

Interactions of blast fungus effectors with small HMAs and the paired rice NLR Pik

Caroline Eleanor Stone

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

Department of Biochemistry and Metabolism
John Innes Centre

September 2024

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there-from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

Plant pathogens can cause extreme crop losses. The blast fungus, *Magnaporthe oryzae*, is a leading threat to rice production. One significant mechanism of immunity to *M. oryzae* is through intracellular nucleotide binding leucine-rich repeat receptors (NLRs), which can detect intracellular translocated pathogen effectors.

In rice, NLRs can function in pairs - for example, the Pik-1/Pik-2 pair. The Pik-1 “sensor” NLR contains an integrated heavy metal associated (HMA) domain which recognises the *M. oryzae* effector AVR-Pik, and the Pik-2 “helper” NLR is required for signal transduction to produce the immune response. A conserved feature of NLR activation is oligomerisation into an immune complex called the resistosome, however, to date there is no structure of a paired NLR with an integrated domain.

Therefore, the first objective of this thesis was to express and purify the Pik resistosome for structural and functional studies. I present progress in expression and purification of Pik NLRs using insect cell culture, and preliminary work towards understanding the oligomeric status and subcellular localisation of the Pik pair in *Nicotiana benthamiana*.

The second part of my work was to investigate the novel *M. oryzae* effector AVR-Mgk1 and its interactions with HMA domain containing proteins of rice. I present a crystal structure of AVR-Mgk1 with the small HMA OsHPPo2 and used surface plasmon resonance to quantify the interaction. Finally, I attempted a proof-of-concept engineering using the Pik-chassis by exchanging the HMA of Pikm-1 with OsHPPo2.

The molecular characterisation of the interaction of AVR-Mgk1 with HMAs adds to our understanding of the different effector binding modes of HMA domains, ultimately contributing towards efforts to engineer immune receptors with integrated domains. Meanwhile, understanding the physical arrangement of the Pik resistosome before and during activation would aid future immune receptor engineering efforts, which may ultimately contribute to securing future crop health and productivity.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Acknowledgements

Although it feels difficult to believe, four years have passed since I started at the John Innes Centre. As I conclude my PhD work, I would like to offer my thanks to those who worked alongside me, mentored me, and supported me during my time in Norwich.

Foremost, I wish to particularly thank Mark Banfield for the supervision of this PhD work, and all the advice, discussion, and support provided over the years. I am very grateful to Christine Faulkner for helpful and incisive contributions as my secondary supervisor, and for the mentorship from my other supervisory committee members Adam Bentham, Joanna Feehan, and Indira Saado who each added uniquely to my projects.

To the Banfield lab members past and present who are too many to thank by name, thank you for the daily discussions and great lab environment.

Thank you, Yu Sugihara, for your collaboration on the AVR-Mgk1 work, and thank you to the entire BlastOff! Team for the interesting discussions.

I thank Paul Schulze-Lefert and Banfield Lab alumna Nitika Mukhi who hosted my research visit to the MPIPZ in Cologne, Germany in June 2023.

To all of those who contribute to the great research environment at the JIC through their work in Lab Support, Horticultural Services, and Research Platforms, thank you. In particular, from the Chatt building I would like to thank Gary Wortley for his tireless work as Lab Manager, and Sarah Tolland and Naomi Wang of the Departmental Office for logistical and emotional support. I have benefitted greatly from expert scientific contributions from David Lawson, Clare Stephenson, and Julia Mundy of the JIC Structural Biology platform, Abbas Maqbool of the JIC Biophysical Analysis platform, Carlo Martins of the JIC Mass Spectrometry platform and Jan Sklenar of the TSL Proteomics platform – thank you.

I am also grateful to those at Sygnature Discovery, Nottingham, who hosted me for a three-month internship in 2023 and taught me about the application of SPR to fragment-based drug discovery.

Finally, I am deeply thankful to my family, and my friends in Norwich and beyond. I profoundly appreciate the numerous and unique expressions of support I have received over the years.

Table of Contents

Abstract	3
Acknowledgements	5
Table of Contents	7
Table of Figures.....	13
Table of Tables	17
Chapter 1: General Introduction.....	19
1.1 Plant disease: a topic of existential significance to humanity	19
1.2 Blast disease threatens cereal crop production	20
1.3 Pathogens secrete effectors to aid colonisation	22
1.3.1 Effectors can target a range of host cellular components.....	23
1.3.2 Effectors exhibit a variety of protein folds.....	23
1.3.3 Effectors can be predicted by sequence or structural homology	24
1.3.4 Effectors can mutate to overcome host immunity and effector repertoire defines host range.....	25
1.4 Plant immune receptors recognise pathogens and trigger defence responses...	26
1.4.1 Cell surface immune receptors	27
1.4.2 NLRs are intracellular immune receptors	27
1.5 Oligomerisation into a resistosome is a hallmark of NLR activation.....	33
1.6 NLR engineering is an opportunity to introduce new resistances	38
1.6.1 Leucine-rich repeat (LRR) domain engineering.....	39
1.6.2 Integrated domain mutation.....	39
1.6.3 Integration of novel domains.....	41
1.6.4 Other mechanisms	41
1.6.5 Further challenges such as autoimmunity can be overcome.....	42
1.7 Aims and objectives of the thesis	43
Chapter 2: Materials and Methods	45
2.1 DNA methods.....	45
2.1.1 Molecular Cloning.....	45

Table of Contents

2.1.2	Gene synthesis	46
2.1.3	Polymerase Chain Reaction.....	47
2.1.4	DNA gel electrophoresis	47
2.1.5	PCR product clean-up for subsequent cloning steps.....	47
2.1.6	DNA sequencing	48
2.1.7	Sequence curation and alignment	48
2.2	Bacterial cell culture and protein purification	48
2.2.1	Media	48
2.2.2	Bacterial strains	49
2.2.3	<i>Escherichia coli</i> cell culture for protein expression	49
2.2.4	Protein purification from <i>E. coli</i>	50
2.3	<i>Agrobacterium</i> infiltration of <i>N. benthamiana</i> for protein expression	51
2.4	Cell death assays in <i>N. benthamiana</i>	52
2.5	Preparation of <i>N. benthamiana</i> samples for SDS-PAGE and western blotting..	52
2.6	SDS-PAGE, BN-PAGE, and western blot	53
2.6.1	SDS-PAGE	53
2.6.2	Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE)	53
2.6.3	Western Blot	54
Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome.....		55
3.1	Introduction.....	55
3.1.1	Challenges in obtaining resistosome structures.....	55
3.1.2	Full length Pik paired NLRs as a target for structural studies	57
3.1.3	Choice of Pik alleles for resistosome reconstitution.....	57
3.1.4	Mitigation of resistosome mediated cell death in heterologous expression systems	58
3.2	Methods	60
3.2.1	SpyTag/SpyCatcher	60
3.2.2	Expression and purification of the Pik pair in <i>N. benthamiana</i>	61
3.2.3	Insect cell culture.....	62

Table of Contents

3.3	Results.....	67
3.3.1	Expression in <i>N. benthamiana</i>	67
3.3.2	Expression in Insect cell culture	78
3.4	Discussion	95
3.4.1	Multiple approaches to improve expression of Pikm-1/Pikm-2 in <i>N. benthamiana</i> produced limited success.....	95
3.4.2	Sf9 culture was challenging, but Sf21 culture offers hope	95
3.4.3	Alternative structural biology approaches could offer new insights	96
3.4.4	Conclusion.....	98
Chapter 4: Probing subcellular localisation and activation of the Pik-NLR pair		99
4.1	Introduction	99
4.1.1	Subcellular localisation	99
4.1.2	Oligomerisation.....	100
4.1.3	Post-translational modification	101
4.2	Methods	102
4.2.1	Agrobacterium infiltration of <i>N. benthamiana</i> for fluorescence microscopy 102	
4.2.2	Confocal microscopy	102
4.3	Results.....	103
4.3.1	Pikm-1-mCherry, Pikm-2-GFP, and CFP-AVR-PikD can be observed by fluorescence microscopy.....	103
4.3.2	Blue-Native PAGE analysis of the Pik NLR pair.....	110
4.4	Discussion	114
4.4.1	Preliminary observations of the localisation of Pikm-1, Pikm-2, and AVR-PikD were made.....	114
4.4.2	BN-PAGE reveals a high molecular weight Pik complex which accumulates upon recognition of AVR-PikD.....	115
4.4.3	Post-translational modifications of the Pik pair upon activation.....	115
4.4.4	Conclusion.....	117
Chapter 5: Interactions of novel <i>M. oryzae</i> effector AVR-Mgk1		119

Table of Contents

5.1	Introduction.....	119
5.1.1	Forward genetics studies reveal resistance provided by <i>Piks</i>	119
5.1.2	Novel <i>M. oryzae</i> effector AVR-Mgk1 is encoded on a mini-chromosome	120
5.1.3	Small differences in sequence can underlie recognition specificity of <i>Pik</i>	120
5.1.4	AVR-Mgk1 is predicted to be a member of the MAX fold family	122
5.1.5	Advances in protein structure prediction provide increasingly valuable research tools	122
5.2	Methods.....	124
5.2.1	Expression and purification of protein complexes from <i>E. coli</i>	124
5.2.2	Mass spectrometry.....	124
5.2.3	Protein-protein interactions.....	125
5.2.4	Protein structure prediction	126
5.2.5	Macromolecular X-ray crystallography.....	127
5.3	Results	130
5.3.1	Expression and purification of AVR-Mgk1 and <i>Piks</i> -HMA.....	130
5.3.2	Direct interaction of AVR-Mgk1 and <i>Piks</i> -HMA demonstrated by analytical gel filtration.....	131
5.3.3	Attempts at AVR-Mgk1/ <i>Piks</i> -HMA structure determination	132
5.3.4	Further attempts at crystallisation of the <i>Piks</i> -HMA and AVR-Mgk1 complex	140
5.3.5	Expression and purification of OsHPPs	143
5.3.6	Direct interaction of AVR-Mgk1 and OsHPPo2 ^{K77} by analytical gel filtration	148
5.3.7	Expression and purification of the AVR-Mgk1/OsHPPo2 ^{K77} complex for crystallisation	149
5.3.8	Purification and crystallisation of AVR-Mgk1 ^{T81} /OsHPPo2 ^{E74}	153
5.3.9	X-ray data collection and structure determination of the AVR-Mgk1 ^{T81} /OsHPPo2 ^{E74} complex.....	160
5.3.10	Biophysical characterisation of AVR-Mgk1 interactions with different HMAs	165

Table of Contents

5.3.11	Engineering the OsHPPo2/o3/o4 HMAs into the Pik chassis to provide proof-of-concept recognition of AVR-Mgk1	171
5.4	Discussion	174
5.4.1	Crystal structures of Piks-HMA and the OsHPPo2/AVR-Mgk1 complex provide experimental validation for predicted protein models.....	174
5.4.2	Crystal structure of OsHPPo2/AVR-Mgk1 reveals a novel interface between a HMA domain and a MAX effector	174
5.4.3	Structural prediction is a valuable tool in experimental structural biology ..	176
5.4.4	Pikm ^{OsHPP} chimeric receptors provide further insight into engineering the Pik chassis	178
5.4.5	AVR-Mgk variants are found in <i>Magnaporthe</i> isolates infecting diverse grass species	180
5.4.6	Conclusion.....	181
Chapter 6: General Discussion.....		183
6.1	Structural mechanisms of the Pik resistosome remain unresolved	183
6.2	The AVR-Mgk1/OsHPPo2 crystal structure confirmed AVR-Mgk1 as a MAX effector and revealed a new mode of HMA binding.....	185
6.3	Summary and outlook.....	187
Appendix		191
AlphaFold2 outputs.....		191
AVR-Mgk1.....		192
OsHPPo2 full length.....		192
OsHPPo3 full length.....		193
OsHPPo4 full length.....		193
AVR-Mgk1/OsHPPo2 ^{K77} complex		194
References		195
List of Abbreviations		213
Endnote.....		217

Table of Figures

Figure 1-1: Schematic of Magnaporthe infecting a rice cell	21
Figure 1-2: NLRs can detect effectors in different ways	32
Figure 1-3: NLRs can operate as singletons, in pairs, and in networks.....	33
Figure 1-4: Oligomeric resistosome and inflammasome complexes	37
Figure 2-1: <i>Nicotiana benthamiana</i> cell death assay scoring scale.....	52
Figure 3-1: SpyTag/SpyCatcher proof-of-concept purifications from <i>E. coli</i> and <i>N. benthamiana</i>	68
Figure 3-2: Medium scale purification of SpyTag-3C-Myc-AVR-PikD and AVR-PikD-FLAG-3C-Spytag from <i>N. benthamiana</i>	69
Figure 3-3: Expression testing of SpyTagged AVR-PikD in <i>N. benthamiana</i> under control of pUbi and Geminivirus expression system	70
Figure 3-4: Western blot expression testing of the combined Pikm-1-HF and Pikm-2-HA construct CS045.....	71
Figure 3-5: Expression testing of genomic Pikm-1-HF and Pikm-2-HA constructs.	72
Figure 3-6: Expression trial of Pikm-1 and Pikm-2 in <i>N. benthamiana</i> by Leaf Expression Systems	74
Figure 3-7: Second expression trial of Pikm-1 in <i>N. benthamiana</i> by Leaf Expression Systems	76
Figure 3-8: Second expression trial of Pikm-2 in <i>N. benthamiana</i> by Leaf Expression Systems	77
Figure 3-9: Preliminary Sf9 expression trials.....	81
Figure 3-10: Sf9 expression time course.....	82
Figure 3-11: Sf9 expression testing of Pikm-2-FLAG alone and in combination with Pikm-1-FLAG	83
Figure 3-12: Sf9 expression testing of Strep-tagged Pik resistosome components.....	85
Figure 3-13: Small scale affinity purification of Pikm-1-FLAG, Pikm-2-FLAG, and 6His-SUMO-Pikm-1 demonstrating expression of these constructs in Sf21 insect cell culture...	87
Figure 3-14: Purification of “activated” 6His-SUMO-Pikm-1 + Pikm-2 ^{L19E/L23E} -FLAG + AVR-PikD-TwinStrep complex.....	90
Figure 3-15: Purification of “inactivated” 6His-SUMO-Pikm-1 + Pikm-2 ^{L19E/L23E} -FLAG complex	91
Figure 3-16: Negative stain micrograph from “active” Pik resistosome purification.....	92
Figure 3-17: Negative stain micrographs of “inactive” Pik resistosome purification	93

Table of Figures

Figure 4-1: <i>N. benthamiana</i> cell death assay of CFP-AVR-PikD, Pikm-1-mCherry, and Pikm-2-GFP	104
Figure 4-2: Confocal fluorescence microscopy of CFP-AVR-PikD, Pikm-1-mCherry, and Pikm-2-GFP	106
Figure 4-3: The Pikm-2 ^{L15E/L19E/L23E} α 1 mutation abolishes cell death in <i>N. benthamiana</i>	108
Figure 4-4: Confocal fluorescence microscopy of Pikm-1-mCherry, Pikm-2 ^{L15E/L19E/L23E} -GFP, and CFP-AVR-PikD	109
Figure 4-5: BN-PAGE of Pikm-1-FLAG, Pikm-2-Strep, +/- Myc-AVR-PikD	111
Figure 4-6: SDS-PAGE and BN-PAGE of Pikm-1-FLAG + Pikm-2-StrepII +/- Myc-AVR-PikD	112
Figure 5-1: The HMA domains of Pikm-1 and Piks-1 have 98% sequence identity.	121
Figure 5-2: Structural prediction of AVR-Mgk1 produced varied results	123
Figure 5-3: Purification of AVR-Mgk1 from SHuffle <i>E. coli</i>	130
Figure 5-4: Purification of Piks-HMA from SHuffle <i>E. coli</i>	131
Figure 5-5: Analytical gel filtration of AVR-Mgk1 and Piks-HMA.	132
Figure 5-6: Gel filtration chromatograph of AVR-Mgk1/Piks-HMA complex.	133
Figure 5-7: Purification of co-expressed AVR-Mgk1 and Piks-HMA complex.	134
Figure 5-8: Time course of formation and diffraction image of crystal used for collecting Piks-HMA data.	135
Figure 5-9: The crystal structure of Piks-HMA is in good agreement with the Pikm-HMA crystal structure.	137
Figure 5-10: Mass spectrometry indicates that AVR-Mgk1 may be underrepresented in the protein sample.	140
Figure 5-11: Purification of AVR-Mgk1/Piks-HMA complex by mixing.	141
Figure 5-12: Intact mass spectra from AVR-Mgk1/Piks-HMA complex purification.	142
Figure 5-13: Yeast two hybrid assay between AVR-Mgk1 and small HMA library.	144
Figure 5-14: Sequence alignment and structural prediction of full length OsHPPo2, OsHPPo3, OsHPPo4.	145
Figure 5-15: <i>E. coli</i> expression testing of OsHPPo2 ^{K77} , OsHPPo3 ^{K77} , and OsHPPo4 ^{K77} ...	146
Figure 5-16: SDS-PAGE from purification of His-GB1-OsHPPo4 ^{K77}	146
Figure 5-17: Purification of OsHPPo2 ^{K77}	147
Figure 5-18: Analytical gel filtration of AVR-Mgk1 and OsHPPo2 ^{K77}	148
Figure 5-19: Purification of AVR-Mgk1/OsHPPo2 ^{K77} complex.	149
Figure 5-20: Intact mass spectra from AVR-Mgk1/OsHPPo2 ^{K77} complex purification.	150
Figure 5-21: SDS-PAGE of limited trypsin proteolysis of the AVR-Mgk1/OsHPPo2 ^{K77} complex.	152

Table of Figures

Figure 5-22: AlphaFold2 multimer predictions of AVR-Mgk1/OsHPPo2 ^{K77}	153
Figure 5-23: Purification of AVR-Mgk1 ^{T81} /OsHPPo2 ^{E74} complex.	154
Figure 5-24: Intact mass spectra of the purified AVR-Mgk1 ^{T81} /OsHPPo2 ^{E74} complex.	155
Figure 5-25: Crystals obtained in the ShotGun1 screen with the AVR-Mgk1 ^{T81} -OsHPPo2 ^{E74} complex.	157
Figure 5-26: Crystals obtained by seeding into the Morpheus screen with AVR-Mgk1 ^{T81} /OsHPPo2 ^{E74}	158
Figure 5-27: Crystals obtained from the ProPlex screen with AVR-Mgk1 ^{T81} /OsHPPo2 ^{E74}	159
Figure 5-28: AVR-Mgk1 ^{T81} /OsHPPo2 ^{E74} crystal and representative diffraction image	160
Figure 5-29: X-ray crystal structure of the AVR-Mgk1 ^{T81} /OsHPPo2 ^{E74} complex.....	161
Figure 5-30: Details of the AVR-Mgk1 ^{T81} /OsHPPo2 ^{E74} interaction.....	162
Figure 5-31: QtPISA analysis of the AVR-Mgk1 ^{T81} /OsHPPo2 ^{E74} interface.....	163
Figure 5-32: SDS-PAGE of proteins used in SPR experiments	166
Figure 5-33: Example sensorgram from an early multi-cycle kinetics SPR experiment... ..	167
Figure 5-34: Single-cycle kinetics SPR sensorgrams.	169
Figure 5-35: Western blot confirming expression of chimeric Pikm-1 ^{HPP} constructs in <i>N. benthamiana</i>	172
Figure 5-36: Chimeric Pik ^{HPP} NLRs are auto-active in combination with either Pikm-2 or Pikp-2.	173
Figure 5-37: AVR-Mgk1 binds to OsHPPo2 in a novel orientation.....	175
Figure 5-38: MAX effectors target small HMAs via different interfaces.....	176
Figure 5-39: The crystal structure and AlphaFold2 prediction of AVR-Mgk1/OsHPPo2 are in good agreement.....	178

Table of Tables

Table 5-1: Data collection and refinement statistics for the Pks-HMA structure.....	138
Table 5-2: Data collection and refinement statistics for the AVR-Mgk ^{T81} /OsHPPo ₂ ^{E74} complex structure	164
Table 5-3: Summary of K _D determined by SPR	170

Chapter 1: General Introduction

In this introduction, I will first provide an overview of the field of plant immunity, focussing on the protein machinery of intracellular immunity and approaches to engineer it for improved plant immunity, before outlining the aims and objectives of the work presented in this thesis.

1.1 Plant disease: a topic of existential significance to humanity

When surveying the natural environment, it is clear that most plants are immune to most pathogens. However, monoculture of significant agricultural crops including cereals is perpetually threatened by disease caused by plant pathogens. Plant disease negatively impacts crop yields and therefore negatively impacts on both local subsistence farmers and on global food security (Yan and Talbot, 2016). Diverse pathogens and pests such as bacteria, oomycetes, viruses, fungi, nematodes, and insects can all infect plants. A significant proportion of cereal crop disease is caused by fungal pathogens such as rusts, powdery mildew, various pathogenic members of the *Fusarium* genus which cause diseases such as Fusarium wilt and Fusarium head blight, and *Magnaporthe* species, the causative agent of blast disease.

Within the context of a growing global population, and global climate change, it is projected that by 2050 we need to increase world food production by at least 60% using the same amount of land (Esse et al., 2020). One key factor in how this could be achieved is through reduction of waste across the food cycle from harvest, storage, and delivery to consumers. However, reduction of post-harvest waste alone will not secure a sufficient food supply; pre-harvest yield loss caused by pests and pathogens is extreme and is exacerbated by the continual evolution of new strains of pathogens that can evade detection by the plant immune system or develop pesticide resistance.

To control fungal pathogens increasing quantities of fungicides are deployed. In 2019 around 350,000 tonnes of fungicides were used globally and this is projected to increase as pesticide use in developing regions of the world continues to rise (Sharma et al., 2019). However, such use of agrochemicals raises concerns about environmental impact, long term sustainability, and remains unaffordable in many regions. To meet the ongoing challenges of crop production without reliance on fungicides in the context of global climate change,

genetic improvements to crops are required (Esse et al., 2020). Genetic modification (GM) refers to the practice of manipulating the genetic material of an organism, including introduction of transgenes from other organisms. A distinction has been made between this and precision breeding, which refers to the introduction of targeted genetic changes which could have been produced through traditional breeding or natural processes. While adoption of GM approaches in agriculture remains hotly debated in certain regions, public and legislative opinion is changing. For example, last year England introduced the Genetic Technology (Precision Breeding) Act 2023, which introduces a new regulatory approach. Plants produced through precision gene editing approaches are treated “proportionately to risk”, which provides a regulatory distinction from plants produced by GM approaches. Irrespective of this recent change, molecular knowledge of plant immune mechanisms has informed and will continue to inform both targeted genetic changes and conventional breeding strategies. These tip the balance in favour of crop health and productivity in the ongoing arms race between crop plant and pathogen. Accordingly, the field of molecular plant-microbe interactions remains a priority within plant sciences and has grown into a productive area of research in the international plant pathology community.

1.2 Blast disease threatens cereal crop production

Rice, *Oryza sativa*, is a staple food for more than half of the global population, with consumption highest in Asia and Africa. *Magnaporthe oryzae* (Syn. *Pyricularia oryzae*) is a hemibiotrophic filamentous fungal pathogen of rice which causes recurrent epidemics of rice blast disease wherever rice is cultivated, threatening individual farmer livelihoods and regional food security (Wang et al., 2014; Yan and Talbot, 2016).

Due to its economic significance and experimental tractability, *M. oryzae* has been adopted as a molecular model for plant pathogenic fungi (Liu et al., 2014). The infection cycle observed in the field is asexual and has been described in many reviews (Wilson and Talbot, 2009; Saleh et al., 2012a; Yan and Talbot, 2016). Briefly, upon landing on a host leaf surface, the three-celled, teardrop shaped conidium (asexual spore) germinates and forms an appressorium structure, generating high turgor pressure to puncture the leaf cuticle (Wilson and Talbot, 2009). Upon entering the host cell, a filamentous primary invasive hypha forms, and from this, invasive hyphae branch out to create a hyphal network (Wilson and Talbot, 2009). The fungal hyphae form extensive interfaces with the host cell along the extra-invasive hyphal membrane, which is formed by invagination of the host cell membrane (Yan and Talbot, 2016). At the tip of the primary invasive hypha a membrane rich structure derived from the plant membrane develops into a structure called the

biotrophic interface complex (BIC). As the invasive hyphae progress to invade the rest of the cell, the BIC is “left behind” near the primary invasion site (Giraldo et al., 2013; Shipman et al., 2017) (Figure 1-1).

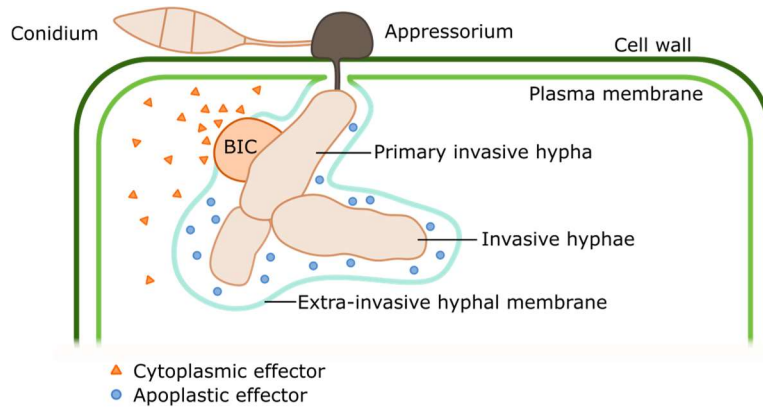


Figure 1-1: Schematic of *Magnaporthe* infecting a rice cell

Cytoplasmic effectors (orange triangles) are preferentially secreted from the biotrophic interface complex (BIC, depicted in orange). Extracellular effectors (blue circles) are secreted from the invasive hyphae. Extra-invasive hyphal membrane (blue) is continuous with the plant plasma membrane (green).

The pit field sites which connect plant cells are exploited by a specialised fungal structure called the transpressorium, allowing the hyphae to traverse between cells (Wilson and Talbot, 2009). As the hyphal network matures, specialised hyphae called conidiophores develop on the leaf surface and produce conidia. Finally, sporulation occurs in high humidity; the mature conidia are dispersed by wind or rain splashes, and the cycle is repeated (Yan and Talbot, 2016).

Multiple factors are implicated in the severity of rice blast disease. Climate factors such as temperature and humidity, the stage of development at which rice is first infected, and what treatments and strategies are available to mitigate infection all contribute to the degree of yield loss. In developed countries, some control of rice blast disease is possible thorough use of resistant rice cultivars and fungicides, however in developing countries these measures are frequently cost prohibitive and so the impacts of rice blast disease are correspondingly more severe. In these cases, crop yield can be reduced by up to 100% in the most acutely infected fields (Wang et al., 2014). Based on data from the 2009-2010 rice harvest and an estimated 10-35% crop loss due to *M. oryzae*, it has been estimated that annual loss of rice yield could have otherwise fed 212 – 742 million people for a year (Savary et al., 2019).

It is also important to note that *M. oryzae* is not limited to rice as its host; different pathotypes are able to infect other grasses. Wheat blast disease, caused by *M. oryzae* pathotype *triticum* (MoT) emerged in Brazil in 1985 (Igarashi, 1986). MoT spread to other South American countries such as Bolivia, Paraguay, and Argentina, and has been observed on two further continents, Asia and Africa, since the mid-2010s. A widespread outbreak of wheat blast disease occurred in Bangladesh in 2016, originating from grain imported from South America (Islam et al., 2016). During this outbreak, an average yield loss of 51 % was experienced in the most impacted region, although some fields experienced up to 100 % loss (Islam et al., 2016). MoT remains pandemic in Bangladesh and continues to threaten wheat production in tropical and sub-tropical regions (Yan and Talbot, 2016). In 2018, wheat blast disease was first observed in Zambia (Tembo et al., 2020). Genetic analysis of the isolates from Zambia and Bangladesh indicated that the outbreaks were caused by the same clonal lineage of MoT, but had occurred independently (Latorre et al., 2023).

1.3 Pathogens secrete effectors to aid colonisation

Plant pathogens including bacteria, fungi, oomycetes and invertebrates secrete effectors to target plant immune processes and manipulate host cell physiology to be more conducive to colonisation (Jones et al., 2016; Zhou and Zhang, 2020). Effectors which are perceived by the plant and activate effector-triggered-immunity were initially defined as avirulence proteins (AVR), nomenclature which can be seen in many effectors discussed in this thesis.

Effectors are typically proteinaceous, although small RNA and small molecule metabolite effectors are also deployed by some pathogens (Wang et al., 2017; Collemare et al., 2019). Fungi are known to produce extracellular vesicles, known to be vehicles for sRNA, used to carry out cross-kingdom RNAi. This relationship is bidirectional - release of extracellular vesicles originating from *Arabidopsis* tissue has been observed following infection with fungal pathogen *Botrytis cinerea*, in addition to detection of signatures of RNAi against *Botrytis* pathogenicity genes (Cai et al., 2018; Cai et al., 2019). However, in this introduction, discussion will remain focussed on proteinaceous effectors, particularly those of fungal phytopathogens.

1.3.1 Effectors can target a range of host cellular components

Different effector families target diverse proteins and cellular compartments to achieve pathogenicity. Apoplastic effectors act on targets outside the host cell, whereas cytosolic effectors must cross the plant cell plasma membrane to reach their targets. Effectors which act in the apoplast are secreted from invasive hyphae, while cytoplasmic effectors are secreted preferentially from the BIC, which possesses specialised secretion machinery (Shipman et al., 2017). In general, fungal effectors possess N-terminal signal peptides which facilitate their secretion from the fungal cell. Upon secretion, the signal peptide is cleaved by proteases present in the apoplast; this should be taken into account when designing constructs for heterologous expression, for example by agroinfiltration.

Apoplastic effectors include chitin binding LysM effectors, which sequester chitin oligomers to prevent their detection by plant extracellular PRRs that monitor the apoplast (Sánchez-Vallet et al., 2015). All cellular compartments are targeted by effectors; including cytosol, nucleus, plasma membrane, chloroplasts, mitochondria, and endoplasmic reticulum, as reviewed by Figueroa et al. (Figueroa et al., 2021). Challenges in determining the function of an effector can be attributed to functional redundancy; knocking out a single effector gene is unlikely to result in a significant pathogenic phenotype (Giraldo and Valent, 2013). Multiple effectors often target similar host functions, maintaining virulence of the pathogen even if one effector is lost. The virulence functions of many effectors remain unknown, however, one novel method to discover host virulence targets was recently described in a pre-print (Haley et al., 2024). In brief, effector protein structures were used as a template and small binding motifs were generated *de novo* using generative protein design tool RFDiffusion (Watson et al., 2023). The plant was then queried for these predicted structural elements using a Foldseek structural search, the results of which can reveal predicted effector binders, which may be virulence targets (van Kempen et al., 2024).

1.3.2 Effectors exhibit a variety of protein folds

Phytopathogen effectors have been classified into several families defined by shared protein folds. An example is the RXLR effector family from oomycetes, which exhibit conservation of a “WY-domain fold”. The WY-fold is composed of three or four short α -helices connected by variable loop regions, despite being encoded by sequences with sparse similarity (Win et al., 2012).

To date, two groups of *Magnaporthe* effectors have been most studied: the *Magnaporthe* AVRs and ToxB-like (MAX) effectors such as AVR-Pia, AVR-Pib, AVR-Pik, and the Zinc

Finger Fold effectors (ZiFs), such as AVR-Pii (De la Concepcion et al., 2024). MAX effectors have been particularly well studied and adopt a β -sandwich fold, normally stabilised by at least one disulfide bond, and containing at least two cystines that are separated by around 40 amino acids (de Guillen et al., 2015). Structures determined to date exhibit variable length of the β -strands and loops, and distinct polymorphic residues on the surface which confer different binding activities while maintaining the same core structure as a chassis (De La Concepcion et al., 2018; Guo et al., 2018; Zhang et al., 2018; Varden et al., 2019; Maidment et al., 2021; Zdrzałek et al., 2024). The virulence activity of most MAX effectors is yet to be determined, but a subset are known to target host small HMA (sHMA) proteins, discussed further in section 1.4.2.2.

1.3.3 Effectors can be predicted by sequence or structural homology

The presence of the N-terminal signal peptide in secreted effectors is a feature that allows for effector prediction through analysis of pathogen genomes. One such example is the oomycete RXLR effectors previously mentioned (Whisson et al., 2007; Win et al., 2012). The tool EffectorP is widely used for this task, and predicts effector proteins from fungal secretomes using machine learning (Sperschneider et al., 2016). The latest iteration of EffectorP can also predict oomycete effectors, and can distinguish between apoplastic and cytoplasmic effectors (Sperschneider and Dodds, 2022).

Rapid mutation of fungal effector proteins means that their sequences can diverge over short periods of evolutionary time, obscuring relationships between them. However, structural folds may be conserved, so putative effectors can be grouped by known or predicted structural features. Structural prediction based on sequence homology is not possible in many cases when comparing sequences with no detectable homology to existing effector structures, which is a limitation of this methodology. The advent of structural prediction large language models (LLMs) has increased the proportion of effectors which can be modelled, with the caveat that some novel folds may be harder to predict and so may still be less accurate and under-represented in the predicted effector repertoire of any given species (Seong and Krasileva, 2021, 2023). Use of trRosetta and AlphaFold2 facilitated discovery and classification of sequence-unrelated structurally similar effector families. Additionally, this work drew attention to the Tin2 fold effectors of *Ustilago maydis* which contain disordered stretches, which are known to mutate at a greater rate and so are hypothesised to compromise plant immunity by shielding the core effector fold from the plant immune system (Seong and Krasileva, 2023).

1.3.4 Effectors can mutate to overcome host immunity and effector repertoire defines host range

Effectors can mutate rapidly to overcome host immune recognition. These changes in effector repertoire impact both the ability of the pathogen to infect, as well as the capacity of the host to detect pathogen invasion. In *Magnaporthe*, the field infection cycle is asexual, therefore recombination is not a typical mechanism of genetic variation in *Magnaporthe* isolates found globally in rice growing regions. Although genetic signatures of sexual reproduction have been observed in south-east Asia, there is also evidence that female-sterile isolates can arise within 10 – 19 generations *in vitro*, implying that in asexually reproducing populations of *Magnaporthe*, reversion to sexual reproduction may not be possible (Saleh et al., 2012a; Saleh et al., 2012b). Nonetheless, in *Magnaporthe* effector gene nucleotide polymorphisms are common and effector genes are rapidly gained and lost, associated with activity of transposable elements (Yoshida et al., 2016). The presence of mini-chromosomes has been correlated with virulence in multiple pathogen-host systems, and changes in virulence can be partially attributed to horizontal transfer of mini-chromosomes (Barragan et al., 2024). The mini-chromosomes of *Magnaporthe* carry signatures of a greater rate of major rearrangement than the core genome, and also contain effector and virulence related loci which encode some MAX effectors such as AVR-PikD and AVR-PikA (Langner et al., 2021). The genome of any given strain of a pathogen will encode many effector proteins; transcriptional profiling of rice plants during *M. oryzae* infection resulted in detection of 546 genes encoding secreted proteins predicted to be effectors (Yan et al., 2023). AVR-Pik effectors are highly diversified due to their co-evolutionary history with the receptors of the rice immune system that recognise them (Kanzaki et al., 2012).

Possession of different effector repertoires between pathogen isolates can account for their host specificity. The host jump that *M. oryzae* made to infect wheat in the 1980s, as described in the previous section, is one such example of effector repertoire of a pathogen dictating host compatibility. But how did *Magnaporthe* overcome the resistance of wheat in Brazil? Wheat cultivars grown in Brazil in the 1980s included those with the *rw3* genotype, which were grown in fields next to varieties with a functional *Rwt3* resistance gene. *Magnaporthe oryzae* *Lolium* isolates with the genotype *PWT3:pwt4* were able to infect *rw3* wheat cultivars, and through random mutations, *PWT3* function was lost in these strains, facilitating their infection of the nearby wheat varieties which still possessed functional *Rwt3*. Wheat cultivars such as Hope, with a *rw3/rwt4* genotype, are vulnerable to *M. oryzae* *Avena* isolates (from oat) which possess both *PWT3* and *PWT4* (Inoue et al., 2017). Additionally, the resistance gene *Rwt6* (*Rmg9*) which corresponds to *Magnaporthe* *PWT6* was determined to be genetically linked to *Rwt3* (Asuke et al., 2021). In contrast to

the two avirulence genes which govern the compatibility of *Lolium* isolates of *Magnaporthe* with wheat cultivars such as Hope, in “standard” cultivars Norin 4 and Chinese Spring, it has been determined that at least five avirulence genes are involved in their incompatibility with *Magnaporthe oryzae Elucine* isolates. As sexual reproduction of *Magnaporthe* has not been observed outside of southeast Asia, the likelihood of *Elucine* isolates evolving to infect Norin 4 or Chinese Spring wheat is low, as it would require mutation of five genes (Saleh et al., 2012a; Saleh et al., 2012b; Asuke et al., 2020).

Viewed from the other side of the evolutionary arms-race, in response to the proliferation of different host adapted pathogens, plants developed diverse intracellular immune receptors, nucleotide binding leucine rich repeat receptors (NLRs); these are discussed in the next section.

1.4 Plant immune receptors recognise pathogens and trigger defence responses

Plants exhibit robust immune defences, achieved without the adaptive immune system of animals or mobile immune cells of chordates. This means that all pathogen detection and defence functions are genetically encoded and are carried out by the innate immune system of each cell. Plants are in contact with a range of microbes with different lifestyles; some are pathogenic, but many are neutral or beneficial, either simply co-occurring in the environment or as symbiotic partners. Therefore, plant immunity must be triggered appropriately - only against pests and pathogens - which requires complex regulation and detection capabilities.

Plant resistance genes (R genes) have been selected for by plant breeders throughout history, and the concept was formalised by Flor, who proposed the gene for gene model. This model states that for a given plant gene that confers resistance to a given pathogen, there will be a corresponding gene in that pathogen that confers a specific pathogenic activity (Flor, 1971). The first R gene to be cloned and published was *Hm1* from *Zea mays* (Johal and Briggs, 1992). In the three decades since, the majority (61%) of cloned R genes have been found to be nucleotide binding, leucine rich repeat receptors (NLRs), but tandem kinases are increasingly demonstrated to function in intracellular immunity (Kourelis and Van Der Hoorn, 2018; Fahima et al., 2024).

1.4.1 Cell surface immune receptors

During pathogen invasion into the plant apoplast, pathogens shed distinctive molecular signatures, known as microbe- or pathogen- associated molecular patterns (MAMPs or PAMPs). MAMPs are typically signatures of molecules essential to microbes, and are therefore common to broad categories of pathogen such as chito-oligosaccharides of fungal cell walls (Mélida et al., 2018), or the flg22 peptide of bacterial flagellin (Robatzek, 2006). Their release into the plant cannot be avoided, and the plant immune system exploits this through detection of PAMPs via plasma membrane localised Pattern Recognition Receptors (PRRs) which monitor the extracellular space. Damage Associated Molecular Patterns (DAMPs) such as oligogalacturonides from damaged plant cell walls are generated upon pathogen attack and can also be recognised by PRRs. Upon recognition of a MAMP/DAMP, pattern-triggered immunity (PTI) is activated. The PRR extracellular receptor domain transduces a signal to the intracellular kinase domain, which initiates signalling pathways such as mitogen activated protein kinase (MAPK) cascades, resulting a variety of defensive responses to reduce further pathogen invasion (Asai et al., 2002). Examples of PTI responses triggered by PRRs include stomata and plasmodesmata closure, reactive oxygen species (ROS) production by NADPH oxidases, cell wall thickening and callose deposition, ion fluxes across the plasma membrane, and transcriptional responses, for example the upregulation of pathogenesis-related (PR) proteins, and are covered by many excellent reviews (Liang and Zhou, 2018; DeFalco and Zipfel, 2021; Dodds et al., 2024). PAMPs are highly conserved and indispensable features of the pathogen and therefore PTI represents a very broad and conserved immune response of the plant.

1.4.2 NLRs are intracellular immune receptors

As introduced in (1.3), pathogens deliver effectors into the plant cell during infection to overcome PTI and promote pathogenesis. Intracellular immune receptors are responsible for monitoring the interior of the plant cell and perceiving pathogen effectors directly or indirectly. Most characterised intracellular receptors are NLR (Nucleotide binding, leucine rich repeat) proteins. Upon recognition of an effector, the receptor is activated, resulting in localised programmed cell death characterised as the Hypersensitive Response (HR), which limits further pathogen spread through neighbouring cells and tissues. This is known as effector triggered immunity (ETI) and is often referred to as the second line of defence in plant immunity. Initially PTI and ETI were considered as separate mechanisms, but there is a growing body of evidence about the interplay between the two systems. This interdependence between the two 'layers' of plant immunity increases the robustness of the

immune response as a whole (Hatsugai et al., 2017; Ngou et al., 2021; Yuan et al., 2021a; Yuan et al., 2021b). Plant genomes contain a great diversity of NLR proteins – for example in rice, Sarris et al. detected 438 NLR gene sequences (Sarris et al., 2016). This proliferation of NLRs evolved in response to host adapted pathogens which each carry a diverse range of effectors (Kanzaki et al., 2012).

1.4.2.1 NLRs have a canonical three-domain architecture

Canonical NLR domain architecture typically consists of three domains: an N-terminal signalling domain, a central nucleotide-binding module, nucleotide-binding Apaf-1, R protein, CED4-shared domain (NB-ARC), and a C-terminal leucine rich repeat domain (LRR). Plant NLRs resemble mammalian nod-like receptors (NLRs); the use of the acronym of NLR for plant NB-LRR proteins came into favour to emphasise the similarities between them, however these are thought to have emerged through convergent evolution (Yue et al., 2012; Urbach and Ausubel, 2017). In this thesis, NLR will refer to plant NLRs, unless specified otherwise.

The N-terminal NLR domains are involved in downstream signalling to cause cell death following NLR activation. In plants, NLRs can typically be classified based on the identity of the N-terminal domain - a coiled-coil (CC) indicates a CC-NLR (or CNL), or a toll/interleukin-1 receptor (TIR) domain indicates a TIR-NLR (or TNL). A subset of CC domains are designated as RPW8-like CC (CC_{RPW8} or CC_R) due to their similarity with Resistance to Powdery Mildew 8 (RPW8); the NLRs containing these are RNLs (Shao et al., 2016).

CC domains can self-associate, or associate with CC domains of other NLRs in the paired NLR context; an example of this is the association of RGA4 and RGA5 via their CC domains to form homo-and hetero-complexes (Cesari et al., 2014). When expressed alone, the CC domain can cause cell death in some cases. Interrogation of CC-NLR sequences has revealed a N terminal MADA motif (consensus sequence MADAxVSFxVxKLxxLLxxEx) and a minimal region of 29 amino acids that has been implicated in cell death (Adachi et al., 2019). Although only around 20 % of monocot and dicot CC-NLRs contain a MADA motif, the rice Pik-2 NLR is included within this minority (Adachi et al., 2019). In the context of oligomerisation and activation of the full length CC-NLR ZAR1, the α 1 helix in particular has been implicated in plasma membrane cation channel formation, discussed further in section 1.5 (Wang et al., 2019a; Bi et al., 2021). Approximately 38 % of Arabidopsis CC-NLRs also contain an “EDVID motif” (Glu-Asp-Val-Ile-Asp) within the CC domain, a motif which is shared with NLRs including ZAR1 and Sr35 (Wróblewski et al., 2018).

TIR domains have been shown to self-associate (Bernoux et al., 2011; Schreiber et al., 2016; Zhang et al., 2017). It was then found that upon oligomerisation, TIR domains form a holoenzyme with nucleotide hydrolase activity, cleaving NAD⁺ to promote cell death via EDS1/PAD4 or EDS1/SAG101 mediated interactions with helper NLRs (Wan et al., 2019; Ma et al., 2020; Martin et al., 2020). Later, it was revealed that TIR-NLR resistosomes also possess ADP-ribosylation activity which provides specific activation of the EDS1/SAG101/NRG1 branch of downstream responses (Jia et al., 2022). However, TIR-NLRs are almost entirely absent from monocot genomes - including rice - although present in gymnosperms and eudicots, suggesting evolutionary loss from the monocot lineage (Tarr and Alexander, 2009).

The NB-ARC (nucleotide-binding adaptor shared by Apaf-1, R proteins, and Ced-4) domain is the most conserved domain, and is a member of the signal transduction ATPases with numerous domains (STAND) superfamily of ATPases. The NB-ARC functions as an “off-on” switch, facilitating resistosome activation upon exchange of ADP for ATP. Nucleotide binding capability is provided by the P loop, GxP, and MHD motifs. ADP binding typically promotes a closed, inactive conformation, whilst binding to ATP promotes an open, active, conformation (Tameling et al., 2006; Steele et al., 2019). ZAR1 was reconstituted in its active state in the presence of dATP, which was not observed to interact with the MHD motif, in contrast to the MHD motif hydrogen bonding to the ADP in the inactivate structure (Wang et al., 2019a). It is thought that ATP hydrolysis can drive the resetting of the NLR into the inactive state, but this is not the case in all NLRs; RPP1 lacks ATP binding motifs and cryo-EM density showed it was unambiguously ADP-bound in its active, oligomerised resistosome conformation (Ma et al., 2020).

Finally, the C-terminal LRR domain is characterised by repeating units of a LxxLxLxxNxL motif, which forms a beta strand, interspersed with variable regions. The repeats assemble into large beta strand-turn-alpha helix folds which form a gently curving ‘horseshoe’ solenoid with a beta sheet extending over the concave side. These provide a large surface area for effector and other interactions, and variability in repeat number provides a variety of LRR lengths. Through autoinhibitory binding to the NB-ARC domain, the LRR has been shown to play a role in stabilising NLRs in their inactive state (Padmanabhan et al., 2009).

1.4.2.2 NLRs can recognise effectors in multiple ways

There are four known modes of effector recognition by NLRs – direct recognition by the LRR, indirect recognition via a host target, indirect recognition via a decoy, and recognition via integrated domain. Direct recognition of a pathogen effector by the NLR occurs in many

cases; often via the LRR domain, or an integrated domain. Just a few examples are the recognition of AVR-Pita of *M. oryzae* by rice NLR Pi-ta (Jia et al., 2000), the recognition of bacterial effector XopQ via the LRR domain of *N. benthamiana* NLR ROQ1 (Martin et al., 2020), the recognition of the stem rust fungal effector AvrSr35 via the LRR of wheat NLR Sr35 (Förderer et al., 2022; Zhao et al., 2022), and finally the recognition of barley powdery mildew effector AVR_{A13} via the LRR and central WHD motif of barley NLR MLA13 (Figure 1-2 A).

In contrast to direct recognition, in other cases NLRs continually monitor the status of a host virulence target, and detect modification resulting from interaction with the effector. This is a form of indirect effector detection known as the “guardee” model and is a way of detecting “modified-self” (Cesari, 2018) (Figure 1-2 B). A variation on the guardee model can occur in which paralogues or alleles of targeted host components emerge which are no longer indispensable in host cell function but are instead mimics, or “decoys”, of host targets. This is known as the decoy model (Cesari et al., 2014). Effector recognition via decoys should be less evolutionarily constrained, as the decoy does not retain the original protein function, allowing for a greater degree of co-evolution with pathogen effectors as they mutate over time (van der Hoorn and Kamoun, 2008). An example of the decoy model is the *Arabidopsis* NLR ZAR1 which can detect the activity of a range of bacterial effectors through its association with a suite of ZRK pseudokinases and PBL kinases (Wang et al., 2015). As a consequence, *Pseudomonas syringae* effectors such as HopZ1a, HopF1, HopX1, HopO1, and HopBA1, *Xanthomonas perforans* effector XopJ4, and *Xanthomonas campestris campestris* effector AvrAC can all be detected by a single NLR. To take one example, ZAR1, in complex with pseudokinase ZRK1, is able to recruit kinase PBL2 only upon uridylation of PBL2 by effector AvrAC. PBL2 qualifies as a decoy, as its presence does not enhance the virulence function of AvrAC (Wang et al., 2015).

In an extension to the decoy model, first proposed by Cesari et al. in 2014, the decoy protein can become incorporated into the NLR as an extra domain (Cesari et al., 2014). Initially called the integrated decoy model, the contemporary term “integrated domain” (ID) encompasses non-canonical domains of NLRs irrespective of the function of the domain beyond its incorporation in an NLR. Integrated domains are widespread, comprising an average of 10 % of the NLRs of any given species, and examples of IDs can be found across both TIR- and CC-NLRs, in singleton, paired, and networked NLRs. (Kroj et al., 2016). Approximately 20 % of angiosperm CC-NLRs contain IDs, often located at the C-terminus (Contreras et al., 2023a). In rice specifically, Sarris et al. determined that out of 438 NLRs, 22 are predicted to contain an integrated “non-canonical” domain (Sarris et al., 2016).

Some examples of IDs include the C-terminally integrated WRKY domain of the *Arabidopsis* TIR-NLR RRS1, which mediates recognition of two different effectors, the *Pseudomonas syringae* pv. *pti* effector AvrRsp4, and the *Solanum americanum* effector PopP2 (Mukhi et al., 2021). Other IDs include the C-terminally integrated Heavy Metal Associated (HMA) domain of the rice CC-NLR RGA5, and BED, thioredoxin, or kinase domains (Grund et al., 2018). The HMA domain adopts an α - β sandwich fold made up of an antiparallel β -sheet and two α -helices which is found across prokaryotes and eukaryotes (Gitschier et al., 1998). The rice sensor CC-NLRs Pik-1 and Pia-2 also contain integrated HMA domains located between the CC- and NB-ARC domains. The location of these integrated HMA domains demonstrates that the modular nature of NLRs can, in some cases, accommodate incorporation of novel domains at positions other than at the termini. However, the jelly-roll/Ig-like domain (C-JID) found at the C-termini of ROQ1 and RPP1 is classified as a post LRR domain rather than an integrated domain (Ma et al., 2020; Martin et al., 2020).

Some integrated domains correspond to host proteins with unknown activities. By investigating these host proteins, the virulence targets of pathogen effectors can be better understood. One well studied class of virulence targets are the small HMA proteins (sHMA), which are HMAs that are not integrated into NLRs but can nonetheless be bound by *Magnaporthe* effectors. Examples of sHMAs are heavy metal associated isoprenylated plant proteins (HIPPs) and heavy metal associated plant proteins (HPPs) (De Abreu-Neto et al., 2013). These are found in vascular plants in expanded families - around 45 HIPPs and 22 HPPs have been identified in *Arabidopsis* (De Abreu-Neto et al., 2013). Each HIPP or HPP contains one or more HMA domains. At their C-terminus, HIPPs carry an isoprenylation motif, CaaX (a = aliphatic amino acid, X = any amino acid) whereas HPPs lack this motif. HIPPs have been implicated in heavy metal tolerance, transcriptional responses to cold and drought, in addition to roles in plant-pathogen interactions, and expression patterns vary across tissues and developmental stage of the plant (De Abreu-Neto et al., 2013). Knowledge of sHMA interactions with effectors can feed forward into engineering approaches; for example the crystal structure of AVR-PikF binding to OsHIPP19 led to the modification of the Pikp-1 interface to include OsHIPP19 binding capabilities (Maidment et al., 2023). This resulted in an engineered receptor capable of responding to the previously unrecognised AVR-PikF allele.

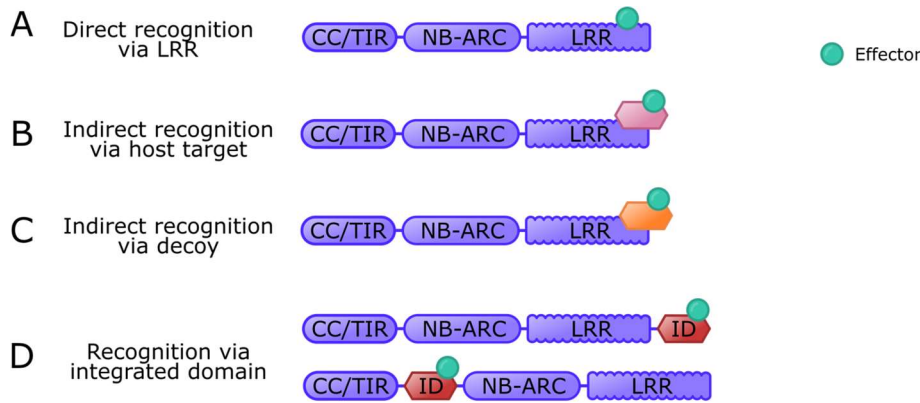


Figure 1-2: NLRs can detect effectors in different ways

A) Direct recognition, typically via the LRR, **B)** indirect recognition via host targets, and **C)** via a decoy, **D)** Recognition via integrated domain at different positions. Domain abbreviations as follows: CC = coiled coil, CC_R = RPW8-like CC, TIR = toll/interleukin-1 receptor, NB-ARC = nucleotide binding adaptor shared by Apaf-1, R proteins, and Ced-4, LRR = leucine rich repeat, ID = integrated domain. Effector depicted as a turquoise circle.

1.4.2.3 NLRs can operate as singletons, in pairs, or in networks

NLRs operate in various configurations, ranging from those which operate alone, to those which are part of large networks (Wu et al., 2017; Bentham et al., 2020). NLRs which operate as a single genetic unit are known as singleton NLRs; examples include ZAR1 and ROQ1 (Figure 1-3 A) (Baudin et al., 2017; Martin et al., 2020). NLRs which function as a pair are termed paired NLRs, and are often tightly genetically linked and are expressed from a shared promoter. One NLR is responsible for detection of the effector and is termed the “sensor” NLR, while the second NLR is required for producing the signalling to produce cell death and is known as the “helper” or “executor” NLR (Figure 1-3 B) (Feehan et al., 2020). Examples of paired NLRs include the *Arabidopsis* RPS4/RRS1, and rice RGA5/RGA4 and Pik-1/Pik-2. Based on insights from the rice NLR pair RGA5/RGA4 and the *Arabidopsis* pair RPS4/RRS1, a mechanism of negative regulation was suggested, whereby the helper NLR (RRS1 and RGA4) expressed alone leads to autoactive cell death and co-expression of the sensor with the helper provides suppression of autoactivity (Williams et al., 2014). In these two paired NLRs, an intact p-loop is required only in the helper NLR to produce cell death in planta (Cesari et al., 2014; Williams et al., 2014). Effector perception by the sensor NLR relieves inhibition of the helper NLR, allowing activation and cell death. However, the rice NLR pair Pik-1/Pik-2 differs in its regulation and appears to follow a cooperative regulation mechanism, as co-expression of all three proteins is required for cell death in *N. benthamiana* (Zdrzałek et al., 2020). Additionally, co-immunoprecipitation experiments

demonstrated that both NLRs and the effector can form a tripartite complex, and that a functional p-loop motif is required in both Pikip-1/Pikip-2 to produce in planta cell death (Zdrzałek et al., 2020). Adding further complexity, in the Solanaceae, there is an expanded repertoire of sensor and helper NLRs which are not genetically linked and operate in a network-like manner (Wu et al., 2017) (Figure 1-3 C). These are the NRCs (NLR-required for cell death) and function in many-to-one and one-to-many sensor and helper configurations, exemplified by the convergence of signalling from many TIR- and CC-NLRs on the helper CC_R-NLR NRG1 (Wu et al., 2017; Castel et al., 2019).

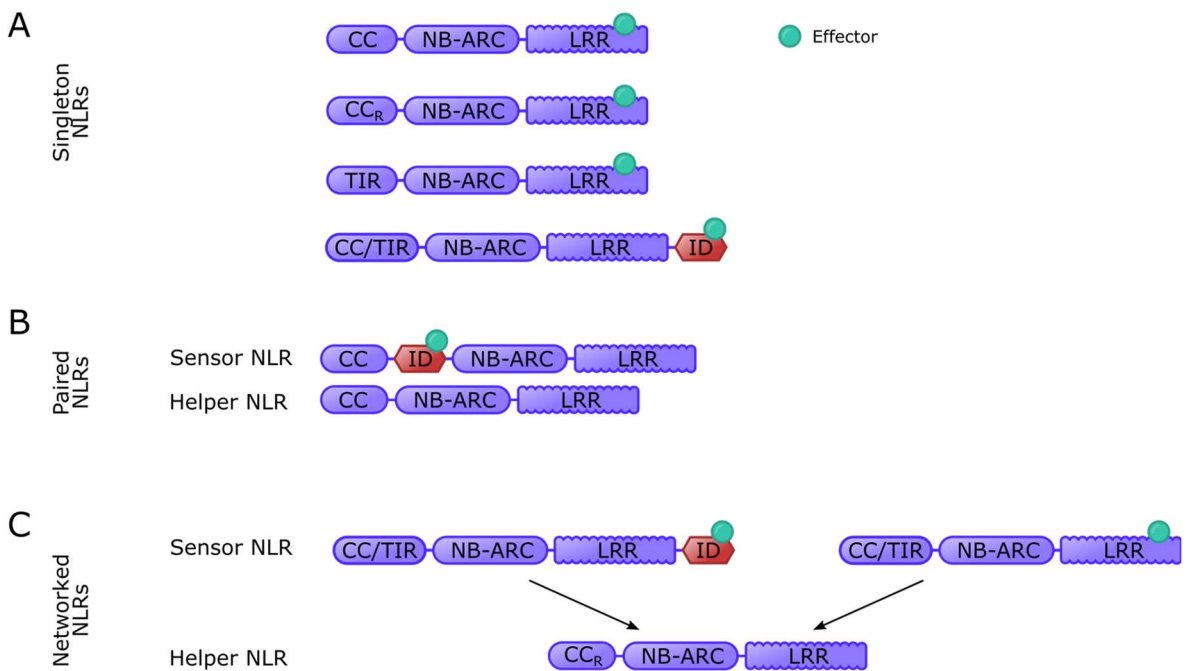


Figure 1-3: NLRs can operate as singletons, in pairs, and in networks

A) Singleton, B) paired, and C) networked NLRs. Domain abbreviations as follows: CC = coiled coil, CC_R = RPW8-like CC, TIR = toll/interleukin-1 receptor, NB-ARC = nucleotide binding adaptor shared by Apaf-1, R proteins, and Ced-4, LRR = leucine rich repeat, ID = integrated domain. Effector depicted as a turquoise circle.

1.5 Oligomerisation into a resistosome is a hallmark of NLR activation

Metazoan NLRs have long been known to function with the immune system by associating into wheel like oligomers, termed “inflammasomes” or “apoptosomes”, upon activation by immunity triggering PAMPs/DAMPs. The oligomerised NLRs form a platform for recruitment of pro-inflammatory caspases, which in turn proteolytically activate cytokines and the pore forming protein gasdermin D, leading to a type of inflammatory cell death

called pyroptosis. One such example is the oligomerisation of NAIP5/NLRC4 in response to flagellin (Tenthorey et al., 2017). Since the early 2000s, oligomerisation has been known to be a feature of R gene activation; examples include the oligomerisation of the NLR N of *Nicotiana tabacum* upon perception of *Tobacco mosaic virus* helicase (Mestre and Baulcombe, 2005). More recently, structural biology approaches have revealed that activated plant NLRs can also assemble into characteristic oligomeric complexes reminiscent of metazoan inflammasomes, and termed “resistosomes” (Xiong et al., 2020).

The last five years have been an exciting time in plant NLR biology as full length NLR structures have been determined by cryo- electron microscopy (cryo-EM) for the first time. Structures of the singleton CC-NLR ZAR1 with co-receptor kinase RKS1 were determined in preactivated and active states in 2019 (Figure 1-4 A) (Wang et al., 2019a; Wang et al., 2019b). This was followed by the first two TIR-NLR resistosome structures, published in late 2020. The first TIR-NLR resistosome resolved was the tetrameric ROQ1 resistosome (Martin et al., 2020) closely followed by the tetrameric TIR-NLR RPP1 resistosome (Figure 1-4 B) (Ma et al., 2020). In the case of both TIR NLRs, oligomerisation brings together the four N-terminal TIR domains to form a holo-enzyme, which catalyses small molecule signalling processes.

These three oligomeric resistosome structures, ZAR1, ROQ1, and RPP1 showed striking similarities to various metazoan inflammasomes (Meunier and Broz, 2017), and comparisons of the ZAR1 resistosome to the Apaf-1 inflammasome, CED-4 apoptosome, and NLRC4 inflammasome have been made (Wang et al., 2019a). However, as noted earlier, the origins of plant and metazoan NLRs are believed to have independent evolutionary origins, with similarities in domain structure and conformation of oligomeric complexes attributed to convergent evolution (Yue et al., 2012; Urbach and Ausubel, 2017).

Together with the kinase RKS1, the NLR ZAR1 guards the host decoy PBL2 which is modified by pathogen effectors. ZAR1 and RKS1 exist in a complex together before activation, and then upon associating with modified PBL2, ZAR1 becomes activated and competent to oligomerise into a pentameric resistosome complex (Wang et al., 2019b). It was shown that the N-terminal alpha helix, α_1 , of ZAR1 becomes solvent exposed upon activation and can associate with the equivalent α_1 of neighbouring ZAR1 monomers, forming a hydrophobic funnel surface that was hypothesised to insert into cell membranes (Wang et al., 2019a). The ZAR1 structure was determined using protein which was purified with N-terminal affinity and solubility tags removed, and association of the “funnel” provided sufficient rigidity to resolve cryo-EM density for that region, providing direct evidence the formation of the α_1 funnel (Wang et al., 2019a). The funnel was later demonstrated to act as a cation channel, facilitating Ca^{2+} influx upon resistosome formation

and insertion into the plasma membrane (Bi et al., 2021). It remains unclear if the channel is directly causing cell death, or if the resulting Ca^{2+} influx is part of a signalling pathway which ends in cell death. A similar mode of activity is suggested for the helper NLR NRG1, the CC domains of which self-associate and form a Ca^{2+} permeable channel when expressed in HeLa cells (Jacob et al., 2021).

Sr35 is a CC-NLR from wheat and its cryo-EM structure was determined in a pentameric resistosome complex with effector AvrSr35 from the wheat stem rust pathogen (Förderer et al., 2022; Zhao et al., 2022). In the study by Förderer et al., although the $\alpha 1$ helix of the N-terminal CC-domain was not resolved by cryo-EM, and despite low sequence identity, the formation of an $\alpha 1$ funnel analogous to that of ZAR1 was hypothesised. Through *Xenopus laevis* oocyte electrophysiology experiments, Ca^{2+} channel activity upon resistosome activation was demonstrated. The structure was obtained in an ATP bound state although no nucleotides were supplemented, which is consistent with its activation into a resistosome. The “EDVID” motif of the CC domain, interacts with the $\text{LRR}^{\text{R-cluster}}$, a series of arginine residues spaced along the LRR domain, each separated by one LRR repeat (correlating to one turn of the superstructure), an interaction which was observed in ZAR1 too (Förderer et al., 2022).

In a further example of a CC-NLR, the cryo-EM structure of the wheat CC-NLR MLA13 was recently reported as a pre-print (Lawson et al., 2024). MLA13 was purified in a stable heterodimer with effector AVR_{A13}-1 of *Blumeria hordei*, the causative agent of powdery mildew. The complex was purified from heterologous expression in *N. benthamiana*, and although not adopting a wheel-like resistosome structure, it may represent an activation intermediate. AVR_{A13}-1 is recognised by the LRR and WHD domain of the NB-ARC, and structural information about the interface in combination with sequence alignment with other MLA alleles was used to make a single amino acid substitution in NLR MLA7 that facilitated a broadening of recognition range to encompass one more effector, AVR_{A13}-V2 (Lawson et al., 2024).

Oligomerisation of helper NLRs from the NRC network of Solanaceous plants has been demonstrated through biochemical means such as blue native-PAGE. For example, the potato late blight pathogen, *Phytophthora infestans*, effector AVR_{Ramr3} is detected by the sensor NLR Rpi-amr3 from *Solanum americanum* and triggers incorporation of helper NLRs NRC2 and NRC4 into higher molecular weight oligomers (Ahn et al., 2023). Recently, cryo-EM structures have been determined of the helper NLR NRC2 from two species in various oligomeric states. *Solanum lycopersicum* NRC2 has been observed to form inactive dimers, tetramers, and higher order autoinhibited oligomers (Ma et al., 2024). A pre-print describing the structure of a hexameric *N. benthamiana* NRC2 resistosome has also

recently been released, in which a NRC2 forms an activated wheel shaped resistosome reminiscent of the ZAR1 pentamer (Madhuprakash et al., 2024). The structures of *S. lycopersicum* were of sufficient resolution to identify inositol hexakisphosphate (IP₆) or pentakisphosphate (IP₅) binding to the inner surface of the LRR domain. When the residues mediating binding to IP₅ or IP₆ were mutated, binding of these small molecules was lost and cell death activity was lost in planta, suggesting a role for inositol phosphates in NRC2 activation. The *N. benthamiana* NRC2 structure was determined in an ATP bound state, whereas no nucleotide binding was reported in the *S. lycopersicum* NRC2 structures, nor were any nucleotides or nucleotide analogues supplemented during the protein purification. The filamentous form of S/NRC2 was reported upon high concentrations of protein, and was previously observed by confocal microscopy upon overexpression of NRC2-GFP in *N. benthamiana* (Duggan et al., 2021). Assuming that the NRC2 is operating by the same mechanism in each species, then taken together these two studies suggest a transient “activation-and-release” mode of action although it remains unclear how transformation between inactive filaments, tetramers, or dimers and an active hexameric form would occur.

Finally, the cryo-EM structure of an autoactivated NRC4 hexameric resistosome was reported, including negative stain 2D class averages providing evidence for the formation of hexameric NRC0, NRC2, and NRC3 resistosomes (Liu et al., 2024). The $\alpha 1$ of the CC-domain was not resolved as a funnel, and similar to Sr35 this is likely to be due to intrinsic flexibility of this helix. Analysis of the dimension of the putative ion conducting pore formed by the activated NRC4 resistosome indicated a larger internal diameter of the intracellular portion of the channel when compared to the ZAR1 or Sr35 resistosomes, although the central portion of the pore shared similar dimensions between all three resistosomes. This can be taken to imply that function as a Ca²⁺ conducting channel may be similar between these three NLRs (Wang et al., 2019a; Förderer et al., 2022; Zhao et al., 2022; Liu et al., 2024).

A general mode of resistosome activation has emerged, in which the NB-ARC domain binds ADP, which upon exchange for ATP facilitates association with other NLRs into an oligomeric resistosome. The formation of the oligomer brings the N-termini into proximity and thereby allows the assembly of the TIR N-terminal holoenzyme or the CC-NLR cation channel.

These structures illustrate that, while diverse, NLR conformations and resistosome architecture follow some unifying principles. However, there is also no structure to date of an NLR with an integrated domain between the N-terminal and the NB-ARC domains, leaving questions as to how this might be arranged in an activated resistosome. Furthermore, all CC-NLR structures determined to date are of singleton NLRs; it remains

unclear what the spatial arrangement of a paired NLR resistosome would be. In the case of paired NLRs which are activated by de-repression, it is possible that the helper NLR forms an oligomeric cartwheel-like resistosome similar to the ZAR1 resistosome (Figure 1-4 A). In the case of paired NLRs where cooperative activation occurs, such as in the Pik pair, it is uncertain whether paired NLR resistosomes would form a multimeric cartwheel-like structure with helper NLRs at the centre with their CC-domains forming a funnel structure. An alternative hypothesis is the formation of a complex resembling the NAIP2/NLRC4 inflammasome in which a single activated sensor NLR templates the oligomerisation multiple copies of the helper NLR (Figure 1-4 C) (Zhang et al., 2015).

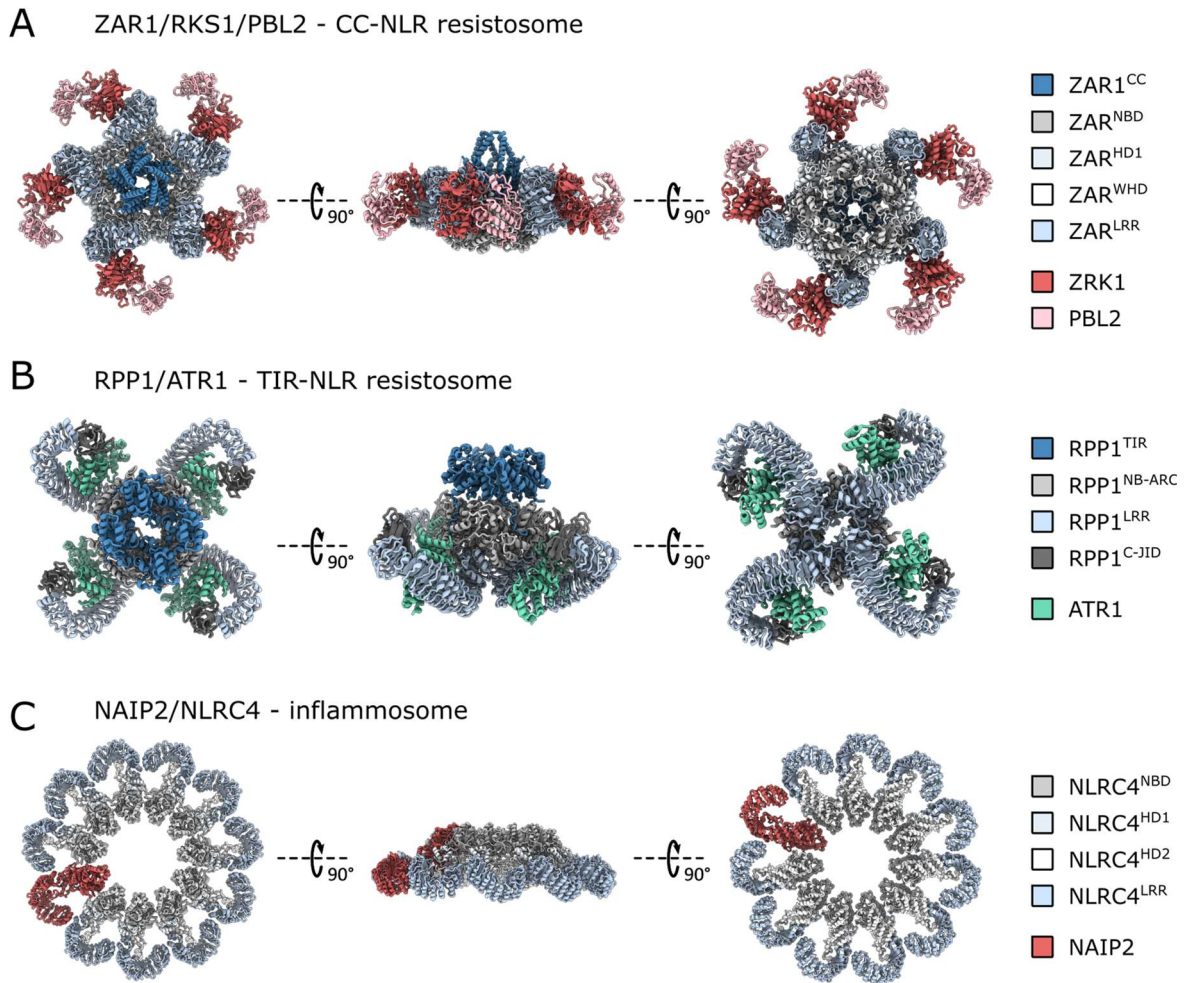


Figure 1-4: Oligomeric resistosome and inflammasome complexes

A) ZAR1/ZRK1/PBL2 - (PDB-ID: 6J5T) (Wang et al., 2019a). **B)** RPP1/ATR1- (PDB-ID: 7CRC) (Ma et al., 2020). **C)** NAIP2/NLRC4 - (PDB-ID: 3JBL) (Zhang et al., 2015) - NAIP2 complex structure not available, so one NLRC4 monomer is coloured red to represent NAIP2. Domains are coloured according to the key. Domain abbreviations as follows: CC = coiled coil, TIR = Toll/interleukin receptor, NBD = nucleotide binding domain, NB-ARC = Nucleotide binding adaptor shared by Apaf-1, R proteins, and Ced-4, HD1,2 = helical domain 1,2, WHD = winged helix domain, LRR = leucine rich repeat, C-JID = C-terminal jelly-roll/Ig-like domain.

Understanding the precise interactions and three-dimensional structure underpinning the formation of NLR resistosomes will be key in informing rational design of synthetic immune receptors, as discussed in the following section. Experimental structures obtained in a variety of biochemical states will provide insight into the NLR conformational changes involved in resistosome assembly and activation. High resolution structures can inform choice of single residue mutations to perform subtle engineering, for example to enhance effector binding affinity or change which effector is recognised. Meanwhile, lower resolution information can still provide a valuable guide as to which surfaces or domains are involved in interactions or are solvent accessible. This information could guide engineering to alter downstream signalling, or guide incorporation of novel integrated domains or novel effector binding surfaces without introducing steric clashes or disrupting interfaces required for complex formation. Structure informs function, therefore structural information is crucial for understanding the mechanisms underlying plant immunity observed in wild-type NLRs. This understanding will pave the way for continued innovation in synthetic plant immunity, contributing to the eventual goal of enhancing crop resistance to pathogens in the field.

1.6 NLR engineering is an opportunity to introduce new resistances

Introduction or manipulation of NLRs provides an opportunity to introduce novel immune recognition capabilities, therefore supporting crop health in the face of rapidly mutating pathogen effector populations. The specificity and magnitude of NLR response can be modulated in a number of different ways, as will be discussed in this section.

Traditional breeding approaches aim to develop resistant crop cultivars by introducing existing NLRs from breeding populations or through backcrossing with wild relatives. However, deploying an unlimited number of NLRs is not feasible as their expression in absence of pathogens can incur a yield penalty. Therefore, deployment of resistance activities must be tailored to the situation, to reach an appropriate compromise between crop health and yield.

The transfer of novel NLRs, including from other species, offers the opportunity to introduce novel resistances not previously found in that cultivar or species. However, many NLRs function only with partner proteins. In the case of paired NLRs, they can easily be introduced as a pair as they are co-localised in the genome, however networked NLRs such as those in the Solanaceae NRC network present a challenge as the exact identity and activity of each node of the network is still being elucidated (Wu et al., 2017).

In an alternative approach to introducing entirely new NLRs, a second method is to engineer existing proteins which have been characterised, to either introduce new recognition specificities, or to fine-tune the magnitude of the response. In the case that post-translational modifications (PTMs) are important in the activation or regulatory mechanisms of an NLR, then maintenance of the site of the PTM, and presence of the relevant protein to add the PTM would be an important factor in producing an engineered NLR which can function as intended. Any binding partners, or downstream signalling network components – even those not yet characterised – should be present and expressed appropriately. This will promote effective incorporation of the engineered NLR into the host immune system, and so is one reason why engineering a NLR that originates from within the same genome as the one it will be deployed in can offer a practical advantage.

NLRs can recognise pathogen effectors directly, or indirectly, by monitoring host proteins or decoys altered by effectors, as previously discussed (1.4.2.2). This diversity in NLR recognition mechanisms means that engineering strategies must also be diverse and should be tailored to the specific biology of the NLR to be engineered. Recent advances in structural understanding of NLRs have facilitated improved engineering efforts including design of targeted mutations, in contrast to early NLR engineering attempts, which often employed random mutagenesis or domain swapping. The examples of engineering presented in the following subsections will provide illustration of this.

1.6.1 Leucine-rich repeat (LRR) domain engineering

The leucine-rich repeat (LRR) domain is responsible for effector recognition in many NLRs, including the wheat stem rust resistance genes Sr33, Sr35, and Sr50. The LRR of Sr33 was resurfaced using a structural prediction and diversity analysis guided approach based on the LRR of Sr50 and the LRRs of other closely related MLA family of NLRs from barley. This engineering enabled Sr33 to recognise the effector AvrSr50, which was previously unrecognised (Tamborski et al., 2022). Following a similar approach, multiple sequence alignments and the cryo-EM structure of Sr35 were used to inform mutation of a barley and a wheat homologue of NLR SH1 to facilitate binding to AvrSr35 (Förderer et al., 2022).

1.6.2 Integrated domain mutation

Modification of integrated domains provides an opportunity to refine their existing recognition specificities, enhancing the range of effectors recognised and their binding affinities. This approach relies on a detailed structural understanding of the domains being

modified so that rational mutations can be made. In the case of the Pik HMA, distinct binding interfaces have been reported upon complex formation with different MAX effectors. AVR-Pik effectors bind on the “Pik” interface, which does not overlap with the “Pia” interface, defined by the location of AVR-Pia binding (Maqbool et al., 2015; Varden et al., 2019).

A further challenge of engineering integrated domains is that some effectors that bind integrated domains have been reported to contact other domains of the NLR too. For example, the *Magnaporthe* effector AVR-Pia is recognised by the integrated HMA domain of the rice NLR RGA5, but a mutated RGA5 lacking any HMA domain was still able to bind AVR-Pia, suggesting that AVR-Pia is binding other regions of RGA5 in addition to the ID (Ortiz et al., 2017). Correspondingly, in a pre-print by Zhang et al., mutations in both the HMA and in the lysine-rich C-terminal region downstream of the HMA of RGA5 were made to confer resistance to AVR-PikD (Zhang et al., 2022).

In another study of RGA5 engineering, RGA5 recognition of AVR-Pia was exchanged for recognition of AVR-Pib. Wildtype RGA5 does not recognise the *M. oryzae* effector Avr-Pib; however, when the binding surface of RGA5 was engineered, RGA5 lost binding to its cognate effector AVR-Pia and gained binding to AVR-Pib. To achieve this, the lysine-rich region at the C-terminal end of the RGA5 HMA domain was mutated, to accommodate a positive patch on the surface of AVR-Pib (Liu et al., 2021). The mutation was inspired by the binding of AVR1-CO39 to RGA5, explored in previous work by Guo et al. (Guo et al., 2018).

In 2022, Cesari et al. reported the engineering of RGA5 to recognise AVR-PikD in addition to its original targets, AVR1-CO39 and AVR-Pia. Recognition in *N. benthamiana* cell death assays was achieved by incorporating Pikp-1 interface features into RGA5-HMA (Cesari et al., 2022). This engineering exploits the different binding sites on different sides of the HMA ID scaffold. RGA5 recognises the *M. oryzae* effectors AVR1-CO39 and AVR-Pia via an interface on one side of its HMA domain, while the HMA of Pikp1 binds the effector AVR-PikD at another interface (Maqbool et al., 2015; De La Concepcion et al., 2018; Guo et al., 2018).

However, activation of cell death upon co-expression of effector proteins and NLRs in *N. benthamiana* does not always correspond to a resistance phenotype in cereal crops, as documented in the 2022 work by Cesari and colleagues. Transgenic rice lines expressing the engineered RGA5 variants exhibited resistance towards *M. oryzae* expressing AVR1-CO39 and AVR-Pia, but were susceptible to *M. oryzae* expressing AVR-PikD, indicating that a high affinity interaction between an effector and the HMA domain is not sufficient to

activate immunity mediated by the RGA4/RGA5 complex. This is likely to be because additional contacts between the effector and RGA5 domains other than the HMA are required for activation (Cesari et al., 2022).

1.6.3 Integration of novel domains

The sHMA OsHIPP19 binds the ‘stealthy’ AVR-Pik alleles AVR-PikC and AVR-PikF, which are not recognised by any Pik NLR allele (Maidment et al., 2021). Through exchange of the HMA domain of Pikp-1 for the HMA domain of sHMA OsHIPP19, a chimeric Pik NLR is produced which has an expanded recognition capacity encompassing the ‘stealthy’ AVR-Pik effectors which were previously not recognised by the wildtype receptor. Furthermore, this chimeric receptor produced AVR-PikC or AVR-PikF dependent cell death in *N. benthamiana*, and resistance in rice against *Magnaporthe* strains carrying AVR-PikC or AVR-PikF (Maidment et al., 2021; Maidment et al., 2023). In another example, which combines the use of integration of novel domains and the use of targeted mutations, the integrated HMA domain of Pikm-1 was exchanged with the HMA of RGA5 to produce Pikm-1^{RGA5}, capable of weak AVR-Pia binding. Targeted mutations were then made, to introduce back the regions of the Pik-interface which facilitate binding of AVR-PikD, AVR-PikC, and AVR-PikF while retaining weak AVR-Pia binding (Bentham et al., 2023). However, incorporation of new domains is not limited to domains found in the host species; in a further synthetic biotechnological extension of the Pik chassis work, the HMA domain of Pik-1 was replaced by a VHH nanobody specific to GFP or RFP. These synthetic receptors, named ‘Pikobodies’ produced cell death in response to GFP or RFP expression (Kourelis et al., 2023). This demonstrates that Pik NLRs can theoretically be engineered to contain effector recognition modules for any secreted protein delivered into plant cells, extending NLR function beyond binding surfaces naturally occurring in plants.

1.6.4 Other mechanisms

There are alternative modes of activating the plant immune system in response to the presence of a pathogen effector. Understanding the biochemical function of an effector offers an opportunity to exploit it and use it to activate plant immunity. For example, the RRS1-R/RPS4 pair are under negative regulation; RRS1-R suppresses RRS1-R^{slh1}, which otherwise causes cell death in combination with RPS4 (Sohn et al., 2014). The SAPo5 effector of *Phytoplasma* targets host transcription factors to the proteasome for degradation and thereby changes host expression patterns to produce developmental changes (Huang

et al., 2021). In subsequent work by Wang et al., the GATA zinc finger transcription factor “degron” signal recognised by SAPo5 was used to target a protein of choice for degradation. This was demonstrated in *N. tabacum* by adding the SAPo5 degron to RRS1-R, and upon secretion of SAPo5 into the cell, RRS1-R was targeted to the proteosome and degraded. Therefore the suppression of RRS1-R^{slh1} was removed and RRS1-R^{slh1} and RPS4 proceed to cause cell death in a SAPo5-dependant manner (Wang et al., 2021).

One additional method through which specific effector activities can be harnessed to activate immunity is via proteolytic activity, if the cleavage site is known. An example of this is *P. syringae* pv. *phaseolicola* effector AvrPphB which cleaves decoy protein PBS1, which is then detected by the *Arabidopsis* NLR RPS5 (Pottinger and Innes, 2020). By replacing the PBS1 cleavage site with the cleavage site of the NIa protease of soybean mosaic virus, the presence of NIa protease now activates RPS5 – an approach which could be generalised (Kim et al., 2016; Pottinger et al., 2020).

1.6.5 Further challenges such as autoimmunity can be overcome

Engineering an NLR which operates as part of a network, or producing a new combination of NLRs through either engineering or traditional breeding can have unintended consequences. The whole NLR network may be perturbed, which can lead to inappropriate immune activation in absence of pathogen effectors; this is an autoimmune phenotype, deleterious to the fitness of the plant. In one example, Sr33 and Sr50 were known to sometimes cause autoactive cell death upon overexpression in *N. benthamiana* (Cesari et al., 2014). In the work of Tamborski and colleagues to engineer Sr33 to recognise AvrSr50, autoactivity of Sr50, but not of Sr33, was observed in wheat protoplasts. The autoactivity was overcome by targeted mutations in the NB-ARC domain of Sr50, to make it more like the NB-ARC of Sr33 (Tamborski et al., 2022). In a second example, in the work of Bentham et al. to engineer the RGA5 HMA into the paired NLR Pik, autoactivity was observed in *N. benthamiana* upon expression of Pikm-1^{RGA5}/Pikm-2. However, by mismatching the allele of the helper NLR Pik-2, and expressing Pikm-1^{RGA5}/Pikp-2, the autoimmune phenotype was overcome. This then facilitated further engineering of the Pik-1-HMA and provides an example of how allelic diversity can provide tools for NLR engineering. In contrast, the combination of Pikp-1/Pikm-2 leads to autoactivity (Bentham et al., 2023).

In this section of the introduction, I have summarised the potential of NLR engineering to generate novel disease resistance, and demonstrated some of the tools available to mitigate autoactivity, should it arise. Many of these examples of NLR engineering were founded on structural understanding of the NLR or effectors in question. As previously discussed,

improved structural and mechanistic knowledge of paired Pik NLRs would contribute to the objective of engineering paired NLRs and NLRs with integrated domains.

1.7 Aims and objectives of the thesis

During my PhD I focussed on two main objectives. The first was to produce recombinantly expressed and purified Pik resistosomes, with the aim to proceed to structural and biophysical analysis through approaches such as cryo-electron microscopy, SAXS, or mass photometry. In Chapter 3, I present progress towards expression and purification of full length Pik-1/Pik-2 in a pre-activated or activated complex using *Nicotiana benthamiana* and insect cell expression systems. During this work I found that expression and purification of the Pik pair was challenging, and therefore I turned to investigate the cell localisation and oligomerisation dynamics of the Pik pair upon activation using Blue-Native PAGE and confocal fluorescence microscopy. This work is presented in Chapter 4.

My next aim was to characterise the novel *Magnaporthe oryzae* effector AVR-Mgk1 and its binding to Piks-HMA and other rice HMA proteins. This work is presented in Chapter 5. I expressed and purified AVR-Mgk1 and HMA proteins individually and in complexes from *E. coli* and performed analytical gel filtration. I set out to determine the structure of AVR-Mgk1 in complex with Piks-HMA using X-ray crystallography, and I present the structures of Piks-HMA, and AVR-Mgk1 in complex with a small HMA. I performed biophysical analysis by surface plasmon resonance. Finally, I attempted a proof-of-concept engineering of Pikm-1 to produce a new source of resistance to AVR-Mgk1 by integrating the small HMAs OsHPPo2/OsHPPo3/OsHPPo4 in place of the wild type HMA domain.

Chapter 2: Materials and Methods

General materials and methods are described here, while materials and methods specific to individual chapters can be found within each results chapter.

2.1 DNA methods

2.1.1 Molecular Cloning

All bacterial and plant expression constructs were assembled using the modular cloning (MoClo) Golden Gate (Engler et al., 2008). Constructs for bacterial expression were prepared in level 1 pOPIN Golden Gate (pOPIN-GG) acceptor vectors (Bentham et al., 2021). Constructs for in planta expression were prepared in lvl1 and lvl2 Golden Gate acceptors. Constructs for insect cell expression were made using the GoldenBac system (Neuhold et al., 2020). The genes of interest were first tagged using Golden Gate modules before PCR amplification of gene and tag together and insertion into GoldenBac entry vectors via In-Fusion cloning (Raman and Martin, 2014). Multiple GoldenBac entry vectors were combined through a Golden Gate reaction to produce a single multigene expression construct.

2.1.1.1 Golden gate cloning

Molecular cloning was principally carried out by the Golden Gate method (Engler et al., 2008). Overhangs for each transcriptional unit are defined so that constructs can be assembled from individual modules (Weber et al., 2011), and the method has been adapted for plants (Engler et al., 2014) and for bacterial expression (Bentham et al., 2021).

Hypertrans (Sainsbury and Lomonosoff, 2008; Sainsbury et al., 2009) and Geminivirus (Mor et al., 2003) expression systems were used during attempts to increase protein expression levels in *N. benthamiana*. The level 2 Golden Gate Hypertrans vector pEAQ-HT contains the modified 5'-untranslated region (UTR) and the 3'-UTR from Cowpea mosaic virus which flank the gene of interest. The Geminivirus expression system exploits DNA replication machinery of the Bean yellow dwarf virus (BeYDV) to produce high copy numbers of the expression vector containing the gene of interest, thereby providing the potential for greatly increased expression levels.

2.1.1.2 Production of Pik-2 CC-domain mutants

Pikm-2 CC-domain mutants L19E/L23E, and L15E/L19E/L23E were produced as analogous mutants to the ZAR1 F9A/L10A/L14A and NRC4 L9A/V10A/L14A mutants,

based on alignment with other CC-NLR CC-domains performed by Adachi et al. (Adachi et al., 2019; Wang et al., 2019a). Work of Dr Enoch Lok Him Yuen, with Dr Hiroaki Adachi, revealed that the L15E mutation was associated with autoactivity, and so variants with and without this mutation were used (Dr Mark Banfield, Dr Hiroaki Adachi, personal communication). The genes were mutated using an adapted Golden Gate method. A pair of primers containing the required mutations were annealed to produce a short replacement gene fragment with unique 4 bp overhangs. The primers were annealed by mixing in equimolar ratio (100 μ M each) before heating to 94 °C for 2 minutes followed by gradual cooling to room temperature. This was either performed using a thermocycler or using a metal heat block which gradually cooled once turned off. The whole of the WT gene and expression vector were amplified with PCR with BpiI cut sites at each end. Upon BpiI digestion in a digestion-ligation reaction, 4 bp overhangs complementary to the amplified gene and backbone were produced, allowing incorporation of the fragment. This approach was followed to produce these mutations in both the wild-type *N. benthamiana* expression construct, and in the Sf9 optimised insect cell expression construct.

2.1.1.3 InFusion Cloning

InFusion Cloning (Takara Biosciences) was performed according to (Raman and Martin, 2014). Acceptor vectors were linearised at the desired point of insertion by PCR amplification. Inserts were then produced by PCR amplification using primers designed with 15 base pair tails with complementarity to the insertion point in the backbone. The exonuclease activity of the DNA polymerase in the InFusion reaction then trims back the 3' ends of the DNA to reveal complementary overhangs which anneal. Upon transformation into *E. coli*, the DNA fragments are ligated.

2.1.2 Gene synthesis

Genes not already held by the Lab were Sf9 expression codon optimised if applicable (Twist Biosciences) and domesticated for the Golden Gate system by introduction of synonymous changes to eliminate any BsaI or BpiI restriction sites before being synthesised as gene fragments (Integrated DNA Technologies gBlocks™, Genewiz FragmentGENE, or Twist Biosciences Gene Fragments). Genes were typically ordered with a flanking BpiI cut site designed with overhangs to allow direct incorporation into Golden Gate level 0 acceptor vectors. Long sequences such as full length NLR genes were synthesised in multiple fragments, each flanked by a BpiI cut site designed with unique overhangs to facilitate simultaneous assembly of all fragments into the appropriate level 0 Golden Gate acceptor vector in a single digestion-ligation reaction.

Genes which were required for N- and C-terminal tagging were typically ordered with a stop codon present, ready to be incorporated into a Golden Gate level 0 plasmid to then produce untagged or N terminal tagged level 1 expression constructs. Level 0 vectors with appropriate overhangs for C terminal tagging were produced by PCR amplification excluding the stop codon, followed by a further digestion-ligation reaction.

All acceptor vectors and level 0 modules containing tags, promoters, and terminators were obtained from TSL SynBio platform, apart from the GoldenBac vectors which were obtained from the Belgian Coordinated Collections of Microorganisms GeneCorner.

2.1.3 Polymerase Chain Reaction

2.1.3.1 Phusion PCR

PCR using Phusion polymerase (New England Biolabs) was carried out to amplify genes for insertion into Golden Gate acceptor vectors or directly into expression vectors (for example, as part of InFusion cloning). These PCR products were required for further cloning steps and so upon completion were subject to immediate Dpn1 digestion. In this way re-transformation of the template plasmid is prevented through digestion of the methylated template DNA.

2.1.4 DNA gel electrophoresis

Agarose gels were cast from molten 1 % w/v agarose made up in TAE buffer (40 mM Tris-acetate pH 8.0, 1.0 mM EDTA), with addition of 1 µL Midori Green dye (Nippon Genetics Europe GmbH) per 25 ml of gel. DNA samples were mixed with 4 x FOG loading dye (12 % ficoll 400 and 0.25 % w/v Orange G), and 1 Kb Plus DNA ladder (New England Biolabs) was included in one lane as a standard. Gel electrophoresis was carried out by applying a voltage of 80 – 120 V in TAE buffer. Gels were visualised using a UV light box (UVitec) or images acquired using a G:box imager (Syngene).

2.1.5 PCR product clean-up for subsequent cloning steps

PCR products required for subsequent cloning steps were purified using a column-based kit (NucleoSpin® gel and PCR clean-up, Macherey-Nagel). First DNA gel electrophoresis was performed with 5 µL of PCR mixture to identify if the PCR had produced a single band of the correct size for the desired product. If multiple bands were produced, the desired band was excised with a clean razor blade over a UV light box, and DNA extracted using a column-based kit. If the PCR had produced the single desired product, then the remaining PCR

mixture was subject to Dpn1 digestion and then the PCR mixture was purified using the same column-based kit.

2.1.6 DNA sequencing

All constructs were confirmed by sequencing. Constructs produced by purely ligation based cloning approaches such as Golden Gate were Sanger sequenced across the boundary between the insert and backbone (Azenta/Genewiz). Plasmids cloned using PCR based steps were Sanger sequenced throughout the body of the gene of interest using multiple primers. As the accessibility of whole plasmid sequencing improved, whole plasmid sequencing (Plasmidsaurus) was used on occasion for sequence confirmation of challenging constructs.

2.1.6.1 Sanger Sequencing

250 μ M DNA and 2.5 μ M sequencing primer were diluted in a 10 μ L final volume and submitted for Sanger sequencing (Genewiz, now Azenta Life Sciences).

2.1.6.2 Whole plasmid sequencing

A minimum of 10 μ L plasmid DNA at 30 ng/ μ L was submitted for whole plasmid sequencing (Plasmidsaurus).

2.1.7 Sequence curation and alignment

All sequence manipulation, in silico construct assembly, and alignment of sequencing results was carried out using Benchling biology software (2020-2024). To confirm incorporation of the correct DNA inserts during cloning, DNA sequencing results were aligned with plasmid maps using the inbuilt implementations of EMBL-EBI alignment algorithms MAFFT (Katoh and Standley, 2013) and Clustal Omega (Sievers and Higgins, 2014). Once confirmed, plasmid maps were downloaded in .gb format for storage.

2.2 Bacterial cell culture and protein purification

2.2.1 Media

2.2.1.1 Lysogeny Broth

Lysogeny Broth (LB, Miller, Formedium) is composed of 1 % w/v tryptone, 0.5 % w/v yeast extract, 1 % w/v sodium chloride, at pH 7.0. LB Agar for growth of bacteria on plates was produced by addition of 1.1 % w/v agar to LB media before autoclaving. Appropriate antibiotics were added to molten LB Agar before dispensing into petri dishes.

2.2.1.2 Auto-induction Media

Auto-induction media (AIM, LB broth base including trace elements, Formedium) is composed of 1 % w/v tryptone, 0.5 % w/v yeast extract, 0.33 % w/v ammonium sulfate, 0.68 % w/v monopotassium phosphate, 0.71 % w/v monosodium phosphate, 0.05 % w/v glucose, 0.2 % w/v α -lactose, 0.015 % w/v magnesium sulfate, 0.003 % w/v trace elements.

2.2.1.3 SOC

Super Optimal broth with Catabolite repression (SOC) media is composed of 2 % w/v tryptone, 0.5 % w/v yeast extract, 0.058 % w/v sodium chloride, 0.0186 % w/v potassium chloride, 0.203 w/v magnesium chloride, 0.246 % w/v magnesium sulfate, 0.36 % w/v glucose, pH 7.0.

2.2.1.4 2 X YT

2 X YT broth (Formedium) is composed of 1.6 % w/v tryptone, 1 % w/v yeast extract, 0.5 % sodium chloride, pH 7.4.

2.2.2 Bacterial strains

Chemically competent cells were prepared by Dr Richard Hughes and Dr Rafal Zdrzalek.

2.2.2.1 Cloning strains

For cloning and growth of *E. coli* for plasmid propagation, chemically competent Stellar *E. coli* (Takara Bio) were used.

2.2.2.2 Bacmid production strain

For production of bacmid DNA for insect cell expression, DH10EMBacY *E. coli* (Geneva biotech) were used.

2.2.2.3 Expression strains for agrobacterium mediated expression in *N. benthamiana*

Agrobacterium tumefaciens GV3101 vir-helper strain was used for agrobacterium mediated expression in *Nicotiana benthamiana* plants.

2.2.2.4 *Escherichia coli* expression strains

SHuffle *E. coli* (Lobstein et al., 2012), BL21 (DE3) *E. coli* (New England Biolabs), and BL21-AI One Shot (Arabinose Inducible) *E. coli* (Invitrogen) were used for protein expression.

2.2.3 *Escherichia coli* cell culture for protein expression

Bacterial expression constructs were transformed into chemically competent expression strains of *E. coli* by heat shock. Typically, proteins were expressed in 8 x 2 L flasks each containing 1 L of LB or auto-induction media (AIM) (Studier, 2005) supplemented with appropriate antibiotics to maintain selection. The cultures were grown at 37 °C (AI) or 30

°C (SHuffle), until the OD₆₀₀ of 0.7 was reached, before adjusting to the expression temperature (18-37 °C, protein dependent). The cultures were then induced by addition of 1mM IPTG or 0.1 % arabinose, before incubation at the expression temperature for 18 hours. The cells were then harvested by centrifugation and the pellets either scraped into plastic bags and flattened, or resuspended in purification buffer A1 (50 mM HEPES pH 8.0, 500 mM NaCl, 30 mM Imidazole, 5 % glycerol, 20 mM glycine) before storage at -80 °C.

2.2.4 Protein purification from *E. coli*

Escherichia coli expressed proteins were typically purified by IMAC-gel filtration followed by 3C tag cleavage, reverse IMAC to remove the tag, and a second round of gel filtration. The following protocol was applied to both individual proteins expressed alone, or to protein complexes co-expressed with an affinity tag on only one member of the complex.

Typically for purification from 8L of *E. coli* culture, the cell pellet was thawed in a total volume of 300 ml in buffer A1 (50 mM HEPES pH 8.0, 500 mM NaCl, 30 mM Imidazole, 5 % glycerol, 20 mM glycine) with addition of 1 cOmplete protease inhibitor tablet (Roche) per 50 ml of buffer. After lysis by sonication, the lysate was clarified by centrifugation at 45000 RCF for 45 minutes at 4 °C. The clarified lysate was applied to a 5 ml Ni-NTA column (Cytiva) connected to an AKTA Xpress or AKTA Go chromatography system controlled by Unicorn software (Cytiva). The Ni-NTA column was washed with buffer A1, before step elution in 100% buffer B1 (50 mM HEPES pH 8.0, 500 mM NaCl, 250 mM Imidazole, 5 % glycerol, 20 mM glycine) followed by application onto a Superdex 75 16/60 or Superdex 75 26/60 gel filtration column equilibrated in buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl). In the case of purification with an AKTA express, the elution from the Ni-NTA column was stored in an internal loop and directly applied onto the gel filtration column. In the case of purification with an AKTA Go, the fractions from Ni-NTA elution were collected and injected onto the gel filtration column manually. The fractions containing the protein of interest were identified by SDS-PAGE, pooled, and 10 ug of 3C protease added per 1 mg of protein and incubated for 16 hours at 4 °C.

At the end of the incubation, imidazole was added to the pooled cleaved protein to a total concentration of 20 mM, and using a peristaltic pump, the protein was applied to a 5 ml Ni-NTA column equilibrated in buffer A1. If a MBP tag was present on the protein, a MBP trap would also be added in tandem with the Ni-NTA column to improve capture efficiency of cleaved tag. The flow through (containing the protein of interest) was collected and concentrated in a centrifugal concentrator (Pierce, Thermofisher, or Amicon, Millipore) of appropriate molecular weight cut-off (MWCO, double the predicted molecular weight of the protein). The protein was then injected onto an appropriate gel filtration column for the

protein size, typically a Superdex 75 16/60 or Superdex 75 26/60, equilibrated in buffer A4. The fractions of interest were identified by SDS-PAGE, pooled, and again concentrated using centrifugal concentrators. Concentrated protein was aliquoted and snap cooled in liquid nitrogen before storage at -80 °C. Protein concentration was assessed by absorbance at 280 nm (NanoDrop One, Thermofisher) and corrected to account for calculated extinction coefficient (ExPASy ProtParam). In the case that the protein contained no aromatic residues, the protein location within the elution fractions was assessed by SDS-PAGE and concentration assessed by absorbance at 205 nm (NanoDrop One, Thermofisher) or by Direct Detect® (Merck-Millipore) which is an infra-red spectrometry-based method.

2.3 *Agrobacterium* infiltration of *N. benthamiana* for protein expression

Plant expression constructs were transformed into *Agrobacterium tumefaciens* GV3101 vir-helper strain by electroporation. An aliquot of *Agrobacterium* GV3101 cells was chilled with the plasmid in an electroporation cuvette on ice for 20 minutes before electroporation (2500 V). Cells were then allowed to recover for one to four hours with addition of 1ml SOC and incubation at 28 °C, before plating onto LB agar plates with rifampicin and gentamycin plus appropriate antibiotics to select for the transformed plasmid, typically kanamycin or carbenicillin. Colonies were isolated and 50ml cultures grown for 48 hours before being resuspended into infiltration buffer (10 mM MgCl₂, 10 mM MES, pH 5.6, 150 µM acetosyringone). Glycerol stocks for storage at -70 °C were prepared by mixing *Agrobacterium* culture with 50% glycerol (v/v).

In all subsequent weeks, when a glycerol stock was available, glycerol stocks were streaked out onto LB agar plates supplemented with appropriate selection antibiotics. Following 48 hours of incubation of 28 °C, the *Agrobacterium* lawn was scraped from the plates using a microbiological loop and resuspended in infiltration buffer.

Agrobacterium carrying the RNA silencing suppressor plasmid p19 was also prepared according to this method. The *Agrobacterium* suspensions were then diluted and mixed to give suspensions with OD₆₀₀ = 0.4 for NLRs and 0.6 for effector expression constructs respectively, and 0.1 for p19. The total OD₆₀₀ of *Agrobacterium* mixtures were made equal within any given experiment by addition of *Agrobacterium* transformed with empty vector control plasmid. Each mixture was infiltrated into whole leaves of 4–5-week-old *N. benthamiana* plants using a 1 ml syringe without a needle. For expression tests and western blotting, whole leaves or leaf disks were gathered at 3 dpi, or at other stated timepoints,

before being flash cooled in liquid nitrogen for storage at -70 °C. For cell death assays, leaves were gathered at 5 d.p.i. For fluorescence microscopy, *Agrobacterium* carrying each construct was infiltrated at OD₆₀₀ = 0.25, and *Agrobacterium* carrying p19 was infiltrated at OD₆₀₀ = 0.1.

2.4 Cell death assays in *N. benthamiana*

For cell death assays, *Agrobacterium* mixtures were infiltrated in spots of around 1 cm diameter, and the positions rotated around the leaf. Cell death experiments were repeated three times from fresh *Agrobacterium* plates grown from glycerol stock. 5 d.p.i, leaves were taken from the plants and bright light and UV photography was performed by Phil Robinson of the JIC photography department. Cell death of *N. benthamiana* results in release of phenolic compounds which fluoresce when illuminated by UV light. Damaged areas are visible as green spots on a red background of healthy tissue. Cell death severity in UV photographs is visually scored against a previously defined seven-point severity scale, where 0 represents no cell death and 6 represents very strong cell death (Figure 2-1) (Maqbool et al., 2015).

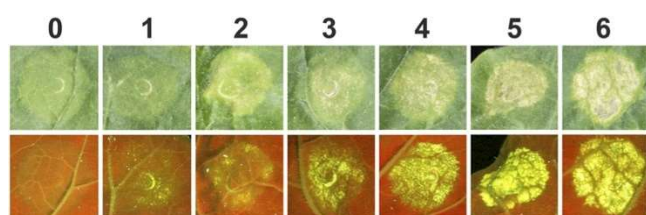


Figure 2-1: *Nicotiana benthamiana* cell death assay scoring scale

Above, in white light, and below, in UV light. Reproduced from Maqbool et al., 2015.

2.5 Preparation of *N. benthamiana* samples for SDS-PAGE and western blotting

Leaf disks were prepared for western blotting as follows. Leaves were powdered by addition of one metal bead per 1.5 ml microtube and agitation for 60 seconds at 1000 RPM in the Geno/Grinder with blocks cooled to liquid N₂ temperature. Alternatively, grinding was performed manually using a small plastic pestle cooled to liquid nitrogen temperature for 45 seconds in a 1.5 ml microtube. Per 100 uL of tissue (2 leaf disks), 100 uL 4x SDS loading dye supplemented with 100 mM DTT was then added. Samples were vortexed vigorously to mix before incubation for 10 minutes shaking at 65 °C and a second centrifugation step at

21k RCF for ten minutes. The supernatant was transferred to a second tube, avoiding the pellet, before a second centrifugation step at 21k RCF for three minutes. Finally, samples were diluted in water and additional 4x SDS loading dye ready for gel loading.

2.6 SDS-PAGE, BN-PAGE, and western blot

2.6.1 SDS-PAGE

SDS-PAGE was performed using pre-cast gels such as 4-20 % Tris-Glycine (NuSep), 4- 20 % gradient RunBlue TEO-Tricine SDS mini gels, 12 % RunBlue SDS gels, 16 % RunBlue SDS gels (Expedeon), or with 15 % gels cast using the TGX FastCast™ Acrylamide kit (Biorad) according to manufacturer's instructions. Samples were mixed with RunBlue 4 x LDS sample buffer supplemented with 10 mM DTT, heated to 95 °C for ten minutes, loaded into gels and run at 120 – 180 V until the dye running front reached the bottom of the gel. Molecular weight standards were included on each gel, typically Abcam extra broad molecular weight pre-stained standard (ab116029). For protein visualisation, gels were stained using Coomassie instant blue stain (such as ReadyBlue™ Protein Gel Stain, Sigma-Aldrich), or otherwise were left un-stained and were immediately processed for western blotting as described below.

2.6.2 Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE)

BN-PAGE was performed according to a protocol developed by Dr Hee-Kyung Ahn, Dr Lida Derevnina, and Dr Mauricio Contreras, and performed with assistance from Josh Bennett. Leaf disks were ground into fine powder using a plastic micro-pestle at liquid nitrogen temperature, before addition of extraction buffer (50 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM MgCl₂, 10% v/v glycerol, 0.2 % v/v Triton X -100, 1 % DTT, 1 % protease inhibitor cocktail (Sigma)). The samples were incubated on ice with extraction buffer for ten minutes with vortexing every two minutes, before centrifugation for 15 minutes at 4 °C at 21k RCF. 25 µL of supernatant was mixed with 12.5 µL NativePAGE 4x Sample Buffer (Invitrogen), 1.25 µL NativePAGE G-250 additive (Invitrogen), and 11.25 µL water, and 5 µL per lane was loaded into the wells of a 4-20 % Mini-PROTEAN® TGX™ Precast Protein Gels (BioRad). 2.5 µL of Native PAGE marker (SERVA Native marker, liquid mix) was included as a standard on every gel. The BN-PAGE was performed using anode buffer Invitrogen™ NativePAGE™ Running buffer (20x) diluted with ultra-pure water to 1x, and dark cathode buffer (1 x Anode buffer + 1 x Invitrogen™ NativePAGE™ Cathode Buffer Additive). After running for 40 minutes at 150 V, the dark cathode buffer was exchanged for light cathode

buffer (1 x Anode buffer + 0.1 x Invitrogen™ NativePAGE™ Cathode Buffer Additive) and 200 V was applied for a further 90 minutes.

Western blot of BN-PAGE samples was performed using PVDF membranes in a Trans-Blot® Turbo™ semi-dry transfer system (Biorad) at 25 V, 1 A, for 30 minutes. The PVDF membrane was then incubated in 8 % v/v acetic acid for 15 minutes before rinsing thoroughly with ultra-pure water and allowing to dry. The dried PVDF membrane was then rehydrated with 100 % ethanol, revealing the position of the bands of the marker, which were then marked with a ball pen. The membrane was then rinsed with TBS-T buffer (Tris-buffered Saline + Tween, made up of 50 mM Tris-HCl, 150 mM NaCl, 0.1 % Tween-20, pH 8.0) to remove ethanol before continuing the western blotting protocol from the point of blocking buffer (as outlined in the section below, 2.6.3 Western Blot)

2.6.3 Western Blot

After SDS-PAGE, protein was transferred from the gel onto a polyvinylidene difluoride (PVDF) membrane by a Trans-Blot® Turbo™ semi-dry transfer system (Biorad) using the high molecular weight protocol (10 minutes, 25 V, 1.3 A). The membrane was immediately placed into a blocking buffer of TBS-T + 5% milk (Tris-buffered Saline + Tween, TBS-T, is made up of 50 mM Tris-HCl, 150 mM NaCl, 0.1 % Tween-20, pH 8.0) and incubated for one hour at 6 °C. After blocking, the membrane was transferred to TBS-T +5% milk plus appropriate HRP-conjugated antibody, for 18 hour incubation at 6 °C. Following this, the membranes were washed in TBS-T with 3 washes of 15 minutes each, before being developed using ECL reagents (Abcam) and imaged using the ImageQuant LAS 500. Membranes were subsequently stained with Ponceau S solution to assess total protein loaded. Antibodies used were anti-HA-HRP (Thermo 26183-HRP), anti-FLAG-HRP (Generon, CPA9020), anti-StrepII-HRP (Merck, 71591-3), and anti-Myc-HRP (Santa Cruz Biotechnologies, 9E10 sc-40), anti-GFP-HRP (Santa Cruz Biotechnology, sc-9996). Inkscape software was used to adjust brightness and contrast of western blot images and assemble into figures.

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre- activated or activated Pik resistosome

3.1 Introduction

As discussed in the main introduction (1.5), over the five years since 2019, a number of resistosome structures have been reported. The first example of a CC-NLR resistosome structure was that of ZAR1 in 2019 (Wang et al., 2019a; Wang et al., 2019b) and since then, a number of CC-NLR structures from monocots have been resolved including the pentameric wheat Sr35 resistosome (Förderer et al., 2022), and the barley MLA13 heterodimer (Lawson et al., 2024). However, to date, neither a structure of a paired monocot CC-NLR resistosome, nor a structure of a CC-NLR with an integrated HMA domain located between the CC and NB-ARC domains have been reported. Understanding the spatial arrangement of the pre-activated or activated Pik resistosome could provide information useful when considering Pik engineering approaches but requires a supply of purified protein of sufficient concentration and purity. This context motivated our work to express and purify the Pik-1/Pik-2 resistosome in a pre-activated or activated state.

3.1.1 Challenges in obtaining resistosome structures

The challenges in obtaining resistosome structures are varied and numerous. Firstly, resistosome complexes may be of low abundance in their native environment, and may not easily accumulate to high levels, precluding direct purification. This is because in the plant cell there may not be a requirement for a large number of active complexes to transduce the signals for cell death or to produce cell death directly. As a consequence, overexpression systems have been used to drive higher levels of NLR protein expression, but often relatively low abundance necessitates purification from a very large amount of tissue. For example, the TIR-NLR ROQ1 resistosome was purified from around 200 g of *N. benthamiana* tissue which is a large amount of starting material.

Second, NLRs cause cell death and so this must be mitigated to maximise accumulation of NLRs. TIR-NLR protein expression in *N. benthamiana* has frequently been performed in the mutant background *eds1-1*, which lacks downstream signalling machinery and therefore

does not produce cell death. In the case of CC-NLRs, there are now multiple reports supporting the mechanism that the CC-NLR resistosome causes cell death directly through formation of a Ca^{2+} channel. Therefore, to abolish cell death without preventing the oligomerisation events that are signatures of active complexes, CC-NLR mutants must be employed, or the channel must otherwise be inhibited.

Third, these complexes may require particular chaperones or post-translational modifications during expression to stabilise the complex. Therefore, prokaryotic expression systems may not be able to produce correctly folded NLR proteins. In the scientific literature there are a number of reports which highlight the difficulties and compromises involved in expression and purification of NLR proteins.

Previous attempts to express full length NLRs in *E. coli* have typically not been successful, although study of domains is possible. Dr John Steele was able to express and purify the NRC1 NB-ARC domain in *E. coli* in quantities suitable for analytical SEC, and he also worked on the LRR domain of R3a using *E. coli* expression. Dr Richard Hughes carried out expression screening at Oxford expression technologies for a panel of constructs of full length, CC, NB-ARC, and LRR domains of the NLRs R3a, R2, I-2, Rpi-blb2, NRC1 and NRC1-like. Expression was attempted in both BL21 *E. coli* and Sf9 insect cells, but out of the constructs that did produce expression, none were uniquely expressed in insect cells as these constructs also could also be expressed in *E. coli* (Doctoral Thesis of John Steele). Previous work in the Banfield Lab was not successful in producing the Pik NLRs in *E. coli*.

The difficulty in producing NLR proteins in *E. coli* also extends beyond the study of plant NLRs. For example, due to difficulties in expression and purification of the mammalian NLR NOD1 in *E. coli*, instead mammalian cell culture was used (Askari et al., 2012). In a second example of mammalian NLR expression, NOD2 was expressed in Sf9 culture, purified, and a crystal structure was obtained (Maekawa et al., 2016). This did provide information into interdomain contacts, but extensive truncations were required to facilitate the crystallisation, including the entire caspase-activation and recruitment domain (CARD) domain. Since the CARD domain is required for activation, this meant that this structure was obtained in an auto-inhibited state. This highlights the difficulty in crystallising large protein complexes and illustrates why cryo-EM is the preferred approach for structural studies of full length NLR complexes.

3.1.2 Full length Pik paired NLRs as a target for structural studies

Despite past difficulties in producing purified Pik NLRs from *E. coli* in quantities suitable for structural biology, the Pik NLR remains an attractive target for structural studies. First, as mentioned above, a structure of a pre-activated or activated Pik NLR would be the first structure of a paired CC-NLR resistosome and would give insight into the stoichiometry and organisation of a paired resistosome. The full-length structure of a Pik resistosome would also provide an example of an integrated domain NLR resistosome and so would provide information on how effector binding to an integrated domain may be able to activate the NLR.

Second, purification of the complex is feasible from a technical standpoint. Upon *Agrobacterium* mediated overexpression, the Pik pair can accumulate in *N. benthamiana* to a level sufficient for co-immunoprecipitation, and a previous study provides evidence for Pik pair hetero-complexes. Zdrzałek et al. observed that Pikp-1/Pikp-2 interact via their CC domains and were co-immunoprecipitated both in absence of an effector and also upon addition of effector AVR-PikD (Zdrzałek et al., 2020). From this, a complex would be expected, but the overall arrangement and stoichiometry of proteins within the complex remains to be determined. The recently published cryo-EM NLR structures demonstrate that insect cell culture and agroinfiltration of *N. benthamiana* are both proven methods of producing sufficient quantities of purified NLR for cryo-EM.

Finally, the technical expertise and instrumentation at the JIC would allow for negative stain and preliminary cryo-EM screening and data collection on site, which would improve the speed and feasibility of the process of optimisation of purification and sample preparation protocols.

3.1.3 Choice of Pik alleles for resistosome reconstitution

In previous work from the Banfield Lab, binding affinities between Pikm-HMA and AVR-Pik alleles were assessed by surface plasmon resonance (SPR), revealing the highest affinity interactions were with AVR-PikD, followed by AVR-PikE, AVR-PikA and AVR-PikC. The same experiment was repeated with Pikp-HMA, and the same order of binding affinity was found. AVR-PikD had a higher binding affinity to Pikm-HMA than to Pikp-HMA (De La Concepcion et al., 2018). In addition, crystal structures of Pikm-HMA bound to AVR-PikD, AVR-PikE, and AVR-PikA show that the largest interface between HMA and effector is

between Pikm-1/AVR-PikD, corresponding to the highest affinity interaction observed. Therefore Pikm/AVR-PikD was selected as a candidate for structural studies.

It might be possible that the sensor NLR Pikm-1 only acts to activate the helper NLR Pikm-2, which then goes on to form a resistosome alone, without the sensor. However, there is evidence that Pikp-1/Pikp-2 form heterocomplexes with and without the presence of AVR-PikD in *N. benthamiana* (Zdrzałek et al., 2020). Therefore I considered it likely that the Pikm-1/Pikm-2 alleles would interact in a similar manner. Taken together, these two pieces of previous work informed the selection of the Pikm-1/Pikm-2 and Pikm-1/Pikm-2/AVR-PikD complexes as targets for expression, purification, and structural studies.

3.1.4 Mitigation of resistosome mediated cell death in heterologous expression systems

Reduced viability of insect cell culture has been reported upon co-expression of CC-NLR resistosome components (Förderer et al., 2022). This is consistent with the proposed mechanism of CC-NLR mediated cell death; the direct formation of Ca^{2+} channels in the plasma membrane induces cell death independently of specific downstream signalling and therefore can still function in a heterologous expression system. Cell death is counterproductive to accumulating as much protein as possible; in other studies this has been mitigated by careful selection of timepoint (collect samples or take measurements before cell death progresses), by activation of the resistosome in vitro (Wang et al., 2019a), creation of cell-death-null mutants of the CC-NLR, or by other means of cell death inhibition.

Förderer et al. used mutations of the EDVID motif and LRR^{R-cluster} to abolish cell death in Sf21 cells, however as these are conserved oligomer stabilisation motifs, they are inappropriate mutants to use for biochemical isolation of an oligomeric resistosome (Förderer et al., 2022). Mitigation of ZAR1 mediated cell death through mutation of the $\alpha 1$ helix has been well documented; the Glu11 on the inner surface of the funnel was demonstrated to be important for the function of the channel, and mutations F9A/L10A/L14A on the outer surface were also found to reduce cell death (Wang et al., 2019a; Bi et al., 2021). The equivalent residues of NRC4, L9A/V10A/L14A, were also demonstrated to have a similar effect in work by Adachi et al. (Adachi et al., 2019).

During expression of CC-NLR proteins in planta, the Ca^{2+} antagonist LaCl_2 has been used to inhibit the calcium channel activities of activated CC-NLR resistosomes and thereby allow

accumulation of these proteins in their activated form without the accompanying cell death (Hu et al., 2020). This circumvents the need to inhibit resistosome activity through NLR mutations. It would be very convenient to use a Ca^{2+} channel inhibitor in insect cell culture to achieve accumulation of activated wild-type NLRs without cell death. However, insect cells are cultured in phosphate buffered media; upon addition of LaCl_2 , insoluble lanthanum phosphate precipitates form. Therefore, LaCl_2 could not be explored as a means of abolishing CC-NLR induced cell death in insect cell culture and instead CC domain mutants were used.

3.2 Methods

3.2.1 SpyTag/SpyCatcher

3.2.1.1 His-SpyCatcher purification from *E. coli*

6His-SpyCatcher was expressed in BL21 *E. coli* in 3 L of auto-induction media as described previously (2.2.3, *Escherichia coli* cell culture for protein expression). The cultures were incubated at 37 °C for 24 hours before harvest. The protein was purified by IMAC as described previously (2.2.4 Protein purification from *E. coli*) but without any tag cleavage or subsequent gel filtration steps. The protein was concentrated in a centrifugal concentrator to a final concentration of 10 mg/ml and was cooled in buffer A4 in 1 mg aliquots.

3.2.1.2 Small scale SpyTag purification from *E. coli*

SpyTag-3C-Myc-GFP was expressed in BL21 *E. coli* in 3 L of auto-induction media as described previously (2.2.3, *Escherichia coli* cell culture for protein expression). The cultures were grown to an OD₆₀₀ = 1.3 and then incubated 18 °C for 20 hours before harvest. All steps of the purification were performed at 4 °C. Clarified SpyTag-3C-Myc-GFP *E. coli* lysate was prepared as described previously (2.2.4 Protein purification from *E. coli*) and collected in a 100 ml glass bottle. Ni-NTA + His-SpyCatcher resin was prepared as follows: 500 µl bed volume of Ni-NTA resin was washed by addition of 50 ml of water, centrifugation at 900 RCF, removal of the supernatant, and addition of 50 ml of buffer A4. 5 mg of His-SpyCatcher was then added and incubated on a roller for 20 minutes. Following this incubation, the resin was separated by centrifugation at 900 RCF and the supernatant removed (containing any unbound His-SpyCatcher). The resin was then incubated with the clarified SpyTag-3C-Myc-GFP *E. coli* lysate for two hours, before retrieval of the resin using a gravity flow column. The resin was then washed with 10 ml of buffer A1 repeated three times, then 10 ml of buffer A4 repeated five times. The resin appeared very green, indicating strong accumulation of GFP. The resin was then resuspended in a 1.5 ml microtube and 40 µg of 3C protease added and gently mixed, before incubation at 21 °C for 90 minutes. The cleaved protein was then collected in a 1 ml elution followed by three successive 0.5 ml elutions. Following this, the resin still exhibited strong GFP fluorescence under UV illumination, and so a further 40 µg of 4C protease was added and incubated for 18 hours at 4 °C before collection of a final 2 ml elution.

3.2.1.3 Medium scale SpyTag affinity purification from *N. benthamiana*

Preparation of Ni-NTA + His-SpyCatcher resin was performed using a 250 µL bed volume of Ni-NTA resin and 5 mg of His-SpyCatcher, resuspended into AGHN buffer (250 mM

NaCl, 100 mM HEPES pH7.5, 5 mM MgCl₂, 10 % glycerol (v/v)) supplemented with 0.1 % (v/v) NP-40, as described previously (3.2.1.2 Small scale SpyTag purification from *E. coli*). 20 g of *N. benthamiana* leaves expressing GFP-FLAG-3C-SpyTag were collected and lysed in lysis buffer (250 mM NaCl, 100 mM HEPES pH7.5, 5 mM MgCl₂, 10 % glycerol (v/v), 0.1 % (v/v) NP-40, 2 % (w/v) Polyvinylpyrrolidone (PVPP), 10 mM DTT, 1 x cOmplete protease inhibitor tablet (1 tablet per 50 mL)) using a blender (Waring) operated in 10 second pulses with 5 second pauses between pulses. The lysed mixture was incubated for twenty minutes at 4 °C before filtering using Miracloth (Millipore) followed by centrifugation at 4000 RCF for 10 minutes at 4 °C. The supernatant was subjected to a second centrifugation step at 50 k RCF for one hour at 4 °C and the clarified lysate incubated with the prepared Ni-NTA + His-SpyCatcher resin. After a two-hour incubation, the resin was retrieved using a gravity flow column and washed with 150 ml AGHN buffer supplemented with 0.1 % (v/v) NP-40, followed by 100 ml AGHN buffer. The resin was collected into a 1.5 ml microtube and 140 µg of 3C protease added before incubation rolling slowly at 4 °C for 18 hours. The cleaved protein was collected in two 1 ml elutions followed by an 0.5 ml elution, and finally buffer B1 (containing imidazole) was used to liberate any remaining complex on the resin. The same approach was used in the attempted isolation of SpyTagged AVR-PikD from *N. benthamiana* lysate.

3.2.2 Expression and purification of the Pik pair in *N. benthamiana*

3.2.2.1 Reconstitution of p35S:Pikm-1-HF + pMAS:Pikm-2-HA using the Pik chassis

The wild-type Pikm-1/Pikm-2 sequences were reconstructed in the Pik chassis acceptor vector by a golden gate reaction. The acceptor vector was developed by Dr Adam Bentham and Dr Mark Youles and contains a mRFP selection cassette replacing amino acids I184 – E263 of the wild-type Pikm-1 sequence. A Pikm-HMA domain expression construct was amplified by PCR using primers that introduced a BsaI restriction site at either flank. Upon digestion of the acceptor vector and HMA insert, CAGA and GATG overhangs were revealed, which facilitated the incorporation of the HMA insert into the chassis, reconstituting the wild-type Pikm-1 sequence.

3.2.2.2 Domestication of Pikm-1 and Pikm-2 genomic sequences

The genomic sequences of the Pikm pair were obtained from the NIH Genebank AB462256.1 (Ashikawa et al., 2008). The sequences were domesticated by elimination of BpiI restriction sites by introduction of synonymous mutations at amino acids D228, V243, L716, and D990 of Pikm-1 and K185, K283, and K486 of Pikm-2. In addition, two BpiI

restriction sites were eliminated from the second intron of Pikm-1, at nucleotide positions 2143 and 2164. The sequences were divided into segments for synthesis; Pikm-1 was synthesised in six sections; four large gene fragments, and one smaller gene fragment and one pair of annealed primers, which were combined to reconstruct a difficult to synthesise region (containing two 10 x poly T regions) which could not be ordered in one fragment. Pikm-2 was divided into two gene fragments. The gene fragments were flanked with BpiI restriction sites, which upon digestion reveal complementary 4 nucleotide overhangs which allow for scarless assembly of the gene fragments into a level 0 Golden Gate acceptor vector using a BpiI mediated Golden Gate reaction.

3.2.3 Insect cell culture

3.2.3.1 Maintenance of Sf9 cell culture

Suspension cultures of Sf9 insect cells were maintained by incubation in plastic Erlenmeyer flasks or plastic T75 cell culture flasks in standard conditions of 26 °C, shaking at 120 rpm. Lids were tightly closed if vented, otherwise lids were left a quarter turn open to allow for some gas exchange. The culture was monitored three times weekly and sub-cultured to maintain the cell density of the culture between 0.5 - 2.0 x 10⁶ cells/ml. Cells were counted by staining with Trypan Blue and using an automated cell counter (Countess II FL, Thermofisher) to quantify live and dead cell counts. Initially, cells were grown without addition of antibiotics, but upon repeated difficulties with contamination, 1 % v/v Penicillin-Streptomycin (Life Technologies Ltd) was added to the media. Culture health was monitored and upon signs such as slowing of division, clumping cells, or contamination, a new vial of cryopreserved stock was retrieved and recovered to replace the old cultures. Sf9 cells were a gift from the Carter Lab (MRC LMB, Cambridge), and from the Maxwell Lab (JIC).

3.2.3.2 Cryopreservation of Sf9 cells

Before cryopreservation, the cell population was expanded by growing in a larger flask and growing to a density of 2 x 10⁶ cells/mL. The cells were centrifuged at 100-200 RCF for 5 minutes at room temperature before gentle resuspension in cryopreservation media (80 % v/v SF900 II media, 10 % v/v fetal bovine serum (FBS), 10 % v/v DMSO) to give a final cell count of 1 x 10⁷ cells/ml, which is 10 x the usual growing concentration. The cells were aliquoted into labelled 1.5 mL or 2 mL cryovials and cooled slowly to -70 °C, ideally in a controlled rate cooling device (E.g. Mr. Frosty Nalgene™ Cryo 1°C Freezing Container) or

otherwise in a sealed polystyrene shell. Once cooled, the cell vials can be transported on dry ice to be stored in the vapor phase of a liquid nitrogen cell storage Dewar.

3.2.3.3 Recovery of cryopreserved Sf9 cells

Before warming the cell vial, a flask containing media at 26 °C was prepared. The cell vial was warmed in a 37 °C water bath with gentle agitation until almost thawed. Then, the vial was removed from the water bath, the outside thoroughly cleaned with ethanol, and the contents added to the flask. The final cell density in the newly seeded flask should be 1×10^6 cells/mL. The cells were monitored daily and when they reached a density of $2-3 \times 10^6$ cells/mL and viability above 95%, they were subcultured as detailed above.

3.2.3.4 Bacmid E.coli transformation, selection, and bacmid purification

Expression constructs in GoldenBac vectors carrying gentamycin resistance (2.1.1) were transformed into DH10EMBacY *E. coli* (Geneva Biotech). GoldenBac entry vector pGB and destination vector pGB-dest carry Gentamycin resistance, and DH10EMBacY *E. coli* carry kanamycin and chloramphenicol resistance in its genome and tetracycline resistance on the T7 transposase helper plasmid. The transformed DH10EMBacY *E. coli* were allowed to recover for at least three hours at 37 °C to allow time for Tn7 transposition to occur, and then plated onto LB agar plates with 50 µg/mL kanamycin, 7 µg/mL gentamicin, 10 µg/mL tetracycline, 100 µg/mL X-gal, and 40 µg/mL IPTG. Successful Tn7 transposition inserts the pGB or pGB-dest plasmid into the baculovirus genome at a site which disrupts a copy of lacZ thereby allowing for blue/white selection. White colonies indicating positive transformants were selected and 5 ml liquid cultures were grown at 37 °C for 16 hours in 2x YT supplemented with 50 µg/ml kanamycin, 7 µg/ml gentamycin, and 10 µg/ml tetracycline to maintain selection. In case of unclear blue/white selection, plates were incubated for one more day, or picked colonies were streaked out onto bacmid selection plates before inoculation to observe for later development of blue coloration while still being able to proceed with bacmid isolation.

Bacmid DNA was purified by alkaline lysis followed by isopropanol precipitation. Alkaline lysis and neutralisation was performed using the reagents from the NucleoSpin® Plasmid isolation kit (Macherey-Nagel), however as the bacmid DNA is large and fragile, the clarified *E. coli* lysate was not applied to the affinity spin columns of the kit and instead was subjected to isopropanol precipitation using 100% isopropanol. After a 10-minute incubation on ice, the DNA was pelleted by centrifugation at 21 k RCF at room temperature. The supernatant was carefully removed, and the translucent pellet was washed three times by gentle application of 70% ethanol at room temperature, inversion, and centrifugation at 21 k RCF for 5 minutes at 4 °C. Finally, the bacmid DNA pellet was air dried for one minute and

dissolved in sterile filtered elution buffer (5 mM Tris/HCl, pH 8.5). Bacmid DNA concentration was assessed by Nanodrop, bacmids were diluted to 1 µg/µL using sterile filtered elution buffer, stored at 4 °C, and used within two weeks.

3.2.3.5 Sf9 transfection and virus production

To produce P1 virus, first Sf9 cells were diluted to 0.5×10^6 cells/mL and 2 ml was dispensed into each well of a six-well plate and allowed to settle and attach to the well. Even if fewer than six-transfections were being performed, each well should contain 2 mL of cell suspension or media to help maintain appropriate humidity. Ideally one well should be left un-transfected as a control to compare cell growth. While cells were attaching, a mixture of 2 µg of DNA and 6 µL FuGene HD transfection reagent (Promega) in 200 µl media was prepared in a microcentrifuge tube. After incubation for 15 minutes, the transfection mixture was added dropwise to the cells, and the transfection plates were incubated at 26 °C in a static incubator in a sealed plastic box in the presence of a damp paper towel to maintain humidity and prevent evaporation of media from the wells. After 2-3 days, YFP fluorescence was visible within transfected cells and after 3-5 days the supernatant was taken and added to liquid Sf9 cultures at $1-2 \times 10^6$ cell/ml density in a 1:100 dilution to generate P2 virus. After a further 3-5 days, the P2 viral supernatant was harvested by centrifugation and stored at 4 °C in the dark, and the cell pellets discarded. The P2 cell pellets can also be used for western blotting as an early indication of protein expression, with the caveat that the timepoint is much later than the timepoint used for protein expression, so the cells are at a later stage of infection and proteins may have degraded. For protein expression, 1 ml of P2 virus was added per 100 ml of cell culture at 2×10^6 cells/mL and the cultures incubated at 26 °C shaking at 120 rpm. After incubation for 48-66 hours cells were harvested by centrifugation, the supernatant discarded, and the cell pellets snap cooled in liquid nitrogen and stored at -80 °C until required for further processing.

3.2.3.6 Preparation of Sf9 cell samples for SDS-PAGE and western blotting

Cell pellets were thawed with 1:4 pellet: lysis buffer ratio (e.g., 50 µL of cell pellet plus 200 µL buffer). Lysis buffer was composed of 50 mM HEPES pH 7.4, 100 mM NaCl, 1 mM DTT, 10% v/v glycerol plus addition of 1 cOmplete protease inhibitor tablet (Roche) per 50 ml. Buffer was aliquoted and stored at -20 °C. Samples were pipetted gently to resuspend into a 1.5 ml microtube, and a plastic micro-pestle used on each sample in turn, for 30 seconds each, on ice. Lysates were centrifuged at 21 K RCF at 4 °C for 45 minutes in benchtop centrifuge. Avoiding the pellet, the clarified lysate was transferred to a new tube, and SDS-PAGE samples prepared using 4x LDS dye in a 1:1 dye: lysate ratio.

3.2.3.7 Affinity purification of the Pik pair from Sf21 cells

Purification of the “inactivated” His-SUMO- Pikm-1 + Pikm-2-FLAG complex was performed as follows. 5.3 g of cell pellet was resuspended in 50 mL of lysis buffer (50 mM HEPES pH 7.4, 300 mM NaCl, 5% v/v glycerol, 0.1% TritonX-100, plus addition of one cOmplete protease inhibitor tablet). Following lysis by sonication, the lysate was clarified by centrifugation at 20 K RCF at 4 °C for 45 minutes. The supernatant was retained and 750 µL bed volume of Ni-NTA resin was added to it and incubated rotating at 4 °C for 90 minutes. The resin had been pre-equilibrated in wash buffer (50 mM HEPES pH 7.4, 300 mM NaCl, 5% v/v glycerol). After incubation, the resin was then retrieved by gentle centrifugation at 100 RCF for 3 minutes and was transferred to a gravity flow column, where it was washed with wash buffer followed by addition of 4 ml volume of elution buffer (50 mM HEPES pH 7.4, 300 mM NaCl, 5% v/v glycerol, 200 mM imidazole). The eluted protein was collected immediately (E₁) before addition of 3 mL elution buffer and transfer of beads back into a tube suitable for incubation rotating at 4 °C. After 10 minutes, eluted protein (E₂) was collected by centrifugation. E₁ and E₂ were pooled, and anti-FLAG resin (pre-equilibrated in wash buffer) was added and incubated together rotating at 4 °C for 2 hours. Following the incubation, the resin was washed three times by addition of 5 ml wash buffer, resuspension, and centrifugation at 100 RCF. Following this, 300 µL elution buffer (50 mM HEPES pH 7.4, 300 mM NaCl, 5% v/v glycerol, 1 mg/mL FLAG peptide) was added and incubated together with the resin rotating at 4 °C for 14 hours. The eluted protein was then collected and a second 300 µL FLAG elution (E₂) was performed for a duration of approximately 2 hours. The 300 µL of FLAG E₁ was then gel filtered over a Superose 6 increase 10/300 column, run at 0.25 ml/min. The gel filtration buffer was 50 mM HEPES, 150mM NaCl, 3% Glycerol. SDS-PAGE was performed on elution fractions.

Purification of the “activated” His-SUMO- Pikm-1 + Pikm-2-FLAG + AVR-PikD-TwinStrep complex was performed as for the “inactivated” complex but with the following differences. The second affinity purification step used Strep-Tactin resin, and the resin was incubated with the Ni-NTA elutions for 16 hours before three washes as performed above. Three successive elutions were then performed by incubating 300 µL Strep elution buffer (50 mM HEPES pH 7.4, 300 mM NaCl, 5% v/v glycerol, 50 mM biotin) with the resin while rotating at 4 °C for 60 minutes.

3.2.3.8 Negative stain electron microscopy

Carbon film copper grids (Electron Microscopy Sciences # CF400-CU-50) were glow discharged (25A, 45 seconds, PELCO EasiGlow) and protein samples were applied to the grid without dilution. 6 µL of sample were applied to the grid and incubated for 60 seconds

before wicking away excess sample from the edge of the grid using filter paper. The grids were then stained with 1 % uranyl acetate for 60 seconds before again wicking away excess stain from the side of the grid using filter paper. While blotting, grids were viewed from an oblique angle so that blotting could be stopped before absolute dryness. Grids were then air dried for one minute before storage in a grid box. Representative micrographs were acquired using a Hitachi HT-7800 TEM operating at an accelerating voltage of 100 kV and fitted with an EMSIS XAROSA camera. Negative stain EM grid making was carried out by Dr Nitika Mukhi and myself, and visualisation by Dr Nitika Mukhi.

3.2.3.9 Cryo-electron microscopy

Grids were prepared using a Vitrobot Mark IV plunge-freezer (Thermofisher) with the chamber set to 22 °C with 100% humidity. 4 µL of protein sample was applied to carbon coated or gold grids before blotting for between 2-3 seconds and then plunge cooling in liquid ethane. Grids were stored under liquid nitrogen until visualisation using a Titan Krios G3i (Thermofisher) electron microscope equipped with a Volta phase plate and operated at 300 kV. Cryo-EM grid preparation, screening, and preliminary processing was performed by staff of the University of Cologne Cryo-Electron Microscopy Platform.

3.3 Results

3.3.1 Expression in *N. benthamiana*

Expression of the Pik pair in *N. benthamiana* is well established in the Banfield Lab and yields sufficient protein for co-IP studies, but for structural work a larger amount of protein of higher purity is required. Therefore, in this section I set out to improve the expression and purification of the Pik pair to obtain quantities sufficient for structural studies.

3.3.1.1 Improving efficiency of the purification by using the high affinity tag system SpyTag/SpyCatcher

Spytag-Spycatcher is an affinity tag system developed from a surface protein from *Streptococcus pyogenes* (Keeble et al., 2019). The surface protein has been split into a small 13-amino acid SpyTag and a larger SpyCatcher module; upon association of the two parts, a spontaneous-isopeptide bond forms between a lysine side chain of SpyCatcher and an aspartate sidechain of SpyTag. Since the isopeptide bond is covalent, the resulting affinity of the interaction is very high. Dr Adam Bentham had suggested that this may be a good affinity tag system to use for purification of the Pik pair. Since the Pik pair is of low abundance even when overexpressed in *N. benthamiana*, a high affinity tag could help to concentrate low abundance proteins onto a relatively small amount of resin. The tag system has been demonstrated to work in *E. coli*, mammalian expression systems, and in planta to tag organelles for microscopy, but I could not find any examples of use of SpyTag/SpyCatcher for protein purification from *N. benthamiana* tissue.

By coating it in purified His-tagged SpyCatcher, Ni-NTA resin can be functionalised as “SpyCatcher resin” and used to capture SpyTagged protein from *E. coli*. As a proof-of-concept I first purified His-SpyCatcher, before functionalising Ni-NTA resin and carrying out a purification test with SpyTag-3C-Myc-GFP expressed in *E. coli* (3.2.1.2). GFP is very amenable to expression in prokaryotic and eukaryotic expression systems and as an additional benefit, can be followed through the stages of a purification by monitoring fluorescence. I was successful in capturing SpyTag-3C-Myc-GFP from *E. coli* lysate and in releasing a pure band of GFP after cleavage from the beads using 3C protease (Figure 3-1 A). I then progressed to expression of GFP-FLAG-3C-SpyTag from *N. benthamiana* (3.2.1.3). This proof of concept expression trial was somewhat successful; GFP-FLAG-3C-SpyTag was captured from *N. benthamiana* lysate, and again was cleaved from the resin by 3C protease, but at a lower efficiency than in the *E. coli* test case (Figure 3-1 B).

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

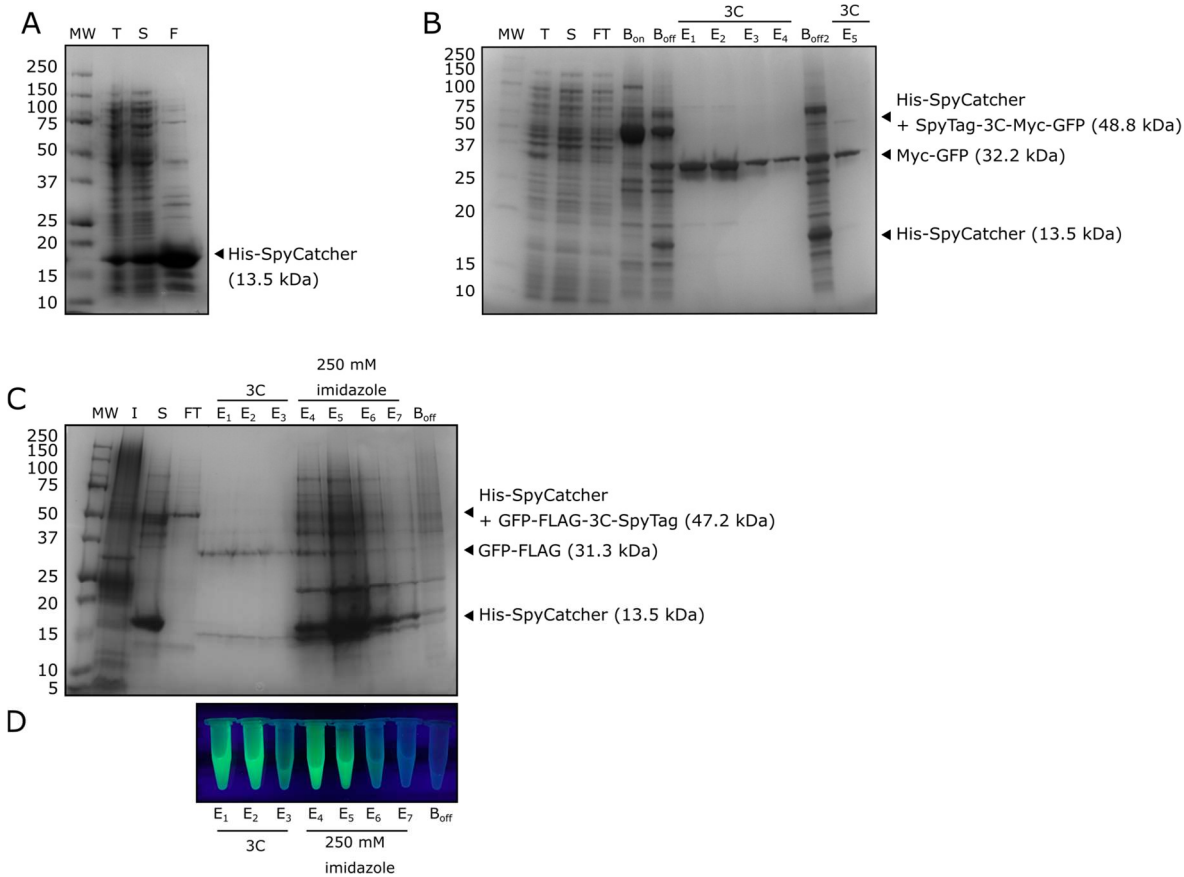


Figure 3-1: SpyTag/SpyCatcher proof-of-concept purifications from *E. coli* and *N. benthamiana*

A) SDS-PAGE of purification of His-SpyCatcher from *E. coli*. T = total lysate, S = soluble lysate, F = final purified protein after IMAC-gel filtration and concentration. **B)** SDS-PAGE of proof-of-concept purification of SpyTag-3C-Myc-GFP expressed in *E. coli* using SpyCatcher-Ni-NTA resin. T = total lysate, S = soluble lysate, FT = flowthrough after application to SpyCatcher resin, B_{on} = beads with protein bound, B_{off1-2} = beads after 3C cleavage, E₁₋₅ = sequential elutions using 3C protease. **C)** SDS-PAGE of Proof-of-concept purification of GFP-FLAG-3C-SpyTag expressed in *N. benthamiana* using SpyCatcher-Ni-NTA resin. I = insoluble material from *E. coli* pellet, S = soluble lysate, FT = flowthrough after application to SpyCatcher resin, E₁₋₃ = sequential elutions using 3C protease, E₄₋₇ = sequential elutions using 250 mM imidazole. B_{off} = beads after 3C cleavage and imidazole elution. **D)** photograph of eluted samples on the UV transilluminator; samples are aligned with the SDS-PAGE in C).

3.3.1.2 SpyTagged AVR-PikD presents a possible method for affinity purification of an activated resistosome from *N. benthamiana*

SpyTag-3C-Myc-AVR-PikD and AVR-PikD-FLAG-3C-Spytag were cloned for in planta expression under both the control of a pUbi10 promotor. A cell death assay in *N. benthamiana* revealed some cell death upon co-expression of SpyTagged AVR-PikD with

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

Pikm-1 and Pikm-2, implying that the SpyTagged AVR-PikD can still be perceived by Pikm-1 and Pikm-2. This suggests that SpyTagged AVR-PikD could be used to capture activated Pik resistosome complexes, and so a test purification of SpyTag-3C-Myc-AVR-PikD and AVR-PikD-FLAG-3C-Spytag was performed using the same method as for the GFP proof-of-concept purification (Figure 3-2).

SDS-PAGE revealed protein bands at a size corresponding to SpyTag-3C-Myc-AVR-PikD (Figure 3-2 A), although they could not be detected by anti-Myc western blot (Figure 3-2 C). SDS-PAGE bands corresponding to AVR-PikD-FLAG-3C-Spytag were also detected (Figure 3-2 B), and their identity confirmed by FLAG western blot. However, the signal was only detected on the samples derived from boiling the Ni-NTA beads to release the captured protein; no protein was detected to have been released by 3C protease treatment.

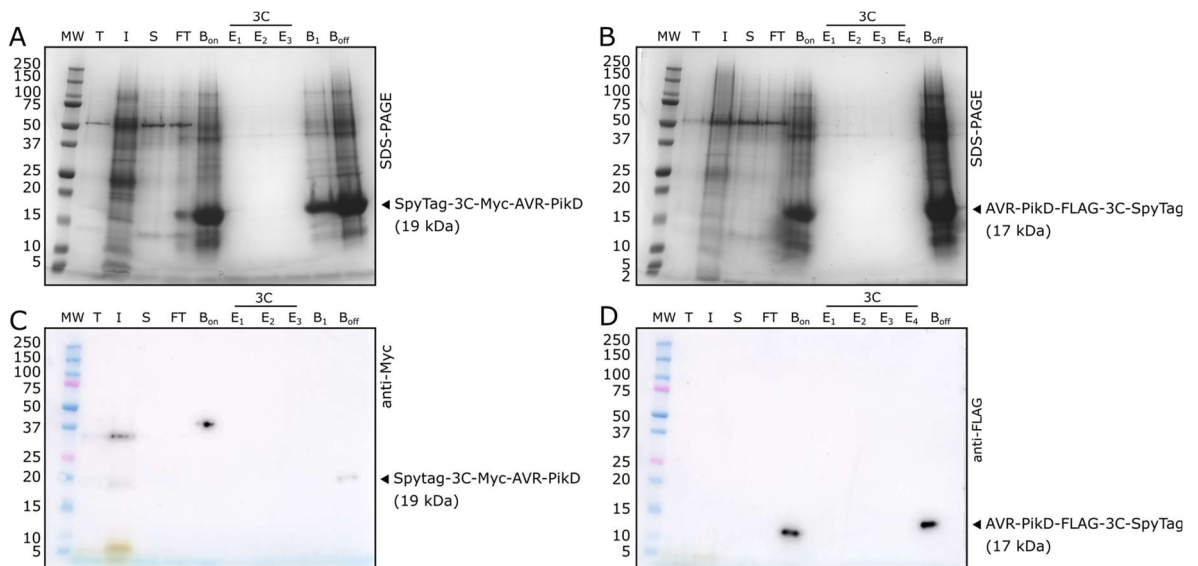


Figure 3-2: Medium scale purification of SpyTag-3C-Myc-AVR-PikD and AVR-PikD-FLAG-3C-Spytag from *N. benthamiana*

A) SDS-PAGE of purification steps of SpyTag-3C-Myc-AVR-PikD. **B)** SDS-PAGE of purification steps of AVR-PikD-FLAG-3C-Spytag. **C)** Anti-Myc western blot of SpyTag-3C-Myc-AVR-PikD purification. **D)** Anti-FLAG western blot of AVR-PikD-FLAG-3C-Spytag purification. T = total lysate, I = insoluble (cell pellet) fraction, S = soluble lysate, FT = flowthrough from binding to resin, B_{on} = beads with captured protein before elution, E₁₋₄ = elution from beads by 3C protease, B₁ = 250 mM imidazole elution, B_{off} = beads after 3C protease and imidazole elution.

To improve the purification of SpyTagged AVR-PikD, an attempt was made to drive higher expression levels by using the Geminivirus system. Geminivirus:SpyTag-3C-Myc-GFP (CS035) was selected as a Myc-tagged positive control as GFP fluorescence had previously been observed in leaves expressing this construct. pUbi:GFP-FLAG-3C-Spytag (CS009) was selected as a FLAG-tagged positive control for the same reason. Neither pUbi:SpyTag-3C-Myc-AVR-PikD (CS042) nor pUbi:AVR-PikD-FLAG-3C-Spytag (CS043) were detectable by western blot (Figure 3-3 A, B), despite some bands of a size corresponding to AVR-PikD-FLAG-3C-Spytag that were visible during the medium scale expression SDS-PAGE and western blot (Figure 3-2 B, D). Incorporation of the same genes and tags into the Geminivirus expression yielded an improvement in expression of Geminivirus:SpyTag-3C-Myc-AVR-PikD (CS050), with the highest band intensity observed 3 d.p.i, closely followed by the 4 d.p.i and reduced by 5 d.p.i. (Figure 3-3 A).

Much improved expression of Geminivirus:GFP-FLAG-3C-SpyTag (CS051) was also detected, with a gradually increasing expression from days 2-4 d.p.i to reach the peak expression on 4 d.p.i. and only slightly reduced by 5 d.p.i. (Figure 3-3 B). These constructs were then available for purification tests.

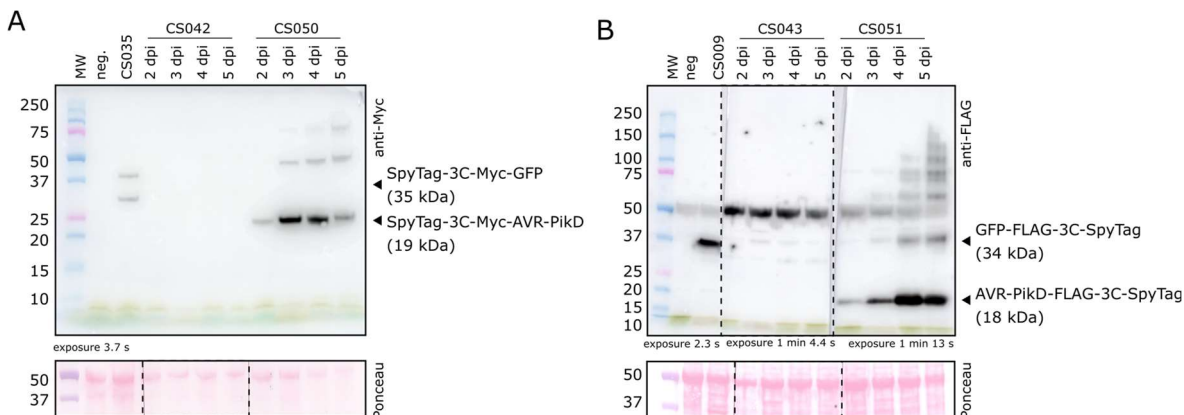


Figure 3-3: Expression testing of SpyTagged AVR-PikD in *N. benthamiana* under control of pUbi and Geminivirus expression system

A) Anti-Myc western blot of clarified *N. benthamiana* lysate expressing SpyTag-3C-Myc-AVR-PikD under control of pUbi (CS042) and the Geminivirus expression system (CS050). CS035 = positive control Geminivirus:SpyTag-3C-Myc-GFP. **B)** Anti-FLAG western blot of clarified *N. benthamiana* lysate expressing AVR-PikD-FLAG-3C-Spytag under control of pUbi (CS043) and under control of the Geminivirus expression system (CS051). Construct CS009 = positive control pUbi:GFP-FLAG-3C-Spytag.

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

3.3.1.3 Reconstitution of the Pikm-1-HF/Pikm-2-HA *N. benthamiana* expression construct

For simultaneous expression of the Pikm-1-HF/Pikm-2-HA pair from the same expression vector, the Pikm-HMA was inserted back into the Pik chassis by a golden gate reaction. The expression of the resulting construct, CS045, was then tested in *N. benthamiana* in comparison to the individual expression constructs pMAS:Pikm-1-HF:t35S (APB001) and pMAS:Pikm-2-HA:t35S (APB005). No bands of the expected size of 118 kDa were observed in the anti-HA western blot for Pikm-2-HA (Figure 3-4 B), but in the anti-FLAG blot bands corresponding to the Pikm-1-HF were detected only in the pMAS:Pikm-1-HF:t35S (APB001) samples and not in CS045 (Figure 3-4 A). Following this, I used APB001 and APB005 for expression of Pikm-1 and Pikm-2 in planta.

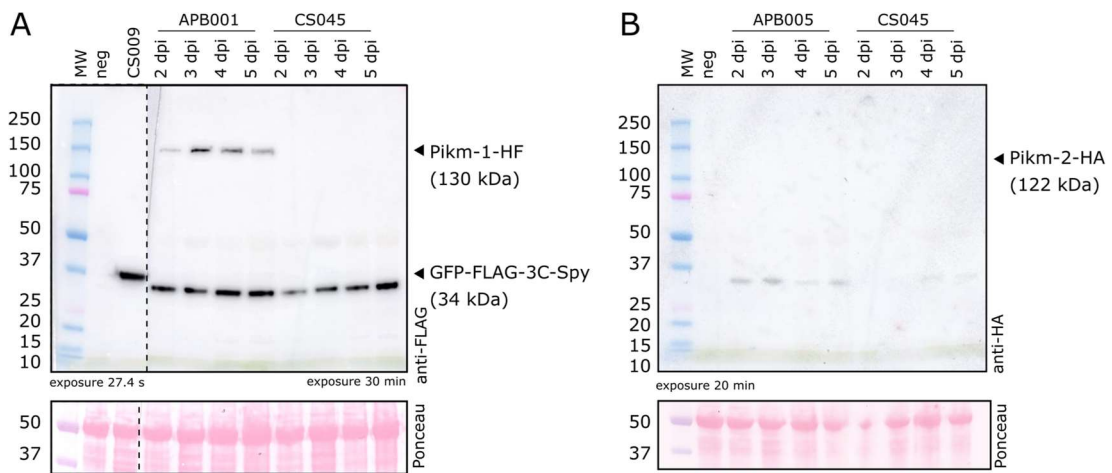


Figure 3-4: Western blot expression testing of the combined Pikm-1-HF and Pikm-2-HA construct CS045

A) Anti-FLAG western blot of pMAS:Pikm-1-HF (APB001) and p35S:Pikm-1-HF + pMAS:Pikm-2-HA (CS045) expression in *N. benthamiana*. CS009 = GFP-FLAG-3C-Spy positive control. **B)** Anti-HA western blot of pMAS:Pikm-2-HA (APB005) and p35S:Pikm-1-HF + pMAS:Pikm-2-HA (CS045) expression in *N. benthamiana*. Arrowheads indicate expected protein molecular weights.

3.3.1.4 Genomic pikm-1 and pikm-2 constructs do not enhance expression in *N. benthamiana*

The genomic sequences of *pikm-1* and *pikm-2* contain two and one introns, respectively. These introns were previously excluded from the sequence of *N. benthamiana* Pik NLR expression constructs. However, anecdotal evidence suggested that higher expression levels of NLRs can be achieved through use of expression constructs containing introns. This is consistent with reports that loss of introns can diminish gene expression, and that in plants in particular, expression enhancing elements can be found within introns (Jo and Choi, 2015).

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

The genomic sequences for Pikm-1 and Pikm-2 were domesticated by removal of BpII restriction sites, and the DNA synthesised and assembled into level 0 Golden Gate acceptor vectors before further cloning into plant expression vectors. The genomic expression vectors were expressed under control of a MAS promotor and 35S terminator in an exact parallel to the previously used expression vectors; the only difference was the inclusion of introns. A *N. benthamiana* cell death experiment was then performed to make a direct comparison between coding sequence and genomic expression vectors in response to AVR-PikD. Effector dependent cell death was observed with the CDS constructs, but not with the genomic constructs (Figure 3-5 A, B). *Nicotiana benthamiana* tissue infiltrated with these constructs was then analysed by western blot. Bands corresponding to Pikm-1-HF were detected in the coding sequence sample if expressed alone, but not when co-expressed with Pikm-2-HA. No band was detectable in the genomic Pikm-1-HF sample (Figure 3-5 C, D). From this, we concluded that the inclusion of introns does not enhance expression of Pikm-1 or Pikm-2 in planta.

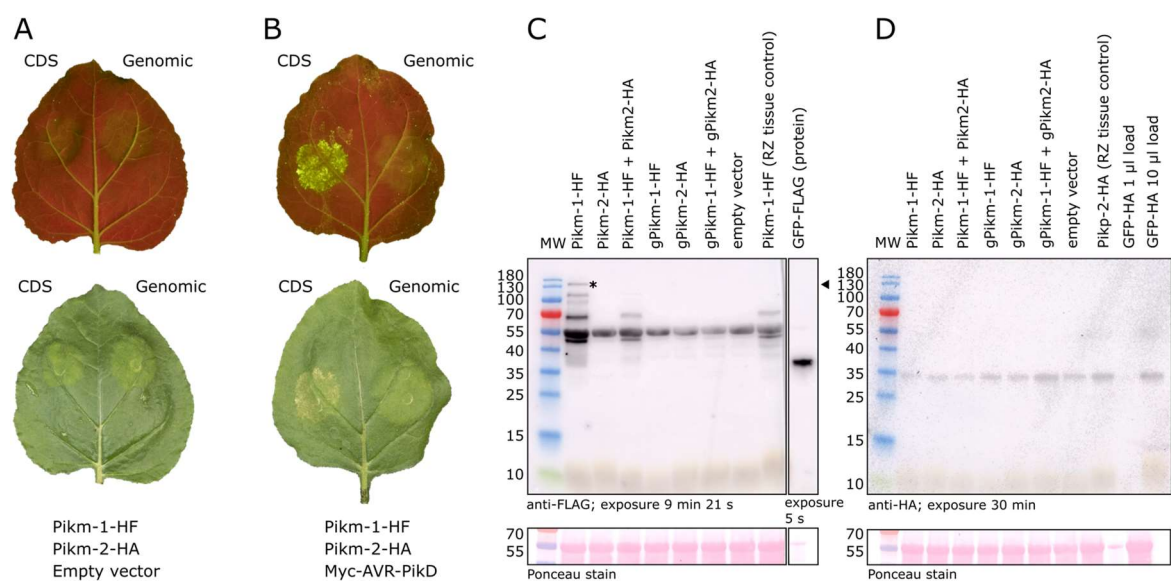


Figure 3-5: Expression testing of genomic Pikm-1-HF and Pikm-2-HA constructs.

Cell death assay in presence of **A)** empty vector control **B)** in presence of AVR-PikD. Left hand side, coding sequence, right hand side, genomic construct. Upper leaf: UV photograph of representative leaf, Lower leaf: white light photograph of same representative leaf, flipped along vertical axis. **C)** Anti-FLAG western blot of *N. benthamiana* leaf material. Pikm-1-HF and Pikm-2-HA = CDS construct. Pikm-2-HF tissue control from Dr Rafal Zdrzalek. GFP-FLAG control protein sample. Arrowhead indicates expected position of Pikm-1-HF band = 130 kDa. * indicates position of Pikm-1-HF. **D)** Anti-HA western blot of *N. benthamiana* leaf material. Pikm-1-HF and Pikm-2-HA = CDS construct. Pikm-2-HF tissue control from Dr Rafal Zdrzalek, GFP-HA control plasmid from Dr Indira Saado. Arrowhead indicates expected protein size of Pikm-2-HA – 118 kDa.

3.3.1.5 Expression feasibility studies with Leaf Expression Systems

The work presented in this section was outsourced to and performed by Leaf Expression Systems.

Agroinfiltration of *N. benthamiana* using the standard Banfield Lab vectors produces low to moderate levels of Pikm-1 and Pikm-2 expression, suitable for cell death assay and co-IP studies. However, protein purification for structural studies would benefit from higher expression levels. Viral vectors have been demonstrated to increase expression levels of other proteins and the SupraVec® vector is advertised to deliver a three-fold improvement in yield in comparison to the pEAQ-HT system, also known as Hypertrans®. Therefore, we commissioned Leaf Expression Systems to carry out an expression feasibility study of the Pikm-1/Pikm-2 pair in the Leaf Expression Systems proprietary SupraVec® pLES_543 vector. We targeted an inactivated complex/individual Pik proteins to avoid complications from cell death, and to allow future in vitro activation assays.

The gene sequences used for *N. benthamiana* expression of the Pik pair in the Banfield Lab were domesticated for Golden Gate cloning from the wildtype plant sequences but were not codon optimised. The workflow of Leaf Expression Systems includes synthesis of codon optimised gene sequences, which represented an opportunity to discover if codon optimisation might improve protein expression. For compatibility with the Leaf Expression Systems purification workflow, Pikm-1 was C-terminally tagged with 6His-FLAG and Pikm-2 was C-terminally tagged with StrepII. These are tags which have previously been successfully used by Leaf Expression Systems, and the C-terminus was chosen as C-terminal tagging of both Pik NLRs has been successful in co-immunoprecipitation experiments (Zdrzałek et al., 2020). The DNA synthesis company subcontracted by Leaf Expression Systems was unable to synthesise the Pikm-1 sequence; therefore, we opted to supply the coding sequence commonly used within the Banfield Lab. As a result, the expression testing was carried out with wildtype, non-codon optimised Pikm-1-6His-FLAG and codon optimised Pikm-2-StrepII, each within the pLES_543 vector. The vectors were transiently expressed in *N. benthamiana* individually and in combination, and tissue samples were harvested at 3- and 6 days post-infiltration. The tissue samples were probed by anti-His and anti-FLAG western blot. Although controls were detected, unfortunately neither Pikm-1-6His-FLAG nor Pikm-2-StrepII could be detected (Figure 3-6). This indicated that neither construct achieved overexpression and accumulation when expressed using the SupraVec® vector.

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

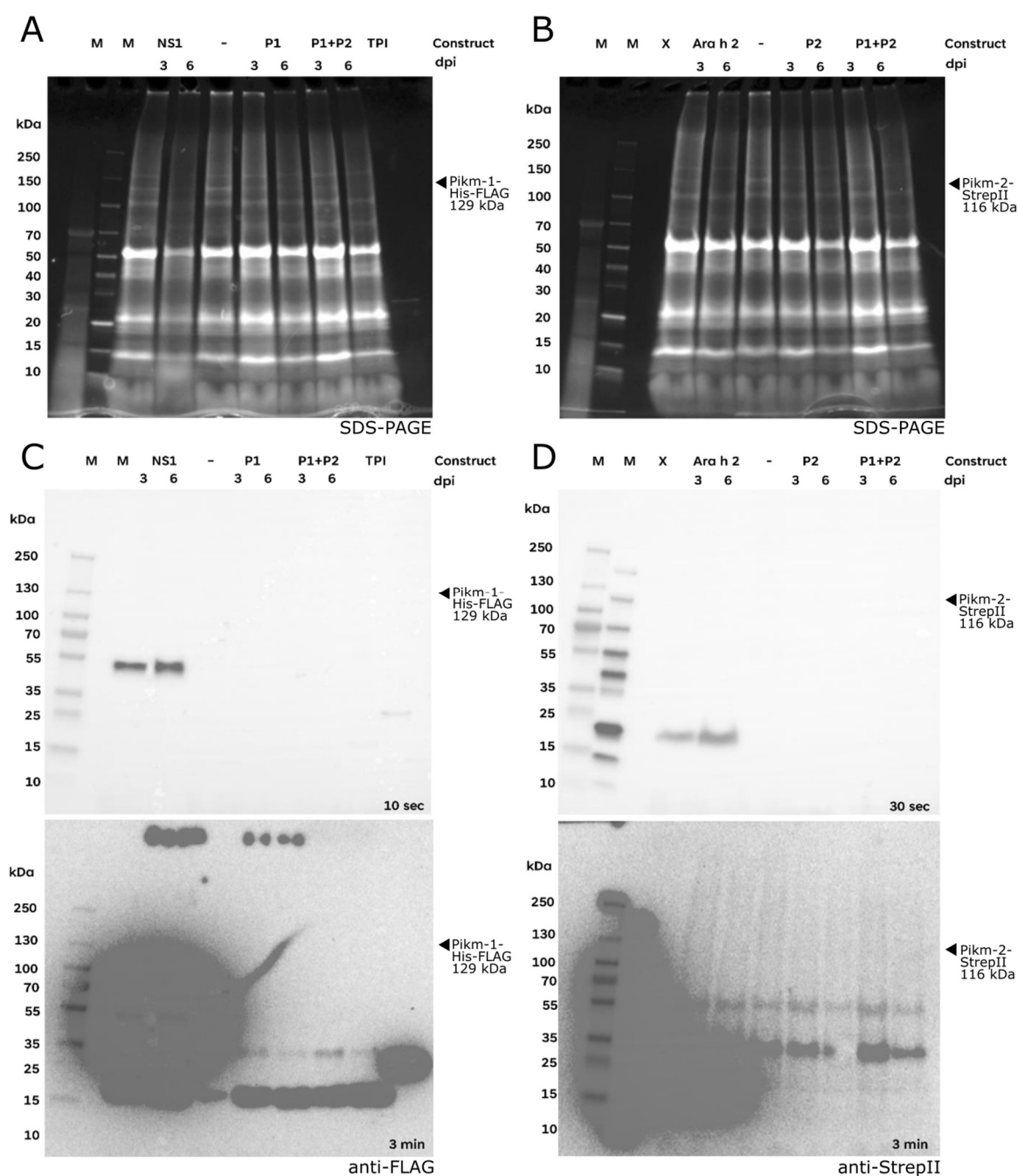


Figure 3-6: Expression trial of Pikm-1 and Pikm-2 in *N. benthamiana* by Leaf Expression Systems

A) SDS-PAGE of Pikm-1 constructs. **B)** SDS-PAGE of Pikm-2 constructs. **C)** Anti-His-HRP western blot of Pikm-1 constructs. **D)** Strep-Tactin®-HRP western blot of Pikm-2 constructs. Samples from 3 and 6 dpi. A short 10-30 sec and long 3 min exposure are displayed for each western blot. M = Marker; - = Uninfiltrated leaf tissue; X = empty well; NS1 = 6His positive control infiltration; TPI = 0.1 μ g 6His-tagged purified protein; Ara h 2 = StrepII positive control infiltration. Expected Sizes: Pikm-1-His-FLAG = 129 kDa, Pikm-2-StrepII = 116 kDa. Arrowheads indicate expected protein band positions. Figure and caption adapted from Leaf Expression Systems Results Report.

The lack of detectable expression using the SupraVec® system was unexpected; therefore, further expression trials were conducted. pMAS:Pikm-1-HF and pMAS:Pikm-2-HA constructs were included as further positive controls, and were infiltrated individually and in combination (Figure 3-7, Figure 3-8). On this occasion, expression of Pikm-1-6His-FLAG in pLES_543 was detected by both anti-His and anti-FLAG western blots, with the strongest band detected in the 6 d.p.i. sample of Pikm-1-6His-FLAG expressed alone (Figure 3-7 D). That expression was detected in the second expression trial but not the first could be a function of inconsistent or low expression levels. The control construct pMAS:Pikm-1-HF was only detected in the 3 d.p.i. sample anti-FLAG western blot as a very faint band (Figure 3-7 D).

Under the same expression conditions, Pikm-2-StrepII in pLES_543 was not detected by anti-StrepII western blot, consistent with the previous experiment. The control construct pMAS:Pikm-2-HA was detected in the anti-HA western blot when expressed alone in both 3 and 6 d.p.i. samples, as well as in the co-expressed pMAS:Pikm-1-HF + pMAS:Pikm-2-HA sample at 3 d.p.i (Figure 3-8 D). Although overall signal was low, the most intense band was in the pMAS:Pikm-2-HA sample at 3 d.p.i, consistent with observations from other Banfield Lab members.

The proprietary pLES_543 vector produced an increase in Pikm-1 overexpression compared to the pMAS vector (Figure 3-7 C, D); both constructs used the same gene sequence and so the difference in expression can be attributed to the difference in expression vector. However, expression of Pikm-1 was not consistent between the two expression trials and no expression of Pikm-2-StrepII was detected in either trial. From this work we conclude that without further optimisation expression with pLES_543 does not represent a suitable method for expression of the Pik pair for structural studies.

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

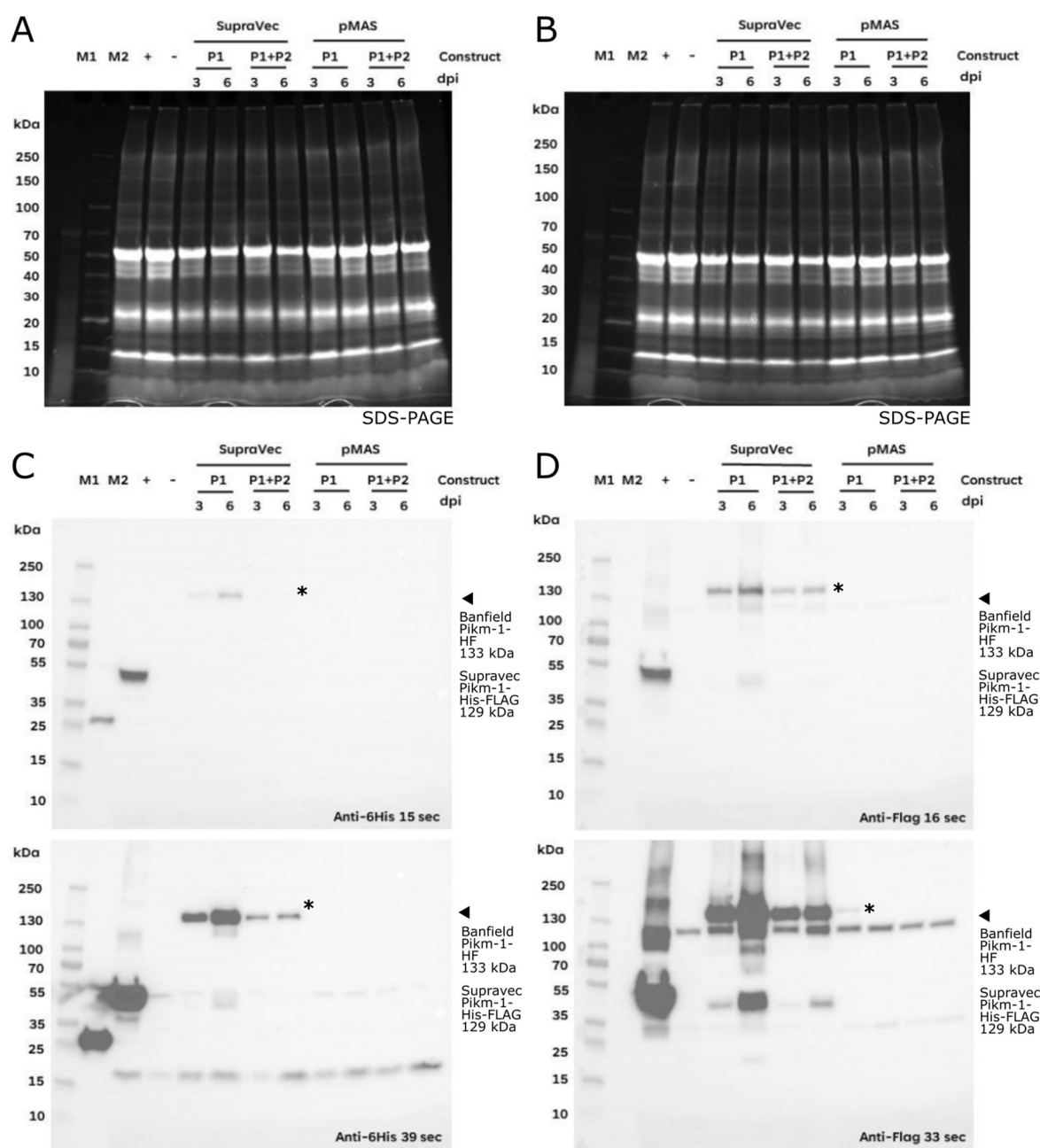


Figure 3-7: Second expression trial of Pikm-1 in *N. benthamiana* by Leaf Expression Systems

A) and B) SDS-PAGE of Pikm-1 constructs expressed alone and in combination with Pikm-2. **C)** Anti-6His-HRP western blot of Pikm-2 constructs. **D)** Anti-FLAG-HRP western blot of Pikm-2 constructs. Samples from 3 and 6 d.p.i. A short 15-16 s and long 30-40 s exposure are displayed for each western blot. P1 = Pikm-1, P2 = Pikm-2. SupraVec = Leaf Expression Systems vector: Pikm-1-His-FLAG - expected size 129 kDa; Pikm-2-StrepII - expected size 116 kDa. pMAS = Banfield Lab binary expression vector. Pikm-1-HF expected size 133 kDa; Pikm-2-HA - expected size 122 kDa. M1 = pre-stained marker; M2 = unstained marker containing TPI = 0.1 μ g 6His-tagged purified protein. '-' = Uninfiltrated leaf tissue, '+' = NS1 6His-Flag positive control infiltration. NS1 = 52 kDa, TPI = 27 kDa. Figure and caption adapted from Leaf Expression Systems Report. * = protein of interest.

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

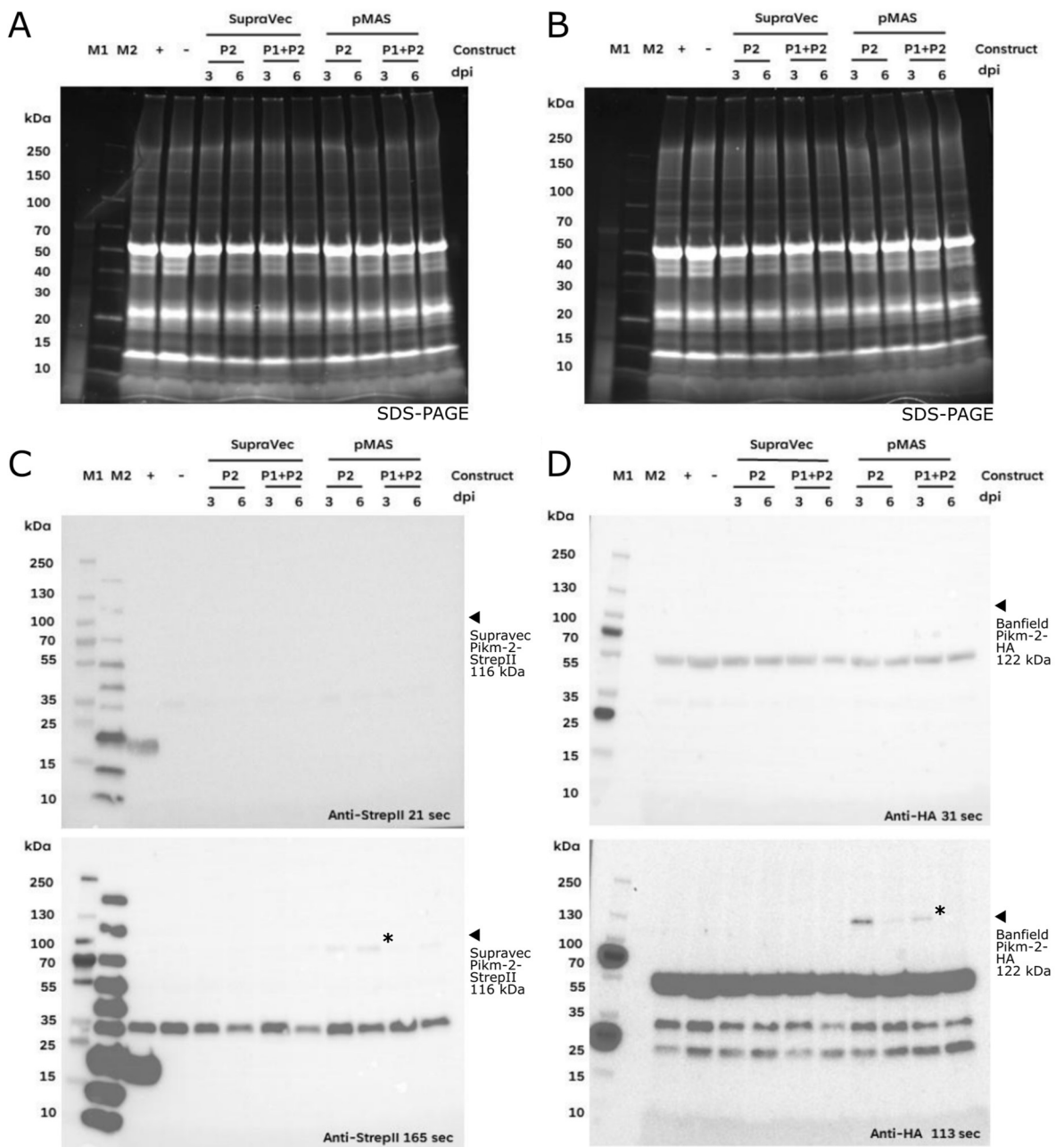


Figure 3-8: Second expression trial of Pikm-2 in *N. benthamiana* by Leaf Expression Systems

A) and B) SDS-PAGE of Pikm-2 constructs expressed alone and in combination with Pikm-1. **C)** Strep-Tactin®-HRP Western blot of Pikm-2 constructs. **D)** Anti-HA-HRP Western blot of Pikm-2 constructs. Samples from 3 and 6 d. SupraVec = Leaf Expression Systems vector: Pikm-1-His-FLAG - expected size 129 kDa; Pikm-2-StrepII - expected size 116 kDa. pMAS = Banfield Lab binary expression vector. Pikm-1-HF expected size 133 kDa; Pikm-2-HA - expected size 122 kDa. M1 = pre-stained marker; M2 = unstained marker containing Ara h 2 = 0.1 µg StrepII-tagged purified protein. '-' = Uninfiltrated leaf tissue, '+' = Ara h 2-StrepII-tagged positive control infiltration. Expected Sizes: Ara h 2 = 19 kDa. Figure and caption adapted from Leaf Expression Systems Results Report. * = protein of interest.

3.3.2 Expression in Insect cell culture

Sf9 cells were a gift from the Carter Lab, MRC-LMB, Cambridge, and the Maxwell Lab, JIC. The control construct Dynein-1 heavy chain was a gift from the Carter Lab. Sf21 work was performed with Dr Nitika Mukhi during a research visit to the Schultze-Lefert Lab, MPIPZ, Germany.

The three most commonly used insect cell lines for recombinant protein expression are Sf9, Sf21, and High Five. Sf9 and Sf21 are derived from ovarian tissue of the moth *Spodoptera frugiperda* and are adapted for suspension culture but can also be grown as adherent cells. Typically, exogenous genes for overexpression are introduced by infection with recombinant baculovirus, which itself is generated by transfection of the insect cells with a large plasmid carrying the baculovirus genome and the gene of interest. Sf9 is a sub-lineage of the Sf21 cell line with a more uniform cell size and is thought to be better adapted to high density suspension cell culture. Sf21 cell culture has been used in the Chai Lab for expression of NLRs such as Zarl, RPP1, and Sr35. Given this precedent, and my own previous experience working with Sf9 cells, I first set out to express the Pikm-1/Pikm-2 pair in Sf9 insect cells.

3.3.2.1 Re-establishing Sf9 culture in the Banfield Lab

To carry out Sf9 expression trials, Sf9 cell culture first had to be reestablished in the lab. The biological safety cabinet and incubator designated for insect cell use were thoroughly cleaned. Cryopreserved Sf9 cell stocks were a gift from the Carter Lab (MRC-LMB, Cambridge), and were grown according to typical protocols summarised in the Methods section of this chapter (3.2.1 Maintenance of Sf9 cell culture).

The Sf9 cell line is not immortal and so periodically the cell culture must be reestablished from a cryo-preserved cell stock. Therefore, initially none of the Sf9 cell culture was discarded so that the culture volume expanded, and from this multiple equivalent vials of cells were prepared for cryo-preservation. At the time of preparation for storage, the cell morphology appeared normal under the microscope, the cultures looked healthy macroscopically, culture viability (proportion of live cells) was 92%, and the cell count was doubling every 24 hours, indicating that the culture was in log-phase growth. The cryo-preserved aliquots were pre-qualified by using one to re-establish a cell culture; the original culture was maintained until it had been proven that the cryo-preserved vials could be successfully used to resume healthy cell culture.

3.3.2.2 Preparation of baculovirus expression vectors

Expression vectors for recombinant baculovirus mediated insect cell expression of Pik NLRs were prepared. In brief, the gene of interest is cloned with the very late *Autographa californica* multiple nucleopolyhedrovirus polyhedrin promoter (AcNPV-polh) and the Simian virus 40 early polyadenylation signal and terminator (SV40). In the Tn7 based method, the promoter, gene, and terminator are flanked by Tn7 transposition sites. When the plasmid carrying these elements is transformed into an *E. coli* strain carrying the modified baculovirus genome with Tn7 insertion sites and a helper plasmid with Tn7 transposase, then site specific Tn7 transposition can occur. The gene of interest, flanked by promoter and terminator, is thus inserted into the baculovirus genome.

Co-expression of genes from a single plasmid is considered optimal, to ensure that every gene is present in each infected insect cell. This would be particularly important in the case that co-expression stabilises the proteins and allows them to accumulate to higher levels. Since Pikm-1 and Pikm-2 share a promoter in the rice genome it is reasonable to assume that they are co-expressed and therefore ensuring that they are also co-expressed in a heterologous system should be advantageous. Therefore, an expression construct should be produced containing multiple genes of interest, each with their own promoter and terminator.

For ease of co-expression of different combinations of tagged Pikm-1 and Pikm-2, with or without AVR-PikD, I used a modular cloning system. The GoldenBac vector set, which can facilitate assembly of 2 to 15 genes into an acceptor vector in a single Golden Gate cloning reaction, had recently been made available (Neuhold et al., 2020). Compared to a previous iteration of this vector set, biGBac which employed Gibson cloning to assemble multi-gene constructs, GoldenBac is more compatible with cloning techniques currently used in the Banfield Lab (Weissmann et al., 2016).

However, the GoldenBac vectors are not entirely compatible with the TSL SynBio Golden Gate vectors, as GoldenBac entry vectors do not contain Golden Gate entry cloning sites, and instead only use Golden Gate reactions to combine the genes from multiple entry vectors into a single acceptor vector. Therefore, genes were cloned with tags using the TSL SynBio Golden Gate system before individual transfer into the GoldenBac entry vectors by InFusion cloning. Individual genes on pGB entry vectors could then be combined into a multi-gene expression vector in pGB-dest by a Golden Gate reaction, which should simplify re-cloning different combinations of tagged genes of interest onto a single expression vector.

To mirror the affinity tags commonly used in the Banfield Lab with Pikm-1 and Pikm-2 in *N. benthamiana* expression, Pikm-1 was first cloned with a C-terminal FLAG tag, and Pikm-2 with a C-terminal HA tag. Since these tags are tolerated at these positions in the *N. benthamiana* system, it was hoped that they might also be expressed in Sf9 cells. AVR-PikD was N-terminally tagged following the same logic; a tag at this position is tolerated in *N. benthamiana* and is still able to trigger cell death, implying that N-terminally tagged AVR-PikD can interact normally with the Pik NLRs.

3.3.2.3 Sf9 expression gave encouraging early results, but ultimately proved challenging

Initially small scale Sf9 expression trials were conducted. The approach taken was to first establish if expression of the Pik pair was possible in Sf9 cells, prioritising detection of all complex components by western blot, before turning to combinations of genes and tags to facilitate purification of a homogenous complex suitable for structural studies.

Taking the results of Zdrzalek et al. into consideration, the ultimate plan for purification of a tripartite complex, was for the final affinity purification to be via a tagged AVR-PikD, in hope of isolating only complexes containing effector (Zdrzałek et al., 2020). If all three members of the complex are associated, the resistosome should be in an activated or pre-activated state.

After expression for 66 hours, Sf9 cell pellets were harvested. The positive control construct, Sf9 codon optimised dynein-1-heavy chain 1 (~530 kDa), was visible on Coomassie stained SDS-PAGE gels of clarified cell lysates as a very high molecular weight overexpression band which cross reacted with the anti-FLAG antibody (Figure 3-9). This demonstrated that bacmid purification, transfection, and viral amplification in insect cells was all working. No other expression constructs produced overexpression bands visible in clarified lysate samples on SDS-PAGE with Coomassie blue staining, but it is quite usual for this to be the case when expressing large proteins in insect cell culture. Therefore His-tagged AVR-PikD alone and Pikm-1-FLAG + Pikm-2-HA + His-3C-AVR-PikD were subjected to small-scale affinity purification with Ni-NTA resin followed by SDS-PAGE analysis (Figure 3-9 A). No significant protein bands were observed at sizes corresponding to AVR-PikD, Pikm-1, or Pikm-2 in the small-scale affinity purification. This was the case in both samples eluted from the resin using imidazole and in samples produced by directly boiling beads in SDS-PAGE sample buffer; therefore, subsequent expression screening focussed on western blot detection.

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

Clarified cell lysates were subjected to western blotting and Pikm-1-FLAG was detected both when expressed alone and in combination on a single plasmid with Pikm-2-HA and His-3C-AVR-PikD (Figure 3-9 B). Pikm-2-HA could not be detected by western blot and was re-cloned with a C-terminal FLAG tag as anti-FLAG western blotting had already been demonstrated to be working in insect cell lysates.

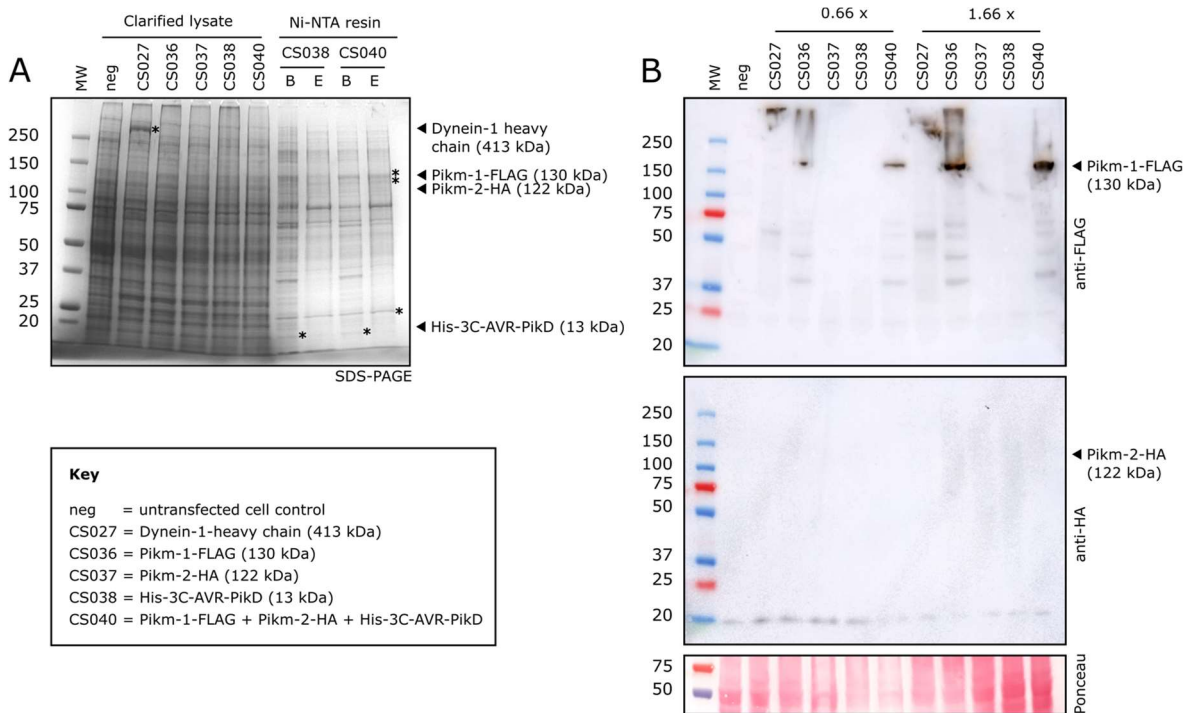


Figure 3-9: Preliminary Sf9 expression trials

A) SDS-PAGE of clarified lysate samples from Sf9 expression trials, and of samples from a small-scale Ni-NTA affinity purification. Sample volume equivalent to 1 μ L of cell pellet was loaded per well. B = Ni-NTA resin boiled in SDS-PAGE sample buffer, E = 250 mM imidazole elution from Ni-NT resin. * indicates location of potential protein bands in the SDS-PAGE. **B)** Anti-FLAG and anti-HA western blot performed on clarified lysate samples, loaded onto the gel at two concentrations (0.66 μ L and 1.6 μ L cell pellet equivalent). Construct numbers are as indicated in the key, arrowheads indicate expected size of proteins.

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

A time course was performed to determine the optimal expression duration, with samples collected 48-, 67-, and 72-hours post-infection. When co-expressed with Pikm-2-HA + His-3C-AVR-PikD, western blot band intensity of Pikm-1-FLAG was highest at the 48 hour timepoint and dropped over the subsequent two timepoints. No expression of Pikm-1-FLAG was detected when co-expressed with Pikm-2-HA + Spy-Myc-AVR-PikD. Subsequently, for protein expression, Sf9 cells were harvested after 48 hours.

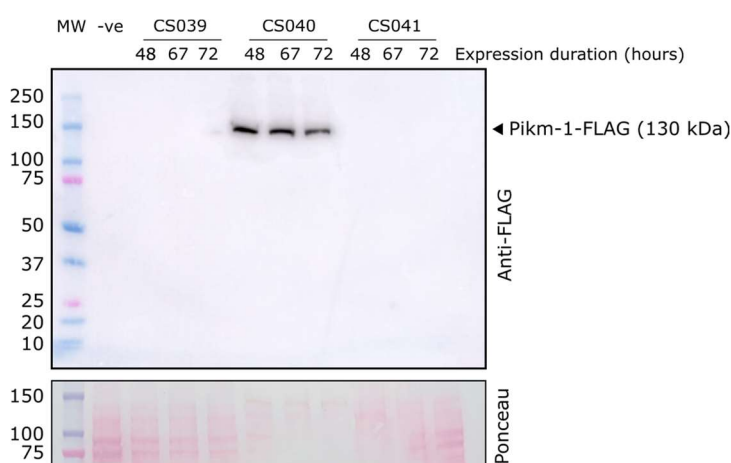


Figure 3-10: Sf9 expression time course

Western blot performed on Sf9 cell samples collected at the indicated timepoints. Samples were as follows: CS039 = Spy-3C-Myc-AVR-PikD. CS040 = Pikm-1-FLAG + Pikm-2-HA + His-3C-AVR-PikD. CS041 = Pikm-1-FLAG + Pikm-2-HA + Spy-3C-Myc-AVR-PikD. Arrowheads indicate position of bands of interest.

Pikm-2-FLAG was subsequently detected by anti-FLAG western blot when expressed alone and in combination with Pikm-1-FLAG by co-infection by two individually generated baculoviruses. The western blot band intensity of Pikm-2-FLAG was less than that of Pikm-1-FLAG, but it is difficult to make any conclusion about relative expression levels as the FLAG epitope may be more or less accessible depending on the topology of each protein. When Pikm-1-FLAG and Pikm-2-FLAG were co-expressed using a single vector carrying both genes, no expression of either Pik was detected.

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

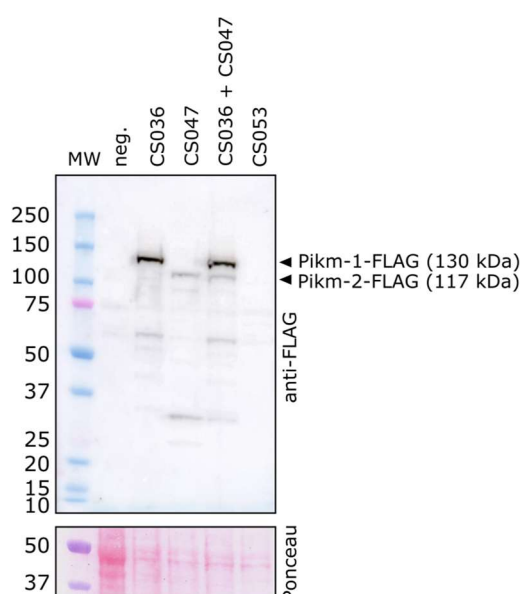


Figure 3-11: Sf9 expression testing of Pikm-2-FLAG alone and in combination with Pikm-1-FLAG

Western blot performed on Sf9 cell samples collected at 48 hours post infection. Samples are as follows: neg. = uninfected cell control, CS036 = Pikm-1-FLAG, CS047 = Pikm-2-FLAG, CS053: Pikm-1-FLAG + Pikm-2-FLAG. Arrowheads indicate the expected protein molecular weights.

Detection of both Pikm-1-FLAG and Pikm-2-FLAG was encouraging, but for large scale purification, cell death mitigation strategies and two different affinity tags would be required to avoid losses from cell death and to allow for more stringent purification. The Pik-2 N-terminal $\alpha 1$ helix mutant Pikm-2^{L19E/L23E} was produced for this purpose by Dr Hiroaki Adachi. These mutations were designed based on sequence alignment with ZAR1 and are expected to be on the outer surface of the $\alpha 1$ helix funnel. The equivalent residues were found to be essential for ZAR1 and NRC4 mediated cell death (2.1.1.2) (Adachi et al., 2019). If these residues impair membrane association of the $\alpha 1$ funnel and cause the complex to remain cytosolic, this would be an advantage for purification as it would eliminate the need for membrane protein purification, which introduces additional challenges. A C-terminal FLAG tag was appended to produce Pikm-2^{L19E/E23E}-FLAG. This construct could not be detected by anti-FLAG western blot, but neither were the previously detected Pikm-1-FLAG and Pikm-2-FLAG constructs. The anti-FLAG western blot contains high molecular weight bands which are present in all samples including those samples in which only AVR-PikD was expressed, indicating some irreproducibility in the western blot process; the reason for this is not known. Myc-AVR-PikD was also not detected by anti-Myc

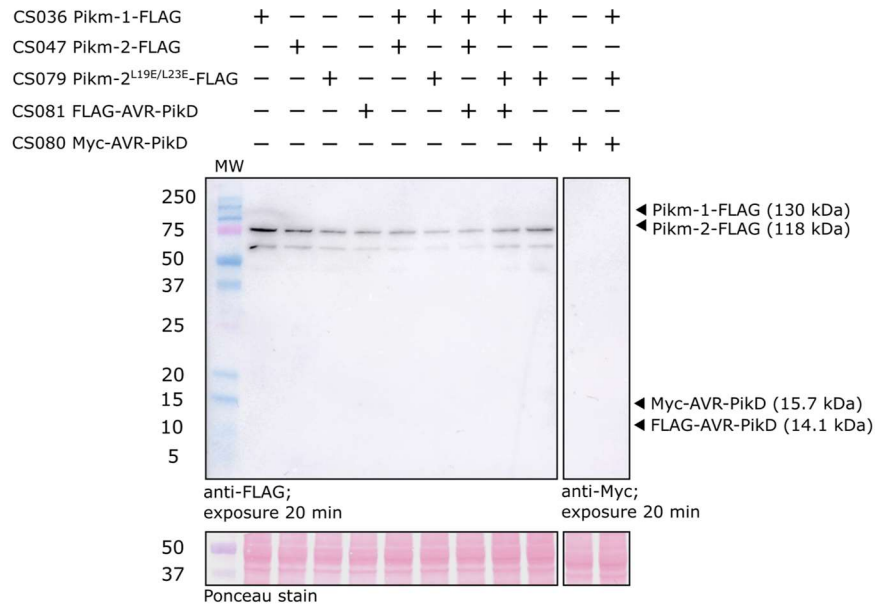
Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

western blot when expressed alone nor when expressed in combination with Pikm-1-FLAG and Pikm-2^{L19E/L23E}-FLAG (Figure 3-12 A).

The Strep-tag is commonly used for purification of proteins expressed in insect cells, so Pikm-2-StrepII-His, StrepII-AVR-PikD and AVR-PikD-TwinStrep constructs were then produced, and their expression tested. Neither N- nor C-terminally Strep-tagged AVR-PikD were detectable by western blot either alone or in combination with Pikm-1-FLAG and Pikm-2^{L19E/L23E}-FLAG. Despite many non-specific anti-Strep bands around the ~100 kDa size, higher intensity bands were detected in the samples containing Pikm-2-StrepII-His at a size that could correspond to that protein (Figure 3-12 B).

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

A



B

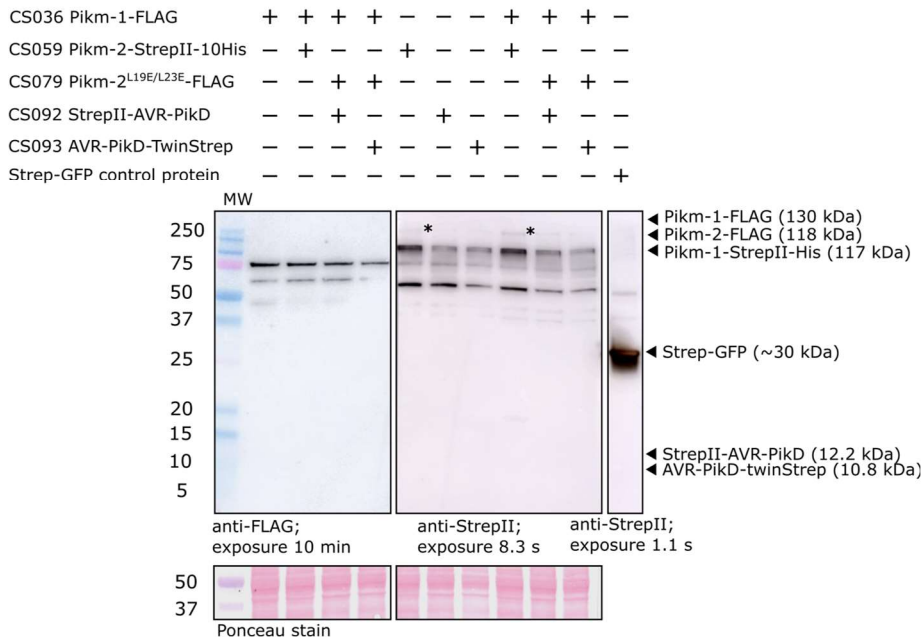


Figure 3-12: Sf9 expression testing of Strep-tagged Pik resistosome components

Western blot performed on Sf9 expression samples collected at 48 hours post-infection. Samples are indicated above each western blot. **A)** Anti-FLAG and anti-Myc western blot. **B)** Anti-FLAG and anti-StrepII western blot. Arrowheads indicate expected protein molecular weights, * denotes bands of interest.

Until this point, only C-terminal tagging of the Pik NLRs had been attempted, to avoid disruption to the $\alpha 1$ helix implicated in the function of other CC-NLRs. However, since I had not achieved high level, reproducible overexpression of Pikm-1 and Pikm-2 in Sf9 cells, I chose to investigate a range of N-terminal tags which may help to raise expression levels. In the study of Sr35, Sr35 was N-terminally tagged with a His-SUMO tag and a C-terminal glutathione S-transferase (GST) tag (Förderer et al., 2022). Taking inspiration from this, I cloned a panel of Pikm-1 with N-terminal tags as follows: 6His-SUMO-3C-Pikm-1, 6His-MBP-3C-Pikm-1, 6His-GB1-3C-Pikm-1, 8His-ZZ-TEV-Pikm-1, 8His-ZZ-3C-Pikm-1, Spytag-3C-Myc-Pikm-1, 10His-StrepII-3C-Pikm-1, Spytag-3C-Pikm-1. I had planned to carry out expression trials followed by anti-His western blot analysis; however, this work was not completed at the JIC. Repeated cell culture contamination and inconsistent anti-FLAG detection of previously detected samples, and the need to focus on other, more productive avenues (presented in Chapter 5) necessitated the deferral of my remaining experimental plans in late 2022. However, the Sf9 expression vector system, along with the insect cell expression constructs and N-terminal tags that I generated, remain available for future work. Upon development of a collaboration in early 2023 these resources were taken forward into a new set of expression and purification trials which are presented in the next section.

3.3.2.4 Sf21 expression offers new hope for Pik NLR expression

A collaboration with Dr Nitika Mukhi, in the Lab of Professor Paul Schulze-Lefert (Max Planck Institute for Plant Breeding Research, MPIPZ, Cologne, Germany) was initiated. Expression constructs Pikm-1-FLAG, Pikm-2-FLAG, and 6His-SUMO-Pikm-1, previously used for Sf9 insect cell expression trials, were shared. It was established that expression and purification of these constructs was possible in Sf21 cells, as exemplified here by a representative small-scale expression test and affinity purification (Figure 3-13). On the back of this, a research visit to MPIPZ in June 2023 was arranged so that I could participate in further purification efforts. Presented here in the remaining part of this results section is the expression, purification, and preparation of electron microscopy samples that was performed together with Dr Nitika Mukhi during my research visit to MPIPZ.

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

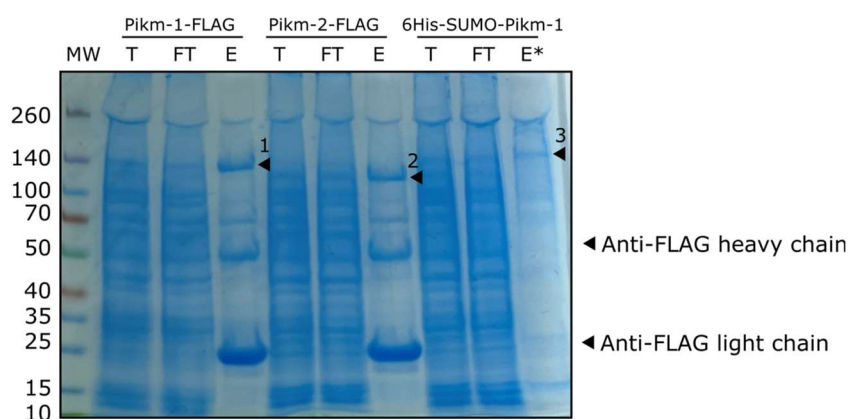


Figure 3-13: Small scale affinity purification of Pikm-1-FLAG, Pikm-2-FLAG, and 6His-SUMO-Pikm-1 demonstrating expression of these constructs in Sf21 insect cell culture.

Numbered arrowheads indicate NLRs; 1) Pikm-1-FLAG expected size 130 kDa, 2) Pikm-2-FLAG expected size 118 kDa, 3) 6His-SUMO-Pikm-1 expected size 140 kDa. MW = molecular weight marker, FT = unbound protein, E = elution from FLAG resin by boiling in SDS-PAGE sample buffer, E* = elution from Ni-NTA resin using 200 mM imidazole. Performed by Dr Nitika Mukhi and included with permission.

Bacmid production and isolation were carried out as previously described (3.2.3.4 Bacmid *E.coli* transformation, selection, and bacmid purification). Sf21 cells were transfected by a method very similar to the transfection of Sf9 cells (3.2.3.5 Sf9 transfection and virus production), apart from use of a different transfection reagent (X-tremeGENE HP transfection reagent, Roche). Virus was amplified by continuing to subculture the infected Sf21 cells, which is an additional difference to the Sf9 cell culture practices I had been using at the JIC. Amplified virus was applied to Sf21 cell culture and after 48 hours, cell pellets were harvested by centrifugation.

Purification of two complexes was undertaken simultaneously: an active complex composed of 6His-SUMO-Pikm-1 + Pikm-2^{L19E/L23E}-FLAG + AVR-PikD-TwinStrep, and an inactivated complex composed of 6His-SUMO-Pikm-1 + Pikm-2^{L19E/L23E}-FLAG. Our strategy to purify the active complex was to perform a Ni-NTA affinity purification to capture 6His-SUMO-Pikm-1, followed by Streptavidin affinity purification to capture only particles containing both Pikm-1 and the effector. Since Pikm-2 is also expected to interact with these two proteins, a tripartite complex was expected from the “active” purification. Our strategy to purify inactivated complex would be to use Ni-NTA affinity purification to capture 6His-SUMO-Pikm-1, followed by FLAG affinity purification to select only those 6His-SUMO-Pikm-1 molecules which are also in complex with Pikm-2^{L19E/L23E}-FLAG. The Pikm-2 N-terminal L19E/L23E mutant was selected for this work to abolish cell death; as in the *N. benthamiana* work these mutants were selected based on alignments performed by Adachi

et al. which determined that Pik-2 L19E/L23E are equivalent residues to L10E/L14E of NRC4 (Adachi et al., 2019; Wang et al., 2019a).

Purification of the activated complex was quite successful up to the stage of biotin elution from the Strep-Tactin resin (Figure 3-14 B). Clear protein bands in SDS-PAGE were visible corresponding to the expected sizes of His-SUMO-Pikm-1 (140 kDa) and Pikm-2^{L19E/L23E}-FLAG (118 kDa), although there was not a single band that corresponded to the expected size of AVR-PikD-TwinStrep (14 kDa). The band corresponding to His-SUMO-Pikm-1 is of a significantly higher intensity than the Pikm-2^{L19E/L23E}-FLAG band. Considering the similar molecular weight of both proteins, this suggests that Pikm-2^{L19E/L23E}-FLAG is present in the sample in a lower than 1:1 ratio to His-SUMO-Pikm-1. This differs to the previous test purification performed by Dr Mukhi in which equal band intensities for His-SUMO-Pikm-1 and Pikm-2^{L19E/L23E}-FLAG were observed.

Purification of the inactivated complex was also successful up to the stage of FLAG peptide elution (Figure 3-15 B) and again SDS-PAGE revealed bands that correspond to the expected sized of both members of the complex. In contrast to the activated complex, after FLAG affinity purification, the relative band intensities between Pikm-1 and Pikm-2 appear equal and so from this we can assume that the stoichiometry of the complex purified at this stage is approximately 1:1.

Both samples were then subjected to SEC. In both the “activated” and “inactivated” purifications, the SEC peak at 9 ml represents the void volume; particles which elute at this volume are too large to access any pores within the stationary phase of the column and therefore exceed the upper size limit that the column can resolve. Any particles eluting at the void volume are likely to be aggregated protein and therefore should be avoided for single-particle electron microscopy.

The void peak of the “activated” prep was larger than the peak at 11.5 ml, suggesting that a large proportion of all protein molecules in that purification were aggregated (Figure 3-14 C, D). A band at around 15 kDa which might represent AVR-PikD-TwinStrep was visible in fractions corresponding to the void volume, among other low molecular weight bands, but by fraction 21 which corresponds to the second peak, the band intensity was much lower, leading us to question whether a tripartite complex was present in the second peak. To investigate if any higher order oligomers were present here, fractions 21 and 22, corresponding to elution volumes of 11.6 ml and 11.9 ml, were selected for negative stain electron microscopy.

In the “inactivated” purification, the relative intensities of the chromatogram were lower, and bands corresponding to the Pik pair were more weakly visible by SDS-PAGE, indicating a lower concentration in this sample (Figure 3-15 C, D). The void peak at 9 ml had a shallower trailing “shoulder” which could indicate a range of high molecular weight aggregate/complex sizes. The peak at 11.5 ml was narrower and monodisperse, and fractions 21, 22, and 23, corresponding to elution volumes of 11.6 ml, 11.9 ml, and 12.1 ml were selected for negative staining. To assess if this protein could be concentrated after SEC without aggregation, fraction 21 was also subjected to centrifugal concentration and this sample was then applied to negative stain grids.

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

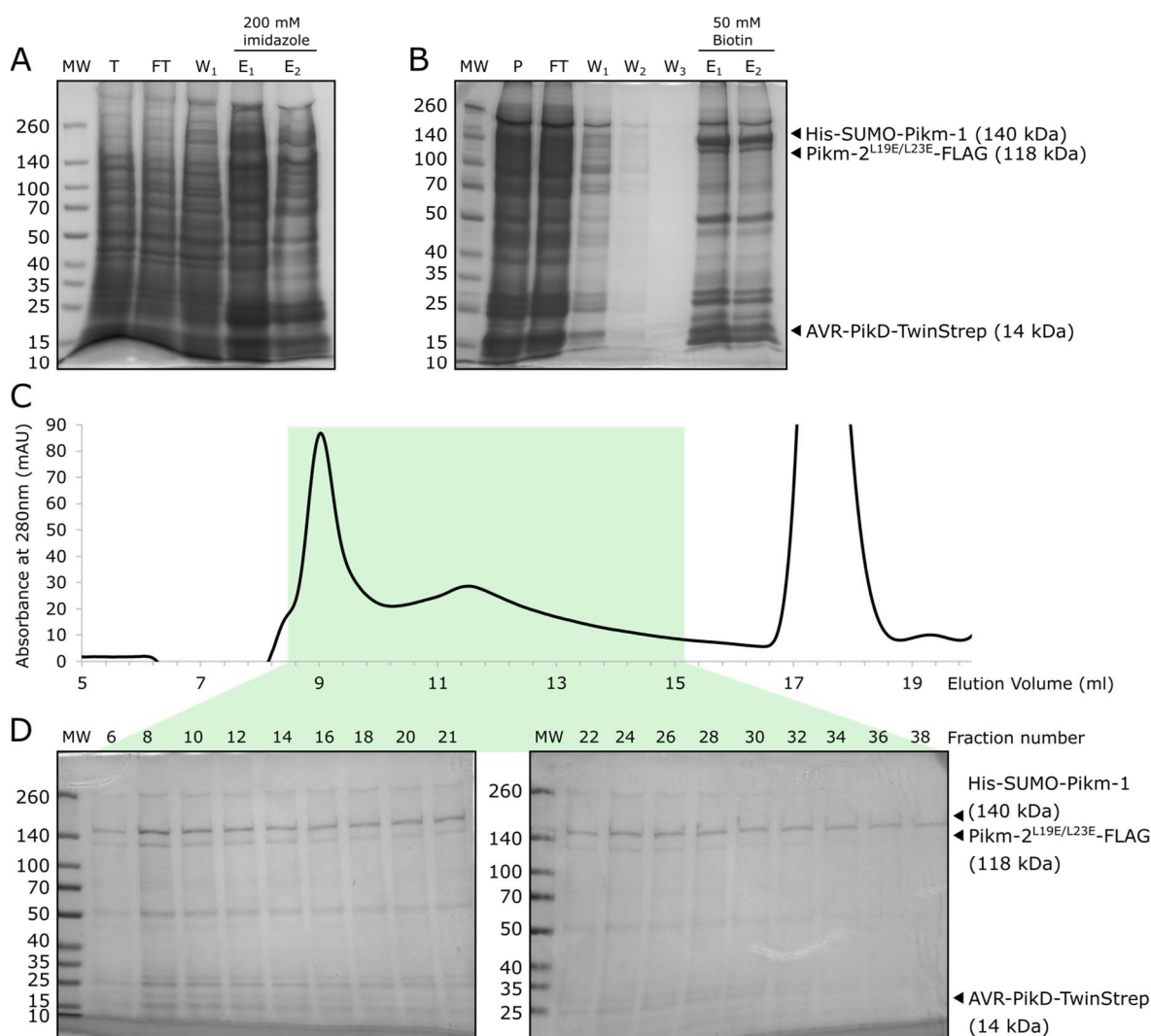


Figure 3-14: Purification of “activated” 6His-SUMO-Pikm-1 + Pikm-2^{L19E/L23E}-FLAG + AVR-PikD-TwinStrep complex

A) Ni-NTA affinity purification. T = total soluble protein, FT = flow through, W₁ = wash, E₁ = 200 mM imidazole elution after 30-minute incubation, E₂ = 200 mM imidazole elution after 10-minute incubation. **B)** Streptavidin affinity purification. P = pooled Ni-NTA elutions, FT = unbound flowthrough, W₁₋₃ = washes 1-3, E₁₋₂ = 50 mM biotin elutions after 1 hour incubation each. **C)** Superose 6 increase 10/30 SEC chromatogram of concentrated biotin eluted protein. Green highlighted area indicates the position of fractions from which SDS-PAGE samples were taken. **D)** SDS-PAGE of SEC elution fractions corresponding to green highlighted region of the chromatogram in C). MW = molecular weight marker. Arrowheads indicate expected protein sizes: His-SUMO-Pikm-1 – 140 kDa, Pikm-2^{L19E/L23E}-FLAG – 118 kDa, AVR-PikD-TwinStrep – 14 kDa.

Chapter 3: Attempts to express and purify full length Pik-NLR proteins for reconstitution of a pre-activated or activated Pik resistosome

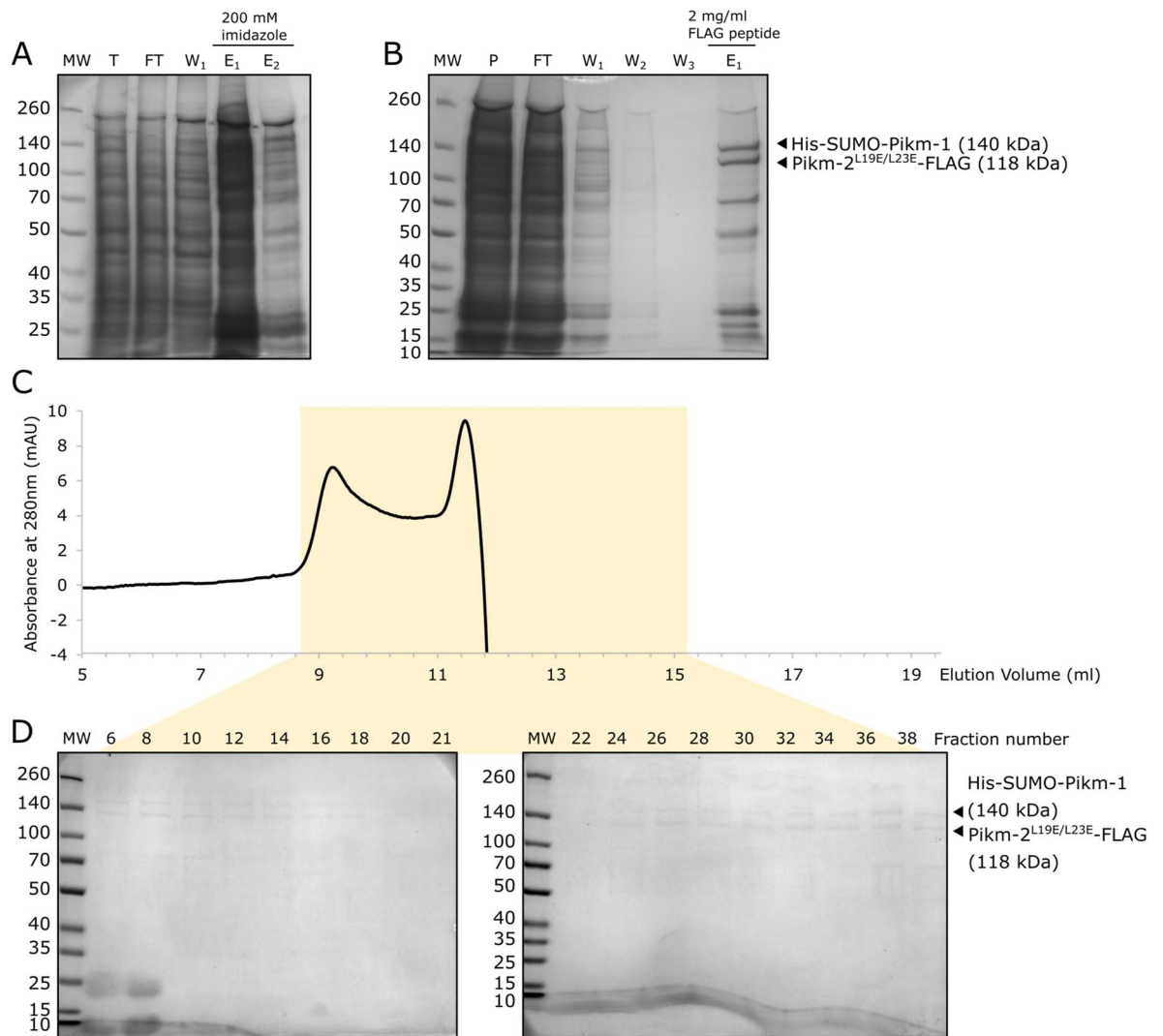
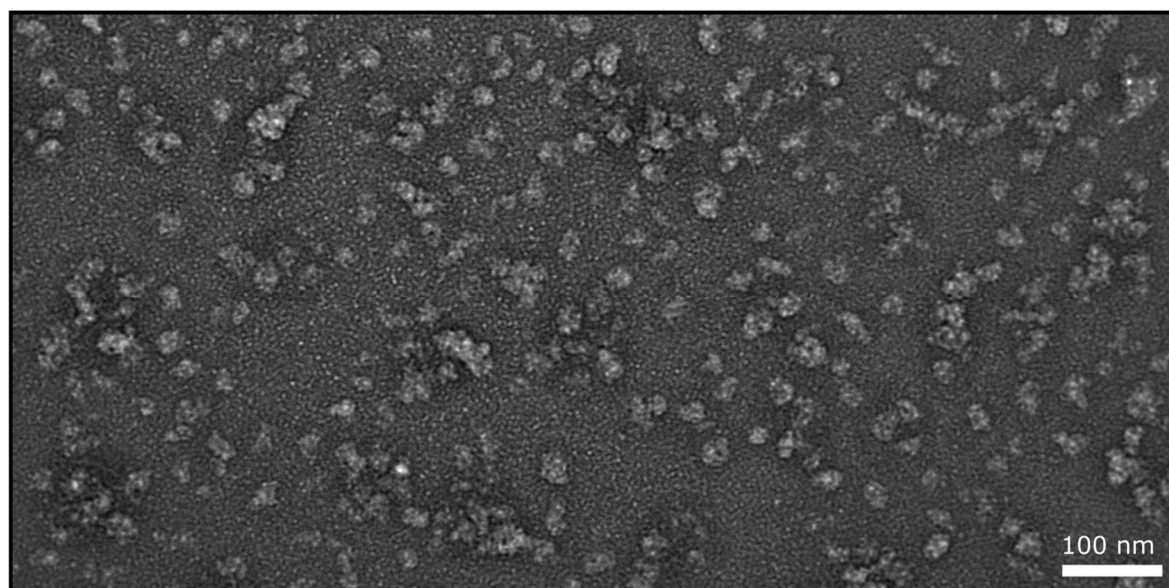


Figure 3-15: Purification of "inactivated" 6His-SUMO-Pikm-1 + Pikm-2^{L19E/L23E}-FLAG complex

A) Ni-NTA affinity purification. T = total soluble protein, FT = flow through, W₁ = wash, E₁ = 200 mM imidazole elution after 30-minute incubation, E₂ = 200 mM imidazole elution after 10-minute incubation. **B)** FLAG affinity purification. P = pooled Ni-NTA elutions, FT = unbound flowthrough, W₁₋₃ = washes 1-3, E₁ = 1 mg/ml FLAG peptide elution after 14-hour incubation. **C)** Superose 6 increase 10/30 SEC chromatogram of concentrated FLAG eluted protein. Yellow highlighted area indicates the position of fractions from which SDS-PAGE samples were taken. **D)** SDS-PAGE of SEC elution fractions corresponding to yellow highlighted region of the chromatogram in C). MW = molecular weight marker. Arrowheads indicate expected protein sizes: His-SUMO-Pikm-1 – 140 kDa, Pikm-2^{L19E/L23E}-FLAG – 118 kDa.

3.3.2.5 Electron microscopy of Pik complexes purified from Sf21 cells

Negative stain grids were produced using 1 % uranyl acetate (3.2.3.8 Negative stain electron microscopy). Particles could be observed on grids prepared from all samples. In the “activated” complex samples, from fractions 21 and 22 from Figure 3-14 C, D, many homogenous particles were visible, with only a few larger aggregates present in each micrograph (Figure 3-16).



Fraction 21: "Active" 6His-SUMO-3C-Pikm-1 + Pikm-2^{L19E/L23E}-FLAG + AVR-PikD-TwinStrep

Figure 3-16: Negative stain micrograph from “active” Pik resistosome purification

Representative negative stain micrographs prepared from SEC fraction 21 from the “Active” 6His-SUMO-Pikm-1 + Pikm-2^{L19E/L23E}-FLAG + AVR-PikD-TwinStrep protein purification. Scale bar = 100 nm.

In the “inactivated” complex samples from Figure 3-15 C, D, particles in fraction 21 appeared quite heterogenous, ranging from around 18 – 26 nm in diameter (Figure 3-17 A). After concentration particles from the same fraction had a much larger diameter of approximately 29 – 35 nm and appeared at a higher density on the grid. While centrifugal concentration succeeded in improving particle density, it also seems to have changed the nature of the particles, possibly representing some form of higher order complex or aggregation formation during concentration (Figure 3-17 B). In fraction 22 of the “inactivated” purification some particles resembling rings were visible, with diameters of around 18 – 22 nm (Figure 3-17 C). Fraction 23 of the “inactivated” purification contained the most homogenous particles with minimal aggregation; among the unconcentrated samples, it had the highest particle density (Figure 3-17 D).

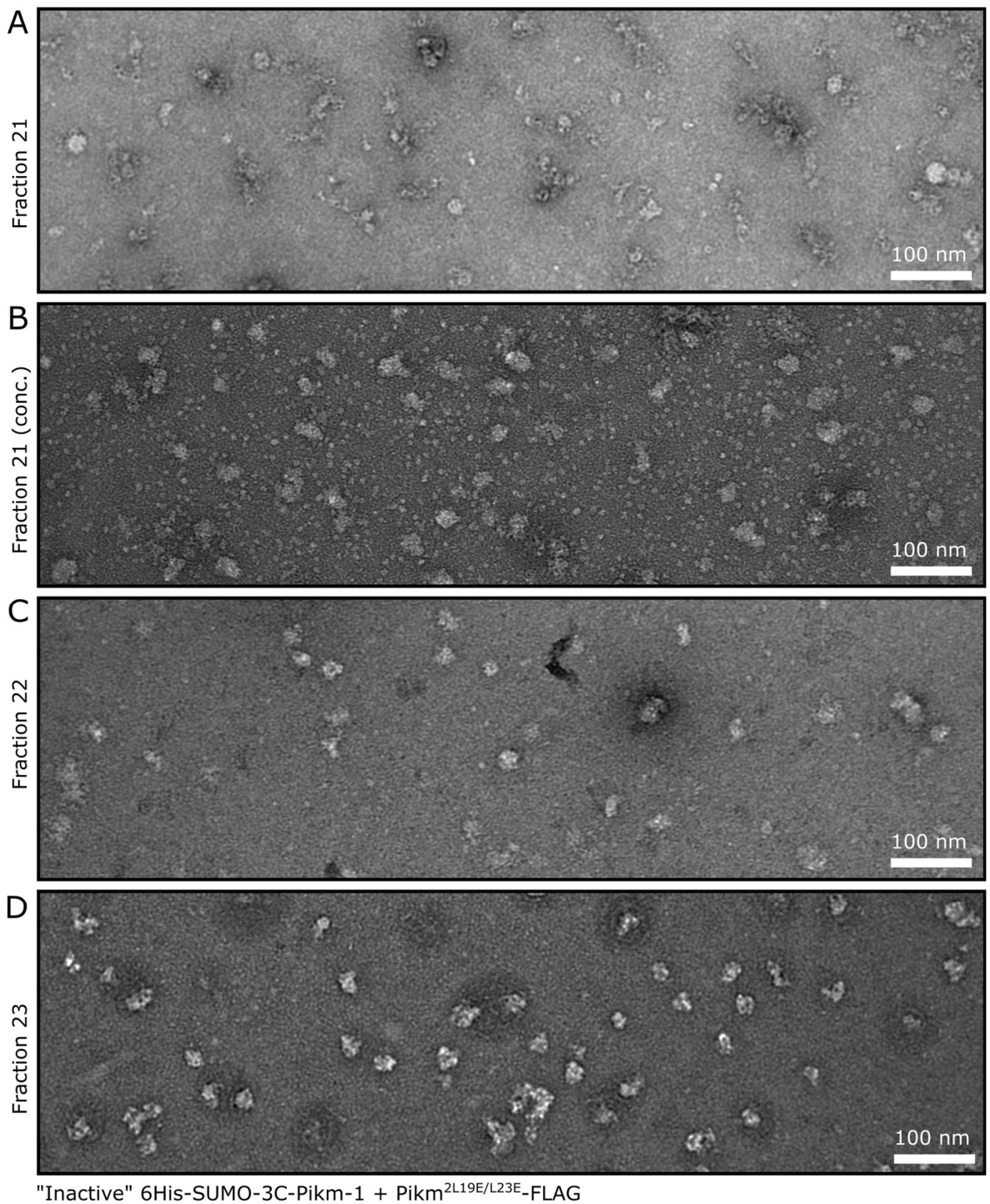


Figure 3-17: Negative stain micrographs of "inactive" Pik resistosome purification

*Representative negative stain micrographs prepared from SEC fractions from the "inactive" 6His-SUMO-Pikm-1 + Pikm-2^{L19E/L23E}-FLAG protein purification. Scale bar = 100 nm. **A)** Fraction 21, **B)** Fraction 21 (concentrated), **C)** Fraction 22, **D)** Fraction 23.*

Based on observed negative stain particle appearance and density, fraction 21 from the “activated” complex was most promising for further study. However, without visual confirmation of the presence of AVR-PikD by SDS-PAGE we did not pursue this sample for cryo-EM sample preparation. Instead, we focussed on fraction 23 from the inactivated complex. To achieve a higher protein concentration for making cryo-electron microscopy grids, the protein was concentrated in a centrifugal concentrator. After storage on ice for 16 hours, the protein samples were applied to cryo-electron microscopy grids and vitrified using a Vitrobot plunge-freezing device (3.2.3.9 Cryo-electron microscopy). The grids were screened on a cryo-electron microscope but unfortunately very few particles were observed. Although a small data set was collected and particles were picked, 2D classification did not produce any defined 2D classes. Therefore, these samples were ultimately proven unsuitable for structure determination and work to optimise purification of a stable Pik resistosome complex continued.

3.4 Discussion

3.4.1 Multiple approaches to improve expression of Pikm-1/Pikm-2 in *N. benthamiana* produced limited success

The Pikm-1/Pikm-2 pair are known to express in *N. benthamiana* by *agrobacterium* mediated expression; this has formed the basis for prior western blot and co-immunoprecipitation studies in the Banfield Lab. However, for structural studies a larger amount of protein is required, and at a higher purity. The novel tag system SpyTag/SpyCatcher was initially promising, but ultimately concerns over expression level in *N. benthamiana* meant that this work was not continued. Genomic expression constructs of Pikm-1 and Pikm-2 did not yield an improvement in expression; they could not be detected by western blot and did not produce cell death when co-infiltrated with AVR-PikD, suggesting they were not expressed. Finally, the Leaf Expression Systems vector pLES_543 produced significantly higher expression of Pikm-1-FLAG on one occasion and no detectable expression on another occasion, but the reason for the inconsistency between trials is unknown. Pikm-2-Strep expression was not detected in this system in either trial. Therefore, outsourcing the expression to Leaf Expression Systems to take advantage of their proprietary viral expression vector technology was not a viable option for ongoing work.

3.4.2 Sf9 culture was challenging, but Sf21 culture offers hope

Unfortunately, the difficulties in maintaining stable Sf9 cell culture in the Lab prevented a full exploration of possible expression constructs and combinations. The detection of Pikm-1-FLAG by western blot provided early encouragement, but this success was not replicated with the other affinity tags as cultures would often succumb to contamination before the viral amplification steps required for protein production could be completed. In most cases, little to no protein expression could be detected by western blot.

If this work were to be repeated at the JIC, I would focus on eliminating contamination. I would replace reagents such as the transfection reagent and adopt the use of antimycotic supplementation to the media in addition to the antibiotic supplementation used. Although anecdotal knowledge says that the use of antibiotics could reduce protein expression yield, a lower yield is vastly preferable to yield loss due to contamination of the cultures. Additionally, if the work were to be repeated, I would invest time at the start of the project into converting the GoldenBac vectors into TSL SynBio compatible Golden Gate acceptor vectors. In this way, the initial combination of tag and gene could be carried out using

Golden Gate assembly directly into the GoldenBac entry vectors, in contrast to the InFusion or other primer-based cloning required in the system as it currently exists. This would replace the two cloning steps required to assemble a GoldenBac entry vector with one single Golden Gate reaction, thereby eliminating the time required to order and receive cloning primers and carry out InFusion cloning. In this way, a week could be saved when making every new gene/tag combination which is valuable when carrying out iterative construct design and testing in a system where around three weeks are already required to amplify virus, carry out protein expression, and carry out diagnostic western blots.

Upon successful expression of the Pik NLRs, I would prioritise performing a membrane fractionation assay and western blot, to determine which subcellular fraction the protein is associating with. I feel this would be prudent, particularly in the case of the activated complex, as other activated CC-NLRs have been reported to associate with the plasma membrane (Bi et al., 2021). By confirming this, it will confirm if purification of the protein should be approached as a membrane protein or a cytosolic protein, informing choices such as type and concentration of detergents used, handling techniques, and considerations when preparing electron microscopy grids.

It is very fortunate that through the establishment of a collaboration it was possible to circumvent some of the difficulties I experienced with Pik expression in Sf9 insect cell culture at the JIC. Through considerable effort of our collaborators, including use of a different insect cell line (Sf21), and different culture conditions (e.g. use of antibiotic/antimycotic supplementation), the constructs which showed some initial promise in Sf9 expression were expressed well in Sf21 cells. Additionally, expression constructs such as 6His-SUMO-Pikm-1 which was produced for Sf9 testing but was ultimately not assessed at the JIC, were able to be tested in the Sf21 system at the MPIPZ. Significant progress in protein expression and purification was achieved at the MPIPZ. After my return to the JIC I remained involved in the project though cloning of different expression constructs for further testing. This work is ongoing; while considerable challenges remain, our collaborators have been instrumental in advancing the work beyond what I could achieve at the JIC.

3.4.3 Alternative structural biology approaches could offer new insights

Other structural biology approaches could offer alternative ways to achieve a Pik resistosome structure. Since protein expression has been demonstrated and final grid

sample preparation has been challenging, use of streptavidin affinity grids could provide a way of concentrating relevant particles on the grid while allowing other proteins to be washed away. In brief, a 2-dimensional streptavidin crystal can be grown on a cryo-EM grid attached to a biotinylated monolayer of lipids laid upon a carbon support. Proteins can be biotinylated in vitro, and subsequently are able to associate with the streptavidin layer. This technique can be particularly helpful in overcoming low particle concentration.

In the case of Pik resistosome capture on affinity grids, a SEC step could still be performed to capture non-aggregated complexes, but the need for a final concentration step would be circumvented. Instead, SEC fractions could be directly applied to the grid and particles concentrated on the surface there. However, the functionalisation of the grid does increase its charge density. Therefore, even after subtraction of the streptavidin crystal signal, the signal to noise ratio of micrographs is decreased, which introduces additional data processing challenges.

Cryogenic electron tomography (Cryo-ET) can be used to image cellular structures and proteins in their native environment and therefore could circumvent the need for any protein purification. A Pik resistosome complex would likely be large enough to be detected by this technique (a hexamer of Pik-1/Pik-2 dimers – approximately 1.4 MDa, a hexamer of Pik monomers – approximately 700 kDa). However, the challenge of this is that it is usually not possible to observe electron microscopy density of individual protein complexes and therefore it would be almost impossible to identify the location of regions of interest within the cellular volume. Cryogenic correlative light and electron microscopy (cryo-CLEM) offers a way to identify regions of interest by first using cryogenic fluorescence microscopy (cryo-FM) to identify regions of interest marked by fluorophores in a vitrified electron microscopy sample, before then moving on to image the exact same position of the sample using cryo-ET. This is achieved through registering features visible in both images and superimposing the two to aid navigation around the sample while performing cryo-ET.

Although fluorophores excited under cryogenic conditions are more likely to be excited to their dark state, therefore producing a lower fluorescence signal, it could be that oligomerisation of activated NLRs tagged with fluorophores could produce a bright enough signal to facilitate localisation of complexes. Use of a bright organic dye such as Alexa fluor 647, which has good performance at liquid nitrogen temperature and could be conjugated to NLRs in-cell via to a SNAP tag could contribute towards this (Panchuk-Voloshina et al., 1999; Bosch et al., 2014). Choice of fluorophore to avoid the 300- to 550-nm range where autofluorescence of plant cells is highest would also help. However, key challenges of cryo-ET are ensuring that the sample is both thin enough and contains a high enough density of

the complex of interest so that sufficient particles can be collected for sub-tomogram averaging. Zar1 resistosomes were overexpressed and observed as puncta at quite high density in *Arabidopsis* protoplasts by total internal reflection fluorescence (TIRF) microscopy, which suggests the approach of Pik overexpression in protoplasts might be feasible (Bi et al., 2021). If thinned by focussed ion beam (FIB) milling, protoplasts might be a suitable sample for cryo-ET, as demonstrated in *A. thaliana* root protoplasts (Sanchez Carrillo et al., 2023). However, FIB-milling is a low throughput and labour-intensive process and use of this technique would require the participation of a laboratory with specific expertise. As the methodology continues to advance, it would be exciting to see a cryo-ET lab take on this challenge.

3.4.4 Conclusion

Ultimately, the goal of expression and purification of the Pik NLRs/resistosome in great enough quantities for structural studies was an ambitious one and was not realised through the *N. benthamiana* and Sf9 insect cell culture methods explored in this chapter. The fact that expression enhancing techniques such as the use of genomic expression constructs, successful for other NLRs, did not give the same results in this case highlights that each protein is unique, and that protein expression and purification outcomes do not always meet expectations. However, despite significant challenges, the ongoing Sf21 expression and purification work by our collaborators offers hope. I am optimistic that one day the full-length structure of a Pik resistosome will be solved, revealing further details of how the activation of this complex occurs. This will lead the way for understanding how Pik resistosomes, and paired CC-NLR resistosomes more widely, can be manipulated to enhance or modulate their effector recognition capabilities. It would also offer new insight into ways that the paired CC-NLRs can be manipulated to produce multiple options of CC-NLR chassis with improved amenability to bioengineering.

Chapter 4: Probing subcellular localisation and activation of the Pik-NLR pair

4.1 Introduction

Determining the subcellular localisation, oligomeric status, and post-translational status of a NLR provides further insight into its mode of action. Is the NLR cytosolic, or does it associate with a subcellular compartment? Does the localisation change upon activation? In the case of paired NLRs, presence of both NLRs is required for activation, but is subcellular localisation of both NLRs the same, or different? Association into an oligomeric resistosome is linked to activation of NLRs, but the configuration and stoichiometry of the resistosome is not determined in the case of many NLRs, including the Pik pair. Finally, post-translational modifications (PTMs) such as phosphorylation, ubiquitination, and SUMOylation are demonstrated to regulate the activation of some NLRs, and so investigation of PTMs is required to gain a full understanding of plant immunity.

4.1.1 Subcellular localisation

CC-NLRs have a proposed mode of activity which relies upon membrane association to form a Ca^{2+} ion channel, therefore membrane localisation would be expected upon activation. Before activation, the presence of sensor NLRs might be expected in any subcellular compartment that pathogen effectors localise to, as the sensor NLR is monitoring for these effectors. For example, the barley CC-NLR MLA10 is found in the nucleus while the *Arabidopsis* CC-NLR RPM1 is found at the plasma membrane (Shen et al., 2007; Gao et al., 2011).

Previous work examined the activated CC-NLR Zar1 localised in oligomers at the plasma membrane of *Arabidopsis* protoplasts (Bi et al., 2021). TIRF microscopy was performed with single fluorophore calibration which allowed a population of puncta with mEGFP fluorescence intensity corresponding to pentamers to be observed. Bleaching experiments were then performed and it was observed that upon bleaching, diffusion of unbleached complexes back into the field of illumination was comparable to diffusion of the plasma membrane aquaporin protein PIP2;1. This suggests that the Zar1 pentamers are able to diffuse in the plasma membrane and this lateral diffusion is more common than cytosolic complexes associating with and dissociating from the plasma membrane. In a study by

Adachi et al, addition of the $\alpha 1$ helix of NRC4 to YFP was sufficient to change localisation of YFP from cytosolic to a plasma membrane localisation, further confirming the membrane association conferred by the $\alpha 1$ helix of CC-NLRs (Adachi et al., 2019).

Extending this paradigm of plasma membrane association, a preprint by Ibrahim et al. provides evidence that some activated CC_R-NLRs can localise to alternative intracellular membranes. The authors demonstrate that the N-terminal domain of the CC_R-NLR NRG1 is predicted to adopt an extended conformation compared to CC-NLRs, sufficient to span double membranes such as those which surround mitochondria and chloroplasts. Fluorescence microscopy confirmed that upon activation, NRG1 localised to organellar membranes of the endoplasmic reticulum, mitochondria, and chloroplasts, but not the Golgi or plasma membrane. This is in contrast to the CC_R-NLR NRC4 which localised to the plasma membrane, but not to the endoplasmic reticulum, Golgi, chloroplasts, or mitochondria (Ibrahim et al., 2024).

4.1.2 Oligomerisation

The current model of CC-NLR activity includes formation of an Ca²⁺ ion channel that acts in the plasma membrane, therefore membrane localisation would be expected upon activation. It is not known if other CC-NLR resistosomes target different membrane bound compartments of the cell, however a recent pre-print, discussed above, demonstrates that not all NLRs act at the plasma membrane (Ibrahim et al., 2024). It has long been known that pre-activation NLRs are found in range of intracellular localisations, to facilitate their monitoring for effector activities across a range of cellular targets. In the case of paired NLRs, both partners might move together, or they might only associate or dissociate upon activation.

Oligomerisation is a common feature of NLR activation. Full-length CC-NLRs structures to date are of singleton NLRs, and so these results do not explicitly offer insight into the mode of oligomerisation of a paired CC-NLR such as Pik-1/Pik-2. Homo-oligomerisation is a feature of CC-NLR activation, as demonstrated in NLRs including MLA10 (Maekawa et al., 2011), Mla13 (Lawson et al., 2024), Sr35 (Förderer et al., 2022; Zhao et al., 2022), Sr33 and Sr50 (Casey et al., 2016), Sr50, and Zarl (Baudin et al., 2017; Wang et al., 2019a; Wang et al., 2019b; Bi et al., 2021). However, it is unclear how a paired CC-NLR resistosome would be physically arranged.

In the co-IP study of Pikp-1/Pikp-2/AVR-PikD, a tripartite activated complex was observed (Zdrzałek et al., 2020). Therefore, transient binding and dissociation models such as Pik-1/AVR-Pik activating the formation of a homo-oligomer of only Pik-2 are less likely and

mixed tripartite complexes are expected. These could take multiple forms, but the two most probable are as follows. First - the association of heterotrimers of Pik-1/Pik-2/AVR-Pik to form higher oligomers with a 1:1:1 molar ratio between the three components. Although both Pik-1 and Pik-2 have a CC-domain, Pik-1 would not be expected to participate in the formation of any α 1 helix “funnel” as observed in Zar1. Second - the Pik-1/Pik-2/AVR-Pik complex could occur from the binding of AVR-Pik to a Pik-1/Pik-2 dimer de-repressing Pik-2 and triggering an oligomerisation of further copies of Pikm-2 into a resistosome. This mode of oligomerisation has not yet been observed in plant NLRs, but would be reminiscent of the metazoan NLR inflammasomes such as NAIP5/NLRC4 (Tenthorey et al., 2017) and NAIP2/NLRC4 (Zhang et al., 2015) where a single effector-activated sensor NLR per inflammasome disk templates the addition of subsequent helper NLRs.

4.1.3 Post-translational modification

Changes in post-translational modification status have been found to be important in activation of TIR-NLRs such as RRS1. Guo et al. reported that when expressed in planta, RRS1 allele RRS1-R is maintained in a repressed state by phosphorylation at a site in the integrated WRKY domain. Upon PopP2 effector interaction with RRS1-R, O-acetylation at the WRKY domain is carried out which prevents the phosphorylation and therefore initiates activation of the NLR pair. The N- and C- termini of RRS1-R are brought into close proximity and the repression of RPS4 -TIR domains is lifted, allowing activation of a RPS4 dependent cell death (Guo et al., 2020).

Taking all of this into account, to further understand the dynamics of the Pik-NLR pair, I therefore set out to determine oligomeric state using blue native-PAGE, and subcellular localisation of the using fluorescence microscopy.

4.2 Methods

4.2.1 *Agrobacterium* infiltration of *N. benthamiana* for fluorescence microscopy

N. benthamiana were infiltrated as for other experiments, but at a lower OD₆₀₀ of 0.1 per construct. *Agrobacterium* transformed with an empty vector control was used to make OD₆₀₀ equal across all infiltration mixtures, typically to a maximum of OD₆₀₀ = 0.75 leaves were harvested into a humidified plastic box, 3 d.p.i, directly before the microscopy session.

4.2.2 Confocal microscopy

Confocal microscopy was performed using a Zeiss 880 Airyscan confocal microscope under control of ZEN (Black) software (Zeiss) with support from Dr Sergio Lopez of the JIC BioImaging platform, and from Dr Indira Saado. *Nicotiana benthamiana* leaves expressing fluorescent proteins of interest were harvested and kept in a humidified plastic box. Each sample was prepared immediately prior to imaging. Leaf disks were prepared using a 4 mm diameter biopsy punch and infiltrated with water by manual vacuum infiltration inside a syringe barrel. Leaf disks from two or three different plants were used. The leaf disks were then placed on glass slides and covered with a 1.5 thickness cover slip before mounting in the microscope and imaging. Samples were visualised with bright light, and relevant laser light for the fluorophores present in the sample: Argon-458 nm (CFP), Argon-488 nm (GFP), and 594 nm (mCherry), and were imaged using a 40X water immersion objective. Z-stacks and images at a single Z-height were obtained. Images were processed using FIJI (ImageJ). The channels were split, for bright field Z-stacks a single Z-slice was selected and for fluorescence Z-stacks a Z-projection of a subset of the Z-stack was produced using the average intensity values for each pixel. Brightness and contrast were adjusted and after application of appropriate look-up-tables to match the fluorophore (CFP = cyan, GFP = green, mCherry = magenta), the channels were merged and the merged micrograph exported. A second version of the same micrograph was exported with a scale bar, to refer to when drawing the vector scale bar. Brightness and contrast of images were adjusted using Microsoft PowerPoint and finally using Inkscape, multi-panel images were assembled, vector scale bars were added, and high quality .png files were exported.

4.3 Results

Dr Indira Saado and Dr Sergio Lopez provided training and assistance with confocal microscopy. BN-PAGE was performed with advice from Josh Bennett, according to a protocol developed by Dr Hee-Kyung Ahn and modified by Dr Mauricio Contreras.

4.3.1 Pikm-1-mCherry, Pikm-2-GFP, and CFP-AVR-PikD can be observed by fluorescence microscopy

For fluorescence microscopy experiments, the Pik pair was C-terminally tagged with monomeric fluorophores mCherry and mEGFP to produce Pikm-1-mCherry and Pikm-2-GFP, and each was placed under the control of a p35S promotor. AVR-PikD was N-terminally tagged with CFP and placed under the control of the pMAS promotor, and all constructs were expressed in *N. benthamiana* by agroinfiltration.

First, a cell death assay was performed to assess the effector response of Pikm-1-mCherry/Pikm-2-GFP in comparison to that of Pikm-1-HF/Pikm-2-HA, and to ensure that the fluorophore tagged constructs CFP-AVR-PikD or Pikm-1-mCherry/Pikm-2-GFP are not visible by UV photography in absence of cell death (as this would prevent scoring of cell death in UV photographs). I observed effector dependent cell death by Pikm-1-HF/Pikm-2-HA in response to AVR-PikD, and to CFP-AVR-PikD, but not in response to AVR-PikF, as expected. From this we conclude that the larger CFP tag is not preventing recognition by the Pik NLR pair. I also observed that when expressed alone, with no NLRs, CFP-AVR-PikD produced cell death areas which scored only 0 or 1, indicating that CFP is not eliciting any cell death response and nor is the CFP significantly visible by UV photography. Expression of Pikm-1-mCherry and Pikm-2-GFP together with no effector resulted in essentially no cell death (occasionally a couple of areas scored 3 or 4, as is often seen in such assays). Pikm-1-mCherry/Pikm-2-GFP responded to Myc-AVR-PikF, Myc-AVR-PikD, and CFP-AVR-PikD with cell death scores which were indistinguishable from that of the Pikm-1-HF/Pikm-2-HA pair. Therefore, the introduction of the larger fluorophore tags did not make any significant difference in response; only effector dependent cell death was observed (Figure 4-1). From this we conclude that these constructs are suitable for fluorescence microscopy.

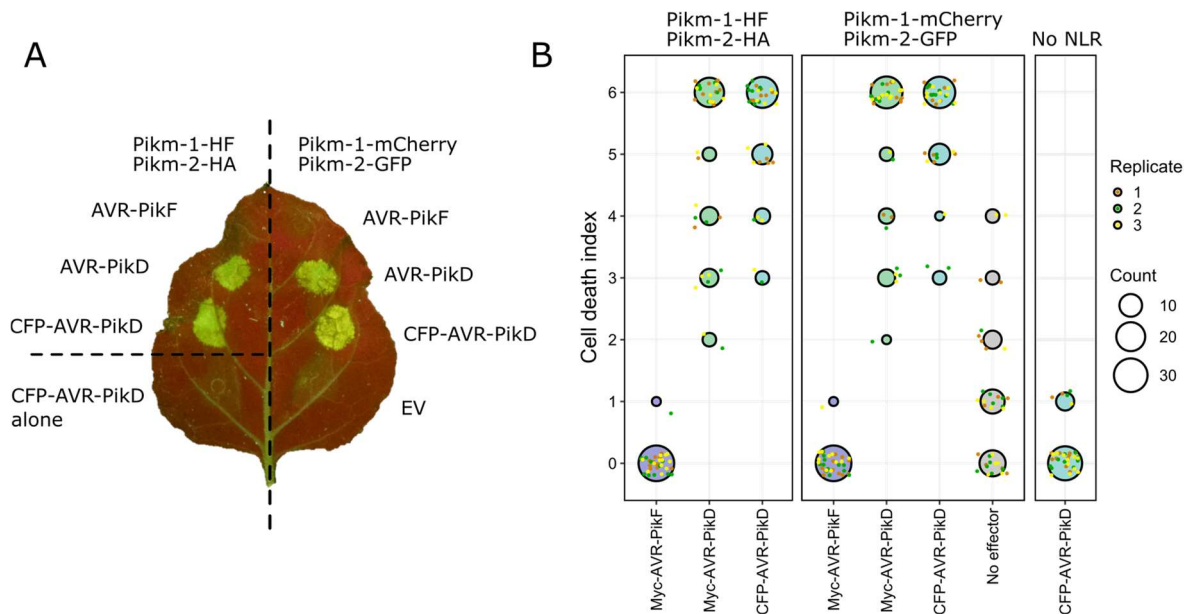


Figure 4-1: *N. benthamiana* cell death assay of CFP-AVR-PikD, Pikm-1-mCherry, and Pikm-2-GFP

A) Representative UV photograph of *N. benthamiana* leaf cell death assay. **B)** Quantification of *N.*

benthamiana cell death according to scale in Maqbool et al. 2015 (Figure 2-1). Experiment was performed in three independent replicates, with a total of 36 leaf spots per condition.

The constructs were then agroinfiltrated into *N. benthamiana* for fluorescence microscopy. The *Agrobacterium* density of each construct at inoculation was reduced to $OD_{600} = 0.25$ to avoid excessive over-expression, and leaves were observed at 3 d.p.i. by confocal microscopy (4.2.2). CFP-AVR-PikD was observed in a cytoplasmic distribution, including in cytoplasmic strands visible crossing the vacuole, and at high intensity in and around the nucleus.

Pikm-1-mCherry was observed at the periphery of the cell, and within the nucleus. Overall fluorescence was low, and so imaging of Pikm-1-mCherry was carried out with higher gain settings. Since increasing the overexpression increased the risk of introducing over-expression artifacts, it was decided not to assess Pikm-1-mCherry over-expression plasmids with alternative promoters (although these have been generated and are a resource available for any future work). Pikm-2-GFP was observed at the cell periphery and at the periphery of the nucleus, but not within the nucleus. The position of the nucleus can also be observed in the bright field images. Upon co-expression of Pikm-1-mCherry and Pikm-2-GFP by co-infiltration, it was difficult to locate cells in which both constructs were simultaneously expressed to a suitable level for visualisation. In the cells which exhibited co-expression, the localisation of Pikm-1-mCherry and Pikm-2-GFP appeared similar; fluorescence corresponding to both proteins was observed at the cell periphery. Nuclei were

not easily observed, and are not pictured in the representative micrograph below. Therefore this experiment is inconclusive as to the potential change in localisation of the NLRs from their previous positions, which were within (Pikm-1-mCherry) and at the periphery (Pikm-2-GFP) of the nucleus in addition to at the cell periphery. Furthermore, in the co-expression of Pikm-1-mCherry and Pikm-2-GFP, the cell periphery appears less well defined, which could be indicative that the cell is losing integrity – which could be due to experimental error (for example, dehydration of the leaf), or due to cell death caused by overexpression of these proteins. Although cell death was not observed on a gross scale in the cell death assay upon co-expression of Pikm-1-mCherry and Pikm-2-GFP, it is possible that these proteins are causing changes which can be observed on a cellular scale.

A single replicate of this experiment was performed and is presented here, and so further inferences can only be made after imaging multiple cells of interest across at least two more biological replicates, although this work provides a baseline for observation of Pikm-1-mCherry and Pikm-2-GFP in *N. benthamiana*.

Chapter 4: Probing subcellular localisation and activation of the Pik-NLR pair

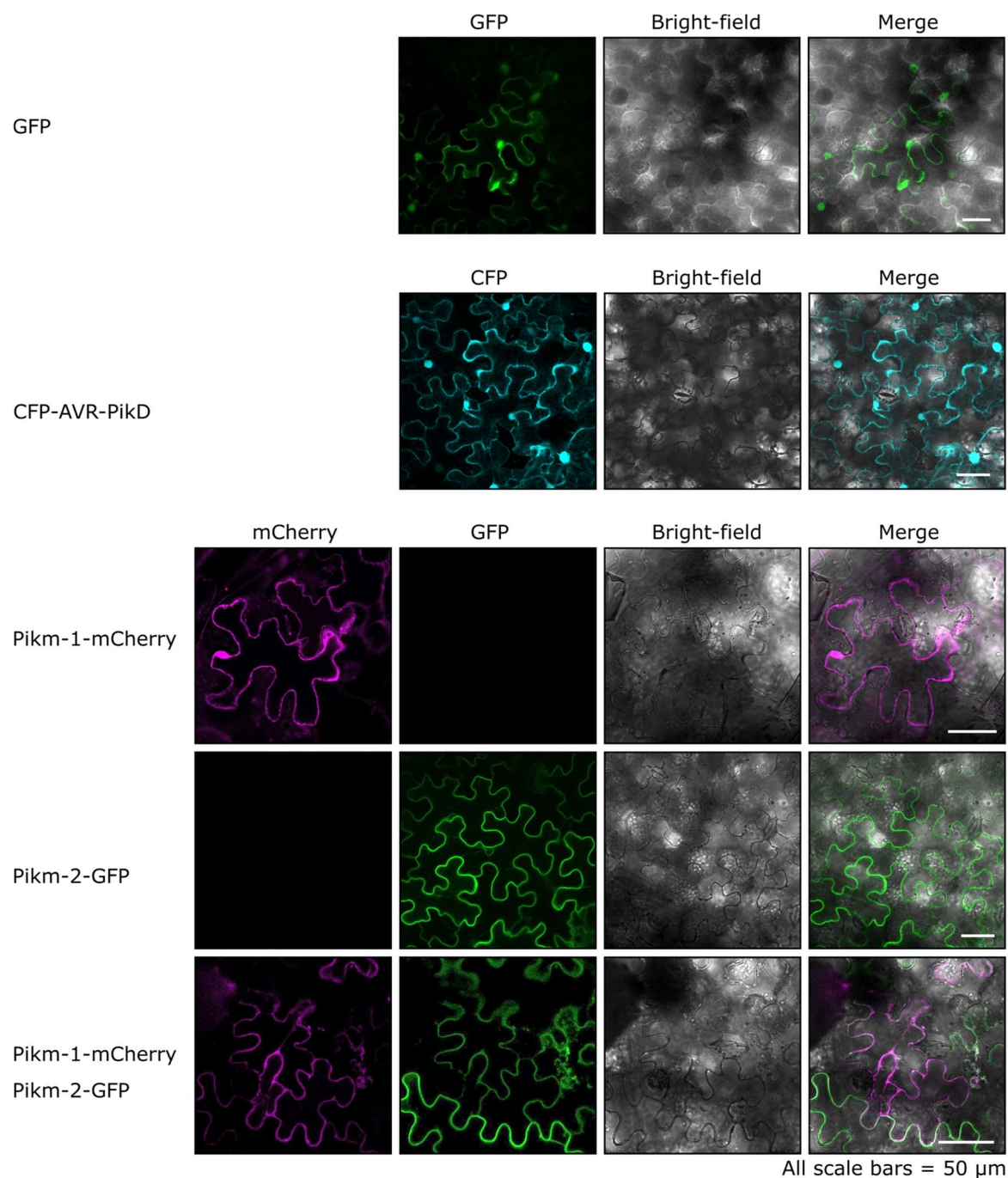


Figure 4-2: Confocal fluorescence microscopy of CFP-AVR-PikD, Pikm-1-mCherry, and Pikm-2-GFP

Fluorescence and bright field micrographs were obtained and are displayed individually in the left-most column. The merged images are displayed in the right-most column. Scale bars each indicate 50 μm .

Experiment was performed with n=1.

Following independent observation of CFP-AVR-PikD, Pikm-1-mCherry and Pikm-2-GFP, and co-expression of Pikm-1-mCherry + Pikm-2-GFP in *N. benthamiana*, the experiment was extended to co-expression of all three proteins together. However, it could reasonably be expected that cell death would begin to occur very shortly after the onset of expression of the wild-type proteins, as they had previously been demonstrated to cause cell death. As a consequence, the fluorescent proteins may be prevented from accumulating to levels detectable by fluorescence microscopy, or cell death phenotypes which disturb normal cell structures might be microscopically visible before cell death areas are visible by eye. Therefore, an approach to abolish cell death without abolishing complex formation or perturbing complex localisation is required.

To abolish cell death, the $\alpha 1$ helix triple mutant Pikm-2^{L15E/L19E/L23E}-GFP was used. *Nicotiana benthamiana* cell death assays were then performed and cell death was not observed when Pikm-2^{L15E/L19E/L23E}-GFP was co-infiltrated with Pikm-1-mCherry and CFP-AVR-PikD or Myc-AVR-PikD, in contrast to the cell death elicited by Pikm-1-HF/Pikm-2-HA in response to the same effectors (Figure 4-3 A, B). Expression of Pikm-2^{L15E/L19E/L23E}-GFP was confirmed by western blot (Figure 4-3 C).

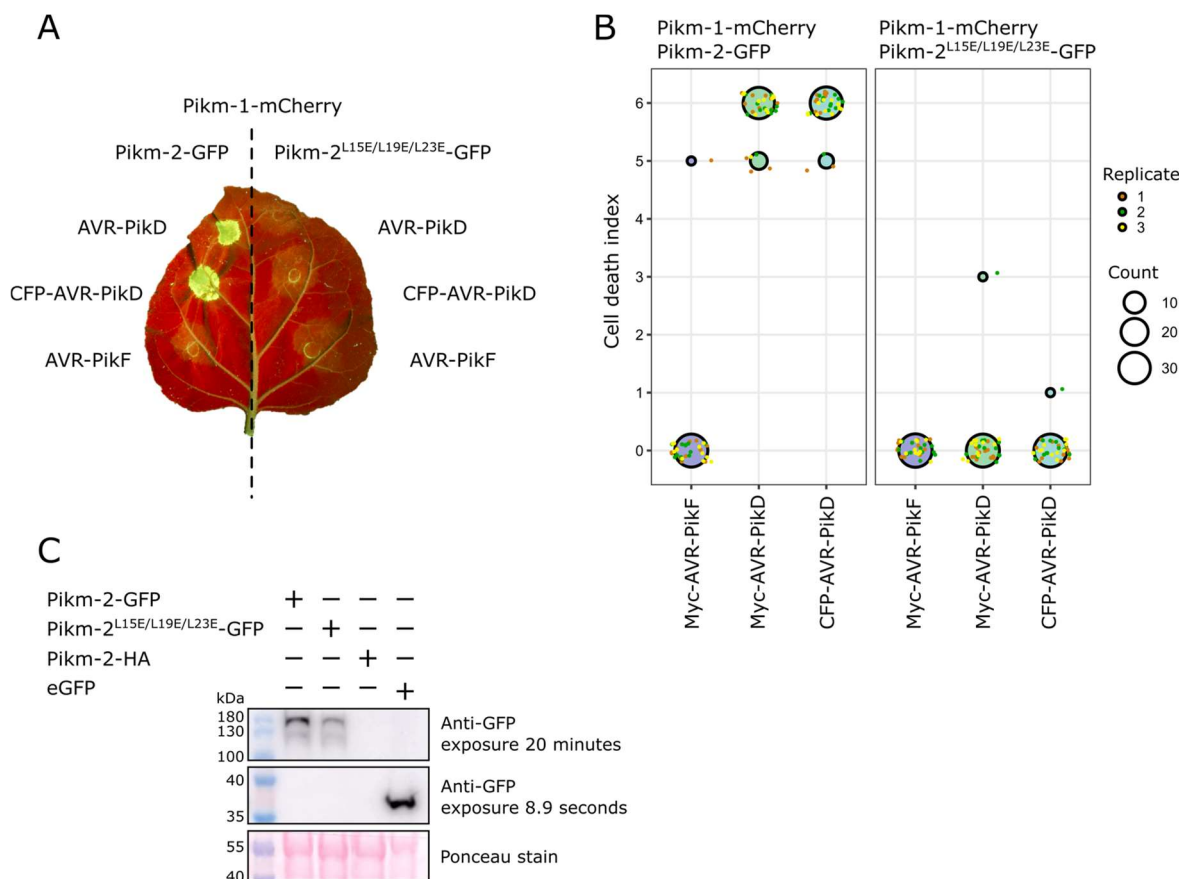


Figure 4-3: The Pikm-2^{L15E/L19E/L23E} α 1 mutation abolishes cell death in *N. benthamiana*

A) Representative UV photograph of *N. benthamiana* leaf cell death assay. **B)** Quantification of *N. benthamiana* cell death according to scale in Maqbool et al. 2015 (Figure 2-1). Cell death experiment was performed in three independent replicates, with a total of 35 leaf spots per condition. **C)** Anti-GFP western blot of protein expression. Pikm-2-HA = negative control, eGFP = positive control. Western blot was performed in triplicate, each replicate using samples expressed in at least two different plants.

Preliminary confocal microscopy of Pikm-2^{L15E/L19E/L23E}-GFP in combination with Pikm-1-mCherry was then performed. The localisation of Pikm-2^{L15E/L19E/L23E}-GFP when co-expressed with Pikm-1-mCherry compared to that of Pikm-2-GFP when co-expressed with Pikm-1-mCherry appears to be comparable. Likewise, the localisation of CFP fluorescence upon co-expression of CFP-AVR-PikD with Pikm-1-mCherry and Pikm-2^{L15E/L19E/L23E}-GFP was cytoplasmic, with fluorescence visible at the cell periphery as well as in and around the nucleus. This pattern of expression appeared similar to when CFP-AVR-PikD was expressed alone.

Chapter 4: Probing subcellular localisation and activation of the Pik-NLR pair

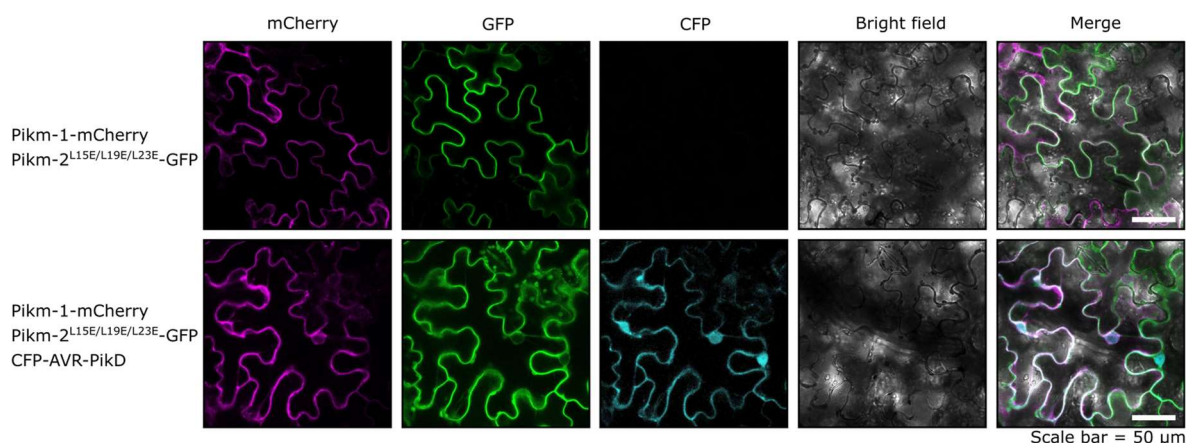


Figure 4-4: Confocal fluorescence microscopy of Pikm-1-mCherry, Pikm-2^{L15E/L19E/L23E}-GFP, and CFP-AVR-PikD

Fluorescence and bright field micrographs were obtained and are displayed individually in the left-most column. The merged images are displayed in the right-most column. Scale bars each indicate 50 μm.

Experiment was performed with n=1.

The aim of these experiments had been to abolish cell death while retaining re-localisation of the NLRs, however use of the of Pikm-2^{L15E/L19E/L23E} mutant was inappropriate for this experiment as these residues are predicted to lie on the outside of the predicted $\alpha 1$ helix funnel and disrupt association with the plasma membrane. Therefore, the lack of re-localisation observed is consistent with the mutants used. Instead, if wild-type proteins are used, a time course could be performed to ascertain if there is a time-point at which fluorescence microscopy can detect the labelled proteins before any microscopically visible cell death processes become apparent. Alternatively, use of LaCl_2 to abolish cell death by blocking Ca^{2+} channels could be explored – although this will also inhibit other ion channels and so may produce physiological responses which interfere with Pik complex visualisation. Therefore, the most elegant option would be to perform screening of $\alpha 1$ funnel inner surface mutants, as these may be able to deliver specific disruption of Pikm-2 $\alpha 1$ funnel channel function without disrupting the assembly of the Pik resistosome, or the overall physiology of the cell.

4.3.2 Blue-Native PAGE analysis of the Pik NLR pair

Blue-Native PAGE has been used to analyse the oligomeric status of multiple NLRs, for example the oligomerisation of NRC2 in response to AVR_{Ram3} perception by Rpi-amr3 (Ahn et al., 2023), of NRG1.2 in response to Avr-Rps4 (Feehan et al., 2023), and of NRC4 in response to Potato virus X capsid protein perception by NLR Rx (Contreras et al., 2023b). I performed preliminary BN-PAGE experiments to see if oligomerisation of the Pik NLR pair could be observed pre- and post-activation by AVR-PikD.

To correspond to tags used by collaborators, and to provide affinity tags which can be used for both detection by western blot and for affinity purification, Pikm-1-FLAG and Pikm-2-StrepII were cloned. SDS-PAGE western blots were performed to assess expression of the constructs (Figure 4-5 A) simultaneously with probing of the oligomeric status via BN-PAGE western blot (Figure 4-5 B).

First, the response of Pikm-1-FLAG or Pikm-2-StrepII to Myc-AVR-PikD was assessed. When expressed alone or with Myc-AVR-PikD, Pikm-1-FLAG was detected in SDS-PAGE - anti-FLAG western blot as a band between 130 – 180 kDa, which corresponds to its expected size as a monomer of 130 kDa. In BN-PAGE - anti-FLAG western blot, when expressed alone, Pikm-1-FLAG was detected as a band between 272 – 545 kDa, with a low-density smear of protein above and below extending in a wide molecular weight range from around 200 – 600 kDa. This could correspond to an oligomer of pre-activated Pikm-1-FLAG. In the BN-PAGE anti-FLAG western blot of co-expressed Myc-AVR-PikD + Pikm-1-FLAG, the same band was detected at a higher intensity, suggesting accumulation, but no significant change in the molecular weight of the band was detected relative to when Pikm-1-FLAG alone was expressed, suggesting that the oligomeric status of the protein had not changed sufficiently to be detected through the smear of different molecular weight complexes.

SDS-PAGE - anti-Strep western blot revealed a band slightly below the 130 kDa marker, which could correspond to the expected molecular weight of Pikm-2-Strep (118 kDa). However, this band was detected in all samples, including those in which Pikm-2-Strep was not agroinfiltrated, and so this band is more likely to represent a non-specific cross reactivity and so for this reason the anti-Strep western blots should be interpreted with caution. Additionally, an intense non-specific band was detected at ~ 55 kDa in all samples which is known to correspond to Rubisco. This same band is present in all samples during BN-PAGE at a molecular weight of ~ 146 kDa and is used as an internal loading control during Ponceau staining (displayed underneath each western blot).

Myc-AVR-PikD was detected in the SDS-PAGE - anti-Myc western blot at a size corresponding to just above the 15 kDa marker, which is consistent with the expected

molecular weight of 15.6 kDa. However, Myc-AVR-PikD was not detected in the BN-PAGE - anti-Myc western blot and therefore the status of its association with potential higher order oligomeric assemblies of either Pikm-1 or Pikm-2 could not be determined. Since AVR-PikD could not be detected by BN-PAGE - anti-Myc western blot, it was not probed for in subsequent BN-PAGE experiments.

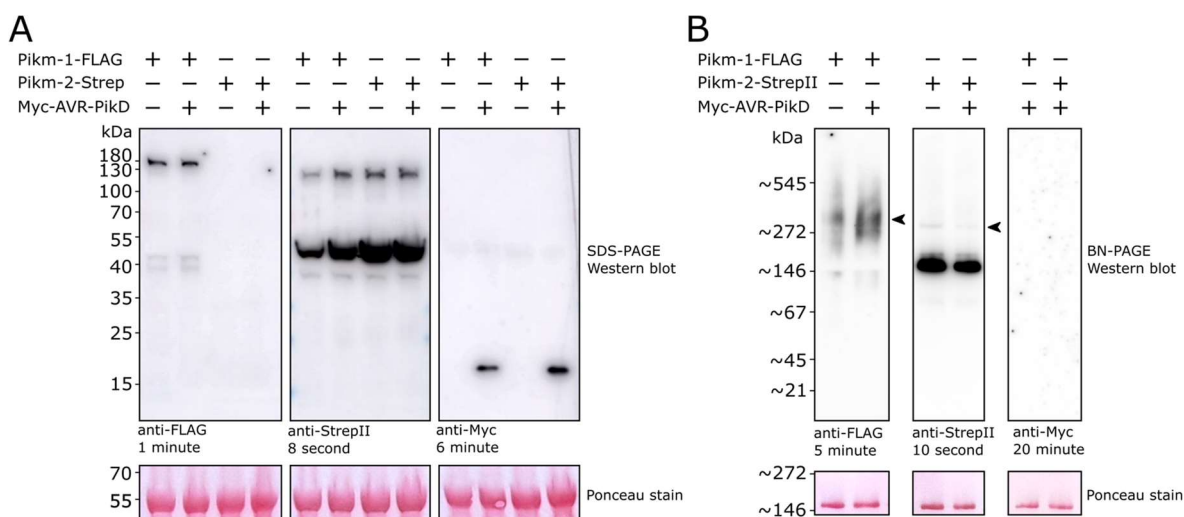


Figure 4-5: BN-PAGE of Pikm-1-FLAG, Pikm-2-Strep, +/- Myc-AVR-PikD

A) SDS-PAGE – Western blot of samples used in this experiment. B) BN-PAGE – Western blot. Arrowheads denote the bands of interest. Antibody and exposure time are indicated underneath each western blot. The intense band in both anti-StrepII western blots is of a size directly corresponding to the high intensity Rubisco bands used as a loading control when detected by Ponceau staining. Experiment was performed with $n=1$.

Subsequently, a BN-PAGE experiment was performed to assess Pikm-1-FLAG + Pikm-2-Strep +/- Myc-AVR-PikD. SDS-PAGE - anti-Myc western blot confirmed the expression of Myc-AVR-PikD only in the sample in which all three proteins were co-expressed. Myc-AVR-PikD was not probed for in the BN-PAGE western blots.

In the SDS-PAGE – anti-FLAG western blot, Pikm-1-FLAG was detected in both the sample with and without Myc-AVR-PikD, although the band intensity was much greater in the sample including Myc-AVR-PikD. This trend was also observed in the corresponding BN-PAGE – anti-FLAG western blot, suggesting that the co-expression of AVR-PikD results in accumulation of Pikm-1-FLAG protein. In the BN-PAGE – anti-FLAG western blot where all three proteins were co-expressed (Figure 4-6 B, lane 2), there was a large smear of protein from ~ 200 kDa up to >545 kDa, which could be accounted for by Pikm-1-FLAG being part of complexes of multiple oligomeric states. Within this smeared region of the gel,

there were three bands of higher density at sizes corresponding to approximately 300 kDa, 400 kDa and 700 kDa.

These bands could correspond to Pikm-1-FLAG protein alone or in complexes of multiple oligomeric states, possibly a heterotrimer of Pikm-1-FLAG/Pikm-2-Strep/Myc-AVR-PikD (~ 263 kDa), and a dimer (~ 526 kDa) and trimer (~ 789 kDa) of such heterotrimers. Other configurations are also possible, for example an inhibited heterodimer of Pikm-1/Pikm-2 which upon AVR-PikD perception becomes uninhibited and nucleates oligomerisation of multiple copies of Pikm-2, resulting in a heteropentamer made up of one copy each of AVR-PikD and Pikm-1, and multiple copies of Pikm-2.

Similarly to Pikm-1-FLAG, Pikm-2-Strep was detected in both the SDS-PAGE anti-Strep western blot and the BN-PAGE – anti-Strep western blot, and in both cases the band intensity was much higher in the sample where all three components are co-expressed. However, since the anti-Strep western blot exhibited non-specificity in the previous experiment, these results cannot be considered reliable.

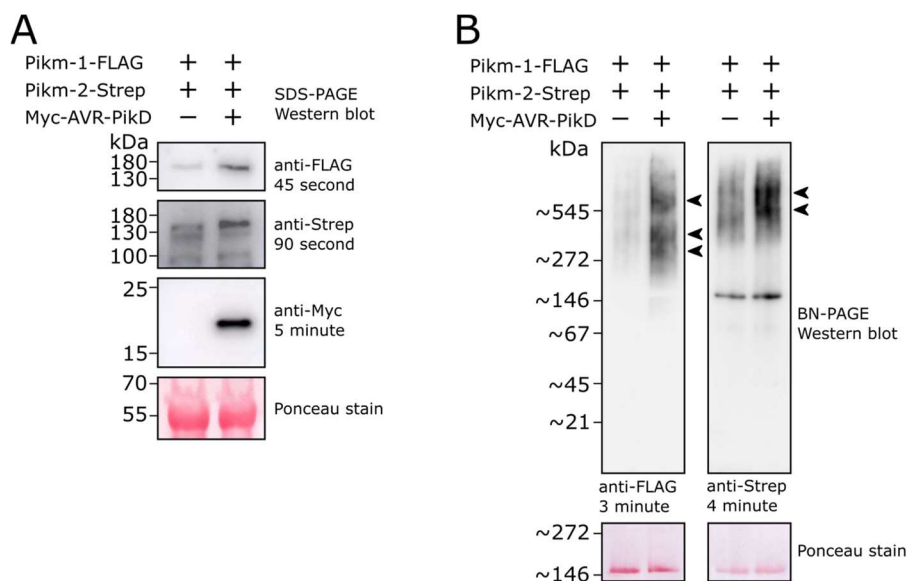


Figure 4-6: SDS-PAGE and BN-PAGE of Pikm-1-FLAG + Pikm-2-StrepII +/- Myc-AVR-PikD

A) SDS-PAGE – Western blot of samples used in this experiment. B) BN-PAGE – Western blot. Arrowheads denote the bands of interest. Antibody and exposure time are indicated underneath each western blot. The ~ 146 kDa band on the BN-PAGE - anti-StrepII western blot is of a size directly corresponding to the high intensity Rubisco bands used as a loading control when detected by Ponceau staining. Experiment was performed with n=1.

It is important to note that both of these BN-PAGE experiments have only been performed once and therefore, these data must be approached with caution. Additionally, the judgement of molecular weights from a BN-PAGE western blot is subjective and so it is not possible to make any conclusions about the exact identity or composition of any given oligomer by precise addition of the expected molecular weights of individual proteins that may be part of the complex.

However, gross changes in complex migration can be observed in these preliminary data. Upon co-expression of Pikm-1-FLAG and Myc-AVR-PikD, a BN-PAGE anti-FLAG western blot band at a molecular weight corresponding to ~ 300 kDa was detected. Upon co-expression of all three constructs, a high molecular weight (> 545 kDa) complex accumulates. Both complexes must contain Pikm-1-FLAG as both bands were observed by BN-PAGE – anti-FLAG western blot. Although not observed on the same western blot, I am confident this does represent a change in oligomeric status for two reasons. First, the presence of the Rubisco band at ~ 146 kDa provides confidence in the assignment of the identity of the bands in the molecular weight marker. Second, the change in migration (from ~ 300 kDa to > 545 kDa) is large relative to the spacing of the molecular weight markers, which provides further confidence that this shift in complex size is real.

The presence or oligomerisation status of Pikm-2-Strep is difficult to conclude by BN-PAGE - anti-Strep western blot, as nonspecific binding was observed in the earlier SDS-PAGE western blot (Figure 4-5 A). However, successful expression of all three proteins can be inferred by the early stages of macroscopically visible cell death that were observed at the time of harvest, 2 d.p.i, and the bands, if genuine indication of Pikm-2-Strep, are compatible with accumulation of a higher order oligomeric Pik complex upon co-expression of CFP-AVR-PikD.

4.4 Discussion

4.4.1 Preliminary observations of the localisation of Pikm-1, Pikm-2, and AVR-PikD were made

Confocal microscopy of the Pik pair was used to obtain preliminary information about the localisation of Pikm-1, Pikm-2, and AVR-PikD. Cell death assays confirmed that the fluorophore-tagging of Pikm-1, Pikm-2, and AVR-PikD constructs did not change the ability of Pikm-1/Pikm-2 to produce cell death in response to AVR-PikD. Western blots should also be performed using anti-GFP and anti-RFP antibodies, to confirm that full length Pikm-1-mCherry and Pikm-2-GFP have been produced, and there is not any free GFP or RFP. When expressed individually, wild-type Pikm-1-mCherry was observed at the cell periphery and within the nucleus, whereas wild-type Pikm-2-GFP was observed at the cell periphery and surrounding the nucleus but not within it. These localisations did not appear to change upon co-expression of Pikm-1-mCherry and Pikm-2-GFP. CFP-AVR-PikD was observed with a cytosolic distribution, including in some cytoplasmic strands spanning the vacuole, and strong fluorescence was observed in the nucleus. This nuclear localisation could explain why Pikm-1-mCherry, which monitors for AVR-Pik effectors, is also present in this subcellular compartment. However, without western blot confirmation, the possibility that free RFP is entering the nucleus while Pikm-1-mCherry remains outside the nucleus cannot be excluded. The detection of all three proteins demonstrates that the Pik NLRs are amenable to fluorescence microscopy, however, I did observe that it was difficult to locate cells with co-expression of the Pik pair, and so future work would benefit from co-expression of both Pik NLRs from the same plasmid to increase the proportion of cells expressing both genes.

Due to inappropriate choice of mutant to abolish cell death, the work to visualise all three co-localised proteins was not informative. Alternative strategies to mitigate the cell death caused by the activated Pik resistosome were discussed in the results section of this chapter (4.3.1). Once strategies to overcome Pik mediated cell death are established, a key extension of this work would be to incorporate expression markers for subcellular features such as the plasma membrane and endoplasmic reticulum. At higher magnification, it might be possible to quantify the co-localisation of the plasma membrane marker and Pik fluorophore to distinguish between cytosolic and membrane localisation. It might also be interesting to perform visualisation in protoplasts, or to use bimolecular fluorescent complementation (i.e. split-GFP), although this technique does have the disadvantage that the inherent affinity between the two parts of the split-fluorophore may drive the interaction between the proteins being assayed.

4.4.2 BN-PAGE reveals a high molecular weight Pik complex which accumulates upon recognition of AVR-PikD

BN-PAGE – anti-FLAG western blot was performed on clarified cell lysates of *N. benthamiana* expressing Pikm-1-FLAG +/- Myc-AVR-PikD and Pikm-1-FLAG + Pikm-2-Strep +/- AVR-PikD. In both cases, addition of AVR-PikD led to increased detection of Pikm-1-FLAG. This could be due to increased accumulation of this protein, or perhaps that binding with AVR-PikD facilitated a change in conformation which better exposed the FLAG epitope to detection by anti-FLAG antibody. As this experiment has been performed once, it is important to perform at least two further biological replicates to determine if the higher density bands are reproduced in subsequent experiments. To overcome the apparent non-specific binding of the anti-Strep antibody, western blot blocking and washing conditions could be optimised, or the tag on Pikm-2 exchanged for an alternative epitope tag such as the V5 tag. An important future direction of this work would then be to perform co-immunopurification (co-IP) of different pre- and post-activation complexes. The advantage conferred by co-IP followed by BN-PAGE and western blot is that only complexes associated with any given complex component should be present in the sample. For example, by purifying using the tag on AVR-PikD, only complexes containing AVR-PikD will be isolated. By probing with antibodies to detect Pikm-1 and Pikm-2, the identity of proteins (though not their relative abundance) could be determined.

Immunopurification also offers the opportunity to reduce non-specific background signal through washing of the affinity resin while the sample is bound. In addition, the process is expected to concentrate tagged protein on the affinity resin, and therefore may offer the additional benefit of producing BN-PAGE gels in which the Pik bands can be detected by Coomassie blue dye and submitted for analysis by mass spectrometry. If the protein complexes can be eluted from the affinity resin at sufficient purity, it could also be possible to analyse the size distribution of the complexes by mass photometry which could offer insight into the stoichiometry of the complexes.

4.4.3 Post-translational modifications of the Pik pair upon activation

A further aspect of the Pik pair pre- and post-activation which is not yet understood is that of post-translational modifications. Informative approaches could include proteomic studies, co-immunoprecipitation of activated Pik complexes and mass spectrometry or western blotting using antibodies which recognise post-translational modifications.

Each method confers particular advantages and disadvantages. Proteomics provides a holistic view of the proteome but requires development of specialised pipelines for data analysis, as exemplified in a recent phosphoproteomics study performed using rice tissue infected with *Magnaporthe* (Doctoral Thesis of Dr Neftaly Cruz-Mireles). These data were collected at 32 hours post-infection to capture proteins involved in cell-to-cell movement from the primary infected cell to neighbouring cells and could be interrogated for information on Pik phosphorylation status. Pik-1/Pik-2 dependent cell death can be observed within 48 hours in *N. benthamiana* assays, and so Pik-1/Pik-2 phosphorylation might be occurring around the 32-48 hours post-infiltration timepoint in rice. However, Pikm-1/Pikm-2 are not present in the rice cultivar CO-39 used in that study, and so were not found among the 2538 phosphorylated rice proteins detected (Dr Neftaly Cruz-Mireles, personal communication). A similar experiment could be conducted using a rice cultivar containing Pik alleles during infection by a *Magnaporthe* strain carrying recognised AVR-Pik effectors, sampled at a range of infection time points between 32 – 48 hours. Alternatively, immunopurification-mass spectrometry analysis using tissue from heterologous expression of the Pik pair and AVR-PikD in *N. benthamiana* could be performed. This has potential to provide precise information on the identity of modified residues, but would require purification of Pik protein at sufficient purity and yield to accurately excise an SDS-PAGE gel band containing the Pik protein. It is well established that heterologous expression of the Pik pair in *N. benthamiana* produces cell death in response to AVR-Pik effectors, so we know that activation is occurring. However, caution should still be taken when interpreting information about PTMs obtained from a heterologous expression system, as they may be different to the native situation in rice. In contrast, western blotting analysis does not require purification but would rely on accessibility of any post-translational modification epitopes to their cognate antibodies and would provide almost no evidence on the identity of modified residues without follow-up mutational analysis.

As affinity purification of a high yield of individual Pik proteins or a Pik complex had not been achieved in my hands, the approach of co-IP followed by immunodetection was selected. Although anti-phosphoserine/threonine western blots were planned, unfortunately no experiments were performed due to time constraints. A positive control phosphorylated NLR could be provided by RRS1 which is maintained in a repressed, phosphorylated state in planta (Guo et al., 2020). In absence of an appropriate NLR control, a clarified lysate sample could instead be used as a more general control as the presence of numerous phosphorylated proteins is expected in cell lysate.

4.4.4 Conclusion

In summary, I have presented preliminary work towards determining both intracellular localisation and oligomerisation status of the Pik pair. However, many questions remain regarding these aspects. I remain very interested in the future developments of this research, as it will contribute towards understanding the mechanisms of Pik NLR activation.

Chapter 5: Interactions of novel *M. oryzae* effector AVR-Mgk1

5.1 Introduction

The work in this chapter is grounded in the genetics work carried out by the longstanding collaborators of the Banfield Lab, the Ryohei Terauchi Lab, (Iwate Biotechnology Research Center, Japan). Their expertise in growing rice cultivars and *Magnaporthe* isolates and performing genetic dissection of resistance to specific strains of *Magnaporthe* facilitated their discovery of the novel effector AVR-Mgk1 and is now published (Sugihara et al., 2023).

5.1.1 Forward genetics studies reveal resistance provided by *Piks*

The *Pik* locus in the rice genome contains at least six alleles, which are typically genetically defined in relation to the specific isolates of *Magnaporthe* that they provide resistance against. However, *Piks* had been an exception to this, as there were no reports of a specific resistance provided by *Piks*.

Two *japonica*-type cultivars of rice, Hitomebore and Moukoto, differ in their response to *M. oryzae* isolates TH30 and O23. Hitomebore displays resistance against *M. oryzae* isolates carrying TH30 and O23; in contrast, Moukoto is susceptible to both of these isolates. Forwards genetics experiments were initiated in search of resistance genes that are responsible for the differential resistance.

Using recombinant-inbred-lines (RILs) of rice originating from a cross between the resistant and susceptible rice varieties, the source of resistance was narrowed down to two loci of interest; on chromosome 1, a locus associated with TH30 and O23 resistance, and on chromosome 11, a locus associated with O23 resistance. The chromosome 1 locus was found to contain the NLR *Pish*, and the chromosome 11 locus contained the *Pik* allele *Piks*. RILs such as RIL #58 which had the Hitomebore resistant *Piks* allele and the Moukoto susceptible *Pish* allele were resistant to O23 and susceptible to TH30, implicating *Piks* in resistance against *M. oryzae* isolates carrying O23 (Sugihara et al., 2023).

5.1.2 Novel *M. oryzae* effector AVR-Mgk1 is encoded on a mini-chromosome

To further understand host specificity determining factors of *Pik*, Dr Yu Sugihara and colleagues carried out a screen to search for *M. oryzae* effectors that are recognised by rice strains carrying *Piks* (Sugihara et al., 2023). The candidate source of resistance was narrowed down to a mini-chromosome assembly of *M. oryzae* O23. On this mini-chromosome two copies of a novel effector candidate, AVR-Mgk1 (*Magnaporthe* gene recognised by *Pik*) were found in a head-to-head conformation. The status of AVR-Mgk1 as an effector recognised by *Piks* was then confirmed by punch inoculation assays of *Magnaporthe* carrying AVR-Mgk1 on rice plants. Further inoculation experiments determined that AVR-Mgk1 is widely recognised by *Pik* NLRs; rice cultivars carrying *Pikm*, *Pikp*, and *Pik** also displayed resistance to AVR-Mgk1. However, phylogenetic analysis using the TRIBE-MCL algorithm placed AVR-Mgk1 into a different tribe to the AVR-*Pik* effectors (Sugihara et al., 2023).

5.1.3 Small differences in sequence can underlie recognition specificity of *Pik*

The recognition specificity of different *Pik* alleles can vary greatly, even when the differences between alleles are subtle. The *Pik* pair *Piks*-1/*Piks*-2 does not recognise the *M. oryzae* effectors AVR-*PikA*, AVR-*PikB*, AVR-*PikC*, AVR-*PikD*, or AVR-*PikE*, whereas *Pikm*-1 does recognise AVR-*PikD*, AVR-*PikE*, and AVR-*PikA* but not AVR-*PikC* or B. This specificity is attributed to only two amino acid polymorphisms in the HMA domain of *Piks*-1 and *Pikm*-1; the rest of *Piks*-1 and *Pikm*-1, as well as the entirety of *Piks*-2 and *Pikm*-2 are completely identical (Figure 5-1 A, B).

The amino acid differences between *Piks*-1 and *Pikm*-1 are located at the binding interface between *Pikm*-HMA and AVR-*PikD* (Figure 5-1 B, C) (De La Concepcion et al., 2018). At position 229, at the start of beta strand β_3 , the uncharged glutamine of *Pikm*-1 (Q229) is substituted for the negatively charged glutamic acid of *Piks*-1 (E229). At position 261, just after the end of beta strand β_4 , the hydrophobic valine of *Pikm*-1 (V261) is substituted for the smaller, but still hydrophobic alanine of *Piks*-1 (A261). Based on the *Pikm*-HMA/AVR-*PikD* binding interface, these polymorphisms are found at binding interfaces 2 and 3 respectively (Figure 5-1 C)

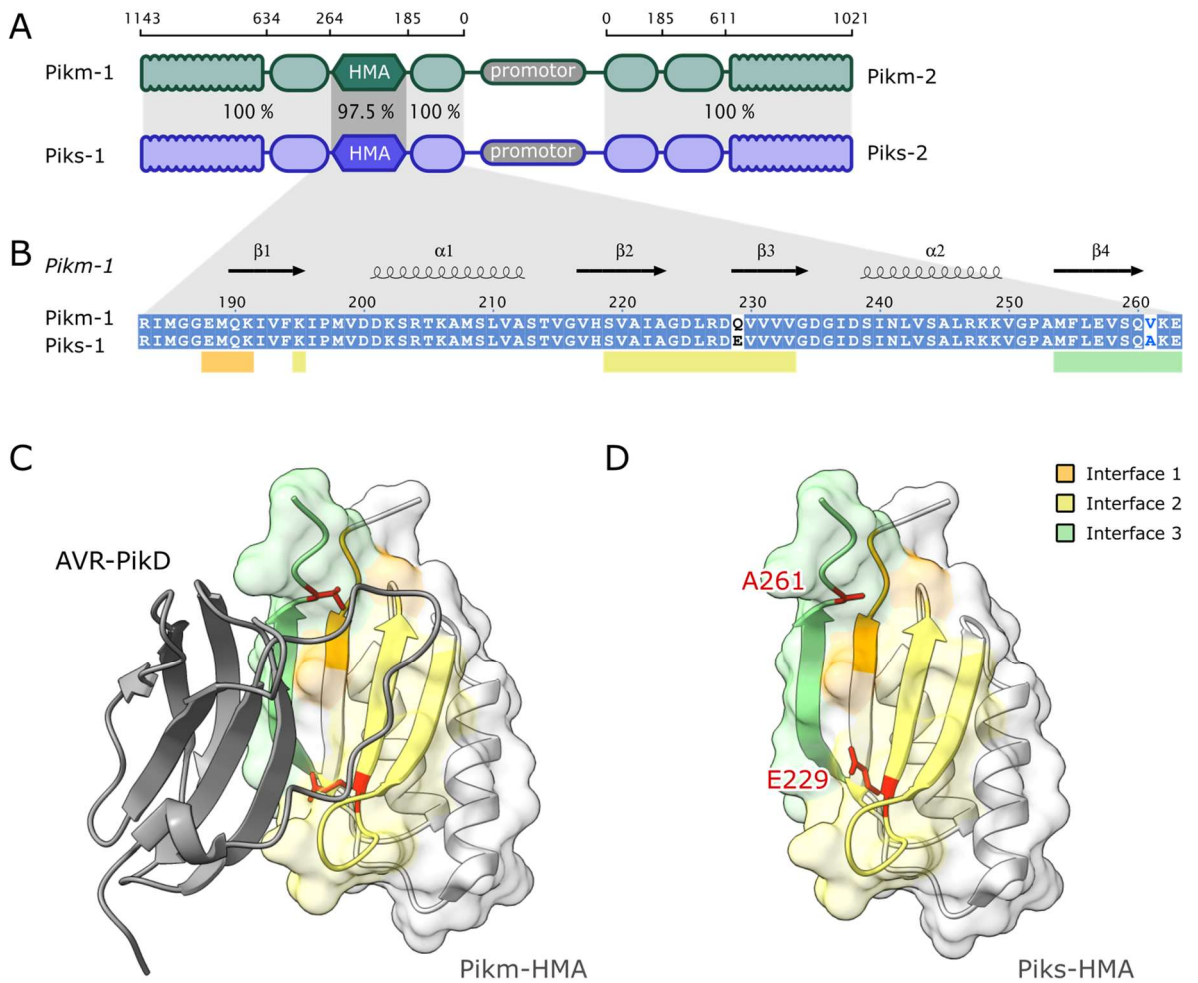


Figure 5-1: The HMA domains of Pikm-1 and Piks-1 have 98% sequence identity.

A) Schematic of the Pikm-1/Pikm-2 and Piks-1/Piks-2 NLR pairs indicating percentage sequence identity between equivalent domains. **B)** Sequence alignment of Pikm-HMA and Piks-HMA. Residue numbering is with respect to full length Pikm-1 and Piks-1. Secondary structure elements of Pikm-HMA are displayed above the alignment (PDB: 6FUD, (De La Concepcion et al., 2018)). AVR-PikD binding interfaces are highlighted below the alignment according to the colour key. Pairwise sequence alignment performed using EMBOSS Needle and visualised using ESPrnt 3.0. Regions in blue with white letters indicate complete sequence agreement, blue letters indicate similarity between amino acids, and black letters indicate dissimilar amino acids. Residue numbering displayed above the alignment is relative to the full length Pikm-1/Piks-1 sequence. **C)** The crystal structure of Pikm-HMA/AVR-PikD (PDB: 6FUD, (De La Concepcion et al., 2018)) displayed with polymorphic residues highlighted. Pikm-HMA is coloured pale grey and AVR-PikD is coloured dark grey. **D)** Polymorphic residues of Piks-HMA displayed mapped onto the crystal structure of the Pikm-HMA structure and highlighted in red.

5.1.4 AVR-Mgk1 is predicted to be a member of the MAX fold family

Despite being recognised by multiple alleles of the Pik NLR, AVR-Mgk1 has only around 10% amino acid sequence similarity with other AVR-Pik effectors (Sugihara et al., 2023). AVR-Mgk1 also shares some of the hallmarks of a MAX effector as defined by de Guillen, Ortiz-Vallejo et al; approximately 6 predicted beta-sheets, two cysteines approximately 40 amino acids apart which would be predicted to form a disulfide bond, and a N-terminal signal sequence (de Guillen et al., 2015). The MAX effector family is a sequence unrelated family; therefore, it was of great interest to determine if AVR-Mgk1 was indeed a MAX effector. Additionally, Pik NLRs recognise MAX effectors via different interfaces on their integrated HMAs and it was unknown which HMA interface might be responsible for AVR-Mgk1 binding.

5.1.5 Advances in protein structure prediction provide increasingly valuable research tools

Structural prediction large language models (LLMs) are continually developing. At the time that I became involved in this project in early 2021, protein structure prediction model performance was improving at a rapid rate. AlphaFold had recently emerged at CASP14, exceeding the performance of RoseTTAFold, and then later in 2021 became available for public use in its updated iteration, AlphaFold2 (Jumper et al., 2021; Pereira et al., 2021; Evans et al., 2022).

In August 2021, RoseTTAFold (Baek et al., 2021) produced AVR-Mgk1 structural predictions which appeared to adopt a MAX fold (Figure 5-2 A). In contrast, early releases of AlphaFold2 produced a variety of elongated predicated AVR-Mgk1 structures which featured extensive coiled or unstructured regions (Figure 5-2 B).

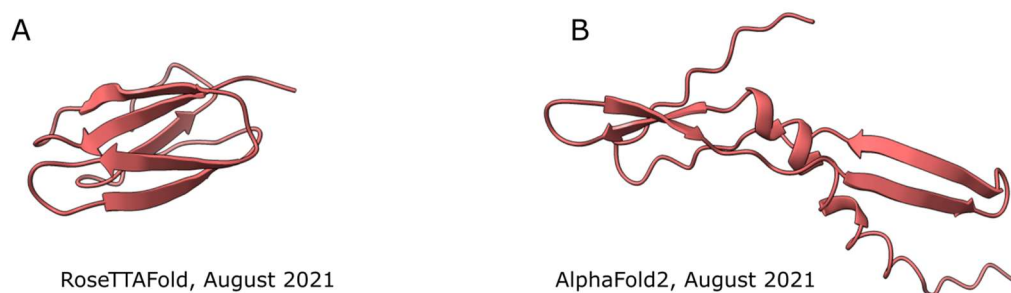


Figure 5-2: Structural prediction of AVR-Mgk1 produced varied results

Structural predictions of AVR-Mg1 carried out in August 2021 A) Using RoseTTAFold, B) using AlphaFold2. Models are superimposed using ChimeraX matchmaker command which executes a sequence alignment between the structures and then fits the aligned residue pairs in 3D space.

Fungal effectors display a range of novel sequences. In keeping with this, the sequence of AVR-Mgk1 has high novelty and when a BLAST search was performed, no similar protein sequences were retrieved. This means that AlphaFold2 and RoseTTAFold, which perform sequence alignments as part of their prediction procedures, will have less prior information when attempting a structural prediction of AVR-Mgk1 when compared to other targets which may be part of large protein families which include previously determined structures.

For the reasons outlined above, the decision was made to pursue an experimentally determined structure of AVR-Mgk1. To date, in vitro work on the *Pik* NLR has largely been done by studying HMA domains in isolation, not in the context of the full-length protein. This is for two reasons. First, that expression and purification of full-length *Pik* has remained elusive despite the efforts of others, and myself as documented in Chapter 3. The second reason is that macromolecular X-ray crystallography as employed in many of these previous studies is most likely to succeed with small, stable, and inflexible proteins. Following these considerations, the construct design offering the greatest chance of success would be to use the domains required for interaction and exclude as much as possible of any flexible surrounding sequences, or domains that may be flexible relative to each other. Therefore, following the precedent set over the past decade of *Pik*-HMA research, in this chapter I set out to study the interaction between AVR-Mgk1 and *Piks*-HMA using protein X-ray crystallography.

5.2 Methods

5.2.1 Expression and purification of protein complexes from *E. coli*

In the case of producing the AVR-Mgk1/Piks-HMA, AVR-Mgk1/OsHPPo2^{K77}, and AVR-Mgk1^{T81}/OsHPPo2^{E74} complexes, both proteins were expressed with affinity tags, then purified individually to the point of tag cleavage with 3C. Once the tags had been removed by reverse IMAC, the two proteins of the complex were mixed and allowed to associate on ice for up to an hour, before concentration in a centrifugal concentrator and application to the final gel filtration column as detailed above. This is the method which is described in the doctoral thesis of Dr Josephine Maidment.

5.2.2 Mass spectrometry

5.2.2.1 Intact liquid chromatography-mass spectrometry (LC-MS)

Intact mass spectrometry and analysis was performed by Dr Carlo Martins of the JIC mass spectrometry platform by LC-MS using a Synapt G2 Si mass spectrometer coupled to an Acquity UPLC system (Waters). Protein samples in buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl), were diluted before application to an Aeris WIDEPORE 3.6 μ m C4 column, 2.1 mm x 50 mm (Phenomenax) and eluted with a linear gradient of acetonitrile in water with 0.1% formic acid from 5 % to 95 % in 5 minutes with a flow rate of 0.4 ml/min. The mass spectrometer was operated with a capillary voltage of 2.4 kV and a cone voltage of 40 V in positive MS-TOF and resolution mode, under the control of the Masslynx 4.1 software (Waters). Calibration was performed in the m/z range of 50 - 2000 using sodium formate according to manufacturer instructions. Every 30 seconds leucine-enkephalin peptide (0.5 μ M in 50% methanol, 0.1 % formic acid, Waters) was infused at 10 μ L/min as an internal standard to calibrate against drift. The data were then processed in Masslynx 4.1 software; first spectra were combined using the background subtract and smooth options, before determination of protein mass by deconvolution using the MaxEnt1 option.

5.2.2.2 Liquid chromatography-mass spectrometry (LC-MS)

Electrospray ionisation mass spectrometry and analysis was performed by Jan Sklenar of the TSL proteomics platform using an Orbitrap Fusion mass spectrometer coupled to a nanoflow UHPLC system U3000 (both Thermofisher Scientific), according to previously published method (Li et al., 2023). Analysis of peptide fragments was carried out using the Mascot server v2.8 (Perkins et al., 1999). AVR-Mgk1 and Piks-HMA sequences were uploaded to the server. The target false discovery rate was set to 1 %, and both carbamidomethyl and oxidation peptide modifications were included in the search.

Monoisotopic masses were searched for with a peptide mass tolerance of ± 10 ppm and fragment mass tolerance of ± 0.6 Da. Results were exported and summary tables prepared using Microsoft Excel.

5.2.3 Protein-protein interactions

5.2.3.1 Analytical Size Exclusion Chromatography (SEC)

Proteins were diluted to 120 μ M in buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl), incubated on ice for 30 minutes, before centrifugation at 21 K RCF for five minutes at 4 °C. Proteins were then injected by glass syringe (Hamilton) run over a Superdex™ 75 10/300 size exclusion column using an AKTA purifier system (both Cytiva) at 4 °C and a flow rate of 0.4 ml/min. Fractions of 0.5 ml were collected and when applicable, retained for SDS-PAGE analysis. Chromatograms were exported and displayed using Microsoft Excel.

5.2.3.2 Surface Plasmon Resonance (SPR)

SPR was performed using a Biacore 8K instrument (Cytiva) equipped with a Series S CM5 Sensor Chip (Cytiva). The sample compartment was maintained at 4 °C and the flow cell at 25 °C. HMA domain proteins Piks-HMA, Pikm-HMA, OsHPPo2, OsHPPo3, OsHPPo4, and effector proteins AVR-PikF, AVR-PikD, and AVR-Mgk1 were individually purified from *E. coli* and the affinity tags cleaved. AVR-PikD protein was purified by Dr Rafal Zdrzalek and AVR-PikF protein was purified by Dr Adam Bentham.

Before amine coupling, pH scouting was carried out. The HMA proteins were diluted in 10 mM sodium acetate buffers at pH 4.0, 4.5, 5.0, and 5.5 (Cytiva). The buffer pHs chosen for immobilisation were as follows: Piks-HMA (pH 4.0), Pikm-HMA (pH 4.5), OsHPPo2, OsHPPo3, OsHPPo4 (pH 5.5). Immobilisation was carried out using the Amine Coupling Kit (Cytiva). Briefly, the chip surface was activated by N-hydroxysuccinimide (NHS) and 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC), before addition of HMAs to the chip. Application of HMAs on the chip was repeated to top up some channels and reach a total of ~ 1100 RU $\pm$ 280 RU. To complete the amine coupling, 1 M ethanolamine-HCl pH 8.5 was applied to the chip in a separately controlled blocking step.

In initial experiments a two-phase regeneration was carried out (1 M NaCl, 20 mM Sodium acetate, pH 4.0, for 60 seconds at 100 μ L/min, followed by 1 M NaCl, 20 mM glycine-HCl pH 3.5, for 30 seconds at 100 μ L/min) however not all channels were returning to baseline response unit level, indicating that effectors remained bound to the HMAs on the chip. Thereafter, a higher stringency denaturing regeneration buffer was used after each cycle (1.83 M MgCl₂, 0.92 M urea, and 1.83 M guanidine-HCl), over both flow cells for 60 seconds

at 100 $\mu\text{L}/\text{min}$ followed by a system wash and HBS-EP+ for 600s to allow time for protein refolding. The regeneration buffer was adapted from 3x ionic regeneration buffer stock (Andersson et al., 1999) by it using without three-fold dilution (Bentham et al., 2023) and by omission of Potassium thiocyanate.

In pilot experiments a Multi Cycle Kinetics approach was followed. However, this was changed to a Single Cycle Kinetics approach due to the slow and incomplete dissociation of effectors from the chip, and due to the introduction of the higher stringency regeneration buffer. There are fewer regenerations required in a SCK experiment, which minimises the potential for accumulated damage to the HMAs on the chip from long exposure to harsh regeneration reagents.

The running buffer, also used to dilute the effector proteins, was HBS-EP+ (10 mM HEPES pH 7.4m 150 mM NaCl, EDTA 3 mM, Tween20 0.05%). Single Cycle Kinetics experiments were conducted to determine the affinity and kinetics of the HMA/effector interactions. Purified effector proteins were sequentially injected over the chip at 0 nM followed by 8 increasing concentrations for a contact time of 110 s each, before a dissociation phase of 600 s. The flow rate was 95 $\mu\text{L}/\text{min}$. The effectors were diluted in a three-fold dilution series (1000 nM top concentration for all analytes apart from Pikm-HMA – AVR-PikF, Piks-HMA – AVR-PikD, and OsHPPo4^{K77} – AVR-PikD where a top analyte concentration of 3000 nM was used) and a 0 nM condition was also included. During the experiment, each cycle was repeated four times.

The data were analysed using Biacore evaluation software Kinetics model 1:1 binding with default parameters for k_a , k_d , R_{max} , t_c , RI , and with the modification of global fit for the drift parameter. The data from the first cycle for each effector was excluded from the analysis as the response was higher in the first run and the three subsequent runs were highly reproducible. SPR data were exported in .txt format and were plotted using R 4.3.2, RStudio (Posit, 2023), using R scripts from Dr Adam Bentham which principally employ the tidyverse collection of R packages and the R package ggplot2 (Wickham, 2016).

5.2.4 Protein structure prediction

Protein structure prediction was performed using RoseTTAFold via the Robetta service of the Baker lab (accessed August 2021) (Baek et al., 2021), and using AlphaFold2 (Jumper et al., 2021) and AlphaFold2 multimer (Evans et al., 2022) via the ColabFold implementation of AlphaFold2 hosted on Google CoLaboratory (accessed between August 2021 and August 2024) (Mirdita et al., 2022). PAE, MSA, pLDDT plots for each prediction are available as an appendix.

5.2.5 Macromolecular X-ray crystallography

5.2.5.1 Crystallisation screening

Aliquots of purified protein were thawed from the -80 °C freezer, centrifuged at 21K RCF for 5 minutes at 4 °C, and the soluble protein in the supernatant transferred to a new tube. The protein was dispensed into commercially available crystallisation screens, using an Oryx Nano or Oryx8 robot (Douglas Instruments) under the control of WASPRUN and WASP software (Douglas Instruments). The sitting drop vapor diffusion method was performed using MRC 2 drop crystallisation plates, allowing for two concentrations of the dispensed protein for each screen condition.

Optimisation screens were designed in XSTEP software (Douglas Instruments) and dispensed using the Oryx Nano or Oryx8 instruments. Following this, protein was dispensed into the screen as above.

Seed stocks were prepared by collecting the contents of the reservoir and droplet into a 1.5 ml microtube and vigorously mixing with a pipette and vortexing to fragment the seed material. After setting up screens with the seed material, the remaining seed stock was aliquoted into small tubes, flash cooled in liquid N₂ and stored at -70 °C.

Plates were kept at 20 °C and imaged regularly to monitor any crystal development using Rock Imager (Formulatrix) or were observed manually under a dissecting microscope. The screens used were BCS, JCSG-plus™, MIDASplus™, Morpheus® I HT-96, PACT premier™, ProPlex™, SG1, and Structure (Molecular Dimensions), KISS (JIC custom screen), and PEGs suite (Qiagen).

Piks-HMA (3 mg/ml) crystals were obtained in Morpheus® HT-96 condition B6 [0.09 M Halogens mix (0.3M Sodium fluoride; 0.3M Sodium bromide; 0.3M Sodium iodide), 0.1 M Buffer System 2 pH 7.5 (Sodium HEPES; MOPS (acid)), 50 % v/v Precipitant Mix 2 (40% v/v Ethylene glycol; 20 % w/v PEG 8000)].

OsHPPo2/AVR-Mgk1 (24.8 mg/ml) crystals were obtained in ProPlex condition H1 (2.0 M Sodium chloride; 0.1 M Sodium citrate pH 6.0).

5.2.5.2 Crystal harvesting

Crystals were cooled in liquid nitrogen in LithoLoops (Molecular Dimensions) by Dr Clare Stevenson or Julia Mundy of the JIC Structural Biology Platform, or by me. The Morpheus screen is formulated with a sufficient concentration of small polyols to provide cryoprotection, all other screens were cryoprotected using well solution with addition of 20 or 30% v/v Ethylene glycol before cooling.

5.2.5.3 Data collection and processing

Cooled crystals were shipped to Diamond Light Source, Oxfordshire, UK, where data collection was carried out at microfocus macromolecular crystallography beamlines I04 and I24. All remote access beamtime was facilitated by Professor David Lawson of the JIC Structural Biology Platform with support from Dr Clare Stevenson and Julia Mundy.

X-ray data were recorded using a Eiger2 16M detector (Dectris) on beamline I04 (Piks-HMA), and a Eiger2 9M detector (Dectris) on beamline I24 (AVR-Mgk1/OsHPPo2). A Cryojet cooler (Oxford Instruments) was used to maintain the crystal temperature at 100 K.

Data were processed on-the-fly by Fast_DP, and xia2 (Winter, 2010) for preliminary assessment of quality, before being downloaded and processed within the CCP4 software suite (Agirre et al., 2023). Data integration and scaling was performed using DIALS (Winter et al., 2018). The data were then merged and reduced using AIMLESS (Evans, 2011; Evans and Murshudov, 2013)

For the Piks-HMA structure, molecular replacement of two copies of Piks-HMA was performed using Mr BUMP (Keegan and Winn, 2007). Pikm-HMA from the previously determined structure of Pikm-HMA/AVR-PikA (De La Concepcion et al., 2018) was suitable for use as a search model for molecular replacement due to its very high sequence identity with Piks-HMA. PHASER (McCoy et al., 2007) and BUCCANEER (Cowtan, 2006) were then used to model the second two copies of Piks-HMA into the remaining density, before successive rounds of refinement using REFMAC (Vagin et al., 2004) and manual model building using Coot (Emsley et al., 2010).

For the AVR-Mgk1/OsHPPo2 complex, data merging and reduction by AIMLESS was followed by molecular replacement of AVR-Mgk1 and OsHPPo2 by PHASER. An AlphaFold2 (Jumper et al., 2021; Mirdita et al., 2022) structural prediction of AVR-Mgk1^{T81} and the Pikm-HMA crystal structure (derived from PDB: 6FUD, as previously) were used for molecular replacement. BUCCANEER was then used to build the AVR-Mgk1^{T81} and OsHPPo2^{E74} sequences by chain tracing before successive rounds of refinement using REFMAC and manual model building using Coot.

5.2.5.4 Visualisation and protein structure analysis.

Advx software (<https://www.scripps.edu/tainer/arvai/advx.html>) was used to display representative diffraction images from the Piks-HMA and AVR-Mgk1/OsHPPo2 datasets. To improve contrast, five 0.1 ° images were summed. Molecular visualisation during model building was performed using Coot (Emsley et al., 2010). Experimentally determined and predicted protein structures were displayed for figure making using UCSF ChimeraX (Goddard et al., 2018; Pettersen et al., 2021). To help with identifying individual secondary

structure elements, a python script made available by the UCSF Resource for Biocomputing, Visualization and Informatics was employed in ChimeraX, (downloaded from https://rbvi.github.io/chimerax-recipes/label_ss/label_ss.html). Interface analysis was performed using the QtPISA server (Krissinel and Henrick, 2007; Krissinel, 2010).

5.3 Results

Mass spectrometry was performed by Jan Sklenar and Dr Carlo Martins. Crystals were harvested and shipped to Diamond Light Source by Dr Clare Stevenson and Julia Mundy. Remote access beamtime at Diamond Light Source was facilitated by Professor David Lawson with assistance from Dr Clare Stevenson and Julia Mundy. Dr Adam Bentham and Professor Mark Banfield provided support and advice in processing of the diffraction data. Surface Plasmon resonance was carried out with support from Dr Abbas Maqbool and Dr Adam Bentham.

5.3.1 Expression and purification of AVR-Mgk1 and Piks-HMA

To commence in vitro studies of the interaction between Piks-HMA and AVR-Mgk1, both proteins were individually expressed and purified. The 24 N-terminal amino acids of AVR-Mgk1 were excluded from the sequence as these are predicted to be a secretion signal peptide and would be cleaved upon secretion from *Magnaporthe*. AVR-Mgk1 was cloned with a N-terminal tandem His-GB1 affinity tag and 3C protease site and was expressed in SHuffle *E. coli*. It was purified by IMAC and gel filtration, followed by tag cleavage and removal by reverse IMAC, and finally gel filtration (2.2.4 *Protein purification from E. coli*) (Figure 5-3).

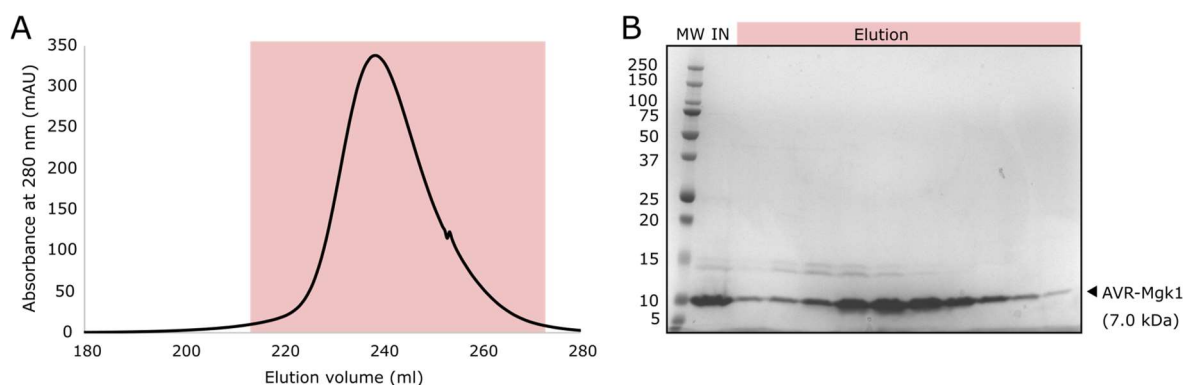


Figure 5-3: Purification of AVR-Mgk1 from SHuffle *E. coli*.

A) Chromatogram of final gel filtration step, **B)** SDS-PAGE of final gel filtration input and eluted fractions. IN = input sample, Elution = eluted fractions corresponding to the area of the chromatogram highlighted in light red.

Similarly, Piks-HMA was cloned with a N-terminal tandem His-GB1 affinity tag and 3C protease site and expressed in SHuffle *E. coli*. Piks-HMA was purified in the same way as

AVR-Mgk1 (2.2.4 Protein purification from *E. coli*) (Figure 5-4). As Piks-HMA lacks aromatic amino acids, it has no absorbance at 280 nm and consequently no significant UV absorbance is detected during chromatography steps once the tag has been cleaved (Figure 5-4).

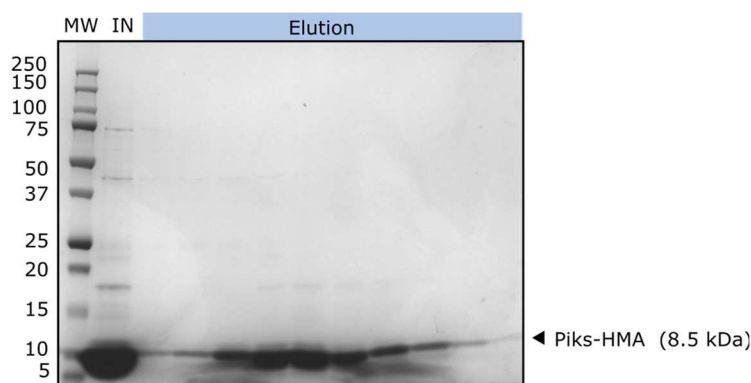


Figure 5-4: Purification of Piks-HMA from SHuffle *E. coli*.

SDS-PAGE of final gel filtration input and eluted fractions. MW = molecular weight marker (kDa), IN = input sample, Elution = eluted fractions.

5.3.2 Direct interaction of AVR-Mgk1 and Piks-HMA demonstrated by analytical gel filtration

To assess if AVR-Mgk1 and Piks-HMA interact in vitro, I performed an analytical gel filtration assay (Figure 5-5). Equimolar quantities of AVR-Mgk1 and Piks-HMA were mixed and incubated on ice for 30-60 minutes before gel filtration on a Superdex 75 10/300 column. The eluted protein from each gel filtration run was fractionated and SDS-PAGE performed on fractions corresponding to the regions around peak absorbance for each were run. Due to lack of aromatic residues in Piks-HMA, the presence of this protein was observed only by SDS-PAGE.

Alone, AVR-Mgk1 is observed as a monodisperse peak at an elution volume of 15.2 ml. Piks-HMA is observed in SDS-PAGE fractions corresponding to elution volume of around 12.5-13 ml. This difference in elution volumes is large considering the similar predicted molecular weight of each protein (AVR-Mgk1, 7.0 kDa; Piks-HMA 8.5 kDa) but my observation of Piks-HMA is consistent with the previous observation of Dr Juan Carlos De la Concepcion that the very similar Pikm-HMA also eluted at around 12.5 ml (Doctoral thesis of Dr Juan Carlos De la Concepcion, p55).

In contrast to the elution of AVR-Mgk1 alone at 15.2 ml, after incubation together AVR-Mgk1 and Piks-HMA elute in a monodisperse peak at 13.2 ml. This is consistent with complex formation bringing AVR-Mgk1 into a larger molecular weight species with a correspondingly lower elution volume. AVR-Mgk1 and Piks-HMA are of similar molecular weight and so could not be individually resolved on a 16% SDS-PAGE gel.

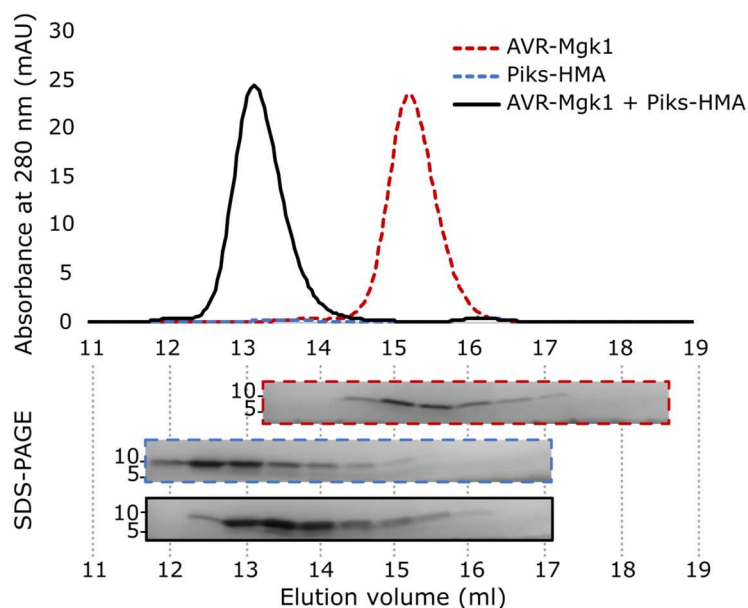


Figure 5-5: Analytical gel filtration of AVR-Mgk1 and Piks-HMA.

Above: chromatograms of analytical gel filtration. Below: SDS-PAGE of corresponding fractions. Red dashed line – AVR-Mgk1, blue dashed line Piks-HMA, black line AVR-Mgk1/Piks-HMA. Numbers on the left hand slide correspond to molecular weight standards (kDa).

5.3.3 Attempts at AVR-Mgk1/Piks-HMA structure determination

Following demonstration of complex formation between AVR-Mgk1 and Piks-HMA in vitro by analytical gel filtration, I then turned towards obtaining an X-ray crystal structure of these two proteins in complex. Quantities of AVR-Mgk1/Piks-HMA complex sufficient for crystallisation trials were prepared by mixing the previously individually purified AVR-Mgk1 and Piks-HMA proteins in 1:1 molar ratio, incubating for one hour on ice, before gel filtration over a Superdex 75 26/60 gel filtration column. The complex eluted as a single peak at a volume of 220 ml (Figure 5-6). This is 20 ml earlier than the 240 ml elution volume observed for AVR-Mgk1 alone (Figure 5-3), indicating that complex formation had occurred and could be resolved on the preparative scale Superdex 75 26/60 column. The peak fractions were identified by SDS-PAGE, collected, and concentrated in a 10 kDa molecular weight cut-off centrifugal concentrator. Two sparse matrix crystallisation screens were then

set up with proteins at two concentrations, 11.7 mg/ml, and 5.8 mg/ml. After two weeks of observation these screens yielded no suitable hits but based on the higher proportion of drops with precipitated protein, subsequent screens were set up with lower concentrations of protein.

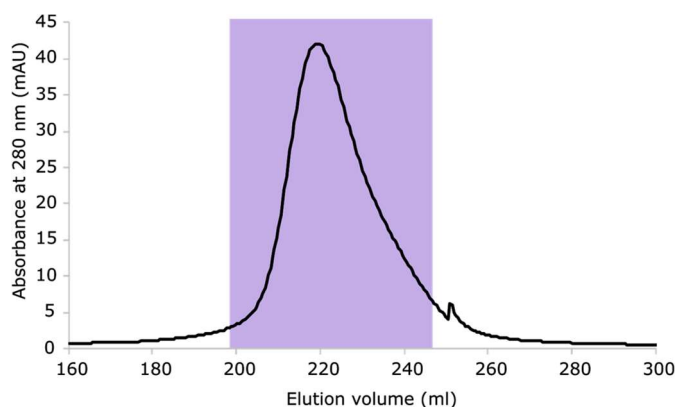


Figure 5-6: Gel filtration chromatograph of AVR-Mgk1/Piks-HMA complex.

Produced by mixture of previously individually purified AVR-Mgk1 and Piks-HMA. The fractions corresponding to the entirety of the peak (highlighted in purple) were collected and concentrated.

For a more efficient large-scale expression and purification of the AVR-Mgk1/Piks-HMA complex, AVR-Mgk1 was re-cloned with no affinity tags into an expression vector for co-expression with His-GB1-Piks-HMA. Expression and purification were performed, yielding 4.6 mg of purified protein (Figure 5-7). As in the analytical gel filtration experiment, once the affinity tag is cleaved, Piks-HMA cannot be separated from AVR-Mgk1 by SDS-PAGE.

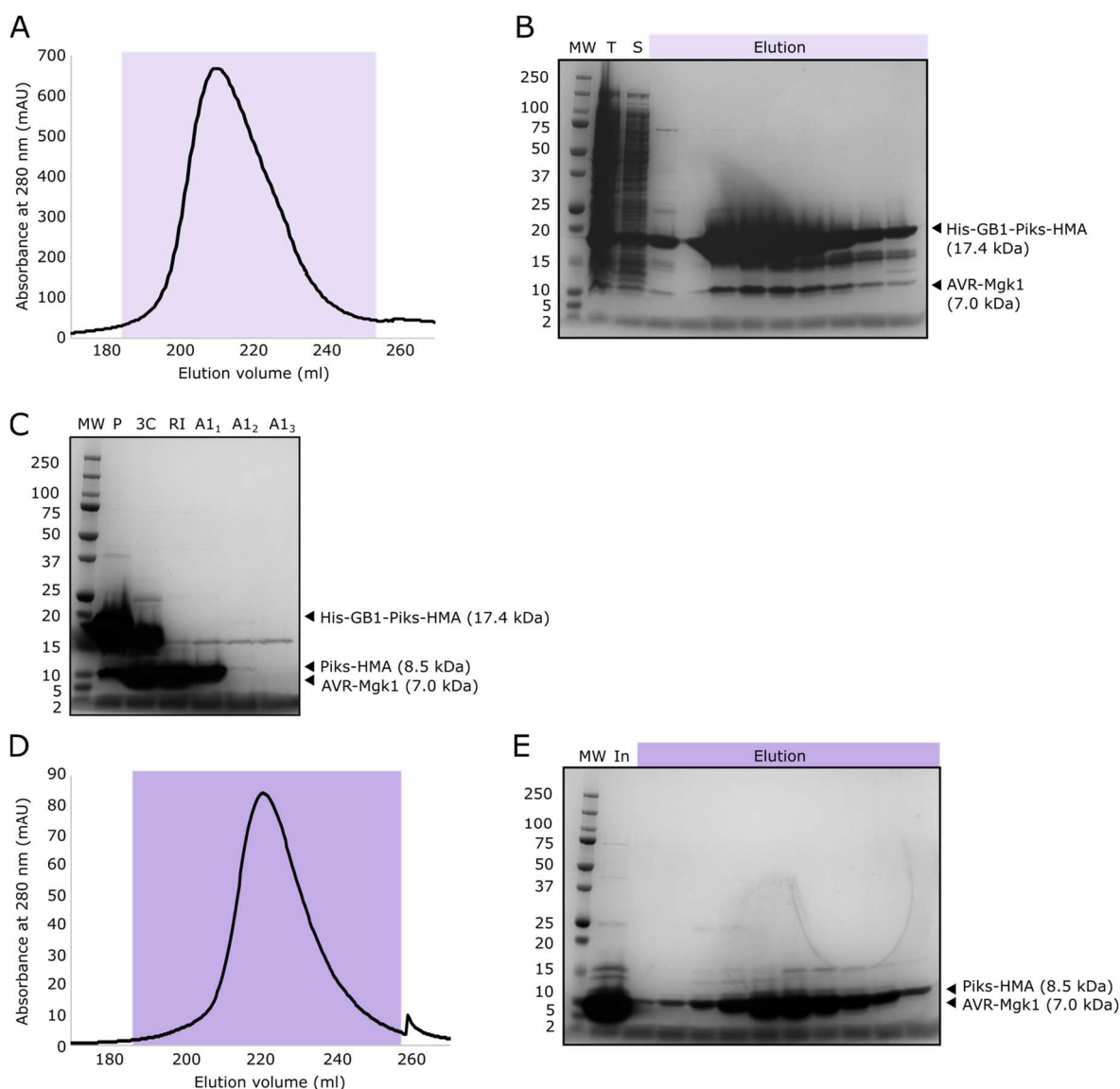


Figure 5-7: Purification of co-expressed AVR-Mgk1 and Piks-HMA complex.

A) Chromatogram of first gel filtration following tandem IMAC-Gel filtration. Fractions highlighted in light purple **B)** SDS-PAGE of T = total protein, S = soluble protein, elution fractions highlighted in light purple **C)** SDS-PAGE of tag cleavage P = pooled fractions from first gel filtration peak, 3C = 3C protease treated, RI = Reverse IMAC flow through, A1 = further column washes with buffer A1. **D)** Final gel filtration chromatogram. Region highlighted in purple corresponds to the fractions highlighted in panel D, **E)** SDS-PAGE of second gel filtration. In = concentrated input sample from RI and A1, Fractions highlighted in purple. MW = molecular weight (kDa).

Further crystallization screens were performed using purified AVR-Mgk1/Piks-HMA at 6 mg/ml and 3 mg/ml in buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl). The JIC custom screen KISS and three commercially available screens were used; Morpheus, Shotgun, and JCSG plus. Plate-like crystals with somewhat ragged edges, measuring 200-350 μ m formed

in condition B6 of the Morpheus screen within 48 hours in the 6 mg/ml condition and within five days in the 3 mg/ml condition (Figure 5-8). The crystallisation conditions were [0.09 M Halogens mix (0.3M Sodium fluoride; 0.3M Sodium bromide; 0.3M Sodium iodide), 0.1 M Buffer System 2 pH 7.5 (Sodium HEPES; MOPS (acid)), 50 % v/v Precipitant Mix 2 (40% v/v Ethylene glycol; 20 % w/v PEG 8000)].

The crystals were cooled and shipped to Diamond Light Source (Oxfordshire, UK), where X-ray diffraction data were collected at the variable microfocus beamline Io4. A total of 3600 x 0.1° images were recorded to a maximum resolution of 1.9 Å at a wavelength of 0.9795 Å (representative diffraction image Figure 5-8C). The space group was *P*1 with cell parameters $a = 35.022$ Å, $b = 35.022$ Å, $c = 61.995$ Å, $\alpha = 92.03^\circ$, $\beta = 99.67^\circ$, $\gamma = 103.79^\circ$. Data collection and refinement statistics are summarized in Table 5-1.

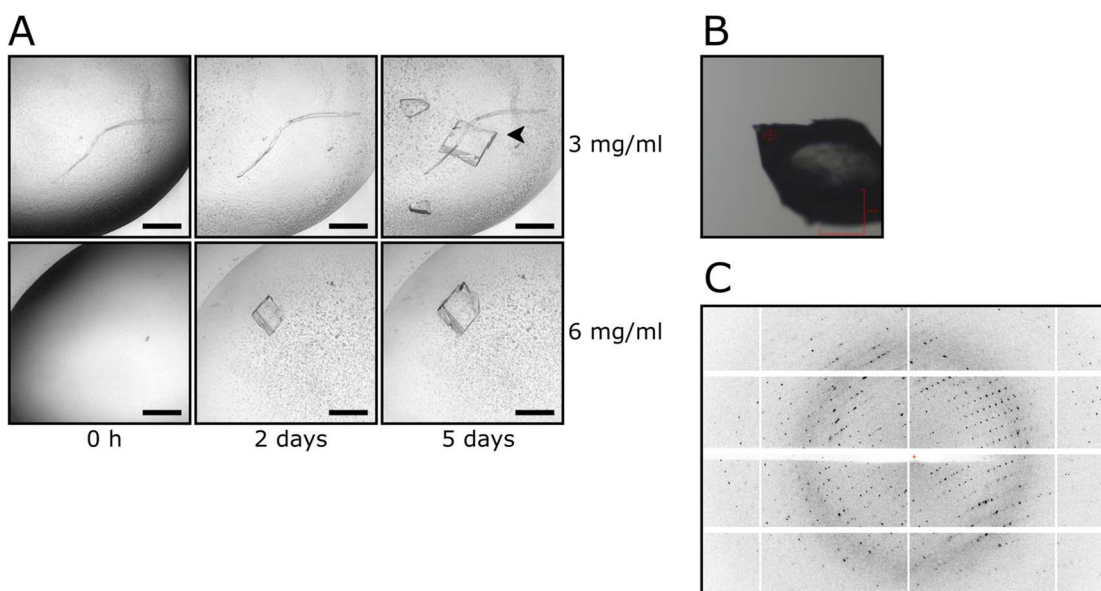


Figure 5-8: Time course of formation and diffraction image of crystal used for collecting Pika-HMA data.

A) The crystal yielding the ultimately used data set is indicated with an arrowhead. The crystallisation condition was Morpheus screen condition B6. Scale bars are 250 μ m. **B)** The crystal from which the useful dataset was obtained, pictured within the loop when mounted in the beamline. **C)** A representative diffraction image from this crystal, exhibiting a high density of clearly defined and separated diffraction spots. The image is cropped, and to improve contrast for display, five 0.1° images were summed.

Sequences of AVR-Mgk1 and Pika-HMA had been uploaded to the Diamond server ahead of the data collection session, so during the session the automatic software pipeline invoked MrBUMP to carry out molecular replacement. A molecular replacement solution was

obtained, with two copies of Piks-HMA fitted into the electron density. An additional two regions of unmodelled electron density remained, of a size that could correspond to AVR-Mgk1. After further work using BUCANEER to attempt to build AVR-Mgk1 into the vacant density, it soon became apparent that AVR-Mgk1 was not an appropriate fit and instead a further two copies of Piks-HMA were present. Ultimately, it was determined that the asymmetric unit of the crystal contained four copies of Piks-HMA, rather than two copies of the AVR-Mgk1/Piks-HMA complex. Further details of data processing can be found in (5.2.5.3 Data collection and processing).

Although not the structure that we had been seeking to determine, the Piks-HMA crystal structure (Figure 5-9 A) experimentally confirms that Piks-HMA shares its structure with Pikh-HMA apart from the two polymorphic residues and the precise position of some side chains and the loop between $\alpha 1$ and $\beta 1$ (visible at the bottom of the proteins as they are oriented in Figure 5-9). The root-mean-square deviation of atomic positions (RMSD) is 0.550 Å when calculated using 69 out of 74 residues, indicating close agreement between the two structures. The side chain density for polymorphic residue E229 allows unambiguous distinction from Q229 of Pikh-HMA (Side chains shown in stick representation in Figure 5-9 B). However, this is not possible at the C-terminus of the protein where density is weaker - therefore, the side chain position of A261 of Piks-HMA cannot be distinguished from V261 of Pikh-HMA with great confidence. However, taken together with later mass spectrometry results where peptides covering the C-terminus of Piks-HMA were detected, confirming the presence of A261 in Piks-HMA, we can conclude that the Piks-HMA was produced and that it shares the HMA fold with Pikh-HMA.

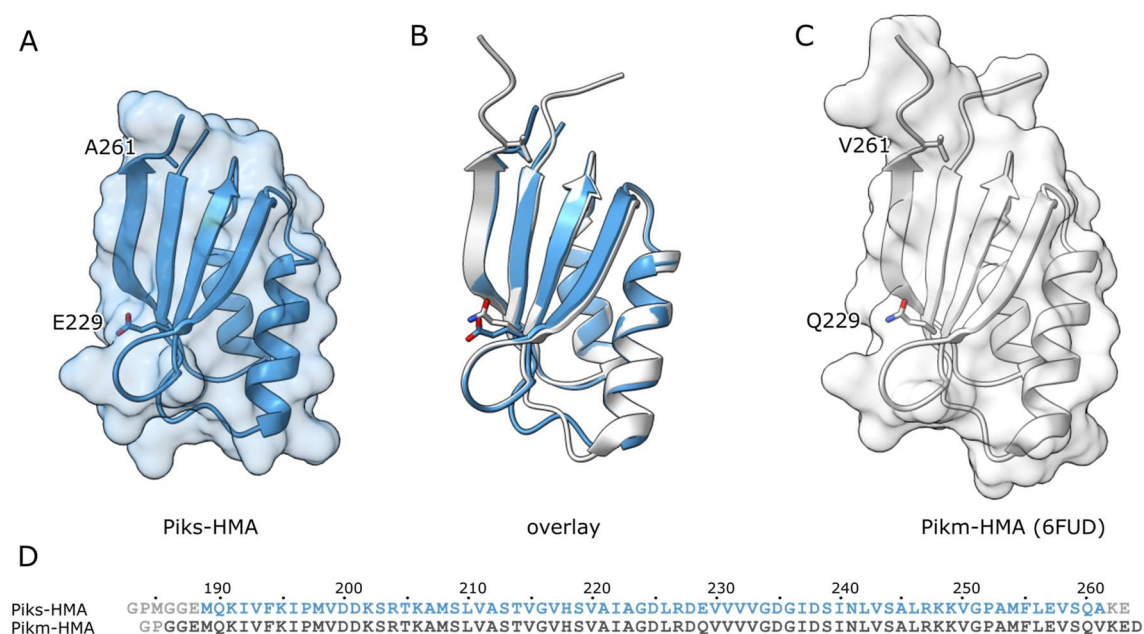


Figure 5-9: The crystal structure of Piks-HMA is in good agreement with the Pikm-HMA crystal structure.

A) Ribbon representation of Piks-HMA, with the side chains of residues E229 and A261 in stick representation. **B)** Overlay of Piks-HMA and Pikm-HMA structures. (Pikm-HMA, PDB: 6FU9) (De La Concepcion et al., 2018). **C)** Ribbon representation of Pikm-HMA, with the side chains of residues Q229 and V261 in stick representation. On side chains, oxygen atoms are depicted in red and nitrogen in dark blue. **D)** Sequence alignment of Piks-HMA and Pikm-HMA indicating residues present in each structure. Piks-HMA residues present in the structure are shown in blue, Pikm-HMA residues present in the structure are shown in dark grey, residues which were unable to be modelled are shown in light grey. Residue numbering is relative to the full length Pikm-1 and Piks-1 sequence. The initial GP residues in both sequences are the 3C protease scar left as a consequence of the purification strategy and are not part of the Pik sequences.

Table 5-1: Data collection and refinement statistics for the Pks-HMA structure

Data collection statistics	
Beamline	I04, Diamond Light Source, UK
Wavelength (Å)	0.9795
Detector	Eiger2 16M
Resolution range (Å)	34.11-1.94 (1.90-1.90)
Space group	<i>P</i> 1
Cell dimensions (<i>a</i> , <i>b</i> , <i>c</i>) (Å)	35.022, 35.230, 61.995
Cell angles (α , β , γ) (°)	92.03, 99.67, 103.79
Total no. of observations	79,589 (5,357)
Unique reflections	21,825 (1,421)
Multiplicity	3.6 (3.8)
Mean(<i>I</i> / σ <i>I</i>)	10.7 (2.3)
Completeness (%)	97.9 (96.6)
<i>R</i> _{merge}	0.048 (0.387)
<i>R</i> _{meas} .	0.056 (0.451)
CC(1/2)	0.999 (0.960)
Wilson B value (Å ²)	31.1
Refinement and model building statistics	
Resolution range (Å)	34.13-1.90
Reflections: working/free	20736 (1091)
<i>R</i> _{work} / <i>R</i> _{free}	0.211 / 0.266
Ramachandran plot favoured/allowed/outliers (%)	98.25/1.75/0.00
RMSD	
Bond length deviation (Å)	0.0164
Bond angle deviation (°)	1.73
No. atoms	
Amino acids	2190
Ions	2
Waters	67
B-factors	
Amino acids	23.13
Ions	36.18
Waters	39.59
MolProbity Score	0.75

Values in parentheses are for the outer resolution shell. RMSD = root mean square deviation

Lack of AVR-Mgk1 in the crystal structure could be attributed to several factors. It might be possible that in the case of a weak interaction between AVR-Mgk1 and Piks-HMA, Piks-HMA may make contacts with other Piks-HMA molecules and form a crystal to the exclusion of AVR-Mgk1. Although this might be considered an unlikely explanation in light of the interaction of the complex in the earlier analytical gel filtration experiment and particularly in light of the later surface plasmon resonance data which indicates a 56 nM affinity between AVR-Mgk1 and Piks-HMA, it is possible that the high ionic strength of the crystallisation solution disrupted the complex. It is also possible that AVR-Mgk1 degraded in the crystallisation screen or was lost from the protein purification at a stage after the cleavage of the tag, as from this point AVR-Mgk1 cannot be distinguished from Piks-HMA by SDS-PAGE.

To determine if both proteins were present in the purified AVR-Mgk1/Piks-HMA complex and to provide an indication as to their relative abundance, the protein sample used when setting up crystallisation screens was subjected to SDS-PAGE, and the Coomassie blue stained band was cut and submitted to the TSL Proteomics platform. Electrospray Ionisation Mass Spectrometry was performed by Dr Jan Sklenar (5.2.2.2 Liquid chromatography-mass spectrometry (LC-MS)). Piks-HMA and AVR-Mgk1 were both readily detected with high confidence, with multiple fragments covering the entire sequence from N- to C-terminus, and multiple incidences of each fragment, making up the top two protein hits disregarding keratin contaminants. However, peptides attributable to AVR-Mgk1 were detected at a much lower incidence compared to Piks-HMA peptides. In light of the crystal structure of Piks-HMA alone and not in complex with AVR-Mgk1, it is likely that AVR-Mgk1 is underrepresented in the protein sample. Since AVR-Mgk1 is clearly visible on SDS-PAGE at the beginning of the purification, it is possible that the protein was largely lost during the purification process. As AVR-Mgk1 and Piks-HMA cannot be separated on an SDS-PAGE gel after the removal of affinity tags, the loss of AVR-Mgk1 from the sample could occur at any stage after tag cleavage and remain undetected by SDS-PAGE.

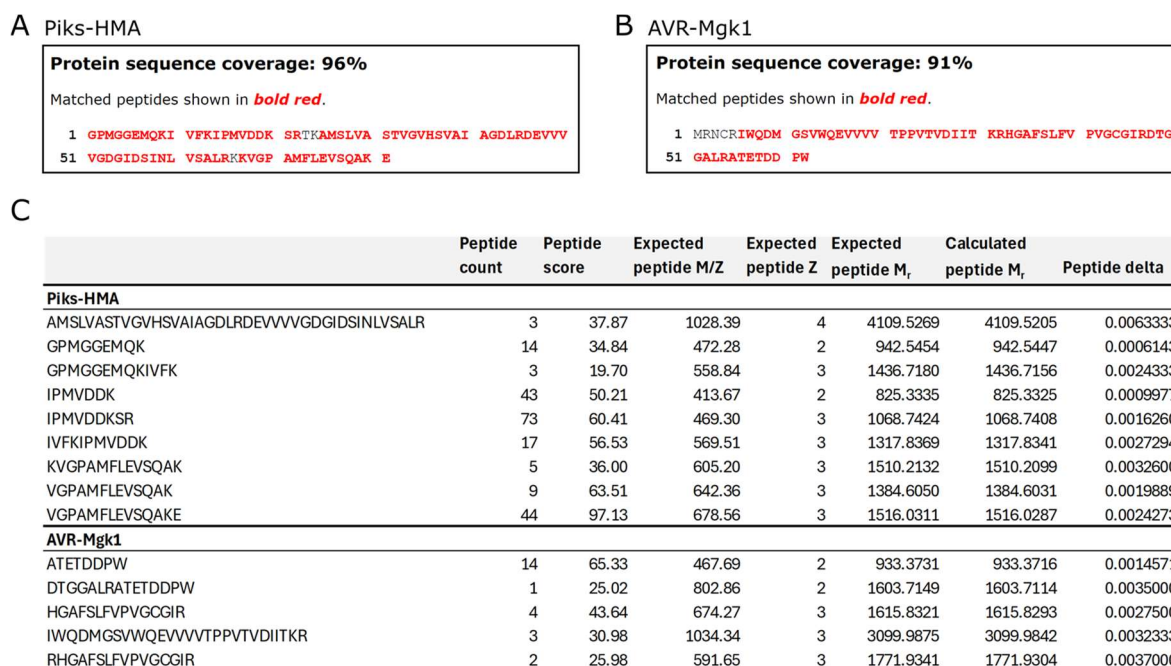


Figure 5-10: Mass spectrometry indicates that AVR-Mgk1 may be underrepresented in the protein sample.

ESI-TRAP MS performed by Jan Sklenar. **A)** Protein sequence coverage of Piks-HMA - detected peptides are displayed in red. **B)** Protein sequence coverage of AVR-Mgk1 – detected peptides are displayed in red. **C)** Table displaying peptides detected from Piks-HMA and AVR-Mgk1. Peptide count = number of instances of that peptide detected, Peptide score = significance score derived from peptide prevalence and likelihood above random noise, Expected peptide Z = expected peptide charge, Expected peptide M_r = expected peptide relative molecular mass, Calculated peptide M_r = calculated peptide relative molecular mass, peptide delta = difference between expected and observed peptide relative molecular mass.

5.3.4 Further attempts at crystallisation of the Piks-HMA and AVR-Mgk1 complex

For further crystallisation attempts, I prioritised optimising the purification method to ensure that a robust complex containing both AVR-Mgk1 and Piks-HMA was produced and present in the final protein sample. To improve the likelihood that the AVR-Mgk1/Piks-HMA complex forms and persists through the purification, each protein was individually expressed and purified to the point of tag removal. After tag removal the proteins were combined and concentrated together before a final gel filtration step (Figure 5-11). This is the same method as was used by Dr Josephine Maidment to purify OsHIPP19-HMA and AVR-Pia (Doctoral thesis of Dr Josephine Maidment).

As previously discussed, Piks-HMA lacks aromatic amino acids and so it cannot be detected by UV absorption at 280 nm. Therefore, its concentration must be calculated by monitoring absorbance of the peptide bond at 205 nm, or by more time-consuming methods such as

Bradford assay or Direct Detect. The A205 programme of the Nanodrop One was found to be unreliable and Bradford Assay or Direct Detect is inconvenient to carry out during the protein purification, so to ensure the maximum possible yield of complex, the entirety of purified AVR-Mgk1 was added to the purified Piks-HMA. This often resulted in an excess of AVR-Mgk1. The ability to easily and entirely separate excess AVR-Mgk1 from the AVR-Mgk1/Piks-HMA complex by gel filtration is owed to the superior resolution provided by a new Superdex 75 16/60 gel filtration column.

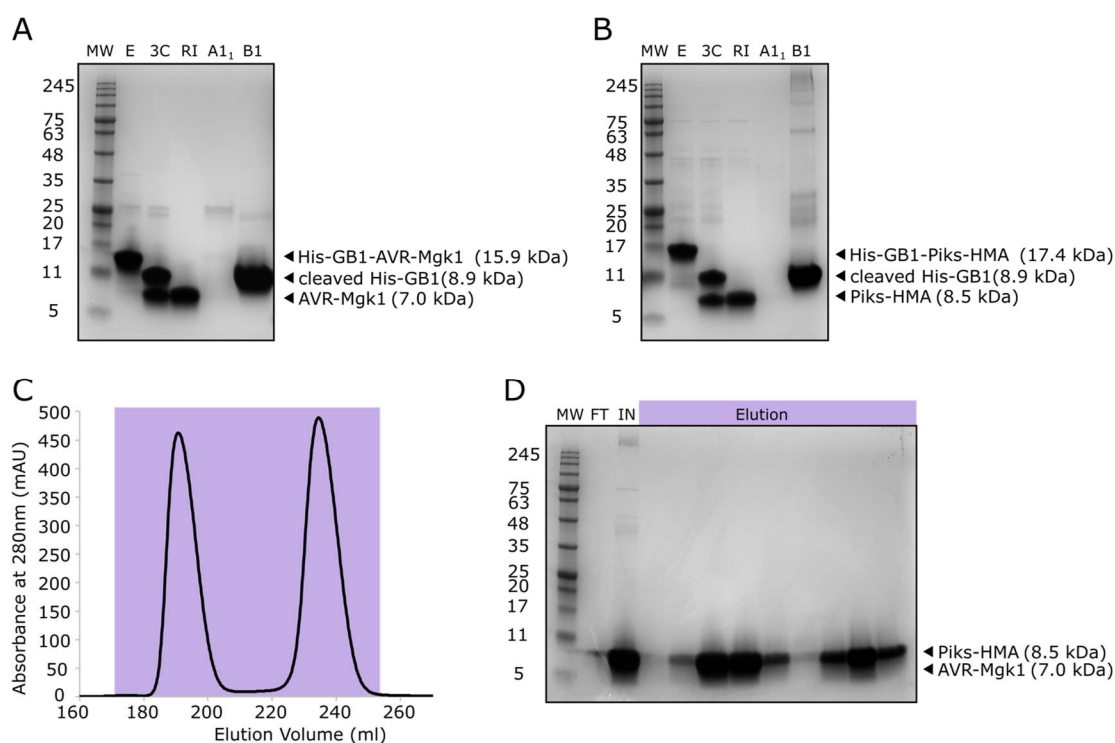


Figure 5-11: Purification of AVR-Mgk1/Piks-HMA complex by mixing.

A) SDS-PAGE of 3C tag cleavage and reverse IMAC of AVR-Mgk1. **B)** SDS-PAGE of 3C tag cleavage and reverse IMAC of Piks-HMA. **C)** Chromatogram of final gel filtration; region highlighted in light purple corresponds to the elution fractions highlighted in panel D). MW = molecular weight, E = pooled elution from IMAC-gel filtration, 3C = 3C protease treated protein, RI = flowthrough from reverse IMAC, A1-3 = washes with buffer A1, B1 = wash with buffer B1. FT = flow-through from centrifugal concentrator before final gel filtration. IN = gel filtration input. **D)** SDS-PAGE of final gel filtration of complex formed between AVR-Mgk1 and Piks-HMA.

The purification of the AVR-Mgk1/Piks-HMA complex was confirmed by intact mass spectrometry (Figure 5-12) (5.2.2 Mass spectrometry). From the final gel filtration that separates the AVR-Mgk1/Piks-HMA complex from AVR-Mgk1, in the first peak, masses of 7038.10 Da and 8526.30 Da were observed (Figure 5-11 D). These correspond very well to

the predicted peptide masses of AVR Mgk1 (7040.08 Da) and Piks-HMA (8526.00 Da.) The 2.0 Da difference in mass between the predicted and observed mass for AVR-Mgk1 can be explained by the loss of two H atoms during the formation of the expected disulfide bond. In the second peak, a single mass of 7038.00 was observed, which again corresponds exactly to AVR-Mgk1 (predicted peptide mass 7040.08 Da) once disulfide bond formation is accounted for.

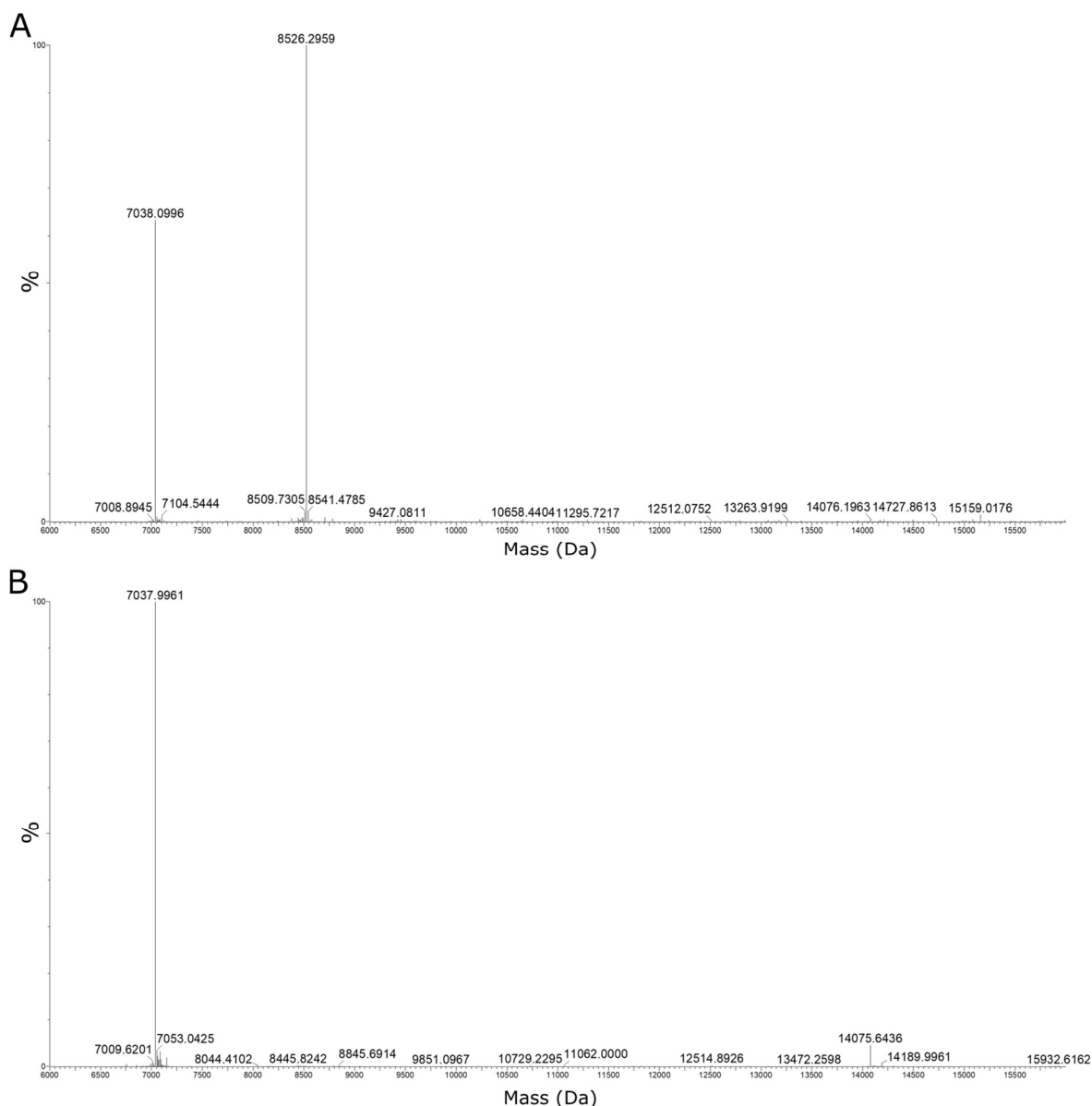


Figure 5-12: Intact mass spectra from AVR-Mgk1/Piks-HMA complex purification.

LC-MS carried out by Dr Carlo Martins. **A)** Masses of 7038.00 Da and 8526.30 Da observed in the first gel filtration peak (Figure 5-11 D), corresponding to AVR-Mgk1 (predicted peptide mass 7040.08 Da) and Piks-HMA (predicted peptide mass 8526.00 Da). **B)** Mass of 7038.00 Da detected in the second gel filtration peak (Figure 5-11 D), corresponding to AVR-Mgk1 alone (predicted peptide mass 7040.08 Da).

The purified complexes were then used to set up further crystallisation screens. AVR-Mgk1/Piks-HMA at concentrations of 27 mg/ml and 13.5 mg/ml was dispensed into the Morpheus, SG1, and JCSG screens initially, later followed by KISS, MIDAS, and ProPlex screens. Although a good mixture of precipitated and clear drops was observed, suggesting that the range of protein concentrations used were appropriate, no crystals were observed over a number of weeks. Finally, Structure, PEGS, BCS, and PACT screens were set up, again with AVR-Mgk1/Piks-HMA at 27 mg/ml and 13.5 mg/ml. Some needle-like crystals were observed in Structure screen, well A6-1 (0.1 M Sodium acetate pH 4.6; 8 % w/v PEG 4000). These could be used for further optimisation, or if their growth was able to be reproduced in the appropriate flat-bottomed plate, could even be suitable for in-situ data collection at beamline VMXi at Diamond Light Source. However, these possibilities were ultimately not pursued due to simultaneous progress in the work presented in the later section (5.3.7).

5.3.5 Expression and purification of OsHPPs

Concurrently with the AVR-Mgk1/Piks-HMA work presented in the previous section, Dr Yu Sugihara was investigating the possibility of interactions between AVR-Mgk1 and small HMAs. He performed a yeast two hybrid screen to assess AVR-Mgk1 binding against a library of sHMAs including a range of HPPs and HPP proteins (Figure 5-13). The sHMA library was produced for a study into interactors of AVR-PikD, and was prepared from cDNA prepared from rice cultivar Sasanishiki inoculated with *M. oryzae* strain Sasa2 (Dr. Yu Sugihara, personal communication) (Oikawa et al., 2024).

The screen revealed OsHPPo2, OsHPPo3, and OsHPPo4 as interactors of AVR-Mgk1. These three proteins are sHMA proteins which share a predicted HMA domain with integrated HMAs but have the distinguishing feature of an additional long C-terminal extension which is predicted to be disordered. Comparison of the HMA domain sequences only reveals a 59.7 – 61 % sequence identity between the OsHPP-HMAs and Piks-HMA. I ordered the full length OsHPPo2, OsHPPo3, and OsHPPo4 genes including domestication for golden gate cloning (2.1.2 Gene synthesis) and cloned them for expression and purification in *E. coli* as alternative co-crystallisation partners for AVR-Mgk1.

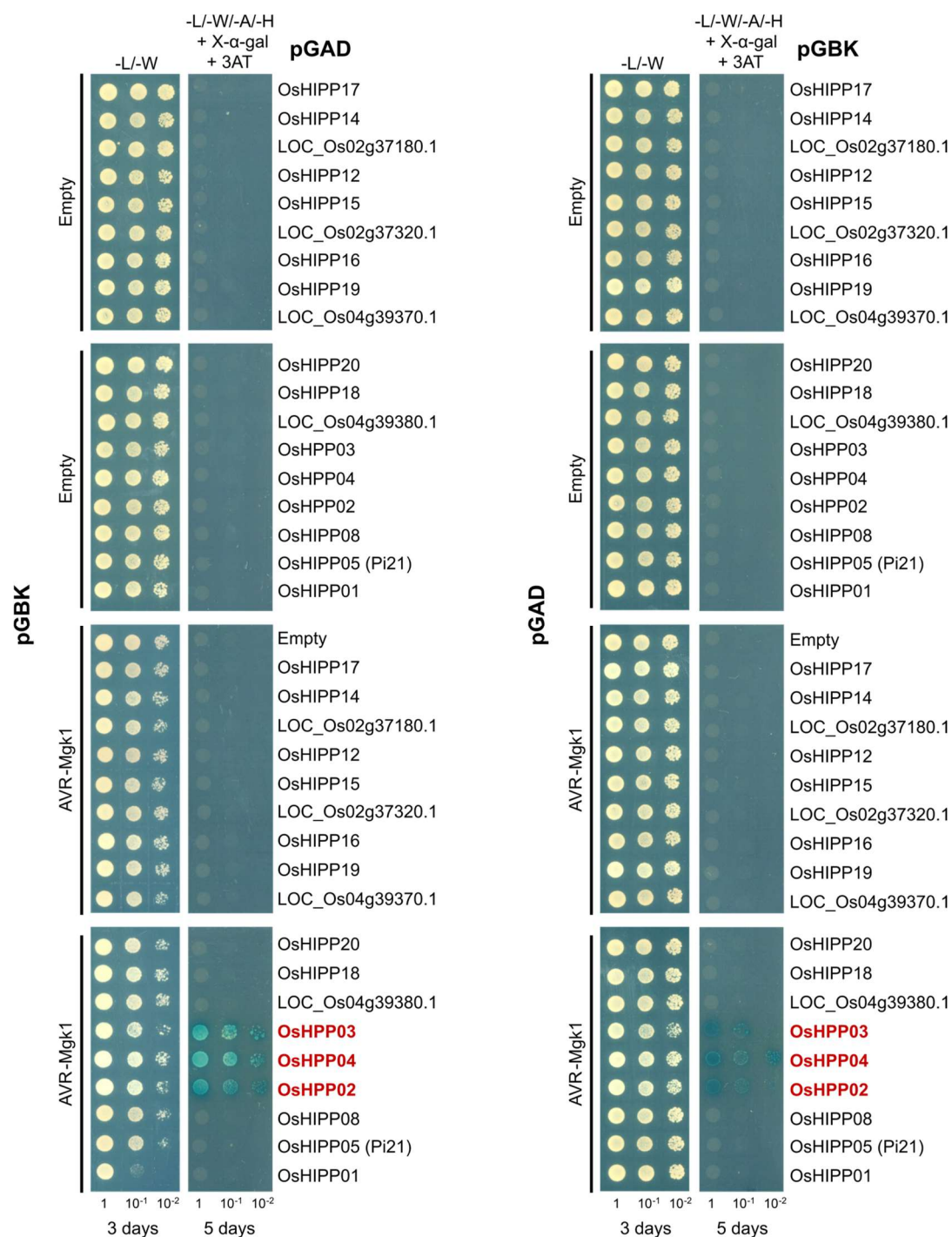


Figure 5-13: Yeast two hybrid assay between AVR-Mgk1 and small HMA library.

Yeast two hybrid assay performed by Dr. Yu Sugihara, reproduced here with permission. The assay was carried out with AVR-Mgk1 in both the bait and the prey position of the assay, and binding between AVR-Mgk1 and OsHPP02, OsHPP03, OsHPP04 was revealed in both cases. -L/-W = basal growth media lacking leucine (L) and tryptophan (W) to control the growth rate; -L/-W/-A/-H = basal growth media lacking leucine (L), tryptophan (W), adenine (A) and histidine (H), and containing X-α-gal for selection and 3AT for increased stringency of selection.

A sequence alignment was made between expression constructs of Pikm-HMA and OsHIPP19-HMA previously used for crystallography, and the full length OsHPP02, OsHPP03, and OsHPP04 sequences (Figure 5-14 A). The secondary structure elements of Pikm-HMA (derived from the crystal structure of Pikm-HMA/AVR-PikD; PDB: 6FUD) are displayed along the bottom of the alignment.

Clear domain boundaries were apparent and so truncations of OsHPP02, OsHPP03, and OsHPP04 to residue K77 were designed. As an additional feasibility check, an AlphaFold2 structural prediction of each of the full length OsHPP sequences was made (Figure 5-14 B) (5.2.4 Protein structure prediction). Each of the OsHPPs was predicted to form a core HMA fold (high confidence) with a long C-terminal tail (low confidence, typically indicative of disordered or flexible residues). The truncation to residue K77 on each of the OsHPPs did not interfere with the final beta sheet of the HMAs in the predicted structures, giving further support to the proposed truncations.

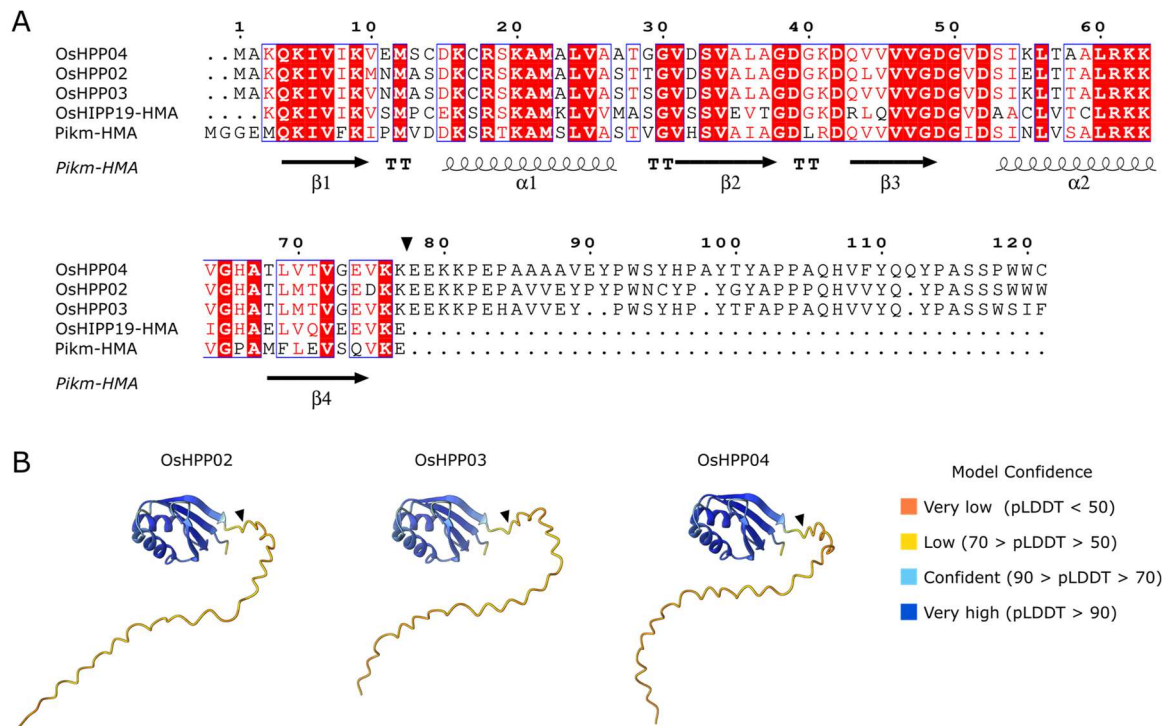


Figure 5-14: Sequence alignment and structural prediction of full length OsHPP02, OsHPP03, OsHPP04.

A) Sequence alignment of OsHPP02, OsHPP03, OsHPP04, Pikm-HMA and OsHIPP19-HMA. Residue numbering is with respect to OsHPP04. Secondary structure elements of Pikm-HMA are displayed below the alignment (PDB: 6FUD, (De La Concepcion et al., 2018)). Alignment performed using Clustal Omega and visualised using ESPrpt 3.0. Black arrowhead indicates the position of the K77 truncation. **B)** AlphaFold2 predictions of full length OsHPP02, OsHPP03, and OsHPP04. Models are coloured according to pLDDT score (blue = high confidence, yellow and red = low confidence) PAE, MSA, and pLDDT, plots are available as an Appendix. Black arrowhead indicates the position of the K77 truncation. TT = turn-turn.

The three truncated OsHPPs were each cloned with a His-MBP and a His-GB1 N-terminal tag for expression screening (Figure 5-15). Overexpression bands were visible for every construct when expressed in arabinose inducible *E. coli* but were fainter or absent when expressed in SHuffle *E. coli*.

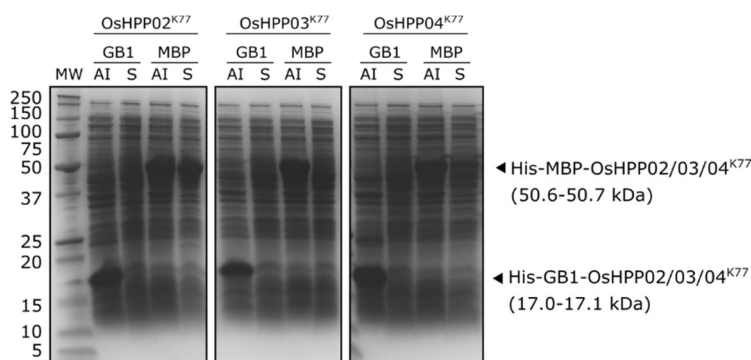


Figure 5-15: *E. coli* expression testing of OsHPP02^{K77}, OsHPP03^{K77}, and OsHPP04^{K77}

SDS-PAGE of clarified cell lysates of OsHPP02^{K77}, OsHPP03^{K77}, and OsHPP04^{K77} constructs with His-GB1, or His-MBP N-terminal tags expressed in both Arabinose Inducible (AI) and SHuffle (S) *E. coli* expression strains.

I first opted to express and purify His-GB1-OsHPP04^{K77} from arabinose inducible *E. coli* due to the high-density overexpression band in the expression test. The His-GB1 tagged construct was chosen over the His-MBP tagged construct despite the approximately equal band intensities because the His-GB1 tag is of a lower molecular weight than the His-MBP tag and so a larger proportion of the His-GB1-OsHPP04^{K77} overexpression band should be composed of OsHPP04^{K77}. Expression at 8L scale was successful and IMAC-gel filtration worked well, however the cleaved affinity tag was not fully removed by reverse-IMAC and due to the similar size between OsHPP04^{K77} (8.0 kDa) and the cleaved His-GB1 tag (8.9 kDa) these two species could not be separated by the final gel filtration step (Figure 5-16).

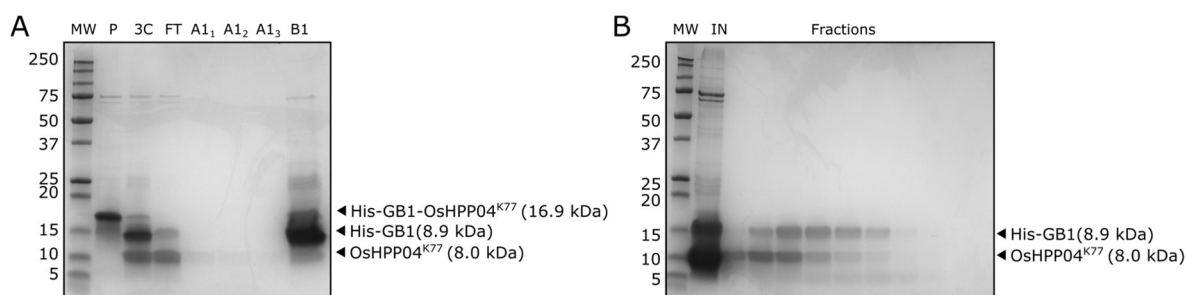


Figure 5-16: SDS-PAGE from purification of His-GB1-OsHPP04^{K77}.

A) P = pooled and concentrated peak from IMAC-gel filtration, 3C = 3C protease treated, FT = flowthrough from reverse-IMAC, A1-3 = elution with buffer A1, B1 = elution with buffer B1. **B)** Final gel filtration step; IN = concentrated input sample; Fractions = eluted fractions.

To overcome the difficulties of separating the similarly sized His-GB1 tag and OsHPPo4^{K77}, purification of His-MBP-OsHPPo2^{K77} was carried out (Figure 5-17). The molecular weight difference between cleaved His-MBP tag and OsHPPo2^{K77} should allow total separation of the two species by gel filtration. Choice of a different OsHPP may also help to overcome any protein interaction issues intrinsic to OsHPPo4. However, in the first purification an MBP column was not used during the reverse IMAC stage, and so some cleaved His-MBP tag persisted through the second gel filtration step and co-eluted with the OsHPPo2^{K77} (Figure 5-17 C). A final application of the purified protein through a MBP column largely removed the His-MBP contamination (Figure 5-17 D).

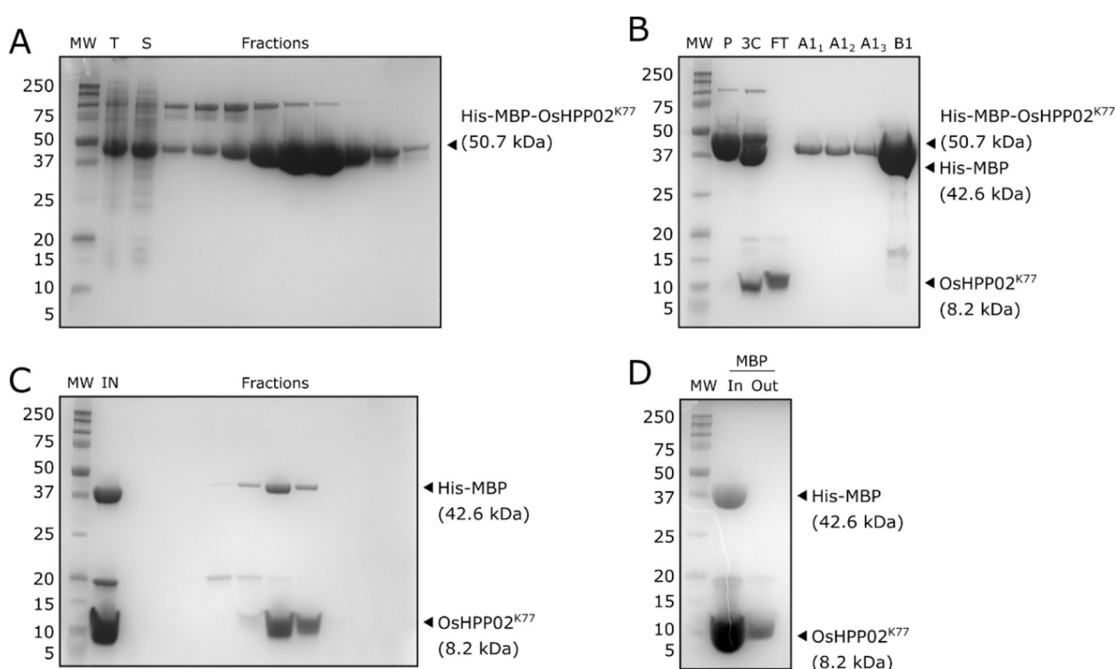


Figure 5-17: Purification of OsHPP02^{K77}.

SDS-PAGE of all stages of purification. Chromatograms not displayed due to lack of absorbance at 280 nm.

A) IMAC-gel filtration. T = total lysate, S = soluble lysate. **B)** P = pooled and concentrated peak from IMAC-gel filtration, 3C = 3C protease treated, FT = flowthrough from reverse-IMAC, A1₁₋₃ = elution with buffer A1, B1 = elution with buffer B1. **C)** Final gel filtration step; IN = concentrated input sample; Fractions = eluted fractions. **D)** MBP column: In = input, Out = output.

5.3.6 Direct interaction of AVR-Mgk1 and OsHPP02^{K77} by analytical gel filtration

With purified OsHPP02^{K77} in hand, the interaction between AVR-Mgk1 and OsHPP02, previously observed in a yeast-two-hybrid experiment, could then be investigated in vitro through an analytical gel filtration experiment. As described previously (5.2.3.1 Analytical Size Exclusion Chromatography (SEC)) the proteins were mixed in equimolar ratio before gel filtration on a Superdex 75 10/30 column. Elution volume of the proteins was observed via absorbance at 280 nm. Alone, AVR-Mgk1 was observed as a monodisperse peak at 15.4 ml. The peak elution volume of OsHPP02^{K77} cannot be accurately determined by absorbance at 280 nm due to lack of aromatic amino acids, although a very small peak was observed at 13.6 ml which could correspond to OsHPP02^{K77}.

After incubation of AVR-Mgk1 and OsHPP02^{K77} together, a single monodisperse peak was observed at 13.6 ml. The shift in elution volume can be attributed to the incorporation of AVR-Mgk1 into a larger protein complex, with a correspondingly lower elution volume than AVR-Mgk1 alone. This experiment confirmed direct interaction between AVR-Mgk1 and OsHPP02^{K77} in vitro, foundational in the pursuit of a crystal structure of AVR-Mgk1 in complex with a HMA domain protein.

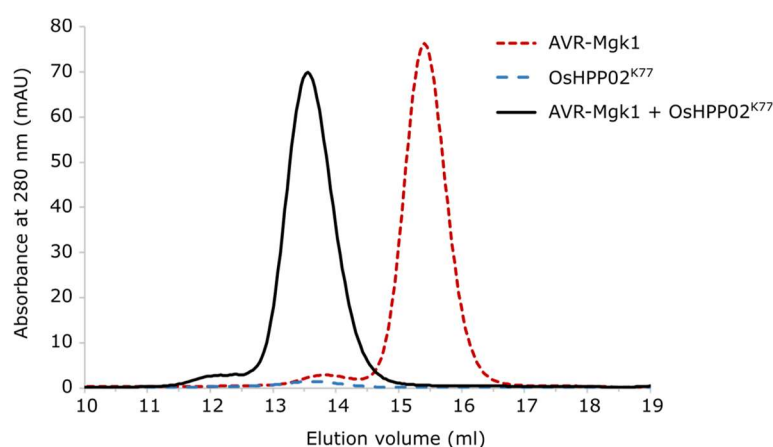


Figure 5-18: Analytical gel filtration of AVR-Mgk1 and OsHPP02^{K77}.

Red dashed line = AVR-Mgk1, blue dashed line = OsHPP02^{K77}, black line = AVR-Mgk1/OsHPP02^{K77}.

5.3.7 Expression and purification of the AVR-Mgk1/OsHPP02^{K77} complex for crystallisation

The AVR-Mgk1/OsHPP02^{K77} complex was expressed and purified using the same strategy as for the AVR-Mgk1/Piks-HMA complex (2.2.4 Protein purification from *E. coli*). To ensure presence of both proteins in the final purified complex, each protein was expressed and purified individually up to the removal of affinity tags, and the proteins were mixed and concentrated together before a final gel filtration step (Figure 5-19).

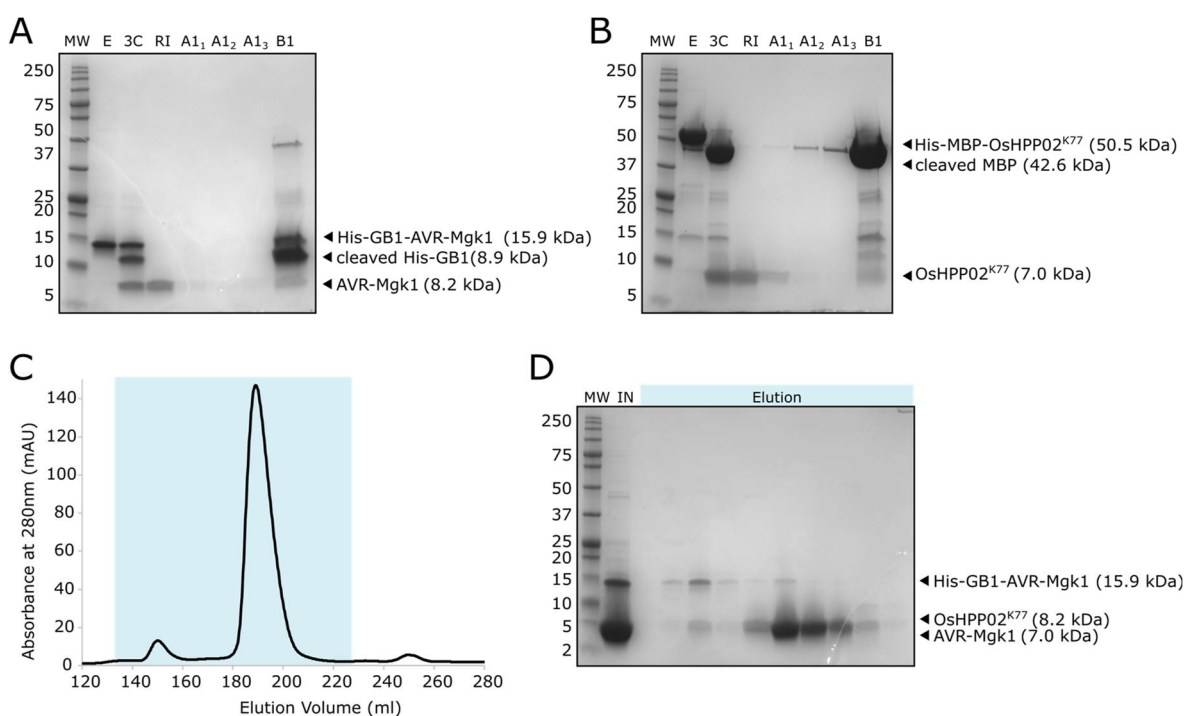


Figure 5-19: Purification of AVR-Mgk1/OsHPP02^{K77} complex.

A) SDS-PAGE of GB1-AVR-Mgk1 purification; **B)** SDS-PAGE of His-MBP-OsHPP02^{K77} purification; **C)** Chromatogram of final gel filtration of the AVR-Mgk1/OsHPP02^{K77} complex. **D)** SDS-PAGE of final gel filtration after mixing and concentrating the individually purified AVR-Mgk1 and OsHPP02^{K77}, Elution corresponds to the region of the chromatogram highlighted blue in C). MW = molecular weight, E = pooled and concentrated elution from IMAC-gel filtration peak, 3C = 3C protease treated protein, RI = reverse IMAC flow through, A1-3 = sequential elutions with buffer A1, B1 = elution with buffer B1. IN = concentrated input sample for final gel filtration.

The identity and composition of the purified AVR-Mgk1/OsHPP02^{K77} complex was confirmed by intact mass spectrometry, performed by Dr Carlo Martins (Figure 5-20). The experimentally determined masses (7038.19 Da and 8188.14 Da) correspond reasonably

well with the expected monoisotopic mass of 7035.52 Da for AVR-Mgk1 and 8182.31 Da for OsHPP02^{K77}, confirming presence of these two proteins.

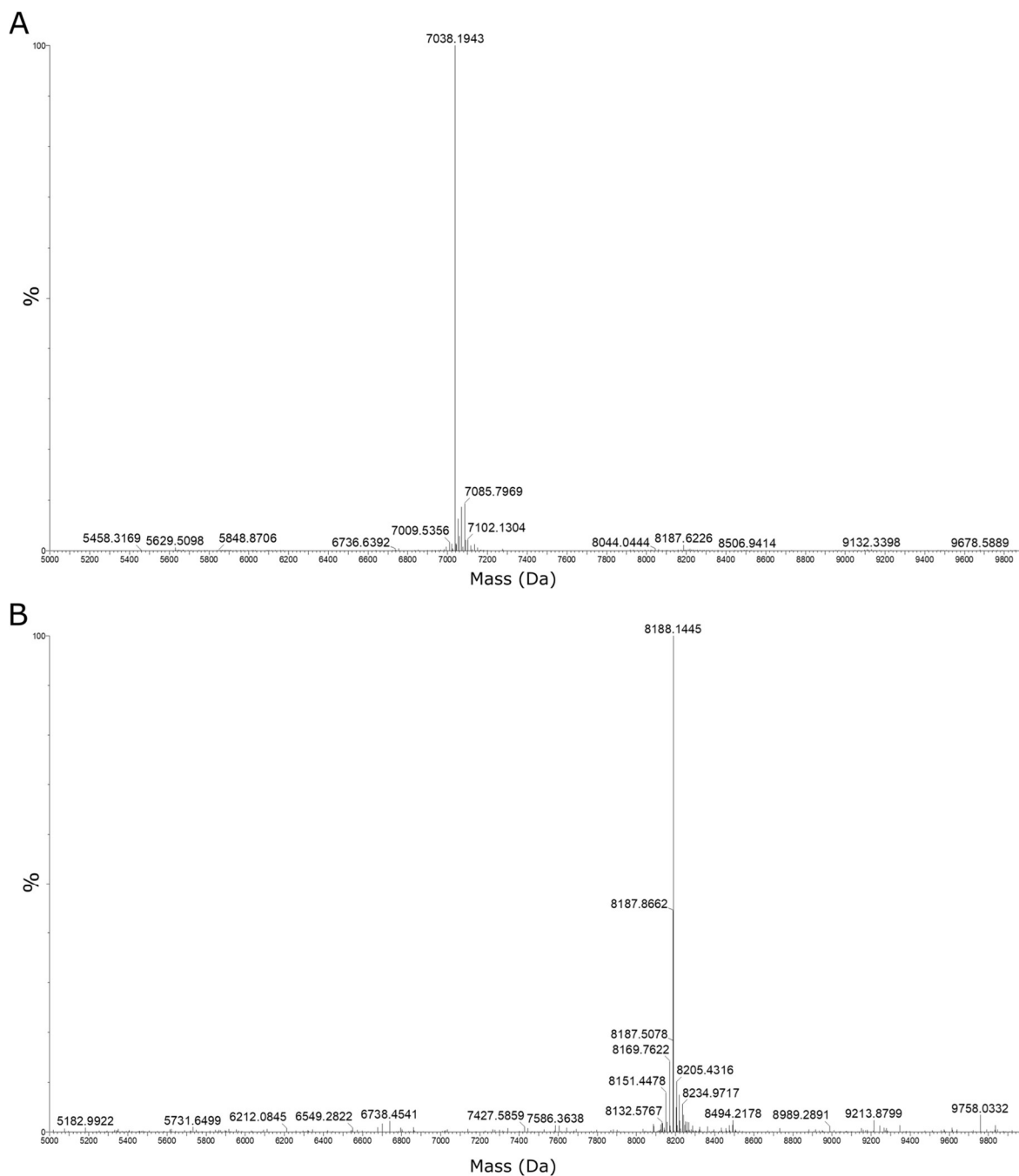


Figure 5-20: Intact mass spectra from AVR-Mgk1/OsHPP02^{K77} complex purification.

A) Mass of 7038.19 Da, corresponding to AVR-Mgk1 (predicted peptide mass 7035.52 Da). **B)** Mass of 8188.14 Da, corresponding to OsHPP02^{K77} (predicted peptide mass 8182.31 Da).

A variety of initial sparse matrix crystallisation screens (PEGs, ShotGun 1, and Morpheus screens) were set up with the AVR-Mgk1/OsHPPo2^{K77} complex at two concentrations of 10 mg/ml and 20 mg/ml in buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl). Observation over the first week after setting up revealed a majority of drops contained precipitated protein and so further six crystallisation screens (PEGs, Shotgun 1, Morpheus, KISS, ProPlex, JCSG) were set up at lower concentrations of 5 mg/ml and 10 mg/ml.

Observation of these plates at a three-week timepoint revealed a variety of granular precipitates, and semi-crystalline “walnut” bi-lobed rounded spherulites with uneven and crinkly surfaces. Some of the precipitates from the PEGs screen conditions B5 and B6, and ProPlex condition E9 were taken and used for seeding into the Morpheus screen, although without success.

Spherulites can form when crystal lattice formation is interrupted by incorporation of proteolytic fragments which are amorphous and prevent the crystal lattice from growing. If a dominant species of proteolytic fragment is causing the spherulite formation, truncating this fragment from the protein could alleviate the formation of spherulites. Therefore, limited trypsin proteolytic digest and SDS-PAGE was performed to investigate if there were any dominant proteolytic products of the AVR-Mgk1/OsHPPo2^{K77} complex (Figure 5-21). Trypsin hydrolyses peptide bonds at the C-terminal side of lysine and arginine residues. In the case of a proteolytic product running as a single band on SDS-PAGE, then it could then be excised from the gel and identified by mass spectrometry.

Dr Rafal Zdrzalek successfully used this approach in the crystallisation of the PWL2/OsHIPP43 complex. He identified a 10-residue truncation at the C-terminus of PWL2 which facilitated subsequent crystallisation and structure determination (Zdrzalek et al., 2024). Contrastingly, in the case of AVR-Mgk1/OsHPPo2^{K77} complex, a smear of proteolytic products of different sizes was observed on SDS-PAGE. This leads to the conclusion that trypsin digestion does not reveal specific truncations to use in AVR-Mgk1/OsHPPo2^{K77} construct optimisation (Figure 5-21). However, use of an alternative protease with a different amino acid specificity could reveal a different candidate truncation. For example, α -chymotrypsin selectively catalyses the hydrolysis of the peptide bond on the C-terminal side of leucine, and aromatic amino acids phenylalanine, tryptophan, and tyrosine.

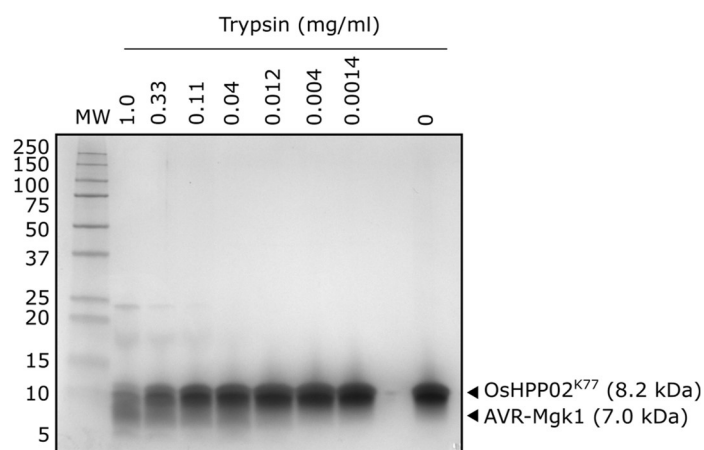


Figure 5-21: SDS-PAGE of limited trypsin proteolysis of the AVR-Mgk1/OsHPP02^{K77} complex.

As an alternative to further limited proteolysis experiments, I took a structure prediction approach to guide my choice of truncations of the AVR-Mgk1/OsHPP02^{K77} complex. I used AlphaFold2 multimer (Evans et al., 2022; Mirdita et al., 2022) to predict the structure of the AVR-Mgk1/OsHPP02^{K77} complex (5.2.4, PAE, MSA, and pLDDT plots are available as an Appendix). AVR-Mgk1 and OsHPP02^{K77} were each predicted with high confidence (denoted by blue colouring) in the core of the predicted model as a MAX effector fold and HMA fold, respectively (Figure 5-22). Some residues at the N and C termini of each protein were modelled with lower confidence scores (denoted by orange and yellow colouring), indicating potential flexibility, and therefore highlighting them as candidates for truncation (Figure 5-22). The glycine-proline scars left by 3C protease at the N-terminus of both AVR-Mgk1 and OsHPP02^{K77} were a consequence of the purification strategy and so could not be avoided. However, at the C-terminus the following truncations were made, based on the predicted structures. Four residues (DDPW) were truncated from AVR-Mgk1 to produce AVR-Mgk1^{T81} and three residues (DKK) were truncated from OsHPP02^{K77} to produce OsHPP02^{E74}. The residues to be truncated are indicated in Figure 5-22 with a dashed line and scissors icon.

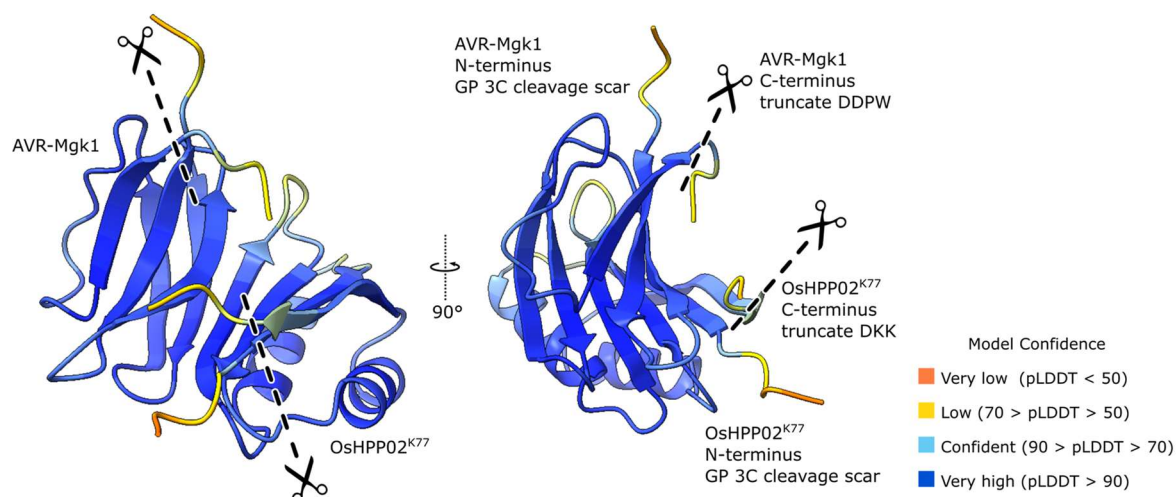


Figure 5-22: AlphaFold2 multimer predictions of AVR-Mgk1/OsHPP02^{K77}.

The position of the N-terminal truncations to produce AVR-Mgk1^{T81} and OsHPP02^{E74} are indicated with the scissor and dotted line symbols. Predicted models are coloured according to pLDDT score (blue = high confidence, yellow and red = low confidence (pLDDT score, a measure of confidence in C α position)). PAE, MSA, and pLDDT, plots available as an Appendix.

5.3.8 Purification and crystallisation of AVR-Mgk1^{T81}/OsHPP02^{E74}

The truncated AVR-Mgk1^{T81}/OsHPP02^{E74} complex was expressed and purified by the same method as the AVR-Mgk1/OsHPP02^{K77} complex (Figure 5-23) (2.2.4 Protein purification from *E. coli*). At the point of mixing the two individually purified proteins, AVR-Mgk1^{T81} was added in excess to avoid limiting the final yield of complex. The excess AVR-Mgk1^{T81} was separated from the AVR-Mgk1^{T81}/OsHPP02^{E74} complex by gel filtration (Figure 5-23 C,D).

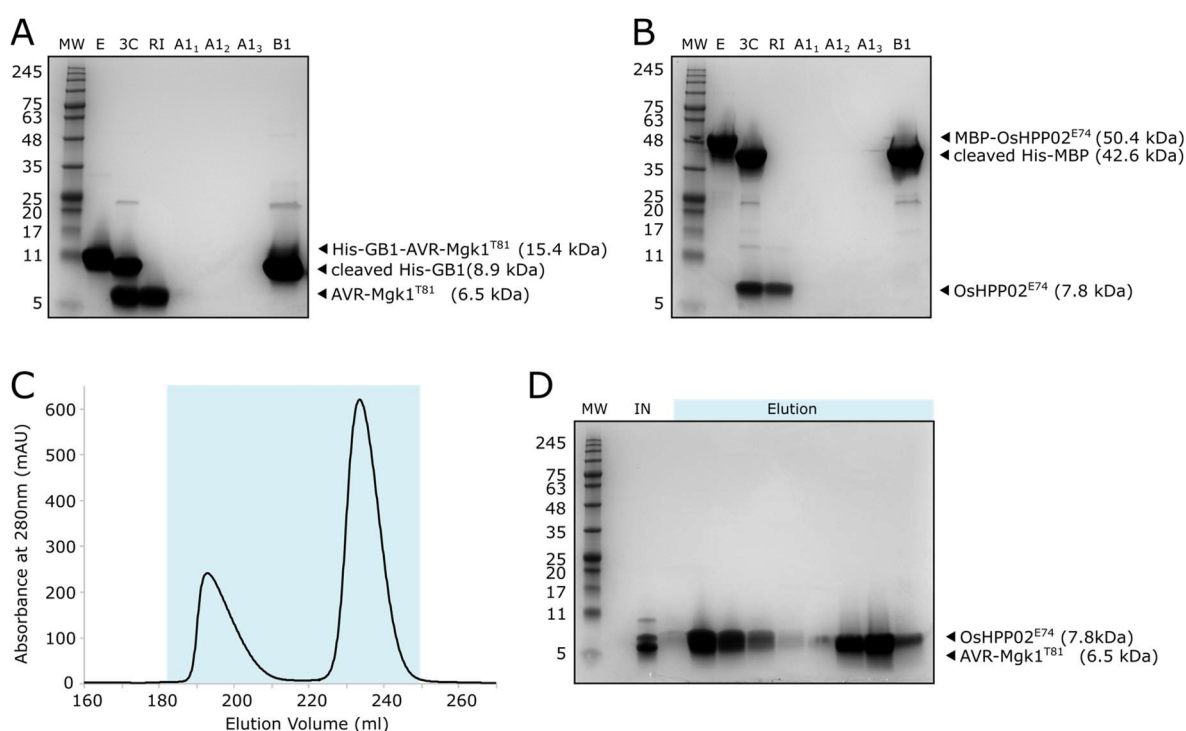


Figure 5-23: Purification of AVR-Mgk1^{T81}/OsHPP02^{E74} complex.

A) SDS-PAGE of 3C tag cleavage and reverse IMAC of AVR-Mgk1^{T81}, **B)** SDS-PAGE of 3C tag cleavage and reverse tandem MBP trap and IMAC of OsHPP02^{E74}, **C)** Chromatogram of final gel filtration; region highlighted in pale blue corresponds to the elution fractions highlighted in panel D). **D)** SDS-PAGE of final gel filtration of complex formed between AVR-Mgk1^{T81} and OsHPP02^{E74}.

MW = molecular weight, E = pooled elution from IMAC-GF peak, 3C = 3C protease treated sample, RI = flowthrough from reverse IMAC/tandem MBP trap and IMAC, A₁₋₃ = washes with buffer A1, B1 = wash with buffer B1. IN = gel filtration input.

As previously, the presence of both proteins in the complex was verified by intact mass spectrometry (Figure 5-24). A mass of 6524.55 Da was experimentally determined, corresponding to AVR-Mgk1^{T81} (predicted peptide mass of 6526.58 Da); the difference of 2.0 Da is perfectly accounted for by the loss of two H atoms during formation of the expected disulphide bond. A further mass of 7816.43 Da was experimentally determined, corresponding to OsHPP02^{E74} (predicted peptide mass of 7816.10 Da).

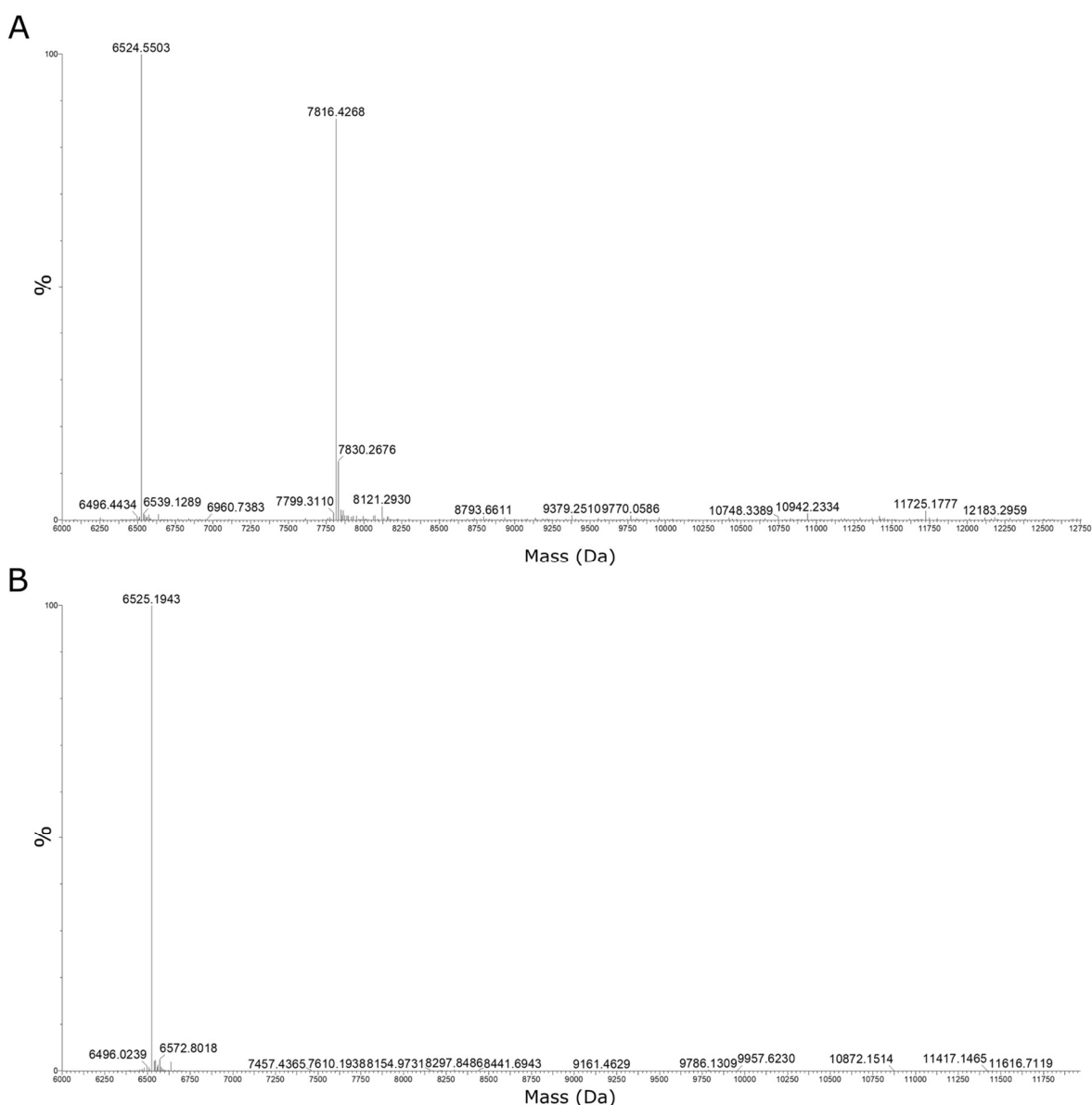


Figure 5-24: Intact mass spectra of the purified AVR-Mgk1^{T81}/OsHPP02^{E74} complex.

A) Mass spectrum from the first gel filtration peak (Figure 5-23 E) - masses of 6524.55 Da and 7816.43 Da were observed, corresponding to AVR-Mgk1^{T81} (predicted peptide mass 6526.58 Da) and OsHPP02^{E74} (predicted peptide mass 7816.18 Da). **B)** Mass spectrum from the second gel filtration peak (Figure 5-23 E). A single mass of 6524.55 Da was detected, corresponding to AVR-Mgk1^{T81} alone (predicted peptide mass 6526.58 Da).

Having now purified and confirmed the identity of the AVR-Mgk1^{T81}/OsHPPo2^{E74} complex, it was dispensed into Morpheus, SG1, and JCSG screens for crystallisation trials at concentrations of 12.4 mg/ml and 24.8 mg/ml, in buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl). No crystals were obtained from either the Morpheus or JCSG screens, but in well H8 of SG1 several crystals were obtained (0.1 M Sodium acetate trihydrate; 0.1 M Bis-

Tris pH 5.5; 17 % w/v PEG 10000). These crystals had UV fluorescence, indicating that they contained protein, and had clear faces and edges, albeit with a slightly rounded appearance.

In attempt to reproduce these crystals and potentially improve their morphology, an optimisation screen was set up to explore crystallisation conditions adjacent to the original condition (Sodium acetate trihydrate 0.1M, Bis-tris 0.1M, pH 5-8, PEG 10000 10-25 % w/v) (5.2.5.1 Crystallisation). Conditions were chosen so that the original SG1 well H8 conditions were almost exactly reproduced in one of the central wells of the optimisation plate. As previously, proteins were at concentrations of 12.4 mg/ml and 24.8 mg/ml, in buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl). After observation for several weeks, no crystals similar to the original hit had formed, however there was crystalline material in some of the wells with a distinctive appearance of a double ended bundle of rounded needles, reminiscent of a wheat sheaf.

Both the initial crystals obtained in SG1 H8 and the crystalline material obtained in the optimisation screen were fragile when handled during harvesting. At a remote access session at Beamline Io4 of Diamond Light Source, it was confirmed that, although composed of protein, the crystals from SG1 were not suitable for data collection. They produced an indistinct and smeared diffraction pattern, indicating a disordered and damaged crystal lattice. The “wheat sheaf” sample obtained from the SG1 H8 optimisation plate would never have been a feasible sample for protein structure determination due to the macroscopically visible overlapping crystal lattices, however X-ray diffraction confirmed that it was composed of protein (Figure 5-25) and so I felt it would be reasonable to use this crystalline material for seeding.

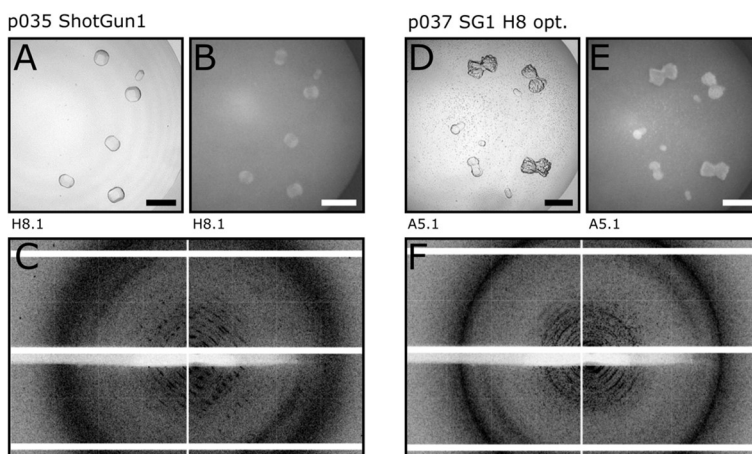


Figure 5-25: Crystals obtained in the ShotGun1 screen with the AVR-Mgk1^{T81}-OsHPP02^{E74} complex.

A) ShotGun1 condition H8 in white light **B)** as in A, in UV. **C)** X-ray diffraction image from one of the crystals in A and B. **D)** from condition A5 of the ShotGun1 optimisation plate, in white light **E)** as D, in UV. **F)** X-ray diffraction image from one of the crystalline bundles in D and E. Scale bars = 200 μ m.

Next, three further initial crystal screens (BCS, PACT, PEGs) were set up using the truncated AVR-Mgk1^{T81}/OsHPP02^{E74} complex at 24.8 and 12.4 mg/ml, again in buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl). These screens were selected for their richness in different PEG conditions, like the original hit in Shotgun condition H8. These screens yielded several semi-crystalline forms, including bi-lobed rough spherical “walnuts”, some with facets, some needle bi-lobed “propellers”, and more of the bundle of needles “wheat sheaf” morphology, mainly in the BCS and PACT screens.

In pursuit of single crystals, “wheat sheaf” bundled needle crystalline material was harvested from well C7-2 of the previous PACT screen as a seed stock, and dispensed into a new PACT crystal screen with AVR-Mgk1^{T81}/OsHPP02^{E74} at concentrations of 12.3 mg/ml and 6.15 mg/ml, in buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl). A second identical screen was set up without addition of the seed stock. The protein concentrations were halved for two reasons. First, because it was observed that in the previous PACT screen with AVR-Mgk1^{T81}/OsHPP02^{E74} at 24.8 mg/ml and 12.4 mg/ml, higher protein concentration wells which contained crystalline material with a fine precipitate surrounding it usually also had a hit in the lower concentration well. Second, introduction of seed material would accelerate the process of crystal formation as nucleation points are available, therefore the protein concentration should be reduced to slow the rate of crystal formation to encourage the formation of a more ordered crystal lattice. Ultimately, these screens yielded some crystalline material though no suitable single crystals.

As a final seeding attempt, crystalline material from well C7-1 of the PACT screen was taken and seeded into a Morpheus screen with AVR-Mgk1^{T81}-OsHPP02^{E74} at concentrations of 18 mg/ml and 9 mg/ml in buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl). After two weeks, the seeded screen yielded a range of crystal morphologies which appeared ideal for X-ray diffraction, clustered around row C (Figure 5-26). These conditions are centred around the NPS precipitant mix which contains sodium nitrate, disodium hydrogen phosphate and ammonium sulfate, at pH 6.5, 7.5, and 8.5. An identical screen without seeding was also set up but yielded no crystals. A variety of crystals from the seeded screen were cooled and sent to Diamond Light Source Beamline i24 for remote access X-ray diffraction where all were determined to be salt crystals. Cross-seeding between screens with different precipitants is a risk factor for salt crystal formation; for this reason, no further cross-seeding attempts were made.

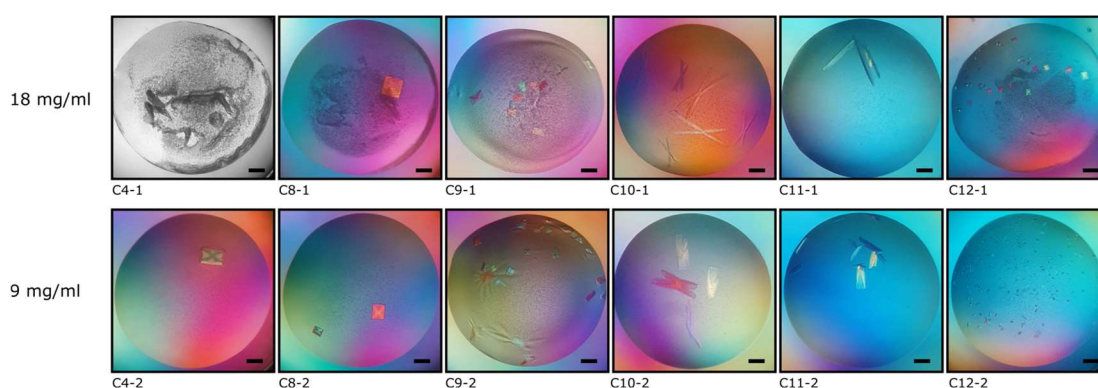


Figure 5-26: Crystals obtained by seeding into the Morpheus screen with AVR-Mgk1^{T81}/OsHPP02^{E74}

Morpheus crystallisation screen set up with AVR-Mgk1^{T81}/OsHPP02^{E74} (18 mg/ml and 9 mg/ml) and seeding with seed stock obtained from PACT condition C7. Well numbers indicated below each image; bright field images, some taken with polarising filter, black and white images acquired using the Formulatrix Rock Imager. Scale bar = 200 μm.

A new set of previously unexplored preliminary screens were then set up for the AVR-Mgk1^{T81}-OsHPP02^{E74} complex (KISS, MIDAS, ProPlex) at two concentrations, 24.8 mg/ml and 12.4 mg/ml. As previously, proteins were in buffer A4, (20 mM HEPES pH 7.5, 150 mM NaCl). All these screens all yielded interesting hits of varying quality; in KISS, rounded semi-crystalline material, in MIDAS, very fine needle crystals which were growing from the edge of the drop, and in ProPlex, a range of crystals many of which appeared to have a hexagonal prismatic shape, some tapered, and of varying sizes. Of particular interest were ProPlex wells E12.1, F1.2, and H1.1 and H1.2 which contained large and medium sized crystals of around 200 μ m length (Figure 5-27).

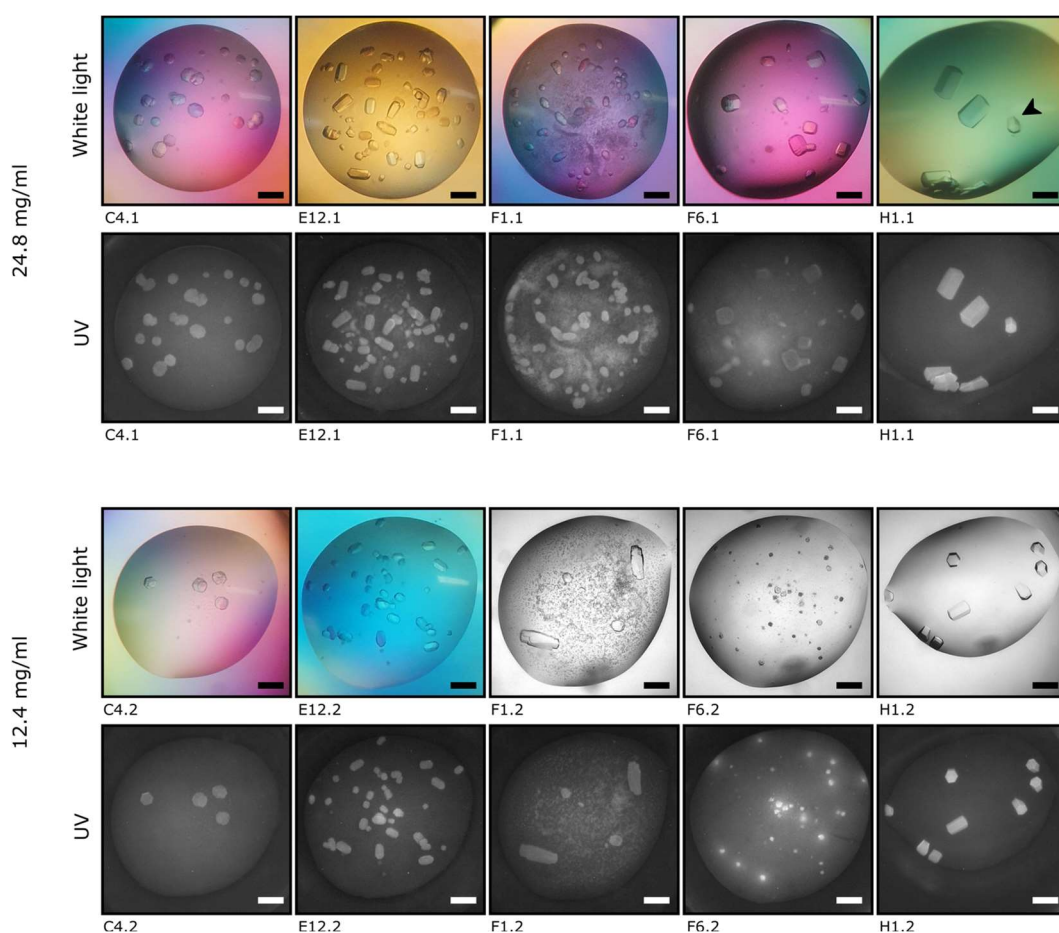


Figure 5-27: Crystals obtained from the ProPlex screen with AVR-Mgk1^{T81}/OsHPP02^{E74}

Photographs of selected crystals obtained from the ProPlex crystallisation screen with AVR-Mgk1^{T81}/OsHPP02^{E74}. Well numbers indicated below each image; upper row are bright field images, some taken with polarising filter, black and white images acquired using the Formulatrix Rock Imager; lower row are UV images acquired with the Formulatrix Rock Imager. Black arrowhead indicates the smallest crystal of H1.1, which produced the data set used for structure determination Scale bar = 200 μ m.

5.3.9 X-ray data collection and structure determination of the AVR-Mgk1^{T81}/OsHPPo2^{E74} complex

Diffraction data were collected from crystals from ProPlex wells E12.2 (0.2 M Ammonium sulfate, 0.1 M MES pH 6.5, 20 % w/v PEG 8000) and H1.1 (2.0 M Sodium chloride, 0.1 M Sodium citrate pH 6.0). Protein was suspended in buffer A4, (20 mM HEPES pH 7.5, 150 mM NaCl). As previously, these crystals were cooled and sent to Beamline I24, Diamond Light Source (Oxfordshire, UK). The smallest crystal from well H1.1 diffracted and a total of 3600 x 0.1° images were recorded to a maximum resolution of 2.0 Å at a wavelength of 0.6199 Å. Further details of data collection can be found in the methods section (5.2.5.3).

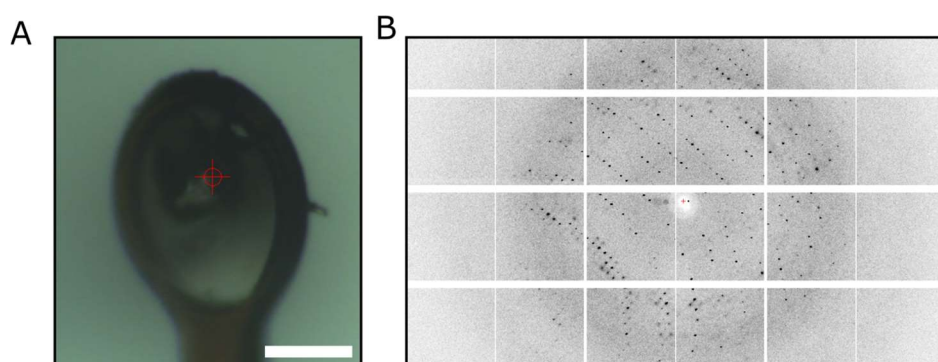


Figure 5-28: AVR-Mgk1^{T81}/OsHPPo2^{E74} crystal and representative diffraction image

A) Image of crystal mounted in loop immediately prior to data collection. Scale bar = 100 μ m, **B)** A representative diffraction image from this crystal, exhibiting a high density of clearly defined and separated diffraction spots. The image is cropped, and to improve contrast for display, five 0.1 ° images were summed.

The dataset was solved by molecular replacement, using Pkm-HMA from the previously determined crystal structure of Pkm-HMA and AVR-PikD (PDB: 6FUD) (De La Concepcion et al., 2018) and an AlphaFold2 predicted structure of AVR-Mgk1^{T81} as search models. The space group was $P3_121$ with cell parameters $a = 71.57$ Å, $b = 71.57$ Å, $c = 68.85$ Å, $\alpha = 90^\circ$, $\beta = 90^\circ$, $\gamma = 120^\circ$. Data collection and refinement statistics are summarized in Table 5-2.

The structure revealed a 1:1 complex between AVR-Mgk1^{T81} and OsHPPo2^{E74} (Figure 5-29). All 57 residues of the AVR-Mgk1^{T81} construct (R25-T81), and 71 residues of the OsHPPo2^{E74} construct (A2-V72) were able to be modelled (Figure 5-29 A). AVR-Mgk1 adopts a clearly defined β -sandwich MAX fold, made up of a three stranded β -sheet opposing a two-stranded β -sheet, connected by an unstructured loop between β -strand 4 and 5. The final predicted β -sheet of AVR-Mgk1 was not resolved, perhaps due to truncation of too many residues leading to more flexibility; however is at the opposite side of AVR-Mgk1 to the interface with OsHPPo2. The unstructured loop of AVR-Mgk1 is uninterrupted, in contrast

to other MAX effectors such as AVR-Pik, AVR-Pia, ToxB, and AVR1CO39 in which the unstructured loop is disrupted by a short β -strand. The two cysteine residues of AVR-Mgk1, Cys27 and Cys67, are linked by a clearly resolved disulfide bond (Figure 5-30 B). Therefore, this structure provides further experimental confirmation for the expected single disulfide bond as previously inferred by intact mass spectrometry of purified AVR-Mgk1 and AVR-Mgk1^{T81} proteins.

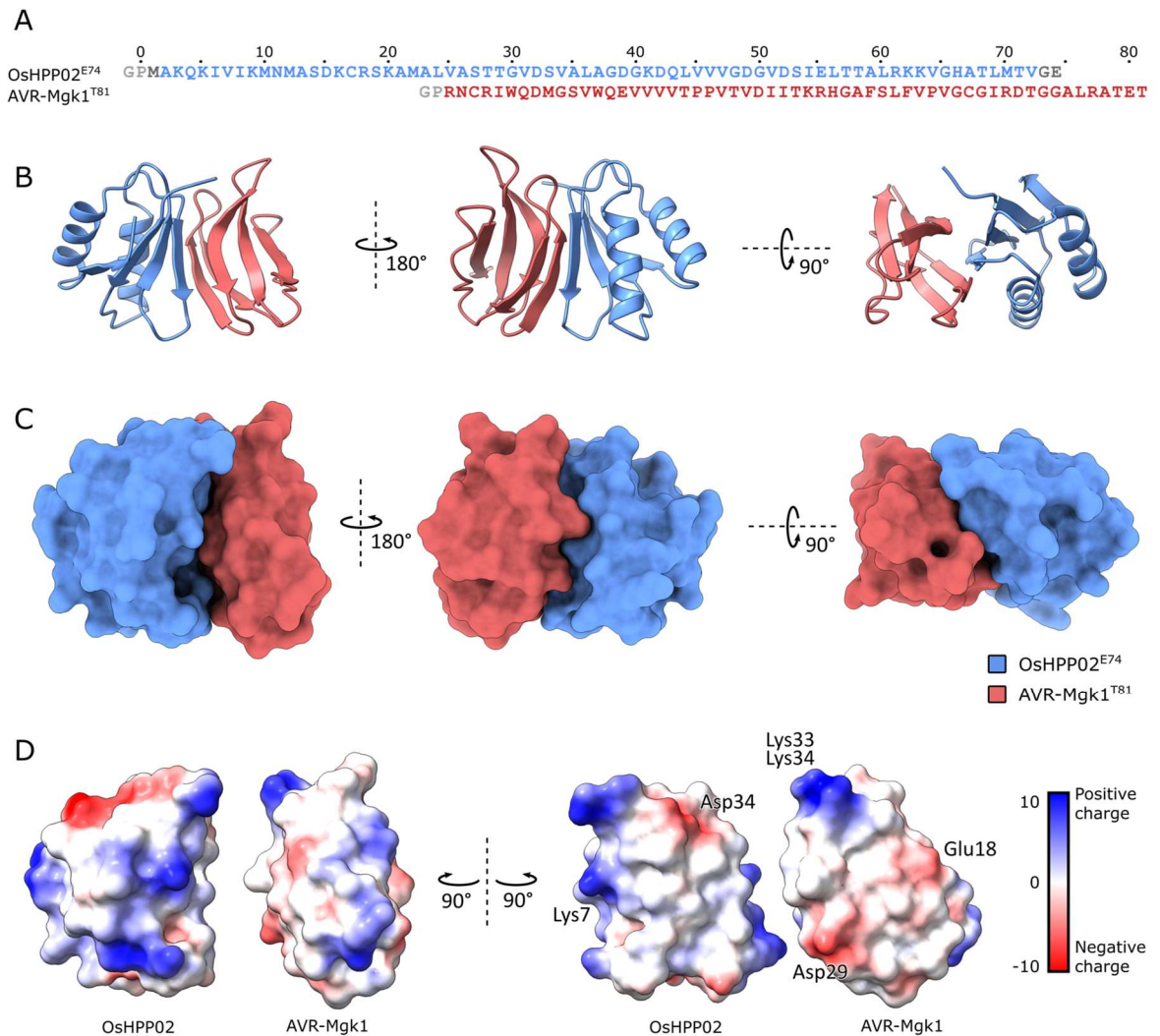


Figure 5-29: X-ray crystal structure of the AVR-Mgk1^{T81}/OsHPP02^{E74} complex

A) Amino acid sequences of both proteins. GP residues at the N-terminus of each protein (indicated in light grey) are a 3C protease cleavage scar resulting from the cleavage of affinity tags required for purification and could not be resolved. Residues M1, G73, E74 of OsHPP02^{E74} could not be resolved (indicated in dark grey). **B)** Ribbon representation of the complex, **C)** surface representation of the complex. Throughout panels **A-C**, AVR-Mgk1^{T81} is coloured in red, and OsHPP02^{E74} is coloured in blue. **D)** Surface representation coloured according to charge, where blue = positive charge and red = negative charge. With lateral separation (left), and with 90° rotation (right) to reveal the interaction interface.

OsHPPo2^{E74} adopts a typical HMA fold, with a β -sheet made up of three β -strands and backed by a pair of α -helices. The AVR-Mgk1^{T81} and OsHPPo2^{E74} complex is made up of a continuous antiparallel five stranded β -sheet which spans the two proteins, with three β -strands contributed by AVR-Mgk1^{T81} and two by OsHPPo2^{E74}. Eight intermolecular hydrogen bonds stabilise the complex: five between the backbone of β 2 of OsHPPo2^{E74} and β 3 of AVR-Mgk1^{T81}, directly stabilising the continuous antiparallel beta sheet that spans the binding interface. The remaining three hydrogen bonds are found between the backbone and side chain of Asp32, on β 3 of OsHPPo2^{E74}, with Gln31 and Gln38 of AVR-Mgk1 respectively. Finally, a salt bridge is formed between Lys5 of OsHPPo2 and Glu39 of AVR-Mgk1.

The interaction between OsHPPo2/AVR-Mgk1 is also mediated by charge complementarity between a positive patch located on the β 2- β 3 loop of AVR-Mgk1, produced by residues Lys33 and Lys34, and negatively charged Asp34 located on the α 1- α 2 loop of OsHPPo2 (Figure 5-29).

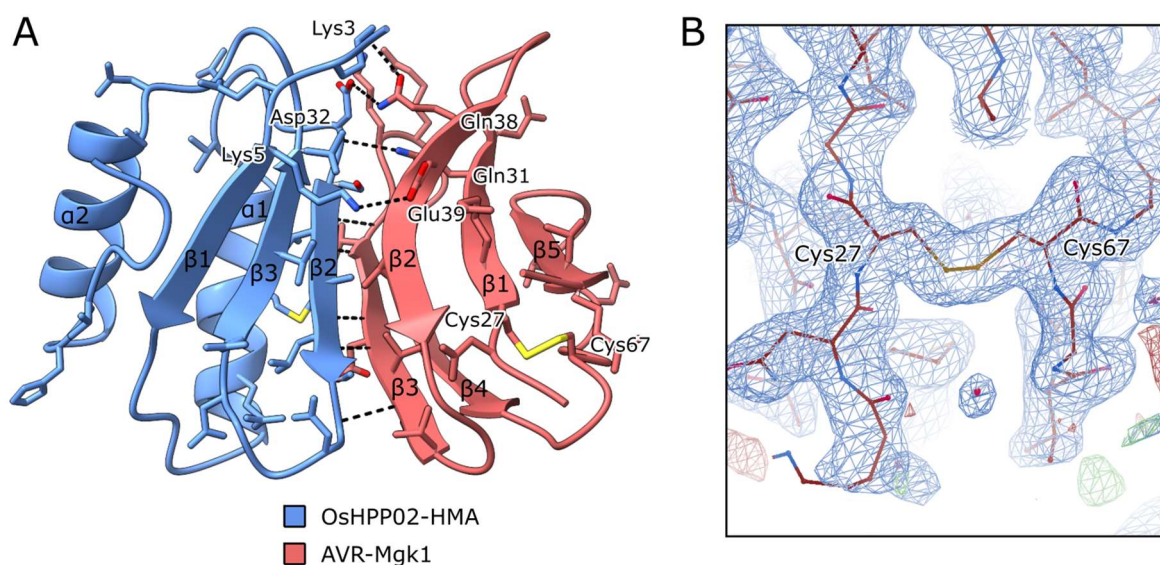
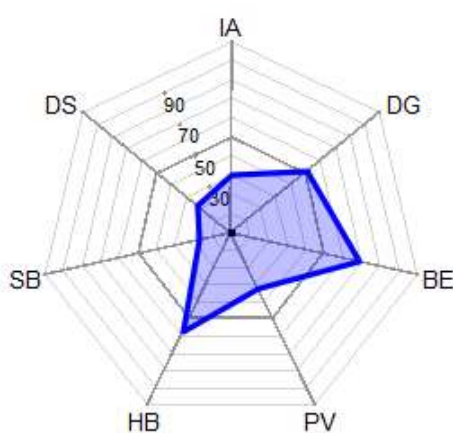


Figure 5-30: Details of the AVR-Mgk1^{T81}/OsHPPo2^{E74} interaction

A) The AVR-Mgk1^{T81}/OsHPPo2^{E74} complex in ribbon representation. Hydrogen bonds and salt bridges are indicated by black dashed lines. Secondary structure elements are labelled. **B)** Detail view of AVR-Mgk1, centred on the disulfide bond between Cys27 and Cys67. Electron density map shown in blue mesh, contoured to 0.192 e/Å² (1.12 rmsd), model in ball and stick representation.

QtPISA was used to analyse the interfaces (Krissinel and Henrick, 2007; Krissinel, 2010). The output reported that 15.5 % of the residues of AVR-Mgk1 and 12.9 % of the residues of OsHPPo2 are buried in the within the interface ($550.4 \text{ \AA}^2$ and $559.9 \text{ \AA}^2$ respectively). The interface area is therefore a modest $555.1 \text{ \AA}^2$, which is almost half that of the interface between Pikp-HMA/AVR-PikD ($1031.0 \text{ \AA}^2$) (Maqbool et al., 2015). Interface analysis parameters are reported below (Figure 5-31).



Interface Parameters

IA: Interface Area	555.1 $\text{\AA}^2$
DG: Delta G	-6.2 kcal/mol
BE: Binding Energy	-10.6 kcal/mol
PV: Hydrophobic P-value	0.2787
HB: Hydrogen Bonds	9
SB: Salt Bridges	1
DS: Disulfide Bonds	0

Figure 5-31: QtPISA analysis of the AVR-Mgk1^{T81}/OsHPPo2^{E74} interface

QtPISA radar plot ranks the likelihood of the interface being part of a biological interface based on statistical analysis which compares the interface with all interfaces in the PDB (the larger the area, the greater the probability). The interface is ranked across seven categories which are plotted on the radar beams, defined in the table below the plot.

Table 5-2: Data collection and refinement statistics for the AVR-Mgk1^{T81}/OsHPP02^{E74} complex structure

Data collection statistics	
Beamline	I24, Diamond Light Source, UK
Wavelength (Å)	0.6199
Detector	Eiger2 9M
Resolution range (Å)	46.06-2.05 (2.00-2.00)
Space group	<i>P</i> ₃ 21
Cell dimensions (<i>a</i> , <i>b</i> , <i>c</i>) (Å)	71.57, 71.57, 68.85
Cell angles (α , β , γ) (°)	90.0, 90.0, 120.0
Total no. of observations	286,566 (21,246)
Unique reflections	14,186 (1,025)
Multiplicity	20.7 (20.7)
Mean(<i>I</i> / σ <i>I</i>)	21.6 (2.2)
Completeness (%)	100.0 (100.0)
<i>R</i> _{merge}	0.082 (4.336)
<i>R</i> _{meas} .	0.084 (4.444)
CC(1/2)	1.000 (0.767)
Wilson B value (Å ²)	42.3
Refinement and model building statistics	
Resolution range (Å)	46.11-2.00
Reflections: working/free	13477 (709)
<i>R</i> _{work} / <i>R</i> _{free}	0.202 / 0.260
Ramachandran plot favoured/allowed/outliers (%)	99.19/0.81/0.00
RMSD	
Bond length deviation (Å)	0.0164
Bond angle deviation (°)	2.70
No. atoms	
Amino acids	967
Waters	66
B-factors	
Amino acids	38.75
Waters	53.6
MolProbity Score	1.77

Values in parentheses are for the highest resolution shell. RMSD = root mean square deviation

5.3.10 Biophysical characterisation of AVR-Mgk1 interactions with different HMAs

Surface plasmon resonance (SPR) is a biophysical technique which can be used to probe biomolecular interactions, including protein-protein interactions, without need for labelling. One protein, termed the ligand, is bound to the sensor chip, and the second protein, a potential binding partner, the analyte, is flowed over the surface of the chip through microfluidic channels at a steady rate. Binding affinities and kinetics (association and dissociation rates) can be obtained. To investigate the interaction between AVR-Mgk1 and both integrated HMAs and small HMAs, SPR experiments were conducted with support from Dr Clare Stevenson, Dr Abbas Maqbool, and Dr Adam Benthams, using the Biacore 8K instrument (Cytiva). The HMAs Piks-HMA, Pikm-HMA, OsHPPo2^{K77}, OsHPPo3^{K77}, and OsHPPo4^{K77} were covalently coupled to the sensor chip as ligands, and the effectors AVR-Mgk1, AVR-PikD, and AVR-PikF were flowed over as analytes, which is consistent with previous work (Maqbool et al., 2015; Benthams et al., 2023). Previously purified AVR-PikF and AVR-PikD protein was a gift from Dr Adam Benthams and Dr Rafal Zdrzalek respectively. OsHPPo3^{K77} and OsHPPo4^{K77} were expressed and purified by the same method as OsHPPo2^{K22} (Figure 5-32 A, B). A₂₈₀ chromatograms are not displayed as these proteins lack aromatic amino acids and therefore do not have absorbance at 280 nm. Finally, SDS-PAGE of all proteins used in the SPR experiment was performed (Figure 5-32 C).

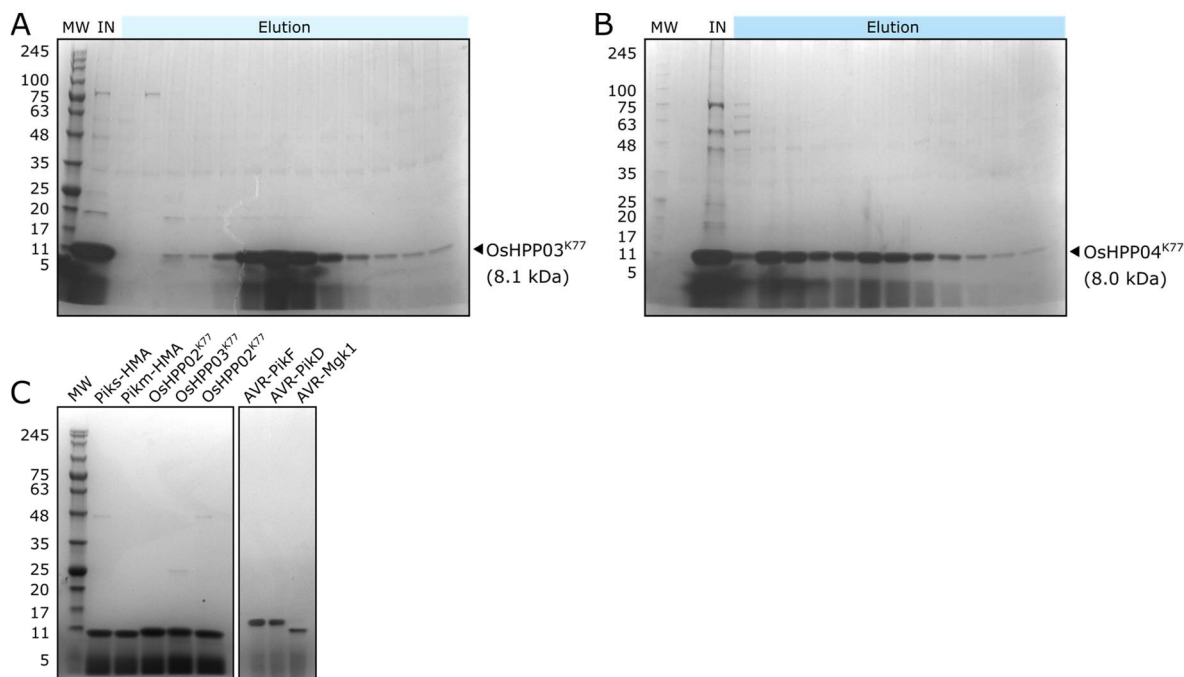


Figure 5-32: SDS-PAGE of proteins used in SPR experiments

A) SDS-PAGE of final gel filtration of the purification of OsHPP03^{K77}, **B)** SDS-PAGE of final gel filtration of the purification of OsHPP04^{K77}, **C)** SDS-PAGE of all HMA and effector proteins used in the surface plasmon resonance experiments. MW = molecular weight, IN = input sample into final gel filtration, Elution = gel filtration elution.

A CM5 sensor chip was used, which is coated in a dense yet flexible matrix of carboxymethylated dextran which extends around 100 nm from the chip surface and allows for some movement of the immobilised ligands. An irreversible covalent coupling reaction is employed to link amine groups of a protein to the chip matrix. It is important to immobilise the correct density of ligand on the chip as different applications of SPR all have different requirements. In the case of protein-protein interaction experiments, kinetics measurements are best determined using a sensor chip with a low ligand density, to avoid any mass transport effects, and affinity measurements are best made using a chip with a moderate ligand density. To facilitate an appropriate level of ligand loading on the surface, the correct buffer conditions for coupling should be determined. Therefore, careful pH and concentration scouting experiments were carried out before undertaking the irreversible coupling reaction. pH 4.5, 5.0, and 5.5 buffers were assessed, and it was determined that pH 5.5 was best for OsHPPo2, OsHPPo3, and OsHPPo4, pH 4.5 was best for Pikm-HMA, and pH 4.0 was best for Piks-HMA.

Initial experiments were carried out using the multi-cycle kinetics (MCK) experimental design favoured in previous studies on Pik-HMA domains (Maqbool et al., 2015; Bentham et al., 2023). The shorter cycle times of MCK experiments carry the advantage of lower drift, but the drawbacks include that slow interactions may not reach equilibrium during the cycle, and also that effective regeneration conditions are required. It was found that the response of each successive cycle was lower, which can indicate instability of the ligand on the chip surface, or that regeneration conditions are unsuitable. If regeneration is too harsh, HMA activity can be lost from the chip surface, or if regeneration is not stringent enough, effectors can remain bound to the HMA after regeneration and therefore the binding capacity available in subsequent cycles is diminished. After inspection of the sensorgrams it was apparent that after regeneration the response was not returning to the starting level, so higher stringency regeneration conditions would be required.

In addition, a very slow dissociation of AVR-Mgk1 from OsHPPo2^{K77}, OsHPPo3^{K77}, and OsHPPo4^{K77}, was observed. To accurately capture the dynamics of the interaction, a dissociation of at least 10% of the $R_{\max}$ should be observed. By changing to a single cycle kinetics strategy, the overall number of dissociation phases is reduced and so the dissociation duration could be extended to 600 seconds, while the total experimental duration remained reasonable. This would allow more time for AVR-Mgk1 to dissociate at the end of the cycle.

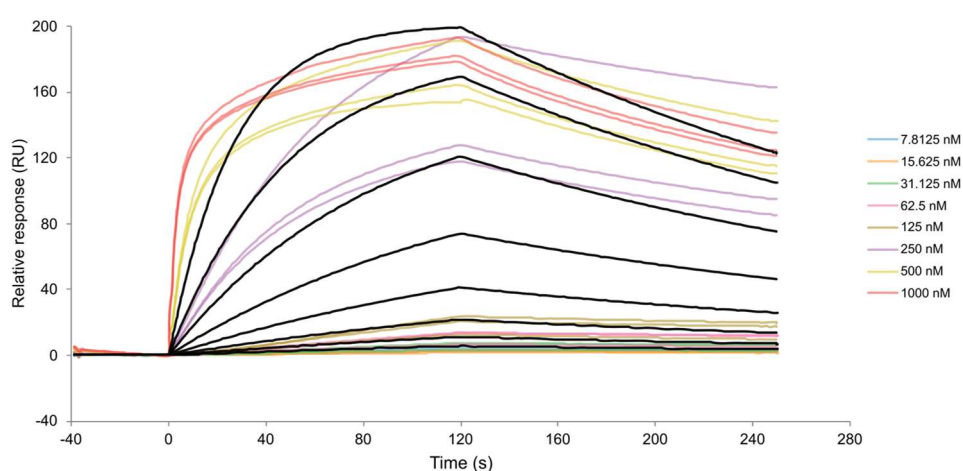


Figure 5-33: Example sensorgram from an early multi-cycle kinetics SPR experiment.

Multicycle kinetics experiments suffered from deteriorating $R_{\max}$ with each successive cycle. Example sensorgrams of analyte AVR-Mgk1 addressing the OsHPPo2^{K77} channel. 8 AVR-Mgk1 concentrations were used, in a two-fold dilution series from a top concentration of 1000 nM. Data are coloured by concentration, and the thick black lines represent the best fit obtained by the Biacore evaluation software.

The “ionic regeneration” 3x stock solution published in a Cytiva study of buffer space for SPR experiments (Andersson et al., 1999) was employed undiluted by Dr Adam Bentham and Dr Juan Carlos de la Concepcion in their SPR study of HMA mutants of Pikm-HMA and Pikp-HMA (Bentham et al., 2023). The regeneration buffer composition used by Bentham, De la Concepcion et al. was therefore 0.46 M Potassium thiocyanate, 1.83 M MgCl₂, 0.92 M urea, and 1.83 M guanidine-HCl which provided a strong denaturing environment.

The same buffer with the omission of potassium thiocyanate was successfully used in this study (1.83 M MgCl₂, 0.92 M urea, and 1.83 M guanidine-HCl). A series of regeneration tests were performed with this modified denaturing regeneration buffer to ensure that all analyte was removed after each cycle. Ultimately this regeneration buffer was found to be highly effective when applied to the chip at 100 ul/min for 60 seconds and followed by a 600 second application of running buffer which allows time for refolding of any denatured HMA proteins.

Although effective in regeneration tests of relatively short duration, the performance of the HMA chip when exposed to regular washes in this stringent modified denaturing regeneration solution over the course of a full 12–18-hour MCK experiment had not been assessed. There was a risk that with repeated application over a relatively long experimental period, HMA integrity could deteriorate. Multi-cycle kinetics experiments require regeneration to be performed between each cycle, and therefore to reduce the total number of washes required, a single-cycle kinetics experimental design was adopted for all future experiments.

Further experiments were performed to optimise the sample application order and duration, analyte layout in the plates, and analyte concentration ranges. First a blank cycle was performed, followed by AVR-PikF, AVR-PikD, and AVR-Mgk1 in order. This was repeated four times followed by a final blank cycle, and the data from the first repeat excluded so that only the latter three cycles were analysed. This was because the first cycle was found to have a slightly higher response, and the subsequent cycles were all very consistent with each other. This could be attributed to a chip conditioning effect from the first analyte series. Analyte concentration range always included a 0 nM concentration at the beginning of each series, before 8 subsequent concentrations progressing up a three-fold dilution series produced from a top concentration of 1000 nM or 3000 nM. To improve the accuracy of K_D determination, the analyte concentration range was selected to reach as close as possible to 10 times above and 10 times below the preliminary K_D determined in earlier experiments.

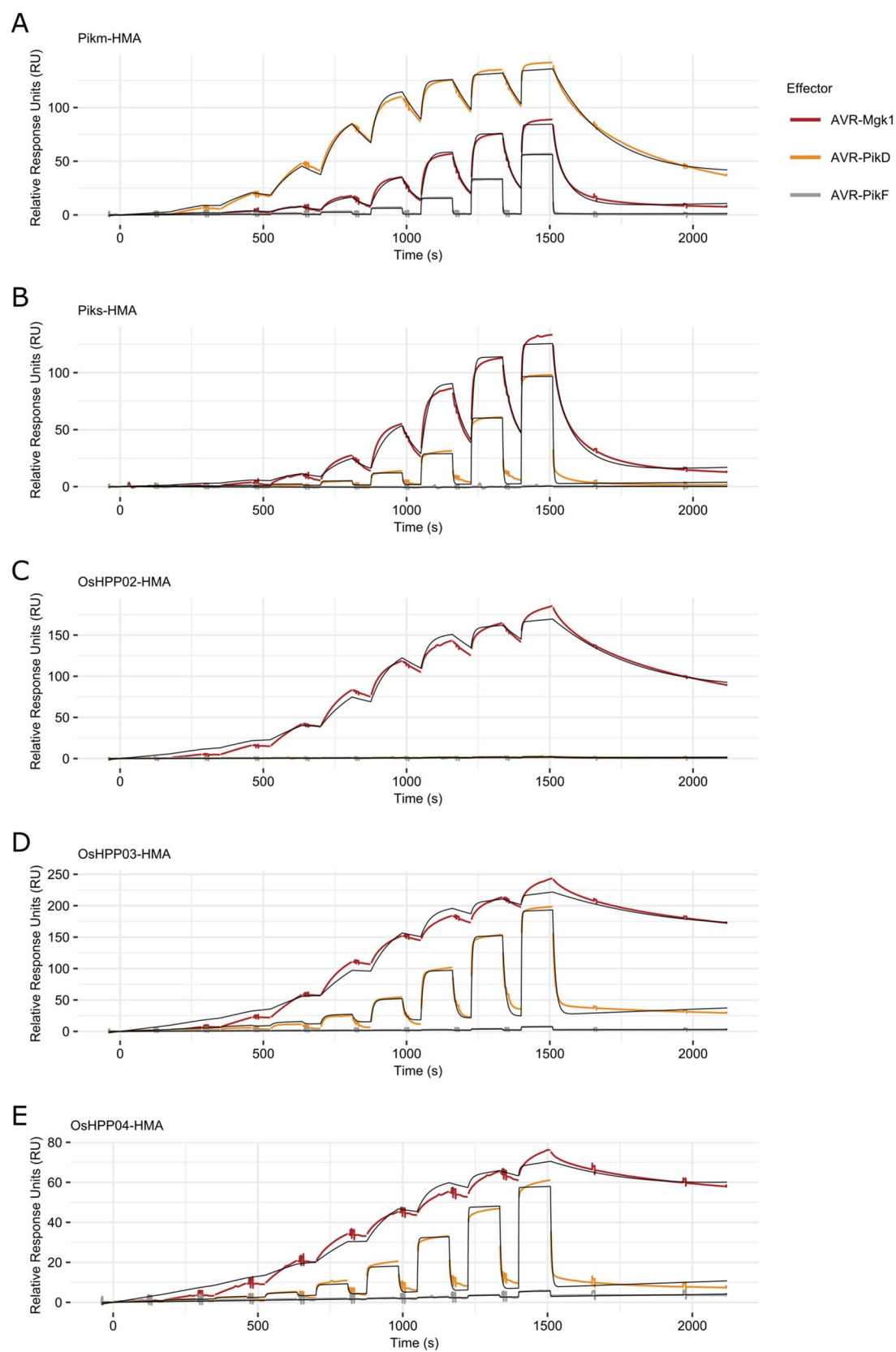


Figure 5-34: Single-cycle kinetics SPR sensorgrams.

Each panel represents a different channel, **A)** Pikm-HMA, **B)** Piks-HMA, **C)** OsHPP02-HMA, **D)** OsHPP03-HPP, **E)** OsHPP04. Each HMA was addressed in turn with effectors AVR-PikF (grey), AVR-PikD (orange), AVR-Mgk1

(red). Effectors were injected at 0 nM and then in 8 steps of increasing concentration from 0.46 nM - 1000 nM, apart from Pikm-HMA/AVR-PikF, Piks-HMA/AVR-PikD, and OsHPP04^{K77}/AVR-PikD where effectors were injected at 0 nM and then in 8 increasing steps from 1.37 nM - 3000 nM. A representative sensorgram is displayed for each effector, and the Biacore single cycle kinetics line of best fit is displayed in black.

In agreement with the literature, and confirming its role as the negative control analyte, AVR-PikF did not interact with Piks-HMA or Pikm-HMA. Some response was recorded as AVR-PikF was injected over the Pikm-HMA surface, although the square shape of the sensorgram indicates a bulk effect and not a true binding response (Figure 5-34 A). AVR-PikF also did not interact with OsHPP02^{K77}, OsHPP03^{K77}, or OsHPP04^{K77}.

AVR-PikD interacted with Pikm-HMA with high affinity, confirming the positive control for binding in this experiment ($K_D = 8$ nM). This K_D is in close agreement with the previously determined $K_D = 10$ nM binding affinity (Bentham et al., 2023)). AVR-PikD was also able to interact with Piks-HMA ($K_D = 612$ nM), OsHPP03^{K77} ($K_D = 120$ nM), and OsHPP04 ($K_D = 165$ nM). No interaction was recorded between AVR-PikD and OsHPP02^{K77}. These results are consistent with the yeast two hybrid and AlphaScreen assays performed by Oikawa et al. in which interaction was observed between AVR-PikD and OsHPP03 and OsHPP04, but not with OsHPP02 (Oikawa et al., 2024).

Finally, AVR-Mgk1 was observed binding to all HMAs investigated in this experiment. AVR-Mgk1 interacts with both Piks-HMA ($K_D = 56$ nM) and Pikm-HMA ($K_D = 66$ nM) at similar affinities in contrast to AVR-PikD which displays binding approximately two orders of magnitude stronger to Pikm-HMA ($K_D = 8$ nM), compared to Piks-HMA ($K_D = 612$ nM). AVR-Mgk1 also displayed equal binding affinities to each of the OsHPPs investigated: OsHPP02^{K77} ($K_D = 17$ nM), OsHPP03^{K77} ($K_D = 17$ nM), and OsHPP04^{K77} ($K_D = 21$ nM). The shallow curve in the sensorgram during dissociation of AVR-Mgk1 from the OsHPPs channels indicates the slow dissociation rate (Figure 5-34 C-E).

Table 5-3: Summary of K_D determined by SPR

	Piks-HMA	Pikm-HMA	OsHPP02 ^{K77}	OsHPP03 ^{K77}	OsHPP04 ^{K77}
AVR-PikF	No interaction	No interaction	No interaction	No interaction	No interaction
AVR-PikD	612 nM	8 nM	No interaction	120 nM	165 nM
AVR-Mgk1	56 nM	66 nM	17 nM	17 nM	21 nM

5.3.11 Engineering the OsHPPo2/o3/o4 HMAs into the Pik chassis to provide proof-of-concept recognition of AVR-Mgk1

Prior work led by the Banfield Lab developed the NLR pair Pik as a platform for engineering intracellular immunity. It has been demonstrated that the recognition specificity of Pik can be modified through manipulation of the HMA domain in a variety of ways. The HMA of Pik-1 can be mutated to mimic the HMAs of other Pik-1 alleles, such as in (De La Concepcion et al., 2018), or can be exchanged for novel domains such as the small HMA OsHIPP19 (Maidment et al., 2021; Maidment et al., 2023), the HMA of RGA5 (Bentham et al., 2023), or protein domains from outside the plant lineage such as nanobodies (Kourelis et al., 2023). In a further mode of regulation available to the Pik NLR pair, if Pikm-1 is expressed in combination with Pikp-2, the immune response is dampened in comparison to when Pikm-1 is co-expressed with Pikm-2 (Bentham et al., 2023).

To provide further proof-of-concept supporting use of the Pik-1 chassis for engineering effector recognition through exchange of small HMA domains, recognition of AVR-Mgk1 by chimeric Pik NLRs was investigated. As discussed previously (5.3.10), when assessed by SPR the interaction between OsHPPo2, OsHPPo3, and OsHPPo4 HMA and AVR-Mgk1 is of higher affinity than the interaction between Pikm-HMA and AVR-Mgk1 or Piks-HMA and AVR-Mgk1. From this, the hypothesis was formed that Pikm-1 mediated recognition of AVR-Mgk1 would be strengthened after exchange of the wild-type Pikm-1-HMA for the HMA domain of OsHPPo2, OsHPPo3, or OsHPPo4 to form the chimeric NLRs Pikm-1^{HPPo2}, Pikm-1^{HPPo3}, or Pikm-1^{HPPo4}.

Chimeric NLR proteins Pikm-1^{HPPo2}, Pikm-1^{HPPo3}, or Pikm-1^{HPPo4} were generated by exchanging the Pikm-1 HMA for the HMA of OsHPPo2, OsHPPo3, and OsHPPo4 in the Pik chassis developed by Dr Adam Bentham and Dr Mark Youles. In this chassis, Pikm-1 with a C-terminal FLAG tag under the control of the constitutive viral promotor double 35S promotor is combined with either Pikm-2 or Pikp-2 with a C-terminal HA tag under the control of the MAS (*Agrobacterium* derived mannopine synthetase) promotor (2.1.1.1). The HMA domain of Pikm-1 can be exchanged through a golden gate cloning reaction, and a new protein sequence inserted in that position of the protein. The constructs inserted were OsHPPo2^{K77}, OsHPPo3^{K81} (due to an error in primer design), and OsHPPo4^{K77}. Effector constructs AVR-PikD (obtained from Dr Rafal Zdrzalek) and AVR-Mgk1 each with a N-terminal 4xMyc epitope tag and under the control of the At10Ubi promotor (*Arabidopsis* Ubiquitin promotor) were also prepared. These constructs were transformed into *Agrobacterium* and infiltrated into *N. benthamiana* leaves individually for expression testing by western blot.

Expression of the chimeric Pikm-1^{HPP02}, Pikm-1^{HPP03}, or Pikm-1^{HPP04}, and of the effectors was confirmed by western blot (Figure 5-35). Expression of Pikm-2-HA and Pkcp-2-HA was not confirmed by western blot, however these sequences were present on the same plasmid as the chimeric Pikm-1^{HPP}, and control cell death areas infiltrated with the same construct did produce cell death indicating that these proteins were expressed.

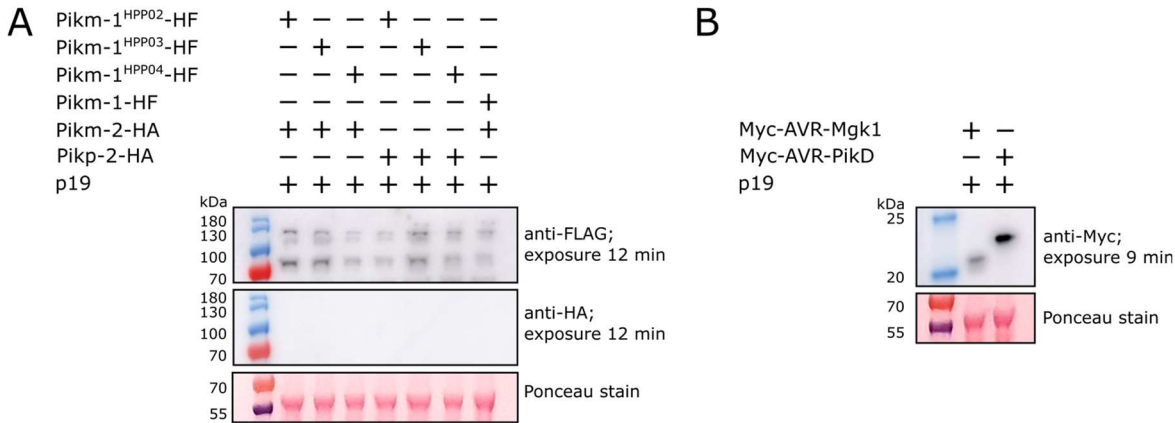


Figure 5-35: Western blot confirming expression of chimeric Pikm-1^{HPP} constructs in *N. benthamiana*

A) Pikm-1^{HPP02}-FLAG, Pikm-1^{HPP03}-FLAG, or Pikm-1^{HPP04}-FLAG in combination with Pikm-2-HA or Pikp-2-HA. **B)** Myc-AVR-Mgk1 and Myc-AVR-PikD.

The constructs were then co-infiltrated in combinations of each chimeric Pikm-1 in tandem with either Pikm-2 or Pikp-2, to explore the response to co-infiltration with AVR-Mgk1, AVR-PikD, or an empty vector control (Figure 5-36).

Co-expression of chimeric Pikm-1^{HPP02}, Pikm-1^{HPP03}, or Pikm-1^{HPP04} with Pikm-2 produced effector independent cell death. This is in contrast to the Pikm-1 + Pikm-2 control, which displays no cell death in the absence of effectors (Figure 5-36 A, B). The auto-activity of the chimeric Pikm-1^{HPP} receptors in the presence of Pikm-2 means that any cell death observed when these receptors are co-infiltrated with the effectors cannot be attributed to a response to the effector.

The combination of Pikm-1 with Pikp-2 is known to reduce the intensity of cell death responses and previously has been successfully used to dampen auto-activity induced by NLR engineering of Pikm-1 to recognise an expanded effector repertoire (Bentham et al., 2023). However, these data demonstrate that the auto-activity induced by the chimeric Pikm-1^{HPP02}, Pikm-1^{HPP03}, or Pikm-1^{HPP04} with Pikm-2 was not mitigated by combination with Pikp-2 and auto-activity was still observed (Figure 5-36 C,D). Ultimately, due to time constraints, these engineering attempts were not pursued further.

Chapter 5: Interactions of novel *M. oryzae* effector AVR-Mgk1

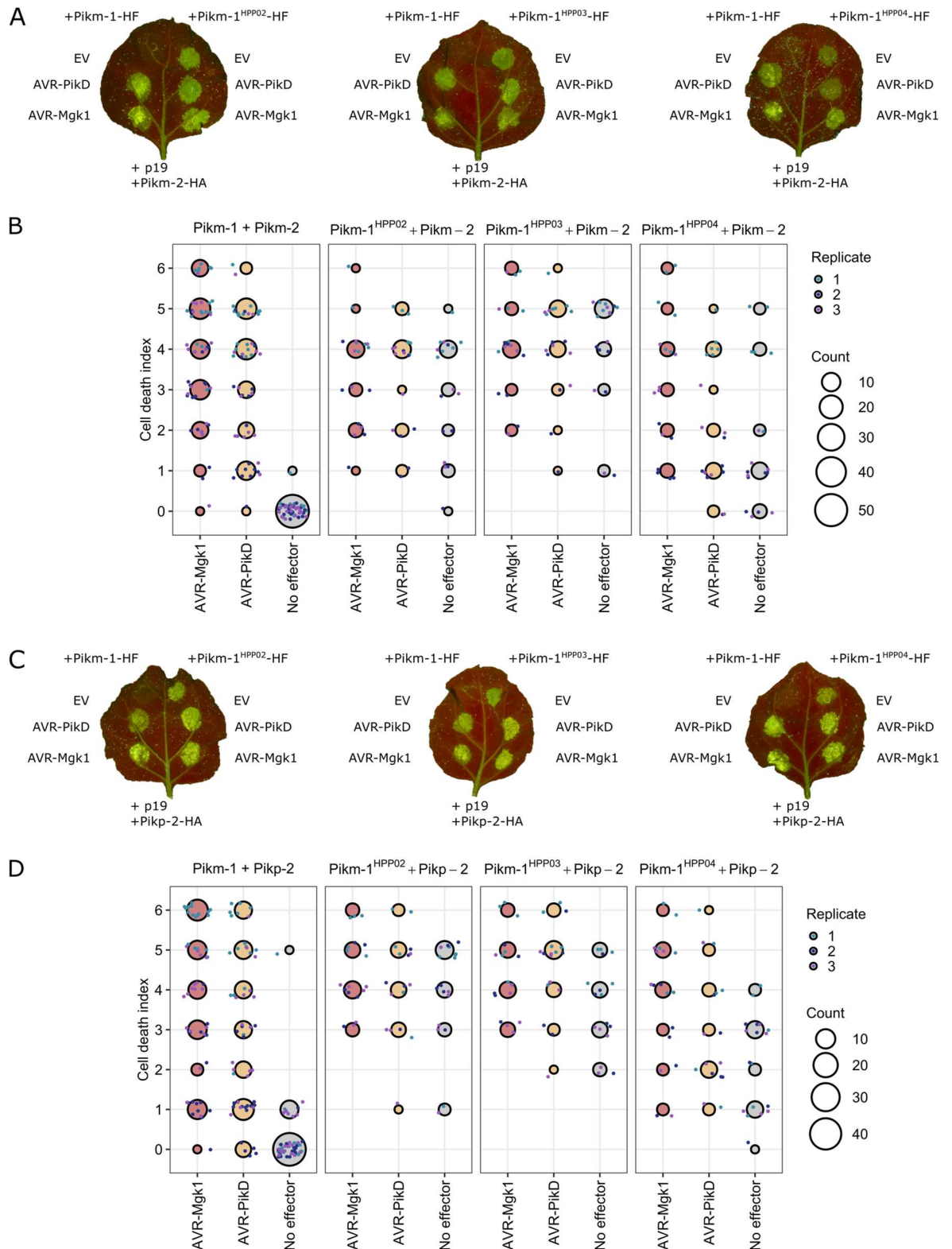


Figure 5-36: Chimeric Pik^{HPP} NLRs are auto-active in combination with either Pikm-2 or Pikp-2.

Cell death in *N. benthamiana* resulting from *Pikm*^{HPP02}, *Pikm*^{HPP03}, *Pikm*^{HPP04} chimeric NLRs in combination with *Pikm*-2 or *Pikp*-2, in response to AVR-Mgk1, AVR-PikD, or no effector. **A), B)**, Chimeric *Pikm*-1^{HPP} + *Pikm*-2 **C), D)**, Chimeric *Pikm*-1^{HPP} + *Pikp*-2 **A), C)**, UV photographs of representative leaves **B), D)**, Cell death dot plot of results obtained from a minimum of 18 leaves across three independent replicates for each.

5.4 Discussion

5.4.1 Crystal structures of Piks-HMA and the OsHPPo2/AVR-Mgk1 complex provide experimental validation for predicted protein models

In this chapter the crystal structures of both Piks-HMA and the complex between AVR-Mgk1^{T81}/OsHPPo2^{E74} were presented. Each structure provides experimental confirmation to my and others' previous structural prediction and modelling work. The crystal structure of Piks-HMA presented in this chapter is in good agreement with the previously obtained experimental structures of Pikm-HMA. Accordingly, the Piks-HMA crystal structure is also in good agreement with the modelled Piks-HMA structure produced through substitution of Pikm-HMA with the two polymorphic amino acids Q229E and V261A. The modest differences seen between experimental structure and predicted structures can be ascribed to several reasons. First, the Pikm-1/AVR-PikD complexes which forms the basis of some of the structural predictions was crystallised under different conditions to the Piks-HMA, in a different space group. Second, Piks-HMA was not part of a complex, which may have otherwise helped to stabilise N- and C-terminal residues into a fixed conformation within the crystal lattice, allowing them to be better resolved. Finally, the Pikm-HMA crystal structure used for modelling purposes was obtained as part of a complex with AVR-PikD, and therefore the side chains of the residues involved in the interaction, including Q229 and V261, may have changed conformation slightly as they accommodate AVR-PikD.

5.4.2 Crystal structure of OsHPPo2/AVR-Mgk1 reveals a novel interface between a HMA domain and a MAX effector

After experiencing difficulty in crystallising AVR-Mgk1 with the HMA domain from its cognate Piks NLR, the opportunity arose to instead co-crystallise AVR-Mgk1 with an sHMA protein. The AVR-Mgk1^{T81}/OsHPPo2^{E74} crystal structure fulfils the original objective of determining if AVR-Mgk1 shares the MAX effector fold with the AVR-Pik effectors; AVR-Mgk1 was observed to adopt a clearly defined MAX fold with one disulfide bond.

The binding position observed on the surface of the OsHPPo2-HMA is intermediate between the “Pik-interface” observed between Pikm-HMA/AVR-PikA and the “Pia-interface” exemplified by the interaction between Pikp-HMA/AVR-Pia (De La Concepcion et al., 2018; Varden et al., 2019). A continuous antiparallel beta-sheet running across the interface between HMA and MAX effector is a common feature of these complexes.

However, each binding interface displays a different topology; OsHPPo2 $\beta 2$ interacts with AVR-Mgk1 $\beta 3$, whereas Pikip-HMA $\beta 2$ interacts with AVR-Pia $\beta 2$, and Pikm-HMA $\beta 4$ interacts with AVR-PikD $\beta 3$. The overall effect is that in contrast to the orientation of the Pikm-HMA/AVR-PikD and Pikip-HMA/AVR-Pia structures where the $\beta 1$ strand of HMA and effector are oriented in parallel, AVR-Mgk1 is rotated 180 ° on the surface of the HMA so that the $\beta 1$ strand in OsHPPo2 and AVR-Mgk1 run antiparallel to each other (Figure 5-37).

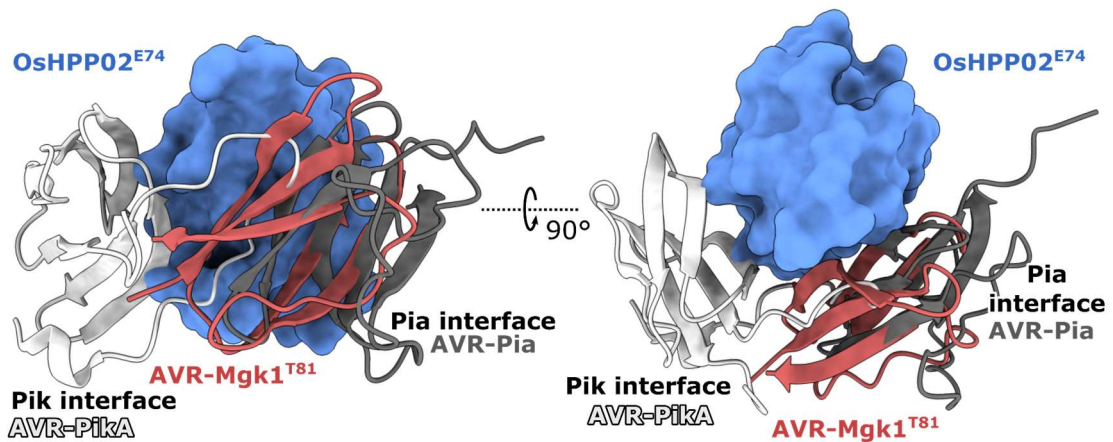


Figure 5-37: AVR-Mgk1 binds to OsHPP02 in a novel orientation

Side and top view of an overlay of the OsHPP02^{E74}/AVR-Mgk1^{T81} structure with previously published Pikm-HMA/AVR-PikA (PDB-ID: 6FUD) and Pikip-HMA/AVR-Pia (PDB-ID: 6Q76) structures which exemplify the “Pik-interface” and “Pia-interface” respectively (De La Concepcion et al., 2018; Varden et al., 2019). OsHPP02^{E74} is displayed in blue surface representation, and the effectors are in ribbon representation with colours: AVR-Mgk1^{T81} = red, AVR-PikA = white, and AVR-Pia = dark grey. The structures were aligned with respect to the HMA only, using “matchmaker” function in ChimeraX. For clarity, only the OsHPP02 HMA is displayed.

Furthermore, the complex provides an additional example of the diversity of MAX effector binding to HMA domain containing proteins, including small HMAs that are not integrated into immune receptors and are presumably host targets. As one such example, the OsHPPo2/AVR-Mgk1 structure sits alongside other sHMA/MAX effector structures, OsHIPP19/AVR-PikF and OsHIPP43/ PWL2 (Maidment et al., 2021; Zdrzałek et al., 2024) (Figure 5-38). The interface between OsHPPo2/AVR-Mgk1 appears very similar to the central interface between OsHIPP43/PWL2, however the OsHIPP43 $\beta 2$ interacts with PWL2 $\beta 2$, in contrast to $\beta 2$ of OsHPPo2 which interacts with AVR-Mgk1 $\beta 3$. Again, the overall effect is that AVR-Mgk1 is rotated 180 ° relative to PWL2 (Figure 5-38).

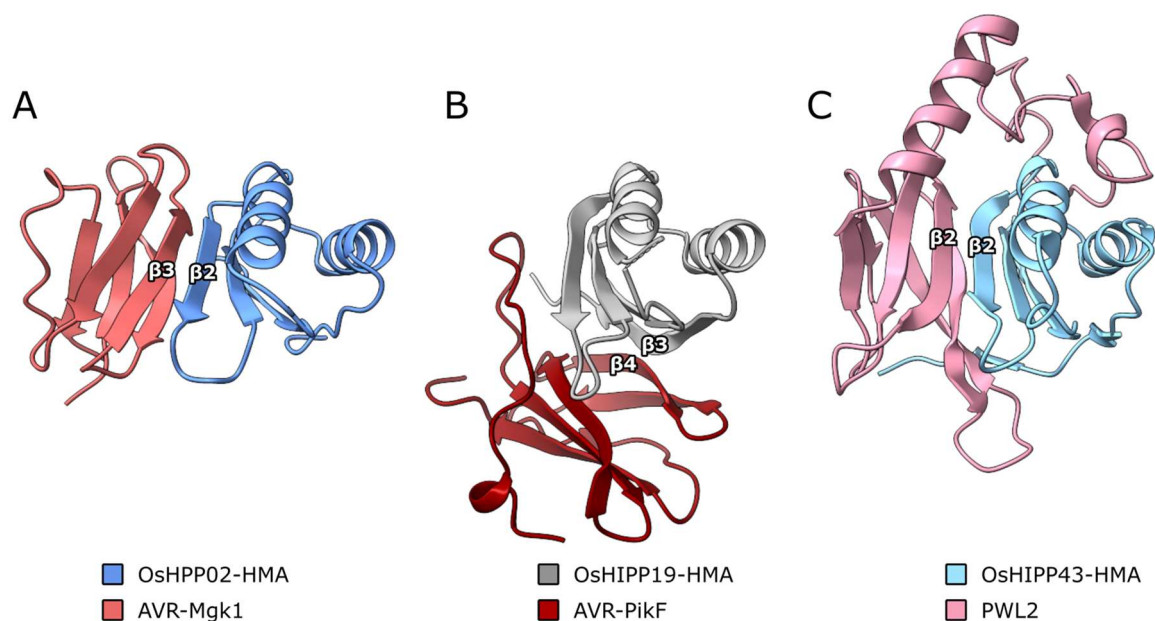


Figure 5-38: MAX effectors target small HMAs via different interfaces

A) *OsHPP02* β2 interacts with *AVR-Mgk1* β3. **B)** *OsHIPP19* β2 interacts with *AVR-PikF* β4 (PDB-ID: 7B1I) (Maidment et al., 2021). **C)** *OsHIPP43* β2 interacts with *PWL2* β2 (PDB-ID: 8R7A) (Zdrzałek et al., 2024). The structures were aligned with respect to the HMA only, using “matchmaker” function in ChimeraX, to highlight the relative orientations of the binding interfaces.

5.4.3 Structural prediction is a valuable tool in experimental structural biology

At the time that the project commenced, RoseTTAfold was the best performing protein structure prediction large language model. Shortly afterwards, AlphaFold2 was made publicly available. Although the performance of AlphaFold2 was superior in the CASP14 assessment, initially it did not perform well at AVR-Mgk1 prediction, producing extended and unstructured domains that appeared implausible, despite a high prediction confidence score. However, within a few months’ prediction plausibility had improved greatly and a recognisable MAX effector fold was reliably produced. Therefore, from this point on, AlphaFold2 was used for structural prediction tasks. Capabilities continued to rapidly progress with complex prediction officially added to AlphaFold2 in late 2021 (Evans et al., 2022).

As part of the process of producing the experimental structures presented in this thesis, protein prediction models were used in two ways. First, predicted structures were used to display regions of anticipated flexibility (Figure 5-14). Full length *OsHPP02*, *OsHPP03*, and *OsHPP04* predictions reported low confidence scores for the structure of the C-terminal

tail, indicating likely flexibility, and supporting our hypothesis that inclusion of the C-terminus would likely be incompatible with crystallisation. Later, a predicted structure of AVR-Mgk1/OsHPPo2^{K77} was produced and examined for regions of potential flexibility at the C-termini (Figure 5-22), which then formed the basis for further truncations of the proteins, and ultimately resulted in successful crystallisation, X-ray diffraction, and structure determination of the complex.

The second use of structural prediction in this project was for molecular replacement during structure determination. Phasing using computationally predicted protein structures is less expensive and faster than experimental phasing through methods such as single or multiple anomalous diffraction, which requires inclusion of a few heavy atoms. A predicted model of AVR-Mgk1^{T81} was successfully used for molecular replacement. The prediction of AVR-Mgk1 alone was used in order avoid introducing bias into the refinement by providing a model with a predicted interface with OsHPPo2.

Ultimately, the experimental AVR-Mgk1^{T81}/OsHPPo2^{E74} structure also aligned well with the predicted AVR-Mgk1/OsHPPo2^{K77} structure produced by AlphaFold2 (Figure 5-39 B). The binding position of AVR-Mgk1 relative to the HMA was very well predicted; the RMSD between the 65 atom pairs common to both models is 0.602 Å. The hydrogen bonding between the continuous antiparallel β -sheet between AVR-Mgk1 and OsHPPo2 was well predicted, and the residues contributing to the network of hydrogen bonding between the Gln38 of AVR-Mgk1 and Lys3 and Asp32 of OsHPPo2^{E74} at the top of the complex were also reasonably well predicted (Figure 5-39). However, differences in side chain positions illustrate some biases of structural predictions. The AlphaFold2 output reports the most likely residue position for each residue, resulting in flexible regions such as termini and lysine residues being predicted in full, often making additional contacts across the complex interface. However, in reality, these flexible regions may occupy a large range of possible positions, and therefore cannot always be well resolved by crystallography. The absence of such residues from a crystal structure could be viewed as a better reflect the true state of these flexible residues in solution rather than fixed in one position as in a structural prediction. One such example is the predicted interactions of the C-termini of both proteins; the AVR-Mgk1 C-terminus was predicted to interact with OsHPPo2 via Lys76 at the C-terminus and Lys5 on β 1. However, the AVR-Mgk1/OsHPPo2^{K77} complex could not be crystallised until the C-termini were truncated, suggesting that although AF2 positioned the termini interacting with each other, in solution these regions had enough flexibility to perturb crystallisation (Figure 5-39).

In the publication by my collaborator Dr Yu Sugihara describing his discovery of AVR-Mgk1, structural prediction was employed to designate AVR-Mgk1 as a predicted MAX effector

(Sugihara et al., 2023). The crystal structure obtained in this work was shared and provided experimental validation to this. It is interesting to note that despite low sequence identity with other known fungal effectors, structural prediction LLMs are able to make an accurate prediction of AVR-Mgk1 alone and in complex.

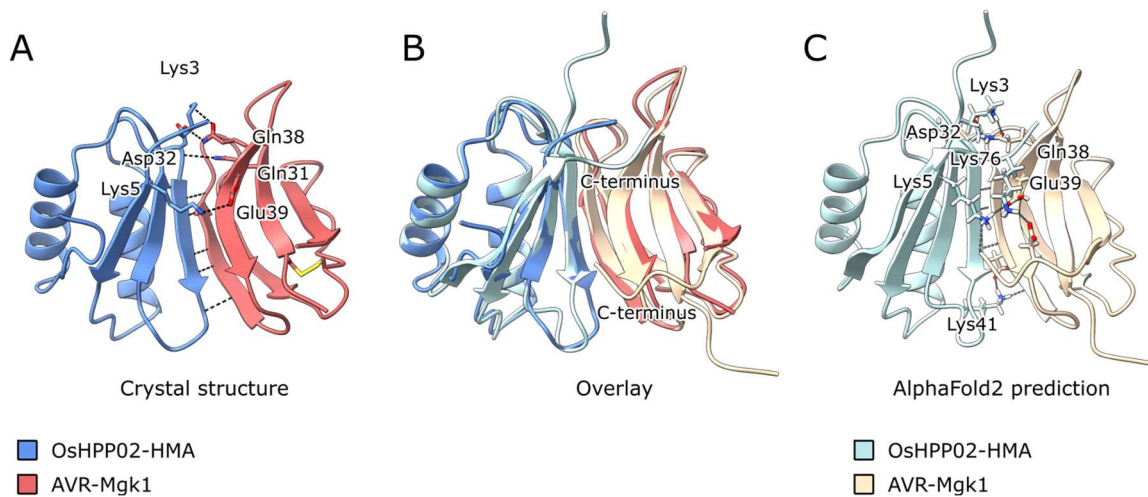


Figure 5-39: The crystal structure and AlphaFold2 prediction of AVR-Mgk1/OsHPP02 are in good agreement.

A) Crystal structure of AVR-Mgk1^{T81}/OsHPP02^{E74}, B) overlay of experimental and predicted structures, C) AlphaFold2 prediction of AVR-Mgk1/OsHPP02^{K77}.

5.4.4 Pikm^{OsHPP} chimeric receptors provide further insight into engineering the Pik chassis

Nicotiana benthamiana cell death assays were carried out to provide further support for use of the Pik-chassis as platform for engineering resistance against diverse effectors, in this case by complete substitution of the HMA domain. Substituting the HMA domain of Pikm-1 with the HMA domain of OsHPP02/03/04, was expected to produce a novel AVR-Mgk1 resistance in Pikm-1 when co-expressed with helpers Pikm-2 or Pikip-2. Unfortunately, effector-independent cell death was observed .

The Pik chassis has been successfully used to produce engineered receptors Pikm-1^{RGA5}, Pikip-1^{OsHIPP19}, and Pikm-1^{OsHIPP43} by incorporation of the RGA5, OsHIPP19, or OsHIPP43 HMA domains, conferring recognition of previously unrecognised effectors (Bentham et al., 2023; Maidment et al., 2023; Zdrzałek et al., 2024). In some of these cases, such as that of Pikm-1^{OsHIPP43}, co-expression with the mismatched helper allele Pikip-2 was required to overcome autoactivity (Zdrzałek et al., 2024). However, the chimeric Pikm^{OsHPP} receptors

produced in this study displayed effector independent cell death, regardless of their co-expression with Pikm-2 or mismatched Pikp-2 helper NLRs.

Autoactivity in engineered Pikp-1^{OsHIPP19} was overcome through use of P-loop mutants in the NB-ARC domain (Maidment et al., 2023). P-loop mutations were previously demonstrated to abolish effector-dependent cell death in Pikp-1, but in the case of Pikp-1^{OsHIPP19}, a K296R P-loop mutant retained some capacity to cause effector-dependent cell death while autoactive, effector-independent cell death was lost (De La Concepcion et al., 2018; Maidment et al., 2023). In the same work, Maidment et al. demonstrated the potential of using chimeric HMA domains to control autoactivity when incorporating domains which have not co-evolved with the NLR. By mutating seven amino acids of the MxCxxC metal binding motif in the β 1- α 1 loop of the Pikp-1-HMA, to revert them from the engineered OsHIPP19 sequence back to the wild-type Pikp-1 sequence, autoactivity was abolished, an approach first described by Bialas et al. (Bialas et al., 2021; Maidment et al., 2023). Perhaps use of the P-loop mutant and/or the metal binding loop chimera to produce Pikp-1^{OsHPP02/03/04}/Pikp-2 would mitigate the effector-independent cell death observed in the Pikm-1^{OsHPP02/03/04}.

In work by Bentham et al., the importance of maintaining the same size of the HMA after engineering was emphasised, as autoactivity was observed after deletion of a single amino acid from the inserted HMA domain, Δ G186 in the Pikm-1/Pikm-2 context (Bentham et al., 2023). In the work presented in this thesis, the differences in the N-terminus were not accounted for and so as a result the Pikm-1^{HPP02-K77} and Pikm-1^{HPP04-K77} constructs were four amino acids shorter than the wild-type Pikm-1-HMA which is 81 amino acids long. The Pikm-1^{HPP03-K81} construct was produced with an additional four amino acids (EEKK) at the C-terminus compared to Pikm-1^{HPP02-K77} and Pikm-1^{HPP04-K77}. The inserted OsHPP03 HMA is the same length as wildtype Pikm-1-HMA, although the domain boundaries differ when aligned; regardless, effector-independent cell death was observed in all three cases.

The OsHPP02/03/04-HMA is smaller than other exogenous domains which have been successfully substituted into Pik chassis in the Pikm-1/Pikp-2 combination, such as nanobody LaM-4 at 144 amino acids long (Bentham et al., 2023; Kourelis et al., 2023). Since the size and dimensions of the inserted OsHPP HMA domains are very similar to those of the wildtype Pikm-1-HMA, it seems highly unlikely that difference in size is the sole reason for auto-activity observed in the Pikm^{HPP02/03/04} chimeric NLRs.

Taking these prior results into account, I consider it most likely that the effector-independent cell death of Pikm-1^{OsHPP02/03/04} could be attributed to the HMA domain boundaries used. Therefore, future work should focus on careful design and screening of

different OsHPPo2/o3/o4 HMA boundaries. Structural prediction could be used to ensure that the boundaries do not impinge upon secondary structure elements of the source HMA. An alternative approach would be to perform modification of the AVR-Mgk1 binding interface only, transferring OsHPPo2 residues responsible for AVR-Mgk1 binding across to the Pikm-1-HMA at the equivalent structural position. This more sensitive approach could avoid disturbing other interactions that the HMA domain may be making with other domains of Pikm-1 or Pikm-2/Pikp-2.

5.4.5 AVR-Mgk variants are found in *Magnaporthe* isolates infecting diverse grass species

Although an interesting case study in producing novel recognition capacity in the Pik chassis through substitution of an entire HMA domain, unfortunately engineering AVR-Mgk1 recognition is of limited use agriculturally, for two reasons. First, existing resistance is provided by *Piks-1* and *Pikm-1*, and second, *AVR-Mgk1* itself is narrowly distributed across strains of *Magnaporthe* (Sugihara et al., 2023).

To further investigate the relative evolutionary relationships of the AVR-Mgk1 effector, Dr Yu Sugihara identified 16 AVR-Mgk sequences across diverse *Magnaporthe* lineages and calculated pairwise protein sequence identity (Sugihara et al., 2023). Among the AVR-Mgk effectors found in *Magnaporthe* isolates that infect rice, *AVR-Mgk3* has the highest sequence identity to *AVR-Mgk1* (51%). However, there are other AVR-Mgk effectors found within *M. oryzae* species which have unknown hosts or that infect other grasses such as *Urochloa mutica*. To visualise AVR-Mgk effector distribution across *Magnaporthe* lineages, Dr Yu Sugihara then produced a phylogeny of 155 *Magnaporthe* lineages displaying the presence and identity of *AVR-Mgk* alleles (Sugihara et al., 2023).

In contrast to the narrowly distributed *AVR-Mgk1*, *AVR-Mgk3* is found across a broader range of *Magnaporthe* strains, including the majority of those strains infecting rice. This makes *AVR-Mgk3* an attractive target to engineer recognition and resistance against. The retention of *AVR-Mgk3* across many *Magnaporthe* strains could be taken to imply that it carries out a significant activity during infection and therefore cannot be lost from the *Magnaporthe* genome without incurring a fitness penalty.

The structural and biophysical understanding of AVR-Mgk1 recognition by OsHPPo2-HMA presented in this thesis offers potential as a platform for engineering a HMA that can recognise AVR-Mgk3, assuming that auto-activity in the Pik chassis could be overcome. However, even without auto-activity of the chimeric receptor, this will be challenging

because of the low sequence similarity between *AVR-Mgk1* and *AVR-Mgk3* (51% sequence similarity). Therefore, it is likely that extensive modification of the OsHPP02-HMA would be required to convert it from an AVR-Mgk1 binder to an AVR-Mgk3 binder.

Finally, one major reason why *AVR-Mgk3* may not be a suitable candidate for engineering novel recognition of *Magnaporthe* strains is that it is likely to not be expressed as protein. Strand specific RNA-seq data suggests that *AVR-Mgk3* may be non-coding RNA (Dr Yu Sugihara, personal communication). This status as non-coding RNA could be confirmed with techniques such as 3' RACE-PCR. The wide distribution of *AVR-Mgk3* could therefore represent an unexpressed transcript with no selection pressure acting upon it, in contrast to it being an indispensable effector protein which is retained due to a fitness benefit it confers. Ultimately, intracellular immunity mediated by NLRs relies upon perception of effector proteins. Therefore, if *AVR-Mgk3* is not expressed it cannot be a suitable target for engineering immunity against *Magnaporthe*.

5.4.6 Conclusion

In this chapter, my objective was to explore the structure and interactions of the novel effector AVR-Mgk1. I determined the structure of AVR-Mgk1, confirming it as a MAX effector despite its low sequence similarity with AVR-Pik effectors. Biophysical analysis revealed binding affinities of AVR-Mgk1 in response to integrated and small HMAs, in comparison to the responses of AVR-PikD to the same HMA panel. Preliminary attempts at NLR engineering were made and although effector-independent cell death resulted, future avenues to explore to mitigate this have been considered.

Chapter 6: General Discussion

Awareness of plant immunity traits is critical when designing crop cultivars which can maintain high performance in the future. Global trade accelerates the circulation of pathogens between countries, and in the face of global climate change, the geographical ranges of pathogens are changing. As pathogens continue to evolve and change their effector repertoires, crop breeding must remain ahead of these changes. Furthermore, pathogens may make host jumps to infect new hosts, bringing the potential need to produce a new cultivar with multiple resistances not found in current breeding lines. Biotechnology will be key in responding to this, through production of resistance genes tailored to provide an appropriate and effective immune response to novel pathogen effectors. For this reason, it is important to understand both how effectors target host proteins, and how effectors are recognised by the wild-type immune receptors of crop plants.

The overall aim of this thesis has been to better understand how blast fungus effectors interact with the paired rice NLR Pik, and with small HMAs. To address this, I pursued a structural approach, employing eukaryotic protein expression systems to attempt full length Pik resistosome purification, and *E. coli* expression of individual HMA domains and X-ray crystallography of HMA domains in complex with the novel *Magnaporthe* effector AVR-Mgk1. Both projects offered potential to provide insights which could contribute towards the engineering of NLRs, and my results reveal some novel features of these interactions. Here, I outline the main questions I sought to address during my PhD, and for each, I provide a summary of the data presented in this thesis and discuss the implications of my findings.

6.1 Structural mechanisms of the Pik resistosome remain unresolved

A signature of intracellular plant immunity is the formation of oligomeric resistosome complexes upon NLR activation. At the time of commencing this PhD work, one cryo-EM structure, of the ZAR1 resistosome, had been published, shortly followed by the structures of TIR-NLR resistosomes of ROQ1 and RPPS1 (Wang et al., 2019a; Wang et al., 2019b; Ma et al., 2020; Martin et al., 2020). Since then, many cryo-EM structures of resistosomes have been published, revealing conserved details of NLR activation and effector recognition (Förderer et al., 2022; Lawson et al., 2024; Liu et al., 2024; Ma et al., 2024; Madhuprakash

et al., 2024). However, there has still not been a publication revealing the structure of a paired NLR resistosome, leaving us without structural detail of how NLRs such as the Pik pair of rice are activated. In addition, Pik-1 contains an integrated HMA domain located between the CC- and NB-ARC domains, through which *Magnaporthe* effectors are recognised (Maqbool et al., 2015). To date, there are no full-length structures of an NLR with an integrated domain at this location, and so the precise details of how effector binding to the integrated domain results in immune activation are unknown. For these reasons, I pursued expression and purification of full length Pikm-1/Pikm-2 protein, to subject to cryo-electron microscopy, presented in Chapter 3.

I determined that expression of the Pik pair by agroinfiltration in *N. benthamiana* was not improved by use of a genomic expression construct containing introns, nor was expression consistently improved through use of a proprietary viral over-expression vector system which we accessed on a fee-for-service basis (Figure 3-5, Figure 3-7, Figure 3-8). To attempt expression in a different eukaryotic system, I re-established Sf9 insect cell culture in the Banfield Lab. I obtained preliminary western blot evidence of expression (Figure 3-11), but was unable to scale this up and attempt purification as unfortunately the Sf9 insect cell culture suffered from frequent contaminations and I made the decision to prioritise other elements of my doctoral work. However, an opportunity to continue this work then arose through a collaboration with Dr Nitika Mukhi. Through significant effort including use of a different cell line and culture techniques, and many expression trials, she was able to express and purify full length Pikm-1 and Pikm-2 in Sf21 cells. As a visitor to the Schulze-Lefert Lab in June 2023, together with Dr Mukhi I participated in further purification of an active 6His-SUMO-Pikm-1/Pikm-2-FLAG/AVR-PikD-TwinStrep complex and an inactive 6His-SUMO-Pikm-1/Pikm-2-FLAG complex (Figure 3-14 and Figure 3-15). We examined both purified complexes by negative stain EM and undertook a preliminary cryo-EM sample screening. Although sample preparation still requires optimisation for cryo-EM, these are significant advances and suggest that a structure of a Pik pair resistosome may be achievable.

In addition to pursuing Pik NLR protein purification, I also probed the cellular localisation and oligomeric status of the Pik pair using confocal microscopy and BN-PAGE in Chapter 4. CC-NLRs such as ZAR1 and Sr35 have been observed at the plasma membrane acting as a Ca²⁺ ion channel, but it is not yet established if this is also the case for the Pik pair (Bi et al., 2021; Förderer et al., 2022). At the time of writing, a report of a CC_R-NLR localising to chloroplast, endoplasmic reticulum, and mitochondrial membranes has been released, highlighting the potential diversity in localisation of activated NLRs, and emphasising the importance of determining the localisation of the Pik pair (Ibrahim et al., 2024). Confocal

microscopy revealed that Pikm-1-mCherry localised to the cell and nuclear periphery, and within the nucleus, whereas Pikm-2-mEGFP was distributed at the cell and nuclear periphery, but not within the nucleus (Figure 4-2). In contrast to this, CFP-AVR-PikD produced a strong nuclear fluorescence signal and also appeared at the cell periphery, suggesting a cytoplasmic distribution (Figure 4-2). This establishes the Pik NLRs and AVR-PikD as amenable to fluorescence microscopy, and future work should focus on mitigating cell death when all three proteins are co-expressed, without perturbing protein surfaces implicated in membrane association. BN-PAGE was performed and broad bands representing a range of high molecular weight species were observed. Although the migration of the band did not change upon addition of AVR-PikD, the band intensity appeared greater, suggesting accumulation of a Pik complex (Figure 4-6). However, both the confocal microscopy and BN-PAGE experiments must be repeated to establish the reproducibility of these preliminary observations. These experiments must also be optimised further including through use of markers for the confocal microscopy and incorporation of an IP step before the BN-PAGE.

6.2 The AVR-Mgk1/OsHPPo2 crystal structure confirmed AVR-Mgk1 as a MAX effector and revealed a new mode of HMA binding

Novel *Magnaporthe* effector AVR-Mgk1 was predicted to adopt a MAX effector fold based on sequence features, but due to its low sequence similarity with AVR-Pik effectors, and with no structural evidence, this assignment was not certain. Therefore, in Chapter 5 I set out to determine the structure of AVR-Mgk1 in complex with Piks-1-HMA. I expressed and purified individual Piks-HMA and AVR-Mgk1 protein from *E. coli* and performed analytical gel filtration demonstrating the interaction between the two proteins (Figure 5-4, Figure 5-5). I then set out to purify the complex at a larger scale, and obtained protein which I used to carry out crystallisation screens. I obtained a crystal which diffracted and the structure of Piks-HMA alone was determined (Figure 5-9). I then turned to crystallising AVR-Mgk1 with HMA domains of small HMAs OsHPPo2, OsHPPo3, or OsHPPo4. I was able to demonstrate binding between AVR-Mgk1 and OsHPPo2^{K77} by analytical gel filtration, and therefore proceeded to purify these two proteins in complex (Figure 5-18). To ensure complex formation, I expressed and purified each protein individually, then mixed them before a final gel filtration step (Figure 5-11). To optimise crystallisation, truncated constructs AVR-Mgk1^{T81} and OsHPPo2^{E74} were designed, guided by predicted structures of the complex (Figure 5-22). Further crystallisation trials yielded crystals which diffracted X-

rays, and the structure of the AVR-Mgk1^{T81}/OsHPPo2^{E74} complex was determined, confirming AVR-Mgk1 as a MAX effector (Figure 5-29).

The position and orientation of AVR-Mgk1 on the surface of OsHPPo2 was novel in comparison to the Pikm-HMA/AVR-PikD and Pikp-HMA/AVR-Pia binding interfaces, and the orientation of AVR-Mgk1 was rotated by 180 ° relative to AVR-Pik or AVR-Pia, due to the interaction between OsHPPo2 β 2 and AVR-Mgk1 β 3. The interface between AVR-Mgk1^{T81}/OsHPPo2^{E74} is largely composed of an extended antiparallel β -sheet which runs across the two proteins, with three β -strands contributed by AVR-Mgk1^{T81} and two by OsHPPo2^{E74} (Figure 5-30). QtPISA analysis reveals a relatively small interaction interface of 555.1 Å², which is half the area of other sHMA/MAX effector complexes such as OsHIPP19-HMA/AVR-PikF (1045.3 Å) (Maidment et al., 2021). Mutational analysis to perturb the interaction between AVR-Mgk1^{T81}/OsHPPo2^{E74} would provide confirmation of the interaction interface. However, since a significant contributor to the interaction between the two proteins is the extensive hydrogen bonding stabilising the continuous β -sheet at the interface, designing OsHPPo2 mutations to abolish the interaction between the two proteins without breaking the β -strands might be challenging. Structural prediction tools could be employed to screen potential mutations, excluding those which are predicted to disrupt the overall structure of the HMA.

Subsequently, surface plasmon resonance was used to assess the binding affinities between AVR-Mgk1 and OsHPPo2^{K77} and compare these with other HMA-effector pairs. This revealed that the strength of interaction between AVR-Mgk1 and OsHPPo2, OsHPPo3, or OsHPPo4 is greater than the affinity between AVR-Mgk1 and Pikm-HMA or AVR-Mgk1 and Piks-HMA (Figure 5-34).

Finally, this new information on binding affinities inspired a proof-of-concept Pik NLR engineering attempt. In this case, the HMA domains from OsHPPo2, OsHPPo3, and OsHPPo4 were incorporated into Pikm-1 with the goal of enhancing the AVR-Mgk1 recognition capabilities of Pikm-1. This would provide further examples of integration of domains in place of the Pikm-1-HMA to produce novel recognition capacities. Expression of these chimeric Pikm-1^{OsHPP} NLRs was confirmed by western blot, but unfortunately when expressed in *N. benthamiana* by agroinfiltration, effector-independent cell death was observed (Figure 5-36). The autoactive phenotype was not relieved by sensor/helper allelic mismatching, which has been helpful in other cases, such as the engineering of the Pikm-1^{RGA5} and Pikm-1^{OsHIPP43} (De La Concepcion et al., 2018; Bentham et al., 2023; Zdrzałek et al., 2024). Other approaches to mitigate autoactivity such as mutation of the P-loop or substitution of wild-type metal binding motif in the α 1- β 1 loop might rescue these integrations. Furthermore, it is possible that post-translational modifications are

modulating the activities of the Pik resistosome, and so future discoveries about the conformation and activation of the Pik resistosome may reveal additional reasons for why the Pikm-1^{HPP02/03/04} integrations produced effector-independent cell death. This is another example of how structural and mechanistic knowledge can benefit the objective of engineering NLRs, including those which operate as a paired NLR and those with integrated domains.

6.3 Summary and outlook

In this work I have used different heterologous expression systems to pursue expression and purification of the full length Pikm-1/Pikm-2 resistosome. Through collaboration, significant progress towards obtaining a cryo-EM structure has been made, presented in Chapter 3; this is work which is ongoing. Preliminary experiments to obtain information about cellular localisation and oligomeric status of the Pik pair were presented in Chapter 4. To understand the structure and interactions of AVR-Mgk1, in Chapter 5 I used in planta assays, X-ray crystallography, and biophysical techniques. I confirmed the identity of the *M. oryzae* effector AVR-Mgk1 as a MAX effector, documented its novel interface with OsHPP02, performed biophysical analysis of its interactions with integrated and small HMAs, and attempted proof-of-concept engineering of Pikm-1 based on these biophysical insights.

The modular nature of NLRs with IDs makes them an attractive target for engineering. The idea of a NLR chassis which can accept host target modules without any customised engineering and provide conversion of a binding event between an effector and a host target into an immune response is very appealing. This may not be possible in reality, as every NLR and domain to be integrated have distinct properties. The Pik chassis has this potential, and there are many examples of its successful engineering by domain exchanges and through more subtle interventions (De La Concepcion et al., 2018; Bentham et al., 2023; Maidment et al., 2023; Zdrzałek et al., 2024). However, we do not yet fully understand the physical constraints for engineering into the Pik chassis as we lack detailed structural information about the conformation of a Pik resistosome complex pre- and during activation. Therefore, there are also engineering cases where functionality was not as expected, such as the integration of the small HMAs OsHPP02, OsHPP03, and OsHPP04 into the Pikm-1 chassis as documented in this thesis. Some strategies to overcome autoactivity have been developed, but many of these are specific to the NLR and domain to be integrated, such as allelic mismatch of the Pik pair, or mutation of specific residues to tune autoactivity. For this reason, future efforts to engineer the Pik pair, other NLR pairs,

or NLRs with IDs would all benefit from insights obtained through elucidation and analysis of a Pik resistosome structure.

Recently, another class of intracellular immune receptors, the tandem kinases, have been coming into focus. Like NLRs, some tandem kinases contain integrated domains (Fahima et al., 2024). As more details of tandem kinase effector recognition and activation are determined, these receptors will become an increasingly attractive target for engineering. Generative AI capabilities are rapidly improving, including programmes such as RFdiffusion which can generate amino acid sequences of proteins with the ability to bind a specified protein surface (Watson et al., 2023). As structural and molecular biology of intracellular immune receptors continues to advance, rational design of novel integrated domains from all sources will be further enabled, including those produced through *de novo* methods. This may ultimately translate into novel recognition capabilities in the field, although agronomic deployment also depends on regulatory and consumer acceptance of GE and GM, which vary globally.

The Genetic Technology (Precision Breeding) Act 2023 paves the way for cultivation of genome edited crops in England. Meanwhile, the European Sustainable Agriculture Through Genome Editing (EU-SAGE) consortium maintains a global online database of genome-edited crops (<https://www.eu-sage.eu/genome-search>). At the time of writing, the database contains 955 edited traits reported in the peer-reviewed literature, of which 31.7 % are in rice (303 entries). Fifteen of these entries relate to resistance to blast disease, and while their inclusion in this database is no indication that these specific traits have entered or will enter agricultural use, it illustrates the broader trend in use of genome editing to enhance rice performance, including in immune traits that can confer improved blast disease resistance.

Accumulating multiple R genes into a single cultivar, each contributing a different range of resistances, can produce a wider spectrum of resistance. This reduces the need for fungicide application in the field, and also reduces the likelihood that a pathogen strain will be able to overcome resistance, thereby increasing resistance durability. By constantly evaluating the effector repertoires of prevalent and emerging strains of *M. oryzae*, aided by open science resources such as Open Rice Blast (<http://openriceblast.org/>), breeders can evaluate which R gene combinations are likely to be most effective. The process of accumulating multiple R genes into a single cultivar can be undertaken using conventional breeding techniques, but genome editing accelerates the breeding process and can produce results not easily achieved through traditional means.

Finally, returning to the longstanding arms race between crop plant and pathogen, it can be argued that the conflict is escalating. With an increasing global population, the stakes become higher – maintaining agricultural productivity is essential for human wellbeing, and yet this must be balanced with responsible environmental stewardship. Factors including global climate change, pathogen distribution through international trade, and natural mutation of pathogens expose crops to pathogens carrying ever-changing effector repertoires, perpetuating the arms race. Regardless, innovation in agronomy, crop breeding, and fundamental research is pushing the balance towards sustainable food production. Improved understanding of the mechanisms of plant immunity will continue to contribute towards the goal of modulating the plant immune system to secure future crop health.

Appendix

AlphaFold2 outputs

AlphaFold2 was used for protein structure prediction (Pereira et al., 2021; Mirdita et al., 2022). Query sequences are reproduced here in FASTA format. The following outputs are reproduced here:

Predicted Aligned Error (PAE) – A plot describing the expected difference between the predicted and true structures, if they were aligned at each residue in turn. Plotted on a scale from 0 to 30, where low scores, coloured in blue = low expected difference between the predicted and true structures, while high scores, coloured in red = high expected difference between the predicted and true structures. Five plots displayed, one for each of the five models produced per AlphaFold2 run.

Multiple sequence alignment (MSA) – AlphaFold2 queries protein sequence databases and produces a per-residue alignment of sequences similar to the query sequence, in order to infer co-evolutionary relationships. A deep MSA (hundreds to thousands of similar sequences) is more likely to result in a high quality prediction, whereas a shallow MSA (no or a low number of similar sequences) means that the prediction is made with much less prior information and so may be of lower accuracy.

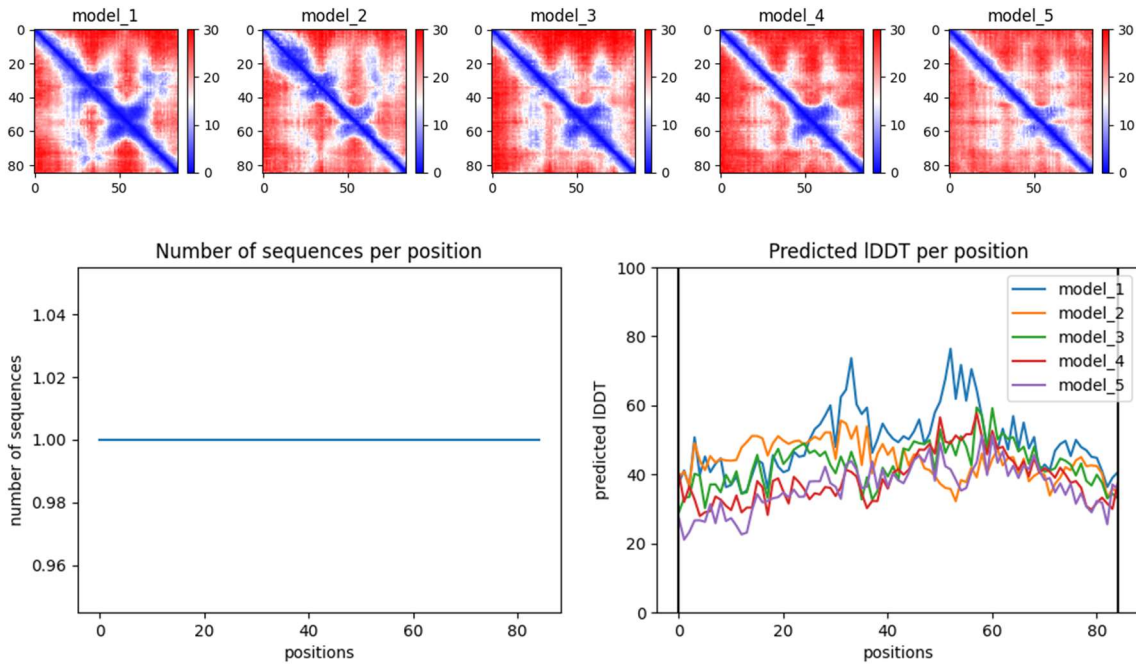
Local Distance Difference Test (pLDDT) – A per-residue confidence estimate, scored on a scale from 0-100. This value is stored in the B-factor column of structure files, but different to B-factors, low values correspond to low confidence while higher pLDDT indicates higher confidence that the predicted C α position is accurate.

Appendix

AVR-Mgk1

>AVR-Mgk1

MRNCRIWQDMGSVWQEYVVVTPPVTVDIITKRHGAFSLFVPVGC GIRD TG GALRATETDDPW

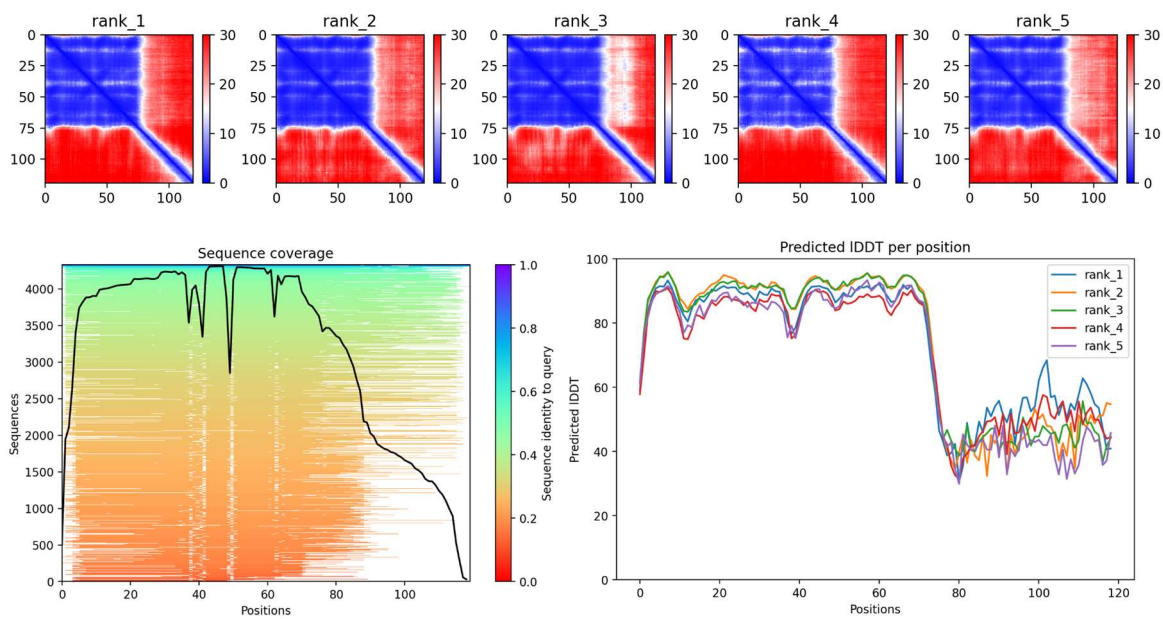


OsHPPo2 full length

>AVR-Mgk1

MAKQKIVIKMNASDKCRSKAMALVASTTGVD SVALAGDGKDQLVVVG DGVDSIELTTALRKKVGHATLMTVG

EDKKEKKPEPAVVEYPYPWNCYPYGYAPPPQHVVYQYPASSSWWW

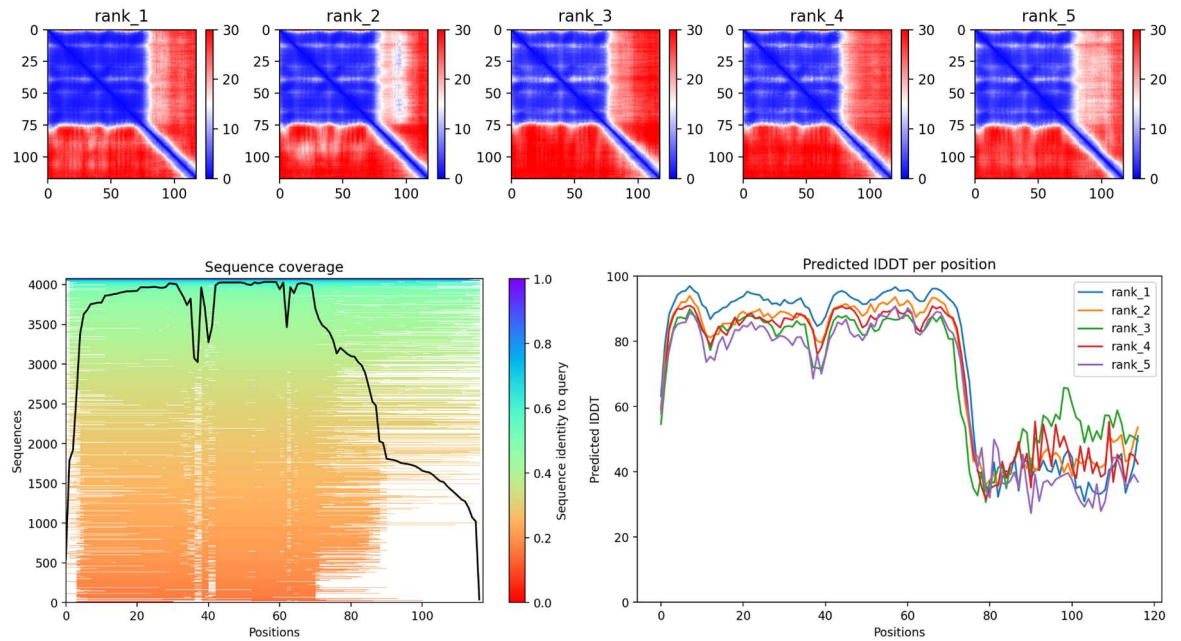


Appendix

OsHPPo3 full length

>OsHPP03

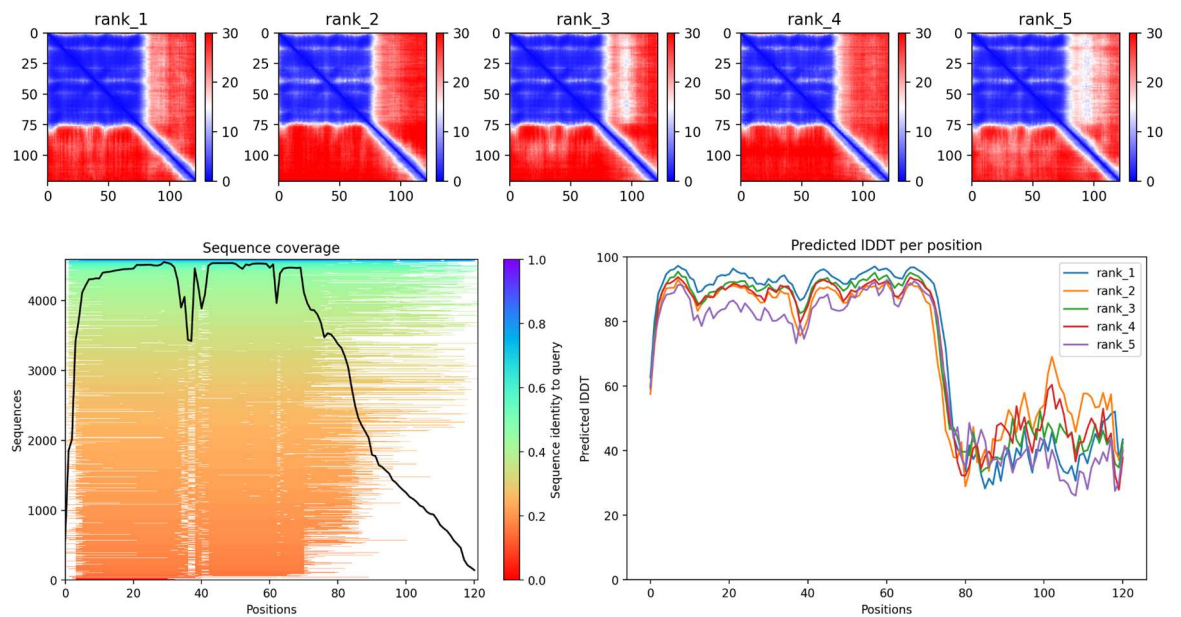
MAKQKIVIKVNMA SDKCRSKAMALVASTSGVDSVALAGDGKDQVVVVG DGVD SIKLT TALRKKVGHATLMTVG
EVKKEEKKPEHAVVEYPWSYHPYTFAPPAQHVVYQYPASSWSIF



OsHPPo4 full length

>OsHPP04

MAKQKIVIKVEMSCDKCRSKAMALVAATGGVDSVALAGDGKDQVVVVG DGVD SIKLTAALRKKVGHATLTVG
EVKKEEKKPEPAAA AVEYPWSYHPAYTYAPPAQHVFYQQYPASSPWWC



Appendix

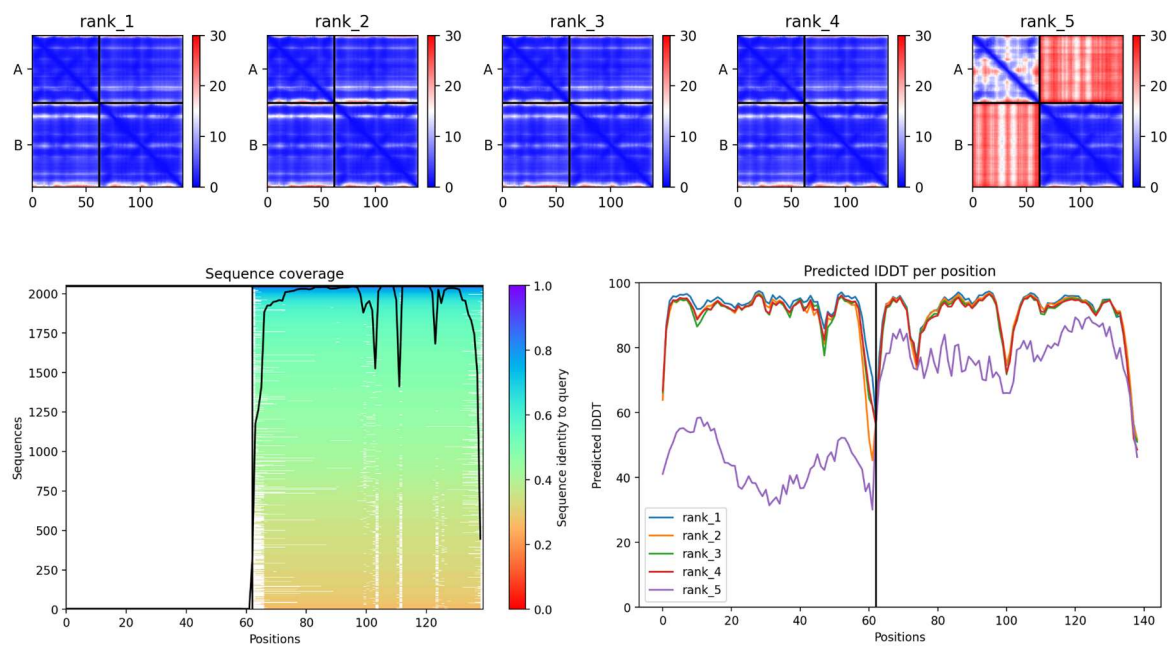
AVR-Mgk1/OsHPPo2^{K77} complex

>AVR-Mgk1

MRNCRIWQDMGSVWQEYVVVTPPVTVDIITKRHGAFSLFVPVGCGRDTGGALRATETDDPW

>OsHPP02

MAKQKIVIKMNMASDKCRSKAMALVASTTGVDSVALAGDGKDQLVVVGDSIELTTALRKKVGHATLMTVG
EDKK



References

- Adachi, H., Contreras, M.P., Harant, A., Wu, C.-H., Derevnina, L., Sakai, T., Duggan, C., Moratto, E., Bozkurt, T.O., Maqbool, A., *et al.* (2019). An N-terminal motif in NLR immune receptors is functionally conserved across distantly related plant species. *eLife* 8.
- Agirre, J., Atanasova, M., Bagdonas, H., Ballard, C.B., Basle, A., Beilstein-Edmands, J., Borges, R.J., Brown, D.G., Burgos-Marmol, J.J., Berrisford, J.M., *et al.* (2023). The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallographica Section D* 79, 449-461.
- Ahn, H.-K., Lin, X., Olave-Achury, A.C., Derevnina, L., Contreras, M.P., Kourelis, J., Wu, C.-H., Kamoun, S., and Jones, J.D.G. (2023). Effector-dependent activation and oligomerization of plant NRC class helper NLRs by sensor NLR immune receptors Rpi-amr3 and Rpi-amr1. *The EMBO Journal* 42, e111484.
- Andersson, K., Areskoug, D., and Hardenborg, E. (1999). Exploring buffer space for molecular interactions. *J Mol Recognit* 12, 310-315.
- Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.L., Gomez-Gomez, L., Boller, T., Ausubel, F.M., and Sheen, J. (2002). MAP kinase signalling cascade in Arabidopsis innate immunity. *Nature* 415, 977-983.
- Ashikawa, I., Hayashi, N., Yamane, H., Kanamori, H., Wu, J., Matsumoto, T., Ono, K., and 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.
- Askari, N., Correa, R.G., Zhai, D., and Reed, J.C. (2012). Expression, purification, and characterization of recombinant NOD1 (NLRC1): A NLR family member. *J Biotechnol* 157, 75-81.
- Asuke, S., Nishimi, S., and Tosa, Y. (2020). At Least Five Major Genes Are Involved in the Avirulence of an Eleusine Isolate of *Pyricularia oryzae* on Common Wheat. *Phytopathology*® 110, 465-471.
- Asuke, S., Umehara, Y., Inoue, Y., Vy, T.T.P., Iwakawa, M., Matsuoka, Y., Kato, K., and Tosa, Y. (2021). Origin and Dynamics of Rwt6, a Wheat Gene for Resistance to Nonadapted Pathotypes of *Pyricularia oryzae*. *Phytopathology*® 111, 2023-2029.
- Baek, M., Dimaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., *et al.* (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, eabj8754.
- Barragan, A.C., Latorre, S.M., Malmgren, A., Harant, A., Win, J., Sugihara, Y., Burbano, H.A., Kamoun, S., and Langner, T. (2024). Multiple Horizontal Mini-chromosome

References

- Transfers Drive Genome Evolution of Clonal Blast Fungus Lineages. *Mol Biol Evol* 41.
- Baudin, M., Hassan, J.A., Schreiber, K.J., and Lewis, J.D. (2017). Analysis of the ZAR1 Immune Complex Reveals Determinants for Immunity and Molecular Interactions. *Plant Physiol* 174, 2038-2053.
- Bentham, A.R., De la Concepcion, J.C., Benjumea, J.V., Kourelis, J., Jones, S., Mendel, M., Stubbs, J., Stevenson, C.E.M., Maidment, J.H.R., Youles, M., *et al.* (2023). Allelic compatibility in plant immune receptors facilitates engineering of new effector recognition specificities. *The Plant Cell* 35, 3809-3827.
- Bentham, A.R., De La Concepcion, J.C., Mukhi, N., Zdrzałek, R., Draeger, M., Gorenkin, D., Hughes, R.K., and Banfield, M.J. (2020). A molecular roadmap to the plant immune system. *J Biol Chem* 295, 14916-14935.
- Bentham, A.R., Youles, M., Mendel, M.N., Varden, F.A., Concepcion, J.C.D.I., and Banfield, M.J. (2021). pOPIN-GG: A resource for modular assembly in protein expression vectors. *bioRxiv*, 2021.2008.2010.455798.
- Bernoux, M., Ve, T., Williams, S., Warren, C., Hatters, D., Valkov, E., Zhang, X., Ellis, J.G., Kobe, B., and Dodds, P.N. (2011). Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. *Cell Host Microbe* 9, 200-211.
- Bi, G., Su, M., Li, N., Liang, Y., Dang, S., Xu, J., Hu, M., Wang, J., Zou, M., Deng, Y., *et al.* (2021). The ZAR1 resistosome is a calcium-permeable channel triggering plant immune signaling. *Cell* 184, 3528-3541.e3512.
- Białas, A., Langner, T., Harant, A., Contreras, M.P., Stevenson, C.E.M., Lawson, D.M., Sklenar, J., Kellner, R., Moscou, M.J., Terauchi, R., *et al.* (2021). Two NLR immune receptors acquired high-affinity binding to a fungal effector through convergent evolution of their integrated domain. *eLife* 10, e66961.
- Bosch, Peter J., Corrêa, Ivan R., Sonntag, Michael H., Ibach, J., Brunsveld, L., Kanger, Johannes S., and Subramaniam, V. (2014). Evaluation of Fluorophores to Label SNAP-Tag Fused Proteins for Multicolor Single-Molecule Tracking Microscopy in Live Cells. *Biophys J* 107, 803-814.
- Cai, Q., He, B., Weiberg, A., Buck, A.H., and Jin, H. (2019). Small RNAs and extracellular vesicles: New mechanisms of cross-species communication and innovative tools for disease control. *PLoS Pathog* 15, e1008090.
- Cai, Q., Qiao, L., Wang, M., He, B., Lin, F.-M., Palmquist, J., Huang, S.-D., and Jin, H. (2018). Plants send small RNAs in extracellular vesicles to fungal pathogen to silence virulence genes. *Science* 360, 1126-1129.

References

- Casey, L.W., Lavrencic, P., Benthall, A.R., Cesari, S., Ericsson, D.J., Croll, T., Turk, D., Anderson, P.A., Mark, A.E., Dodds, P.N., *et al.* (2016). The CC domain structure from the wheat stem rust resistance protein Sr33 challenges paradigms for dimerization in plant NLR proteins. *Proceedings of the National Academy of Sciences* *113*, 12856-12861.
- Castel, B., Ngou, P.M., Cevik, V., Redkar, A., Kim, D.S., Yang, Y., Ding, P., and Jones, J.D.G. (2019). Diverse NLR immune receptors activate defence via the RPW8-NLR NRG1. *New Phytol* *222*, 966-980.
- Cesari, S. (2018). Multiple strategies for pathogen perception by plant immune receptors. *New Phytol* *219*, 17-24.
- Cesari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y., Shimamoto, K., Dodds, P., Terauchi, R., and Kroj, T. (2014). The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. *The EMBO Journal* *33*, 1941-1959.
- Cesari, S., Xi, Y., Declerck, N., Chalvon, V., Mammri, L., Pugnière, M., Henriquet, C., De Guillen, K., Chochois, V., Padilla, A., *et al.* (2022). New recognition specificity in a plant immune receptor by molecular engineering of its integrated domain. *Nature Communications* *13*.
- Collemare, J., O'Connell, R., and Lebrun, M.-H. (2019). Nonproteinaceous effectors: the terra incognita of plant–fungal interactions. *New Phytol* *223*, 590-596.
- Contreras, M.P., Lüdke, D., Pai, H., Toghiani, A., and Kamoun, S. (2023a). NLR receptors in plant immunity: making sense of the alphabet soup. *EMBO Rep* *24*, e57495.
- Contreras, M.P., Pai, H., Selvaraj, M., Toghiani, A., Lawson, D.M., Tumtas, Y., Duggan, C., Yuen, E.L.H., Stevenson, C.E.M., Harant, A., *et al.* (2023b). Resurrection of plant disease resistance proteins via helper NLR bioengineering. *Science Advances* *9*, eadg3861.
- Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D* *62*, 1002-1011.
- De Abreu-Neto, J.B., Turchetto-Zolet, A.C., De Oliveira, L.F.V., Bodanese Zanettini, M.H., and Margis-Pinheiro, M. (2013). Heavy metal-associated isoprenylated plant protein (HIPP): characterization of a family of proteins exclusive to plants. *FEBS J* *280*, 1604-1616.
- de Guillen, K., Ortiz-Vallejo, D., Gracy, J., Fournier, E., Kroj, T., and Padilla, A. (2015). Structure Analysis Uncovers a Highly Diverse but Structurally Conserved Effector Family in Phytopathogenic Fungi. *PLoS Path* *11*, e1005228.

References

- De La Concepcion, J.C., Franceschetti, M., Maqbool, A., Saitoh, H., Terauchi, R., Kamoun, S., and Banfield, M.J. (2018). Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen. *Nature Plants* 4, 576-585.
- De la Concepcion, J.C., Langner, T., Fujisaki, K., Yan, X., Were, V., Lam, A.H.C., Saado, I., Brabham, H.J., Win, J., Yoshida, K., *et al.* (2024). Zinc-finger (ZiF) fold secreted effectors form a functionally diverse family across lineages of the blast fungus *Magnaporthe oryzae*. *PLoS Path* 20, e1012277.
- DeFalco, T.A., and Zipfel, C. (2021). Molecular mechanisms of early plant pattern-triggered immune signaling. *Mol Cell* 81, 3449-3467.
- Dodds, P.N., Chen, J., and Outram, M.A. (2024). Pathogen perception and signaling in plant immunity. *The Plant Cell* 36, 1465-1481.
- Duggan, C., Moratto, E., Savage, Z., Hamilton, E., Adachi, H., Wu, C.H., Leary, A.Y., Tumtas, Y., Rothery, S.M., Maqbool, A., *et al.* (2021). Dynamic localization of a helper NLR at the plant-pathogen interface underpins pathogen recognition. *Proc Natl Acad Sci U S A* 118.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486-501.
- Engler, C., Kandzia, R., and 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., and Marillonnet, S. (2014). A Golden Gate Modular Cloning Toolbox for Plants. *ACS Synthetic Biology* 3, 839-843.
- Esse, H.P., Reuber, T.L., and Does, D. (2020). Genetic modification to improve disease resistance in crops. *New Phytol* 225, 70-86.
- Evans, P.R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* 67, 282-292.
- Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D* 69, 1204-1214.
- Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., *et al.* (2022). Protein complex prediction with AlphaFold-Multimer. *bioRxiv*, 2021.2010.2004.463034.
- Fahima, T., Reveguk, T., Fatiukha, A., Potapenko, E., Reveguk, I., Sela, H., Klymiuk, V., Li, Y., Pozniak, C., Wicker, T., *et al.* (2024). Atlas of tandem kinase proteins across the plant kingdom. Preprint, Nature Portfolio.

References

- Feehan, J.M., Castel, B., Bentham, A.R., and Jones, J.D.G. (2020). Plant NLRs get by with a little help from their friends. *Curr Opin Plant Biol* 56, 99-108.
- Feehan, J.M., Wang, J., Sun, X., Choi, J., Ahn, H.K., Ngou, B.P.M., Parker, J.E., and Jones, J.D.G. (2023). Oligomerization of a plant helper NLR requires cell-surface and intracellular immune receptor activation. *Proc Natl Acad Sci U S A* 120, e2210406120.
- Figuerola, M., Ortiz, D., and Henningsen, E.C. (2021). Tactics of host manipulation by intracellular effectors from plant pathogenic fungi. *Curr Opin Plant Biol* 62, 102054.
- Flor, H.H. (1971). Current Status of the Gene-For-Gene Concept. *Annu Rev Phytopathol* 9, 275-296.
- Förderer, A., Li, E., Lawson, A.W., Deng, Y.-n., Sun, Y., Logemann, E., Zhang, X., Wen, J., Han, Z., Chang, J., *et al.* (2022). A wheat resistosome defines common principles of immune receptor channels. *Nature* 610, 532-539.
- Gao, Z., Chung, E.H., Eitas, T.K., and Dangl, J.L. (2011). Plant intracellular innate immune receptor Resistance to *Pseudomonas syringae* pv. *maculicola* 1 (RPM1) is activated at, and functions on, the plasma membrane. *Proc Natl Acad Sci U S A* 108, 7619-7624.
- 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., and Valent, B. (2013). Two distinct secretion systems facilitate tissue invasion by the rice blast fungus *Magnaporthe oryzae*. *Nature Communications* 4, 1996.
- Giraldo, M.C., and Valent, B. (2013). Filamentous plant pathogen effectors in action. *Nature Reviews Microbiology* 11, 800-814.
- Gitschier, J., Moffat, B., Reilly, D., Wood, W.I., and Fairbrother, W.J. (1998). Solution structure of the fourth metal-binding domain from the Menkes copper-transporting ATPase. *Nat Struct Biol* 5, 47-54.
- Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin, T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* 27, 14-25.
- Grund, E., Tremousaygue, D., and Deslandes, L. (2018). Plant NLRs with Integrated Domains: Unity Makes Strength. *Plant Physiol* 179, 1227-1235.
- Guo, H., Ahn, H.-K., Sklenar, J., Huang, J., Ma, Y., Ding, P., Menke, F.L.H., and Jones, J.D.G. (2020). Phosphorylation-Regulated Activation of the *Arabidopsis* RRS1-R/RPS4 Immune Receptor Complex Reveals Two Distinct Effector Recognition Mechanisms. *Cell Host & Microbe* 27, 769-781.e766.

References

- Guo, L., Cesari, S., de Guillen, K., Chalvon, V., Mammri, L., Ma, M., Meusnier, I., Bonnot, F., Padilla, A., Peng, Y.-L., *et al.* (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* *115*, 11637-11642.
- Haley, O.C., Harding, S., Sen, T.Z., Woodhouse, M.R., Kim, H.-S., and Andorf, C. (2024). Application of RFDiffusion to predict interspecies protein-protein interactions between fungal pathogens and cereal crops. *bioRxiv*, 2024.2009.2017.613523.
- Hatsugai, N., Igarashi, D., Mase, K., Lu, Y., Tsuda, Y., Chakravarthy, S., Wei, H.L., Foley, J.W., Collmer, A., Glazebrook, J., *et al.* (2017). A plant effector-triggered immunity signaling sector is inhibited by pattern-triggered immunity. *EMBO J* *36*, 2758-2769.
- Hu, M., Qi, J., Bi, G., and Zhou, J.-M. (2020). Bacterial Effectors Induce Oligomerization of Immune Receptor ZAR1 In Vivo. *Molecular Plant* *13*, 793-801.
- Huang, W., MacLean, A.M., Sugio, A., Maqbool, A., Busscher, M., Cho, S.-T., Kamoun, S., Kuo, C.-H., Immink, R.G.H., and Hogenhout, S.A. (2021). Parasitic modulation of host development by ubiquitin-independent protein degradation. *Cell* *184*, 5201-5214.e5212.
- Ibrahim, T., Yuen, E.L.H., Wang, H.-Y., King, F.J., Toghiani, A., Kourelis, J., Vuolo, C., Adamkova, V., Castel, B., Jones, J.D., *et al.* (2024). A helper NLR targets organellar membranes to trigger immunity. *bioRxiv*, 2024.2009.2019.613839.
- Igarashi, S. (1986). Pyricularia em trigo. 1. Ocorrência de Pyricularia sp no estado do Paraná. *Fitopatol Bras* *11*, 351-352.
- 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., *et al.* (2017). Evolution of the wheat blast fungus through functional losses in a host specificity determinant. *Science* *357*, 80-83.
- 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., *et al.* (2016). Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *BMC Biol* *14*, 84.
- Jacob, P., Kim, N.H., Wu, F., El-Kasmi, F., Chi, Y., Walton, W.G., Furzer, O.J., Lietzan, A.D., Sunil, S., Kempthorn, K., *et al.* (2021). Plant "helper" immune receptors are Ca²⁺-permeable nonselective cation channels. *Science* *373*, 420-425.
- Jia, A., Huang, S., Song, W., Wang, J., Meng, Y., Sun, Y., Xu, L., Laessle, H., Jirschitzka, J., Hou, J., *et al.* (2022). TIR-catalyzed ADP-ribosylation reactions produce signaling molecules for plant immunity. *Science* *377*, eabq8180.
- Jia, Y., McAdams, S.A., Bryan, G.T., Hershey, H.P., and Valent, B. (2000). Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J* *19*, 4004-4014.

References

- Jo, B.-S., and Choi, S.S. (2015). Introns: The Functional Benefits of Introns in Genomes. *Genomics Inform* 13, 112-118.
- Johal, G., and Briggs, S. (1992). Reductase activity encoded by the HM1 disease resistance gene in maize. *Science* 258, 985-987.
- Jones, J.D.G., Vance, R.E., and Dangl, J.L. (2016). Intracellular innate immune surveillance devices in plants and animals. *Science* 354, aaf6395.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., *et al.* (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*.
- Kanzaki, H., Yoshida, K., Saitoh, H., Fujisaki, K., Hirabuchi, A., Alaux, L., Fournier, E., Tharreau, D., and 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.
- Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 30, 772-780.
- Keeble, A.H., Turkki, P., Stokes, S., Khairil Anuar, I.N.A., Rahikainen, R., Hytonen, V.P., and Howarth, M. (2019). Approaching infinite affinity through engineering of peptide-protein interaction. *Proc Natl Acad Sci U S A*.
- Keegan, R.M., and Winn, M.D. (2007). Automated search-model discovery and preparation for structure solution by molecular replacement. *Acta Crystallogr D Biol Crystallogr* 63, 447-457.
- Kim, S.H., Qi, D., Ashfield, T., Helm, M., and Innes, R.W. (2016). Using decoys to expand the recognition specificity of a plant disease resistance protein. *Science* 351, 684-687.
- Kourelis, J., Marchal, C., Posbeyikian, A., Harant, A., and Kamoun, S. (2023). NLR immune receptor-nanobody fusions confer plant disease resistance. *Science* 379, 934-939.
- Kourelis, J., and Van Der Hoorn, R.A.L. (2018). Defended to the Nines: 25 Years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. *The Plant Cell* 30, 285-299.
- Krissinel, E. (2010). Crystal contacts as nature's docking solutions. *J Comput Chem* 31, 133-143.
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372, 774-797.

References

- Kroj, T., Chanclud, E., Michel-Romiti, C., Grand, X., and Morel, J.B. (2016). Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytol* **210**, 618-626.
- Langner, T., Harant, A., Gomez-Luciano, L.B., Shrestha, R.K., Malmgren, A., Latorre, S.M., Burbano, H.A., Win, J., and Kamoun, S. (2021). Genomic rearrangements generate hypervariable mini-chromosomes in host-specific isolates of the blast fungus. *PLoS Genet* **17**, e1009386.
- Latorre, S.M., Were, V.M., Foster, A.J., Langner, T., Malmgren, A., Harant, A., Asuke, S., Reyes-Avila, S., Gupta, D.R., Jensen, C., *et al.* (2023). Genomic surveillance uncovers a pandemic clonal lineage of the wheat blast fungus. *PLoS Biol* **21**, e3002052.
- Lawson, A.W., Flores-Ibarra, A., Cao, Y., An, C., Neumann, U., Gunkel, M., Saur, I.M.L., Chai, J., Behrmann, E., and Schulze-Lefert, P. (2024). The barley MLA13-AVR_{A13} heterodimer reveals principles for immunoreceptor recognition of RNase-like powdery mildew effectors. *bioRxiv*, 2024.2007.2014.603419.
- Li, H., Wang, J., Kuan, T.A., Tang, B., Feng, L., Wang, J., Cheng, Z., Sklenar, J., Derbyshire, P., Hulin, M., *et al.* (2023). Pathogen protein modularity enables elaborate mimicry of a host phosphatase. *Cell* **186**, 3196-3207.e3117.
- Liang, X., and Zhou, J.-M. (2018). Receptor-Like Cytoplasmic Kinases: Central Players in Plant Receptor Kinase-Mediated Signaling. *Annu Rev Plant Biol* **69**, 267-299.
- Liu, F., Yang, Z., Wang, C., You, Z., Martin, R., Qiao, W., Huang, J., Jacob, P., Dangl, J.L., Carette, J.E., *et al.* (2024). Activation of the helper NRC4 immune receptor forms a hexameric resistosome. *Cell* **187**, 4877-4889.e4815.
- Liu, W., Liu, J., Triplett, L., Leach, J.E., and Wang, G.-L. (2014). Novel Insights into Rice Innate Immunity Against Bacterial and Fungal Pathogens. *Annu Rev Phytopathol* **52**, 213-241.
- Liu, Y., Zhang, X., Yuan, G., Wang, D., Zheng, Y., Ma, M., Guo, L., Bhadauria, V., Peng, Y.-L., and Liu, J. (2021). A designer rice NLR immune receptor confers resistance to the rice blast fungus carrying noncorresponding avirulence effectors. *Proceedings of the National Academy of Sciences* **118**, e2110751118.
- Lobstein, J., Emrich, C.A., Jeans, C., Faulkner, M., Riggs, P., and 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.
- Ma, S., An, C., Lawson, A.W., Cao, Y., Sun, Y., Tan, E.Y.J., Pan, J., Jirschitzka, J., Kümmel, F., Mukhi, N., *et al.* (2024). Oligomerization-mediated autoinhibition and cofactor binding of a plant NLR. *Nature*.

References

- Ma, S., Lapin, D., Liu, L., Sun, Y., Song, W., Zhang, X., Logemann, E., Yu, D., Wang, J., Jirschitzka, J., *et al.* (2020). Direct pathogen-induced assembly of an NLR immune receptor complex to form a holoenzyme. *Science* 370, eabe3069.
- Madhuprakash, J., Toghani, A., Contreras, M.P., Posbeyikian, A., Richardson, J., Kourelis, J., Bozkurt, T.O., Webster, M., and Kamoun, S. (2024). A disease resistance protein triggers oligomerization of its NLR helper into a hexameric resistosome to mediate innate immunity. *bioRxiv*, 2024.2006.2018.599586.
- Maekawa, S., Ohto, U., Shibata, T., Miyake, K., and Shimizu, T. (2016). Crystal structure of NOD2 and its implications in human disease. *Nature Communications* 7, 11813.
- Maekawa, T., Cheng, W., Spiridon, L.N., Töller, A., Lukasik, E., Saijo, Y., Liu, P., Shen, Q.H., Mićluta, M.A., Somssich, I.E., *et al.* (2011). Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. *Cell Host Microbe* 9, 187-199.
- Maidment, J.H.R., Franceschetti, M., Maqbool, A., Saitoh, H., Jantasuriyarat, C., Kamoun, S., Terauchi, R., and Banfield, M.J. (2021). Multiple variants of the fungal effector AVR-Pik bind the HMA domain of the rice protein OSHIPP19, providing a foundation to engineer plant defense. *J Biol Chem* 296, 100371.
- Maidment, J.H.R., Shimizu, M., Bentham, A.R., Vera, S., Franceschetti, M., Longya, A., Stevenson, C.E.M., De la Concepcion, J.C., Białas, A., Kamoun, S., *et al.* (2023). Effector target-guided engineering of an integrated domain expands the disease resistance profile of a rice NLR immune receptor. *eLife* 12, e81123.
- Maqbool, A., Saitoh, H., Franceschetti, M., Stevenson, C., Uemura, A., Kanzaki, H., Kamoun, S., Terauchi, R., and Banfield, M. (2015). Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife* 4.
- Martin, R., Qi, T., Zhang, H., Liu, F., King, M., Toth, C., Nogales, E., and Staskawicz, B.J. (2020). Structure of the activated ROQ1 resistosome directly recognizing the pathogen effector XopQ. *Science* 370, eabd9993.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography* 40, 658-674.
- Mélida, H., Sopena-Torres, S., Bacete, L., Garrido-Arandia, M., Jordá, L., López, G., Muñoz-Barrios, A., Pacios, L.F., and Molina, A. (2018). Non-branched β -1,3-glucan oligosaccharides trigger immune responses in *Arabidopsis*. *The Plant Journal* 93, 34-49.
- Mestre, P., and Baulcombe, D.C. (2005). Elicitor-Mediated Oligomerization of the Tobacco N Disease Resistance Protein. *The Plant Cell* 18, 491-501.

References

- Meunier, E., and Broz, P. (2017). Evolutionary Convergence and Divergence in NLR Function and Structure. *Trends Immunol* 38, 744-757.
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat Methods* 19, 679-682.
- Mor, T.S., Moon, Y.S., Palmer, K.E., and Mason, H.S. (2003). Geminivirus vectors for high-level expression of foreign proteins in plant cells. *Biotechnol Bioeng* 81, 430-437.
- Mukhi, N., Brown, H., Gorenkin, D., Ding, P., Bentham, A.R., Stevenson, C.E.M., Jones, J.D.G., and Banfield, M.J. (2021). Perception of structurally distinct effectors by the integrated WRKY domain of a plant immune receptor. *Proceedings of the National Academy of Sciences* 118, e2113996118.
- Neuhold, J., Radakovics, K., Lehner, A., Weissmann, F., Garcia, M.Q., Romero, M.C., Berrow, N.S., and Stolt-Bergner, P. (2020). GoldenBac: a simple, highly efficient, and widely applicable system for construction of multi-gene expression vectors for use with the baculovirus expression vector system. *BMC Biotechnol* 20.
- Ngou, B.P.M., Ahn, H.-K., Ding, P., and Jones, J.D.G. (2021). Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature* 592, 110-115.
- Oikawa, K., Fujisaki, K., Shimizu, M., Takeda, T., Nemoto, K., Saitoh, H., Hirabuchi, A., Hiraka, Y., Miyaji, N., Białas, A., *et al.* (2024). The blast pathogen effector AVR-Pik binds and stabilizes rice heavy metal-associated (HMA) proteins to co-opt their function in immunity. *bioRxiv*, 2020.2012.2001.406389.
- Ortiz, D., de Guillen, K., Cesari, S., Chalvon, V., Gracy, J., Padilla, A., and 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.
- Padmanabhan, M., Cournoyer, P., and Dinesh-Kumar, S.P. (2009). The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. *Cell Microbiol* 11, 191-198.
- Panchuk-Voloshina, N., Haugland, R.P., Bishop-Stewart, J., Bhalgat, M.K., Millard, P.J., Mao, F., Leung, W.-Y., and Haugland, R.P. (1999). Alexa Dyes, a Series of New Fluorescent Dyes that Yield Exceptionally Bright, Photostable Conjugates. *Journal of Histochemistry & Cytochemistry* 47, 1179-1188.
- Pereira, J., Simpkin, A.J., Hartmann, M.D., Rigden, D.J., Keegan, R.M., and Lupas, A.N. (2021). High-accuracy protein structure prediction in CASP14. *Proteins* 89, 1687-1699.
- Perkins, D.N., Pappin, D.J.C., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551-3567.

References

- Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* 30, 70-82.
- Posit (2023). RStudio: Integrated Development Environment for R. Posit Software.
- Pottinger, S.E., Bak, A., Margets, A., Helm, M., Tang, L., Casteel, C., and Innes, R.W. (2020). Optimizing the PBS1 Decoy System to Confer Resistance to Potyvirus Infection in Arabidopsis and Soybean. *Molecular Plant-Microbe Interactions®* 33, 932-944.
- Pottinger, S.E., and Innes, R.W. (2020). RPS5-Mediated Disease Resistance: Fundamental Insights and Translational Applications. *Annu Rev Phytopathol* 58, 139-160.
- Raman, M., and Martin, K. (2014). One solution for cloning and mutagenesis: In-Fusion® HD Cloning Plus. *Nat Methods* 11, iii-v.
- Robatzek, S. (2006). Ligand-induced endocytosis of the pattern recognition receptor FLS2 in Arabidopsis. *Genes Dev* 20, 537-542.
- Sainsbury, F., and Lomonossoff, G.P. (2008). Extremely High-Level and Rapid Transient Protein Production in Plants without the Use of Viral Replication. *Plant Physiol* 148, 1212-1218.
- Sainsbury, F., Thuenemann, E.C., and Lomonossoff, G.P. (2009). pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol J* 7, 682-693.
- Saleh, D., Milazzo, J., Adreit, H., Tharreau, D., and Fournier, E. (2012a). Asexual reproduction induces a rapid and permanent loss of sexual reproduction capacity in the rice fungal pathogen *Magnaporthe oryzae*: results of in vitro experimental evolution assays. *BMC Evol Biol* 12, 42.
- Saleh, D., Xu, P., Shen, Y., Li, C., Adreit, H., Milazzo, J., Ravigné, V., Bazin, E., NottéGhem, J.-L., Fournier, E., *et al.* (2012b). Sex at the origin: an Asian population of the rice blast fungus *Magnaporthe oryzae* reproduces sexually. *Mol Ecol* 21, 1330-1344.
- Sánchez-Vallet, A., Mesters, J.R., and Thomma, B.P. (2015). The battle for chitin recognition in plant-microbe interactions. *FEMS Microbiol Rev* 39, 171-183.
- Sanchez Carrillo, I.B., Hoffmann, P.C., Barff, T., Beck, M., and Germain, H. (2023). Preparing *Arabidopsis thaliana* root protoplasts for cryo electron tomography. *Frontiers in Plant Science* 14.
- Sarris, P.F., Cevik, V., Dagdas, G., Jones, J.D.G., and Krasileva, K.V. (2016). Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biol* 14.

References

- Savary, S., Willocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution* 3, 430-439.
- Schreiber, K.J., Bentham, A., Williams, S.J., Kobe, B., and Staskawicz, B.J. (2016). Multiple Domain Associations within the Arabidopsis Immune Receptor RPP1 Regulate the Activation of Programmed Cell Death. *PLoS Path* 12, e1005769.
- Seong, K., and Krasileva, K.V. (2021). Computational Structural Genomics Unravels Common Folds and Novel Families in the Secretome of Fungal Phytopathogen *Magnaporthe oryzae*. *Molecular Plant-Microbe Interactions®* 34, 1267-1280.
- Seong, K., and Krasileva, K.V. (2023). Prediction of effector protein structures from fungal phytopathogens enables evolutionary analyses. *Nature Microbiology* 8, 174-187.
- Shao, Z.-Q., Xue, J.-Y., Wu, P., Zhang, Y.-M., Wu, Y., Hang, Y.-Y., Wang, B., and Chen, J.-Q. (2016). Large-Scale Analyses of Angiosperm Nucleotide-Binding Site-Leucine-Rich Repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns. *Plant Physiol* 170, 2095-2109.
- Sharma, A., Kumar, V., Shahzad, B., Tanveer, M., Sidhu, G.P.S., Handa, N., Kohli, S.K., Yadav, P., Bali, A.S., Parihar, R.D., *et al.* (2019). Worldwide pesticide usage and its impacts on ecosystem. *SN Applied Sciences* 1, 1446.
- Shen, Q.H., Saijo, Y., Mauch, S., Biskup, C., Bieri, S., Keller, B., Seki, H., Ulker, B., Somssich, I.E., and Schulze-Lefert, P. (2007). Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses. *Science* 315, 1098-1103.
- Shipman, E.N., Jones, K., Jenkinson, C.B., Kim, D.W., Zhu, J., and Khang, C.H. (2017). Nuclear and structural dynamics during the establishment of a specialized effector-secreting cell by *Magnaporthe oryzae* in living rice cells. *BMC Cell Biol* 18, 11.
- Sievers, F., and Higgins, D.G. (2014). Clustal Omega. *Current Protocols in Bioinformatics* 48, 3.13.11-13.13.16.
- Sohn, K.H., Segonzac, C., Rallapalli, G., Sarris, P.F., Woo, J.Y., Williams, S.J., Newman, T.E., Paek, K.H., Kobe, B., and Jones, J.D. (2014). The nuclear immune receptor RPS4 is required for RRS1SLH1-dependent constitutive defense activation in *Arabidopsis thaliana*. *PLoS Genet* 10, e1004655.
- Sperschneider, J., and Dodds, P.N. (2022). EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. *Molecular Plant-Microbe Interactions®* 35, 146-156.
- Sperschneider, J., Gardiner, D.M., Dodds, P.N., Tini, F., Covarelli, L., Singh, K.B., Manners, J.M., and Taylor, J.M. (2016). EffectorP: predicting fungal effector proteins from secretomes using machine learning. *New Phytol* 210, 743-761.

References

- Steele, J.F.C., Hughes, R.K., and Banfield, M.J. (2019). Structural and biochemical studies of an NB-ARC domain from a plant NLR immune receptor. *PLoS One* *14*, e0221226.
- Studier, F.W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression Purif* *41*, 207-234.
- Sugihara, Y., Abe, Y., Takagi, H., Abe, A., Shimizu, M., Ito, K., Kanzaki, E., Oikawa, K., Kourelis, J., Langner, T., *et al.* (2023). Disentangling the complex gene interaction networks between rice and the blast fungus identifies a new pathogen effector. *PLoS Biol* *21*, e3001945.
- Tamborski, J., Seong, K., Liu, F., Staskawicz, B., and Krasileva, K.V. (2022). Engineering of Sr33 and Sr50 plant immune receptors to alter recognition specificity and autoactivity. *bioRxiv*, 2022.2003.2005.483131.
- Tameling, W.I., Vossen, J.H., Albrecht, M., Lengauer, T., Berden, J.A., Haring, M.A., Cornelissen, B.J., and Takken, F.L. (2006). Mutations in the NB-ARC domain of I-2 that impair ATP hydrolysis cause autoactivation. *Plant Physiol* *140*, 1233-1245.
- Tarr, D.E., and Alexander, H.M. (2009). TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC Res Notes* *2*, 197.
- Tembo, B., Mulenga, R.M., Sichilima, S., M'Siska K, K., Mwale, M., Chikoti, P.C., Singh, P.K., He, X., Pedley, K.F., Peterson, G.L., *et al.* (2020). Detection and characterization of fungus (*Magnaporthe oryzae* pathotype *Triticum*) causing wheat blast disease on rain-fed grown wheat (*Triticum aestivum* L.) in Zambia. *PLoS One* *15*, e0238724.
- Tenthorey, J.L., Haloupek, N., López-Blanco, J.R., Grob, P., Adamson, E., Hartenian, E., Lind, N.A., Bourgeois, N.M., Chacón, P., Nogales, E., *et al.* (2017). The structural basis of flagellin detection by NAIP5: A strategy to limit pathogen immune evasion. *Science* *358*, 888-893.
- Urbach, J.M., and Ausubel, F.M. (2017). The NBS-LRR architectures of plant R-proteins and metazoan NLRs evolved in independent events. *Proceedings of the National Academy of Sciences* *114*, 1063-1068.
- Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G.N. (2004). REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* *60*, 2184-2195.
- van der Hoorn, R.A., and Kamoun, S. (2008). From Guard to Decoy: a new model for perception of plant pathogen effectors. *Plant Cell* *20*, 2009-2017.
- van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J., and Steinegger, M. (2024). Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* *42*, 243-246.

References

- Varden, F.A., Saitoh, H., Yoshino, K., Franceschetti, M., Kamoun, S., Terauchi, R., and 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. *J Biol Chem* 294, 13006-13016.
- Wan, L., Essuman, K., Anderson, R.G., Sasaki, Y., Monteiro, F., Chung, E.-H., Osborne Nishimura, E., Diantonio, A., Milbrandt, J., Dangl, J.L., *et al.* (2019). TIR domains of plant immune receptors are NAD⁺-cleaving enzymes that promote cell death. *Science* 365, 799-803.
- Wang, G., Roux, B., Feng, F., Guy, E., Li, L., Li, N., Zhang, X., Lautier, M., Jardinaud, M.-F., Chabannes, M., *et al.* (2015). The Decoy Substrate of a Pathogen Effector and a Pseudokinase Specify Pathogen-Induced Modified-Self Recognition and Immunity in Plants. *Cell Host & Microbe* 18, 285-295.
- Wang, J., Hu, M., Wang, J., Qi, J., Han, Z., Wang, G., Qi, Y., Wang, H.-W., Zhou, J.-M., and 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., *et al.* (2019b). Ligand-triggered allosteric ADP release primes a plant NLR complex. *Science* 364, eaav5868.
- Wang, M., Weiberg, A., Dellota, E., Jr., Yamane, D., and Jin, H. (2017). Botrytis small RNA Bc-siR37 suppresses plant defense genes by cross-kingdom RNAi. *RNA Biol* 14, 421-428.
- Wang, S., Huang, W., Duxbury, Z., Hogenhout, S.A., and Jones, J.D.G. (2021). Novel effector recognition capacity engineered into a paired NLR complex. *bioRxiv*, 2021.2009.2006.459143.
- Wang, X., Lee, S., Wang, J., Ma, J., Bianco, T., and Ji, Y. (2014). Current Advances on Genetic Resistance to Rice Blast Disease. In (InTech).
- Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahern, W., Borst, A.J., Ragotte, R.J., Milles, L.F., *et al.* (2023). De novo design of protein structure and function with RFdiffusion. *Nature* 620, 1089-1100.
- Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE* 6, e16765.
- Weissmann, F., Petzold, G., VanderLinden, R., Huis in 't Veld, P.J., Brown, N.G., Lampert, F., Westermann, S., Stark, H., Schulman, B.A., and Peters, J.-M. (2016). biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. *Proceedings of the National Academy of Sciences* 113, E2564-E2569.

References

- 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., *et al.* (2007). A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* **450**, 115-118.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York).
- Williams, S.J., Sohn, K.H., Wan, L., Bernoux, M., Sarris, P.F., Segonzac, C., Ve, T., Ma, Y., Saucet, S.B., Ericsson, D.J., *et al.* (2014). Structural basis for assembly and function of a heterodimeric plant immune receptor. *Science* **344**, 299-303.
- Wilson, R.A., and Talbot, N.J. (2009). Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nature Reviews Microbiology* **7**, 185-195.
- Win, J., Krasileva, K.V., Kamoun, S., Shirasu, K., Staskawicz, B.J., and Banfield, M.J. (2012). Sequence divergent RXLR effectors share a structural fold conserved across plant pathogenic oomycete species. *PLoS Pathog* **8**, e1002400.
- Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**, 186-190.
- Winter, G., Waterman, D.G., Parkhurst, J.M., Brewster, A.S., Gildea, R.J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I.D., *et al.* (2018). DIALS: implementation and evaluation of a new integration package. *Acta Crystallographica Section D* **74**, 85-97.
- Wróblewski, T., Spiridon, L., Martin, E.C., Petrescu, A.J., Cavanaugh, K., Truco, M.J., Xu, H., Gozdowski, D., Pawłowski, K., Micheltore, R.W., *et al.* (2018). Genome-wide functional analyses of plant coiled-coil NLR-type pathogen receptors reveal essential roles of their N-terminal domain in oligomerization, networking, and immunity. *PLoS Biol* **16**, e2005821.
- Wu, C.-H., Abd-El-Haliem, A., Bozkurt, T.O., Belhaj, K., Terauchi, R., Vossen, J.H., and Kamoun, S. (2017). NLR network mediates immunity to diverse plant pathogens. *Proceedings of the National Academy of Sciences* **114**, 8113-8118.
- Xiong, Y., Han, Z., and Chai, J. (2020). Resistosome and inflammasome: platforms mediating innate immunity. *Curr Opin Plant Biol* **56**, 47-55.
- Yan, X., and Talbot, N.J. (2016). Investigating the cell biology of plant infection by the rice blast fungus *Magnaporthe oryzae*. *Curr Opin Microbiol* **34**, 147-153.
- Yan, X., Tang, B., Ryder, L.S., MacLean, D., Were, V.M., Eseola, A.B., Cruz-Mireles, N., Ma, W., Foster, A.J., Osés-Ruiz, M., *et al.* (2023). The transcriptional landscape of plant infection by the rice blast fungus *Magnaporthe oryzae* reveals distinct families of temporally co-regulated and structurally conserved effectors. *The Plant Cell* **35**, 1360-1385.

References

- Yoshida, K., Saunders, D.G., Mitsuoka, C., Natsume, S., Kosugi, S., Saitoh, H., Inoue, Y., Chuma, I., Tosa, Y., Cano, L.M., *et al.* (2016). Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. *BMC Genomics* 17, 370.
- Yuan, M., Jiang, Z., Bi, G., Nomura, K., Liu, M., Wang, Y., Cai, B., Zhou, J.M., He, S.Y., and Xin, X.F. (2021a). Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature* 592, 105-109.
- Yuan, M., Ngou, B.P.M., Ding, P., and Xin, X.-F. (2021b). PTI-ETI crosstalk: an integrative view of plant immunity. *Curr Opin Plant Biol* 62, 102030.
- Yue, J.-X., Meyers, B.C., Chen, J.-Q., Tian, D., and Yang, S. (2012). Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *New Phytol* 193, 1049-1063.
- Zdrzałek, R., Kamoun, S., Terauchi, R., Saitoh, H., and Banfield, M.J. (2020). The rice NLR pair Pikp-1/Pikp-2 initiates cell death through receptor cooperation rather than negative regulation. *PLOS ONE* 15, e0238616.
- Zdrzałek, R., Xi, Y., Langner, T., Bentham, A.R., Petit-Houdenot, Y., De la Concepcion, J.C., Harant, A., Shimizu, M., Were, V., Talbot, N.J., *et al.* (2024). Bioengineering a plant NLR immune receptor with a robust binding interface toward a conserved fungal pathogen effector. *Proceedings of the National Academy of Sciences* 121, e2402872121.
- Zhang, L., Chen, S., Ruan, J., Wu, J., Tong, A.B., Yin, Q., Li, Y., David, L., Lu, A., Wang, W.L., *et al.* (2015). Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science* 350, 404-409.
- Zhang, X., Bernoux, M., Bentham, A.R., Newman, T.E., Ve, T., Casey, L.W., Raaymakers, T.M., Hu, J., Croll, T.I., Schreiber, K.J., *et al.* (2017). Multiple functional self-association interfaces in plant TIR domains. *Proceedings of the National Academy of Sciences* 114, E2046-E2052.
- Zhang, X., He, D., Zhao, Y., Cheng, X., Zhao, W., Taylor, I.A., Yang, J., Liu, J., and Peng, Y.L. (2018). A positive-charged patch and stabilized hydrophobic core are essential for avirulence function of AvrPib in the rice blast fungus. *Plant J* 96, 133-146.
- Zhang, X., Liu, Y., Yuan, G., Wang, D., Zhu, T., Wu, X., Ma, M., Guo, L., Guo, H., Bhadauria, V., *et al.* (2022). The effector recognition by synthetic sensor NLR receptors requires the concerted action of multiple interfaces within and outside the integrated domain. *bioRxiv*, 2022.2008.2017.504349.
- Zhao, Y.-B., Liu, M.-X., Chen, T.-T., Ma, X., Li, Z.-K., Zheng, Z., Zheng, S.-R., Chen, L., Li, Y.-Z., Tang, L.-R., *et al.* (2022). Pathogen effector AvrSr35 triggers Sr35 resistosome assembly via a direct recognition mechanism. *Science Advances* 8, eabq5108.

References

Zhou, J.-M., and Zhang, Y. (2020). Plant Immunity: Danger Perception and Signaling. *Cell* 181, 978-989.

List of Abbreviations

Standard one- and three-letter codes are used for amino acids

2xYT	2x Yeast Extract Tryptone medium
A ₂₈₀	Absorbance at 280 nm
AcNPV-polh	Very late <i>Autographa californica</i> multiple nucleopolyhedrovirus polyhedrin promoter
AI	Artificial Intelligence
AI	Arabinose Inducible <i>E. coli</i>
AIM	Auto-induction Media
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BIC	Biotrophic Interface Complex
BN-PAGE	Blue-Native PAGE
CC	Coiled-coil
CCP4	Collaborative Computational Project No. 4 – Software for Macromolecular X-Ray Crystallography
CC-NLR	Coiled-Coil-NLR, a subcategory of NLR
CC _R -NLR	RPW8-type-CC-NLR, a subcategory of NLR
C-JID	C-terminal jelly-roll/Ig-like domain
DAMP	Damage associate molecular pattern
dATP	Deoxyadenosine triphosphate
DMSO	Dimethyl sulfoxide
d.p.i	Days post infiltration (<i>N. benthamiana</i>) or infection (insect cell work)
DTT	Dithiothreitol

List of Abbreviations

EDTA	Ethylenediaminetetraacetic acid
HA	Haemagglutinin epitope tag
HF	HellFire affinity tag - (6x His-3x FLAG)
HR	Hypersensitive Response
HRP	Horseradish peroxidase
HMA	Heavy Metal Associated domain
IPTG	Isopropyl β -D-1-thiogalactopyranoside
LB	Lysogeny Broth
LLM	Large Language Model
LysM	Lysine Motif
mCherry	A monomeric RFP (red fluorescent protein) variant
mEGP	Monomeric Enhanced Green Fluorescent Protein
NAD ⁺	Nicotinamide adenine dinucleotide
NB-ARC	Nucleotide Binding adaptor shared by Apaf-1, R proteins, and Ced-4
NBD	Nucleotide Binding Domain
NLR	Nucleotide binding, Leucine Rich Repeat protein
NLR	Nod-Like Receptor of animals – this use will be specified
pLDDT	A per-residue measure of local confidence of AlphaFold prediction accuracy
MCK	Multi-Cycle Kinetics
MAMP	Microbe associated molecular pattern
MAX	<i>Magnaporthe</i> AVR _s and ToxB-like effector
MBP	Maltose Binding Protein affinity tag
MoT	<i>M. oryzae</i> pathotype triticum
OD ₆₀₀	Optical Density at 600 nm, a measure of bacterial culture turbidity
p35S	35S promotor
PDB	Protein Data Bank

List of Abbreviations

PAMP	Pathogen Associated Molecular Pattern
PISA	Protein Interfaces, Surfaces and Assemblies
pMAS	Mannopine Synthase Promotor
PTI	Pattern-Triggered Immunity
PTM	Post-Translational Modification
pUbi	<i>Arabidopsis</i> Ubiquitin 10 promotor
pv.	Pathovar
PVDF	Polyvinylidene fluoride, a membrane for western blotting
RCF	Relative Centrifugal Force
RNAi	RNA interference
RSMD	Root Mean Square Deviation
SCK	Single-Cycle Kinetics
Sf9	Clonal lineage derived from <i>Spodoptera frugiperda</i> cells
Sf21	<i>Spodoptera frugiperda</i> insect cell line
SOC	Super Optimal broth with Catabolite repression
SPR	Surface Plasmon Resonance
sRNA	small RNA
SV40	Simian virus 40 early polyadenylation signal and terminator
t35S	35S terminator
TBS-T	Tris-Buffered Saline + 0.1 % Tween-20
TIR-NLR	Toll/Interleukin-1 Receptor-NLR, a subcategory of NLR
Tn7	A bacterial transposon

Endnote

Generative AI was used in the preparation of this thesis as follows:

ChatGPT-4 was employed in July-September 2024, during the editing of the introduction of this thesis, to demonstrate alternative ways of writing. Original sentences or short paragraphs written by me and only containing concepts present in the published literature were submitted to ChatGPT-4 and output was used to inspire different phrasing while retaining the meaning of the original sentence.

AlphaFold2 and RoseTTAFold were employed for structural prediction between 2021-2024. Protein sequences were submitted to AlphaFold2 and RoseTTAFold, and predicted structures generated were used for preparation of figures demonstrating the performance of structural prediction models, for designing protein expression constructs of OsHPPo2, OsHPPo3, OsHPPo4 and OsHPPo2/AVR-Mgk1 for X-ray crystallography, and during structure determination of the OsHPPo2/AVR-Mgk1 complex to provide a model of AVR-Mgk1 for molecular replacement.