control).

One hypothesis to explain the lack of resistance of the transgenic plants to *M. oryzae* carrying AVR-PikD is that the HF or HA epitope tags interfered with appropriate folding and/or activity of the NLR proteins. Tagging proteins with an epitope enables the protein to be detected by Western blot, confirming that the protein has been produced in the system of interest. Pik-1 and Pik-2 have been successfully produced in *N. benthamiana* with C-terminal HF and HA epitope tags, respectively, and are functional, as indicated by effector-dependent HR-like cell death in the assays described in chapters 5 and 6. However, epitope tagging has been reported to affect protein folding, stability and activity. This is exemplified by work on the receptor-like kinase FLS2, where C-terminal fusions of different epitope tags resulted in highly variable immune signalling in *Arabidopsis thaliana* (Hurst et al., 2018). For the work described here, therefore, epitope tags were not included in the constructs.

The previous constructs for rice transformation only contained the coding sequence for the NLR proteins. Introns are widely reported to increase gene expression and protein production in plants. This phenomenon, known as intron-mediated enhancement, has been attributed to an increase in mRNA stability, mRNA production and nuclear export, and promotion of translation, and is reviewed in (Shaul, 2017). However, transforming rice with just the coding sequence of certain NLR proteins has led to successful protein production (Zhou et al., 2019). Due to the additional complexity of including the introns

in the Golden Gate level 2 constructs, the constructs generated in the work described here contain the coding sequence only.

The cauliflower mosaic virus (CaMV) 35S promoter is a widely used, strong and constitutive promoter. Over 60 % of approved commercial GM events contain the 35S promoter (Wu et al., 2014), and it has been successfully used to drive transgene expression in rice (Terada and Shimamoto, 1990). However, NLR proteins are typically maintained at low levels in the plant, and overexpression of NLR proteins can cause autoactivation, with negative consequences for the plant (Holt et al., 2002, Tao et al., 2000). For example, overexpression of the tomato NLR protein-encoding gene Pto under the control of the 35S promoter led to the formation of small necrotic lesions on the leaves (Tang et al., 1999). I therefore hypothesised that generating constructs with the expression of Pik-1 and Pik-2 being driven by the native promoter may result in appropriate regulation of the NLR proteins. Pik-1 and Pik-2 are found in a head-to-head orientation in the genome, which suggests co-regulation. However, use of the native promoter likely relies on appropriate control of expression by endogenous proteins. As both promoters have advantages and disadvantages, it was decided to generate two constructs for each Pik-1 variant, with the expression of the Pik NLR proteins driven by either the CaMV 35S promoter or the native Pik promoter.

The remainder of this chapter describes the cloning of constructs for rice and barley transformation, and the progress made so far towards stable transgenic rice and barley plants producing the engineered NLR proteins Pikp-1$^{SNK\text{-}EKE}$ and Pikp-1$^{HIPP19\text{-}mbl7}$. This work is ongoing, and future research will test whether stable transgenic plants are resistant to *M. oryzae* carrying AVR-PikC and AVR-PikF.

7.2 Results

7.2.1 Cloning Golden Gate level 2 constructs with the Pik genes under the 35S promoter

Golden Gate cloning (Engler et al., 2014, Engler et al., 2008, Weber et al., 2011) was used to assemble the multigene constructs used for rice and barley transformation. Chapter 2 of this thesis provides an overview of this cloning method. A schematic representation

of the assembly of the level 2 Golden Gate constructs with the Pik genes under the control of the 35S promoter is shown in figure 7.1.

Level 1 constructs for Pikp-1:HF and Pikp-2:HA, both driven by a 35S promoter and with a 35S terminator, were provided by Marina Franceschetti. Level 1 constructs for Pikp-1$^{SNK-EKE}$:HF and Pikp-1$^{HIPP19-mbl7}$:HF had been generated previously for *N. benthamiana* cell death assays. The 35S promoter sequence and Pik coding sequence from each of the four level 1 constructs were amplified by PCR to give DNA fragments flanked by appropriate overhangs and *BsaI* restriction sites. The PCR products were combined with a level 0 module containing a 35S terminator (TSL SynBio code pICSL51277) and a level 1 acceptor plasmid (position 2 for Pikp-1, Pikp-1$^{SNK-EKE}$ and Pikp-1$^{HIPP19-mbl7}$, position 3 for Pikp-2) in digestion-ligation (dig-lig) reactions. The resulting level 1 constructs contained untagged Pikp-1, Pikp-1$^{SNK-EKE}$, Pikp-1$^{HIPP19-mbl7}$ or Pikp-2 flanked by 35S promoter and terminator sequences.

A level 1 (position 1) construct containing a hygromycin resistance cassette (consisting of the hygromycin resistance gene *hph* driven by the 35S promoter and followed by the nopaline synthase (nos) terminator (from *Agrobacterium tumefaciens*)) was supplied by Marina Franceschetti. Hygromycin is an antibiotic produced by *Streptomyces hygroscopicus* which targets protein synthesis in both prokaryotes and eukaryotes. The resistance gene *hph* from *E. coli* encodes a hygromycin phosphotransferase which phosphorylates and detoxifies hygromycin (Gritz and Davies, 1983), and is widely used as selectable marker for plant transformation (Miki and McHugh, 2004, van den Elzen et al., 1985).

The level 1 construct containing the hygromycin resistance cassette was combined with the level 1 (position 3) construct containing 35S::Pikp-2 and one of the level 1 (position 2) constructs containing 35S::Pikp-1, 35S::Pikp-1$^{SNK-EKE}$ and 35S::Pikp1$^{HIPP19-mbl7}$. Digestion-ligation reactions assembled the level 1 constructs into a level 2 acceptor plasmid. The resulting level 2 constructs contain the hygromycin resistance cassette, 35S::Pikp-1/Pikp-1$^{SNK-EKE}$/Pikp-1$^{HIPP19-mbl7}$ and 35S::Pikp-2.

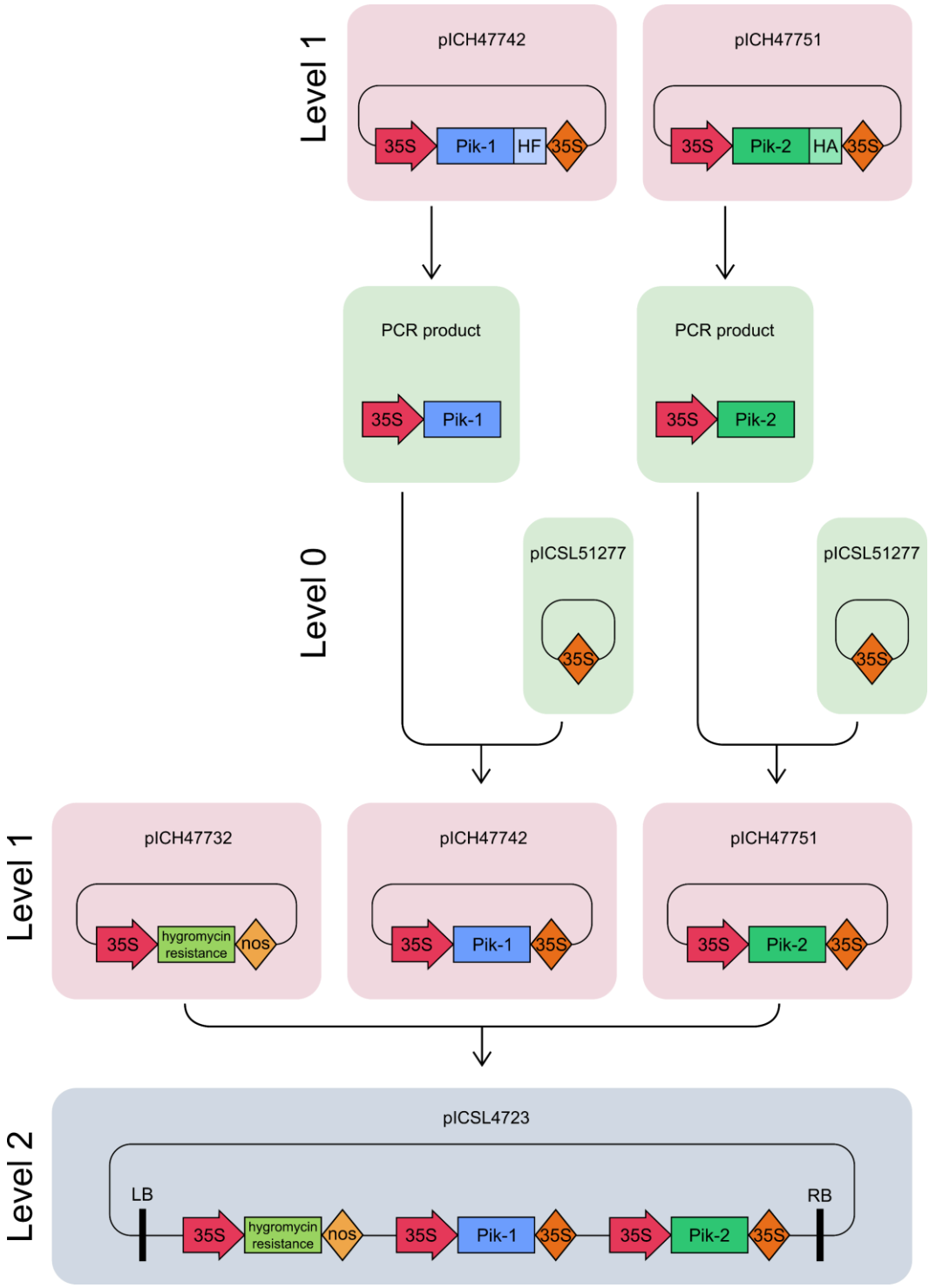


Figure 7.1 Schematic representation of cloning pipeline to generate level 2 constructs (35S promoter) for stable transformation of barley and rice.

Promoters are represented as arrows, coding sequences as rectangles, and terminators as diamonds. pIC codes refer to the Sainsbury Laboratory reference codes for standard parts or acceptor plasmids.

7.2.2 Cloning Golden Gate level 2 constructs with the Pik genes under the native promoter

A schematic representation of the assembly of the level 2 Golden Gate constructs with the Pik genes under the control of the native rice promoter is shown in figure 7.2. A level 0 module containing the rice native Pik promoter was supplied by Marina Franceschetti. The Pik-1/Pik-2 coding sequences and 35S terminator sequences were PCR amplified from level 1 35S::Pik-1 and 35S::Pik-2 constructs and cloned into the universal level 0 acceptor pUAP1 via a digestion-ligation reaction. A second digestion-ligation reaction was carried out to assemble a level 1 (position 2) construct containing Pik-1 and Pik-2 in a head-to-head orientation, with the native rice promoter sequence separating the two coding sequences. This level 1 (position 2) construct was combined with the level 1 (position 1) hygromycin resistance cassette described in section 7.2.1 and the level 2 acceptor plasmid in a digestion-ligation reaction to assemble a level 2 construct containing a hygromycin resistance cassette and Pikp-1/Pikp-1$^{SNK\text{-}EKE}$/Pikp-1$^{HIPP19\text{-}mbl7}$ and Pikp-2 in a head-to-head orientation under the control of the native rice promoter.

7.2.3 The presence of the transgene in transformed barley plants has been confirmed by PCR

Rice (Oryza sativa *cv. Nipponbare*) *transformation and all subsequent work with the transgenic rice plants was carried out by Motoki Shimizu and colleagues at Iwate Biotechnology Research Center. Barley* (Hordeum vulgare *cv. Golden Promise*) *transformation was carried out by the Sainsbury Laboratory plant transformation platform, led by Dr Matthew Smoker. Regenerated plantlets were supplied in jiffy pots.*

Following transfer to soil, leaf tissue was sampled from regenerated barley plants. Genomic DNA was isolated from the leaf tissue (see Materials and Methods). PCR was carried out with primers specific to the hygromycin resistance cassette to investigate the incorporation of the transgene into the barley genome. Genomic DNA from untransformed barley (*Hordeum vulgare* cv. Golden Promise) was used as a negative control, and the level 2 plasmid was used as a positive control. Nearly all barley plants gave a band of the expected size for the amplified portion of the hygromycin resistance

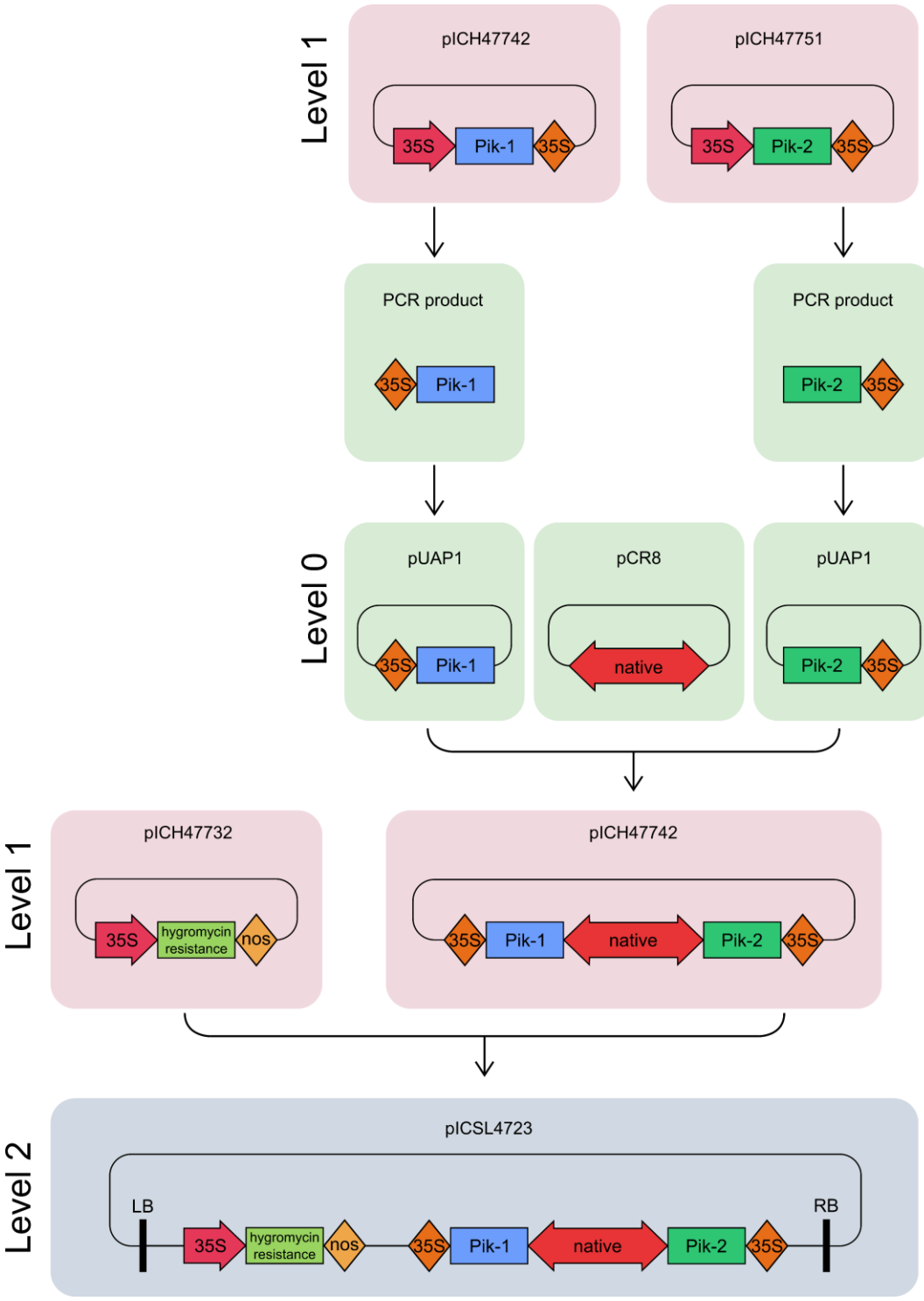


Figure 7.2 Schematic representation of cloning pipeline to generate level 2 constructs (native promoter) for stable transformation of barley and rice.

Promoters are represented as arrows, coding sequences as rectangles, and terminators as diamonds. pIC codes refer to the Sainsbury Laboratory reference codes for standard parts or acceptor plasmids.

cassette, indicating that the transgenes have been successfully introduced into the barley plants. An example of an agarose gel arising from the PCR is shown in figure 7.3.

7.2.4 Cloning Golden Gate level 2 constructs to transform K60$^{\Delta Pikp\text{-}1}$ rice plants

The rice cultivar Nipponbare contains a Pik-1-like gene which is non-functional due to significant indels in and around the HMA domain. The Nipponbare Pik haplotype is referred to as N-type, while the functional Pik-1 gene found in K60 rice is referred to as K-type. While the N-type Pik-1 cannot trigger resistance in response to AVR-PikD, and therefore any response to AVR-PikD can be attributed to the K-type Pik-1 transgene, the effect of the presence of the N-type Pik with regards to regulation of the expression/activity of the transgene is unknown. Our collaborators at Iwate Biotechnology Research Center (Japan) are generating K60 rice lines where Pikp-1 has been knocked out (K60$^{\Delta Pikp\text{-}1}$). These plants can then be complemented with the engineered Pik-1 variant.

Golden Gate cloning was used to generate level 2 constructs containing a selectable marker and either WT Pikp-1, Pikp-1$^{SNK\text{-}EKE}$ or Pikp-1$^{HIPP19\text{-}mbl7}$ under the control of the CaMV 35S promoter. The hygromycin resistance cassette had been used to select for plants where Pikp-1 had been successfully knocked out, therefore a bialaphos resistance cassette was incorporated into the constructs generated to transform the K60$^{\Delta Pikp\text{-}1}$ rice plants. Bialaphos is a protoxin produced by several *Streptomyces* species; processing in the plant releases glufosinate which inhibits glutamine synthetase. The product of the resistance gene *bar* from *S. hygroscopicus* acetylates glufosinate, preventing its inhibitory activity (Block et al., 1987).

A schematic representation of the assembly of the level 2 Golden Gate constructs for transformation of K60$^{\Delta Pikp\text{-}1}$ rice lines is shown in figure 7.4. A level 1 (position 1) construct containing a bialaphos resistance cassette (consisting of the bialaphos resistance gene *bar* driven by the nopaline synthase (nos) promoter and followed by the nos terminator) was combined with a level 1 (position 2) construct containing either WT Pikp-1, Pikp-1$^{SNK\text{-}EKE}$ or Pikp-1$^{HIPP19\text{-}mbl7}$ and the level 2 acceptor plasmid in a digestion-

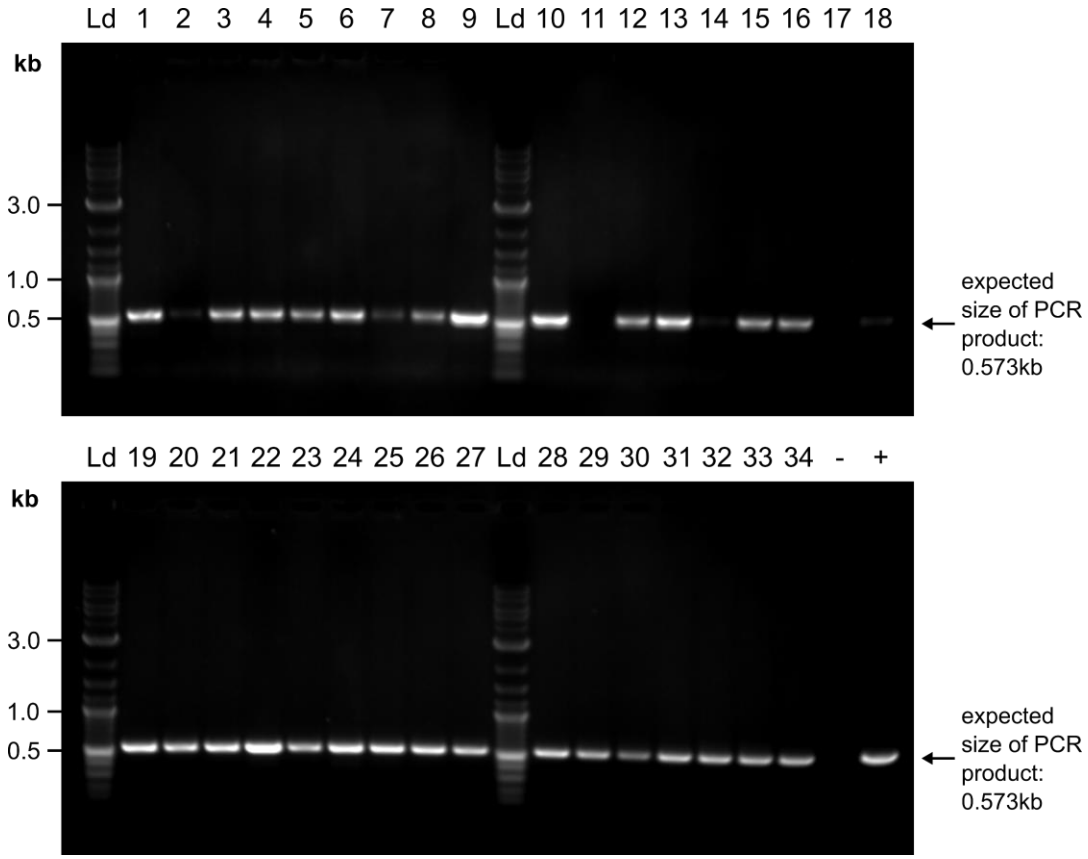


Figure 7.3 Agarose gel electrophoresis of products from PCR to confirm presence of hygromycin resistance cassette in transformed barley plants.

Samples 1-10 were from plants transformed with WT Pikp-1 under the 35S promoter. Samples 11-34 were from plants transformed with Pikp-1$^{SNK\text{-}EKE}$ under the 35S promoter. Genomic DNA from untransformed barley (Hordeum vulgare cv. Golden Promise) was used as a negative control (-), and the level 2 plasmid (WT Pikp-1 under the 35S promoter) was used as a positive control (+). Nearly all samples gave rise to a band of the expected size for the amplified fragment from the hygromycin resistance cassette.

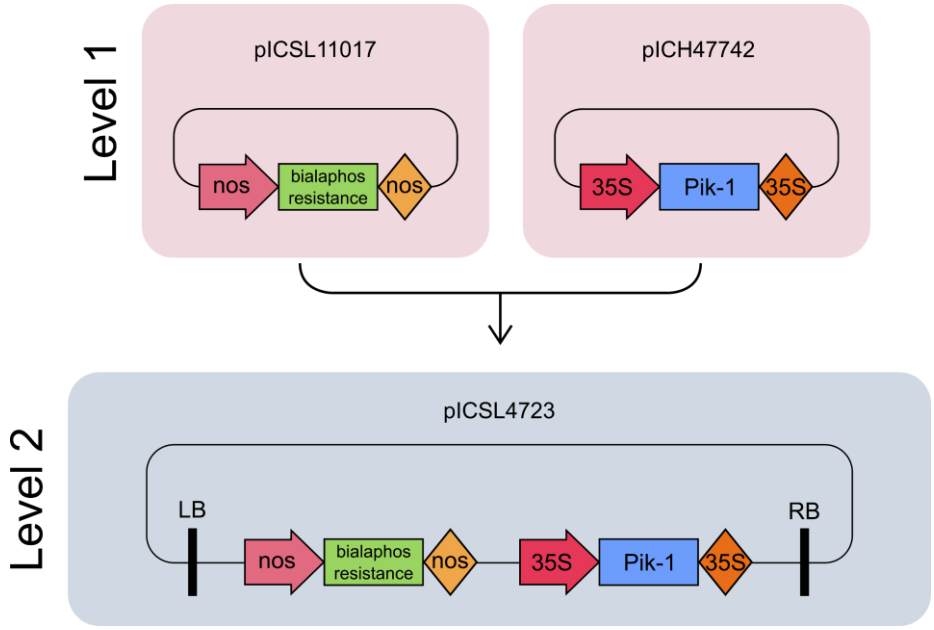


Figure 7.4 Schematic representation of cloning pipeline to generate level 2 constructs for stable transformation of rice (cv. K60 with Pikp-1 knocked out).

Promoters are represented as arrows, coding sequences as rectangles, and terminators as diamonds. pIC codes refer to the Sainsbury Laboratory reference codes for standard parts or acceptor plasmids.

ligation reaction. The resulting level 2 constructs contain the bialaphos resistance cassette, and Pikp-1/Pikp-1$^{SNK-EKE}$/Pikp-1$^{HIPP19-mbl7}$ driven by the CaMV 35S promoter. These constructs have been passed to our collaborators at Iwate Biotechnology Research Center for transformation of the K60$^{\Delta Pikp-1}$ rice plants.

7.3 Discussion

At the time of writing, the T_0 barley (*Hordeum vulgare* cv. Golden Promise) and rice (*Oryza sativa* cv. Nipponbare) plants are setting seed. Future work will characterise the resistance profiles of the T_1 and T_2 plants against *Magnaporthe oryzae* (Sasa2) transformed with AVR-PikD, AVR-PikC or AVR-PikF.

As a negative control, transgenic plants will be inoculated with untransformed *M. oryzae* (Sasa2); in the absence of an AVR-Pik effector, the transgenic plants should be susceptible to infection. As a positive control, plants transformed with Pikp-1 constructs will be inoculated with *M. oryzae* carrying AVR-PikD. It would be expected that these plants would be resistant to the fungus; if they are susceptible, this indicates an issue with either the expression of the transgene or the production, stability and/or function of the NLR proteins.

If the gene is not expressed, alternative promoters could be trialled. For example, the Maize Ubi and rice Act-1 are commonly used constitutive promoters for monocots. As introns can influence the accumulation and stability of transcripts and the rate at which they are exported from the nucleus, including the Pikp-1 and Pikp-2 introns in the construct for transformation may enhance gene expression. If the transgenes are expressed, but the plants are susceptible to *M. oryzae* carrying AVR-PikD, the protein may not be produced or accumulated sufficiently in the plant. As the proteins are not epitope-tagged, and no specific antibodies exist against the Pikp-1 or Pikp-2 proteins, it will not be possible to determine Pikp-1/Pikp-2 protein levels in the transgenic plants. As well as enhancing mRNA levels, introns can increase the rate of translation, influencing protein levels; including introns may aid protein accumulation.

As discussed in section 7.2.4, the Nipponbare rice background may interfere with successful production and function of Pikp-1 and Pikp-2. The K60$^{\Delta Pikp-1}$ rice plants

transformed with either WT Pikp-1 or one of the engineered Pik-1 variants avoids any potential effect of the N-type Pik haplotypes on the transgenes.

If plants transformed with either Pikp-1$^{SNK\text{-}EKE}$ or Pikp-1$^{HIPP19\text{-}mbl7}$ show resistance to AVR-PikC and AVR-PikF, I will have successfully engineered an NLR protein to recognise effectors that currently evade detection by rice immune receptors in the field. Identifying novel resistance genes effective against virulent pathogen populations is challenging, and often pathogen populations can evolve rapidly to overcome deployed resistance genes. Rational design of synthetic immune receptors extends the genetic "toolkit" available to breeders, providing novel resistance genes which can be used to reduce crop losses to plant pathogens.

8

General discussion

8

General discussion

Arable farmers across the world are engaged in a constant battle against plant pathogens. Increased globalisation has increased the movement of pathogens around the planet, while a changing climate has altered their host ranges. Crop losses due to disease jeopardise farmers' livelihoods and constrain food production.

In the ancient Chinese text "The Art of War", Sun Tzu writes that "if you know the enemy, and you know yourself, you need not fear the result of a hundred battles". In the context of plant-pathogen interactions, this quote highlights the importance of research into the fundamental processes underpinning pathogen infection and host resistance. The work described in this thesis contributes to our understanding of the molecular interactions between a staple food crop, rice, and the devastating rice blast fungus *Magnaporthe oryzae*. Importantly, it outlines how this fundamental knowledge can be used to engineer NLR immune receptors with novel recognition specificities.

Studying the interactions between effector proteins and their host targets offers insight into how pathogens manipulate their host to cause disease, and can also inform attempts to engineer crop resistance. In this thesis, I have shown that two *Magnaporthe oryzae* effector proteins interact with the HMA domain of OsHIPP19 at different interfaces. The existence of multiple *M. oryzae* effectors targeting HIPPs is intriguing and hints at a role for these proteins in plant immunity. At present, the mechanisms of action of AVR-Pik and AVR-Pia are unknown. These effectors may be redundant and perform a similar function despite interacting with OsHIPP19 at different interfaces. Redundancy among effectors would give greater scope for effector gene loss, for example to evade recognition by host immune receptors. This would be particularly important for effectors involved in core virulence functions.

Alternatively, AVR-Pik and AVR-Pia may have complementary (non-redundant) functions. Over 100 HIPPs and HPPs have been identified in rice (de Abreu-Neto et al., 2013)(Ryohei Terauchi, personal communication), and it is plausible that the two effectors target different subsets of these proteins. While HIPPs contain an isoprenylation motif at their C-terminus, which is postulated to target these proteins to endomembranes, HPPs lack this isoprenylation motif and may therefore display different localisation patterns. Different subcellular localisations of these effectors following secretion into the host cell may affect the subset of HIPPs and HPPs that they interact with during infection.

AVR-Pik and AVR-Pia may have antagonistic effects if they are produced at different stages of the infection process. *M. oryzae* is a hemibiotrophic fungus and can exist biotrophically for a prolonged period before switching to the necrotrophic stage of its life cycle. These two distinct developmental phases are likely to require different effectors with divergent functions; during biotrophic growth, suppression of necrosis is key for continued invasion, while during the necrotrophic stage, effectors may instead promote cell death. Interestingly, the effector AVRPiz-t is reported to contribute to both the biotrophic and necrotrophic stages of *Magnaporthe oryzae* infection by targeting different host proteins (Wang et al., 2016, Tang et al., 2017, Park et al., 2012, Park et al., 2016). Future work could use RNAseq analysis to probe the expression profiles of AVR-Pik and AVR-Pia during infection.

Several HIPPs have been described as susceptibility genes for pathogen infection (Radakovic et al., 2018, Cowan et al., 2018, Fukuoka et al., 2009). Disruption of susceptibility genes has been proposed as a strategy to engineer resistance (Pavan et al., 2010), however knocking out susceptibility genes can induce pleiotropic phenotypes in the host plant. An alternative approach would involve making targeted mutations in susceptibility factors that prevent their interaction with an effector protein. The crystal structures of AVR-PikF/OsHIPP19-HMA and AVR-Pia/OsHIPP19-HMA could inform mutagenesis of OsHIPP19-HMA to prevent the protein from interacting with these effectors proteins without compromising the function of OsHIPP19 in uninfected plants.

The discovery of diverse non-canonical protein domains integrated into the typical NLR protein architecture has offered insights into how effectors contribute to disease and how

NLR proteins activate effector-triggered immunity. Studies of effector binding to integrated domains have provided insights into the molecular mechanisms underpinning the recognition of specific effectors by their cognate NLR protein(s) (Maqbool et al., 2015, De la Concepcion et al., 2018, Guo et al., 2018, Sarris et al., 2015). Integrated domains can also be used to identify host proteins targeted by pathogen effectors. Following the identification of multiple NLR proteins containing an integrated BED domain, Kroj et al. identified a BED domain containing protein (ZBED) which they subsequently demonstrated had a role in plant immunity (Kroj et al., 2016). While the effector recognised by the BED domain-containing NLR proteins was not identified in this case, this research highlighted the potential of integrated domains for identifying novel proteins involved in plant immunity.

The evolution of NLR proteins containing integrated domains is an area of continued research. Several studies have investigated the distribution of NLR proteins with integrated domains across plant species (Sarris et al., 2016, Kroj et al., 2016, Bailey et al., 2018). While the molecular mechanisms underlying these integration events are undetermined, an amino acid motif immediately upstream of integrated domains in multiple NLR proteins has been hypothesised to be important for the integration process (Bailey et al., 2018).

The work described in this thesis contributes to our understanding of integrated domains in NLR proteins. The observation that AVR-Pik binds to OsHIPP19-HMA at a globally similar interface to Pikp-HMA or Pikm-HMA provides additional support for the hypothesis that integrated domains mediate effector recognition by resembling host targets. Likewise, the interface between AVR-Pia and OsHIPP19-HMA closely resembles that formed between AVR-Pia and Pikp-HMA (Varden et al., 2019) and AVR1-CO39/RGA5-HMA (Guo et al., 2018).

Intriguingly, surface plasmon resonance experiments showed that AVR-Pik interacts more strongly with OsHIPP19-HMA than with the integrated Pik-1 HMA domains in surface plasmon resonance experiments. This finding is supported by structural information, as the complex between OsHIPP19-HMA and AVR-PikF involves a larger total interface area and more intermolecular hydrogen bonds and salt bridges than any of the complexes between integrated Pik-HMA domains and AVR-Pik effectors solved

to date. This result contrasts with the widely-held perspective that the integrated domains of NLR proteins will bind effector proteins with higher affinity than their virulence targets.

The results presented in this thesis demonstrate that all AVR-Pik alleles interact with the HMA domain of OsHIPP19, while only a subset bind to the integrated HMA domain of Pikp-1 or Pikm-1. AVR-PikC and AVR-PikF are not recognised by any Pik alleles characterised to date. The discovery that both stealthy effector alleles could interact with the HMA domain of OsHIPP19 provided a foundation for engineering Pik-1 variants to bind and respond to these stealthy effector alleles.

During this project, I used two complementary approaches, both informed by the effector target OsHIPP19, to engineer NLR proteins with novel recognition specificities. Both Pikp-1$^{SNK\text{-}EKE}$ and Pikp-1^{HIPP19_mbl7} trigger HR-like cell death in response to AVR-PikC and AVR-PikF when transiently expressed in the model plant *Nicotiana benthamiana*. At present, it is not known whether these engineered NLR proteins confer resistance in rice to *Magnaporthe oryzae* strains expressing AVR-PikC or AVR-PikF, however work to generate stable transgenic rice and barley expressing these engineered immune receptors is ongoing.

The deployment of individual resistance genes creates a strong selection pressure for virulent pathogen genotypes. Pyramiding or stacking multiple resistance genes in a crop variety, or planting crop varieties each carrying different resistance genes, should extend the longevity of each resistance gene. It should also be considered how simple it would be for a pathogen to evade recognition by the engineered NLR protein. A point mutation in the effector at interface 3, such as the conversion of one of the residues forming the pocket of negative surface charge to an amino acid with a positively charged side chain, may be sufficient to disrupt the interaction of the effector with Pikp-1$^{SNK\text{-}EKE}$. While the comparison between the binding interfaces of OsHIPP19-HMA/AVR-PikF and Pikm-HMA/AVR-PikA highlighted that additional intermolecular interactions at interface 3 are likely to contribute towards the higher affinity and broader specificity of OsHIPP19-HMA for the AVR-Pik effectors, differences at other interfaces could also play a role. It may therefore be more challenging for the effector to mutate to avoid recognition by Pikp-1^{HIPP19_mbl7} than Pikp-1$^{SNK\text{-}EKE}$.

An additional consideration is that mutations in the effector to avoid binding to Pikp-1$^{SNK\text{-}EKE}$ or Pikp-1^{HIPP19_mbl7} could also prevent interaction with the putative virulence target OsHIPP19. This is particularly likely for mutations that disrupt the interaction with Pikp-1^{HIPP19_mbl7}, as the effector binding interfaces of Pikp-1^{HIPP19_mbl7} and OsHIPP19 are virtually identical. The virulence activity of AVR-Pik is undetermined; while *M. oryzae* isolates lacking AVR-Pik effector alleles have been identified, the allelic series of AVR-Pik indicates selection for mutation rather than loss. Disruption of the interaction between AVR-Pik and OsHIPP19-HMA could therefore reduce the virulence of the pathogen. Future work could investigate mutations in AVR-Pik that disrupt the interaction with Pikp-1$^{SNK\text{-}EKE}$ and Pikp-1^{HIPP19_mbl7} and test the effect of these mutations on the interaction between the effector and OsHIPP19. This would enable conclusions to be drawn about the durability of these engineered NLR proteins.

HIPPs and HPPs have been reported to be susceptibility factors for infection by other pathogens (Radakovic et al., 2018, Cowan et al., 2018, Fukuoka et al., 2009). Incorporating HMA domains from other HIPPs/HPPs into the Pikp-1 backbone (retaining the seven amino acids from Pikp-1 in the β1-α1 loop to prevent autoactivity) could generate Pik-1 variants which respond to HIPP/HPP-targeting effectors from different pathogens. However, the position of the HMA domain in Pik-1 may limit the success of this approach, as the adjacent domains may occlude possible effector binding sites. This strategy may be more successful if novel HMA domains were incorporated at the C-terminus of the protein, perhaps using RGA5 as a scaffold.

In the long term, it would be highly desirable to be able to generate custom synthetic NLR proteins, incorporating a variety of integrated domains, to confer resistance to multiple pathogens. Pikp-HMA and OsHIPP19-HMA share 51 % amino acid sequence identity and are structurally very similar. Despite this, the Pikp-1^{HIPP19} chimera was autoactive, exemplifying the challenges in engineering novel NLR proteins by incorporating novel integrated domains and illustrating how integrated domains have co-evolved with other domains in the NLR receptor to deliver an appropriate, effector-dependent response. The cause of this autoactivity is uncertain; developing an understanding of how binding of an effector to an integrated domain translates into

activation of downstream signalling processes is likely to help dissect this autoactive phenotype.

The structures of the NLR protein ZAR1 pre- and post-activation have revealed key insights into the conformational changes that occur during the activation of this NLR protein and demonstrated that activation involves oligomerisation of ZAR1 to form a pentameric resistosome (Wang et al., 2019b, Wang et al., 2019a). ZAR1 functions as a singleton NLR protein; whether paired NLR proteins form similar oligomeric structures is yet to be determined. Paired NLR proteins can function through negative regulation of the constitutively active helper by the sensor, or through cooperative activation of effector-triggered immunity. These different modes of action may affect the composition and stoichiometry of NLR protein oligomers. ZAR1 lacks an integrated domain, although it could be argued that the associated pseudokinase RKS1 functions in a comparable manner. The effect of integrated domains, particularly towards the N-terminus, on the activation of CC-NLR proteins is an area for future study. Based on the important conformational changes at the N-terminus of ZAR1 during activation, integrating novel protein domains at the C-terminus of the NLR protein may be more likely to result in appropriate activation of immune signalling. The mechanistic link between pentamerisation of ZAR1 and the onset of HR-like cell death remains to be determined and little is known about the signalling components downstream of CC-NLR proteins. Understanding how activation of an NLR protein is connected to observed HR-like cell death and resistance phenotypes could inform future efforts to rationally engineer NLR proteins.

If it is possible to engineer NLR immune receptors with novel recognition specificities, there will be significant obstacles associated with their commercialisation and deployment in the field. As synthetic immune receptors, by definition, are not present in nature, they cannot be incorporated into elite crop cultivars by conventional breeding techniques and must be introduced via genetic modification. Current genome editing technologies use engineered site-specific nucleases to modify specific genes at determined locations in the genome. These nucleases include the transcription activator-like effector nucleases (TALENs), zinc finger nucleases (ZFNs) and the CRISPR-associated endonuclease Cas9. Each of these nucleases induces a double-stranded break

in the target DNA sequence which is then repaired by either non-homologous end joining (NHEJ) or homologous recombination. Genome editing techniques have been successfully used in rice (reviewed by (Mishra et al., 2018)).

The engineered NLR protein Pikp-1$^{SNK\text{-}EKE}$ only differs from WT Pikp-1 in three amino acid positions would involve a maximum of five nucleotide substitutions. Base editing is an emerging technique based upon the CRISPR/Cas9 system which enables substitution of one nucleotide for another without introducing a double strand DNA break (Kang et al., 2018). Pikp is present in a number of elite rice cultivars; base editing could be used to introduce the Ser258Glu, Asn261Lys and Lys262Glu mutations directly into commercial rice varieties. As there are many more nucleotide differences between Pikp-1 and Pikp-1^{HIPP19_mbl7}, targeted gene insertion would be required to introduce this engineered NLR protein into rice.

At the time of writing, the European Union regulations require crop varieties developed using genome editing techniques to be subject to the same regulatory frameworks as GMOs containing transgenes from other species. By contrast, the United States of America's Coordinated Framework for Regulation of Biotechnology excludes edited crops which do not contain recombinant DNA, lack pesticidal activity, and do not provoke any food safety issues distinct from traditionally bred crops, from regulation by the United States Department of Agriculture. Unfortunately, many African nations are following the legislative position of the European Union on crop varieties developed by genome editing. The EU is a major export market for large parts of Africa, and even though there is no legislative barrier to the import of edited crops, there are serious concerns that public mistrust in genetic modification will lower the value of exported food products.

The position of the European Union on edited crops is frustrating. Mutations, indels, gene and chromosome rearrangements and even transfer of genetic material between organisms all occur in nature, meaning that genome edited crops can be virtually indistinguishable from naturally occurring crop variants. This also poses a conundrum for regulatory bodies; if you cannot differentiate between edited and "natural" plants, these regulations cannot be enforced. While genome editing can lead to off-target effects, these can be screened for and eliminated by self- or backcrossing; crop varieties

generated by classical random mutagenesis methods are more likely to contain multiple undesirable mutations but are not considered GMOs under current regulations. There is a disconnect between scientific research and legislation which needs to be addressed by effective communication between scientists, policy makers and the general public. Moreover, I would argue that there is a broader problem with how science is viewed in society. Science is typically seen as the exclusive domain of professional scientists, rather than a public good belonging to the wider community. By promoting a sense of ownership of scientific research amongst the general public and encouraging participation in discussions on science-related issues, we ultimately enable technological advances, such as genome editing, to be exploited to their full potential.

In conclusion, the work described in this thesis contributes to our understanding of the molecular basis of virulence and recognition in rice blast disease. I present two complementary approaches to engineer NLR proteins with novel recognition specificities, both informed by the structural and biochemical characterisation of the interaction between an effector and its host target. Rationally engineered NLR proteins offer new opportunities to control virulent pathogens; if legislative obstacles and perceptual barriers can be overcome, engineered NLR proteins could contribute to global food security in the future.

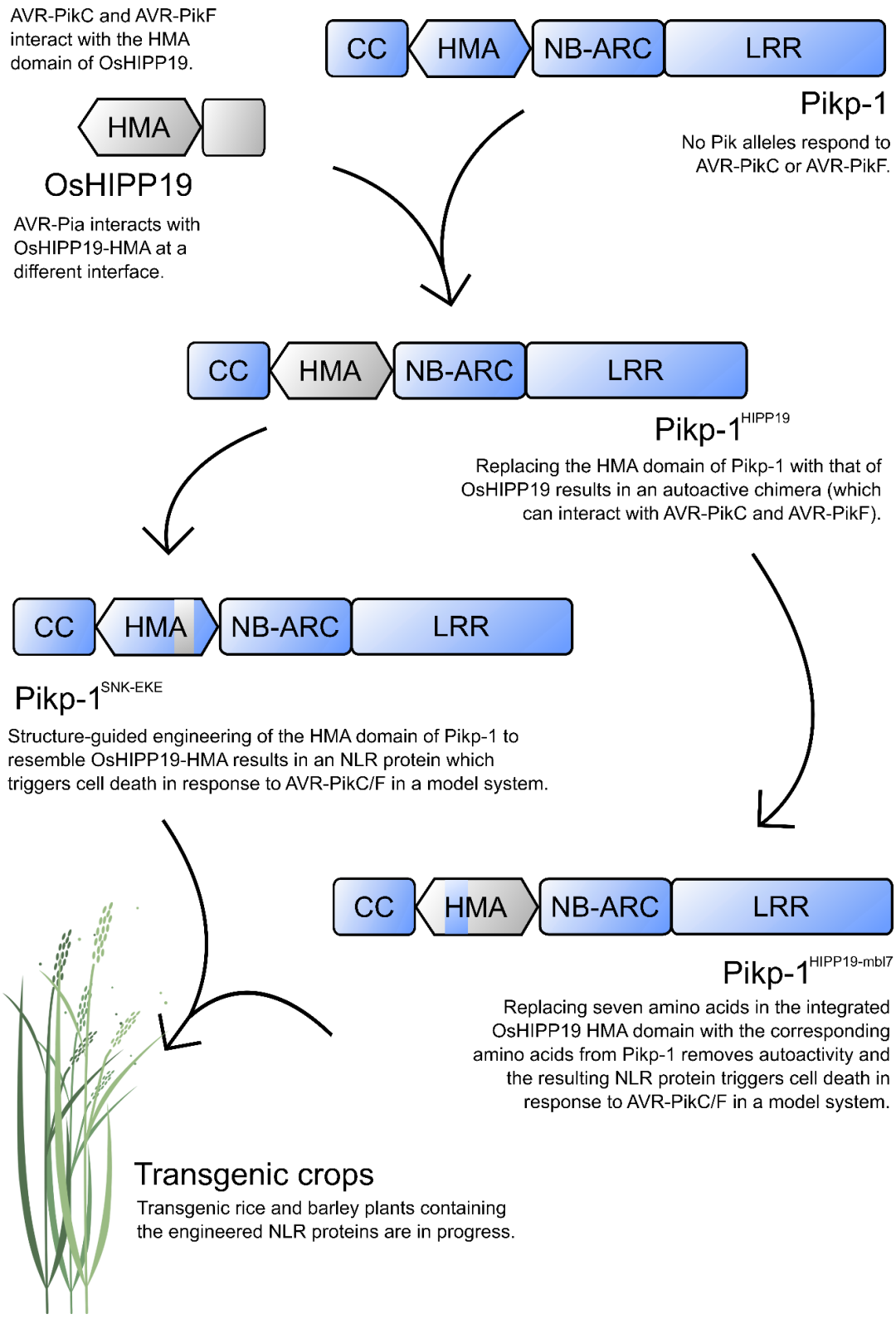


Figure 8.1 Graphical summary of the work presented in this thesis.

References

References

AARTS, N., METZ, M., HOLUB, E., STASKAWICZ, B. J., DANIELS, M. J. & PARKER, J. E. 1998. Different requirements for EDS1 and NDR1 by disease resistance genes define at least two R gene-mediated signaling pathways in Arabidopsis. *Proceedings of the National Academy of Sciences, USA,* 95**,** 10306-11.

ADE, J., DEYOUNG, B. J., GOLSTEIN, C. & INNES, R. W. 2007. Indirect activation of a plant nucleotide binding site-leucine-rich repeat protein by a bacterial protease. *Proceedings of the National Academy of Sciences, USA,* 104**,** 2531-2536.

ARNESANO, F., BANCI, L., BERTINI, I., HUFFMAN, D. L. & O'HALLORAN, T. V. 2001. Solution structure of the Cu(I) and apo forms of the yeast metallochaperone, Atx1. *Biochemistry (Washington),* 40**,** 1528-1539.

ARORA, S., STEUERNAGEL, B., GAURAV, K., CHANDRAMOHAN, S., LONG, Y., MATNY, O., JOHNSON, R., ENK, J., PERIYANNAN, S., SINGH, N., ASYRAF MD HATTA, M., ATHIYANNAN, N., CHEEMA, J., YU, G., KANGARA, N., GHOSH, S., SZABO, L. J., POLAND, J., BARIANA, H., JONES, J. D. G., BENTLEY, A. R., AYLIFFE, M., OLSON, E., XU, S. S., STEFFENSON, B. J., LAGUDAH, E. & WULFF, B. B. H. 2019. Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nature Biotechnology,* 37**,** 139-143.

ASHIKAWA, I., HAYASHI, N., ABE, F., WU, J. & MATSUMOTO, T. 2012. Characterization of the rice blast resistance gene Pik cloned from Kanto51. *Molecular Breeding,* 30**,** 485-494.

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**,** 2267-2276.

BAILEY, B. A., DEAN, J. F. & ANDERSON, J. D. 1990. An ethylene biosynthesis-inducing endoxylanase elicits electrolyte leakage and necrosis in *Nicotiana tabacum* cv Xanthi leaves. *Plant Physiology,* 94**,** 1849-1854.

BAILEY, P. C., SCHUDOMA, C., JACKSON, W., BAGGS, E., DAGDAS, G., HAERTY, W., MOSCOU, M. & KRASILEVA, K. V. 2018. Dominant integration locus drives continuous diversification of plant immune receptors with exogenous domain fusions. *Genome Biology,* 19**,** 23-23.

BALLINI, E., NGUYEN, T. T. & MOREL, J. B. 2013. Diversity and genetics of nitrogen-induced susceptibility to the blast fungus in rice and wheat. *Rice,* 6**,** 32.

BALTRUS, D. A., NISHIMURA, M. T., ROMANCHUK, A., CHANG, J. H., MUKHTAR, M. S., CHERKIS, K., ROACH, J., GRANT, S. R., JONES, C. D. & DANGL, J. L. 2011. Dynamic evolution of pathogenicity revealed by sequencing and comparative genomics of 19 *Pseudomonas syringae* isolates. *PLOS Pathogens,* 7**,** e1002132.

BARTH, O., VOGT, S., UHLEMANN, R., ZSCHIESCHE, W. & HUMBECK, K. 2009. Stress induced and nuclear localized HIPP26 from *Arabidopsis thaliana* interacts via its heavy metal associated domain with the drought stress related zinc finger transcription factor ATHB29. *Plant Molecular Biology,* 69**,** 213-226.

BASHANDY, H., JALKANEN, S. & TEERI, T. H. 2015. Within leaf variation is the largest source of variation in agroinfiltration of *Nicotiana benthamiana*. *Plant Methods,* 11**,** 47.

BENDAHMANE, A., FARNHAM, G., MOFFETT, P. & BAULCOMBE, D. C. 2002. Constitutive gain-of-function mutants in a nucleotide binding site–leucine rich repeat protein encoded at the Rx locus of potato. *The Plant Journal,* 32**,** 195-204.

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**,** e45.

BIELSKA, E., HIGUCHI, Y., SCHUSTER, M., STEINBERG, N., KILARU, S., TALBOT, N. J. & STEINBERG, G. 2014. Long-distance endosome trafficking drives fungal effector production during plant infection. *Nature Communications,* 5**,** 5097.

BLOCK, M. D., BOTTERMAN, J., VANDEWIELE, M., DOCKX, J., THOEN, C., GOSSELE, V., MOVVA, N. R., THOMPSON, C., MONTAGU, M. V. & LEEMANS, J. 1987. Engineering herbicide resistance in plants by expression of a detoxifying enzyme. *The EMBO Journal,* 6**,** 2513-2518.

BOHNERT, H. U., FUDAL, I., DIOH, W., THARREAU, D., NOTTEGHEM, J. L. & LEBRUN, M. H. 2004. A putative polyketide synthase/peptide synthetase from *Magnaporthe grisea* signals pathogen attack to resistant rice. *The Plant Cell,* 16**,** 2499-2513.

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.

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**,** 35834-35842.

BULL, P. C. & COX, D. W. 1994. Wilson disease and Menkes disease: new handles on heavy-metal transport. *Trends in Genetics,* 10**,** 246-252.

BULL, P. C., THOMAS, G. R., ROMMENS, J. M., FORBES, J. R. & COX, D. W. 1993. The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. *Nature Genetics,* 5**,** 327-337.

BÜTTNER, D. 2016. Behind the lines-actions of bacterial type III effector proteins in plant cells. *FEMS Microbiology Reviews,* 40**,** 894-937.

CAPELL, B. C., ERDOS, M. R., MADIGAN, J. P., FIORDALISI, J. J., VARGA, R., CONNEELY, K. N., GORDON, L. B., DER, C. J., COX, A. D. & COLLINS, F. S. 2005. Inhibiting farnesylation of progerin prevents the characteristic nuclear blebbing of Hutchinson-Gilford progeria syndrome. *Proceedings of the National Academy of Sciences, USA,* 102**,** 12879-12884.

CARTER, M. E., HELM, M., CHAPMAN, A. V. E., WAN, E., RESTREPO SIERRA, A. M., INNES, R. W., BOGDANOVE, A. J. & WISE, R. P. 2019. Convergent evolution of effector protease recognition by Arabidopsis and barley. *Molecular Plant-Microbe Interactions,* 32**,** 550-565.

CASTEL, B., NGOU, P. M., CEVIK, V., REDKAR, A., KIM, D. S., YANG, Y., DING, P. & JONES, J. D. G. 2019. Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *New Phytologist,* 222**,** 966-980.

CENTURY, K. S., HOLUB, E. B. & STASKAWICZ, B. J. 1995. NDR1, a locus of *Arabidopsis thaliana* that is required for disease resistance to both a bacterial and a fungal pathogen. *Proceedings of the National Academy of Sciences, USA,* 92**,** 6597-6601.

CESARI, S., BERNOUX, M., MONCUQUET, P., KROJ, T. & DODDS, P. N. 2014a. A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. *Frontiers in Plant Sciences,* 5**,** 606.

CESARI, S., KANZAKI, H., FUJIWARA, T., BERNOUX, M., CHALVON, V., KAWANO, Y., SHIMAMOTO, K., DODDS, P., TERAUCHI, R. & KROJ, T. 2014b. The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *The EMBO Journal,* 33**,** 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 recognises the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. *The Plant Cell,* 25**,** 1463-1481.

CHEN, J., PENG, P., TIAN, J., HE, Y., ZHANG, L., LIU, Z., YIN, D. & ZHANG, Z. 2015. Pike, a rice blast resistance allele consisting of two adjacent NBS–LRR genes, was identified as a novel allele at the Pik Locus. *Molecular Breeding,* 35.

CHEN, V. B., ARENDALL, W. B., III, 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 Crystallographica Section D: Biological Crystallography,* 66**,** 12-21.

CHINCHILLA, D., ZIPFEL, C., ROBATZEK, S., KEMMERLING, B., NUERNBERGER, T., JONES, J. D. G., FELIX, G. & BOLLER, T. 2007. A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature,* 448**,** 497-500.

CHUMLEY, F. G. & VALENT, B. 1990. Genetic analysis of melanin-deficient, nonpathogenic mutants of *Magnaporthe grisea*. *Molecular Plant-Microbe Interactions,* 3**,** 135-143.

CLARKE, S. 1992. Protein isoprenylation and methylation at carboxyl-terminal cysteine residues. *Annual Review of Biochemistry,* 61**,** 355-386.

COLLARD, B. C. & MACKILL, D. J. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society of London, Series B,* 363**,** 557-572.

COLLEMARE, J., PIANFETTI, M., HOULLE, A. E., MORIN, D., CAMBORDE, L., GAGEY, M. J., BARBISAN, C., FUDAL, I., LEBRUN, M. H. & BOHNERT, H. U. 2008. *Magnaporthe grisea* avirulence gene ACE1 belongs to an infection-specific gene cluster involved in secondary metabolism. *New Phytologist,* 179**,** 196-208.

COWAN, G. H., ROBERTS, A. G., JONES, S., KUMAR, P., KALYANDURG, P. B., GIL, J. F., SAVENKOV, E. I., HEMSLEY, P. A. & TORRANCE, L. 2018. Potato mop-top virus co-opts the stress sensor HIPP26 for long-distance movement. *Plant Physiology,* 176**,** 2052-2070.

COWTAN, K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography,* 62**,** 1002-1011.

CRUZ, C. D. & VALENT, B. 2017. Wheat blast disease: danger on the move. *Tropical Plant Pathology,* 42**,** 210-222.

DE ABREU-NETO, J. B., TURCHETTO-ZOLET, A. C., DE OLIVEIRA, L. F., ZANETTINI, M. H. & MARGIS-PINHEIRO, M. 2013. Heavy metal-associated isoprenylated plant protein (HIPP): characterization of a family of proteins exclusive to plants. *FEBS Journal,* 280**,** 1604-1616.

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**,** e1005228.

DE JONG, J. C., MCCORMACK, B. J., SMIRNOFF, N. & TALBOT, N. J. 1997. Glycerol generates turgor in rice blast. *Nature,* 389**,** 244-245.

DE LA CONCEPCION, J. C., FRANCESCHETTI, M., MACLEAN, D., TERAUCHI, R., KAMOUN, S. & BANFIELD, M. J. 2019. Protein engineering expands the effector recognition profile of a rice NLR immune receptor. *eLife,* **8,** e47713.

DE LA CONCEPCION, J. C., FRANCESCHETTI, M., MAQBOOL, A., SAITOH, H., TERAUCHI, R., KAMOUN, S. & BANFIELD, M. J. 2018. Polymorphic residues in an allelic rice NLR expand binding and response to effectors of the blast pathogen. *Nature Plants,* **4,** 576-585.

DESLANDES, L. & GENIN, S. 2014. Opening the *Ralstonia solanacearum* type III effector tool box: insights into host cell subversion mechanisms. *Current Opinion in Plant Biology,* **20,** 110-117.

DEYRUP, A. T., KRISHNAN, S., COCKBURN, B. N. & SCHWARTZ, N. B. 1998. Deletion and site-directed mutagenesis of the ATP-binding motif (P-loop) in the bifunctional murine ATP-sulfurylase/adenosine 5'-phosphosulfate kinase enzyme. *Journal of Biological Chemistry,* **273,** 9450-9456.

DJIAN-CAPORALINO, C., PALLOIX, A., FAZARI, A., MARTEU, N., BARBARY, A., ABAD, P., SAGE-PALLOIX, A.-M., MATEILLE, T., RISSO, S., LANZA, R., TAUSSIG, C. & CASTAGNONE-SERENO, P. 2014. Pyramiding, alternating or mixing: comparative performances of deployment strategies of nematode resistance genes to promote plant resistance efficiency and durability. *BMC Plant Biology,* **14,** 53.

DODDS, P. N., LAWRENCE, G. J., CATANZARITI, A. M., TEH, T., WANG, C. I., AYLIFFE, M. A., KOBE, B. & ELLIS, J. G. 2006. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proceedings of the National Academy of Sciences, USA,* **103,** 8888-8893.

ELERT, E. 2014. Rice by the numbers: A good grain. *Nature,* **514,** S50-S51.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography,* **66,** 486-501.

ENGLER, C., KANDZIA, R. & MARILLONNET, S. 2008. A one pot, one step, precision cloning method with high throughput capability. *PLOS One,* **3,** e3647.

ENGLER, C., YOULES, M., GRUETZNER, R., EHNERT, T.-M., WERNER, S., JONES, J. D. G., PATRON, N. J. & MARILLONNET, S. 2014. A Golden Gate modular cloning toolbox for plants. *ACS Synthetic Biology,* **3,** 839-843.

EPPO 2019. EPPO Global Database (available online). https://gd.eppo.int.

EVANS, P. 2006. Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography,* **62,** 72-82.

EVANS, P. R. & MURSHUDOV, G. N. 2013. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography,* 69**,** 1204-1214.

FIGUEROA BETTS, M., MANNING, V. A., CARDWELL, K. B., PANDELOVA, I. & CIUFFETTI, L. M. 2011. The importance of the N-terminus for activity of Ptr ToxB, a chlorosis-inducing host-selective toxin produced by *Pyrenophora tritici-repentis*. *Physiological and Molecular Plant Pathology,* 75**,** 138-145.

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**,** 186-194.

FUJISAKI, K., ABE, Y., ITO, A., SAITOH, H., YOSHIDA, K., KANZAKI, H., KANZAKI, E., UTSUSHI, H., YAMASHITA, T., KAMOUN, S. & TERAUCHI, R. 2015. Rice Exo70 interacts with a fungal effector, AVR-Pii, and is required for AVR-Pii-triggered immunity. *The Plant Journal,* 83**,** 875-887.

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**,** 998-1001.

GARMAN, E. F. & OWEN, R. L. 2006. Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography,* 62**,** 32-47.

GHISLAIN, M., BYARUGABA, A. A., MAGEMBE, E., NJOROGE, A., RIVERA, C., ROMÁN, M. L., TOVAR, J. C., GAMBOA, S., FORBES, G. A., KREUZE, J. F., BAREKYE, A. & KIGGUNDU, A. 2019. Stacking three late blight resistance genes from wild species directly into African highland potato varieties confers complete field resistance to local blight races. *Plant Biotechnology Journal,* 17**,** 1119-1129.

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.

GITSCHIER, J., MOFFAT, B., REILLY, D., WOOD, W. I. & FAIRBROTHER, W. J. 1998. Solution structure of the fourth metal-binding domain from the Menkes copper-transporting ATPase. *Nature Structural & Molecular Biology,* 5**,** 47-54.

GRITZ, L. & DAVIES, J. 1983. Plasmid-encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase gene and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* 25**,** 179-88.

GROSSE-KUNSTLEVE, R. W., SAUTER, N. K., MORIARTY, N. W. & ADAMS, P. D. 2002. The computational crystallography toolbox: crystallographic algorithms in a reusable software framework. *Journal of Applied Crystallography,* 35**,** 126-136.

GUBAEVA, E., GUBAEV, A., MELCHER, R. L. J., CORD-LANDWEHR, S., SINGH, R., EL GUEDDARI, N. E. & MOERSCHBACHER, B. M. 2018. 'Slipped Sandwich' model for chitin and chitosan perception in Arabidopsis. *Molecular Plant-Microbe Interactions,* 31**,** 1145-1153.

GUO, L., CESARI, S., DE GUILLEN, K., CHALVON, V., MAMMRI, L., MA, M., MEUSNIER, I., BONNOT, F., PADILLA, A., PENG, Y.-L., LIU, J. & KROJ, T. 2018. Specific recognition of two MAX effectors by integrated HMA domains in plant immune receptors involves distinct binding surfaces. *Proceedings of the National Academy of Sciences, USA,* 115**,** 11637-11642.

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

HARWOOD, W. A., BARTLETT, J. G., ALVES, S. C., PERRY, M., SMEDLEY, M. A., LEYLAND, N. & SNAPE, J. W. 2009. Barley transformation using *Agrobacterium*-mediated techniques. *Methods in Molecular Biology,* 478**,** 137-147.

HATSUGAI, N. & KATAGIRI, F. 2018. Quantification of plant cell death by electrolyte leakage assay. *Bio-protocol,* 8**,** e2758.

HAYAFUNE, M., BERISIO, R., MARCHETTI, R., SILIPO, A., KAYAMA, M., DESAKI, Y., ARIMA, S., SQUEGLIA, F., RUGGIERO, A., TOKUYASU, K., MOLINARO, A., KAKU, H. & SHIBUYA, N. 2014. Chitin-induced activation of immune signaling by the rice receptor CEBiP relies on a unique sandwich-type dimerization. *Proceedings of the National Academy of Sciences, USA,* 111**,** E404-E413.

HAYASAKA, T., FUJII, H. & ISHIGURO, K. 2008. The role of silicon in preventing appressorial penetration by the rice blast fungus. *Phytopathology,* 98**,** 1038-1044.

HAYASAKA, T., FUJII, H. & NAMAI, T. 2005. Silicon content in rice seedlings to protect rice blast fungus at the nursery stage. *Journal of General Plant Pathology,* 71**,** 169-173.

HEESE, A., HANN, D. R., GIMENEZ-IBANEZ, S., JONES, A. M. E., HE, K., LI, J., SCHROEDER, J. I., PECK, S. C. & RATHJEN, J. P. 2007. The receptor-like kinase SERK3/BAK1 is a central regulator of innate immunity in plants. *Proceedings of the National Academy of Sciences, USA,* 104**,** 12217-12222.

HELM, M., QI, M., SARKAR, S., YU, H., WHITHAM, S. A. & INNES, R. W. 2019. Engineering a decoy substrate in soybean to enable recognition of the soybean mosaic virus NIa protease. *Molecular Plant-Microbe Interactions,* 32**,** 760-769.

HENRICI, R. C., PECEN, T. J., JOHNSTON, J. L. & TAN, S. 2017. The pPSU plasmids for generating DNA molecular weight markers. *Scientific Reports,* 7**,** 2438.

HENSEL, G., VALKOV, V., MIDDLEFELL-WILLIAMS, J. & KUMLEHN, J. 2008. Efficient generation of transgenic barley: the way forward to modulate plant–microbe interactions. *Journal of Plant Physiology,* 165**,** 71-82.

HITTALMANI, S., PARCO, A., MEW, T. V., ZEIGLER, R. S. & HUANG, N. 2000. Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. *Theoretical and Applied Genetics,* 100**,** 1121-1128.

HOLT, B. F., BOYES, D. C., ELLERSTRÖM, M., SIEFERS, N., WIIG, A., KAUFFMAN, S., GRANT, M. R. & DANGL, J. L. 2002. An evolutionarily conserved mediator of plant disease resistance gene function is required for normal *Arabidopsis* development. *Developmental Cell,* 2**,** 807-817.

HOU, S., LIU, Z., SHEN, H. & WU, D. 2019. Damage-associated molecular pattern-triggered immunity in plants. *Frontiers in Plant Science,* 10**,** 646.

HU, G., RICHTER, T. E., HULBERT, S. H. & PRYOR, T. 1996. Disease lesion mimicry caused by mutations in the rust resistance gene rp1. *The Plant Cell,* 8**,** 1367-1376.

HUA, L., WU, J., CHEN, C., WU, W., HE, X., LIN, F., WANG, L., ASHIKAWA, I., MATSUMOTO, T., WANG, L. & PAN, Q. 2012. The isolation of Pi1, an allele at the Pik locus which confers broad spectrum resistance to rice blast. *Theoretical and Applied Genetics,* 125**,** 1047-1055.

HUANG, J., SI, W. N., DENG, Q. M., LI, P. & YANG, S. H. 2014. Rapid evolution of avirulence genes in rice blast fungus *Magnaporthe oryzae*. *BMC Genetics,* 15.

HURST, C. H., TURNBULL, D., MYLES, S. M., LESLIE, K., KEINATH, N. F. & HEMSLEY, P. A. 2018. Variable effects of C-terminal fusions on FLS2 function: not all epitope tags are created equal. *Plant Physiology,* 177**,** 522-531.

INCARDONA, M. F., BOURENKOV, G. P., LEVIK, K., PIERITZ, R. A., POPOV, A. N. & SVENSSON, O. 2009. EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *Journal of Synchrotron Radiation,* 16**,** 872-879.

INOUE, Y., VY, T. T. P., YOSHIDA, K., ASANO, H., MITSUOKA, C., ASUKE, S., ANH, V. L., CUMAGUN, C. J. R., CHUMA, I., TERAUCHI, R., KATO, K., MITCHELL, T., VALENT, B., FARMAN, M. & TOSA, Y. 2017. Evolution of the wheat blast fungus through functional losses in a host specificity determinant. *Science,* 357**,** 80-83.

IRRI 2018. World Rice Statistics Online Query Facility. http://ricestat.irri.org:8080/wrsv3/entrypoint.htm: IRRI.

IRRI, RICE, A. & CIAT 2010. Global Rice Science Partnership (GRiSP). IRRI, Africa Rice, CIAT.

ISLAM, M. T., CROLL, D., GLADIEUX, P., SOANES, D. M., PERSOONS, A., BHATTACHARJEE, P., HOSSAIN, M. S., GUPTA, D. R., RAHMAN, M. M.,

MAHBOOB, M. G., COOK, N., SALAM, M. U., SUROVY, M. Z., SANCHO, V. B., MACIEL, J. L. N., NHANIJÚNIOR, A., CASTROAGUDÍN, V. L., REGES, J. T. D. A., CERESINI, P. C., RAVEL, S., KELLNER, R., FOURNIER, E., THARREAU, D., LEBRUN, M.-H., MCDONALD, B. A., STITT, T., SWAN, D., TALBOT, N. J., SAUNDERS, D. G. O., WIN, J. & KAMOUN, S. 2016. Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *BMC Biology,* 14**,** 84.

ISLAM, M. T., KIM, K. H. & CHOI, J. 2019. Wheat blast in Bangladesh: the current situation and future impacts. *The Plant Pathology Journal,* 35**,** 1-10.

IWATA, M. 2001. Probenazole - a plant defence activator. *Pesticide Outlook,* 12**,** 28-31.

JAYARAMAN, J., CHOI, S., PROKCHORCHIK, M., CHOI, D. S., SPIANDORE, A., RIKKERINK, E. H., TEMPLETON, M. D., SEGONZAC, C. & SOHN, K. H. 2017. A bacterial acetyltransferase triggers immunity in *Arabidopsis thaliana* independent of hypersensitive response. *Scientific Reports,* 7.

JIA, Y., MCADAMS, S. A., BRYAN, G. T., HERSHEY, H. P. & VALENT, B. 2000. Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *The EMBO Journal,* 19**,** 4004-4014.

JIANG, S., YAO, J., MA, K.-W., ZHOU, H., SONG, J., HE, S. Y. & MA, W. 2013. Bacterial effector activates jasmonate signaling by directly targeting JAZ transcriptional repressors. *PLOS Pathogens,* 9**,** e1003715.

JIN, S. G., KOMARI, T., GORDON, M. P. & NESTER, E. W. 1987. Genes responsible for the supervirulence phenotype of *Agrobacterium tumefaciens* A281. *Journal of Bacteriology,* 169**,** 4417-4425.

JONES, J. D. G. & DANGL, J. L. 2006. The plant immune system. *Nature,* 444**,** 323-329.

KAKU, H., NISHIZAWA, Y., ISHII-MINAMI, N., AKIMOTO-TOMIYAMA, C., DOHMAE, N., TAKIO, K., MINAMI, E. & SHIBUYA, N. 2006. Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proceedings of the National Academy of Sciences, USA,* 103**,** 11086-11091.

KANG, B.-C., YUN, J.-Y., KIM, S.-T., SHIN, Y., RYU, J., CHOI, M., WOO, J. W. & KIM, J.-S. 2018. Precision genome engineering through adenine base editing in plants. *Nature Plants,* 4**,** 427-431.

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**,** 894-907.

KAWASHIMA, C. G., GUIMARAES, G. A., NOGUEIRA, S. R., MACLEAN, D., COOK, D. R., STEUERNAGEL, B., BAEK, J., BOUYIOUKOS, C., MELO BDO, V., TRISTAO, G., DE OLIVEIRA, J. C., RAUSCHER, G., MITTAL, S., PANICHELLI,

L., BACOT, K., JOHNSON, E., IYER, G., TABOR, G., WULFF, B. B., WARD, E., RAIRDAN, G. J., BROGLIE, K. E., WU, G., VAN ESSE, H. P., JONES, J. D. & BROMMONSCHENKEL, S. H. 2016. A pigeonpea gene confers resistance to Asian soybean rust in soybean. *Nature Biotechnology,* 34**,** 661-665.

KHAN, M., SETO, D., SUBRAMANIAM, R. & DESVEAUX, D. 2018. Oh, the places they'll go! A survey of phytopathogen effectors and their host targets. *The Plant Journal,* 93**,** 651-663.

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**,** 684-687.

KRISSINEL, E. 2010. Crystal contacts as nature's docking solutions. *Journal of Computational Chemistry,* 31**,** 133-143.

KRISSINEL, E. & HENRICK, K. 2007. Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology,* 372**,** 774-797.

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**,** 618-626.

LEE, A. H.-Y., HURLEY, B., FELSENSTEINER, C., YEA, C., CKURSHUMOVA, W., BARTETZKO, V., WANG, P. W., QUACH, V., LEWIS, J. D., LIU, Y. C., BÖRNKE, F., ANGERS, S., WILDE, A., GUTTMAN, D. S. & DESVEAUX, D. 2012. A bacterial acetyltransferase destroys plant microtubule networks and blocks secretion. *PLOS Pathogens,* 8**,** e1002523.

LEONELLI, L., PELTON, J., SCHOEFFLER, A., DAHLBECK, D., BERGER, J., WEMMER, D. E. & STASKAWICZ, B. 2011. Structural elucidation and functional characterization of the *Hyaloperonospora arabidopsidis* effector protein ATR13. *PLOS Pathogens,* 7**,** e1002428.

LI, J., WANG, Q., LI, C., BI, Y., FU, X. & WANG, R. 2019a. Novel haplotypes and networks of AVR-Pik alleles in *Magnaporthe oryzae*. *BMC Plant Biology,* 19**,** 204-204.

LI, Z., HUANG, J., WANG, Z., MENG, F., ZHANG, S., WU, X., ZHANG, Z. & GAO, Z. 2019b. Overexpression of *Arabidopsis* nucleotide-binding and leucine-rich repeat genes RPS2 and RPM1(D505V) confers broad-spectrum disease resistance in rice. *Frontiers in Plant Science,* 10**,** 417.

LO PRESTI, L., LANVER, D., SCHWEIZER, G., TANAKA, S., LIANG, L., TOLLOT, M., ZUCCARO, A., REISSMANN, S. & KAHMANN, R. 2015. Fungal effectors and plant susceptibility. *Annual Review of Plant Biology,* 66**,** 513-545.

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.

LONGYA, A., CHAIPANYA, C., FRANCESCHETTI, M., MAIDMENT, J. H. R., BANFIELD, M. J. & JANTASURIYARAT, C. 2019. Gene duplication and mutation in the emergence of a novel aggressive allele of the AVR-Pik effector in the rice blast fungus. *Molecular Plant-Microbe Interactions,* 32**,** 740-749.

MACHO, A. P. & ZIPFEL, C. 2014. Plant PRRs and the activation of innate immune signaling. *Molecular Cell,* 54**,** 263-272.

MAQBOOL, A., SAITOH, H., FRANCESCHETTI, M., STEVENSON, C. E. M., 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**,** e08709.

MATTHEWS, B. W. 1968. Solvent content of protein crystals. *Journal of Molecular Biology,* 33**,** 491-497.

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**,** 658-674.

MCNICHOLAS, S., POTTERTON, E., WILSON, K. S. & NOBLE, M. E. M. 2011. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D: Biological Crystallography,* 67**,** 386-394.

MEW, T., HIBINO, H., SAVARY, S., VERA CRUZ, C., OPULENCIA, R. & HETTEL, G. 2018. Rice diseases: biology and selected management practices. Los Baños (Philippines): International Rice Research Institute.

MIKI, B. & MCHUGH, S. 2004. Selectable marker genes in transgenic plants: applications, alternatives and biosafety. *Journal of Biotechnology,* 107**,** 193-232.

MISHRA, R., JOSHI, R. K. & ZHAO, K. 2018. Genome editing in rice: recent advances, challenges, and future implications. *Frontiers in Plant Science,* 9.

MOTTALEB, K. A., SINGH, P. K., SONDER, K., KRUSEMAN, G., TIWARI, T. P., BARMA, N. C. D., MALAKER, P. K., BRAUN, H.-J. & ERENSTEIN, O. 2018. Threat of wheat blast to South Asia's food security: An ex-ante analysis. *PLOS One,* 13**,** e0197555.

MURSHUDOV, G. N., SKUBÁK, 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: Biological Crystallography,* 67**,** 355-367.

NEMRI, A., SAUNDERS, D., ANDERSON, C., UPADHYAYA, N., WIN, J., LAWRENCE, G., JONES, D., KAMOUN, S., ELLIS, J. & DODDS, P. 2014. The genome sequence

and effector complement of the flax rust pathogen *Melampsora lini*. *Frontiers in Plant Science,* 5**,** 98.

OH, C.-S. & MARTIN, G. B. 2011. Effector-triggered immunity mediated by the Pto kinase. *Trends in Plant Science,* 16**,** 132-140.

ORBACH, M. J., FARRALL, L., SWEIGARD, J. A., CHUMLEY, F. G. & VALENT, B. 2000. A telomeric avirulence gene determines efficacy for the rice blast resistance gene Pi-ta. *The Plant Cell,* 12**,** 2019-2032.

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**,** 156-168.

OSBOURN, A. E. 1996. Preformed antimicrobial compounds and plant defense against fungal attack. *The Plant Cell,* 8**,** 1821-1831.

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**,** 229-235.

PAN, Q., WENDEL, J. & FLUHR, R. 2000. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *Journal of Molecular Evolution,* 50**,** 203-213.

PARK, C.-H., CHEN, S., SHIRSEKAR, G., ZHOU, B., KHANG, C. H., SONGKUMARN, P., AFZAL, A. J., NING, Y., WANG, R., BELLIZZI, M., VALENT, B. & WANG, G.-L. 2012. The *Magnaporthe oryzae* effector AvrPiz-t targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern-triggered immunity in rice. *The Plant Cell,* 24**,** 4748-4762.

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**,** e1005529.

PAVAN, S., JACOBSEN, E., VISSER, R. G. & BAI, Y. 2010. Loss of susceptibility as a novel breeding strategy for durable and broad-spectrum resistance. *Molecular Breeding,* 25**,** 1-12.

PENNISI, E. 2010. Armed and dangerous. *Science,* 327**,** 804-805.

PETIT-HOUDENOT, Y., LANGNER, T., HARANT, A., WIN, J. & KAMOUN, S. 2019. A clone resource of *Magnaporthe oryzae* effectors that share sequence and structural similarities across host-specific lineages. *Zenodo*.

PINTA, W., TOOJINDA, T., THUMMABENJAPONE, P. & SANITCHON, J. 2013. Pyramiding of blast and bacterial leaf blight resistance genes into rice cultivar

RD6 using marker assisted selection. *African Journal of Biotechnology,* 12**,** 4432-4438.

R CORE DEVELOPMENT TEAM 2018. R: A language and environment for statistical computing. . Vienna, Austria: R Foundation for Statistical Computing.

RADAKOVIC, Z. S., ANJAM, M. S., ESCOBAR, E., CHOPRA, D., CABRERA, J., SILVA, A. C., ESCOBAR, C., SOBCZAK, M., GRUNDLER, F. M. W. & SIDDIQUE, S. 2018. Arabidopsis HIPP27 is a host susceptibility gene for the beet cyst nematode *Heterodera schachtii*. *Molecular Plant Pathology,* 19**,** 1917-1928.

RON, M. & AVNI, A. 2004. The receptor for the fungal elicitor ethylene-inducing xylanase is a member of a resistance-like gene family in tomato. *The Plant Cell,* 16**,** 1604-1615.

SAKULKOO, W., OSÉS-RUIZ, M., OLIVEIRA GARCIA, E., SOANES, D. M., LITTLEJOHN, G. R., HACKER, C., CORREIA, A., VALENT, B. & TALBOT, N. J. 2018. A single fungal MAP kinase controls plant cell-to-cell invasion by the rice blast fungus. *Science,* 359**,** 1399-1403.

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**,** 1-19.

SARRIS, P. F., DUXBURY, Z., HUH, S. U., MA, Y., SEGONZAC, C., SKLENAR, J., DERBYSHIRE, P., CEVIK, V., RALLAPALLI, G., SAUCET, S. B., WIRTHMUELLER, L., MENKE, F. L., SOHN, K. H. & JONES, J. D. 2015. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell,* 161**,** 1089-1100.

SAUNDERS, D. G. O., WIN, J., CANO, L. M., SZABO, L. J., KAMOUN, S. & RAFFAELE, S. 2012. Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. *PLOS One,* 7**,** e29847.

SAUR, I. M. L., BAUER, S., KRACHER, B., LU, X., FRANZESKAKIS, L., MÜLLER, M. C., SABELLECK, B., KÜMMEL, F., PANSTRUGA, R., MAEKAWA, T. & SCHULZE-LEFERT, P. 2019. Multiple pairs of allelic MLA immune receptor-powdery mildew AVRA effectors argue for a direct recognition mechanism. *eLife,* 8**,** e44471.

SAUTER, N. K., GROSSE-KUNSTLEVE, R. W. & ADAMS, P. D. 2004. Robust indexing for automatic data collection. *Journal of Applied Crystallography,* 37**,** 399-409.

SAVARY, S., WILLOCQUET, L., PETHYBRIDGE, S. J., ESKER, P., MCROBERTS, N. & NELSON, A. 2019. The global burden of pathogens and pests on major food crops. *Nature Ecology and Evolution,* 3**,** 430-439.

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, USA,* 107**,** 17421-17426.

SECK, P. A., DIAGNE, A., MOHANTY, S. & WOPEREIS, M. C. S. 2012. Crops that feed the world 7: Rice. *Food Security,* 4**,** 7-24.

SHAO, F., GOLSTEIN, C., ADE, J., STOUTEMYER, M., DIXON, J. E. & INNES, R. W. 2003. Cleavage of *Arabidopsis* PBS1 by a bacterial type III effector. *Science,* 301**,** 1230-1233.

SHAUL, O. 2017. How introns enhance gene expression. *International Journal of Biochemistry and Cell Biology,* 91**,** 145-155.

SHIMIZU, T., NAKANO, T., TAKAMIZAWA, D., DESAKI, Y., ISHII-MINAMI, N., NISHIZAWA, Y., MINAMI, E., OKADA, K., YAMANE, H., KAKU, H. & SHIBUYA, N. 2010. Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. *The Plant Journal,* 64**,** 204-214.

SOHN, K. H., SEGONZAC, C., RALLAPALLI, G., SARRIS, P. F., WOO, J. Y., WILLIAMS, S. J., NEWMAN, T. E., PAEK, K. H., KOBE, B. & JONES, J. D. G. 2014. The nuclear immune receptor RPS4 is required for RRS1(SLH1)-dependent constitutive defense activation in *Arabidopsis thaliana*. *PLOS Genetics,* 10**,** e1004655.

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**,** 2094-2110.

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**,** 743-761.

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**,** 252-260.

STACEY, G. & SHIBUYA, N. 1997. Chitin recognition in rice and legumes. *Plant and Soil,* 194**,** 161-169.

STERGIOPOULOS, I., VAN DEN BURG, H. A., OKMEN, B., BEENEN, H. G., VAN LIERE, S., KEMA, G. H. & DE WIT, P. J. 2010. Tomato Cf resistance proteins mediate recognition of cognate homologous effectors from fungi pathogenic on dicots and monocots. *Proceedings of the National Academy of Sciences, USA,* 107**,** 7610-7615.

STRUG, I., UTZAT, C., CAPPIONE, A., 3RD, GUTIERREZ, S., AMARA, R., LENTO, J., CAPITO, F., SKUDAS, R., CHERNOKALSKAYA, E. & NADLER, T. 2014. Development of a univariate membrane-based mid-infrared method for protein

quantitation and total lipid content analysis of biological samples. *Journal of Analytical Methods in Chemistry,* 2014**,** 657079.

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

SUN, W., ZHANG, J., FAN, Q., XUE, G., LI, Z. & LIANG, Y. 2010. Silicon-enhanced resistance to rice blast is attributed to silicon-mediated defence resistance and its role as physical barrier. *European Journal of Plant Pathology,* 128**,** 39-49.

TAKKEN, F. L. W., ALBRECHT, M. & TAMELING, W. I. L. 2006. Resistance proteins: molecular switches of plant defence. *Current Opinion in Plant Biology,* 9**,** 383-390.

TAKKEN, F. L. W. & GOVERSE, A. 2012. How to build a pathogen detector: structural basis of NB-LRR function. *Current Opinion in Plant Biology,* 15**,** 375-384.

TAMELING, W. I. L., ELZINGA, S. D. J., DARMIN, P. S., VOSSEN, J. H., TAKKEN, F. L. W., HARING, M. A. & CORNELISSEN, B. J. C. 2002. The tomato R gene products I-2 and Mi-1 are functional ATP binding proteins with ATPase activity. *The Plant Cell,* 14**,** 2929-2939.

TAMELING, W. I. L., VOSSEN, J. H., ALBRECHT, M., LENGAUER, T., BERDEN, J. A., HARING, M. A., CORNELISSEN, B. J. C. & TAKKEN, F. L. W. 2006. Mutations in the NB-ARC domain of I-2 that impair ATP hydrolysis cause autoactivation. *Plant Physiology,* 140**,** 1233-1245.

TANG, M. Z., NING, Y. S., SHU, X. L., DONG, B., ZHANG, H. Y., WU, D. X., WANG, H., WANG, G. L. & ZHOU, B. 2017. The Nup98 homolog APIP12 targeted by the effector AvrPiz-t is involved in rice basal resistance against *Magnaporthe oryzae*. *Rice,* 10.

TANG, X., XIE, M., KIM, Y. J., ZHOU, J., KLESSIG, D. F. & MARTIN, G. B. 1999. Overexpression of Pto activates defense responses and confers broad resistance. *The Plant Cell,* 11**,** 15-29.

TAO, Y., YUAN, F., LEISTER, R. T., AUSUBEL, F. M. & KATAGIRI, F. 2000. Mutational analysis of the Arabidopsis nucleotide binding site–leucine-rich repeat resistance gene RPS2. *The Plant Cell,* 12**,** 2541-2554.

TEHSEEN, M., CAIRNS, N., SHERSON, S. & COBBETT, C. S. 2010. Metallochaperone-like genes in *Arabidopsis thaliana*. *Metallomics,* 2**,** 556-564.

TERADA, R. & SHIMAMOTO, K. 1990. Expression of CaMV35S-GUS gene in transgenic rice plants. *Molecular and General Genetics,* 220**,** 389-392.

UNDERWOOD, W. 2012. The plant cell wall: a dynamic barrier against pathogen invasion. *Frontiers in Plant Science,* 3**,** 85.

VALENT, B., FARMAN, M., TOSA, Y., BEGEROW, D., FOURNIER, E., GLADIEUX, P., ISLAM, M. T., KAMOUN, S., KEMLER, M., KOHN, L. M., LEBRUN, M.-H., STAJICH, J. E., TALBOT, N. J., TERAUCHI, R., THARREAU, D. & ZHANG, N. 2019. *Pyricularia graminis-tritici* is not the correct species name for the wheat blast fungus: response to Ceresini et al. (MPP 20:2). *Molecular Plant Pathology,* 20**,** 173-179.

VAN DEN ELZEN, P. J. M., TOWNSEND, J., LEE, K. Y. & BEDBROOK, J. R. 1985. A chimaeric hygromycin resistance gene as a selectable marker in plant cells. *Plant Molecular Biology,* 5**,** 299-302.

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**,** 2009-2017.

VAN ESSE, H. P., REUBER, L. & VAN DER DOES, D. 2019. GM approaches to improve disease resistance in crops. *New Phytologist*.

VAN LAREBEKE, N., ENGLER, G., HOLSTERS, M., VAN DEN ELSACKER, S., ZAENEN, I., SCHILPEROORT, R. A. & SCHELL, J. 1974. Large plasmid in *Agrobacterium tumefaciens* essential for crown gall-inducing ability. *Nature,* 252**,** 169-170.

VAN OOIJEN, G., MAYR, G., KASIEM, M. M. A., ALBRECHT, M., CORNELISSEN, B. J. C. & TAKKEN, F. L. W. 2008. Structure-function analysis of the NB-ARC domain of plant disease resistance proteins. *Journal of Experimental Botany,* 59**,** 1383-1397.

VARDEN, F. A., SAITOH, H., YOSHINO, K., FRANCESCHETTI, M., KAMOUN, S., TERAUCHI, R. & BANFIELD, M. J. 2019. Cross-reactivity of a rice NLR immune receptor to distinct effectors from the rice blast pathogen *Magnaporthe oryzae* provides partial disease resistance. *Journal of Biological Chemistry*.

VONRHEIN, C., FLENSBURG, C., KELLER, P., SHARFF, A., SMART, O., PACIOREK, W., WOMACK, T. & BRICOGNE, G. 2011. Data processing and analysis with the autoPROC toolbox. *Acta Crystallographica Section D: Biological Crystallography,* 67**,** 293-302.

WAGNER, S., STUTTMANN, J., RIETZ, S., GUEROIS, R., BRUNSTEIN, E., BAUTOR, J., NIEFIND, K. & PARKER, JANE E. 2013. Structural basis for signaling by exclusive EDS1 heteromeric complexes with SAG101 or PAD4 in plant innate immunity. *Cell Host & Microbe,* 14**,** 619-630.

WAN, J., ZHANG, X.-C. & STACEY, G. 2008. Chitin signaling and plant disease resistance. *Plant Signaling and Behaviour,* 3**,** 831-833.

WANG, J., HU, M., WANG, J., QI, J., HAN, Z., WANG, G., QI, Y., WANG, H. W., ZHOU, J. M. & CHAI, J. 2019a. Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science,* 364**,** eaav5870.

WANG, J., WANG, J., HU, M., WU, S., QI, J., WANG, G., HAN, Z., QI, Y., GAO, N., WANG, H. W., ZHOU, J. M. & CHAI, J. 2019b. Ligand-triggered allosteric ADP release primes a plant NLR complex. *Science,* 364**,** eaav5868.

WANG, R., NING, Y., SHI, X., HE, F., ZHANG, C., FAN, J., JIANG, N., ZHANG, Y., ZHANG, T., HU, Y., BELLIZZI, M. & WANG, G. L. 2016. Immunity to rice blast disease by suppression of effector-triggered necrosis. *Current Biology,* 26**,** 2399-2411.

WANG, X., LEE, S., WANG, J., MA, J., BIANCO, T. & JI, Y. 2014. Current advances on genetic resistance to rice blast disease. *In:* YAN, W. (ed.) *Rice: Germplasm, Genetics and Improvement.* IntechOpen.

WANG, Y., ZHANG, Q., SUN, M. A. & GUO, D. 2011. High-accuracy prediction of bacterial type III secreted effectors based on position-specific amino acid composition profiles. *Bioinformatics,* 27**,** 777-784.

WEBER, E., ENGLER, C., GRUETZNER, R., WERNER, S. & MARILLONNET, S. 2011. A modular cloning system for standardized assembly of multigene constructs. *PLOS One,* 6**,** e16765.

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. 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature,* 450**,** 115-118.

WHITMORE, L. & WALLACE, B. A. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research,* 32**,** 668-73.

WHITMORE, L. & WALLACE, B. A. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers,* 89**,** 392-400.

WICKHAM, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*, Springer International Publishing.

WILLIAMS, S. J., SORNARAJ, P., DECOURCY-IRELAND, E., MENZ, R. I., KOBE, B., ELLIS, J. G., DODDS, P. N. & ANDERSON, P. A. 2011. An autoactive mutant of the M flax rust resistance protein has a preference for binding ATP, whereas wild-type M protein binds ADP. *Molecular Plant-Microbe Interactions,* 24**,** 897-906.

WILSON, R. A. & TALBOT, N. J. 2009. Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nature Reviews: Microbiology,* 7**,** 185-195.

WIN, J., MORGAN, W., BOS, J., KRASILEVA, K. V., CANO, L. M., CHAPARRO-GARCIA, A., AMMAR, R., STASKAWICZ, B. J. & KAMOUN, S. 2007. Adaptive

evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *The Plant Cell,* 19**,** 2349-2369.

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: Biological Crystallography,* 67**,** 235-242.

WINN, M. D., MURSHUDOV, G. N. & PAPIZ, M. Z. 2003. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods in Enzymology,* 374**,** 300-321.

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, USA,* 114**,** 8113-8118.

WU, C.-H., KRASILEVA, K. V., BANFIELD, M. J., TERAUCHI, R. & KAMOUN, S. 2015. The "sensor domains" of plant NLR proteins: more than decoys? *Frontiers in Plant Science,* 6**,** 134.

WU, Y., WANG, Y., LI, J., LI, W., ZHANG, L., LI, Y., LI, X., LI, J., ZHU, L. & WU, G. 2014. Development of a general method for detection and quantification of the P35S promoter based on assessment of existing methods. *Scientific Reports,* 4**,** 7358.

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**,** 1573-1591.

YOSHIOKA, K., NAKASHITA, H., KLESSIG, D. F. & YAMAGUCHI, I. 2001. Probenazole induces systemic acquired resistance in *Arabidopsis* with a novel type of action. *The Plant Journal,* 25**,** 149-157.

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**,** 1017-1028.

YUE, J.-X., MEYERS, B. C., CHEN, J.-Q., TIAN, D. & YANG, S. 2012. Tracing the origin and evolutionary history of plant nucleotide-binding site–leucine-rich repeat (NBS-LRR) genes. *New Phytologist,* 193**,** 1049-1063.

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**,** 321-334.

ZHAI, C., ZHANG, Y., YAO, N., LIN, F., LIU, Z., DONG, Z., WANG, L. & PAN, Q. 2014. Function and interaction of the coupled genes responsible for Pik-h encoded rice blast resistance. *PLOS One,* 9**,** e98067.

ZHANG, L., CHEN, S., RUAN, J., WU, J., TONG, A. B., YIN, Q., LI, Y., DAVID, L., LU, A., WANG, W. L., MARKS, C., OUYANG, Q., ZHANG, X., MAO, Y. & WU, H. 2015a. Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science,* 350**,** 404-409.

ZHANG, L., YAO, J., WITHERS, J., XIN, X.-F., BANERJEE, R., FARIDUDDIN, Q., NAKAMURA, Y., NOMURA, K., HOWE, G. A., BOLAND, W., YAN, H. & HE, S. Y. 2015b. Host target modification as a strategy to counter pathogen hijacking of the jasmonate hormone receptor. *Proceedings of the National Academy of Sciences, USA,* 112**,** 14354-14359.

ZHOU, Z., PANG, Z., ZHAO, S., ZHANG, L., LV, Q., YIN, D., LI, D., LIU, X., ZHAO, X., LI, X., WANG, W. & ZHU, L. 2019. Importance of OsRac1 and RAI1 in signalling of nucleotide-binding site leucine-rich repeat protein-mediated resistance to rice blast disease. *New Phytologist,* 223**,** 828-838.

ZSCHIESCHE, W., BARTH, O., DANIEL, K., BOEHME, S., RAUSCHE, J. & HUMBECK, K. 2015. The zinc-binding nuclear protein HIPP3 acts as an upstream regulator of the salicylate-dependent plant immunity pathway and of flowering time in *Arabidopsis thaliana*. *New Phytologist,* 207**,** 1084-1096.

Appendix

Publications

This thesis gave rise to the following publications:

LONGYA, A., CHAIPANYA, C., FRANCESCHETTI, M., **MAIDMENT, J. H. R.**, BANFIELD, M. J. & JANTASURIYARAT, C. 2019. Gene duplication and mutation in the emergence of a novel aggressive allele of the AVR-Pik effector in the rice blast fungus. *Molecular Plant-Microbe Interactions,* 32**,** 740-749

VARDEN, F. A.*, DE LA CONCEPCION, J. C.*, **MAIDMENT, J. H. R.*** & BANFIELD, M. J. 2017. Taking the stage: effectors in the spotlight. Current Opinion in Plant Biology, 38, 25-33.

* indicates that these authors contributed equally.

Table of DNA primers used in this project

Table A DNA primers used in this project.

For DNA primers used to generate PCR products for In-Fusion® cloning, sequences homologous to In-Fusion® sites are highlighted in red. For DNA primers used to generate PCR products for Golden Gate assembly, BsaI restriction sites are highlighted in blue, BpiI restriction sites are highlighted in purple and the resulting overhangs are highlighted in green. Where DNA primers have been designed to incorporate a mutation into the template sequence, the substituted nucleotides are underlined.

Primer ID	Sequence (5' to 3')	Purpose
M13F	TGTAAAACGACGGCCAGT	Forward sequencing primer for pCR8
M13R	CAGGAAACAGCTATGACC	Reverse sequencing primer for pCR8
1182	GAACCCTGTGGTTGGCATGCACATAC	Reverse sequencing primer for GG level 1 acceptors
1183	CTGGTGGCAGGATATATTGTGGTG	Forward sequencing primer for GG level 1 acceptors
1184	GTGGTGTAAACAAATTGACGC	Forward primer for GG level 2 acceptor (pICSL4723)

Primer ID	Sequence (5' to 3')	Purpose
1185	GGATAAACCTTTTCACGCCC	Reverse primer for GG level 2 acceptor (pICSL4723)
54	CACCACCTTCTGATAGGCAG	Reverse sequencing primer for pOPIN-M
54a	TAATACGACTCACTATAGGG	Forward sequencing primer for pOPIN vectors
64	GGTGAAATCATGCCGAACATC	Forward sequencing primer near end of MBP tag in pOPIN-M
676	CAAGGGGCTTCATGATGTCC	Reverse sequencing primer for pOPIN-M
192	TAGTTATTGCTCAGCGGTGG	Reverse sequencing primer for pOPIN-A
1893	GTGAAAGTTGGAACCTCTTACGTG	Forward sequencing primer for pUAP1
1894	CACATGTTCTTTCCTGCGTTATCC	Reverse sequencing primer for pUAP1
1076	TGGAGGCAAAACAACGG	Sequencing primer for Pikp-1
1077	GGTCGTCTGGTCAGGAGG	Sequencing primer for Pikp-1

Primer ID	Sequence (5' to 3')	Purpose
1078	ATTTGGAGATTTTGTATGTGG	Sequencing primer for Pikp-1
1097	GAAGTGCGTTCCCAAGG	Sequencing primer for Pikp-1
1734	CTCGGATCTGATGAACTC	Sequencing primer for Pikp-1
1735	GGCAGGTGTCCTACCTG	Sequencing primer for Pikp-1
1966	GCTGTCGTCAGCATTGGTTG	Sequencing primer for Pikp-1
1967	GTATAAGGAGATATCAACTGAGTCACAC	Sequencing primer for Pikp-1
1736	GATGTACTGGATGTCACCAC	Sequencing primer for Pikp-2
1737	CTCCGTGAGGTGCATC	Sequencing primer for Pikp-2
1094	GTCTCAGAACTATGACCAAGG	Sequencing primer for Pikp-2
1095	GTTACAATGATTTGCCTGCG	Sequencing primer for Pikp-2

Primer ID	Sequence (5' to 3')	Purpose
1096	CCTCAAGAAAAAAGTAAGAAGC	Sequencing primer for Pikp-2
1974	GGAAACAACGGACGATGATC	Sequencing primer for Pikp-2
1482	TTGATTTGCTATTGGCTAGG	Sequencing primer for native promoter
1483	GACCACGAGTCAGTCAACG	Sequencing primer for native promoter
1484	AGGTAAAATCTCAGCGTTAGG	Sequencing primer for native promoter
1995	CTGATTCTATTGGATGCATGCATTG	Sequencing primer for native promoter
1996	GCATATAGGCAATTAACTATGACCTC	Sequencing primer for native promoter
2119	CAGGACTGAGCCTCAAGAACCAC	Sequencing primer for bialaphos resistance cassette
2120	CTGTTGCCGGTCTTGCGATGATTATC	Sequencing primer for bialaphos resistance cassette
1507	CGATAGTGGAAACCGACGC	Sequencing primer for hygromycin resistance cassette

Primer ID	Sequence (5' to 3')	Purpose
1897	GAGAGGCCTACGCAGCAAG	Sequencing primer for hygromycin resistance cassette
1898	CTGAACTCACCGCGACGTC	Sequencing primer for hygromycin resistance cassette
1964	CCGATGCAAAGTGCCGATAAAC	Sequencing primer for hygromycin resistance cassette
1965	CAACAATGTCCTGACGGACAATG	Sequencing primer for hygromycin resistance cassette
2073	CCATTCGGACCGCAAGGAATCGGTCAATAC	Sequencing primer for hygromycin resistance cassette
2074	GAGTACTTCTACACAGCCATCGGTCCAGAC	Sequencing primer for hygromycin resistance cassette
2073	CCATTCGGACCGCAAGGAATCGGTCAATAC	Forward primer for checking barley transformation (hygromycin resistance cassette)
2074	GAGTACTTCTACACAGCCATCGGTCCAGAC	Reverse primer for checking barley transformation (hygromycin resistance cassette
1235	AAGGTCTCACAGAAAAACAGGGCTAAAGCAAAAGATCGTGATC	Forward primer for cloning the HMA domain of OsHIPP19 into pCR™8 for incorporation into Pikp-1

Primer ID	Sequence (5' to 3')	Purpose
1236	TTGGTCTCACGTCCTTCTCCTTCACCTCCTCCA	Reverse primer for cloning the HMA domain of OsHIPP19 into pCR™8 for incorporation into Pikp-1
1248	AAGGTCTCACAGAAAAACAGGGCTAAAGCAAAAAATCGTG	Forward primer for cloning Pikp-HMA into pCR™8
1257	TTGGTCTCACGTCTTTCTTTGCTTGGCTGACCT	Reverse primer for cloning Pikp-HMA^{N261K} into pCR™8
1513	TTGGTCTCACGTCTTCCTTTGCTTGGCTGACCT	Reverse primer for cloning Pikp-HMA^{NK-KE} into pCR™8
1514	TTGGTCTCACGTCTTCCTTTGCTTGCTCGACCT	Reverse primer for cloning Pikp-HMA$^{SNK-EKE}$ into pCR™8
1515	TTGGTCTCACGTCTTTATTTGCTTGCTCGACCT	Reverse primer for cloning Pikp-HMA^{S258E} into pCR™8
1517	GTTGCGCTCGTAGGTGCTCTAAGAGACAAGATAGAGG	Internal primer for mutating D224 to A
1518	GGTTGCGCTCGTAGGTAAGCTAAGAGACAAGATAGAGG	Internal primer for mutating D224 to K
1404	CTTGGATTGCCAGGTGGAGGCAGAACAACGGTTGCCAGAG	Internal primer for mutating K296 (p-loop motif) to R

Primer ID	Sequence (5' to 3')	Purpose
1393	ATGGTCTCAAATGGAGTTGGTGGTAGGTGCTTCCG	Forward primer for cloning the CC domain of Pikp-2 into pCR™8
1394	ATGGTCTCATCCCCACAGGCTCCTTTATACCAATGATCTGAGG	Reverse primer for cloning the CC domain of Pikp-2 into pCR™8
1395	ATGGTCTCAGGGATGAAGACGGTCATGGAGGAGCTTG	Forward primer for cloning the NBARC domain of Pikp-2 into pCR™8
1396	ATGGTCTCACAGGCCTTTTGTTGAACTTCCACGATTGGATCC	Reverse primer for cloning the NBARC domain of Pikp-2 into pCR™8
1397	ATGGTCTCACCTGAACTTGGCTCAAGTGAGATCACTGACG	Forward primer for cloning the LRR domain of Pikp-2 into pCR™8
1398	ATGGTCTCACGAAGCAGTGACGATGCCATCAACAAATACTTTTG	Reverse primer for cloning the LRR domain of Pikp-2 into pCR™8
1407	GTTTTGGAGGTGTGGGAAGGACTACCATTGCCACAGC	Internal primer for mutating K217 (p-loop motif) to R

Primer ID	Sequence (5' to 3')	Purpose
1575	GCAATGGGAAGGTAAAAACGTTTCAAGTTCATGTCATGGTTCTTGAATACATCATG	Internal primer for mutating D559 (MHD motif) to V
1229	AAGGTCTCAAATGGAAACGGGCAACAAATATATAG	Forward primer to clone AVR-Pik (for Nt tagging) into pCR™8
1232	TTGGTCTCTAAGCTTAAAAGCCGGGCCTTTTTTTCC	Reverse primer to clone AVR-Pik (for Nt tagging) into pCR™8
1888	AAGAAGACTTCTCAGGAGATCTGGATTTTAGTACTGGATTTTGGTTTTAGGAATTAG	Forward primer for cloning 35Ster:Pik-1 into pUAP1
1889	TTGAAGACAACTCGGATGGAGGCGGCTGCCATGGCCGTAAC	Reverse primer for cloning 35Ster:Pik-1 into pUAP1
1890	AAGAAGACTTCTCAAAGATGGAGTTGGTGGTAGGTGCTTCC	Forward primer for cloning Pik-2:35Ster into pUAP1
1891	TTGAAGACAACTCGAGCGATCTGGATTTTAGTACTGGATTTTGGTTTTAGGAATTAG	Reverse primer for cloning Pik-2:35Ster into pUAP1
1759	AAGGTCTCAGGAGGTCAACATGGTGGAGCAC	Forward primer for cloning 35S:Pik-1 or 35S:Pik2 into a level 1 acceptor

Primer ID	Sequence (5' to 3')	Purpose
1760	AAGGTCTCTAAGCTTAACTAGTAGTTTCTGTTTGAATTTCAATATCTGCTACTCG	Reverse primer for cloning 35S:Pik-1 into a level 1 acceptor
1762	AAGGTCTCTAAGCTTAAGCAGTGACGATGCCATCAACAAATAC	Reverse primer for cloning 35S:Pik-2 into a level 1 acceptor
1412	AGGAGATATACCATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGG	Forward primer for cloning MBP (or MBP tagged protein) into pOPIN-A
1439	GTGATGGTGATGTTTTTATTTCTCTTTAACTTCTTCAACTTGCACCAGTTC	Reverse primer for cloning OsHIPP19-HMA into pOPIN-A
1640	AAGTTCTGTTTCAGGGCCCGGGTCTGAAACAAAAAATCG	Forward primer for cloning Pikp-HMA^{SNKEKE} into pOPIN-M
1641	ATGGTCTAGAAAGCTTTAATCTTCTTTTGCTTGCTCGACTTGCAG	Reverse primer for cloning Pikp-HMA^{SNKEKE} into pOPIN-M