

**John Innes Centre**

# **Recognition of rice blast effectors by paired immune receptors**

**Juan Carlos De la Concepción**

A thesis submitted to the University of East Anglia for the degree of  
**Doctor of Philosophy**

**John Innes Centre  
September 2019**

© 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 therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.



***A Mis Padres***



# Abstract

Plant diseases are a major burden for global agriculture. Pathogens deliver an array of molecules, termed effectors, to successfully colonize the host. Effectors can be recognized by intracellular immune receptors from the Nucleotide-binding Leucine-rich Repeat (NLR) family, activating immune responses that restrict pathogen growth. Engineering NLR receptors to enhance disease resistance has been a major biotechnological goal for over two decades. However, this has proven challenging as we still lack mechanistic knowledge on pathogen recognition and immune activation by NLRs. Recently, unconventional domains integrated in NLRs have been found to bind pathogen effectors and mediate immunity, opening new possibilities to engineer pathogen recognition.

Here, I combined molecular and structural biology to understand how integrated Heavy Metal Associated (HMA) domains in the rice NLR Pik mediate recognition to the rice blast pathogen *Magnaporthe oryzae*, and how they co-adapt with their acceptor NLR following integration. By obtaining the crystal structure of five complexes between HMA domains from allelic Pik resistance proteins and the AVR-Pik effectors from the rice blast pathogen, I could define the mechanistic basis of the arms-race co-evolution between plants and pathogens. Moreover, I showed that structural information can guide the engineering of disease resistance by the extending the pathogen recognition specificity of the Pikp NLR using structure-informed mutagenesis.

I also investigated the co-evolutionary dynamics between Pik receptors and how integrated HMA domains play a crucial role in regulation of immune responses. This uncovered unprecedent details on the co-adaptation between integrated domains and their NLR chassis, with broad implications from plant immunity. Altogether, my results advance the understanding on how integrated domains mediate pathogen recognition, their role in immune activation and the co-evolutionary events that underpin immune regulation. Finally, I applied the techniques developed here to a different rice blast effector, AVR-Pii, paving the way for new discoveries in effector biology and plant immunity.

The findings presented in this thesis link mechanistic and evolutionary research in plant immunity and have the potential to inform the design of immune receptors with enhanced capabilities, an important goal in global agriculture.

## **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.



## Author's declaration

The research described in this thesis was conducted entirely at the John Innes Centre. All the data described are original and were obtained by the author, except where specific acknowledgement has been made below and in the text. I had the pleasure of working with the following people throughout the course of this research:

Dr. Marina Franceschetti (John Innes Centre) and I worked together on the projects that gave rise to the results in Chapter 3 and Chapter 4.

Javier Vega Benjumea (University of Sevilla) joined the lab through the ERASMUS+ programme and helped me to produce the mutants and performed some of the cell death assays that are presented in Chapter 5 and 6.

Neftaly Cruz Mireles (The Sainsbury laboratory) joined the laboratory as a rotation student and helped to set up part of the protein purification and in vivo assays presented in Chapter 7.

Victor Sanchez de Medina joined the lab through the ERASMUS+ programme and contributed to fine-tuning the purification of plant Exo70 proteins presented in Chapter 7.

Part of the work contained in this thesis has contributed to the following publications, which are included in the appendices:

**De la Concepcion, J. C.\***, Franceschetti, M.\*., MacLean, D., Terauchi, R., Kamoun, S., Banfield, M. J. (2019). "Protein engineering expands the effector recognition profile of a rice NLR immune receptor." *eLife* 8: e47713.

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

\* indicates that these authors contributed equally.



## Acknowledgements

I would like to start by thanking my supervisor, Professor Mark Banfield, for the support and patience over the last four years. Thanks for sharing your invaluable knowledge during this time and thank you for your ability to keep me focused on my project while letting me explore new ideas. My thanks also go to the past and present members of the Banfield lab, for insightful comments and productive discussions. I would like to thank these people that contributed to my work, particularly Dr. Marina Franceschetti. I was immensely lucky to work side-by-side with her; her expertise and willingness to help have been a source of inspiration and wisdom. I was also very happy to supervise three good friends during my PhD, and would like to thank Javier Vega, Victor Sanchez and Neftaly Cruz for their multiple contributions to this work.

I also would like to thank Professor Sophien Kamoun and the past and present members of the Kamoun lab. I will be always grateful to him for giving me the opportunity to join his group five years ago, and for considering me a member of the lab even when I left. The wisdom shared by him and his lab shaped me as a scientist.

My gratitude also goes to Professor Rob Field for sparing his time to monitor the course of my PhD as a committee member. Thank you as well to Professor Stephen Bornemann for his guidance throughout the rotation programme. I also want to thank Professor Dave Lawson and Dr. Clare Stevenson (JIC protein crystallography and biophysical analysis platforms) for all the help across my PhD, the work presented here could not be possible without their contribution. Also, thanks to the support staff at JIC, especially Andrew David and Phil Robinson from Bio-Imaging, Tracy Hannant, the media kitchen and horticultural services. The John Innes Centre is such an amazing place to do research because of their hard work.

Finally, I would like to acknowledge the major contributions throughout my PhD by Dr. Sibyl Batey, colleague, partner and mentor. All these years of advice, understanding, experiences and fun made me improve as a scientist and as a person. Especially, thanks for helping me with my “barrier language”. I just hope to have the opportunity to give all this back to you one day.

## Agradecimientos

Esta tesis está dedicada a mis padres, por sus años de sacrificio, paciencia y apoyo incondicional. También quiero agradecérselo a mi familia, que hace que siempre tenga un lugar al que volver. A Jaime y Silvia, por su amistad desde que era un niño y por acogerme cada vez que vuelvo como si nunca me hubiera ido. Llegar hasta aquí tampoco hubiera sido posible sin la gente que he encontrado por el camino, no siempre fue fácil, y el apoyo de muchísimas personas me hizo no tirar la toalla. A todos los compañeros de la Uni, por los largos días de estudio y risas. En especial a los Mosqueperros, Alex, Rafa y Javi. Puede que ya no nos veamos tanto, pero los llevo siempre conmigo.

A Sole, por el apoyo mutuo estos años, con sus buenos y malos momentos, y porque, aunque nos vemos poco, parece que el tiempo no pase. Al final no nos cogieron en el Trinity College, pero sobrevivimos a UK y acabamos la tesis, que no es poco.

He tenido la inmensa suerte de tener a Roger y Neftaly como compañeros de laboratorio (aunque fuera en el de al lado). La felicidad contagiosa de Roger, su entusiasmo y sus ideas han sido una inspiración constante en mi día a día. Los consejos y el positivismo de Nef me han guiado todo este tiempo, incluso cuando estábamos lejos. Espero que sigamos acumulando charlas, experiencias y anécdotas juntos.

Quiero acabar agradeciéndole sobre todo a mi familia en Norwich. A los que estuvieron todo el camino y los que acaban de llegar. No escribo nombres por si alguno se me queda en el tintero. Tener el apoyo, consejos y ayuda de todos vosotros hace la vida más llevadera. Gracias por hacer que esté en casa, aunque viva a mil kilómetros de ella. Quien hubiera dicho que el grupo crecería tanto y con tan buena gente. ¡Lo que es la vida! Que el futuro nos traiga más bodorrios y fiestas en el tapas!

Os quiero,

JC

## Table of Contents

<b>Abstract</b> .....	v
<b>Author's declaration</b> .....	vii
<b>Acknowledgements</b> .....	ix
<b>Table of Contents</b> .....	xi
<b>Index of Figures</b> .....	xvii
<b>Index of Tables</b> .....	xxiii
<b>1. Introduction</b> .....	3
1.1 Plant pathogens are a threat to global agriculture .....	3
1.2 The cereal killer <i>Magnaporthe oryzae</i> .....	4
1.3 Pathogen virulence is determined by the repertoire of effectors.....	4
1.4 Structural biology is a powerful tool to inform research in effector biology .....	7
1.5 The plant immune system.....	9
1.6 Intracellular Nucleotide-Binding Leucine-Rich Repeat (NLR) receptors have a modular architecture .....	10
1.7 Plant NLRs use multiple strategies to recognize and respond to pathogen effectors .....	11
1.8 NLR immune receptors can act together in pairs and networks .....	13
1.9 Activation of plant NLRs triggers conformational changes that lead to cell death signalling .....	15
1.10 Regulation of NLR immune responses and autoimmunity.....	19
1.11 Paired NLRs have unconventional domains integrated into their architecture...	20
.....	20
1.12 NLR receptors are targets for engineering disease resistance.....	22
1.13 Rice disease resistance locus <i>Pik</i> encodes an NLR pair and recognizes a rice blast effector through an integrated domain .....	23
1.14 The rice NLR pair <i>Pii</i> monitors the host exocyst complex to mediate resistance to <i>M. oryzae</i> .....	24
1.15 Aims and objectives.....	25
<b>2. Materials and methods</b> .....	29
2.1 Molecular biology.....	29
2.1.1 Polymerase chain reaction .....	29
2.1.2 Site-directed mutagenesis .....	29
2.1.3 In-Fusion™ cloning .....	29
2.1.4 TOPO® cloning.....	30
2.1.5 Golden gate cloning.....	30

2.1.6	Plasmid transformation into <i>E. coli</i> and <i>A. tumefaciens</i> .....	30
2.1.7	Co-transformation of yeast cells for Y2H.....	31
2.2	Computational methods .....	31
2.2.1	Protein interface analyses.....	31
2.2.2	Root Mean Squared Deviation (RMSD) calculation.....	32
2.2.3	Phylogenetic analyses .....	32
2.3	SDS-PAGE and Western blot .....	32
2.4	Protein production and purification.....	33
2.4.1	Protein production in <i>E. coli</i> .....	33
2.4.2	Production of protein complexes .....	35
2.5	Protein-protein interactions .....	35
2.5.1	Yeast-2-hybrid (Y2H) .....	35
2.5.2	In planta co-immunoprecipitation (Co-IP).....	36
2.5.3	Analytical gel filtration (AGF).....	37
2.5.4	Isothermal titration calorimetry (ITC) .....	37
2.5.5	Surface plasmon resonance (SPR).....	37
2.6	Spectroscopic techniques .....	38
2.6.1	Small-angle X-ray scattering (SAXS) .....	38
2.6.2	X-ray crystallography .....	38
2.7	Cell death assays .....	40
2.7.1	Cell death scoring: UV autofluorescence .....	41
2.7.2	Cell death scoring: Ion leakage.....	41
<b>3.</b>	<b>Polymorphic residues in an allelic NLR expand binding and response to effectors of the blast pathogen .....</b>	<b>45</b>
3.1	Introduction.....	47
3.2	Pikm-mediated cell death in <i>N. benthamiana</i> recapitulates the rice allele-specific immune responses to AVR-Pik effectors.....	48
3.3	Allele-specific effector responses in planta correlate with direct Pik-HMA interactions .....	51
3.4	Pikm-HMA can be produced and purified from <i>E. coli</i> .....	52
3.5	Pikm-HMA has tighter binding affinities for AVR-Pik effectors than Pikp-HMA in vitro .....	55
3.6	Production, purification and crystallization of Pik-HMA/AVR-Pik complexes ..	60
3.7	Structures of Pik-HMAs in complex with AVR-Pik effectors reveal multiple interaction surfaces .....	65
3.8	The structure of Pikp-HMA bound to AVR-PikE uncovers the mechanistic basis of how polymorphic effectors escape Pikp recognition.....	70

3.9	Structural changes at interface two underpin differential effector responses by Pikm .....	72
3.10	Interactions across interface three are the determinants of the extended binding of Pikm to polymorphic AVR-Pik effectors .....	73
3.11	Mutations at separate interfaces have differential effects on Pik-HMA/effector binding .....	75
3.12	Mutations at separate interfaces have differential effects on Pikp and Pikm mediated immune phenotypes .....	77
3.13	Discussion .....	80
<b>4.</b>	<b>Protein engineering expands the effector recognition profile of a rice NLR immune receptor.....</b>	<b>85</b>
4.1	Introduction.....	87
4.2	Structure-informed engineering expands Pikp effector recognition in <i>N. benthamiana</i> cell death assays .....	89
4.3	The engineered Pikp-HMA <sup>NK-KE</sup> mutant shows increased binding to effector variants <i>in vivo</i> compared to Pikp-HMA .....	93
4.4	In vitro binding affinity to AVR-Pik effectors is increased in Pikp-HMA <sup>NK-KE</sup> ..	95
4.5	The engineered sensor NLR Pikp-1 <sup>NK-KE</sup> has extended association to AVR-Pik effector variants <i>in planta</i> .....	97
4.6	The engineered Pikp-HMA <sup>NK-KE</sup> adopts a more favourable conformation at the binding interface with the AVR-Pik effectors.....	98
4.7	Mutations in AVR-Pik effectors at the engineered interface impacts binding <i>in vivo</i> .....	105
4.8	Reduced binding to the AVR-Pik effectors mutants disrupts cell death responses <i>in planta</i> .....	107
4.9	Discussion .....	109
<b>5.</b>	<b>A single amino acid polymorphism determines pairing specificity in the Pik receptor pair.....</b>	<b>115</b>
5.1	Introduction.....	117
5.2	Rice Pik alleles forms two phylogenetically distinct groups .....	119
5.3	Pikm-2 polymorphisms are required for efficient cell death response to AVR-Pik effectors in <i>N. benthamiana</i> .....	123
5.4	Maladapted Pik pair Pikp-1/Pikm-2 triggers constitutive cell death responses in <i>N. benthamiana</i> .....	124
5.5	Pik autoactivation is linked to immune signalling.....	126
5.6	A single amino acid polymorphism in Pik-2 has a major role in cell death responses to AVR-Pik effectors.....	128
5.7	NLR specialization and autoimmunity are linked to the same polymorphism .... .....	131

5.8	Constitutive cell death responses in autoactive mutants are dependent on the P-loop and MHD motifs .....	133
5.9	The P-loop motif in the sensor NLR Pik-1 is important, but not essential, for cell death responses in <i>N. benthamiana</i> .....	135
5.10	The Glu230 polymorphism has evolved in modern rice .....	137
5.11	Sensor/helper hetero-pairing alters protein accumulation in Pik NLRs .....	141
5.12	Co-evolved and maladapted Pik pairs form heterocomplexes .....	142
5.13	Point mutations in Pik-2 helper NLR do not alter association with the sensor NLR .....	144
5.14	Sensor/helper association of Pik NLR pairs is independent of NLR activation ..	147
5.15	Sensor and helper NLRs preferentially associate to their co-evolved pair ...	149
5.16	Pik-2 polymorphisms underpin helper/sensor association preference .....	153
5.17	P-loop and MHD motifs are essential for Pikp-2 suppression of constitutive cell death .....	155
5.18	Discussion .....	158
<b>6.</b>	<b>The integrated HMA domain determines allelic compatibility in the rice NLR pair Pik</b> .....	163
6.1	Introduction.....	165
6.2	Pik-HMA is not required for cell death signalling .....	167
6.3	Pikm-HMA acts as a suppressor of Pikm-2 constitutive cell death .....	170
6.4	Chimeric Pik-HMAs uncover the regions involved in cell death repression and allelic incompatibility .....	171
6.5	Two amino acid polymorphisms in Pik-HMA underpin de-repression of constitutive cell death and allelic incompatibility .....	174
6.6	Discussion .....	180
<b>7.</b>	<b>The structure of rice blast effector AVR-Pii bound to host Exo70 uncovers a new fungal effector fold</b> .....	187
7.1	Introduction.....	189
7.2	AVR-Pii specifically associates with OsExo70F-2 and OsExo70F-3 in vivo.194	
7.3	Plant Exo70 proteins can be successfully produced and purified in <i>E. coli</i> .195	
7.4	Purified proteins conserve the characteristic Exo70 protein fold .....	197
7.5	Rice blast effector AVR-Pii can be produced and purified from <i>E. coli</i> .....	199
7.6	AVR-Pii binds OsExo70F-2 and OsExo70F-3 with high affinity in vitro.....	202
7.7	Crystallization and preliminary structure of AVR-Pii bound to rice OsExo70F-2 .....	204
7.8	OsExo70F-2 adopts a conserved Exo70 fold .....	206
7.9	AVR-Pii adopts a new fungal effector fold, based on a Zinc-Finger motif ....	209

7.10	AVR-Pii binds OsExo70s in a hydrophobic pocket.....	211
7.11	Mutations at the effector interface abrogate binding to rice Exo70s .....	215
7.12	OsExo70F-2 and OsExo70F-3 binds to rice RIN4 proteins in vivo .....	217
7.13	The Pii-NOI integrated domain does not bind OsExo70F-3 in vitro .....	219
7.14	Discussion and future experiments .....	221
<b>8.</b>	<b>Discussion .....</b>	<b>227</b>
8.1	The quest for engineered disease resistance .....	227
8.1.1	Multiple routes for NLR engineering.....	227
8.1.2	Extending effector recognition beyond nature.....	229
8.1.3	Future experiments and limitations of structure-guided engineering ....	230
8.2	The study of integrated domains: mechanistic research with evolutionary perspective.....	231
8.2.1	Co-evolution of NLRs as a Bateson-Dobzhansky-Muller evolution.....	231
8.2.2	Integrated domains shape the evolution and regulation of NLRs .....	232
8.2.3	Autoimmunity as a research tool.....	233
8.3	Effectors as tools for biological research.....	234
8.3.1	The new protein fold of <i>M. oryzae</i> AVR-Pii enables the discovery of new fungal effectors and integrated domains .....	234
8.3.2	Applications of effector tools .....	235
8.4	Closing remarks.....	236
<b>References .....</b>	<b>237</b>	
<b>Appendices .....</b>	<b>263</b>	



## Index of Figures

<b>Figure 1.1.</b> Rice blast effector from the MAX family share a conserved fold .....	8
<b>Figure 1.2.</b> NLR receptors use different strategies to detect pathogen effectors .....	12
<b>Figure 1.3.</b> Two different mechanisms of NLR pair activation.....	15
<b>Figure 1.4.</b> The formation of the apoptosome.....	16
<b>Figure 1.5.</b> The formation of the NAIP/NLRC4 inflammasome .....	17
<b>Figure 1.6.</b> The formation of the ZAR1 resistosome .....	18
<b>Figure 1.7.</b> Hybrid necrosis symptoms in wheat .....	20
<b>Figure 1.8.</b> NLRs integrate domains that mimic effector virulence targets into their architecture.....	21
<b>Figure 2.1.</b> Cell death scoring scale for UV quantification .....	41
<b>Figure 3.1.</b> Diversity of Pik NLR and AVR-Pik effectors.....	48
<b>Figure 3.2.</b> The Pik-mediated cell death response to AVR-Pik effector variants in <i>N. benthamiana</i> phenocopies the allelic resistance specificity in rice .....	51
<b>Figure 3.3.</b> Pikm-HMA shows increased binding to AVR-Pik variants in vivo.....	52
<b>Figure 3.4.</b> Pikm-HMA can be efficiently purified by IMAC + Gel filtration.....	53
<b>Figure 3.5.</b> 6xHis:MBP tag can be efficiently cleaved.....	54
<b>Figure 3.6.</b> Final gel filtration step yields a pure Pikm-HMA domain .....	54
<b>Figure 3.7.</b> Pikm-HMA forms a complex with AVR-PikD, AVR-PikE and AVR-PikA in vitro .....	55
<b>Figure 3.8.</b> Binding affinity between Pik-HMA domains and effectors in vitro correlates with in planta responses .....	57
<b>Figure 3.9.</b> Multi-cycle kinetics data for the interaction of Pikm-HMA with AVR-Pik effectors.....	58
<b>Figure 3.10.</b> Multi-cycle kinetics data for the interaction of Pikp-HMA with AVR-Pik effectors.....	59
<b>Figure 3.11.</b> Pikm-HMA can be co-purified with AVR-Pik effectors.....	60
<b>Figure 3.12.</b> 6xHis:MBP tag can be efficiently removed from the Pik-HMA/AVR-Pik complex .....	61
<b>Figure 3.13.</b> Final gel filtration step yields a pure Pikm-HMA domain complexed with AVR-Pik .....	61
<b>Figure 3.14.</b> Pik-HMA/AVR-Pik complexes can be successfully crystallized.....	62
<b>Figure 3.15.</b> Pik-HMA/AVR-Pik complexes adopt a similar fold .....	66
<b>Figure 3.16.</b> Analyses of Pik-HMA binding interfaces with AVR-Pik effectors using QtPISA .....	67

<b>Figure 3.17.</b> Pik-HMA domains form three interaction interfaces with AVR-Pik effectors .....	69
<b>Figure 3.18.</b> Lys262 in Pikp-HMA and Pikm-HMA make numerous interactions at the HMA/AVR-Pik interface three.....	70
<b>Figure 3.19.</b> His46Asn polymorphism in AVR-PikE disrupt interaction in Pikp-HMA/AVR-Pik interface two .....	71
<b>Figure 3.20.</b> The Pikp-HMA chain adopts an unfavourable conformation on interface three with AVR-PikE. ....	72
<b>Figure 3.21.</b> AVR-Pik polymorphisms in interface two alters binding to Pikm-HMA ....	73
<b>Figure 3.22.</b> Pikm-HMA adopts a different conformation at interface three compared to Pikp-HMA .....	74
<b>Figure 3.23.</b> Differential conformation at interface three is the structural determinant for Pikm extended recognition .....	75
<b>Figure 3.24.</b> Effector mutations at positions 46 and 53 perturb interactions with Pikp-HMA and Pikm-HMA.....	76
<b>Figure 3.25.</b> Mutations at different interfaces in the Pik-HMA/effector complexes have differential effects on cell death phenotypes.....	78
<b>Figure 3.26.</b> The reduction in cell death for AVR-Pik mutants is not due lower protein accumulation levels.....	79
<b>Figure 4.1.</b> Sequence alignment of Pikp-1 and Pikm-1 HMA domains.....	89
<b>Figure 4.2.</b> High-throughput cell death screening reveal the effects of mutations at Pikp-HMA interface two and three.....	90
<b>Figure 4.3.</b> NK-KE mutation at Pikp-HMA/AVR-Pik binding interface three extends cell-death phenotype to AVR-PikE and AVR-PikA .....	91
<b>Figure 4.4.</b> Pikp <sup>NK-KE</sup> displays a cell death phenotype comparable to Pikm but consistently elevated.....	92
<b>Figure 4.5.</b> Pikp-HMA <sup>NK-KE</sup> displays increased binding to AVR-Pik variants in vivo.....	94
<b>Figure 4.6.</b> Binding of Pikp-HMA <sup>NK-KE</sup> to the AVR-Pik effectors is consistently higher compared to Pikp-HMA in vitro.....	95
<b>Figure 4.7.</b> Binding of the Pikp-HMA <sup>NK-KE</sup> domain to the AVR-Pik effectors is consistently higher compared to Pikm-HMA in vitro.....	96
<b>Figure 4.8.</b> Pikp-1 <sup>NK-KE</sup> associates with AVR-Pik effector alleles in planta .....	97
<b>Figure 4.9.</b> Pik-HMA <sup>NK-KE</sup> /AVR-Pik complexes can be successfully crystallized .....	99
<b>Figure 4.10.</b> NK-KE mutations do not produce major structural changes in the Pik-HMA/AVR-Pik complexes .....	99
<b>Figure 4.11.</b> Lys261 in Pikp-HMA <sup>NK-KE</sup> plays a central role in the interaction at interface three .....	101

<b>Figure 4.12.</b> Pikp-HMA <sup>NK-KE</sup> adopts a different conformation across interface three, compared to Pikp-HMA.....	102
<b>Figure 4.13.</b> Pikp-HMA <sup>NK-KE</sup> binds AVR-PikE with a more favourable conformation across interface three, compared to Pikp-HMA .....	103
<b>Figure 4.14.</b> Pik-HMA has the same conformation across interface two in the Pikp-HMA and Pikp-HMA <sup>NK-KE</sup> complexes with AVR-PikD or AVR-PikE .....	104
<b>Figure 4.15.</b> Effector mutations at interface three perturb interactions with Pikp-HMA <sup>NK-KE</sup> in vivo.....	106
<b>Figure 4.16.</b> Effector mutations at interface two and three disrupt association to Pikp-1 <sup>NK-KE</sup> in planta .....	107
<b>Figure 4.17.</b> Effector mutations at interface two and three abrogate Pikp-1 <sup>NK-KE</sup> -mediated cell death in <i>N. benthamiana</i> .....	108
<b>Figure 4.18.</b> Pik-HMA <sup>NK-KE</sup> and AVR-PikC can be co-crystallized .....	110
<b>Figure 5.1.</b> The HMA domain in Pik-1 is the most variable region in the Pik NLR pair .....	119
<b>Figure 5.2.</b> Pik-2 polymorphisms are not evenly distributed in the rice cultivars .....	120
<b>Figure 5.3.</b> Pik-2 NLR sequences cluster in two distinct phylogenetic clades .....	121
<b>Figure 5.4.</b> Pik-1 NLR sequences cluster in two distinct phylogenetic clades .....	122
<b>Figure 5.5.</b> Pikm-2 is required for Pikm extended effector recognition in <i>N. benthamiana</i> .....	124
<b>Figure 5.6.</b> Pikm-2 triggers constitutive cell death in the presence of Pikp-1 .....	125
<b>Figure 5.7.</b> Pikm-2 autoactivation is dependent on P-loop and MHD motifs .....	127
<b>Figure 5.8.</b> Schematic representations of Pik-2 mutation at polymorphic sites .....	128
<b>Figure 5.9.</b> A single Pik-2 polymorphism has a major impact in cell death response to AVR-Pik effectors .....	130
<b>Figure 5.10.</b> Pik-2 alleles and mutants have similar levels of protein accumulation in planta.....	131
<b>Figure 5.11.</b> The pik-2 polymorphism at position 230 is the determinant for constitutive cell death .....	133
<b>Figure 5.12.</b> Pikp-2 Asp230Glu autoactivation is dependent on P-loop and MHD motifs .....	134
<b>Figure 5.13.</b> Mutations in Pik-2 P-loop and MHD motifs do not affect protein accumulation.....	135
<b>Figure 5.14.</b> The Pik-1 P-loop motif is important but not essential for Pik mediated cell death .....	136
<b>Figure 5.15.</b> A mutation in the Pik-1 P-loop motif does not affect protein accumulation .....	137

<b>Figure 5.16.</b> Evolutionary history of Pik-2 polymorphism at position 230 .....	138
<b>Figure 5.17.</b> Reversion to ancestral state of Pikm-2 Glu230 abolish autoimmunity...	140
<b>Figure 5.18.</b> Glu230Gly mutation does not affect Pik-2 protein accumulation .....	141
<b>Figure 5.19.</b> Pik NLRs display increased protein accumulation when co-expressed together in planta .....	142
<b>Figure 5.20.</b> Co-adapted and maladapted Pik NLR pairs associate in planta .....	143
<b>Figure 5.21.</b> Pik-2 mutants associate with Pikp-1 in planta.....	144
<b>Figure 5.22.</b> Pik-2 mutants associate with Pikm-1 in planta.....	145
<b>Figure 5.23.</b> Reversion to ancestral state in polymorphism 230 does not abrogate association with Pik-1 alleles .....	146
<b>Figure 5.24.</b> P-loop and MHD mutations do not affect Pik-2 association to Pik-1 ....	147
<b>Figure 5.25.</b> P-loop mutations does not affect Pik-1 association to Pik-2.....	148
<b>Figure 5.26.</b> Schematic representations of Pik NLR competition assays .....	149
<b>Figure 5.27.</b> Pikp-2 supress Pikm-2 mediated cell death .....	150
<b>Figure 5.28.</b> Pikm-2 cannot overcome suppression by Pikp-2 .....	151
<b>Figure 5.29.</b> Pikp-2 outcompetes Pikm-2 association to Pikp-1 .....	152
<b>Figure 5.30.</b> Pikp-2 partially supress cell death mediated by Pikp-2 .....	154
<b>Figure 5.31.</b> Cell death suppression by Pikp-2 requires an intact P-loop motif .....	156
<b>Figure 5.32.</b> Cell death suppression by Pikp-2 requires an intact MHD motif.....	157
<b>Figure 5.33.</b> Schematic representation of the evolutionary model proposed in this chapter.....	159
<b>Figure 6.1.</b> Pik-HMA is not required for cell death signalling.....	169
<b>Figure 6.2.</b> Pikm-HMA supresses Pik cell death signalling in the absence of effector.....	170
<b>Figure 6.3.</b> Schematic representation of secondary structure domain swaps between Pikm-HMA and Pikp-HMA regions .....	172
<b>Figure 6.4.</b> Pikm-HMA secondary structure regions $\beta$ 1, $\alpha$ 2 and $\beta$ 4 are involved in allelic incompatibility .....	173
<b>Figure 6.5.</b> Secondary regions involved in allelic incompatibility are situated together in the Pik-HMA structure.....	174
<b>Figure 6.6.</b> Pik-HMA regions involved in allelic incompatibility are polymorphic.....	175
<b>Figure 6.7.</b> A single polymorphism is the major determinant of allelic compatibility mediated by Pikm-HMA $\beta$ 1 .....	176
<b>Figure 6.8.</b> Allelic incompatibility mediated by Pikm-HMA $\alpha$ 2 is not determined by a single residue.....	177
<b>Figure 6.9.</b> A single polymorphism is the major determinant of allelic compatibility mediated by Pikm-HMA $\beta$ 4 .....	178

<b>Figure 6.10.</b> Residues involved in allelic incompatibility are situated away from the interaction interface with AVR-Pik effectors .....	179
<b>Figure 6.11.</b> Schematic representations of the differential regulation mechanisms proposed for <i>Pikp</i> and <i>Pikm</i> alleles .....	181
<b>Figure 6.12.</b> Schematic representations of the events proposed during adaptive evolution of the <i>Pik</i> NLR pair .....	183
<b>Figure 7.1.</b> The exocyst is a multimeric complex .....	191
<b>Figure 7.2.</b> Exo70 proteins have diversified in plants .....	192
<b>Figure 7.3.</b> AVR-Pii binds specifically to rice Exo70F-2 and Exo70F-3 <i>in vivo</i> .....	194
<b>Figure 7.4.</b> Plant Exo70 proteins can be efficiently purified by IMAC and gel filtration .....	196
<b>Figure 7.5.</b> The 6xHis:SUMO tag can be efficiently cleaved to yield Exo70 protein ..	196
<b>Figure 7.6.</b> The final gel filtration step yields pure OsExo70F-3 <sup>Δ93</sup> protein .....	197
<b>Figure 7.7.</b> Purified plant Exo70 proteins elute mainly as single peaks in analytical gel filtration analysis .....	198
<b>Figure 7.8.</b> Purified proteins adopt a folding shape similar to a monomeric AtExo70A-1 protein in solution .....	199
<b>Figure 7.9.</b> Rice blast effector AVR-Pii can be efficiently purified by IMAC and gel filtration .....	200
<b>Figure 7.10.</b> 6xHis:MBP tag can be efficiently cleaved .....	200
<b>Figure 7.11.</b> A final gel filtration step yields a pure AVR-Pii protein .....	201
<b>Figure 7.12.</b> Mass spectrometry analysis of AVR-Pii confirms the expected mass ...	201
<b>Figure 7.13.</b> AVR-Pii forms a complex with OsExo70F-3 <i>in vitro</i> .....	202
<b>Figure 7.14.</b> AVR-Pii binds OsExo70F-2 and OsExo70F-3 with high affinity .....	203
<b>Figure 7.15.</b> OsExo70F-2/AVR-Pii complexes form rod-like crystals .....	204
<b>Figure 7.16.</b> Crystal structure of AVR-Pii bound to rice Exo70F-2 .....	205
<b>Figure 7.17.</b> Rice Exo70F-2 adopt a similar fold to other Exo70 proteins .....	206
<b>Figure 7.18.</b> Distribution of α-helices in the sequence of OsExo70F-2 .....	207
<b>Figure 7.19.</b> Rice OsExo70F-2 is structurally similar to yeast and <i>Arabidopsis</i> Exo70 proteins .....	208
<b>Figure 7.20.</b> AVR-Pii adopts a new fungal effector fold .....	210
<b>Figure 7.21.</b> Analysis of OsExo70F-2 binding interface with AVR-Pii effectors using QtPISA .....	211
<b>Figure 7.22.</b> AVR-Pii residue Phe65 mediates the effector interaction with the host target .....	213
<b>Figure 7.23.</b> AVR-Pii residue Tyr64 makes important contributions to AVR-Pii effector/target interaction .....	214

<b>Figure 7.24.</b> Variation in OsExo70 hydrophobic pocket may underpin differences in effector binding .....	214
<b>Figure 7.25.</b> Effector mutations at AVR-Pii binding interface perturb interactions with rice Exo70s in vivo.....	215
<b>Figure 7.26.</b> Effector mutations at AVR-Pii binding interface abrogate interactions with rice Exo70s in vitro.....	216
<b>Figure 7.27.</b> Rice Exo70F-2 and Exo70F-3 bind to Rice RIN4 in vivo.....	218
<b>Figure 7.28.</b> Rice Exo70F-3 but not rice Exo70F-2 bind to Arabidopsis RIN4 in vivo.....	219
<b>Figure 7.29.</b> The NOI domain integrated in Pii-2 NLR does not interact with OsExo70F-3 in vitro.....	220
<b>Figure 7.30.</b> AVR-Pii effector is predicted to sit in the outside of exocyst subcomplex II .....	223
<b>Figure 7.31.</b> Structure superposition of AVR-Pii effector in the exocyst holocomplex .....	224
<b>Figure 8.1.</b> Different strategies for engineering Pik NLRs .....	229

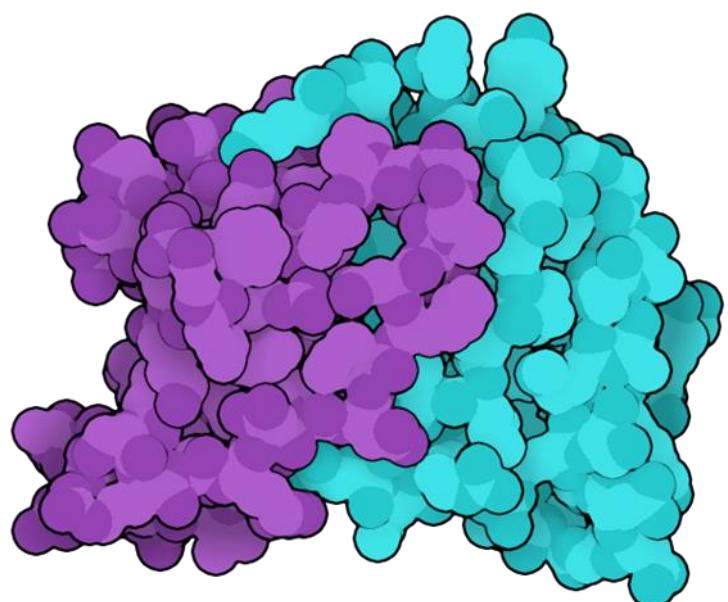
## Index of Tables

<b>Table 3.1.</b> Association and dissociation constants for Pik-HMAs and AVR-Pik effector alleles from the multi-cycle kinetics SPR.....	58
<b>Table 3.2.</b> X-ray data collection and refinement statistics for Pikm-HMA /AVR-PikD, Pikm-HMA /AVR-PikE and Pikm-HMA/AVR-PikA complexes.....	63
<b>Table 3.3.</b> X-ray data collection and refinement statistics for Pikp-HMA/AVR-PikD and Pikp-HMA/AVR-PikE complexes.....	64
<b>Table 3.4.</b> Summary of overlays analysis .....	65
<b>Table 3.5.</b> Summary of interface analysis.....	68
<b>Table 3.6.</b> Summary of the various interactions and phenotypes between Pik NLR alleles and effector variants .....	81
<b>Table 4.1.</b> X-ray data collection and refinement statistics for Pikp-HMA <sup>NK-KE</sup> /AVR-PikD and Pikp-HMA <sup>NK-KE</sup> /AVR-PikE complexes .....	100
<b>Table 4.2.</b> X-ray data collection and refinement statistics for Pikp-HMA <sup>NK-KE</sup> /AVR-PikC complex .....	111



# 1

## Introduction



AVR-PikD/Pikp-HMA

6G10



# 1. Introduction

## 1.1 Plant pathogens are a threat to global agriculture

Plant diseases are a continuous threat to global food security. The combined effect of crop pathogens imposes great economic losses to agriculture and is associated to devastating social and ecological problems worldwide (Bebber and Gurr, 2015; Fisher et al., 2012; Savary et al., 2019). To meet the food demand of an ever-growing population, major crop diseases caused by pathogens are a challenge that remains to be tackled.

Pathogens deliver an array of molecules, termed effectors, which modulate cell processes usually to overcome host defences (Win et al., 2012a). These effectors are major factors for successful colonization of the host by the pathogen, and can be either secreted to the apoplast or translocated inside of the host cell during pathogenesis (Dodds et al., 2009; Giraldo and Valent, 2013; Presti et al., 2015).

Importantly, plants have evolved intracellular immune receptors, mainly of the Nucleotide-binding, Leucine-rich Repeat (NLR) family that perceive translocated effectors or their activities. Recognition of effectors by NLR receptors trigger an immune response that leads to programmed cell death, restricting host colonization by the pathogen. This imposes a high selection pressure on the pathogens that, in turn, rapidly modify their effector repertoire to escape immune recognition. This co-evolutionary dynamic maintains a great diversity of pathogen effectors and plant receptors, shaping the interaction between pathogens and their host.

Filamentous pathogens are considered to be amongst the major threats to agriculture, particularly in the context of global environmental changes that favour the emergence and re-emergence of pathogens (Derevnina et al., 2016; Fisher et al., 2012; Kamoun et al., 2015; McDonald and Stukenbrock, 2016; Velasquez et al., 2018). Some of the most destructive plant diseases are caused by fungal pathogens (Dean et al., 2012). In particular, the causal agent of rice blast disease, *Magnaporthe oryzae*, is consistently listed as the most serious challenge that global agriculture is currently facing (Dean et al., 2012).

The blast fungus has been extensively studied over two decades and many genetic and biochemical tools are currently available (Ebbe, 2007; Talbot, 2003; Wilson and Talbot, 2009). Combined with the availability of the *Magnaporthe* genome (Dean et al., 2005), and that several *M. oryzae* effectors and their cognate plant resistance proteins have

been cloned, makes this pathogen a tractable model system to investigate broad concepts in molecular plant-microbe interactions and plant immunity (Bialas et al., 2018).

### 1.2 The cereal killer *Magnaporthe oryzae*

Rice blast is the most destructive disease of rice, the crop that is the major source of caloric intake for half of the world population (Wing et al., 2018). It is estimated rice blast causes production losses up to 30% of the global harvest (Savary et al., 2019; Talbot, 2003); enough to feed >60 million people annually (Pennisi, 2010). It is a severe problem in the majority of places where rice is grown, particularly in Asian and African countries (Seck et al., 2012; Skamnioti and Gurr, 2009; Talbot, 2003; Wilson and Talbot, 2009). As much of the increase in global population is predicted to occur in these densely populated areas (Wing et al., 2018), reducing rice production losses due to blast disease is a cornerstone to achieving global food security.

Furthermore, the blast fungus is a multi-host pathogen that can infect more than fifty species of grasses, including other staple crops such as wheat, oat, millet and barley (Cruz and Valent, 2017; Gladieux et al., 2018; Wilson and Talbot, 2009). This pathogen can rapidly adapt to different hosts (Huang et al., 2014), and can jump from one host to another (Inoue et al., 2017), creating serious disease outbreaks fuelled by the increase in global plant trade and climatic change (Islam et al., 2016; Islam et al., 2019).

Crop growing techniques and the breeding of natural resistance have provided practical and effective control of blast disease in the past. However, major epidemics still occur frequently, with catastrophic consequences towards food production (Pooja and Katoch, 2014). Since spraying chemicals to restrict the growth of pathogens is costly and dangerous for the environment, the development of durable genetic resistance to improve crop resilience to diseases is central in many plant biotechnology programmes.

### 1.3 Pathogen virulence is determined by the repertoire of effectors

Genomes of different plant pathogens have been shown to encode an armoury of putative effectors (Dean et al., 2005; Duplessis et al., 2011; Haas et al., 2009; Kämper et al., 2006; Levesque et al., 2010; Tyler et al., 2006), and their study has led to the discovery of fundamental concepts in plant-microbe interactions, progressing our understanding in the mechanistic basis of pathogenesis (Bialas et al., 2018).

Effectors are often small proteins and show signatures of rapid diversification, with little or no sequence similarity to proteins of known functions (Franceschetti et al., 2017; Varden et al., 2017). These genes commonly present extreme signatures of selection, which reflects their interaction with the host immune system (Allen et al., 2004; Dodds et al., 2006). In many cases they are situated in dynamically evolving regions of the genome associated with instability and structural variation (Dong et al., 2015; Raffaele et al., 2010; Raffaele and Kamoun, 2012), boosting the evolution of effectors with presence/absence polymorphisms, translocations, duplications and high levels of diversifying selection (Allen et al., 2004; Huang et al., 2014; Inoue et al., 2017; Yoshida et al., 2016). This accelerated evolution underpins the ability of plant pathogens to rapidly evolve to new environments and hosts, breaking disease resistances, and creating a challenge for the management of diseases.

The rapid diversification of pathogen effectors gives rise to new and optimised activities to manipulate metabolic processes and deactivate the host immune system. Some effectors have been shown to have enzymatic activities, which are frequently carried out by previously unknown domains (Franceschetti et al., 2017).

Translocated effectors interact with components of the host (Win et al., 2012a) and have proven to be very precise at discriminating their targets between a large number of related proteins/other factors (Dagdas et al., 2016; Fujisaki et al., 2015). Interestingly, effectors from multiple pathogens seem to convergently target the same proteins, suggesting an important role of these effector hubs in pathogen virulence (Mukhtar et al., 2011; Weßling et al., 2014). Unrelated effectors from the same pathogen can also target similar proteins, and this redundancy would allow for the deletion of effectors that are recognized by the immune system while maintaining a robust manipulation of certain host processes. This may explain why, in many cases, deletion of certain effectors only have minor effects towards pathogenesis (Bialas et al., 2018).

The hypervariability of pathogen effector evolution and function has not only made effector function difficult to predict, but identification of effectors from genomic sequences still remains challenging. However, some effector families from bacterial and filamentous pathogens have conserved features that can be used to inform the discovery of pathogen effectors (Varden et al., 2017). In bacterial pathogens, translocated effectors often have a signal peptide at the N-terminal region that is required for the export by the bacterial secretion system. The conserved characteristics of these regions facilitates the prediction and identification of effector candidates (Wang et al., 2011). Effectors are more highly represented in the genomes of filamentous pathogens, as reflected by their

## Chapter 1

higher copy numbers and variability rates compared to bacterial effectors (Dean et al., 2005; Haas et al., 2009; Kämper et al., 2006). Nevertheless, a common conserved motif required for translocation, named RXLR, has been identified towards the N-terminus of many oomycete effectors (Whisson et al., 2007), and can be used to predict putative oomycete effector families (Wood et al., 2019). However, the C-termini of these RXLR effectors are highly variable and their functions cannot be easily inferred (Win et al., 2007). Identification of effectors in fungal pathogens remains difficult. This is partly because these effectors do not share easily identifiable characteristics, which reflects the diversity exhibited in the delivery of effectors by different fungal pathogen to the host (Giraldo and Valent, 2013).

The rice blast fungus *M. oryzae* has a dynamic genome with signatures of extreme genetic diversity (Dean et al., 2005; Raffaele and Kamoun, 2012) and it contains hundreds of putative effectors. However, only a few of them have been functionally annotated, mainly through their avirulence activity (Zhang and Xu, 2014).

*M. oryzae* effectors are associated to telomeres and regions with transposable elements, and can be also found in accessory chromosomes (Peng et al., 2019). Being located in these regions promotes translocation, duplication, reshuffling and deletion of effector genes, maintaining the dynamic evolution of effectors to avoid detection by the host immune system (Huang et al., 2014; Yoshida et al., 2016). As well as presence/absence polymorphisms, about a third of the candidate effectors of *M. oryzae* show extreme signatures of adaptive evolution with high rations of non-synonymous substitutions, making computational prediction problematic (Bialas et al., 2018; Terauchi and Yoshida, 2010; Yoshida et al., 2016).

The rice blast fungus produces both apoplastic and cytoplasmic effectors (Giraldo and Valent, 2013). While apoplastic effectors are dispersed through the invasive hyphal membrane (EIHM) via conventional secretory pathways, cytoplasmic effectors are delivered by a specialized invasive structure named biotrophic interfacial complex (BIC), involving exocytosis (Giraldo and Valent, 2013). In this work, I will focus on two cytoplasmic effectors from the rice blast fungus, named AVR-Pik and AVR-Pii, which were identified by association genetics (Bialas et al., 2018; Yoshida et al., 2009).

## 1.4 Structural biology is a powerful tool to inform research in effector biology

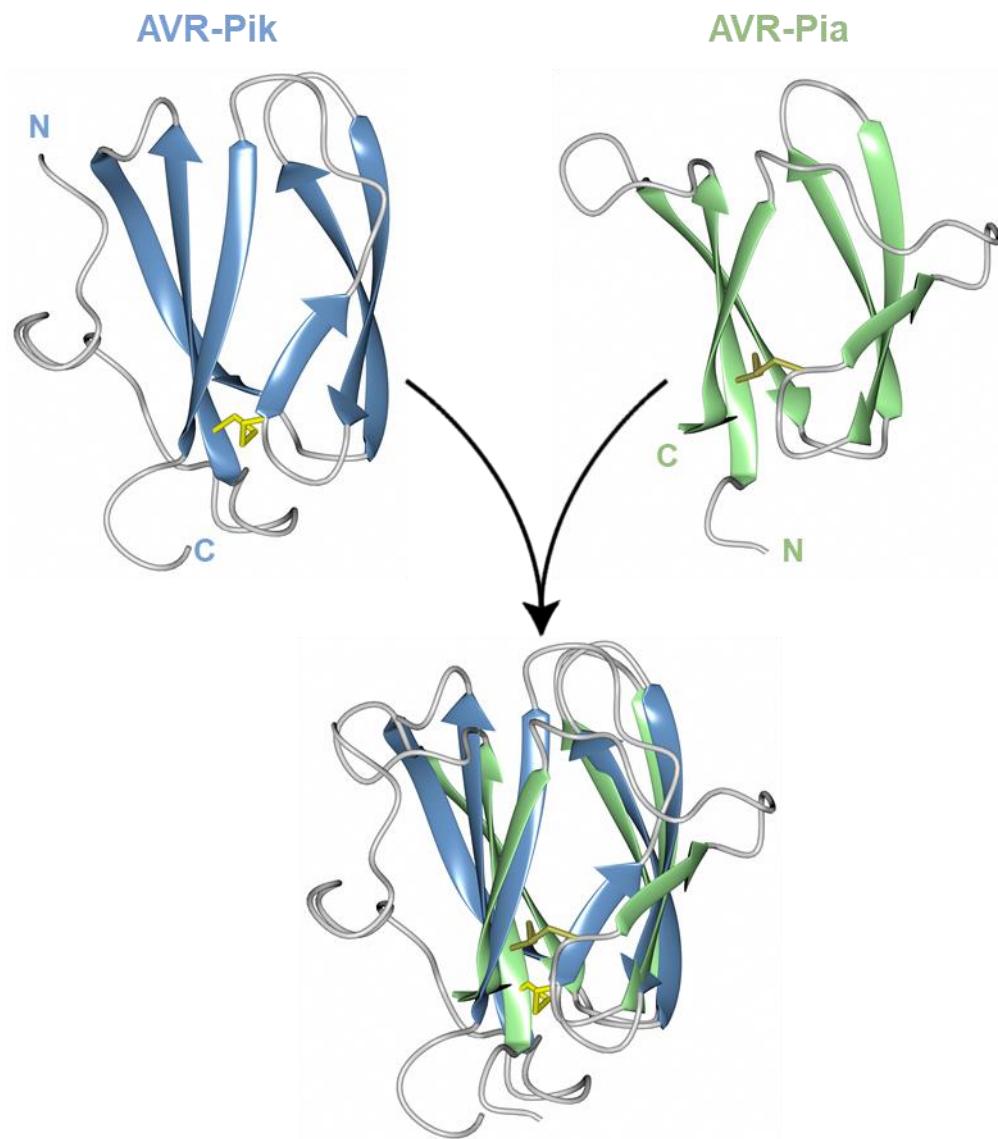
Structural biology has emerged as a powerful tool in effector biology (Stebbins, 2005; Wirthmueller et al., 2013). Structural information can uncover new functions of pathogen effectors (Janjusevic et al., 2006; Lopez et al., 2019) and can inform the description of effector families spanning different plant pathogens (de Guillen et al., 2015; Win et al., 2012b).

Pathogen effectors rarely share similarities to functional proteins, and are extraordinarily diverse. However, similar protein folds between effectors from filamentous pathogens have been found using structural biology (Franceschetti et al., 2017). These conserved folds act as structural scaffold to support protein integrity during the rapid evolution of pathogen effectors (Boutemy et al., 2011).

Structural analysis of the oomycete RXLR class of translocated effectors revealed a conserved structural core, named the WY domain, as is sustained by Trp and Tyr residues in conserved positions (Boutemy et al., 2011; Jiang et al., 2008). Despite the low sequence similarity and the high functional diversity, many oomycete effectors share the WY fold (Chou et al., 2011; Maqbool et al., 2016; Yaeno et al., 2011) or structural variations of this domain (He et al., 2019). The WY domain is present in a variable number of tandem repeats in different effectors, suggesting a function as a structural platform to support the rapid diversification of the effector (Franceschetti et al., 2017).

Bioinformatic searches have predicted that almost half of the *Phytophthora* RXLR effectors contain the WY fold (Win et al., 2012b). Further analysis also revealed this fold in other oomycete pathogens, including effectors lacking the conserved RXLR motif (Wood et al., 2019). Therefore, this domain represents a good predictor for effector candidates in the genomes of oomycete pathogens.

In *M. oryzae*, structural conservation of effectors has also been observed. Unexpectedly, to date, all rice blast effectors whose protein structure is known share a similar overall fold (de Guillen et al., 2015) (**Figure 1.1**). These effectors include AVR-Pik, AVR-Pia, AVR-Pizt, AVR1-CO39 and AVR-Pib (Guo et al., 2018b; Maqbool et al., 2015; Ose et al., 2015; Zhang et al., 2018; Zhang et al., 2013). Although sequence similarity between these proteins is lower than 25%, they form a similar six-stranded  $\beta$ -sandwich fold (**Figure 1.1**). Interestingly, this fold is also shared by a toxin from the fungus *Pyrenophora tritici-repentis* named ToxB (Nyarko et al., 2014).



**Figure 1.1. Rice blast effector from the MAX family share a conserved fold.** Cartoon representation of AVR-Pik and AVR-Pia structure and their superposition, showing the conserved six-strand  $\beta$ -sandwich MAX fold. Cysteines forming disulphide bonds are illustrated by sticks. Amino and carboxyl termini are labelled N and C, respectively. Figure is adapted from (Bialas et al., 2018).

Different effectors present extensions and variations in the length and orientation of the structural elements of the folding core, altering the protein shape and surface characteristics. These differences may underpin the differential binding of *M. oryzae* effectors to host targets, which is reflected, for example, in the binding to different surfaces of integrated HMA domains by AVR-Pik, AVR-Pia and AVR1-CO39 (De la Concepcion et al., 2018; Guo et al., 2018b; Maqbool et al., 2015; Ortiz et al., 2017; Varden et al., 2019).

The structural conservation of effectors led the definition of a new family of fungal effectors with diverse protein sequence named *Magnaporthe* AVRs and ToxB-like (MAX)

effectors (de Guillen et al., 2015). This family is highly expanded in *Magnaporthe* species, accounting for 5% to 10% of the total candidate effectors, and is present in other fungal species such as *Colletotrichum* (de Guillen et al., 2015). These effectors have strongly diversified and the core  $\beta$ -sandwich fold have provided the framework for the rapid evolution and neofunctionalization of the MAX effectors (de Guillen et al., 2015; Yoshida et al., 2016).

*M. oryzae* MAX effectors AVR-Pik, AVR-Pia and AVR1-CO39 seem to share the same type of host targets, as they all bind to integrated HMA domains (Guo et al., 2018b; Maqbool et al., 2015; Ortiz et al., 2017; Varden et al., 2019). However, other MAX effector, AVR-Pizt, bind to different host targets (Park et al., 2012; Park et al., 2016). This confirms that conserved effector folds such as the fungal MAX effectors and the WY described for oomycete pathogen effectors, while intriguing, are not strictly predictive of function (Bialas et al., 2018; Franceschetti et al., 2017).

Bioinformatic analyses have been employed as a tool for discovering and describing pathogen effectors (Sperschneider et al., 2018). However, prediction and classification of pathogen effectors remains challenging, particularly in the case of fungal effectors (Sperschneider et al., 2015; Varden et al., 2017). As protein structures are more conserved than amino acid sequences (Illergård et al., 2009), using structure-guided analysis has proven useful to circumvent the lack of sequence similarity in pathogen effectors, increasing the precision and sensitivity of prediction and characterization of effectors from filamentous pathogens (de Guillen et al., 2015; Win et al., 2012b; Wood et al., 2019).

### 1.5 The plant immune system

As they lack the adaptive immune system found in animals, plants depend on an innate immune system encoded in the germline to defend themselves from pathogens (Jones and Dangl, 2006). This system relies on a complex array of cell surface and intracellular receptors, which monitor the extracellular space and intracellular environment for the presence and activities of microbial pathogens (Jones et al., 2016; Ronald and Beutler, 2010). These receptors trigger signalling responses that restrict pathogen growth or lead to programmed cell death, restricting the spread of the pathogen (Dangl and Jones, 2001; Dodds and Rathjen, 2010; Jones and Dangl, 2006; Morel and Dangl, 1997). Plants also use constitutive and inducible mechanical barriers, such as the cell wall and hydrophobic cuticles, which act to resist pathogen attack (Uma et al., 2011).

Perhaps the best characterised cell surface immune receptors in plants are those that detect the presence of conserved pathogen associated molecular patterns (PAMPs) (Couto and Zipfel, 2016; Zipfel, 2014). These receptors are often described as providing a low level of response to non-adapted pathogens and in many cases are sufficient to maintain plant health. Pathogens and pest then translocate effectors to overcome this layer of defence and suppress plant defences initiated at the cell surface (Jones and Dangl, 2006; Win et al., 2012a).

In a strategy to counteract pathogen effectors, plants have evolved intracellular immune receptors that monitor the intracellular space to detect these proteins or their activities, and activate a strong and efficient cell death response to restrict the spread of the pathogen through the host (Dodds and Rathjen, 2010; Jones and Dangl, 2006). Many of this class of receptor are part of the Nucleotide-Binding, Leucine-rich Repeat (NLR) protein family, an ancient class of multidomain receptors also known to confer innate immunity in animals (Jones et al., 2016). Most plant disease resistance genes cloned to date belong to this family (Jacob et al., 2013; Jones et al., 2016), reflecting their importance. Furthering our understanding of how plant NLRs recognise effectors is a major focus of this thesis.

## **1.6 Intracellular Nucleotide-Binding Leucine-Rich Repeat (NLR) receptors have a modular architecture**

NLR proteins are highly represented in plant genomes (Meyers et al., 2003; Van de Weyer et al., 2019), including bryophytes and liverworts, and are also present in diverse animal clades including mammals (Broz and Dixit, 2016; Meunier and Broz, 2017).

NLRs have a modular architecture formed by different domains that contribute towards protein function (Takken and Goverse, 2012). The NLR family is defined by the presence of a conserved core formed from a Nucleotide-Binding (NB) domain, that plays a central role in NLR activation (Jones et al., 2016; Meunier and Broz, 2017), and a Leucine-Rich Repeat (LRR) domain at their C-terminus (Meunier and Broz, 2017). The NB domain has diversified in plants and animals from a common prokaryotic ancestor (Koonin and Aravind, 2002), and belongs to the STAND AAA+ ATPase superfamily (Leipe et al., 2004). Plant NB domains form a distinct lineage shared with apoptotic proteins such as mammalian APAF-1 and nematode CED-4 and is often referred to as the NB-ARC (Nucleotide-Binding, APAF-1, Resistance, CED4) domain (Yue et al., 2012). The NB-ARC is the most conserved domain in plant NLRs (Meyers et al., 2003) and includes the Walker A motif (also known as the P-loop) and the Walker B motif that are involved in

nucleotide binding and ATP hydrolysis. Mutations at these motifs normally render NLRs inactive and therefore NTP association has long been thought to be essential for NLR activity (Takken and Goverse, 2012; Takken et al., 2006; Tameling et al., 2002; Tameling et al., 2006).

As mentioned above, NLRs also contain a Leucine-Rich Repeat (LRR) domain at their C-terminus (Meunier and Broz, 2017). This domain is primarily thought to carry out regulatory functions as its deletion can cause constitutive NLR activation in several plant and animal proteins (Ade et al., 2007; Faustin et al., 2007). This function was confirmed in the crystal structure of the mammalian NLRC4 receptor, which revealed an autoinhibition mechanism via direct contact between the LRR and the NB domains (Hu et al., 2013). However, this autoinhibitory role of LRR domains might not be so strictly general, as other LRR domains have been shown to interact with regulatory kinases that are necessary for activation (Sharif et al., 2019) and have been implicated in direct interaction with pathogen effectors (Dodds et al., 2006; Jia et al., 2000). Interestingly, some plant NLR receptors have been shown to lack LRR domains (Nandety et al., 2013), but are still functional (Nishimura et al., 2017).

At their N-terminus, NLRs present accessory domains that can be very variable, particularly between plant and animals (Jones et al., 2016). In plants, these domains typically consist of a Coiled Coil (CC) domain or a TOLL/Interleukin-1 Receptor (TIR) domain (Bentham et al., 2018). The presence of one or the other defines the two major lineages of plant NLRs, named CNLs or TNLs, respectively (Bentham et al., 2018). Although these domains are very different, both CC and TIR domains are thought to transduce the receptor signal downstream upon NLR activation (Takken and Goverse, 2012).

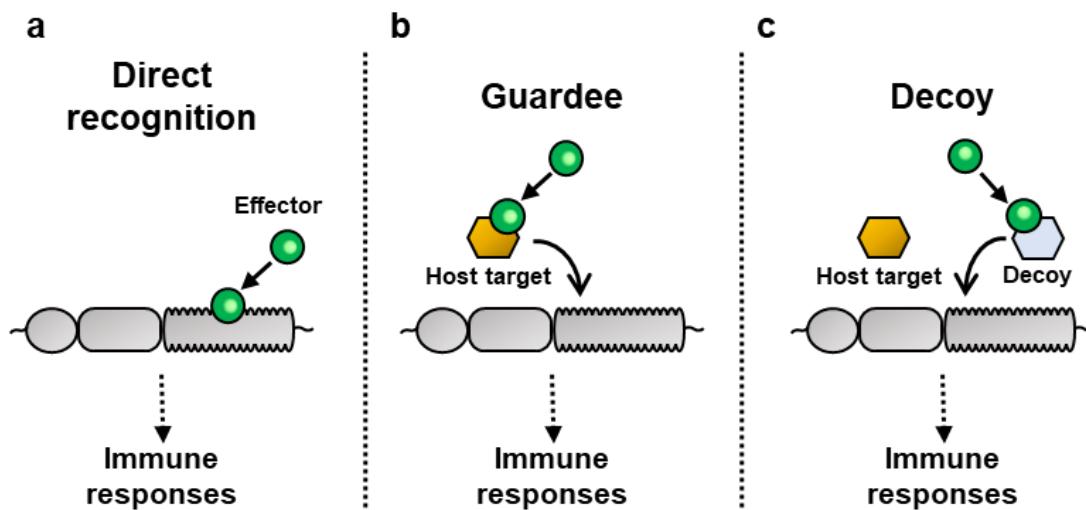
Overall, the modular architecture of NLRs, as well as their diversity, ensures the ability of these receptors to function as a robust immune switch.

### **1.7 Plant NLRs use multiple strategies to recognize and respond to pathogen effectors**

While animal NLRs detect conserved non-self molecular patterns as well as danger signals (Broz and Dixit, 2016), plant NLRs detect pathogen effectors very specifically using different strategies (Cesari, 2018; Kourelis and van der Hoorn, 2018). Three main modes of effector recognition have been proposed, depending on whether NLR recognition to the cognate effector is achieved by direct binding or indirectly through an intermediate host protein (**Figure 1.2**).

Plant NLRs have been shown to directly associate with pathogen effectors, particularly through interactions with the LRR domain (Dodds et al., 2006). This was the case for the rice blast effector AVR-Pita, whose interaction with the LRR domain of the Pi-ta NLR receptor determines effector recognition (Jia et al., 2000) (**Figure 1.2a**). However, in many cases, interactions between pathogen effectors and NLR receptors have not been observed. In such cases, the immune receptors may monitor the status of the virulence target of the effector (Jones and Dangl, 2006). These proteins, termed guardees, are host proteins with important roles in plant processes, commonly functioning in defence against pathogens. As pathogen effectors from many kinds of pathogens targets them (Mukhtar et al., 2011; Weßling et al., 2014), monitoring the status of these hubs is a mechanism to acquire resistance to many diverse pathogens (**Figure 1.2b**).

A prime example of an effector hub is the RPM1-Interacting Protein 4 (RIN4). This protein acts as a regulator of plant immunity, and is located at the plasma membrane (El Kasmi et al., 2017). Extensive research has shown how several pathogen effectors cause post-translational modifications on RIN4, including protein cleavage, altering its role in plant immunity. In turn, these modifications are detected by different NLRs, leading to immune responses (Axtell and Staskawicz, 2003; Chung et al., 2011; Chung et al., 2014; Lee et al., 2015; Liu et al., 2011; Mackey et al., 2002).



**Figure 1.2. NLR receptors use different strategies to detect pathogen effectors.** Schematic representation of different strategies of effector detection by NLRs. **(a)** Direct recognition, the pathogen effector directly binds to the NLR receptor. **(b)** Guardee model, the receptor monitors the status of the virulence targets of pathogen effectors. **(c)** Decoy model, NLR receptors monitors protein that mimic the effector targets whose sole function is activate immunity.

In contrast to the guardee model, some host proteins monitored by plant NLRs have no function other than interacting to pathogen effectors. These serve as non-functional copycats (or decoys) of the real pathogen targets and, upon binding to effectors, activate immune signalling via NLRs (van der Hoorn and Kamoun, 2008) (**Figure 1.2c**). The protein kinase PBS1 is an excellent example of this decoy model. While mutations in PBS1 do not show measurable phenotypical effects, it is targeted for proteolytical cleavage by different effectors, resulting in activation of immune responses mediated by NLRs (Ade et al., 2007; Kim et al., 2016; Shao et al., 2003). A second example are multiple pseudokinases that are targeted by different pathogen effectors (Seto et al., 2017; Wang et al., 2015a) whose modification triggers immune signalling dependent on the NLR ZAR1 (Wang et al., 2019a; Wang et al., 2019b).

### 1.8 NLR immune receptors can act together in pairs and networks

Although many NLRs follow the gene-for-gene dynamic hypothesised by Harold Flor (Flor, 1971), an increasing number of studies have revealed a more complex genetic architecture underpinning immune responses to pathogens.

In many cases, more than a single NLR is required to detect the presence of pathogens and trigger an immune response (Eitas and Dangl, 2010; Wu et al., 2018). When two or more NLRs cooperate to recognize pathogen effectors, their functions tend to be optimised, with one of them specializing in detecting the pathogen, termed sensor, while the other propagating the signalling response downstream, termed helper (Adachi et al., 2019; Jubic et al., 2019) (**Figure 1.7-1**). In plants, TNL and CNL sensors and helpers have been shown to organize in genetically linked pairs, sharing regulatory regions (Eitas and Dangl, 2010; Griebel et al., 2014; Narusaka et al., 2009; Sinapidou et al., 2004), or can be unlinked as part of intricate immune networks where several sensors can signal through diverse helper NLRs (Castel et al., 2019; Wu et al., 2018; Wu et al., 2017).

One of the outstanding questions in NLR biology is how cooperating immune receptors activate immune signalling. To date, two main models of cooperative activation have been proposed, negative regulation and cooperation/transactivation. These are distinguished by the ability of the helper NLR to trigger immune signalling in the absence of the cognate sensor (Bialas et al., 2018) (**Figure 1.3**).

In the case of NLRs working through negative regulation, helper NLRs are intrinsically active and their activity is suppressed by the sensor NLRs until an appropriate stimulus is perceived (**Figure 1.3**). Examples of this model of regulation include the rice NLR pair

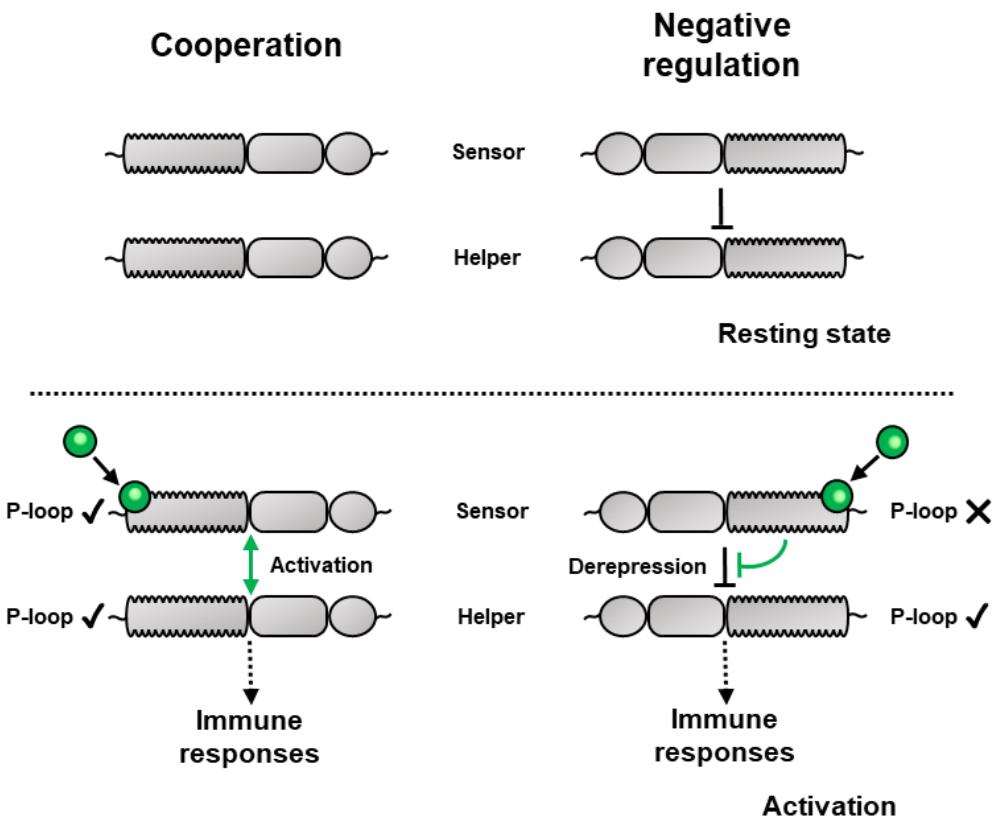
Pia (Cesari et al., 2014b; Cesari et al., 2013), and the Arabidopsis NLR pair RRS1/RPS4 (Narusaka et al., 2009) (Le Roux et al., 2015; Sarris et al., 2015). In each case, binding of the cognate effectors by the sensor NLRs releases negative regulation of the helper NLR to activate immune signalling and disease resistance.

Examples of the cooperation/transactivation model of activation include the rice NLR pair Pik (Bialas et al., 2018) (**Figure 1.3**). Further, a similar case of cooperation occurs in the mammalian NAIP/NLRC4 inflammasomes (Broz and Dixit, 2016). The NLR receptor NAIP acts as a sensor binding bacterial ligands (Kofoed and Vance, 2011; Tenthorey et al., 2017; Zhao et al., 2011). NLRC4 does not bind pathogen ligands, and instead is recruited as helper to amplify and transduce immune signalling (Hu et al., 2015; Tenthorey et al., 2017; Zhang et al., 2015a).

These different mechanisms of activation may shape the evolution of NLR genes. For example, NLRs involved in negative regulation would have a more restricted variability as the release of the constitutively active member will cause cell death and a subsequent loss of fitness. These NLRs are possibly retained as genetically linked pairs to ensure co-regulation (Bialas et al., 2018). In contrast, cooperating NLRs are more likely to evolve at higher rates as they can accommodate more changes.

Interestingly, plant NLRs that activate immune signalling through different mechanisms seem to have a different requirement for the NB P-loop motif, necessary for NLR activation. In cooperating NLR pairs and complexes, both sensor and helper NLRs require an intact P-loop to trigger immune signalling (Tran et al., 2017; Wu et al., 2017) (**Figure 1.3**). However, this might not be generalizable as it seems to not be the case in the mammalian NLRs NAIP/NLRC4, where NAIP P-loop is not required to activate signalling (Kofoed and Vance, 2011). By contrast, in NLR pairs acting through negative regulation an intact P-loop in the sensor is not required (Cesari et al., 2014b; Williams et al., 2014) (**Figure 1.3**). This could suggest these sensors may be merely acting as scaffolds for the helper NLR activities, as in the case of the NAIP/NLRC4 inflammasome (Hu et al., 2015; Tenthorey et al., 2017; Zhang et al., 2015a). These different requirements for immune activation hint at how complex and diverse NLR mechanisms are.

By acting synergistically in pairs and networks, the immune system ensures an efficient, robust and regulated response against a myriad of different pathogens. This also confers to the immune system an evolvability to keep up with rapidly evolving pathogens (Adachi et al., 2019; Wu et al., 2018; Wu et al., 2017).



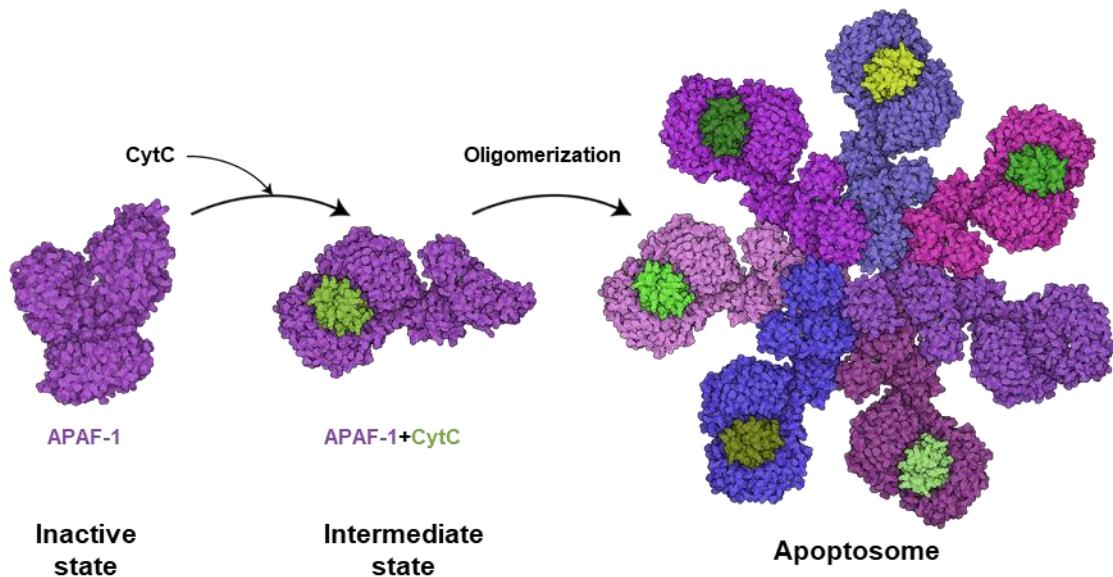
**Figure 1.3. Two different mechanisms of NLR pair activation.** Schematic representation of the NLR activation model by cooperation and negative regulation. In the negative regulation model, the sensor NLR represses the signalling of the helper NLR in the resting state. Upon effector recognition, cooperating NLRs act together to induce immune responses. Effector binding to the negative regulator sensor releases the suppression of the helper NLR, triggering immune responses. A conserved P-loop motif is required in both sensor and helper NLRs in the cooperation model while only an intact P-loop in the helper NLR is required in negative regulation.

## 1.9 Activation of plant NLRs triggers conformational changes that lead to cell death signalling

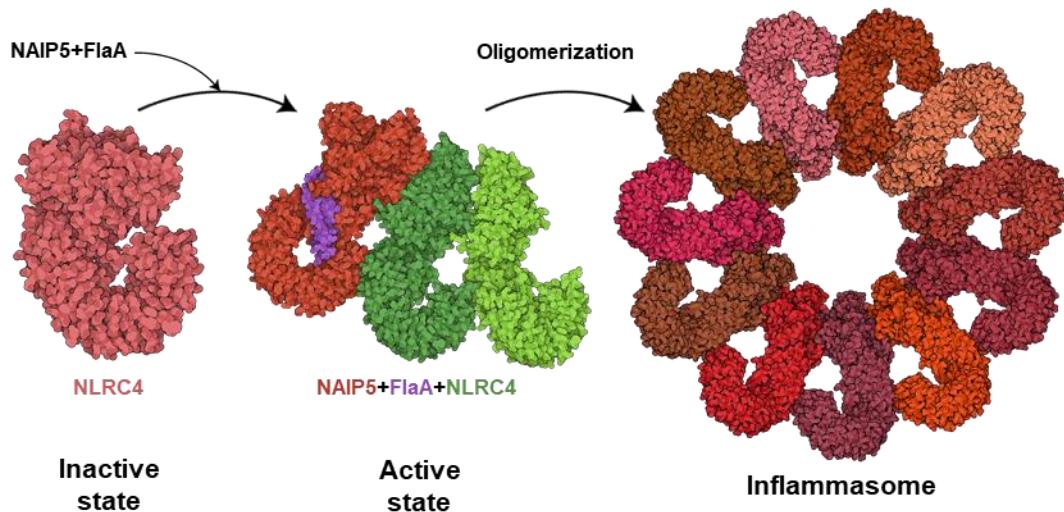
Early biochemical studies established that NLR activation involves ADP/ATP exchange at the nucleotide binding domain. These studies proposed that NB domains are maintained in an inactive, ADP-bound, state by intra and intermolecular interactions (Maekawa et al., 2011; Tameling et al., 2002). The NB domain maintains a constant equilibrium between the association to ADP or ATP, underpinning activation of the NLR (Bernoux et al., 2016; Tameling et al., 2006; Williams et al., 2011). Recognition of pathogen effectors, or their activities, would then switch the ADP-bound inactive state for an ATP-bound active state, possibly via conformational changes that enable ADP/ATP exchange (Bernoux et al., 2016; Takken et al., 2006; Tameling et al., 2002; Tameling et al., 2006; Williams et al., 2011). This explains why mutations in the NB-ARC

that affect ADP/ATP association can render NLRs either constitutively active or inactive (Tameling et al., 2006; Williams et al., 2011).

Largely due to the challenges of producing full length NLRs in vitro, most of the mechanistic understanding about NLR activation has been based on the study of individual domains (Bentham et al., 2018; Casey et al., 2016; Cesari et al., 2016; Maekawa et al., 2011; Zhang et al., 2017a). However, recent advances in structural studies of NLR receptors have begun to shed light on how post-activation events in NLRs lead to immune signalling. Studies of the mammalian APAF-1 apoptosome revealed the importance of oligomerization for activation (Cheng et al., 2016; Yuan and Akey, 2013) (**Figure 1.4**). Further support for this was provided from studies of the mammalian NAIP/NLRC4 inflammasome which, upon pathogen recognition, assembles into a wheel-like architecture (Diebolder et al., 2015; Hu et al., 2015; Tenthorey et al., 2017; Zhang et al., 2015a) (**Figure 1.5**).

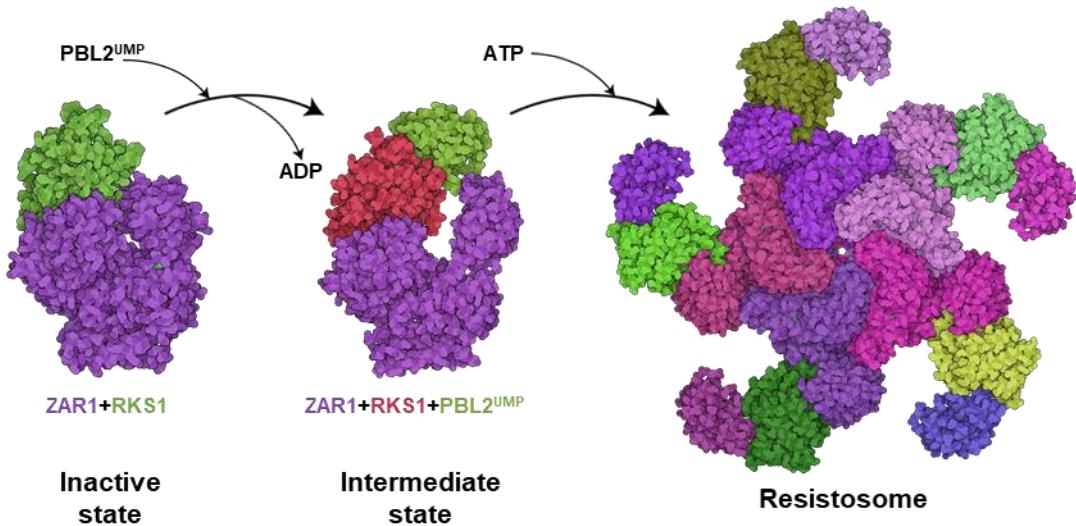


**Figure 1.4. The formation of the apoptosome.** Schematic representation of Apaf-1 activation by cytochrome C. Apaf-1 remains in a resting state and upon binding to CytC molecules, a conformational change promotes the oligomerization of the active apoptosome. Representation was built based on the full length inactive Apaf-1 structure (3SFZ) (Reubold et al., 2011) and the active apoptosome (5JUY) (Cheng et al., 2016) using Illustrate software for biomolecular illustration (Goodsell et al., 2019).



**Figure 1.5. The formation of the NAIP/NLRC4 inflammasome.** Schematic representation of NAIP/NLRC4 activation by bacterial flagellin. NLRC4 remains in a resting state. Binding of NAIP5 to FlaA promotes structural changes that lead to the recruitment of NLRC4 molecules and the oligomerization in a wheel-shaped complex. Representation was built based on the structure of inactive NLRC4 (4KXF) (Hu et al., 2013), the structure of NAIP5-NLRC4-flagellin inflammasome (6B5B) (Tenthorey et al., 2017) and the NLRC4 inflammasome complex (3JBL) (Zhang et al., 2015a) using Illustrate software for biomolecular illustration (Goodsell et al., 2019).

In plants, several studies pointed to the involvement of NLR association in immune signalling (El Kasmi et al., 2017; Huh et al., 2017; Mestre and Baulcombe, 2006; Schreiber et al., 2016). However, it remained unclear whether this association represented the formation of oligomers, as in the case of apoptosomes or inflammasomes. However, two recent studies revealed the mechanistic basis of HOPZ-ACTIVATED RESISTANCE 1 (ZAR1) NLR activation, showing that this NLR forms an oligomeric structure named the “resistosome” on sensing pathogen effector activity. This structure is reminiscent of the mammalian oligomeric apoptosomes and inflammasomes (Wang et al., 2019a; Wang et al., 2019b) (Figure 1.6). Intriguingly, activation of the ZAR1 resistosome has suggested a mechanism for CC domain in immune activation and signalling (Wang et al., 2019a). Upon oligomerization, the CC domains undergo a key conformational change, forming an  $\alpha$ -helical barrel with a funnel shape structure (Burdett et al., 2019; Wang et al., 2019a). This CC structure may mediate insertion into the cell membranes, directly leading to cell death responses associated with immune signalling (Burdett et al., 2019; Wang et al., 2019a).



**Figure 1.6. The formation of the ZAR1 resistosome.** Schematic representation of ZAR1 activation by bacterial effector AVRA-C. ZAR1/RKS1 remains in a resting state. Binding of PBL<sup>UMP</sup> promotes structural changes that lead to the release of ADP. Binding of ATP to the ZAR1/RKS1/PBL2<sup>UMP</sup> intermediate state promotes oligomerization in a wheel-shaped complex. Representation was built based on the structure of inactive ZAR1/RKS1 (6J5W) (Wang et al., 2019b), the structure of ZAR1/RKS1/PBL2<sup>UMP</sup> (6J5V) (Wang et al., 2019b) and the ZAR1 resistosome complex (6J5T) (Wang et al., 2019a) using Illustrate software for biomolecular illustration (Goodsell et al., 2019).

Although undoubtedly a major breakthrough, as mentioned above, many plant NLRs function in pairs or networks (Adachi et al., 2019; Eitas and Dangl, 2010; Jubic et al., 2019; Wu et al., 2018). With current knowledge, it is difficult to see how the activation mechanism of the singleton NLR ZAR1 could be extrapolated to paired NLRs. These type of NLRs may have an activation mechanism more similar to NAIP/NLRC4 inflammasomes.

In contrast to CNLs, recent findings have demonstrated that plant and animal TIR domains harbour a catalytic activity to cleave NADH. NADH depletion, and/or the production of intermediates of this reaction, have been proposed to cause cell death (Essuman et al., 2017; Horsefield et al., 2019; Summers et al., 2016; Wan et al., 2019). However, contrary to CNLs, downstream components of TNL signalling pathways have been identified and there is increasing evidence of direct interaction between these and TNLs (Huh et al., 2017). Therefore, different TIR containing NLRs may act through different mechanisms involving oligomerization, but at present this remains unclear.

## 1.10 Regulation of NLR immune responses and autoimmunity

The formation of inflammasomes and resistosomes may represent an additional layer of control in NLR receptors to fine-tune and regulate efficient responses against pathogens, without the deleterious effects of constitutive immune signalling (Chae et al., 2016; Richard and Takken, 2017). Even small modifications in NLRs can lead to loss of function, or constitutively active phenotypes, mainly caused by incompatible NLR interactions (Bernoux et al., 2016; Stirnweis et al., 2014a; Stirnweis et al., 2014b; Tran et al., 2017). This process is analogous to autoimmune or autoinflammatory diseases caused by undesirable mutations in human NLRs (de Torre-Minguela et al., 2017). Cell death responses mediated by NLR activation carry a fitness cost (MacQueen et al., 2016; Tian et al., 2003). Therefore, plants have evolved different mechanisms to regulate immunity and mitigate negative effects (Deng et al., 2017; Karasov et al., 2017), including positional regulation of NLRs by forming co-regulatory pairs (Bialas et al., 2018; Eitas and Dangl, 2010) or clustering them in specific genome regions (van Wersch and Li, 2019).

At the same time, NLR receptors are the most diverse and fast-evolving protein family in plants (Baggs et al., 2017; Meyers et al., 2003; Yue et al., 2012). This rapid diversification can drive the evolution of interacting NLRs in different trajectories. As a consequence, this can lead to mis-regulation of immune responses, particularly in NLRs that act together to activate immune responses. When these NLRs meet in the same genetic background (i.e. through crossing), they can assemble constitutively active immune complexes, leading to deleterious phenotypes associated with spontaneous cell death including dwarfism, necrosis and lethality (Barragan et al., 2019; Bomblies et al., 2007; Chae et al., 2014; Tran et al., 2017; Vaid and Laitinen, 2019) (**Figure 1.7**). This phenomenon is known as hybrid necrosis, and has major consequences in agriculture as it can restrict the crossing between elite cultivars in breeding programmes (Bomblies and Weigel, 2007; Hermsen, 1963a, b; Yamamoto et al., 2010).

Beyond implications in agriculture, the autoimmune phenotypes produced by the constitutive activation of immune complexes act as a genetic barrier between individuals harbouring incompatible resistance genes (Bomblies and Weigel, 2007). This can lead to reproductive isolation and drive speciation of plant populations with allelic immune receptors, serving as a driver of plant evolution (Bomblies et al., 2007; Bomblies and Weigel, 2007; Ispolatov and Doebeli, 2009).

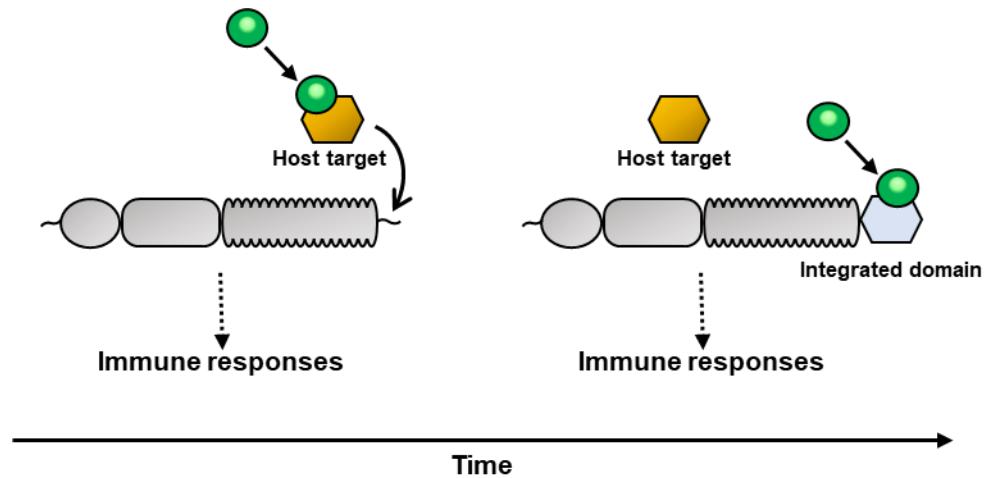


**Figure 1.7. Hybrid necrosis symptoms in wheat.** Hybrid necrosis lead to phenotypes such as dwarfism, chlorosis and spontaneous cell death. Photo credit: CIMMYT

## 1.11 Paired NLRs have unconventional domains integrated into their architecture

NLRs positioned in genetically linked pairs are prevalent in plant genomes (Stein et al., 2018; Van de Weyer et al., 2019; Wang et al., 2019c). Intriguingly, such NLR pairs often carry atypical domains in their architecture (Baggs et al., 2017). Although these atypical domains have been known since early genomic studies (Meyers et al., 2003), their functions remained unclear until recently when functional studies revealed their role in pathogen detection (Cesari et al., 2014a; Wu et al., 2015). This finding added an extra layer of complexity to understanding the NLR immune system.

Many types of unconventional domains have been found in plant NLRs, including kinases, phosphatases, thioredoxins, lectins and transcription factors domains (Bailey et al., 2018; Kroj et al., 2016; Sarris et al., 2016; Steuernagel et al., 2018; Van de Weyer et al., 2019). These are likely derived from host proteins targeted by effectors (Mukhtar et al., 2011; Weßling et al., 2014). During evolution, host targets monitored by the immune system fused to the NLR receptors architecture, becoming sensor domains with the function of binding pathogen effectors (Bialas et al., 2018; Cesari, 2018; Cesari et al., 2014a; Wu et al., 2015) (Figure 1.8). The widespread integration of atypical domains in plant NLRs, and the diversity of integration sites suggests this is a favourably selected and efficient strategy to generate receptor diversity (Bailey et al., 2018; Brabham et al., 2018; Kroj et al., 2016; Sarris et al., 2016).



**Figure 1.8. NLRs integrate domains that mimic effector virulence targets into their architecture.** Schematic representation of the evolution of integrated domains. **(Left)** A host target of pathogen effector is monitored by NLRs. **(Right)** Over the evolution, the host target fuses to the NLR architecture and acts as a binding domain for pathogen effectors.

Most functional studies have shown integrated domains directly bind, or are modified by, pathogen effectors (Guo et al., 2018b; Le Roux et al., 2015; Maqbool et al., 2015; Ortiz et al., 2017; Sarris et al., 2015; Zhang et al., 2017b). Integrated domains are the most diverse and rapidly evolving domains in NLRs (Baggs et al., 2017; Costanzo and Jia, 2010), likely driven by their interaction with pathogen effectors.

Although it is well known how certain effectors bind to NLR integrated domains (Cesari et al., 2014a; De la Concepcion et al., 2018; Guo et al., 2018b; Maqbool et al., 2015; Ortiz et al., 2017; Varden et al., 2019; Zhang et al., 2017b), the contribution of these domains to NLR post-activation mechanisms remains unknown. The structure of the ZAR1 resistosome shed some light on a possible role of integrated domains situated at the C-terminus of NLRs. As the pseudokinase RSK1 is bound to the C-terminus of ZAR1 NLR forming a pre-activation complex, it is acting as a “de facto” integrated domain (Wang et al., 2019a; Wang et al., 2019b). Therefore, C-terminal integrated domains may function in a similar manner as RSK1. However, many integrated domains are placed at the N-terminus or are even inserted between canonical domains of plant NLRs (Bailey et al., 2018; Kroj et al., 2016; Sarris et al., 2016). This suggests that there are still many activation mechanisms led by NLR integrated domains that remain to be uncovered.

To date, integrated domains have only been found in plant NLRs. However, recent reports have demonstrated a similar strategy of pathogen recognition in the mammalian NLRP1 inflammasome. NLRP1 recognizes the lethal factor protease of the anthrax pathogen and leads to immune responses (Chui et al., 2019; Sandstrom et al., 2019). This protease targets kinases implicated in host defences, however, it also cleaves an

N-terminal domain in NLRP1. This cleavage activates NLRP1 and triggers immune responses (Chui et al., 2019; Sandstrom et al., 2019) therefore, acting similarly to an integrated domain. Future functional research in mammalian NLRs may uncover other integrated domains.

## **1.12 NLR receptors are targets for engineering disease resistance**

Due to their established role in disease resistance, NLR immune receptors have been the focus of plant breeding efforts to improve disease resistance (Dangl et al., 2013).

Despite recent advances in novel technologies to rapidly clone resistance genes, breeding new disease resistance into crops, including elite cultivars, is still a slow process (Arora et al., 2019; Rodriguez-Moreno et al., 2017; Steuernagel et al., 2016). These have proven inefficient in many cases due to observations that NLRs can function in pairs and networks (Castel et al., 2019; Wu et al., 2018), and may therefore require a specific genetic context to mediate immune responses to pathogens (Chae et al., 2014; Hurni et al., 2014; Stirnweis et al., 2014a). In addition, NLRs have a very narrow recognition spectrum, recognizing race-specific effectors. In turn, rapidly evolving pathogens modify their effector repertoire and can quickly overcome disease resistance imposed by NLRs (Yoshida et al., 2016).

Synthetic biology offers new opportunities to engineer NLRs, and has emerged as a promising strategy for improving disease resistance. To date, such studies have focussed on random mutagenesis to achieve expanded effector recognition. However, this approach has been largely unsuccessful (Farnham and Baulcombe, 2006; Giannakopoulou et al., 2015; Harris et al., 2013; Segretin et al., 2014).

The recent discovery of diverse NLR integrated domains, which directly bind pathogen effectors, has opened new possibilities for disease engineering (Baggs et al., 2017; Bialas et al., 2018). Recent developments in understanding how effectors bind these domains (De la Concepcion et al., 2019; De la Concepcion et al., 2018; Guo et al., 2018b; Maqbool et al., 2015; Ortiz et al., 2017; Varden et al., 2019; Zhang et al., 2017b), combined with the new knowledge that NLR oligomerisation leads to disease resistance (Wang et al., 2019a; Wang et al., 2019b), will surely inform new strategies to engineer disease resistance in crops.

## 1.13 Rice disease resistance locus Pik encodes an NLR pair and recognizes a rice blast effector through an integrated domain

The Pik disease resistance locus has been historically important in rice breeding and was introgressed multiple times in elite rice cultivars (Kiyosawa, 1969a, b, 1978; Kiyosawa and Murty, 1969). Molecular cloning revealed that this resistance is defined by a receptor pair situated in the NLR-rich chromosome 11 (Ashikawa et al., 2008; Wang et al., 2019c). Pik sensor and helper NLRs are phylogenetically distinct (Stein et al., 2018), and are arranged in a head to head configuration with a shared promotor region (Ashikawa et al., 2008; Bialas et al., 2018).

The sensor NLR Pik-1 and helper NLR Pik-2 are CC-NLRs composed of typical CC, NB-ARC and LRR domains. However, Pik-1 harbours an integrated HMA domain located between the N-terminal CC domain and the NB-ARC (Maqbool et al., 2015). The Pik-1 HMA domain associates with the rice blast effector AVR-Pik, leading to immune responses mediated by Pik-2 (Maqbool et al., 2015).

Interestingly, an integrated HMA domain is also found in the rice NLR pair Pia, although it is positioned at the C-terminus of the sensor NLR Pia-2, after the LRR domain (Cesari et al., 2013). This domain also directly binds *M. oryzae* effectors to mediate recognition and disease resistance (Guo et al., 2018b; Ortiz et al., 2017). However, although activation mechanisms of both NLR pairs remain to be fully elucidated, they show different properties with respect to negative regulation/cooperation, as previously discussed.

The cognate Pik effector, AVR-Pik, was identified at the same time as AVR-Pia and AVR-Pii using an association genetics approach (Yoshida et al., 2009). This effector belongs to the MAX effector family (de Guillen et al., 2015; Maqbool et al., 2015). The function of AVR-Pik in promoting pathogen virulence remains unknown.

The AVR-Pik effectors exist as polymorphic variants in *M. oryzae* populations and display strong signatures of positive selection (Bialas et al., 2018; Li et al., 2019a; Longya et al., 2019; Yoshida et al., 2009). Non-synonymous mutations in AVR-Pik effectors result in amino acid changes. Some *M. oryzae* strains harbour more than one copy of polymorphic AVR-Pik effectors (Longya et al., 2019; Peng et al., 2019). This duplication and positive selection, rather than presence/absence polymorphisms displayed by many

other *M. oryzae* effectors (Yoshida et al., 2016), suggests an important role for AVR-Pik in *M. oryzae* pathology.

The amino acid polymorphisms displayed by AVR-Pik effectors are located at the interaction interface with the Pik integrated HMA. Changes in these positions weaken the interaction with the Pik NLR and allow the strain carrying AVR-Pik variants to avoid immune recognition (Kanzaki et al., 2012; Maqbool et al., 2015)

Pik NLRs are also present as an allelic series in rice. To date, several alleles with differential recognition specificities to *M. oryzae* strains harbouring AVR-Pik effectors have been cloned (Ashikawa et al., 2012; Ashikawa et al., 2008; Hua et al., 2012; Kanzaki et al., 2012; Xu et al., 2008). These alleles have mainly emerged driven by co-evolution with pathogen effectors, and mainly differ in the HMA domain of the sensor NLR Pik-1, reflecting the role of this domain in directly binding to the effectors (Costanzo and Jia, 2010; Kanzaki et al., 2012). These differences in the HMA underpin differential binding and response to AVR-Pik variants by Pik NLR alleles (De la Concepcion et al., 2018).

### **1.14 The rice NLR pair Pii monitors the host exocyst complex to mediate resistance to *M. oryzae***

Of the three rice blast effectors cloned by association genetics in 2009, AVR-Pik, AVR-Pia and AVR-Pii (Yoshida et al., 2009), AVR-Pii has remained the least studied.

AVR-Pii is a small effector that is not predicted to belong to the MAX effector family (de Guillen et al., 2015; Yoshida et al., 2009). Although the function of this effector is currently unknown, it has been shown to bind specifically to a subunit of the plant exocyst complex, Exo70 (Fujisaki et al., 2015). Exo70 proteins form a large and diverse family in plants (Cvrckova et al., 2012) and are involved in multiple processes, including plant immunity (Pečenková et al., 2011; Redditt et al., 2019; Sabol et al., 2017; Stegmann et al., 2012; Stegmann et al., 2013).

It is hypothesised that binding of AVR-Pii to Exo70 is recognized by the paired NLR Pii, leading to disease resistance (Fujisaki et al., 2015; Takagi et al., 2017), although the mechanism is also largely unknown. However, a recent study proposed that recognition of AVR-Pii is achieved by a variation of the integrated domain strategy where an atypical NOI/RIN4 domain in the NLR Pii-2 binds to the Exo70 host target to detect the effector (Fujisaki et al., 2017).

## 1.15 Aims and objectives

The discovery of atypical domains acting as pathogen sensors integrated in plant intracellular immune receptors is a major paradigm shift in plant immunity. Understanding how these domains mediate recognition of pathogen effectors and how this translates to disease resistance has major implications in plant biology and agriculture. In this work, I used the rice resistance *Pik* as a tool to gain a deeper understanding of pathogen recognition mediated by integrated domains.

In chapter 3, by combining structural biology and functional assays to study allelic integrated *Pik*-HMA domains, I uncovered how structural differences in integrated domains underpin differential recognition to pathogen effectors.

With this knowledge in hand, in chapter 4 I engineered the *Pik* integrated domain as a proof-of-concept that integrated domains can be modified to increase the recognition specificities of plant NLR immune receptors.

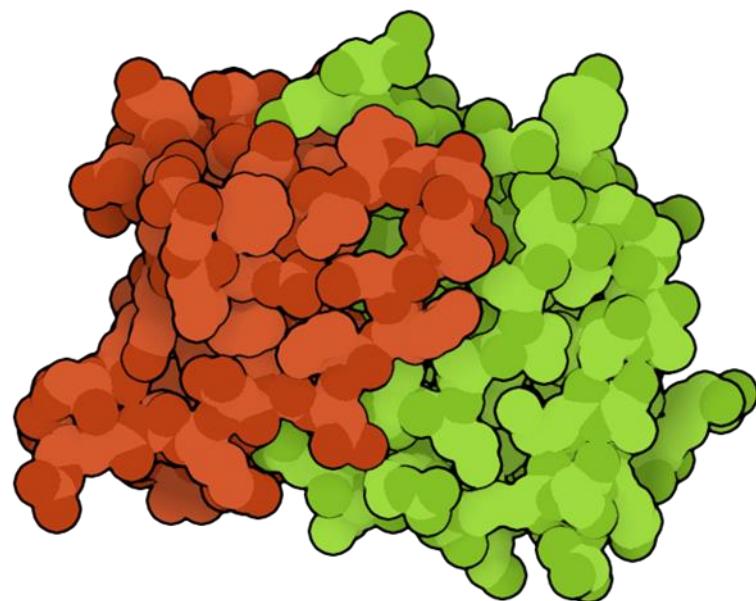
In chapter 5 and 6, I extended mechanistic research of integrated domains to an evolutionary perspective. I use the *Pik* NLR pair to understand how paired receptors evolve and co-adapt to major adaptive events such as the integration of a domain. Furthermore, I showed that integrated domains reciprocally adapt to their acceptor NLRs, possibly acquiring regulatory functions other than binding pathogen effectors.

Finally, I used the knowledge acquired in the rice *Pik* system to advance the understanding of another rice blast resistance gene, *Pii*, whose mechanism of molecular recognition of pathogens can potentially provide important insights into plant immunity and plant processes targeted by pathogens.



# 2

## Materials and methods



AVR-PikE/Pikp-HMA  
6G11



## 2. Materials and methods

### 2.1 Molecular biology

All the techniques for DNA manipulation and cloning, including isolation of plasmid DNA from bacteria, were performed according to the manufacturer's protocols. All DNA constructs were verified by sequencing.

#### 2.1.1 Polymerase chain reaction

Amplification of DNA fragments was performed using Phusion™ high-fidelity DNA polymerase (Thermo Scientific™) according to the manufacturer's protocol, allowing 30 s per kb. PCR annealing temperatures were calculated using  $T_m$  Calculator v1.12.0 (New England Biolabs).

Colony PCR was performed using MyTaq™ DNA polymerase (Bioline) using premixed buffer containing dNTPs, MgCl<sub>2</sub>, and a gel loading dye according to the manufacturer's protocol, with 30 s of amplification per kb. Single, isolated colonies were added to the mixture using a sterile micropipette tip.

#### 2.1.2 Site-directed mutagenesis

Single point mutations were introduced by Phusion PCR amplification (as above) with 5'-phosphorylated primers carrying the desired mutations. The amplification used primers running in opposite directions from the mutation site in the template plasmid. DNA templates were then eliminated by 1 h DpnI (New England Biolabs) restriction digestion at 37 °C. After PCR purification of the amplified products, the DNA sequence was re-ligated using T4 DNA ligase (New England Biolabs) according to the manufacturer's protocol. A 20 µl reaction was incubated overnight at room temperature and 5 µl were transformed into competent *E. coli* DH5α cells.

#### 2.1.3 In-Fusion™ cloning

Genes of interest were cloned into pOPIN vectors (Berrow et al., 2007) for protein production in *E. coli* or into yeast-2-hybrid (Y2H) vectors by In-Fusion™ cloning.

Sequences were amplified with primers containing overhang sequences matching the cloning plasmid. 1 µl 5x In-Fusion® HD enzyme premix was combined with 100 ng of linearized vector and 50-100 ng of purified PCR fragment in a total reaction volume of 5 µl. After incubation at 42 °C for 30 min, the reaction was transformed into *E. coli* DH5α.

### 2.1.4 TOPO® cloning

To generate custom level 0 acceptors for Golden-Gate cloning, the gene of interest was cloned by PCR amplification, flanked by selected overhang sequences and Bsal restriction sites.

A non-templated adenine was added to the 3' end of the PCR product by A-tailing using Taq DNA polymerase (New England Biolabs) according to standard protocols. 4.5 µl A-tailed DNA was combined with 0.5 µl pCRTM8/GW/TOPO® and 1 µl salt solution (1.2 M NaCl, 0.06 M MgCl<sub>2</sub>). The reaction was incubated at room temperature for 30 min and transformed into competent OneShot™ TOP10 *E. coli* cells.

### 2.1.5 Golden gate cloning

To generate constructs for transient expression in *N. benthamiana* I used golden gate cloning using standard parts for plant expression (Engler et al., 2014). All sequences were domesticated to remove Bsal and BbsI restriction sites.

Pik-1 sequences were cloned into pICH47742 with C-terminal 6xHis/3xFlag tags and Pik-2 variants were cloned into pICH47751 with C-terminal 6xHA tag. Expression of all NLRs was driven by the *A. tumefaciens* mannopine synthase (mas) promoter and terminator. AVR-Pik effectors/mutants were cloned into pICH47751 with an N-terminal 4xMyc tag, and expression was driven by the *A. thaliana* Ubi10 promoter with a 35S terminator.

### 2.1.6 Plasmid transformation into *E. coli* and *A. tumefaciens*

Chemically competent *E. coli* cells were incubated with plasmid DNA on ice for 30 min and then at 42 °C for 45 s. After heat-shock, the mixture was incubated on ice for 1 min. Cells were then recovered by the addition of 500 µl of SOC medium (20 g/L tryptone, 5 g/L yeast extract, 0.58 g/L NaCl, 0.186 g/L KCl, 2.03 g/L MgCl<sub>2</sub>, 2.46 g/L MgSO<sub>4</sub> and 3.6 g/L glucose) and growth at 37 °C with agitation (200 rpm) for 1 h. Cells were spun down and resuspended in 50 µl SOC before plating on LB (10 g/L tryptone, 5 g/L yeast extract and 10 g/L NaCl) agar plates containing appropriate antibiotics. Transformed *E. coli* could be observed as isolated colonies on the selective media after incubation at 37 °C overnight.

Electrocompetent *A. tumefaciens* cells were transformed by electroporation. Plasmid DNA was mixed with the bacteria in a pre-chilled 2 mm electroporation cuvette (Geneflow/Cell Projects). A 2500 V pulse was supplied using an Eppendorf®

Electroporator 2510. Cells were then recovered by the addition of 1 ml of LB media and incubated at 28 °C at 200 rpm for 1 h. Cells were spun down and resuspended in 50 µl LB before plating on LB agar plates containing appropriate antibiotics. Transformed *A. tumefaciens* cells could be observed after 60 h of incubation at 28 °C.

### 2.1.7 Co-transformation of yeast cells for Y2H

Desired constructs in pGBKT7 and pGADT7 were co-transformed into chemically competent *Saccharomyces cerevisiae* Y2HGold cells (Takara Bio, USA). The cells were streaked onto YPDA (10 g/L yeast extract, 20 g/L Bactopeptone, 20 g/L Glucose monohydrate, 40 mg/L Adenine hemisulfate) agar and incubated for 72 h at 30 °C. Isolated colonies were then inoculated into 10 ml of liquid YPDA media and incubated overnight in a sealed 50 ml falcon tube with agitation (250 rpm).

Saturated cultures were then diluted in fresh YPDA to an OD<sub>600</sub> of 0.4 and incubated between for 3–4 h and cells were pelleted by spinning at 1600 x g. Cells were subsequently washed twice with 20 ml of sterile H<sub>2</sub>O. Cells were then resuspended in filter sterilised Yeast Resuspension Buffer (100 mM LiAc, 50 mM Tris-HCl, 5 mM EDTA, pH 7.5) and incubated for 10 min.

10 µl of resuspended cells were mixed with 2 µl of ice-cooled denatured carrier DNA (Sigma) and 100–200 ng of appropriate bait and prey plasmids. 70 µl of filter sterilised Yeast Transformation Buffer (40% PEG 3350, 100 mM LiAc, 100 mM Tris-HCl, 10 mM EDTA, pH 7.5) was then added to the mix and incubated for 30 min at 30 °C.

After incubation, 8.8 µl of sterile DMSO was added and the mixture was incubated at 42 °C for 7 min. Yeast cells were then pelleted by centrifugation at 1000 x g and plated on SD media -Leu -Trp. Co-transformed colonies containing both pGBKT7 and pGADT7 - based constructs could typically be isolated after 72 h of incubation at 30 °C.

## 2.2 Computational methods

### 2.2.1 Protein interface analyses

Protein interface analyses were performed using QtPISA (Krissinel, 2015). For each Pik-HMA/AVR-Pik complex, one HMA/effectector assembly was used as a representative example. QtPISA interaction radars were produced as a visual representation of the interface parameters determined by the software.

The area of the polygon correlates with the likelihood that the interface constitutes part of a biological assembly. The scales along the beams compare key interface properties to statistical distributions derived from the Protein Databank. In general, if the radar area is contained within the 50% probability circle then the interface is considered superficial, and its biological relevance is questionable. In cases where the radar area is expanded outside the 50% probability circle, the interface is considered more likely to be significant and biologically relevant (Krissinel, 2015).

### 2.2.2 Root Mean Squared Deviation (RMSD) calculation

RMSDs were calculated using the align and super commands for superposition in PyMOL 2.3. For comparison of Pik HMA structures the align command was used in default mode (allowing for rejection of outliers) giving an all atom executive RMSD. For Exo70 structures the super command was used in default mode (allowing for rejection of outliers) to superpose only C-alpha chains, excluding the first five helices of the yeast and Arabidopsis structures, and the first two helices of the rice structure, to give equivalent regions in the proteins.

### 2.2.3 Phylogenetic analyses

Multiple sequence alignment of publicly available Pik-1 and Pik-2 sequences was performed in Clustal Omega (Sievers and Higgins, 2014). The phylogenetic trees of Pik NLRs were constructed using the maximum likelihood method and Tamura-Nei models (Tamura and Nei, 1993) in MEGA X (Kumar et al., 2018) with bootstrap values based on 1000 iterations. A discrete Gamma distribution (5 categories) was used to model evolutionary rate differences among sites. Phylogenetic trees were represented using Interactive Tree Of Life (iTOL) v4 (Letunic and Bork, 2019).

## 2.3 SDS-PAGE and Western blot

Pre-cast RunBlue™ TEO-Tricine SDS Mini gels (Expedeon) with 10%, 12%, 16% and a gradient of 4% to 20% were used according to the molecular weight of the proteins analysed.

Protein samples were combined with RunBlue™ 4x LDS Sample Buffer supplemented with 100 mM DTT and denatured by incubating at 95 °C for 5 min. Samples were run with 1x RunBlue™ TEO-Tricine SDS running buffer using a variable voltage according to the proteins to be resolved.

To visualize the proteins in the gel, it was incubated with InstantBlue™ Coomassie Protein Stain (Expedeon) at room temperature for 1 h. The protein dye was then rinsed with miliQ water and imaged using a GeneSyst G:Box imaging system (Syngene).

For Western-blot, proteins were transferred from the SDS-PAGE gel to a polyvinylidene difluoride (PVDF) membrane using a Trans-Blot® Turbo™ transfer system (Bio-Rad) according to the guidelines of the manufacturer. The PVDF membrane was previously activated by incubating for 1 min in methanol.

After transference of the proteins, the membrane was incubated at 4 °C for a minimum of an hour in blocking buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1 % Tween®-20, 5% w/v skimmed milk powder) with gentle shaking. After blocking, membranes were incubated overnight with the appropriate antibodies (diluted according to the manufacturer's guidelines).

After incubation with the antibodies, the membranes were washed three to four times with TBS-T buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1 % Tween®-20) for 20 min. Appropriate secondary antibodies diluted in blocking buffer were added to the membrane and incubated for 1 h at room temperature. This incubation was followed by another three to four washes with TBS-T.

To detect the presence of the proteins, 500 µl of each of the two LumiBlue ECL Extreme reagents (Expedeon) were mixed and applied to the membrane. Chemiluminescence was then detected using an ImageQuant LAS 500 spectrophotometer (GE Healthcare).

After Western blot, the total protein load of the western blot was visualized by incubating the membrane with Ponceau red stain (0.1% w/v Ponceau S in 5% v/v acetic acid). Excess stain was removed by several washes with water.

## 2.4 Protein production and purification

### 2.4.1 Protein production in *E. coli*

Pik-HMA, AVR-Pik and AVR-Pii constructs in pOPIN protein expression vectors (Berrow et al., 2007) were transformed into *E. coli* SHuffle cells (Lobstein et al., 2012) while Exo70 proteins were transformed in *E. coli* Rosetta™(DE3) cells. The bacteria were pre-inoculated in 100 ml of LB medium with appropriate antibiotics overnight. 25 ml of the culture was then diluted into 1 l of autoinduction media (AIM) (10 g/L tryptone, 5 g/L yeast extract, 3.3 g/L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 6.8 g/L KH<sub>2</sub>PO<sub>4</sub>, 7.1 g/L Na<sub>2</sub>HPO<sub>4</sub>, 0.5 g/L glucose, 2 g/L α-lactose, 0.15 g/L MgSO<sub>4</sub> magnesium sulphate and 0.03 g/L trace elements) (Studier,

## Chapter 2

2005) with appropriate antibiotics and grown in at 37 °C (30 °C for Shuffle cells) for 6 h and then 16 °C overnight.

Cells were harvested and resuspended in ice-cold lysis buffer (50 mM Tris-HCl pH 8.0, 50 mM glycine, 5% glycerol, 500 mM NaCl and 20 mM imidazole, supplemented with cComplete™ EDTA-free Protease Inhibitor Cocktail). The cells were then disrupted by sonication using a Vibra-Cell™ sonicator (SONICS) with a single 13 mm probe, with the cells chilled on ice. The sonicator was set at 40 % amplitude, with a 1 s pulse followed by a 3 s pause, for 16 min. After the first sonication cell lysate was stirred and followed by another sonication of 8 min.

The soluble fraction of the cell lysate was obtained by centrifuging for 30 min at 36,250 g at 4 °C. The supernatant was transferred to an ÄKTAXpress to carry out immobilised metal affinity chromatography (IMAC) in tandem with gel filtration. IMAC was carried out using 5 ml HisTrap™ HP NTA columns (GE Healthcare). After washing with 100 ml of washing buffer (50 mM Tris-HCl pH 8.0, 50 mM glycine, 5% glycerol, 500 mM NaCl and 20 mM imidazole), proteins were then eluted with 25 ml of elution buffer (50 mM Tris-HCl pH 8.0, 50 mM glycine, 500 mM NaCl, 500 mM imidazole, 5% (v/v) glycerol).

This elution was then loaded onto a gel filtration column equilibrated with gel filtration buffer (20 mM HEPES pH 7.5 and 150 mM NaCl). For purification of Pik-HMA, AVR-Pik and AVR-Pii proteins I used a Superdex™ 75 HiLoad™ 26/600 gel filtration column (GE Healthcare) while for Exo70 proteins I used a Superdex™ 200 HiLoad™ 26/600 column (GE Healthcare). The gel filtration buffer for Exo70 proteins was supplemented with 1 mM TCEP. Protein samples were separated by size and fractionated in 8 ml fractions that were analysed by SDS-PAGE to assess the presence of proteins.

Fractions containing the proteins of interest were pooled together and the affinity/solubility tag was removed by overnight incubation at 4 °C with recombinant human rhinovirus (HRV) 3C protease in a ratio of 10 µg of protease per mg of the purified protein. The mixture of untagged protein, tag and protease was separated by manual affinity chromatography in a 5 ml HisTrap™ HP NTA column (GE Healthcare). Purification involving a MBP solubility tag also included a MBPTrap™ HP dextrin sepharose column (GE Healthcare) attached in tandem. Approximately 20 ml of the mixture was injected on the column(s) and collected as flow-through (FT), this was followed by a wash with 50 ml of washing buffer collecting 10 ml fractions (WT 1 to 5). 20 ml of elution buffer was then injected on the columns and collected as elution (E). The different fractions were analysed by SDS-PAGE for the detection of the proteins and the

fractions of interest were pooled together and concentrated down to 5 ml using 20 ml VivaSpin® concentrators (Sartorius) with an appropriate molecular weight cut-off.

The concentrated sample was injected on an ÄKTAXpress system at 4 °C for a second gel filtration. In the case of Pik-HMA and AVR-Pik proteins the gel filtration was injected onto a Superdex™ 75 HiLoad™ 26/600 gel filtration column (GE Healthcare), for AVR-Pii proteins the column was Superdex™ 75 HiLoad™ 16/600 (GE Healthcare), whereas for Exo70 proteins I used a Superdex™ 200 HiLoad™ 16/600 column (GE Healthcare). All the columns were pre-equilibrated with gel filtration buffer (supplemented with 1 mM TCEP for Exo70 proteins).

After gel filtration, the fractions were analysed by SDS-PAGE. Fractions containing purified proteins were pooled and concentrated using 20 ml, 2 ml and 500 µl VivaSpin® concentrators (Sartorius) with an appropriate molecular weight cut-off, until reaching the desired concentrations for crystallography or biophysical assays, and flash frozen in liquid nitrogen for storage at -70 °C.

#### **2.4.2 Production of protein complexes**

To produce HMA/AVR-Pik complexes in *E. coli*. relevant Pik-HMA domains and AVR-Pik effectors combinations were co-expressed in SHuffle cells following co-transformation of pOPINM:Pik-HMA and pOPINA:AVR-Pik. Cells were grown in AIM, with antibiotic selection for both plasmids. Cells were harvested and processed as described above. As the AVR-Pik constructs in pOPINA did not have an affinity tag, only AVR-Pik effector proteins bound to Pik-HMA were recovered.

For the formation of complexes between OsExo70F-2 and AVR-Pii, each protein was purified separately and mixed on ice. After a minimum of 1 h incubation, proteins were concentrated using 2 ml and 500 µl VivaSpin® concentrators (Sartorius) with a cut-off of 30 kDa, ensuring that unbound AVR-Pii (5.675 kDa) was separated in the flow-through.

### **2.5 Protein-protein interactions**

#### **2.5.1 Yeast-2-hybrid (Y2H)**

To detect protein–protein interactions in yeast, I used the Matchmaker® Gold System (Takara Bio USA). Plasmid DNA encoding Pik-HMA or plant Exo70 proteins in pGBKT7, was co-transformed into chemically competent Y2HGold cells (Takara Bio, USA), along with pGADT7-based plasmids encoding individual AVR-Pik variants/mutants, wild type AVR-Pii and mutants or RIN4 proteins. Single colonies grown on selection plates were

## Chapter 2

inoculated in 5 ml of SD -Leu -Trp overnight at 30 °C, with 200 rpm shaking. Saturated culture was then used to make serial dilutions of OD<sub>600</sub> 1, 1<sup>-1</sup>, 1<sup>-2</sup>, 1<sup>-3</sup>, respectively. 5 µl of each dilution was then spotted on a SD -Leu -Trp plate as a growth control, and on a SD -Leu -Trp -Ade -His plate containing X-α-gal and supplemented with Aureobasidin A (Takara Bio, USA). Plates were imaged after incubation for 60–72 h at 30 °C.

To confirm protein expression in yeast, total protein extracts from transformed colonies were produced by boiling cells for 10 min in LDS Runblue® sample buffer. Samples were centrifuged and the supernatant was subjected to SDS-PAGE prior to Western blotting. The membranes were probed with anti-GAL4 DNA-BD (Sigma) for the proteins in pGBK7 and anti-GAL4 activation domain (Sigma) antibodies for the proteins in pGADT7.

### 2.5.2 In planta co-immunoprecipitation (Co-IP)

Transient gene-expression in planta for Co-IP was performed by delivering t-DNA constructs with *A. tumefaciens* GV3101 strain into 4-week old *N. benthamiana* plants grown at 22–25 °C with high light intensity. *A. tumefaciens* strains carrying the proteins of interest were mixed at 1:1 ratio to a total OD<sub>600</sub> of 0.4 (unless otherwise stated), in agroinfiltration medium (10 mM MgCl<sub>2</sub>, 10 mM 2-(N-morpholine)-ethanesulfonic acid (MES), pH 5.6), supplemented with 150 µM acetosyringone.

Leaf tissue was collected three days post infiltration (dpi), frozen, and ground to fine powder in liquid nitrogen using a pestle and mortar. Leaf powder was mixed with 2x w/v ice-cold extraction buffer (10% glycerol, 25 mM Tris pH 7.5, 1 mM EDTA, 150 mM NaCl, 2% w/v PVPP, 10 mM DTT, 1x protease inhibitor cocktail (Sigma), 0.1% Tween 20 (Sigma)) and vortexed vigorously. After centrifugation at 4,200 x g/4 °C for 20-30 min, the supernatant was passed through a 0.45 µm Minisart® syringe filter.

The presence of each protein in the input was determined by SDS-PAGE/Western blot. Pik-1 proteins were detected by probing the membrane with anti-FLAG M2 antibody (Sigma), Pik-2 proteins with anti-HA antibody 3F10 (Roche) and anti c-Myc monoclonal antibody (Santa Cruz) was used to detect AVR-Pik proteins.

For immunoprecipitation, 1.5 ml of filtered plant extract was incubated with 30 µl of M2 anti-FLAG resin (Sigma) in a rotatory mixer at 4 °C. After 3 h, the resin was pelleted (800 x g, 1 min) and the supernatant removed. The pellet was washed and resuspended in 1 ml of IP buffer (10% glycerol, 25 mM Tris pH 7.5, 1 mM EDTA, 150 mM NaCl, 0.1% Tween 20 (Sigma)) and pelleted again by centrifugation as before. Washing steps were repeated five times. Finally, 30 µl of 1:1 dilution of LDS Runblue® sample buffer and

water supplemented with 100 mM DTT was added to the agarose and incubated for 10 min at 70 °C. The resin was pelleted again, and the supernatant loaded onto SDS-PAGE gels prior to Western blotting. Membranes were probed with anti-FLAG M2 (Sigma) and anti c-Myc (Santa Cruz) monoclonal antibodies.

### **2.5.3 Analytical gel filtration (AGF)**

AGF experiments were conducted at 4 °C using a Superdex 75 10/300 GL column (GE Healthcare) equilibrated in running buffer (20 mM HEPES pH 7.5, 150 mM NaCl) for Pik-HMA/AVR-Pik interactions. Experiments with plant Exo70 proteins were carried out using a Superdex 200 10/300 GL column (GE Healthcare) equilibrated in running buffer supplemented with 1 mM TCEP. Pikm-HMA and the AVR-Pik effectors were mixed in a molar ratio of 1:1 and incubated on ice for 1 h. In each case a sample volume of 110 µl was separated at a flow rate of 0.5 ml/min. Fractions of 0.5 ml were collected for analysis by SDS-PAGE.

### **2.5.4 Isothermal titration calorimetry (ITC)**

Calorimetry experiments were performed using a MicroCal PEAQ-ITC (Malvern, UK). To test the interaction between rice Exo70 proteins and wild type or mutant AVR-Pii effectors, experiments were carried out at 25 °C using a 20 mM HEPES, pH 7.5, 150 mM NaCl buffer. The calorimetric cell was filled with 10 µM Exo70 and titrated with 100 µM AVR-Pii protein. For each ITC run, a single injection of 0.5 µL of ligand was followed by 19 injections of 2 µL each. Injections were made at 120 s intervals with a stirring speed of 750 rpm. Each experiment was repeated three times using different protein preparations, with similar results. The raw titration data were integrated and fitted to a one-site binding model using the built-in software of MicroCal PEAQ ITC.

### **2.5.5 Surface plasmon resonance (SPR)**

SPR experiments were performed on a Biacore T200 system (GE Healthcare) using an NTA sensor chip (GE Healthcare). The system was maintained at 25 °C, and a flow rate of 30 µl/min was used. All proteins were prepared in SPR running buffer (20 mM HEPES pH 7.5, 860 mM NaCl, 0.1% Tween 20).

The equilibrium dissociation constants ( $K_D$ ) for Pikm-HMA binding to AVR-Pik alleles, and Pikp-HMA binding to AVR-PikD, were determined from multicycle kinetics curves using the Biacore T200 BioEvaluation software (GE Healthcare), with a 1:1 or 2:1 fit model respectively. For each kinetic cycle, 30 µl of 0.5 mM NiCl<sub>2</sub> was used to activate

the chip and, after washing, His-tagged AVR-Pik variants were immobilised on the sample cell, giving a response level of  $240 \pm 50$  response units (RUs). Concentrations of Pik-HMA (ranging from 0–200 nM for PikmHMA, and 0–100 nM for Pikp-HMA) were injected over both reference and sample cells for 360 s, and dissociation was recorded for further 120 s. The sensor chip was regenerated between each cycle with 30  $\mu$ l of 0.35 M EDTA. For each measurement, in addition to subtracting the response in the reference cell, a further buffer-only subtraction was made to correct for bulk refractive index changes or machine effects.

For the interactions where it was not possible to accurately determine the  $K_D$  due to the insufficient quality of the data, the level of binding was expressed as a percentage of the theoretical maximum response ( $R_{max}$ ) normalized for the amount of ligand immobilized on the chip. For this, only 3 concentrations (4, 40 and 100 nM) of HMA domain were used. For calculations the binding model to AVR-Pik effectors was 1:1 for Pikm-HMA and 2:1 (HMA/AVR-Pik) for Pikp-HMA and Pikp-HMA<sup>NK-KE</sup>.

SPR data were exported and plotted using R v3.4.3 (<https://www.r-project.org/>) and the function ggplot2 (Wickham, 2009) or Microsoft Excel. Each experiment was repeated a minimum of three times, including internal repeats, with similar results. The proteins used came from three independent preparations for the HMA domains and two independent preparations of the AVR-Pik effectors.

## 2.6 Spectroscopic techniques

### 2.6.1 Small-angle X-ray scattering (SAXS)

SAXS data were collected at the Diamond Light Source, UK, beamline B21. Measurements were made at an energy of 12.4 keV, camera length of 4.018 m, and  $q$  range 0.004–3.8 nm<sup>-1</sup>. Measurements for 60  $\mu$ l of protein solution at ~10 mg/ml were made in HPLC mode using a Superdex™ 200 increase 3.2/300 (GE Healthcare).

Preliminary analysis of the data was carried out using the ATSAS package (Franke et al., 2017). The AtExo70A-1 *ab initio* model was calculated using DAMMIF and the AtExo70A-1 structure body was overlaid with the *ab initio* models using ChimeraX (Goddard et al., 2018).

### 2.6.2 X-ray crystallography

Sitting drop, vapor diffusion crystallization trials were set up in 96 well plates, using an Oryx nano robot (Douglas Instruments, United Kingdom). Plates were incubated at 20

## Chapter 2

°C, and crystals typically appeared after one or two days for Pik-HMA/AVR-Pik complexes and between four to six days for OsExo70F-2/AVR-Pii. For data collection, all crystals were harvested from the screen, and flash-frozen in liquid nitrogen. Crystals obtained in Morpheus® HT-96 screen did not need cryo-protection while the crystals for OsExo70F-2/AVR-Pii were imbibed in the crystallization solution supplemented with 15% ethylene glycol.

Crystals used for data collection appeared in the following conditions:

- (1) Pikm-HMA/AVR-PikD (12.5 mg/ml), Morpheus® HT-96 condition E10 [0.12 M Ethylene glycols (0.3 M Diethylene glycol; 0.3 M Triethylene-glycol; 0.3 M Tetraethylene glycol; 0.3 M Pentaethylene glycol); 0.1 M Buffer system 3 (1 M Tris (base); BICINE) pH 8.5; 50% v/v Precipitant mix 2 (40% v/v Ethylene glycol; 20% w/v PEG 8000)].
- (2) Pikm-HMA/AVR-PikE (6.2 mg/ml), Morpheus® HT-96 condition H6 [0.1 M Amino acids (0.2 M L-sodium glutamate; 0.2 M Alanine (racemic); 0.2 M Glycine; 0.2 M Lysine HCl (racemic); 0.2 M Serine (racemic)); 0.1 M Buffer system 2 (1 M sodium HEPES, MOPS (acid)) pH 7.5; 50% v/v Precipitant mix 2 (40% v/v Ethylene glycol; 20% w/v PEG 8000)].
- (3) Pikm-HMA/AVR-PikA (9.5 mg/ml), Morpheus® HT-96 condition D18 [0.12 M Alcohols (0.2 M 1,6-Hexanediol; 0.2 M 1-Butanol; 0.2 M 1,2-Propanediol; 0.2 M 2-Propanol; 0.2 M 1,4-Butanediol; 0.2 M 1,3-Propanediol); 0.1 M Buffer system 2 (1 M sodium HEPES, MOPS (acid)) pH 7.5; 50% v/v Precipitant mix 4 (25% v/v MPD; 25% v/v PEG 1000; 25% v/v PEG3350)].
- (4) Pikp-HMA/AVR-PikD (15 mg/ml), Morpheus® HT-96 condition D4 [0.12 M Alcohols (0.2 M 1,6-Hexanediol; 0.2 M 1-Butanol; 0.2 M 1,2-Propanediol; 0.2 M 2-Propanol; 0.2 M 1,4-Butanediol; 0.2 M 1,3-Propanediol); 0.1 M Buffer system 1 (1 M Imidazole; MES monohydrate (acid)) pH 6.5; 50% v/v Precipitant mix 4 (25% v/v MPD; 25% v/v PEG 1000; 25% v/v PEG3350)].
- (5) Pikp-HMA/AVR-PikE (28.4 mg/ml), Morpheus® HT-96 condition A4 [0.06 M Divalent (0.3 M Magnesium chloride hexahydrate; 0.3 M Calcium chloride dihydrate); 0.1 M Buffer system 1 (1 M Imidazole; MES monohydrate (acid)) pH 6.5; 50% v/v Precipitant mix 4 (25% v/v MPD; 25% v/v PEG 1000; 25% v/v PEG3350)].
- (6) Pikp-HMA<sup>NK-KE</sup>/AVR-PikD (10 mg/ml), Morpheus® HT-96 condition D4 [0.12 M Alcohols (0.2 M 1,6-Hexanediol; 0.2 M 1-Butanol; 0.2 M 1,2-Propanediol; 0.2 M 2-Propanol; 0.2 M 1,4-Butanediol; 0.2 M 1,3-Propanediol); 0.1 M Buffer system 1 (1 M

## Chapter 2

Imidazole; MES monohydrate (acid)) pH 6.5; 50% v/v Precipitant mix 4 (25% v/v MPD; 25% v/v PEG 1000; 25% v/v PEG3350)].

**(7)** Pikp-HMA<sup>NK-KE</sup>/AVR-PikE (15 mg/ml), Morpheus<sup>®</sup> HT-96 condition A8 [0.06 M Divalent (0.3 M Magnesium chloride hexahydrate; 0.3 M Calcium chloride dihydrate); 0.1 M Buffer system 2 (Sodium HEPES; MOPS (acid)) pH 7.5; 37.5% v/v Precipitant mix 4 (25% v/v MPD; 25% v/v PEG 1000; 25% v/v PEG3350)].

**(8)** Pikp-HMA<sup>NK-KE</sup>/AVR-PikC (13 mg/ml), Morpheus<sup>®</sup> HT-96 condition F8 [0.12 M Monosaccharides (0.2 M D-Glucose; 0.2 M D-Mannose; 0.2 M D-Galactose; 0.2 M D-Fucose; 0.2 M D-Xylose; N-Acetyl-D-Glucosamine); 0.1 M Buffer system 2 (Sodium HEPES; MOPS (acid)) pH 7.5; 37.5% v/v Precipitant mix 4 (25% v/v MPD; 25% v/v PEG 1000; 25% v/v PEG3350)].

**(9)** OsExo70F-2/AVR-Pii (6 mg/ml), [0.3 M Ammonium iodide; 30% v/v PEG3350)].

X-ray data sets were collected at the Diamond Light Source (Oxford, UK). The data were processed using the xia2 pipeline (Winter, 2010) and AIMLESS (Evans and Murshudov, 2013), as implemented in CCP4 (Winn et al., 2011). Pik-HMA/AVR-Pik structures were solved by molecular replacement using PHASER (McCoy et al., 2007) and the Pikp-HMA/AVR-PikD structure (Maqbool et al., 2015). For OsExo70F-2/AVR-Pii, a model of Exo70-F2 was positioned in the electron density by molecular replacement using the crystal structure of AtExo70A-1 (Zhang et al., 2016) with subsequent phasing calculations made by Professor David Lawson (JIC crystallography platform).

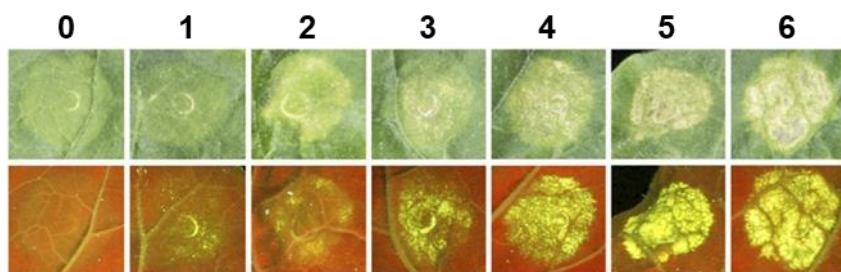
Final structures were obtained through iterative cycles of manual rebuilding and refinement using COOT (Emsley et al., 2010) and REFMAC5 (Murshudov et al., 2011), as implemented in CCP4 (Winn et al., 2011). Structures were validated using the tools provided in COOT and MOLPROBITY (Chen et al., 2010).

## 2.7 Cell death assays

*A. tumefaciens* GV3101 cells, carrying the relevant Pik-1 and Pik-2 proteins were resuspended in induction media (10 mM MES pH 5.6, 10 mM MgCl<sub>2</sub> and 150 μM acetosyringone) and mixed with either AVR-Pik effectors, mCherry or empty vector, and P19 at OD<sub>600</sub> 0.4, 0.4, 0.6 and 0.1, respectively. 4-week old *N. benthamiana* plants were spot-inoculated using a needless syringe. Leaves were collected at 5 dpi to measure UV autofluorescence or ion leakage as a proxy for cell death.

### 2.7.1 Cell death scoring: UV autofluorescence

Detached leaves were imaged at 5 dpi from the abaxial side of the leaves for UV fluorescence images. Photos were taken using a Nikon D4 camera with a 60 mm macro lens, ISO set 1600 and exposure ~10 s at F14. The filter was a Kodak Wratten No. 8 and white balance was set to 6250 degrees Kelvin. Blak-Ray® longwave (365 nm) B-100AP spot light lamps were moved around the subject during the exposure to give an even illumination. Each spot was scored from 0 to 6 using a cell death scoring scale (**Figure 2.1**). Scoring data for each sample was plotted as dotplots using R v3.4.3 (<https://www.r-project.org/>) and the graphic package ggplot2 (Wickham, 2009). The size of the centre dot at each cell death value is directly proportional to the number of replicates in the sample with that score. All individual data points are represented as dots.



**Figure 2.1. Cell death scoring scale for UV quantification.** Example images used for scoring Pik mediated cell death in *N. benthamiana*. Top row shows the infiltrated spot in the adaxial part of the leave under daylight, bottom row shows the corresponding spot from the abaxial side of the leave under UV light.

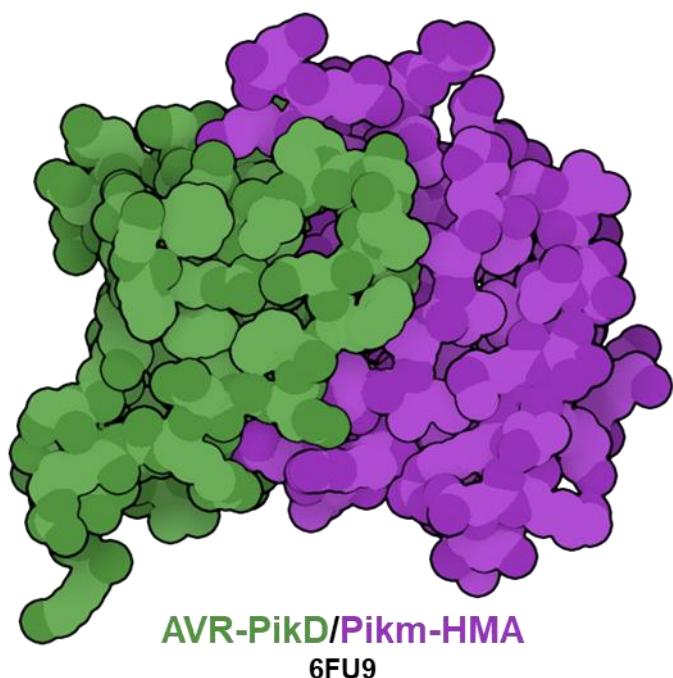
### 2.7.2 Cell death scoring: Ion leakage

For ion leakage quantification, plants were infiltrated with the relevant constructs on two different leaves. After 5 dpi, leaves were detached and two leaf discs with a diameter of 8 mm (one disc per leaf spot) were collected and floated in 1.5 mL of Milli-Q water. Conductivity ( $\mu\text{S}/\text{cm}$ ) was measured immediately after transferring the leaf disc to water (time zero) using a LAQUAtwin EC-33 conductivity meter (Horiba UK Ltd). Leaf discs were then incubated for 6 h at room temperature with gentle shaking before measuring the final conductivity. Conductivity data for each sample was exported and plotted using R v3.4.3 (<https://www.r-project.org/>) and the function ggplot2 (Wickham, 2009).

## Chapter 2

# 3

## **Polymorphic residues in an allelic NLR expand binding and response to effectors of the blast pathogen**



## Chapter 3

### 3. Polymorphic residues in an allelic NLR expand binding and response to effectors of the blast pathogen

#### Abstract

Due to the accelerated co-evolution between plants and pathogens, plant NLR receptors that directly engage with pathogen effectors often occur as allelic series with differential pathogen specificities. The determinants of this specificity remain largely unknown and its understanding may lead to the design of immune receptors with improved capacities. Here, I combined molecular biology and X-ray crystallography to understand the basis of the differences in pathogen specificity between two alleles of the NLR pair Pik. Rice cultivars with the NLR allele Pikm are resistant to blast strains carrying three AVR-Pik effector variants, whilst plants with the allele Pikp respond to only one. Immune specificity is encoded at the integrated HMA domain of Pik-1. Variation at the binding interface with the effector enable Pikm-HMA to bind the three recognised effectors with high affinity, whereas Pikp-HMA only binds one. By obtaining five high resolution structures of complexes between Pik-HMA alleles and different AVR-Pik variants, I could uncover the structural determinants of expanded effector recognition by Pikm. This work reveals how co-evolution shapes the response profiles of an NLR receptor and how the emergence of new immune specificities is driven by natural selection.

## Chapter 3

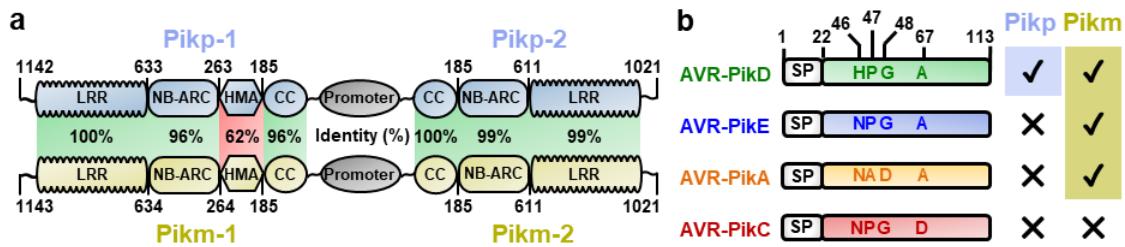
### 3.1 Introduction

Perception of pathogen effectors inside cells by NLR immune receptors, and the subsequent response triggered by this recognition event, imposes a high selection pressure in the pathogen. This process drives the emergence of new effector variants that escape NLR recognition. As a consequence, pathogen effectors often show strong signatures of selection, including high levels of non-synonymous polymorphisms and presence/absence profiles (Allen et al., 2004; Dodds et al., 2006; Raffaele et al., 2010; Yoshida et al., 2016).

The rice blast effector AVR-Pik is a remarkable example of an effector with extreme signatures of positive selection. To date, all known nucleotide polymorphisms found in this effector result in amino acid changes (Bialas et al., 2018; Huang et al., 2014; Longya et al., 2019; Yoshida et al., 2009). These newly evolved polymorphisms allow rice blast strains carrying these AVR-Pik variants to escape from immune recognition (Kanzaki et al., 2012; Longya et al., 2019).

The rice NLR Pik recognizes AVR-Pik by direct binding via the integrated HMA domain in the sensor NLR Pik-1 (**Figure 3.1a**), resulting in disease resistance (Maqbool et al., 2015). The Pik NLRs are present as an allelic series with differential immune specificities for *M. oryzae* strains. The emergence of Pik alleles has likely been driven by co-evolution with pathogen effectors (Kanzaki et al., 2012). Consistent with being the region that binds the effectors, the integrated HMA domain is the most polymorphic region in the Pik NLR alleles (Costanzo and Jia, 2010) (**Figure 3.1a**).

Rice plants with the NLR allele Pikp are resistant to *M. oryzae* strains carrying the effector variant AVR-PikD, but rice cultivars expressing the allele Pikm respond to strains with AVR-PikD, AVR-PikE, or AVR-PikA (Kanzaki et al., 2012) (**Figure 3.1b**). Previous work established the structural basis of AVR-PikD recognition by Pikp (Maqbool et al., 2015). Furthermore, the crystal structure of Pikp-HMA bound to AVR-PikD shows that effector polymorphisms map to the receptor/effector binding interface. These polymorphisms weaken the interaction of the AVR-Pik variants and confer evasion of Pikp immune recognition (Maqbool et al., 2015), confirming they are adaptive. However, it was unknown how Pikm can extend the recognition to these effectors. Given that Pikp and Pikm NLR alleles mostly differ in the HMA domain of the sensor NLR Pik-1 (**Figure 3.1a**), this is an excellent system for understanding the mechanistic basis of allelic variation in effector recognition specificities by plant NLRs.



**Figure 3.1. Diversity of Pik NLR and AVR-Pik effectors.** (a) Schematic representations of Pik NLR alleles. The sensor NLR (Pikp-1) and helper NLR (Pikp-2) share a common promoter and the same overall domain architecture. Pikp-1/Pikp-2 (top) are shown in ice blue, and Pikm-1/Pikm-2 (bottom) are shown in gold. Pairwise protein sequence identity between each domain is indicated, highlighting diversification of the integrated HMA domain. (b) Schematic representations of AVR-Pik variants. Amino acid polymorphisms are shown as a single letter code along with their Pikp- or Pikm-mediated response profiles in rice (Kanzaki et al., 2012). SP indicates Signal Peptide.

In this chapter, I aimed to understand the basis of the extended recognition phenotype in the NLR allele Pikm. In turn, this would reveal how host-pathogen co-evolution has driven the emergence of new receptor specificities. I combined the reconstruction of Pik-HMA/AVR-Pik protein complexes in vitro and the recapitulation of Pik immune responses in *N. benthamiana*, to correlate immune activation and receptor/effectector binding affinities. Using X-ray crystallography, I obtained the structures of Pikp-HMA in complex with the recognised effector AVR-PikD and with the unrecognised effector AVR-PikE. Comparison of these two structures uncovered the atomic details of how polymorphic effectors escape NLR immune recognition, deepening the understanding of the arms-race between plants and pathogens. Furthermore, I obtained the structures of Pikm-HMA in complex with AVR-PikD, AVR-PikE and AVR-PikA, which allowed me to define the interaction interfaces involved in extended pathogen recognition phenotypes by Pikm. This knowledge has the potential to contribute to the rational design of plant NLRs with improved recognition capabilities.

### 3.2 Pikm-mediated cell death in *N. benthamiana* recapitulates the rice allele-specific immune responses to AVR-Pik effectors

*N. benthamiana* is an established heterologous system for the investigation of effector recognition by plant NLRs from multiple pathosystems (Farnham and Baulcombe, 2006; Segretin et al., 2014). It was previously established that the NLR pair Pikp triggers a strong cell death response in *N. benthamiana* upon recognition of the rice blast effector AVR-PikD (Maqbool et al., 2015), leading to the accumulation of phenolic compound that

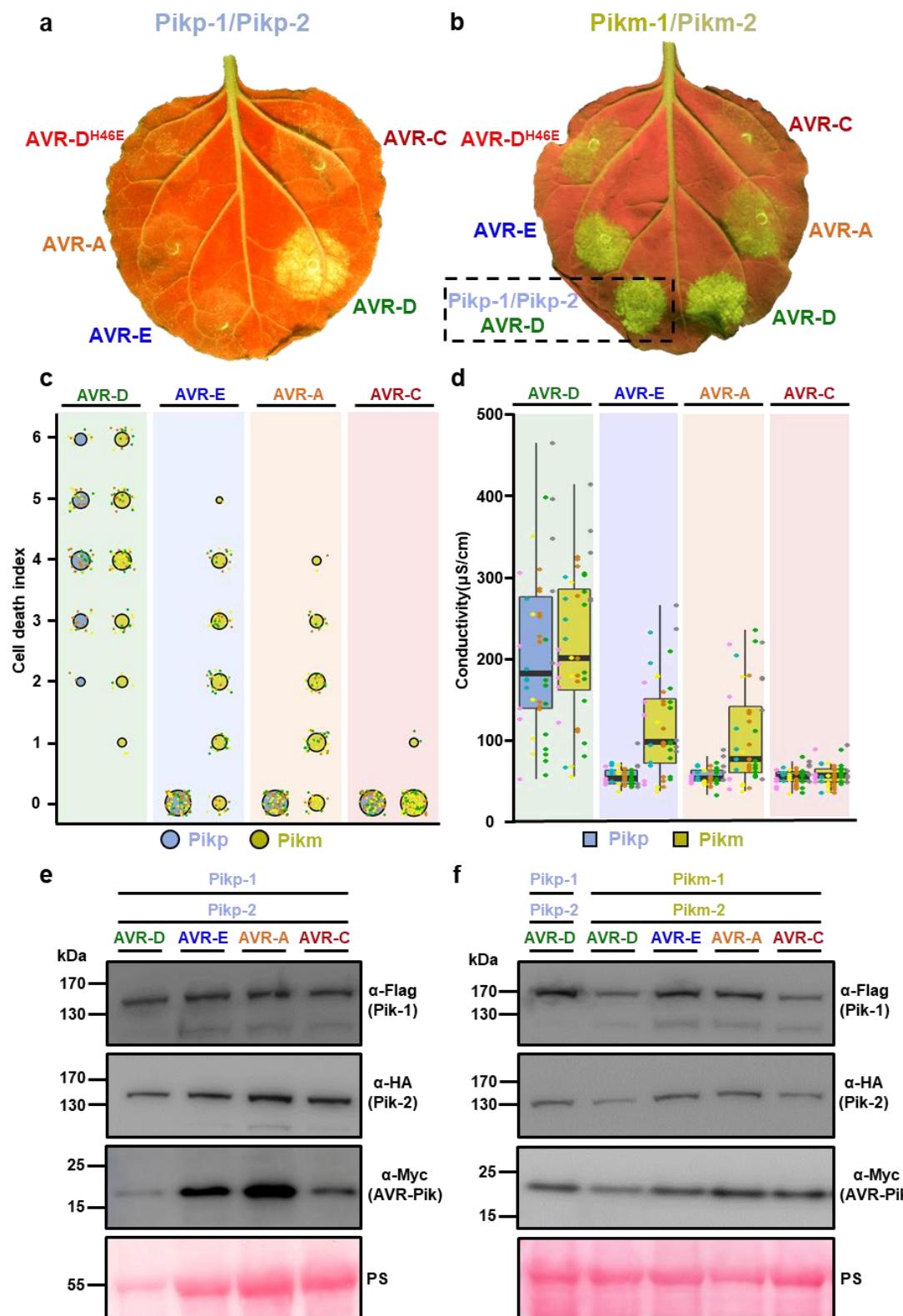
are fluorescent under UV light (Bos et al., 2006). This response phenocopies effector variant-specific resistance in rice, with *Pikp* responding to AVR-PikD, but not AVR-PikE, AVR-PikA, or AVR-PikC (Bialas et al., 2018; Kanzaki et al., 2012). I aimed to test whether this system is suitable for measuring *Pikm*-mediated cell death responses. To this end, I cloned the AVR-Pik effectors variants in golden gate compatible vectors (Engler et al., 2014) to standardize the assay into a more amenable system for high-throughput molecular cloning (see materials and methods), and optimised the cell death assay for these new constructs.

To allow for direct comparison with previously published data, I repeated the assay using the golden gate compatible *Pikp* NLRs generated by Hannah Langlands and the effector variants AVR-PikD, AVR-PikE, AVR-PikA and AVR-PikC in the new golden gate vectors and I scored the cell death intensity in the infiltrated area under UV light. I obtained equivalent results to those previously published (Maqbool et al., 2015), with *Pikp* triggering a cell death response in *N. benthamiana* when co-expressed with AVR-PikD but not AVR-PikE, AVR-PikA or AVR-C (**Figure 3.2a, c**).

To test whether the *N. benthamiana* system recapitulates *Pik* allelic-specific responses to AVR-Pik effectors, I co-transformed the NLR pair *Pikm* with each of the AVR-Pik effector variants. *Pikm* triggered cell death responses to each of AVR-PikD, AVR-PikE, or AVR-PikA, but not to AVR-PikC. (**Figure 3.2b, c**). To avoid subjectivity in the quantification of the cell death, I also performed ion leakage quantification of the cell death in *N. benthamiana*. In this assay, conductivity is measured as a proxy of cell integrity and increased conductivity is correlated with a higher number of cells undergoing cell death (Greenberg and Ausubel, 1993). Quantification of ion leakage also shows a differential cell death response between *Pikp* and *Pikm*, with *Pikm* triggering cell death in the presence of AVR-PikD, AVR-PikE and AVR-PikA, while *Pikp* only does in the presence of AVR-PikD (**Figure 3.2d**). These results match the response of rice cultivars expressing *Pikm* to *M. oryzae* strains encoding the AVR-Pik effectors variants (Kanzaki et al., 2012) (**Figure 3.1b**).

Interestingly, there is a qualitative hierarchy in the level of *Pikm*-mediated cell death in response to the effectors in the order AVR-PikD>AVR-PikE>AVR-PikA (**Figure 3.2c, d**). The expression of each protein was also confirmed by western blot for both *Pikp* and *Pikm* cell death assays (**Figure 3.2e, f**).

These results confirmed that *N. benthamiana* is an amenable system to perform high-throughput cell death assays in different *Pik* NLR alleles.



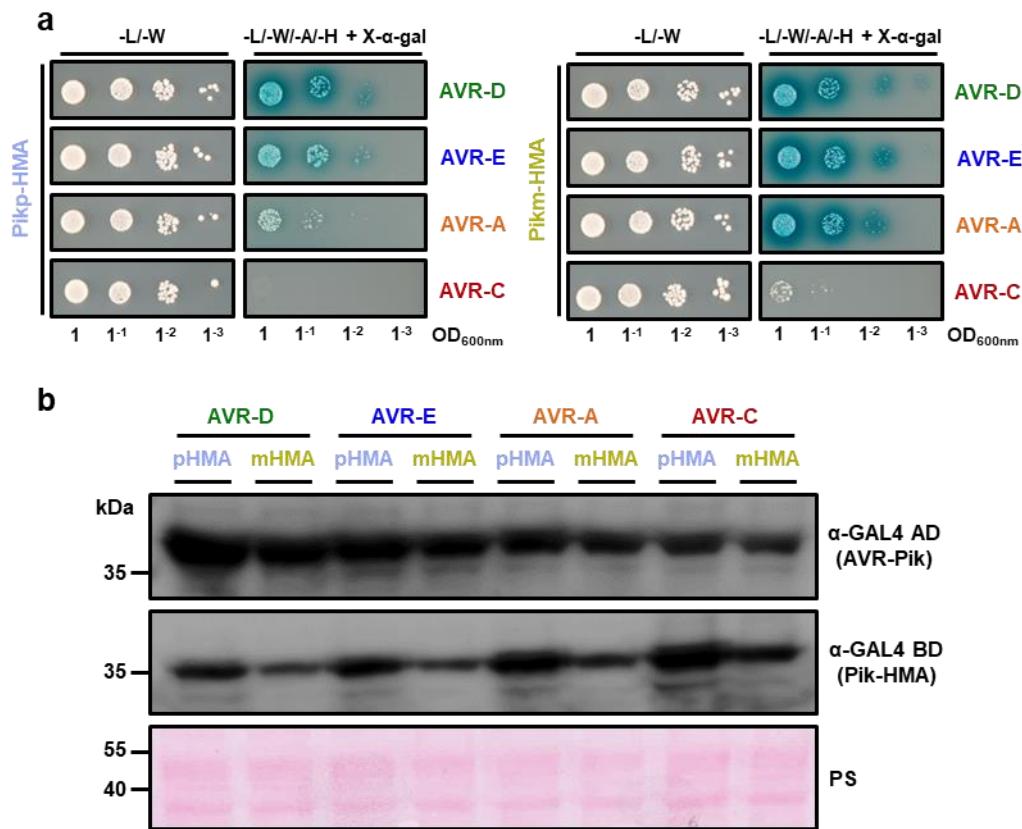
**Figure 3.2. The Pik-mediated cell death response to AVR-Pik effector variants in *N. benthamiana* phenocopies the allelic resistance specificity in rice.** (a, b) Representative leaf images depicting Pikp (a) and Pikm (b) -mediated cell death as autofluorescence under UV-light. Pikp-mediated cell death with AVR-PikD is included as a positive control in the Pikm cell death experiment (surrounded by dashed black line, Pikm-1/Pikm-2 was not infiltrated in this spot). (c) Cell death assay scoring represented as dot plots comparing Pikp and Pikm. The number of repeats was 80 and 90, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score. (d) Ion leakage assay comparing Pikp and Pikm-mediated cell death responses in *N. benthamiana*. The centre line represents the median, the box limits are the upper and lower quartiles, the whiskers extend to the largest value within  $Q1 - 1.5 \times \text{IQR}$  and the smallest value within  $Q3 + 1.5 \times \text{IQR}$ . All the data points are represented as dots with distinct colours for each biological replicate. For each experiment, I performed six biological replicates with 5 or 10 internal repeats, a total of 40 replicates. (e, f) Western blots showing accumulation of AVR-Pik variants, Pikp (e) and Pikm (f) NLRs, Pikp paired NLRs are included as controls in the western blots for Pikm experiments.

### 3.3 Allele-specific effector responses in planta correlate with direct Pik-HMA interactions

I sought to investigate if the differential response profile in planta correlates with a differential binding of the integrated HMA domains of Pikp and Pikm (hereafter Pikp-HMA and Pikm-HMA, respectively) to the AVR-Pik effectors. To this end, I carried out a yeast-2-hybrid (Y2H) assay with Pikp-HMA and Pikm-HMA, and each of the AVR-Pik effectors (**Figure 3.3**).

For Pikm-HMA, there was a comparable growth of yeast on selective plates, and the development of blue colouration with X- $\alpha$ -gal (both indicative of protein:protein interactions) for AVR-PikD, AVR-PikE, and AVR-PikA, but not AVR-PikC (**Figure 3.3a**). While the Y2H assay showed comparable interaction of Pikm-HMA and Pikp-HMA to AVR-PikD, Pikm-HMA had increased interaction with AVR-PikE and a markedly stronger interaction with AVR-PikA (**Figure 3.3a**). No growth was observed with Pikp-HMA and AVR-PikC. All proteins were confirmed to be expressed in yeast by Western blot (**Figure 3.3b**).

This result suggests a correlation between the Pikm extended cell death responses in planta and increased binding of the AVR-Pik effectors to the Pikm-HMA domain in vivo.



**Figure 3.3. *Pikm-HMA* shows increased binding to AVR-Pik variants in vivo.** (a) Yeast-Two-Hybrid assay of *Pikp-HMA* and *Pikm-HMA* with AVR-Pik alleles. Control plate for yeast growth is on the left with quadruple dropout media supplemented with X- $\alpha$ -gal on the right. Growth and development of blue colouration in the selection plate are both indicative of a protein:protein interaction. HMA domains were fused to the GAL4 DNA binding domain, and AVR-Pik alleles to the GAL4 activator domain. Each experiment was repeated a minimum of three times, with similar results. (b) Western blot analysis confirming accumulation of proteins in yeast. Yeast lysate was probed for the expression of the HMA domain with anti-GAL4 DNA binding domain (BD) and AVR-Pik effectors anti-GAL4 activation domain (AD).

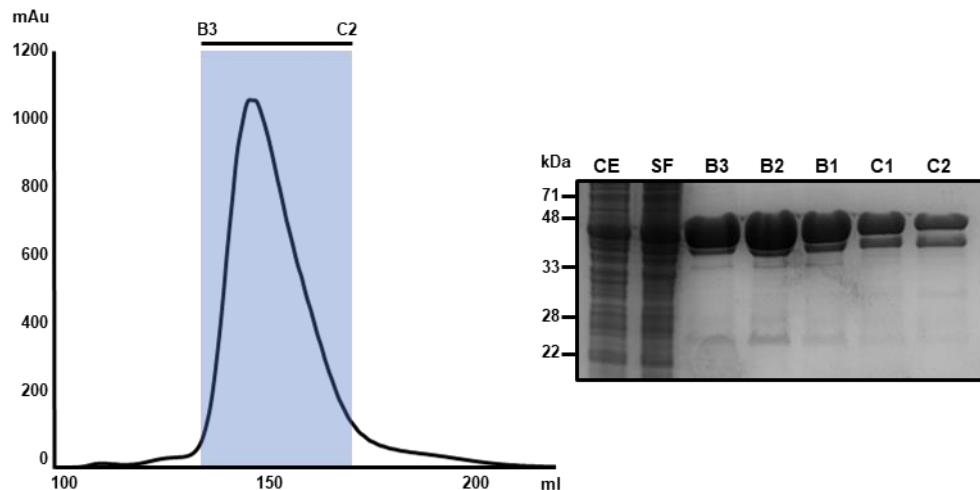
### 3.4 *Pikm-HMA* can be produced and purified from *E. coli*

To enable the study of the AVR-Pik/Pik-HMA interactions in vitro I produced stable *Pikm-HMA* protein in *E. coli* (see details in materials and methods). For this, I used a *Pikm-HMA* construct in the pOPINM vector produced by Dr. Marina Franceschetti. Compared to the previously studied *Pikp-HMA* (Maqbool et al., 2015), this construct has a 5-amino acid extension at the C-terminus (encompassing residues Gly186 - Asp264 of the full-length protein) that enable the heterologous production of soluble protein in *E. coli*.

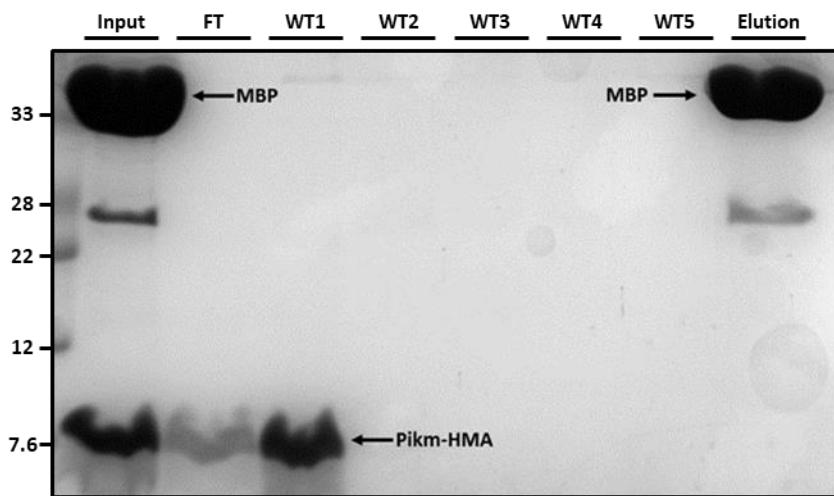
This construct yields a soluble *Pikm-HMA* protein with a N-terminal hexa-histidine (6xHis) tag adjacent to a maltose binding protein (MBP) tag. The histidine tag enables the separation of the fusion protein from the *E. coli* cell lysate by immobilised metal affinity chromatography (IMAC), with further purification achieved by gel filtration in a size

exclusion chromatography column. This process was carried out in tandem using ÄKTAxpress and the elution of the 6xHis:MBP:Pikm-HMA fusion protein could be followed by the UV absorption trace (**Figure 3.4**). Eluted fractions were evaluated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), revealing a protein band close to the expected size for the 6xHis:MBP:Pikm-HMA fusion protein (**Figure 3.4**). These fractions were pooled and concentrated before cleaving the 6xHistidine:MBP tag using 3C protease (see materials and methods). After this, Pikm-HMA can be efficiently separated from MBP tag using HisTrap™ and MBPTrap™ columns attached in tandem (**Figure 3.5**). Pikm-HMA was mainly present in the flow-through (FT) and wash-through (WT) from the columns (**Figure 3.5**). These fractions were combined and concentrated to perform a final gel filtration step. This purification cannot be followed by the UV absorption trace due to the lack of aromatic residues in Pikm-HMA, therefore I analysed all the fractions collected by SDS-PAGE (**Figure 3.6**). Fractions containing purified Pikm-HMA protein were combined and concentrated before performing in vitro experiments.

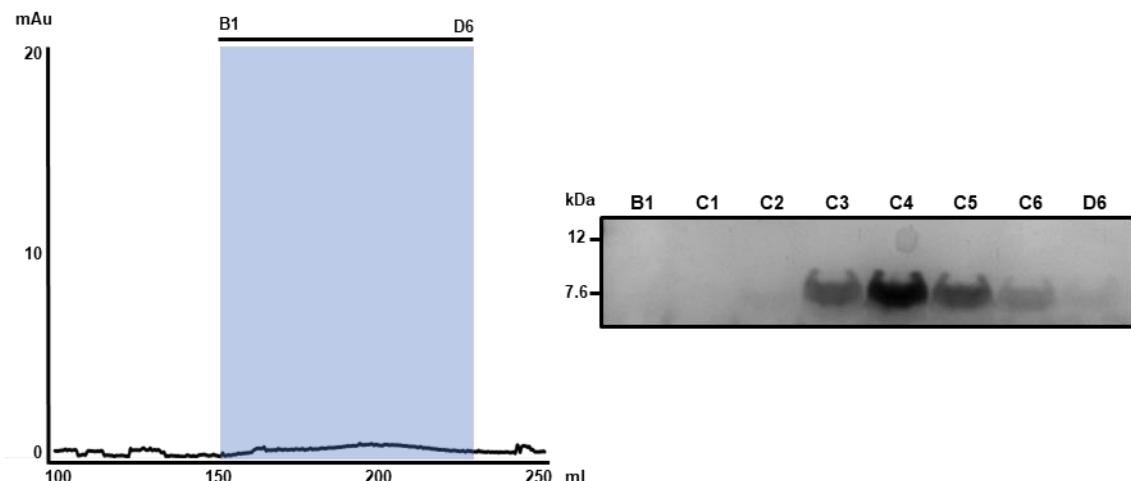
I also cloned the equivalent Pikp-HMA construct (encompassing residues Gly186 - Asp263 of the full-length protein), which also yielded a stable and soluble protein that could be produced and purified using the same method as for Pikm-HMA.



**Figure 3.4. Pikm-HMA can be efficiently purified by IMAC + Gel filtration.** Elution trace of MBP:Pikm-HMA after IMAC + gel filtration. Fractions from B3 to C2 were collected and analysed by SDS-PAGE together with bacterial crude extract (CE) and the soluble fraction (SF) for comparison.



**Figure 3.5. 6xHis:MBP tag can be efficiently cleaved.** SDS-PAGE analysis of fractions collected in Pikm-HMA HisTrap™ and MBPTrap™ tandem purification following 3C protease treatment. The protein treated with 3C protease before purification is shown as an input. Pikm-HMA was successfully purified as a single band in the flow-through (FT) and the first wash-through with A4 buffer (WT1). MBP tag eluted in the elution treatment with B1 buffer containing imidazole.

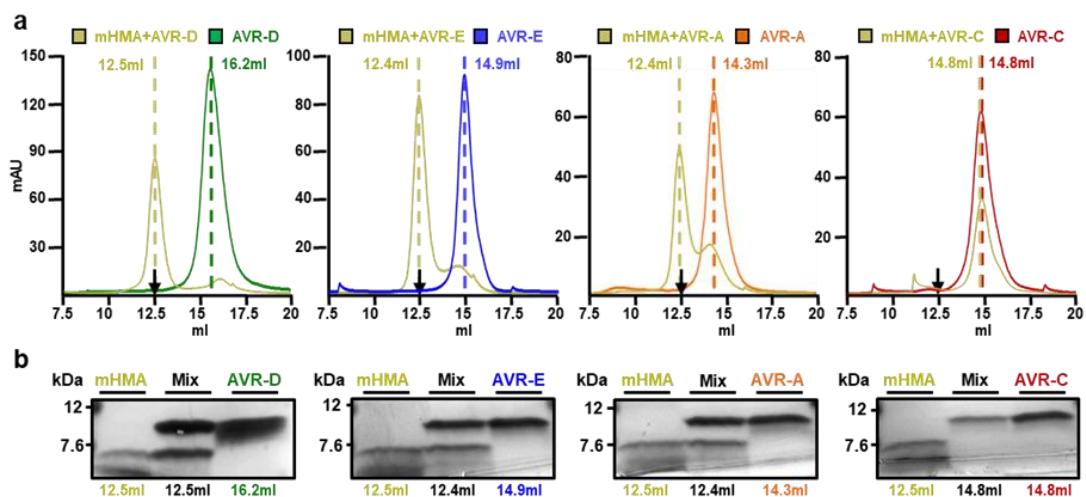


**Figure 3.6. Final gel filtration step yields a pure Pikm-HMA domain.** Elution trace of Pikm-HMA in gel filtration. The absorbance values are almost zero due the lack of aromatic residues. SDS-PAGE analysis of the collected fractions showed Pikm-HMA elutes between the fractions C3 and C6.

### 3.5 Pikm-HMA has tighter binding affinities for AVR-Pik effectors than Pikp-HMA in vitro

To qualitatively evaluate the formation of protein complexes between Pikm-HMA and each of the AVR-Pik effector variants in vitro, I used analytical gel filtration (AGF) by mixing purified Pikm-HMA and AVR-Pik proteins purified as previously published (Maqbool et al., 2015) and observing the volumes at which they elute (see Materials and methods for details). Due to the lack of aromatic residues in Pikm-HMA the elution of this protein cannot be followed by UV detection, therefore I collected all the fraction from AGF and I analysed them by SDS-PAGE. Pikm-HMA was consistently observed in a fraction corresponding to an elution volume of 12.5 ml. Interestingly, this corresponds to a particle size higher than expected for Pikm-HMA. I also ran each of the purified AVR-PikD, AVR-PikE, AVR-PikA and AVR-PikC effector proteins, which were observed at elution volumes of 14.3 to 16.2 ml by UV detection.

In the next step, I mixed the purified Pikm-HMA protein with separately purified AVR-Pik effectors and I incubated them for one hour on ice. Pikm-HMA formed complexes with AVR-PikD, AVR-PikE, or AVR-PikA (as depicted for the earlier elution volume compared with the elution of the effector on its own). Pikm-HMA did not interact with AVR-PikC (**Figure 3.7a**). The fraction corresponding to each peak was analysed by SDS-PAGE to confirm the presence of the proteins (**Figure 3.7b**). These results confirm that Pikm-HMA binds in vitro the AVR-Pik effectors recognized in planta.



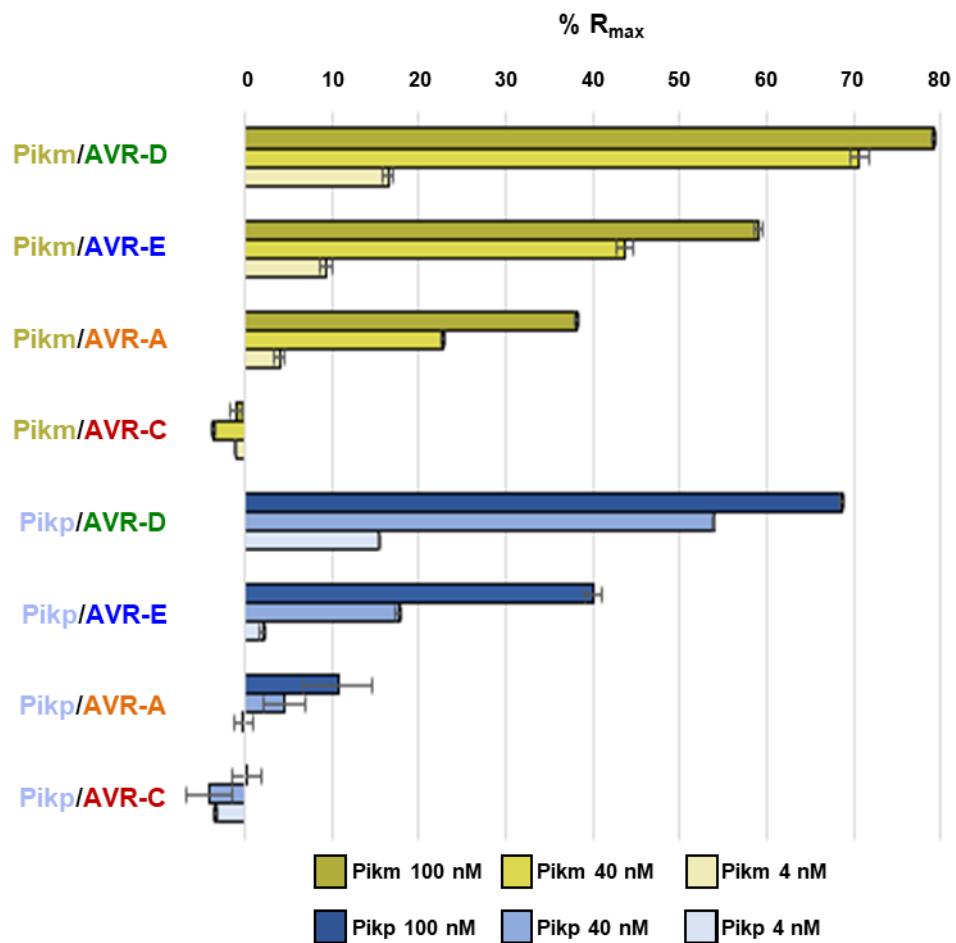
**Figure 3.7. Pikm-HMA forms a complex with AVR-PikD, AVR-PikE and AVR-PikA in vitro.** (a) Analytical gel filtration confirms that Pikm-HMA forms complexes with AVR-PikD, AVR-PikE, and AVR-PikA in vitro, but not AVR-PikC. Earlier elution correlates with increased molecular mass. Retention volumes for peaks are labelled (black arrow indicates Pikm-HMA elution volume, Pikm-HMA does not absorb light at 280 nm). (b) SDS-PAGE showing the proteins eluted under the peaks in the analytical gel filtration experiments.

## Chapter 3

To quantitatively determine the extent to which the extended response of *Pikm* to AVR-PikE and AVR-PikA effectors in *N. benthamiana* is related to the strength of binding to the *Pikm*-HMA, Dr. Marina Franceschetti used Surface Plasmon Resonance (SPR) while I assisted by producing purified proteins and helping to set up the assay (see materials and methods for details).

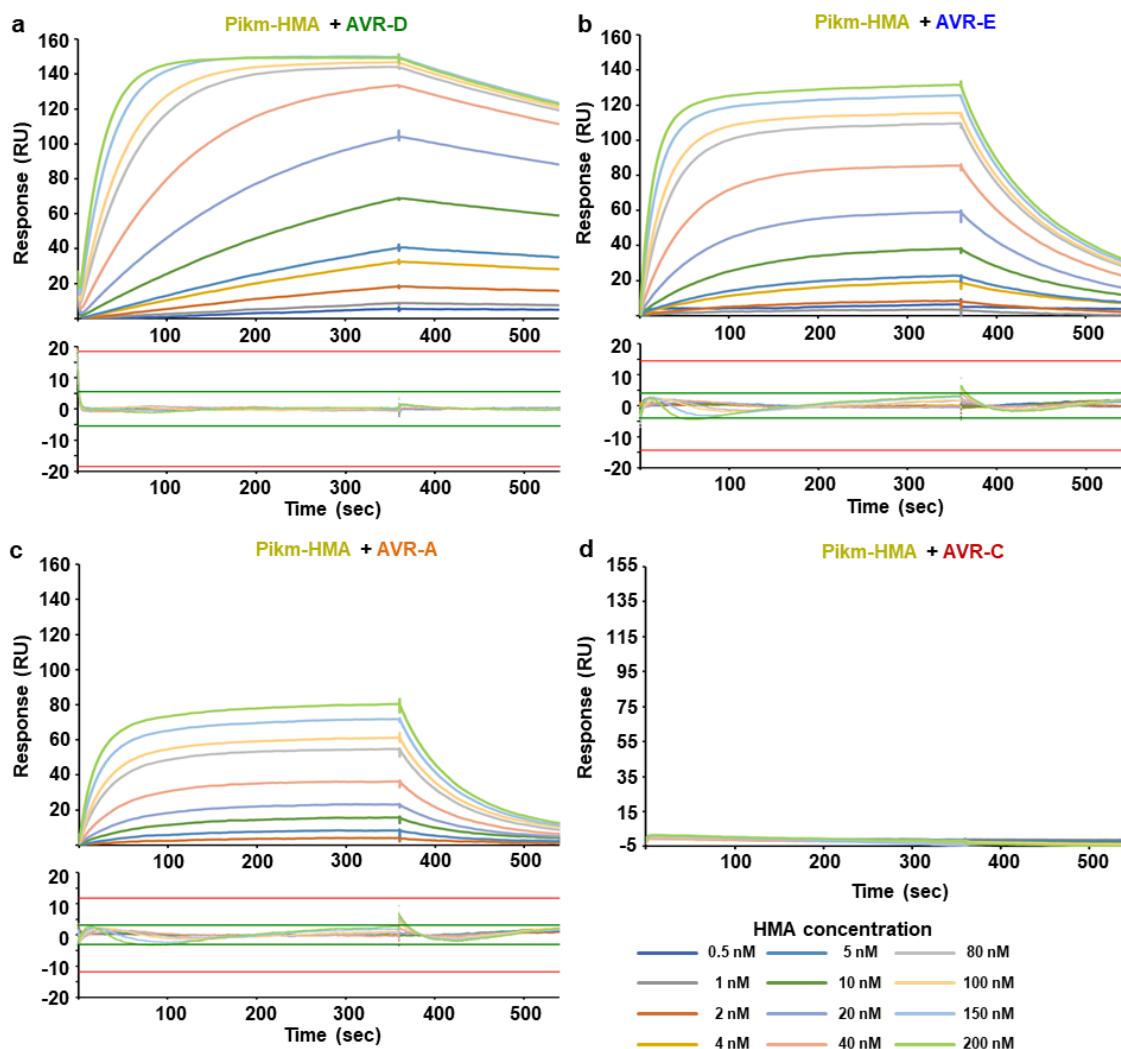
We captured each of the AVR-Pik effectors onto the chip surface and measured response units (RU) after injection of *Pikm*-HMA. For each effector, we measured RUs using three different concentrations of the HMA (4 nM, 40 nM and 100 nM), then we normalised the value to the  $R_{max}$  (theoretical maximum response, assuming a 1:1 effector/*Pikm*-HMA interaction model). This allowed us to rank the order of highest to lowest apparent affinity. To allow comparisons between *Pik* alleles, we also used this method to analyse *Pikp*-HMA binding to AVR-Pik effectors (assuming a 1:2 effector/*Pikp*-HMA interaction model, as previously observed in the crystal structure (Maqbool et al., 2015)). In all the HMA concentrations measured, *Pikm*-HMA consistently binds AVR-Pik effectors with higher apparent affinity (**Figure 3.8**). While binding of *Pikm*-HMA and *Pikp*-HMA to AVR-PikD is similar, there is a marked difference in binding to AVR-PikE and, especially, to AVR-PikA (**Figure 3.8**). No significant binding to AVR-PikC was observed for either *Pikp*-HMA or *Pikm*-HMA (**Figure 3.8**). This increased apparent affinity correlates with the extended *Pikm* cell death response in planta.

To estimate the equilibrium dissociation constant,  $K_D$ , we extended the SPR assay using a multicycle kinetics at concentrations ranging from 0.5nM to 200nM of *Pikm*-HMA. We calculated the RU for each concentration and using a 1:1 interaction model (as observed in the crystal structure, see below), we found *Pikm*-HMA binds AVR-PikD with very high affinity (low  $K_D$ ) (**Figure 3.9a**). Residuals (observed response - response predicted by the model) are shown below the plotted curves and indicate good agreement between model and experimental data as determined by the Biacore T100 evaluation software (**Figure 3.9a**). Binding to AVR-PikE and AVR-PikA have a higher  $K_D$  (and therefore lower binding affinity) (**Figure 3.9b, c**), however, this  $K_D$  value still indicates a high binding affinity for these effectors. We observed no significant binding of *Pikm*-HMA to AVR-PikC (**Figure 3.9d**).



**Figure 3.8. Binding affinity between Pik-HMA domains and effectors in vitro correlates with in planta responses.** %R<sub>max</sub> values (percentage of the theoretical maximum response, assuming a 1:2 (effector:HMA) binding model for Pikp and a 1:1 binding model for Pikm) for Pikp-HMA and Pikm-HMA with the AVR-Pik effectors alleles. Pikp-HMA and Pikm-HMA results are represented by blue and yellow bars, respectively. Results using HMA protein concentration of 4, 40 and 100 nM are plotted. Bars represent the average of three measurements, with the error bars representing the standard deviation.

We then attempted to determine the  $K_D$  for Pikp-HMA. However, Pikp-HMA binds non-specifically to the NTA chip at high concentrations, therefore we had to reduce the Pikp-HMA concentration range to 0.5nM to 100nM. Because of this, we were only able to reliably determine the  $K_D$  for Pikp-HMA bound to AVR-PikD, showing a  $K_D$  comparable to the association between Pikm-HMA and AVR-PikD (Figure 3.10a). We were not able to determine  $K_D$ s for the binding between Pikp-HMA and AVR-PikE and AVR-PikA as the binding affinity was not enough for the curves to reach saturation (even in higher concentrations) (Figure 3.10b, c). Therefore, the Biacore T100 evaluation software reported them of insufficient quality under our assay conditions. The summary of the calculated values can be found in table 3.1.

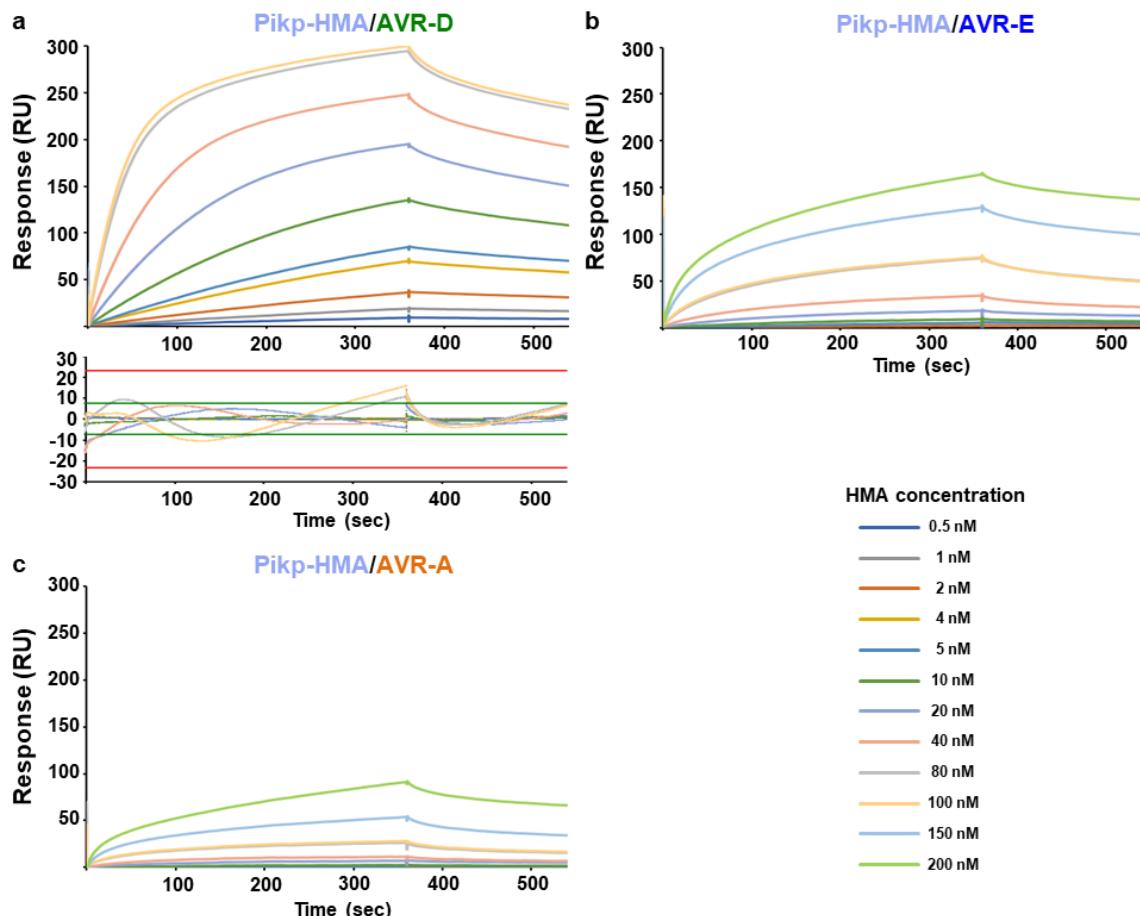


**Figure 3.9. Multi-cycle kinetics data for the interaction of Pikm-HMA with AVR-Pik effectors (a) AVR-PikD, (b) AVR-PikE, (c) AVR-PikA, (d) AVR-PikC.** Response units for each labelled protein concentration are shown with the residuals plot beneath (SPR acceptance guides as determined by Biacore software are shown as green and red lines in the residuals plots). These data have been used to calculate the association/dissociation constants given in **Table 3.1**.

Interaction	$K_a (10^5 \text{ M}^{-1} \text{ s}^{-1})$	SE	$k_d (10^{-3} \text{ s}^{-1})$	SE	$K_D (\text{nM})$
Pikm-HMA/AVR-D	2.195	0.004	1.027	0.002	4.7
Pikm-HMA/AVR-E	2.986	0.010	8.60	0.05	29
Pikm-HMA/AVR-A	1.530	0.009	11.16	0.05	73
Pikm-HMA/AVR-C	N.D.		N.D.		N.D.
Pikp-HMA/AVR-D	2.077	0.002	1.217	0.003	5.9
Pikp-HMA/AVR-E	N.D.		N.D.		N.D.
Pikp-HMA/AVR-A	N.D.		N.D.		N.D.
Pikp-HMA/AVR-C	N.D.		N.D.		N.D.

**Table 3.1. Association and dissociation constants for Pik-HMAs and AVR-Pik effector alleles from the multi-cycle kinetics SPR.** Data obtained by Marina Franceschetti (De la Concepcion et al., 2018). N.D. indicated “not determined” for those Pik-HMA/AVR-Pik pairs whose interaction were of insufficient quality to obtain the parameters.

Altogether, the SPR results collected by Dr. Marina Franceschetti and the interaction monitored by Y2H suggest that the differential binding affinity to the Pik-HMA domains is the source of the allele-specific response profile that is observed in planta.

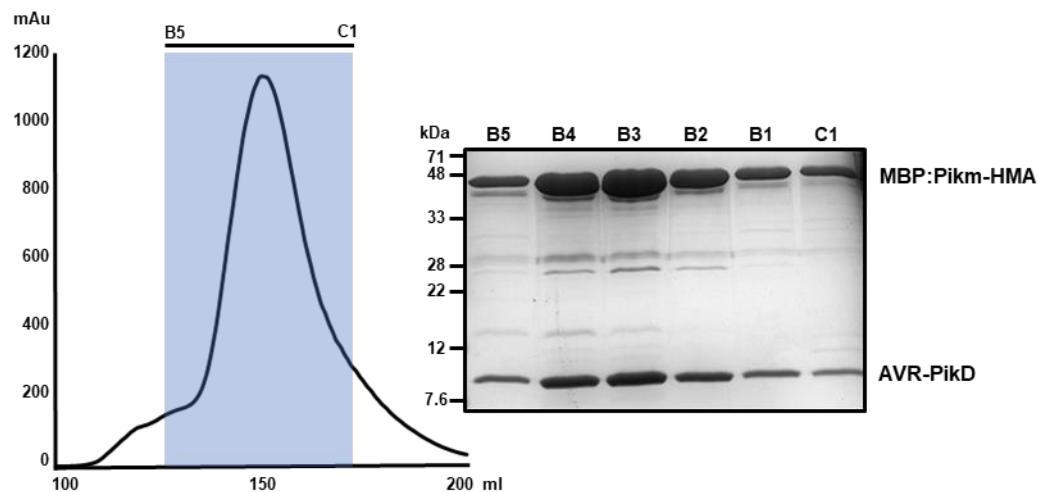


**Figure 3.10. Multi-cycle kinetics data for the interaction of Pikp-HMA with AVR-Pik effectors (a) AVR-PikD, (b) AVR-PikE, (c) AVR-PikA.** Response units for each labelled protein concentration are shown with the residuals plot beneath. The data in panel (a) was used to calculate the association/dissociation constants given in **Table 3.1** (SPR acceptance guides as determined by Biacore software are shown as green and red lines in the residuals plots).

### 3.6 Production, purification and crystallization of Pik-HMA/AVR-Pik complexes

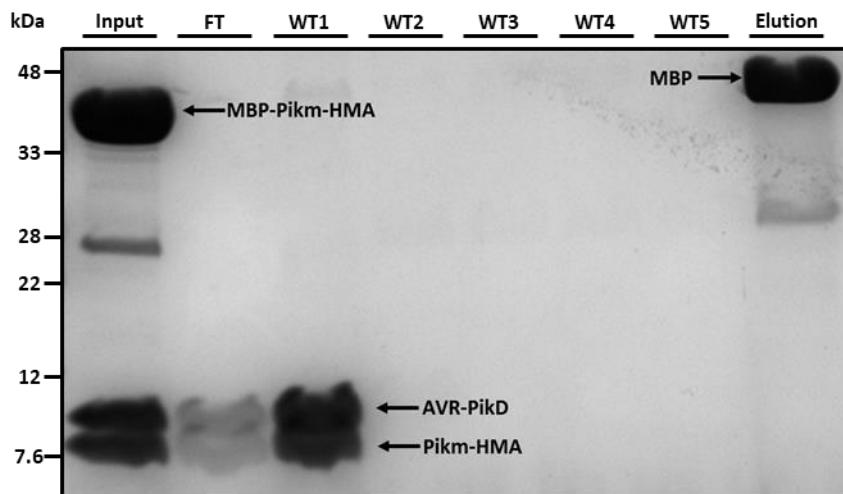
I aimed to use X-ray crystallography to unravel the structural basis of the broader recognition specificity by Pikm. To do this, I used a co-expression and co-purification strategy to obtain complexes between Pikm-HMA and AVR-PikD, AVR-PikE and AVR-PikA. In addition, I also obtained complexes between Pikp-HMA and AVR-PikD and AVR-PikE for comparison (see materials and methods for details).

I used co-transformed *E. coli* cells overexpressing 6xHis:MBP:Pik-HMA and non-tagged AVR-Pik effectors, so after purification I obtained only AVR-Pik effector that is bound to the Pik-HMA. Once produced, protein complexes were purified following the procedure explained above for Pikm-HMA. Protein fractions analysed by SDS-PAGE showed both proteins were present as two separate bands (~50 kDa and ~10 kDa for the MBP tagged HMA protein and the untagged AVR-Pik effector, respectively) (**Figure 3.11**).

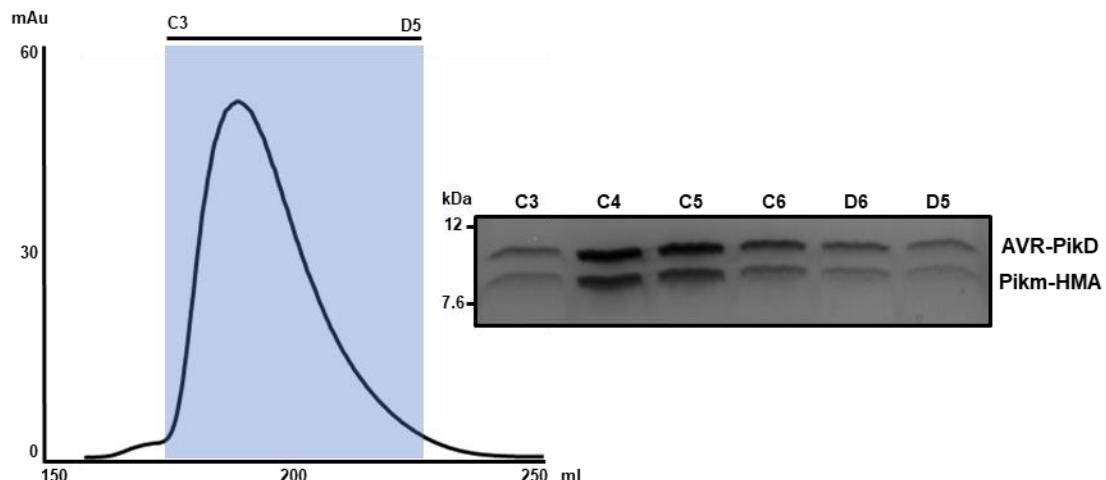


**Figure 3.11. Pikm-HMA can be co-purified with AVR-Pik effectors.** Elution trace of co-purification of MBP:Pikm-HMA and AVR-PikD after IMAC + gel filtration. Fractions from B5 to C1 were collected and analysed by SDS-PAGE showing the presence of both proteins.

After cleavage of the tags and further purification, analysis by SDS-PAGE showed two close bands, corresponding to AVR-Pik effector (~10 kDa) and the Pik-HMA (~8 kDa) (**Figure 3.12**). The fractions containing the complexes were combined and subjected to a final gel filtration. On this occasion, the purification was followed by the UV absorption traces because the AVR-Pik protein bound to the HMA contains aromatic residues (**Figure 3.13**). After gel filtration, I pooled together the fractions containing both proteins and I concentrated them for crystallization trials.



**Figure 3.12. 6xHis:MBP tag can be efficiently removed from the Pik-HMA/AVR-Pik complex.** SDS-PAGE analysis of fractions collected after Pikm-HMA/AVR-PikD HisTrap™ and MBPTrap™ tandem purification following 3C protease treatment. The protein treated with 3C protease before purification is shown as an input. Pikm-HMA in complex with AVR-PikD was successfully purified in the flow-through (FT) and the first wash-through with A4 buffer (WT1). MBP tag eluted in the elution treatment with B1 buffer containing imidazole.



**Figure 3.13. Final gel filtration step yields a pure Pikm-HMA domain complexed with AVR-Pik.** Elution trace of Pikm-HMA/AVR-PikD complex in gel filtration. SDS-PAGE analysis of the collected fractions showed both Pikm-HMA and AVR-PikD co-elutes between the fractions C3 and D5.

I set up several commercial crystallisation screens with the purified complexes and, after 3 to 4 days of incubation, protein crystals appeared in the Morpheus® HT-96 screen (**Figure 3.14**) (see Materials and methods for conditions and concentrations). Protein crystals were harvested by Dr. Clare Stevenson (JIC crystallography platform) and sent to Diamond Light Source (UK). Together with Professor David Lawson (JIC crystallography platform) we collected X-ray diffraction data at 1.2 Å, 1.3 Å, and 1.3 Å

resolution for Pikm-HMA bound to AVR-PikD, AVR-PikE, or AVR-PikA, respectively. Dr. Marina Franceschetti did the subsequent refinement of the protein structures obtained from this dataset. All the details of X-ray data collection, structure solution, and structure completion are given in the Methods and **Table 3.2**.



**Figure 3.14. Pik-HMA/AVR-Pik complexes can be successfully crystallized.** Images of crystal obtained in Morpheus® HT-96 screening for Pikm-HMA in complex with AVR-PikD, AVR-PikE and AVR-PikA.

To allow a structural comparison between Pikm and Pikp alleles, I also obtained crystals of Pikp-HMA, with the 5-amino acid extension at the C-terminus of the HMA, bound to AVR-PikD or AVR-PikE (even though the Pikp-HMA/AVR-PikE combination does not give rise to responses in planta). These crystals diffracted at 1.35 Å and 1.9 Å resolution, respectively. I subsequently solved and refined the protein structures for these complexes. All the details of X-ray data collection, structure solution and refinement are presented in **Table 3.3**.

### Chapter 3

	Pikm-HMA/AVR-D	Pikm-HMA/AVR-E	Pikm-HMA/AVR-A
<b>Data collection statistics</b>			
Wavelength (Å)	0.9168	0.9795	0.9795
Space group	$P2_12_12_1$	$P2_1$	$P2_1$
Cell dimensions			
$a, b, c$ (Å)	37.09, 87.13, 103.84	30.02, 54.35, 48.49	30.07, 54.44, 46.64
$\alpha, \beta, \gamma$ (°)	90.00, 90.00, 90.00	90.00, 90.53, 90.00	90.00, 104.43, 90.00
Resolution (Å)*	29.04-1.20 (1.23-1.20)	23.70-1.30 (1.32-1.30)	34.76-1.30 (1.33-1.30)
$R_{\text{merge}}$ (%)#	4.9 (85.8); 5.2 (97.2)	7.1 (90.4); 7.6 (101.0)	4.3 (28.4); 4.8 (35.2)
$\text{I}/\sigma \text{I}^{\#}$	21.7 (2.3)	12.9 (1.6)	21.5 (4.5)
Completeness (%)#	98.5 (88.4)	99.4 (92.4)	99.3 (91.9)
Unique reflections#	104301 (6772)	38170 (1751)	33854 (2405)
Redundancy#	11.9 (7.9)	7.6 (5.0)	7.8 (5.2)
CC <sup>(1/2)</sup> (%)#	99.9 (64.4)	99.9 (73.9)	99.9 (94.1)
<b>Refinement and model statistics</b>			
Resolution (Å)	29.04-1.20 (1.23-1.20)	23.70-1.30 (1.33-1.30)	34.76-1.30 (1.33-1.30)
$R_{\text{work}}/R_{\text{free}}$ (%)^	14.9/18.4 (25.8/28.1)	14.1/17.9 (23.9/26.9)	12.5/16.2 (17.1/21.8)
No. atoms (Protein)	2632	1386	1294
B-factors (Protein)	22.9	21.4	22.5
R.m.s deviations^			
Bond lengths (Å)	0.019	0.018	0.017
Bond angles (°)	1.902	1.925	1.775
Ramachandran plot (%)**			
Favoured	99.1	98.2	99.4
allowed	0.9	1.8	0.6
Outliers	0	0	0
MolProbity Score	<b>0.92</b> (99 <sup>th</sup> percentile)	<b>0.50</b> (100 <sup>th</sup> percentile)	<b>0.83</b> (100 <sup>th</sup> percentile)

**Table 3.2. X-ray data collection and refinement statistics for Pikm-HMA /AVR-PikD, Pikm-HMA /AVR-PikE and Pikm-HMA/AVR-PikA complexes.** \*The highest resolution shell is shown in parenthesis. # As calculated by xia2 pipeline or Aimless. ^ As calculated by Refmac5. \*\* As calculated by MolProbity.

## Chapter 3

	Pikp-HMA/AVR-D	Pikp-HMA/AVR-E
<b>Data collection statistics</b>		
Wavelength (Å)	0.9763	0.9763
Space group	$P\ 2_1$	$P2_12_12_1$
Cell dimensions		
$a, b, c$ (Å)	40.50, 66.26, 45.74	67.20, 80.18, 105.63
$\alpha, \beta, \gamma$ (°)	90.00, 113.32, 90.00	90.00, 90.00, 90.00
Resolution (Å)*	42.00-1.35 (1.37-1.35)	46.29-1.90 (1.94-1.90)
$R_{\text{merge}}$ (%)#	6.1 (76.4); 6.6 (83.1)	10.0 (100.5); 10.4 (105)
$\text{I}/\sigma$ #	13.4 (2.4)	14.5 (2.6)
Completeness (%)#	99.8 (99.8)	100.0 (99.9)
Unique reflections#	48646 (2371)	45720 (2876)
Redundancy#	6.6 (6.6)	12.0 (12.1)
CC <sup>(1/2)</sup> (%)#	99.8 (82.6)	99.9 (94.7)
<b>Refinement and model statistics</b>		
Resolution (Å)	42.00-1.35 (1.39-1.35)	46.29-1.90 (1.95-1.90)
$R_{\text{work}}/R_{\text{free}}$ (%)^	14.3/18.5 (22.7/23.3)	20.9/22.4 (31.9/31.4)
No. atoms (Protein)	2024	3712
B-factors (Protein)	23.1	36.0
R.m.s deviations^		
Bond lengths (Å)	0.018	0.014
Bond angles (°)	1.919	1.614
Ramachandran plot (%)**		
Favoured	98.2	98.8
allowed	1.8	1.2
Outliers	0	0
MolProbity Score	1.36 (90 <sup>th</sup> percentile)	1.23 (99 <sup>th</sup> percentile)

**Table 3.3. X-ray data collection and refinement statistics for Pikp-HMA/AVR-PikD and Pikp-HMA/AVR-PikE complexes.** \*The highest resolution shell is shown in parenthesis. # As calculated by xia2 pipeline or Aimless. ^ As calculated by Refmac5. \*\* As calculated by MolProbity.

### 3.7 Structures of Pik-HMAs in complex with AVR-Pik effectors reveal multiple interaction surfaces

The structures of these Pik-HMA/AVR-Pik effector complexes display a very similar orientation to each other, and to the previously reported Pikp-HMA/AVR-PikD complex (Maqbool et al., 2015) (Figure 3.15). A summary of the structure superpositions calculated in COOT is provided in Table 3.4. The only major difference we observed is that Pikm-HMA/AVR-Pik structures form a complex with 1:1 stoichiometry, whereas in the Pikp-HMA/AVR-Pik complexes previously reported (Maqbool et al., 2015) and obtained here the stoichiometry is 2:1 (HMA/effectector) with only one molecule of Pikp-HMA interacting with the AVR-Pik effector. This suggest that Pikp-HMA dimerization is most likely an artefact and has no relevance in the activation mechanism.

Professor Mark Banfield performed an overall analysis of the interfaces between Pikm-HMA and the effectors using QtPISA (Krissinel, 2015) (Table 3.16), and we graphically represented the key interface components of each interaction using QtPISA interaction radars (Krissinel, 2015) (Figure 3.16). For Pikm-HMA/AVR-Pik complexes, the overall properties of the interfaces are similar, and the interaction radars reveal a high likelihood that each interface is biologically relevant. However, we observe a trend on reducing the total interface area in the order AVR-PikD>AVR-PikE>AVR-PikA (Figure 3.16 and Table 3.5).

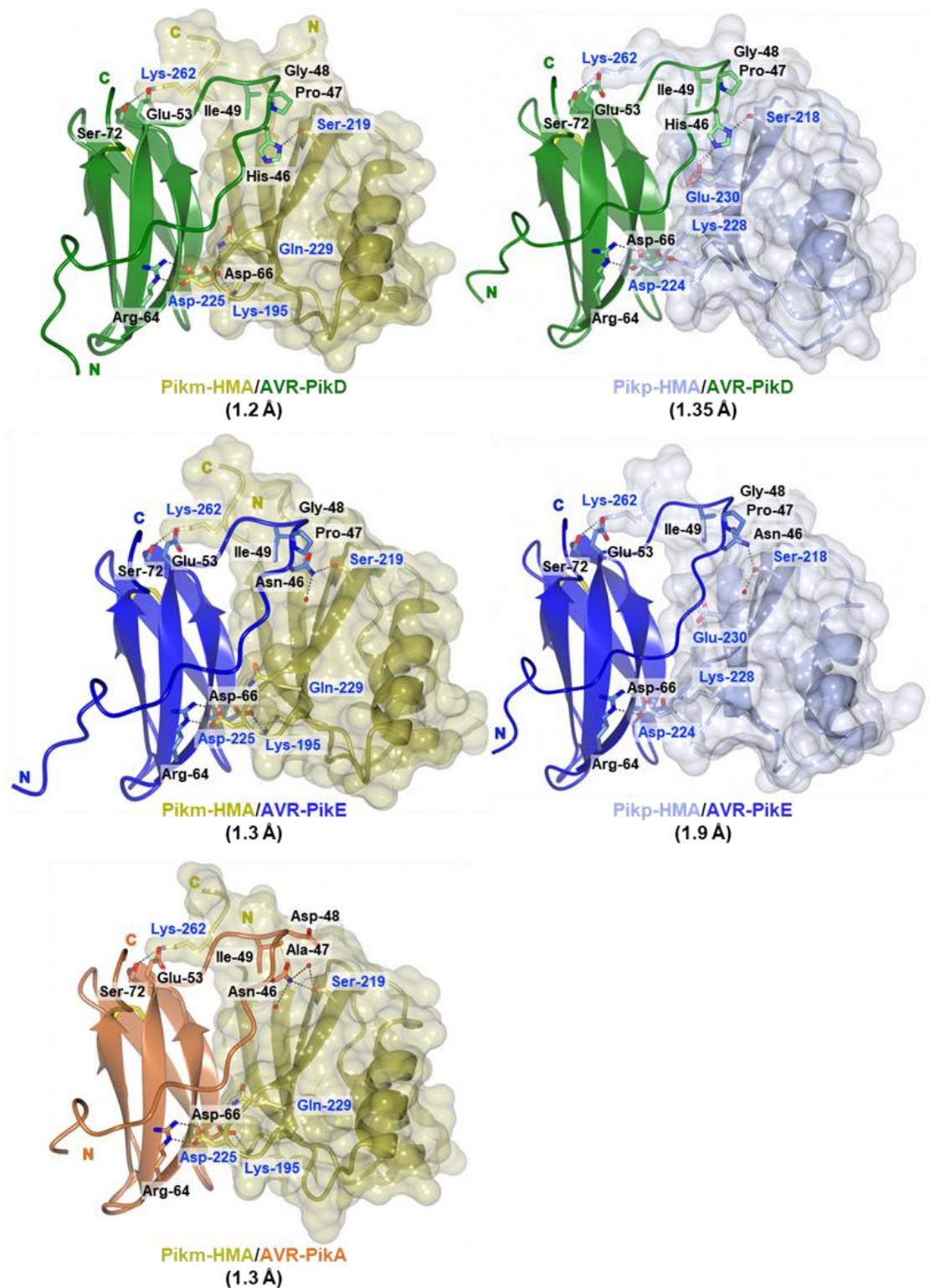
QtPISA analysis comparing Pikp-HMA/AVR-PikD and Pikp-HMA/AVR-PikE structures show that the interface with AVR-PikE has a reduced total binding energy compared to the interface with AVR-PikD, as portrayed by a reduced area of the polygon in Figure 3.16. Furthermore, five out of six interface components are below the 50% threshold, which questions the biological relevance of the Pikp-HMA/AVR-PikE interface.

Complex (compared to <b>Pikm-HMA/AVR-D</b> )	AVR-Pik		HMA
	r.m.s.d. - Å (no. of residues)	r.m.s.d. - Å (no. of residues)	
<b>Pikm-HMA/AVR-E</b>	0.55 (82 of 84)		0.54 (78 of 79)
<b>Pikm-HMA/AVR-A</b>	0.37 (81 of 84)		0.50 (76 of 79)
<b>Pikp-HMA/AVR-D</b>	0.50 (81 of 84)		1.10 (72 of 79)
<b>Pikp-HMA/AVR-E</b>	0.48 (82 of 84)		0.99 (72 of 79)

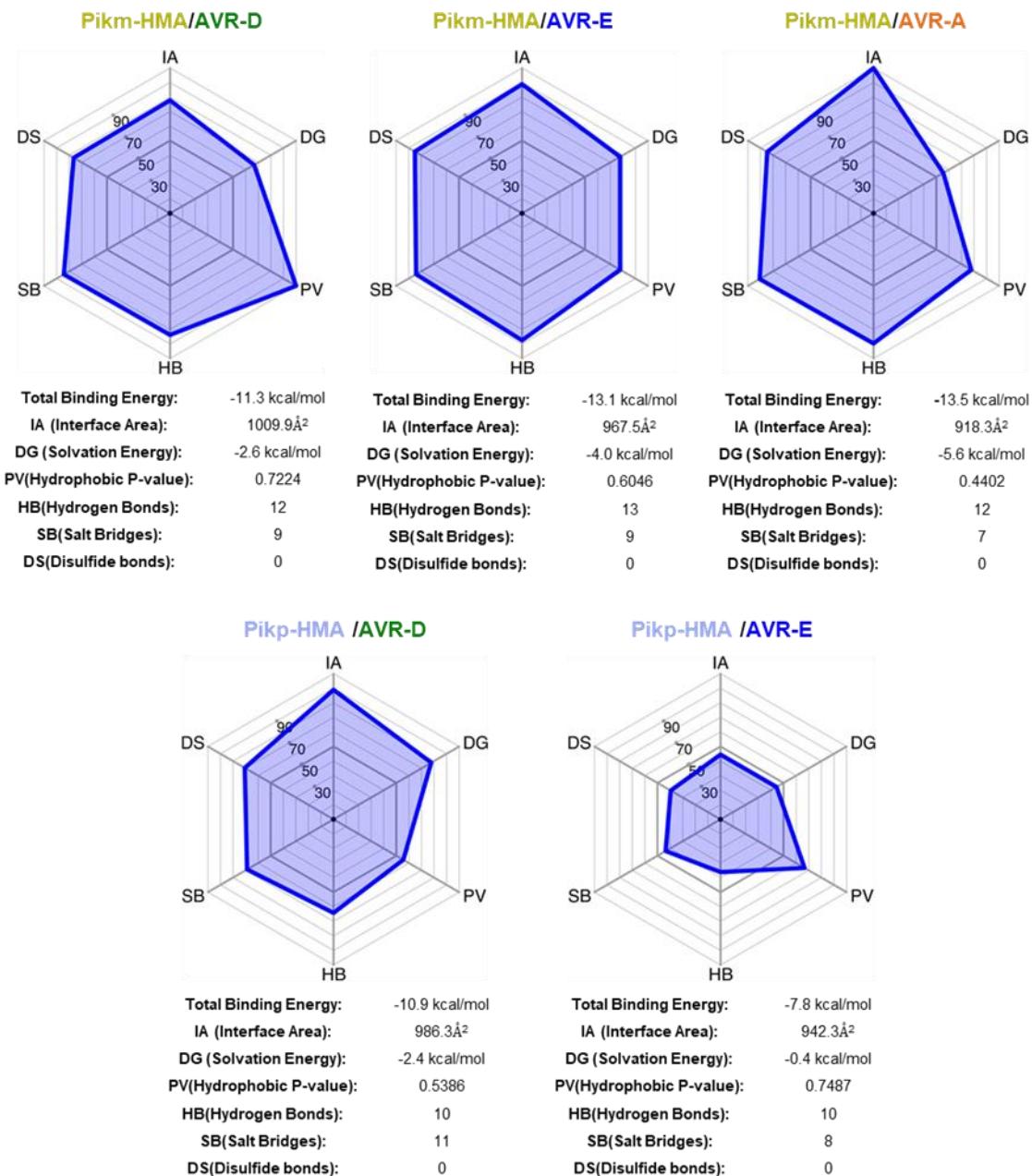
  

Complex (compared to <b>Pikp-HMA/AVR-D</b> )	AVR-Pik		HMA
	r.m.s.d. - Å (no. of residues)	r.m.s.d. - Å (no. of residues)	
<b>Pikp-HMA/AVR-E</b>	0.35 (81 of 83)		0.60 (70 of 76)

Table 3.4. Summary of overlays analysis. Overlays parameters were calculated with SSM in COOT.



**Figure 3.15. Pik-HMA/AVR-Pik complexes adopt a similar fold.** Schematic representations of the structures of Pikm-HMA and Pikp-HMA in complex with AVR-Pik effectors. Pikm-HMA and Pikp-HMA are shown in gold or ice blue cartoon representation, respectively, with selected side chains as sticks; the molecular surface of this domain is also shown. The effectors are coloured as labelled, and shown in cartoon representation, with selected side chains as sticks. Hydrogen bonds/salt bridges are shown as dashed lines and the di-sulphide bond as yellow bars. For clarity, of the two molecules of Pikp-HMA present in the complex, only the one making extensive contacts with the effector is shown.



**Figure 3.16. Analyses of Pik-HMA binding interfaces with AVR-Pik effectors using QtPISA.** For each complex, the QtPISA interaction radar generated with the reference parameter “Total Binding Energy” is shown. In this presentation the internal area of the polygon represents the Total Binding Energy. Also shown are the values obtained for each key interface parameter in the analysis. The QtPISA interaction radars score macromolecular interfaces based on statistical analysis of all interfaces found in the Protein Data Bank (see Methods). The abbreviations in the radars are defined in the tables underneath each panel.

## Chapter 3

Complex	B.S.A. (Å <sup>2</sup> )	% B.S.A. of total	B.S.A. (Å <sup>2</sup> )	% B.S.A. of total	IA* Å <sup>2</sup>	H-bonds	Salt Bridges
<b>Pikm-HMA/AVR-D</b> (chains A and B)	1004.9	18.1	1015.0	20.9	1009.9	12	9
<b>Pikm-HMA/AVR-E</b>	996.2	20.0	939.0	16.3	967.5	13	9
<b>Pikm-HMA/AVR-A</b>	891.2	16.6	945.4	19.8	918.3	12	7
<b>Pikp-HMA/AVR-D</b> (chains B and C)	978.4	17.8	994.2	21.2	986.3	10	11
<b>Pikp-HMA/AVR-E</b> (chains E and F)	969.1	21.3	915.5	17.2	942.3	10	8

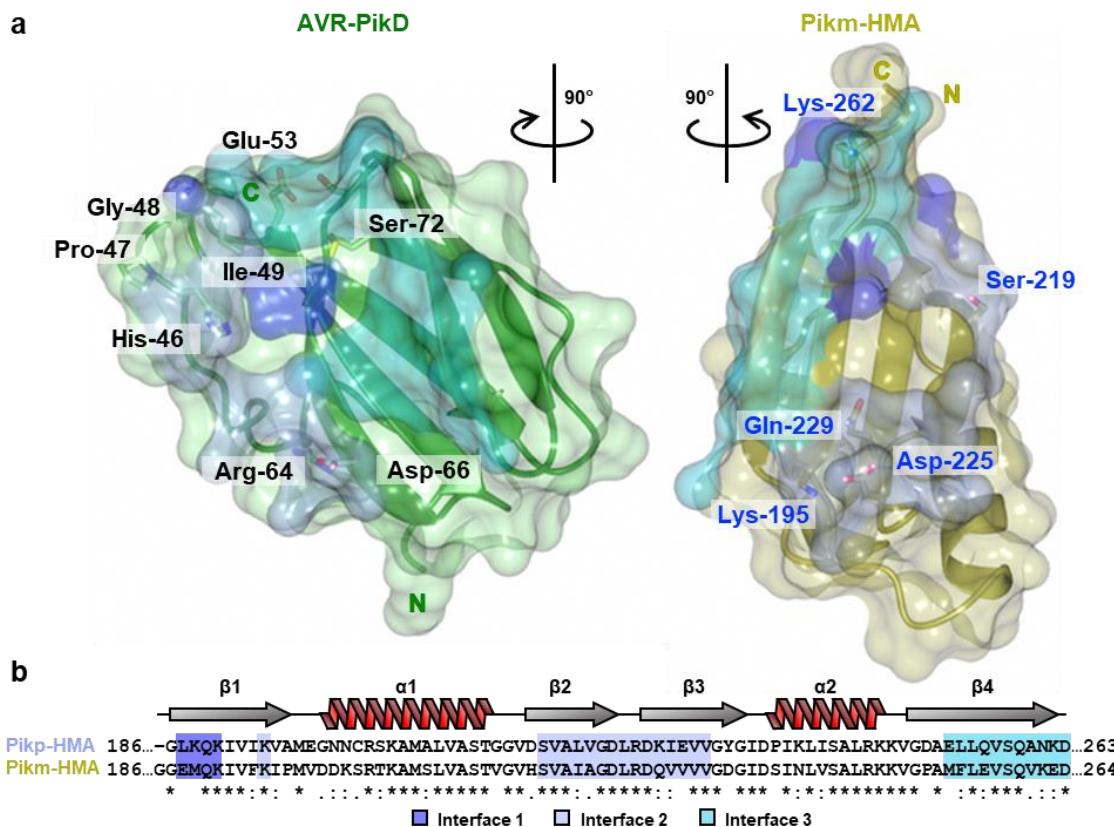
**Table 3.5. Summary of interface analysis.** Parameters were calculated with QtPISA. \*Total interface area is the sum of the B.S.A. (Buried Surface Area) of each component divided by two.

Analysis of these five crystal structures allowed to define three main contact regions (interfaces) between Pik-HMA and the AVR-Pik effectors (**Figure 3.17**).

The first interface spans N-terminal residues Glu188 to Lys191 within the HMA domain. The side chain of Met189 contributes a hydrophobic interaction to the side chain of Ile49 of the effector. Additionally, Lys191 forms a weak hydrogen bond with the main-chain carbonyl group of Thr69 in the AVR-Pik effector. Overall, this interface is a minor component of the Pik-HMA/AVR-Pik interaction.

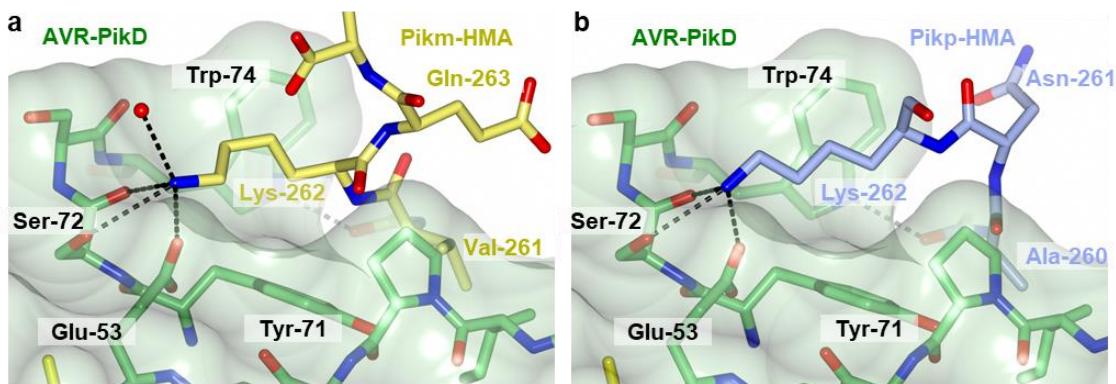
Interface two is more extensive and comprises HMA residues from Ser219 to Val233, situated in the second and the third  $\beta$ -strands, and Lys195 in the first (**Figure 3.17**). This interface was previously described by Maqbool et al. 2015. The interaction is mainly with the effector residues in the N-terminal extension of the conserved MAX effector fold (de Guillen et al., 2015), which includes the highly polymorphic positions 46, 47 and 48 of AVR-Pik (**Figure 3.17 and Figure 3.1b**). The interface also includes salt-bridge/hydrogen bond interactions of Asp225 to Arg64 of AVR-Pik, and Lys195 to Asp66 (**Figure 3.17a**).

Importantly, the work presented here has uncovered a new interface, named interface three, comprising residues Met254 to Asp264 at the C-terminal  $\beta$ -strand (**Figure 3.17**). This interface could not be analysed in previous studies because the construct used for crystallization lacked the last five amino acids of the HMA domain (Maqbool et al., 2015). In this interface, the side-chain of Lys262 from the HMA is inserted into a surface pocket made by residues Glu53, Tyr71, Ser72, and Trp74 from the effector. Here, the Lys262 side-chain makes several interactions, including salt-bridge and hydrogen bonds, with the side-chains of Glu53 and Ser72 from AVR-Pik (**Figure 3.17a, Figure 3.18**).



**Figure 3.17. Pik-HMA domains form three interaction interfaces with AVR-Pik effectors. (a)** The buried surface area of AVR-PikD and Pikm-HMA is shown from the perspective of the partner (change in orientation indicated with arrows). The interaction interfaces surfaces are coloured according to interfaces described in the text (interface one is in dark blue, interface two is in light blue and interface three is in cyan). **(b)** Amino acid sequence alignment of Pikp-1 and Pikm-1 HMA domains. Secondary structure features of the HMA fold are shown above, and the residues located to the interfaces described in the text are highlighted.

In summary, the high-resolution structures of the different Pik-HMA/AVR-Pik complexes allowed us to precisely define the interaction interfaces between the NLR and the effector, uncovering new features that remained hidden in previous studies. The overall conservation between the structures also suggest that small rather than large structural differences are determinants of extended NLR recognition.



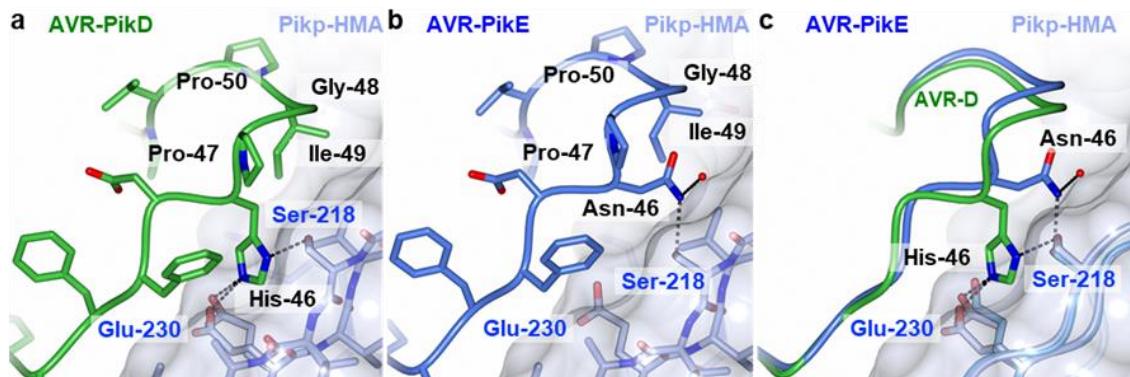
**Figure 3.18. Lys262 in Pikp-HMA and Pikm-HMA make numerous interactions at the HMA/AVR-Pik interface three.** Close-up view of the interactions across interface three in the (a) Pikm-HMA and (b) Pikp-HMA complexes with AVR-PikD, showing different conformations for the C-terminal regions of the HMA domains. Compared to Pikm-HMA, Asn261 of Pikp-HMA loops out from the interface and Lys262 sidechain adopts a different orientation. AVR-PikD is shown in green cartoon, with side chains as sticks, the molecular surface of the effector is also shown. The Pik-HMA domains are coloured as labelled.

### 3.8 The structure of Pikp-HMA bound to AVR-PikE uncovers the mechanistic basis of how polymorphic effectors escape Pikp recognition

As discussed above, polymorphic effector variants AVR-PikD, AVR-PikE and AVR-PikA differ at positions 46, 47 and 48 (Yoshida et al., 2009), which localise at interface two in the Pik-HMA/AVR-Pik complexes (Figure 3.17). These amino acid changes result in loss of Pikp-mediated recognition to AVR-PikE and AVR-PikA (Kanzaki et al., 2012) (Figure 3.1b). Using the structure of a receptor/effectector complex that does not lead to a successful effector recognition event in planta (Pikp-HMA/AVR-PikE), I sought to understand the mechanistic basis of how polymorphic AVR-Pik effectors evade immune recognition in rice cultivars with the NLR allele Pikp.

Previous work showed that Pikp-HMA residues Ser218 and Glu230 coordinate AVR-PikD His46 via hydrogen bonds, and this contact underpins binding and immune recognition of AVR-PikD by Pikp (Maqbool et al., 2015) (Figure 3.19a). A global comparison between Pikp-HMA/AVR-PikD and Pikp-HMA/AVR-PikE complexes, revealed extensive differences at interfaces two and three. At interface two, AVR-PikE Asn46 rotates out of the AVR-PikD His46 binding pocket. This residue still forms a hydrogen bond with Ser218 of Pikp-HMA, however, this is in a very different orientation (Figure 3.19b). This results in AVR-PikE residues Asn46 to Pro50 moving away from Pikp-HMA. Interestingly, the superposition of Pikp-HMA/AVR-PikD and Pikp-HMA/AVR-PikE revealed that changes in interface two are not dramatic (Figure 3.19c). However,

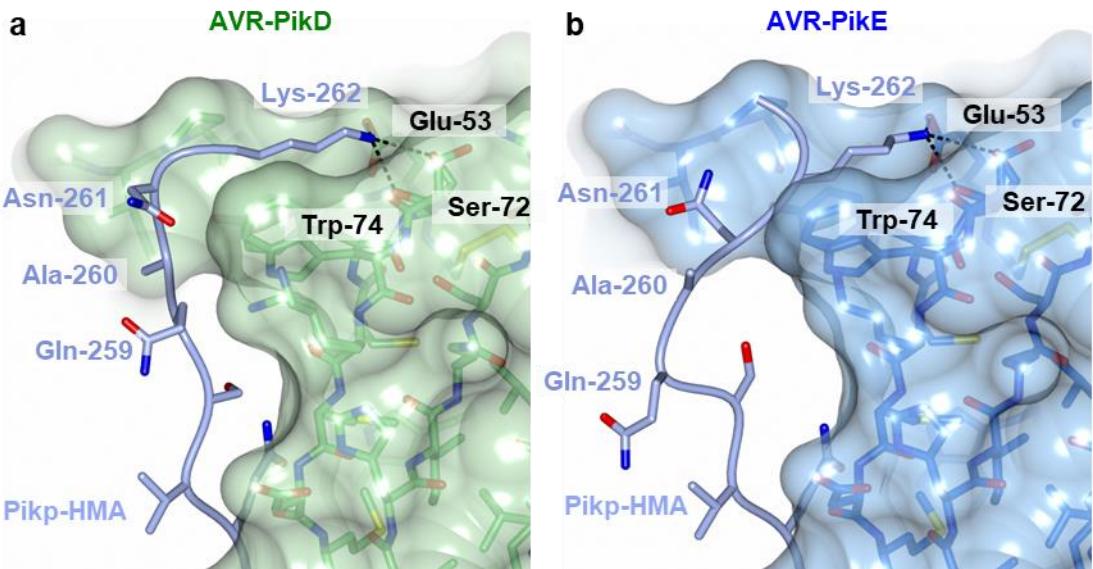
the re-configuration at this interface is propagated to interface three causing a more dramatic difference that disrupts the receptor/effectector interaction.



**Figure 3.19. His46Asn polymorphism in AVR-PikE disrupt interaction in Pikp-HMA/AVR-Pik interface two.** Zoom-in views of the interactions across interface two in the Pikp-HMA complexes with (a) AVR-PikD and (b) AVR-PikE. In each panel the Pikp-HMA domain is shown as ice-blue sticks, the molecular surface is also shown. Effector variant residues are coloured as labelled and shown in Ca-worm with sidechain representation. (c) Superposition of interface two of the Pikp-HMA/AVR-PikD and Pikp-HMA/AVR-PikE structures showing a different occupancy of polymorphism in position 46 of AVR-PikE compared with AVR-PikD.

As discussed above, the side-chain of Lys262 makes several interactions with residues forming a surface pocket at the AVR-Pik effector, particularly Glu53 and Ser72 (**Figure 3.20a**). In the Pikp-HMA/AVR-PikE complex, Lys 262 adopts a similar orientation. However, because of disruption at interface two, Pikp-HMA residues Ser258 to Asn261 must adopt a dramatically different position to enable this. This results in the HMA residues Gln259 and Ala260 looping out from their positions in the Pikp-HMA/AVR-PikD complex (**Figure 3.20b**), which likely affects the binding to the effector.

Altogether, these data unravel how the changes triggered by a single amino acid substitution in the receptor/effectector interaction interface lead to reduced binding to polymorphic AVR-Pik effectors, which ultimately results in loss of immune recognition.



**Figure 3.20. The Pikp-HMA chain adopts an unfavourable conformation on interface three with AVR-PikE.** Schematic view of the different conformations adopted by Pikp-HMA, at interface three in complex with (a) AVR-PikD or (b) AVR-PikE. In each panel, the effector is shown as sticks with the molecular surface also shown and coloured as labelled. Pikp-HMA residues are coloured in ice blue and shown in the Ca-worm with side-chain representation. Residues Gln259 and Ala260 in Pikp-HMA are displaced between the complexes with AVR-PikD or AVR-PikE.

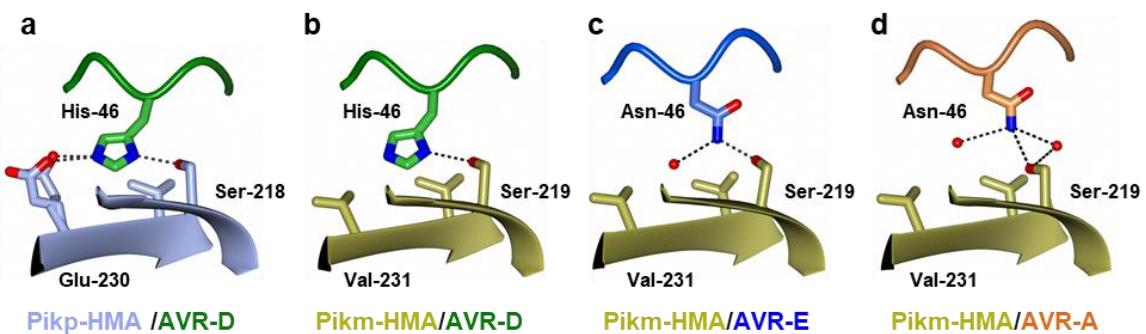
### 3.9 Structural changes at interface two underpin differential effector responses by Pikm

Pikp-HMA and Pikm-HMA differ in some of the residues situated at interface two with the effector (Figure 3.17). The Serine residue equivalent to Pikp-HMA Ser218 is conserved in Pikm and maintains a hydrogen bond with residue His46 in the effector. However, compared with the Pikp-HMA/AVR-PikD complex, Glu230 is replaced by Val231 at the structurally equivalent position. This results in the loss of a hydrogen bond with the side chain of AVR-PikD His46, yet this residue remains in the same position in both complexes (Figure 3.21a, b).

Somewhat surprisingly, the structures of Pikm-HMA in complex with AVR-PikE and AVR-PikA reveal important changes in interface two. In the complex with AVR-PikE, the Asn46 is rotated out of the binding pocket, well away from Pikm-HMA Val231. A water molecule occupies the resulting space, forming a hydrogen bond with Asn46. The AVR-Pik residues Phe44 to Gly48 are pushed away from the HMA, altering the interaction across interface two (Figure 3.21c).

In the complex with AVR-PikA, a hydrogen bond is still formed with Ser219 from the HMA (although in a significantly different orientation). However, AVR-Pik Asn46 is rotated

even further out of the HMA pocket, introducing a second water molecule. Because of these changes, residues Asn46 to Pro50 of AVR-PikA move further away from the HMA (**Figure 3.21d**). Interestingly, polymorphic positions 47 and 48 in AVR-PikA have no direct role in the interaction with Pikm-HMA.



**Figure 3.21. AVR-Pik polymorphisms in interface two alters binding to Pikm-HMA.** Close-up views of the orientation and interactions of AVR-Pik residues in position 46. (a) Pikp-HMA and AVR-PikD His46, (b) Pikm-HMA and AVR-PikD His46, (c) Pikm-HMA and AVR-PikE Asn46 and (d) Pikm-HMA and AVR-PikA Asn 46. Water molecules are shown as red spheres.

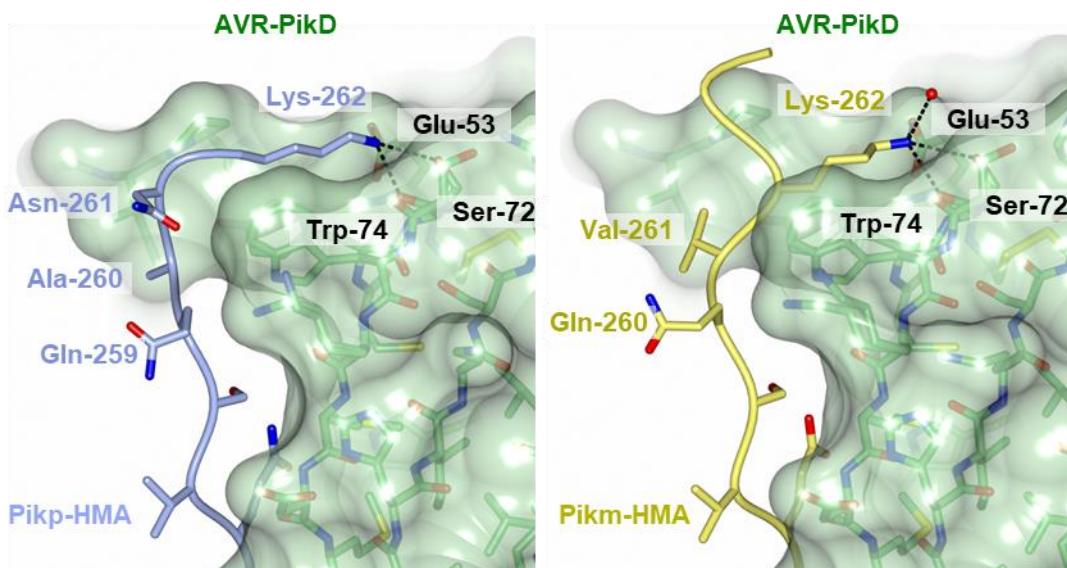
The structural changes at interface two underlie the reduced binding affinity of Pikm-HMA for AVR-PikE and AVR-PikA compared to AVR-PikD (**Figure 3.8, Figure 3.9**), which is ultimately responsible for the hierarchy in cell death response observed *N. benthamiana* (**Figure 3.2a, b**).

Contrary to Pikp, and although interface two in the Pikm-HMA complexes is also disrupted by polymorphisms in AVR-PikE and AVR-PikA, the binding between the Pikm-HMA and the effectors is still sufficient to lead to a successful immune recognition of *M. oryzae* strains carrying AVR-PikE or AVR-PikA (Kanzaki et al., 2012). This indicates that additional protein/protein interactions across interfaces one and three with Pikm-HMA may stabilise the complexes with AVR-PikE and AVR-PikA.

### 3.10 Interactions across interface three are the determinants of the extended binding of Pikm to polymorphic AVR-Pik effectors

Binding interfaces one and three also differ in Pikm compared to Pikp (**Figure 3.17**). As in Pikp, Pikm-HMA Lys262 locates to the binding pocket of the effector containing Glu53 and Ser72 (**Figure 3.18**). However, this residue is shifted one position to the N-terminus in the sequence of Pikm-HMA compared to the structurally equivalent position in Pikp-HMA (note that Pikm-1 has an additional residue in the HMA domain therefore Lys262

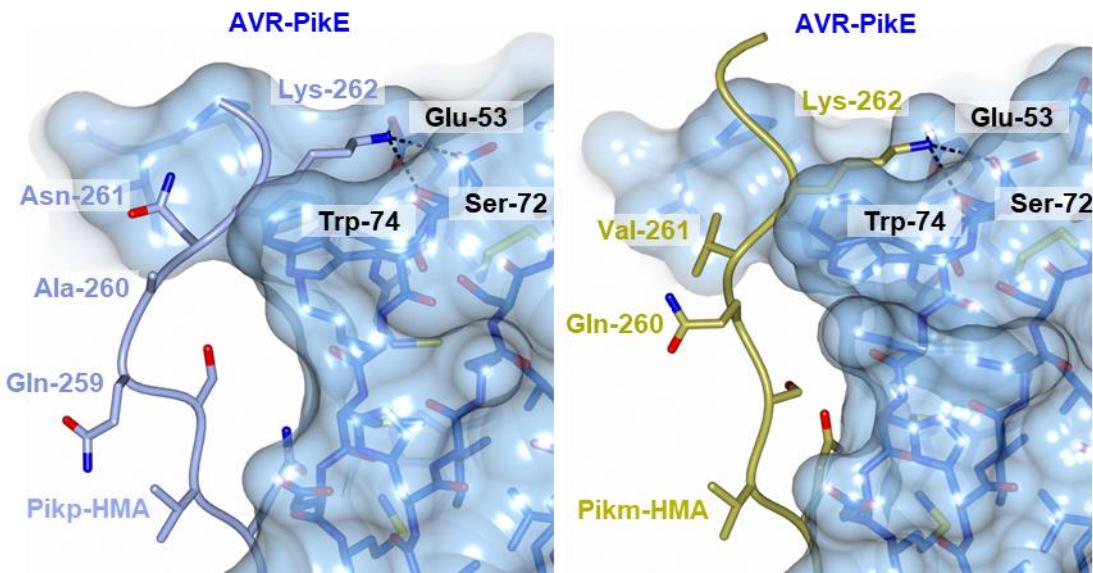
in both Pikp and Pikm have the same numerical order despite occupying a different position). This change results in a different conformation of Pikm-HMA Val 261 and Lys262 when compared to the equivalent Pikp-HMA residues Ala 260 and Asn261 (**Figure 3.22**). As a consequence, Pikm-HMA adopts a more favourable position at this interaction interface with the AVR-Pik effector.



**Figure 3.22. Pikm-HMA adopts a different conformation at interface three compared to Pikp-HMA.** Schematic view of the different conformation adopted by Pikp-HMA and Pikm-HMA, at interface three in complex with AVR-PikD. In each panel, the effector is shown as sticks with the molecular surface also shown and coloured as labelled. Pik-HMAs are shown in Ca-worm with side-chain representation. Pikp-HMA residues are coloured in ice blue and Pikm-HMA are coloured in gold. Residues Gln260 and Val261 in Pikm-HMA adopt a more favourable position compared to the structurally equivalent residues in Pikp-HMA.

When comparing the Pikp-HMA and Pikm-HMA complexes with AVR-PikE, the differences at interface three became more dramatic. As described above, Pikp-HMA must move Asn261 away and adopt an unfavourable conformation to retain Lys262 interactions in the binding pocket of the AVR-PikE effector, affecting the packing of Ala260 and the hydrophobic packing of Lys262 side-chain. (**Figure 3.20b**). However, In Pikm-HMA, the differential position of Lys262 enables a better accommodation and residues Gln260 and Val261 do not move away from the effector interface, ensuring a favourable packing of the HMA chain in the Pikm-HMA/AVR-PikE complex (**Figure 3.23**).

These findings suggest that Pikm has evolved a more robust interaction across interface three. These changes compensate for the loss of interaction to polymorphic AVR-Pik effectors at interface two, maintaining a sufficient level of binding to, ultimately, generate the extended recognition phenotype in Pikm.

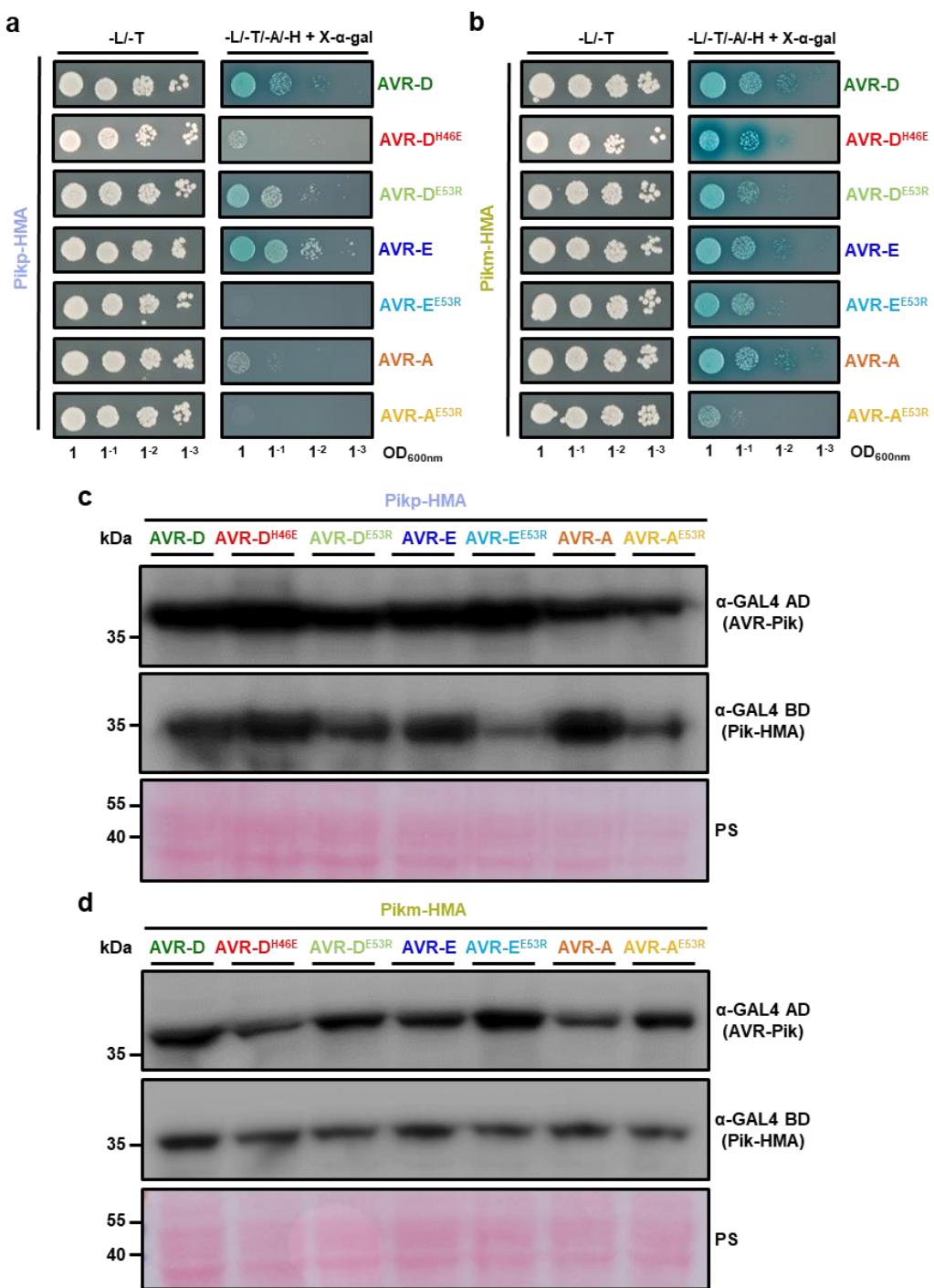


**Figure 3.23. Differential conformation at interface three is the structural determinant for Pikm extended recognition.** Schematic view of the different conformation adopted by Pikp-HMA and Pikm-HMA, at interface three in complex with AVR-PikE. In each panel, the effector is shown as sticks with the molecular surface also shown and coloured as labelled. Pik-HMAs are shown in Ca-worm with side-chain representation. Pikp-HMA residues are coloured in ice blue and Pikm-HMA residues are coloured in gold. Pikp-HMA residues Gln259 and Ala260 are loop out from the interaction interface while the structurally equivalent residues in Pikm-HMA adopt a more favourable position.

### 3.11 Mutations at separate interfaces have differential effects on Pik-HMA/effector binding

Analysis of the Pik-HMA/effector interaction interfaces suggest differential importance of interface two and three in Pikp and Pikm for mediating effector binding. Subsequently, I tested whether mutations at interfaces two and three have differential effects on Pik-HMA binding to AVR-Pik effectors by Y2H.

For interface two, I tested a His46Glu mutation in AVR-PikD (previously characterised in Maqbool et al., 2015). As discussed above, this residue occupies a central position in the interaction at this interface (Figure 3.21a, b). This mutation abrogates the interaction the Pikp-HMA interaction with the effector in Y2H (Figure 3.24a), as previously observed without the C-terminal extension (Maqbool et al., 2015). Interestingly, this mutant retains certain interactions with Pikm-HMA (depicted by growth and blue coloration) (Figure 3.24b), confirming that the loss of interaction at interface 2 in compensated by other interfaces in Pikm.



**Figure 3.24. Effector mutations at positions 46 and 53 perturb interactions with Pikp-HMA and Pikm-HMA.** Yeast-Two-Hybrid assay of (a) Pikp-HMA and (b) Pikm-HMA with AVR-Pik alleles and mutants disrupting either interface two (His46Glu) or interface three (Glu53Arg). Control plate for yeast growth is on the left with quadruple dropout media supplemented with X- $\alpha$ -gal on the right. Growth and development of blue colouration in the selection plate are both indicative of protein:protein interaction. HMA domains were fused to the GAL4 DNA binding domain, and AVR-Pik alleles to the GAL4 activator domain. Each experiment was repeated a minimum of three times, with similar results. (c, d) Western blot analysis confirming accumulation of proteins in yeast for the Y2H assay with (c) Pikp-HMA and (d) Pikm-HMA. Yeast lysate was probed for the expression of HMA domain with anti-GAL4 DNA binding domain (BD) and AVR-Pik effectors anti-GAL4 activation domain (AD).

For interface three, I used Glu53Arg mutants in AVR-PikD, AVR-PikE and AVR-PikA. This mutation localises to the Pik-HMA Lys262 binding pocket. In Pikp, this mutation showed no effect on binding to AVR-PikD as this interaction mostly relies on interface two. However, the Glu53Arg mutant in AVR-PikE abolishes the interaction of this effector with Pikp-HMA in Y2H (**Figure 3.24a**). For Pikm-HMA, little impact on Pikm-HMA interaction in Y2H compared to wild-type was observed, except a reduced interaction to AVR-PikA Glu53Arg (**Figure 3.24b**). Presumably, the mutation at interface three prevents Pikm-HMA from compensating for the disruptions at interface 2 caused by polymorphisms in AVR-PikA. The expression of all the mutants in yeast were confirmed for the Y2H assays with both Pikp-HMA (**Figure 3.24c**) and Pikm-HMA (**Figure 3.24d**).

To quantitatively measure the impact of these mutations, Dr. Marina Franceschetti produced and purified the corresponding AVR-Pik mutants from *E. coli* and tested the interaction by SPR. The results of these experiments are reported in De la Concepcion et al., 2018.

The SPR assay confirmed that AVR-PikD His46Glu mutant abrogates interaction with Pikp-HMA and, despite the retention of binding in Y2H, largely reduces interaction with Pikm-HMA (only ~11% compared to the binding to wild type AVR-PikD) (De la Concepcion et al., 2018). Binding of AVR-Pik Glu53Arg mutants to both Pik-HMA domains is reduced when compared pairwise to each wild-type effector. However, reduction seems to be greater in Pikm-HMA binding than in Pikp-HMA binding (De la Concepcion et al., 2018).

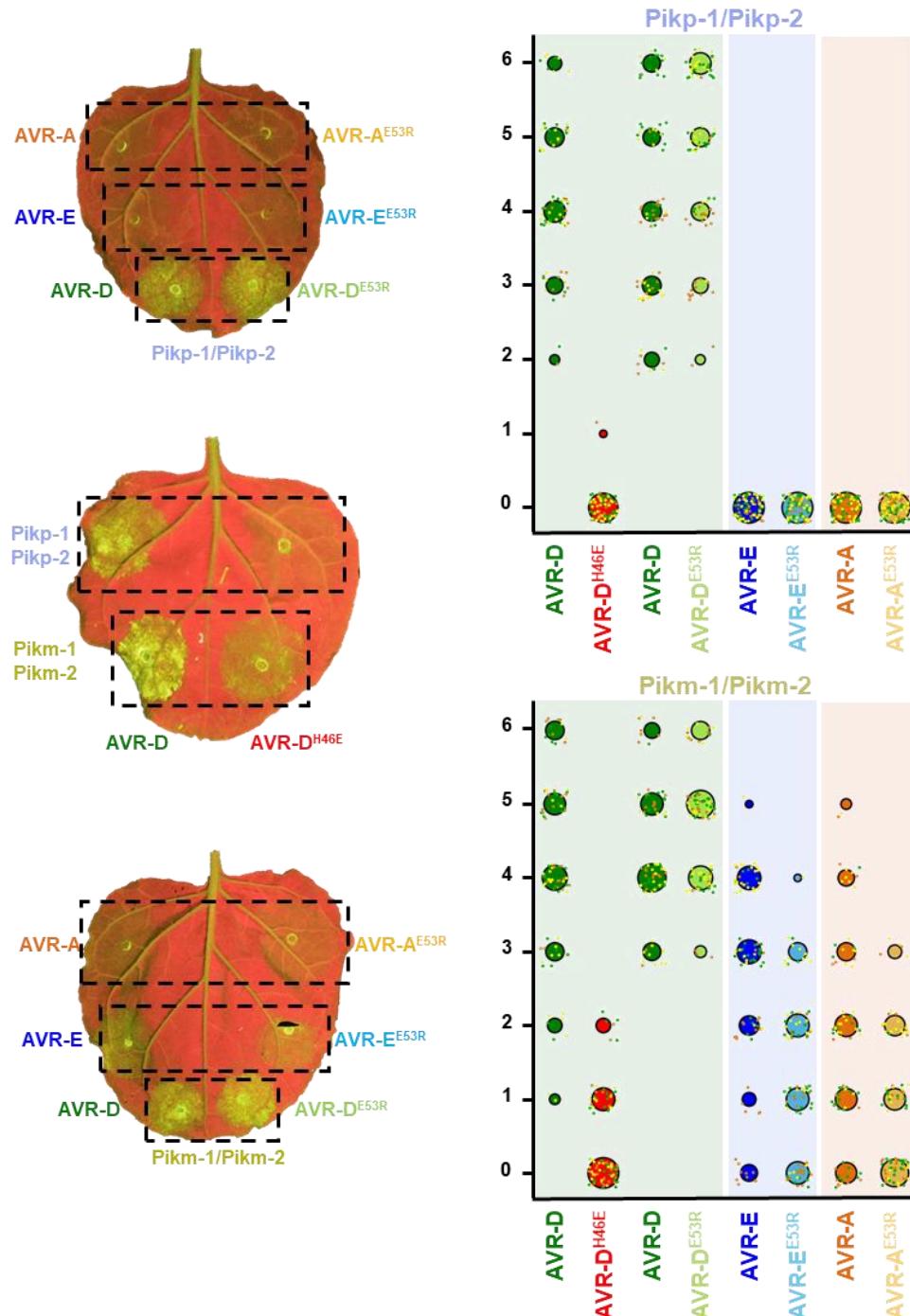
Interaction data by Y2H and SPR point to differential roles for the interaction interfaces used by Pikp-HMA and Pikm-HMA. While Pikp-HMA mostly relies on interface two, this interface is important, but not critical, in Pikm. Pikm-HMA compensates for changes at interface two via improved interactions at interface three.

### **3.12 Mutations at separate interfaces have differential effects on Pikp and Pikm mediated immune phenotypes**

As a proxy to study the differential role that Pik-HMA/AVR-Pik interaction interfaces two and three play in immune recognition by Pik alleles, I performed cell death assays in *N. benthamiana* using the AVR-Pik mutants described above.

In agreement with the lack of direct binding to the AVR-PikD His46Glu mutant, there is no cell death response mediated by Pikp in response to this mutant in planta (**Figure 3.25**). This mutation also largely reduces the Pikm-mediated response, however, the

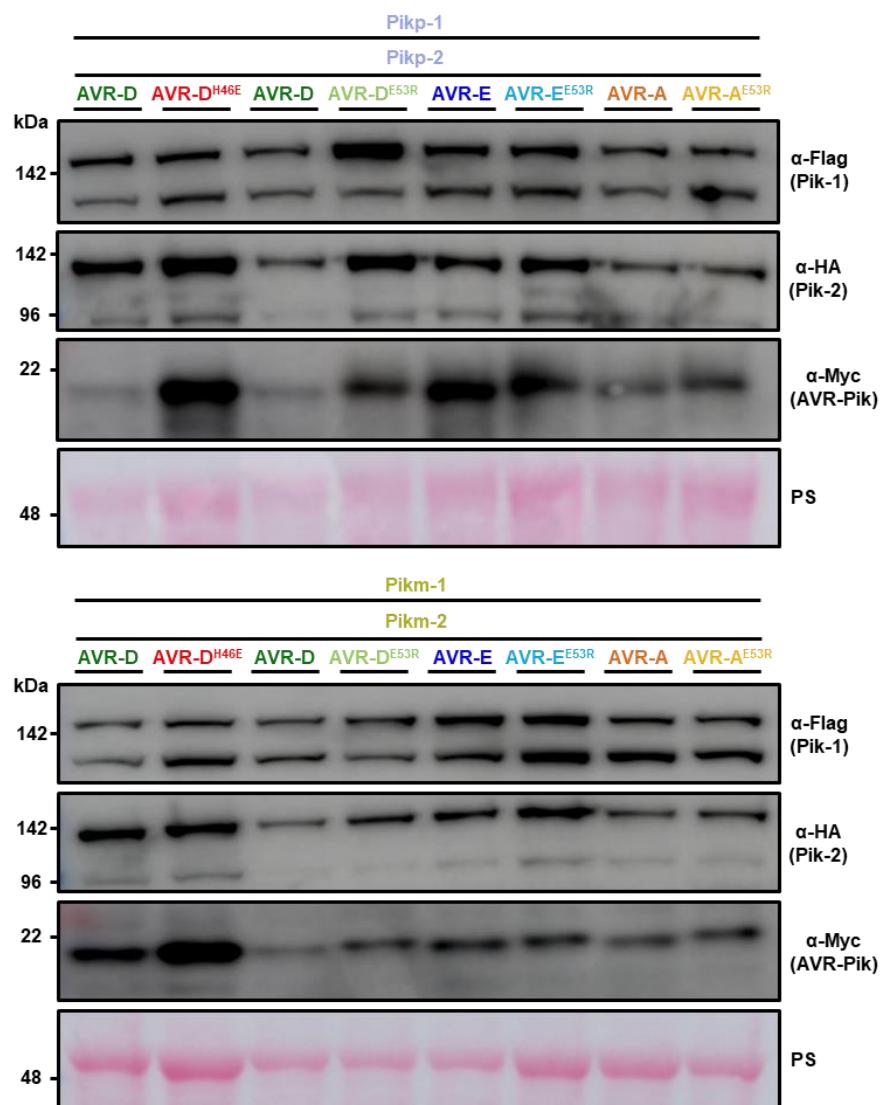
weak binding demonstrated in Y2H and SPR still results in a relatively low response in *N. benthamiana* (**Figure 3.25**).



**Figure 3.25. Mutations at different interfaces in the Pik-HMA/effector complexes have differential effects on cell death phenotypes.** Leaf images showing side-by-side cell death in *N. benthamiana* for the effector mutants at position 46 and 53 and Wild type. For both Pikp-HMA and Pikm-HMA cell death assay scoring represented as dot plots are included. The number of repeats was 90. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

I also compared side by side the AVR-Pik Glu53Arg mutants with the wild-type for both Pikp and Pikm. These mutants show a reduction in the intensity of cell death mediated by Pikm, reflecting the importance of binding at interface three for the recognition of AVR-PikE and AVR-PikA (**Figure 3.25**). Unexpectedly, I could observe a slight increase in cell death response to the Glu53Arg mutant in the AVR-PikD background compared to the wild type in both Pikp and Pikm (**Figure 3.25**). Expression of the effector mutants in planta was confirmed by western blot, to ensure the differences observed are not due changes in protein accumulation (**Figure 3.26**).

The results from the cell death assay further support that binding to the effector in the HMA/AVR-Pik interface three is the determinant of the extended effector recognition phenotype in Pikm.



**Figure 3.26. The reduction in cell death for AVR-Pik mutants is not due lower protein accumulation levels.** Western blots showing expression of AVR-Pik alleles and AVR-Pik mutants in position 46 and 53, as well as, Pikp and Pikm NLRs.

### 3.13 Discussion

Since their discovery in the early 90s, NLR receptors have been intensively studied (Bent et al., 1994; Mindrinos et al., 1994; Whitham et al., 1994). Despite this, the mechanistic basis of how these receptors recognise pathogens and initiate immune signalling is not well understood.

In the last 5 years, the identification of plant NLRs with unconventional domains integrated in their architecture has represented a major breakthrough in the field (Cesari et al., 2014a; Wu et al., 2015). These domains are likely evolved from effector targets and, as such, they directly interact with pathogen effectors in most of the cases studied to date (De la Concepcion et al., 2018; Guo et al., 2018b; Le Roux et al., 2015; Maqbool et al., 2015; Ortiz et al., 2017; Sarris et al., 2015; Varden et al., 2019; Zhang et al., 2017b). Because of their direct interaction with pathogen effectors, and the fact they are widespread across plant genomes (Bailey et al., 2018; Kroj et al., 2016; Sarris et al., 2016), these domains are attractive focus for understanding pathogen recognition by NLRs (Ellis, 2016; Malik and Van der Hoorn, 2016). Thus, the study of integrated domains at the biochemical and structural level will likely shed light onto the mechanism of pathogen recognition by NLRs, enabling the future development of plant cultivars with enhanced disease resistance phenotypes.

In this chapter, I have described five crystal structures of different complexes between rice blast AVR-Pik effectors and the integrated domain of two alleles of the rice NLR Pik with different recognition specificities. Combination of protein:protein interactions in vivo and in vitro, with the study of cell death responses in *N. benthamiana* (summarized in **Table 3.6**), enhances the understanding of the link between pathogen recognition and effector binding to integrated domains. Furthermore, by obtaining the structures of a receptor/effector complexes that have insufficient binding affinity to trigger and effective immune response in planta, I gained structural information about how pathogen effectors evolved to avoid immune recognition, providing unprecedent details to the understanding of the host-pathogen co-evolution.

The integrated HMA domain is the most polymorphic region of the rice Pik-1/Pik-2. After integration, these domains have probably experienced accelerated accumulation of mutations coupled with a strong selective pressure to bind pathogen effectors. This work uncovers the atomic details of how two different alleles of Pik, likely under different selective pressure impose by the presence of different pathogen strains, have undergone

different modifications at the interaction interface with the effector that have led to differential specificity.

This work also uncovered a modified binding interface in *Pikm*-HMA that has evolved more favourable interactions to the polymorphic AVR-Pik effectors. This interface compensates for the disruption of other interfaces, by amino acid changes in the pathogen effectors, ensuring sufficient binding to activate immune responses. This line of data exemplifies how small structural changes in the receptor/effectector binding interfaces can lead to an increased binding affinity to pathogen effectors and, subsequently, to an extended pathogen recognition phenotype. In the following chapter, I will use these insights to establish a proof-of-concept for structure-informed engineering of NLR receptors with extended recognition capabilities.

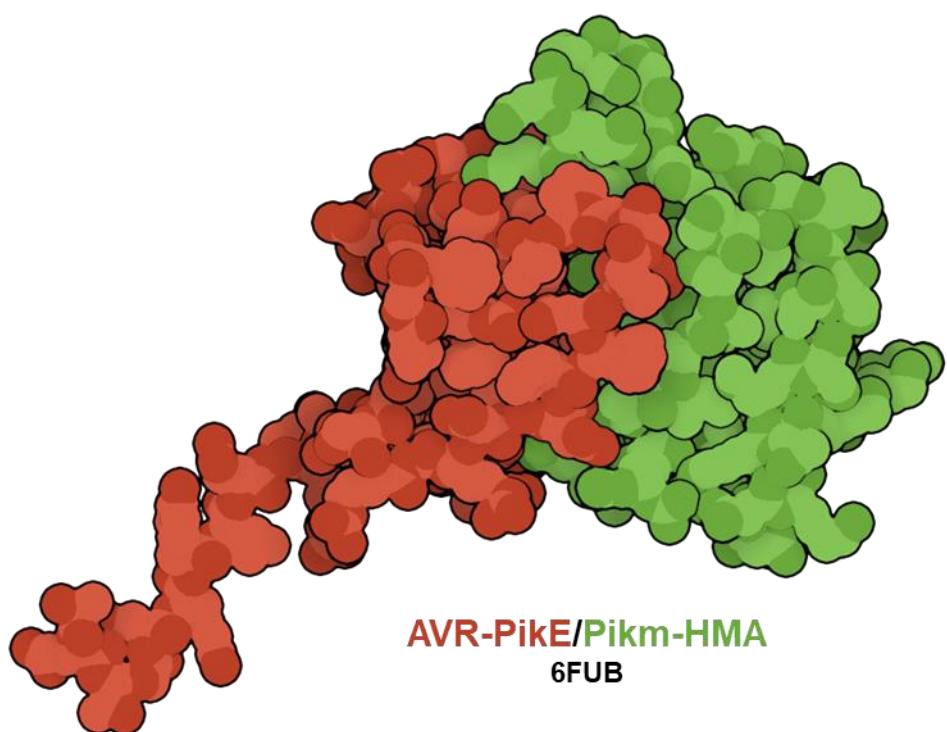
		AVR-D	AVR-E	AVR-A	AVR-C	AVR-D <sup>H46E</sup>	AVR-D <sup>E53R</sup>	AVR-E <sup>E53R</sup>	AVR-A <sup>E53R</sup>
Interaction in	Pikp	+++	++	+	-	+	++	-	-
	Pikm	+++	+++	+++	+	+++	++	++	+
Interaction in	Pikp	+++	++	+	-	-	+++	+	-
	SPR	+++	+++	++	-	+	++	+	-/+
Recognition in	Pikp	+++*	+	(-)	(-)	-*	N.D.	N.D.	N.D.
	rice plants	(+++)	(+++)	(+++)	(-)	N.D.	N.D.	N.D.	N.D.
CD response in	Pikp	+++	-	-	-	-	+++	-	-
<i>N. benthamiana</i>	Pikm	+++	++	+	-	+	+++	+	+

**Table 3.6. Summary of the various interactions and phenotypes between Pik NLR alleles and effector variants.** Interaction values and recognition phenotypes for both *Pikp* and *Pikm* with each AVR-Pik variant obtained in this chapter and by previous studies. Y2H: Yeast-2-Hybrid, SPR: Surface Plasmon Resonance, CD: cell death, N.D: not determined. Recognition in rice plants, when determined, were studied for cultivars K60 (*Pikp*) and Tsuyuake (*Pikm*) in previous studies (those in parenthesis in Kanzaki et al., 2012 and those marked by \* in Maqbool et al., 2015). Note that SPR and Y2H interactions used the isolated HMA domains, while in planta experiments were performed with full length proteins.

## Chapter 3

# 4

## Protein engineering expands the effector recognition profile of a rice NLR immune receptor





## 4. Protein engineering expands the effector recognition profile of a rice NLR immune receptor

### Abstract

Since their discovery in the early 90s, engineering NLR immune receptors to enhance disease resistance has been a scientific goal pursued in plant biotechnology. However, this has proven challenging, in part due to their narrow response specificity but, mainly because the mechanisms of immune response triggered by NLRs are still not fully understood. In the previous chapter, I revealed the structural determinants of effector recognition by the *Pik* NLR receptor, and how subtle structural changes underpin increased recognition specificity in *Pik* natural alleles. Here, I used this information to guide the engineering of the *Pikp* NLR pair to expand the response profile to variants of the rice blast pathogen effector AVR-Pik. By mutating one of the interaction interfaces described previously, *Pikp*-HMA gained increased binding affinity to the AVR-Pik variants. This translated to a gain-of-recognition phenotype to previously unrecognized effectors in *N. benthamiana* cell death assays. I further revealed the mechanistic basis of the expanded effector recognition profile using X-ray crystallography. Overall, the results of this chapter provide a proof-of-concept for engineering NLRs to achieve enhanced disease resistance. This approach has great potential for the application in modern agriculture.

## Chapter 4

## 4.1 Introduction

Plant pathogens are the causative agent of major productivity losses in every commercial crop worldwide (Savary et al., 2019). Since spraying chemicals to restrict the growth of pathogens is costly and can negatively affect the environment, the development of genetic resistance to improve crop resilience to diseases is central in many plant biotechnology programmes.

Plant immune receptors from the NLR family have been targeted for plant improvement in crop breeding programmes early since their discovery (Dangl et al., 2013). However, this has proven difficult to date, partly because NLR receptors characteristically show a narrow specificity, with each receptor recognizing only a particular subset of pathogen effectors. Additionally, NLRs are highly diverse (Meyers et al., 2003) and present multiple different ways to recognize pathogens (Cesari, 2018; Kourelis and van der Hoorn, 2018), which make it difficult to delineate a general improvement strategy.

Most breeding strategies for disease resistance involve the transfer of NLRs between species to acquire resistance, either by conventional crossing or using transgenic technology (Rodriguez-Moreno et al., 2017). Unfortunately, this approach is often unsuccessful because NLRs tend to work in complexes (Bonardi and Dangl, 2012), or as part of networks (Castel et al., 2019; Wu et al., 2017). Therefore, they require a specific genetic context for the successful transfer of disease resistance. Some NLRs have been shown to be suppressed when transferred to a different genetic background (Hurni et al., 2014; Stirnweis et al., 2014a), whereas others activate a constitutive immune response that is deleterious for the plant (Bomblies et al., 2007; Chae et al., 2014; Tran et al., 2017).

Some restricted success on altering NLR performance has been achieved to date, mainly through the use of random mutagenesis. Some studies achieved the extension of recognition profiles in certain plant NLRs, although the translation to disease resistance remained limited (Farnham and Baulcombe, 2006; Giannakopoulou et al., 2015; Harris et al., 2013; Segretin et al., 2014).

Exchange of domains between closely related NLR homologs has been also used as an approach to alter pathogen recognition specificity (Slootweg et al., 2017). Although this approach enabled the partial exchange of immune recognition from viruses to nematodes, it also gave rise to autoactive and loss-of-function phenotypes in the NLR chimeras (Slootweg et al., 2017), highlighting once more how major changes in NLRs can lead to undesired effects.

## Chapter 4

Using a subtler strategy, Kim et al. achieved an exchange in NLR recognition from bacteria to viruses by modifying the cleavage site of a protease effector target (Kim et al., 2016). Although this approach is limited to pathogens that translocate proteases into the host (Giannakopoulou et al., 2016), the conservation of this protease recognition system in crop plants makes this strategy a valuable approach worth pursuing in the future (Carter et al., 2019; Helm et al., 2019).

In recent years, the discovery that NLR integrated domains are widespread across many plant NLRs (Baggs et al., 2017; Bailey et al., 2018; Cesari et al., 2014a; Kroj et al., 2016; Sarris et al., 2016; Wu et al., 2015), together with the study of how these domains interact with effectors to enable pathogen recognition (Cesari et al., 2013; De la Concepcion et al., 2018; Guo et al., 2018b; Le Roux et al., 2015; Maqbool et al., 2015; Ortiz et al., 2017; Sarris et al., 2015; Varden et al., 2019; Zhang et al., 2017b), opened new possibilities for engineering novel pathogen recognition capabilities in plant NLRs.

The integrated HMA domain in the sensor NLR *Pikp*-1 directly associates with the rice blast pathogen effector AVR-PikD, triggering an immune response dependent on the helper NLR *Pikp*-2 (Kanzaki et al., 2012; Maqbool et al., 2015; Yoshida et al., 2009). In the previous chapter, I investigated the structural basis of natural variation in AVR-Pik recognition by *Pik* alleles. Whereas the *Pikp* NLR allele only responds to the effector variant AVR-PikD, the *Pikm* allele responds to AVR-PikD, AVR-PikE, and AVR-PikA (De la Concepcion et al., 2018; Kanzaki et al., 2012). These differences in effector recognition are underpinned by structural differences in the receptor/effector binding interfaces (De la Concepcion et al., 2018). This detailed knowledge of direct effector binding offers the potential for modifying the interactions between *Pik*-HMA and AVR-Pik effectors by resurfacing the interaction interfaces, with a possible impact on immune signalling.

Protein resurfacing changes the properties of solvent-exposed regions, often by mutation, and has been previously used for biotechnological applications (Chapman and McNaughton, 2016). This chapter describes how I used mutations at two residues in *Pikp*-HMA (Asn261Lys, Lys262Glu) to combine favourable interactions identified at the binding interfaces of *Pikp*-HMA and *Pikm*-HMA, generating a *Pik* NLR receptor with an improved recognition profile. These mutations extended the *Pikp* effector recognition profile to AVR-PikE and AVR-PikA in planta. By combining methods to study protein:protein interactions in vitro, in vivo, and in planta, I demonstrated that the *Pikp*<sup>NK-KE</sup> mutant has increased binding to AVR-Pik effectors, establishing a causal link to gain-

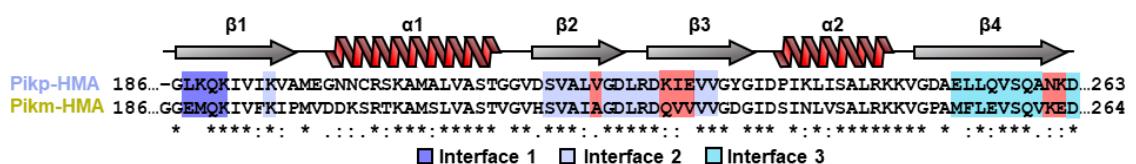
of-function phenotypes in *planta*. Furthermore, I used X-ray crystallography to understand the structural basis of the  $\text{Pikp}^{\text{NK-KE}}$  extended effector recognition phenotype.

This work is a proof-of-concept for the use of structure-informed protein engineering to develop plant NLR immune receptors with new capabilities. The approach developed in this chapter has the potential to lead the improvement of disease resistance in crops, particularly in these cases where the pathogen effector is directly recognized through binding to the NLR receptor.

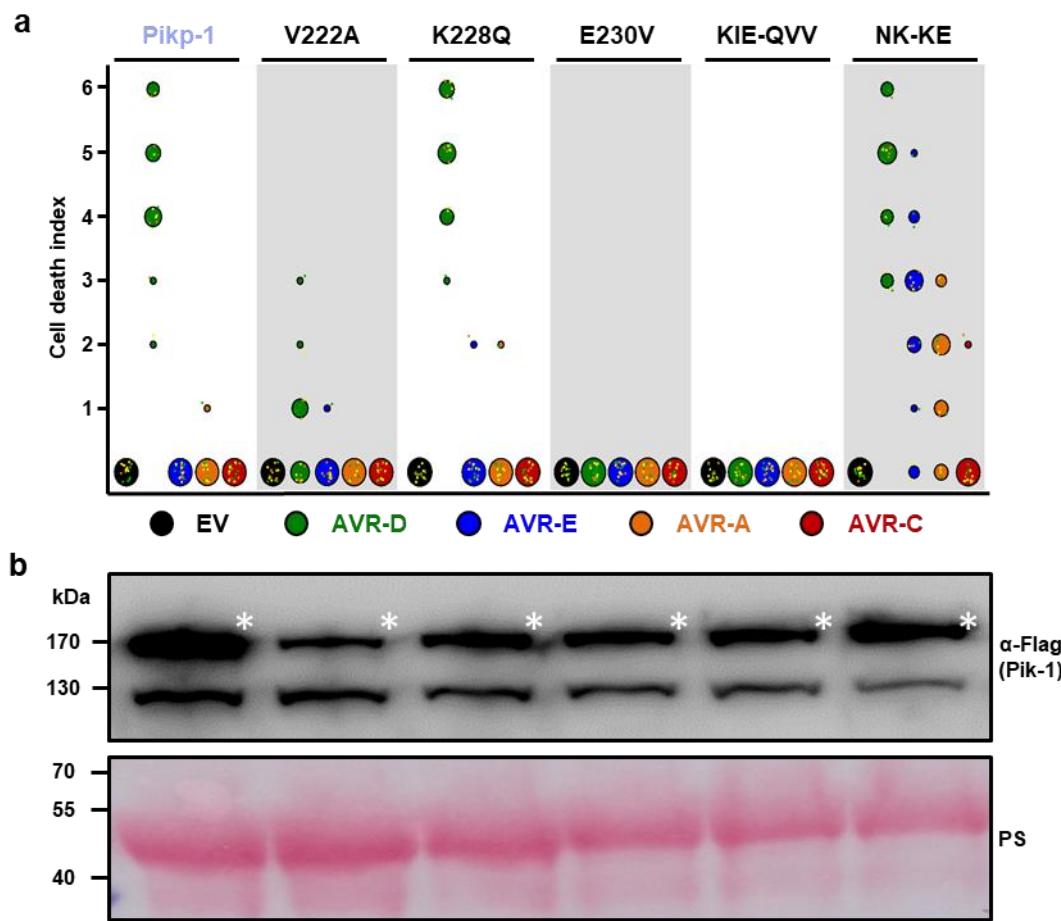
## 4.2 Structure-informed engineering expands **Pikp** effector recognition in *N. benthamiana* cell death assays

Structural investigation of **Pik-HMA/AVR-Pik** in the previous chapter revealed interfaces two and three as the most extensive contact points, with variation at interface three being the determinant for the extended effector recognition by **Pikm** (De la Concepcion et al., 2018). It was hypothesised that this information could be used to identify point mutations to expand the recognition capabilities of **Pikp**.

To investigate this, Dr. Marina Franceschetti performed an unbiased high-throughput cell death screening. **Pikp-1** mutants with residues at interface two and three exchanged by those in the structurally equivalent position in **Pikm-1** (**Figure 4.1, in red**) were co-expressed with **Pikp-2** and different **AVR-Pik** effector variants (**Figure 4.2**). A double mutant in **Pikp** showed an extended cell death response to **AVR-PikE** and **AVR-PikA** (**Figure 4.2, Figure 4.3**). Consistent with the results showed in the previous chapter, these mutations, **Asn261Lys** and **Lys262Glu** (hereafter **Pikp<sup>NK-KE</sup>**), were in two adjacent amino acids contained within interface three (**Figure 4.1**). I carried out a western blot for all the mutants to ensure the differences in cell death responses were not due differential protein accumulation (**Figure 4.2**). I subsequently focused on the characterization of the extended recognition phenotype of the **Pikp<sup>NK-KE</sup>** mutant.

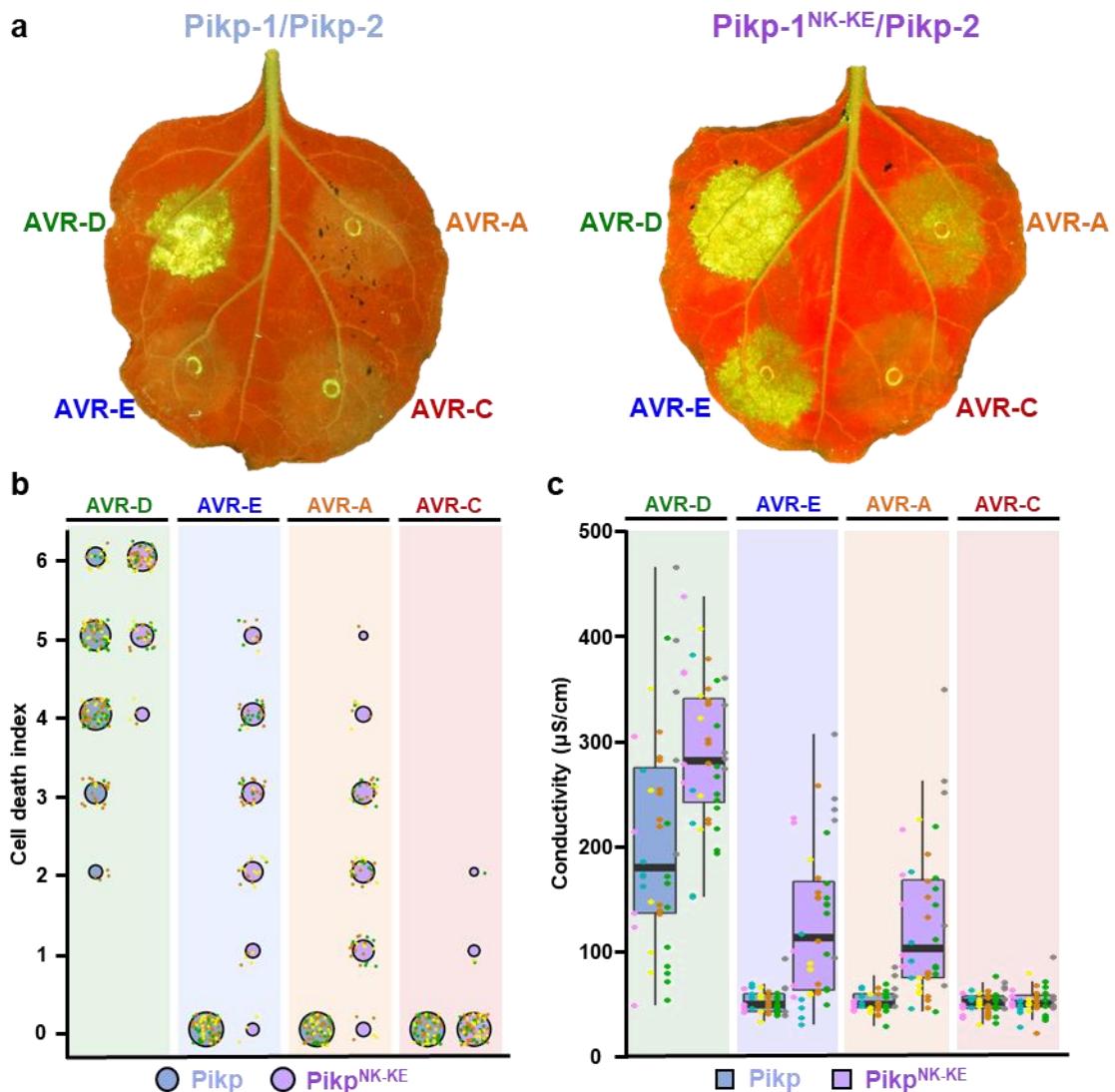


**Figure 4.1. Sequence alignment of **Pikp-1** and **Pikm-1** HMA domains.** Secondary structure features of the HMA fold are shown above, and the residues located to binding interfaces are coloured in purple, ice blue or cyan for interfaces one, two or three, respectively. Key residues from interfaces two and interface three involved in the extended cell death phenotype screening are highlighted in red.



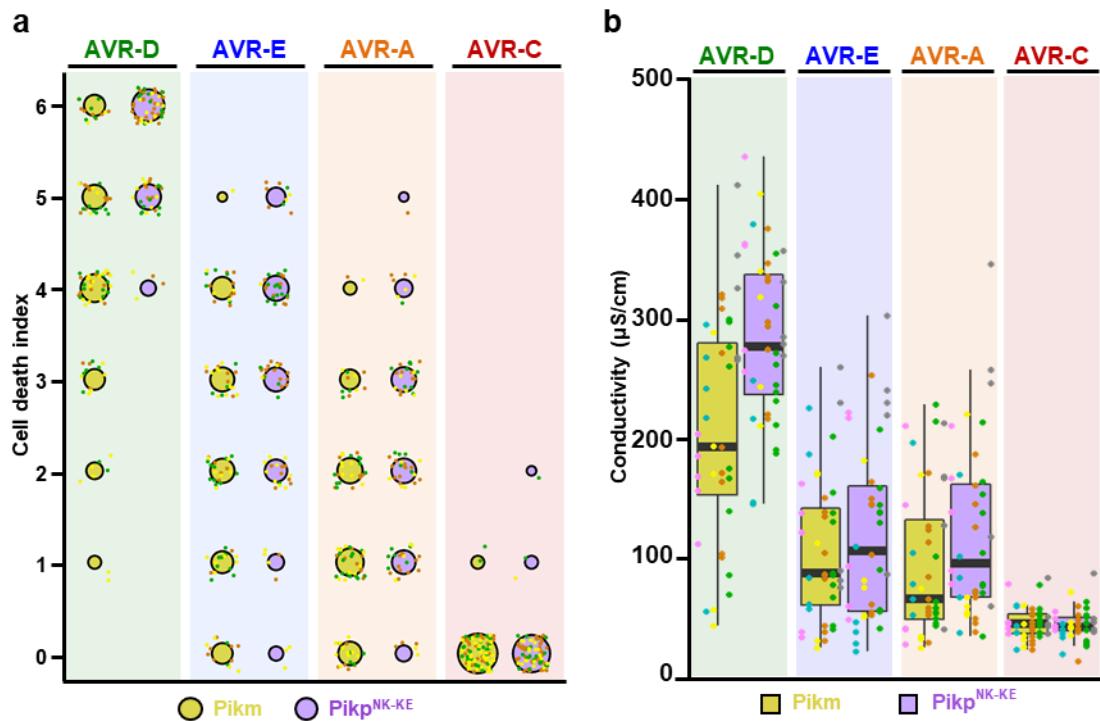
**Figure 4.2. High-throughput cell death screening reveal the effects of mutations at Pikp-HMA interface two and three.** (a) UV scoring of cell death represented as dot plots for Pikp-1 mutants on HMA interface 2 and 3. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score. The number of repeats was 18 for each mutant. (b) Western blot analysis confirming similar levels of Pik-1 protein accumulation in *N. benthamiana*. The asterisks mark the Pik-1 band.

To ensure the robustness of the observed phenotype, I first repeated the cell death assay increasing the number of repeats. This assay was quantified both by scoring UV fluorescence (**Figure 4.3b**) and by quantifying ion leakage (**Figure 4.3c**). Overall, the results showed that  $\text{Pikp}^{\text{NK-KE}}$  triggers a cell death response in the presence of AVR-PikD, AVR-PikE and AVR-PikA, while the wild-type  $\text{Pikp}$  receptor only responds to AVR-PikD (**Figure 4.3**).  $\text{Pikp}^{\text{NK-KE}}$  did not trigger cell death when co-expressed with the stealthy effector AVR-PikC.



**Figure 4.3. NK-KE mutation at Pikp-HMA/AVR-Pik binding interface three extends cell-death phenotype to AVR-PikE and AVR-PikA.** (a) Representative *N. benthamiana* leaf images depicting Pikp and Pikp<sup>NK-KE</sup>-mediated cell death as autofluorescence under UV light. (b) Cell death fluorescence scoring represented as dot plots comparing Pikp and Pikp<sup>NK-KE</sup>. The number of repeats was 80 in each case. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score. For Pikp, data was also shown in the previous chapter (Figure 3.2) (c) Ion leakage quantification comparing Pikp and Pikp<sup>NK-KE</sup>-mediated cell death responses in *N. benthamiana* measured as an increase in conductivity. The centre line represents the median, the box limits are the upper and lower quartiles, the whiskers extend to the largest value within Q1 - 1.5x the interquartile range (IQR) and the smallest value within Q3 + 1.5x IQR. All the data points are represented as dots with distinct colours for each biological replicate. For each experiment, we performed six biological replicates with 5 or 10 internal repeats, a total of 40 replicates. For Pikp, data was also shown in the previous chapter (Figure 3.2).

Similar to *Pikm*, *Pikp*<sup>NK-KE</sup> shows a hierarchy in the effector triggered cell death of AVR-PikD>AVR-PikE>AVR-PikA (**Figure 4.3**). I further compared *Pikp*<sup>NK-KE</sup> and *Pikm* cell death responses with AVR-Pik effectors. Results of both, UV scoring and ion leakage quantification showed that cell death responses to AVR-Pik D, AVR-PikE and AVR-PikA mediated by *Pikp*<sup>NK-KE</sup> are equivalent compared to those mediated by *Pikm*, although they seem consistently elevated (**Figure 4.4a, b**). This is also observed when comparing the cell death of *Pikp*<sup>NK-KE</sup> with *Pikp* in response to AVR-PikD (**Figure 4.3**).



**Figure 4.4. *Pikp*<sup>NK-KE</sup> displays a cell death phenotype comparable to *Pikm* but consistently elevated.** (a) Cell death assay scoring represented as dot plots comparing *Pikm* and *Pikp*<sup>NK-KE</sup>. The number of repeats was 90 and 80, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score. For *Pikm*, data was shown in the previous chapter (**Figure 3.2**) (b) Ion leakage assay comparing *Pikm* and *Pikp*<sup>NK-KE</sup>-mediated cell death responses in *N. benthamiana*. The centre line represents the median, the box limits are the upper and lower quartiles, the whiskers extend to the largest value within  $Q1 - 1.5 \times \text{IQR}$  and the smallest value within  $Q3 + 1.5 \times \text{IQR}$ . All the data points are represented as dots with distinct colours for each biological replicate. For each experiment, we performed six biological replicates with 5 or 10 internal repeats, a total of 40 replicates. For *Pikm*, data was shown in the previous chapter (**Figure 3.2**).

These results show that a double mutation, Asn261Lys/Lys262Glu, at the HMA/AVR-Pik binding interface three, expands the recognition profile of Pikp. Therefore, by modifying the structural determinants of receptor/effectector interaction we can manipulate the recognition profile of NLR receptors.

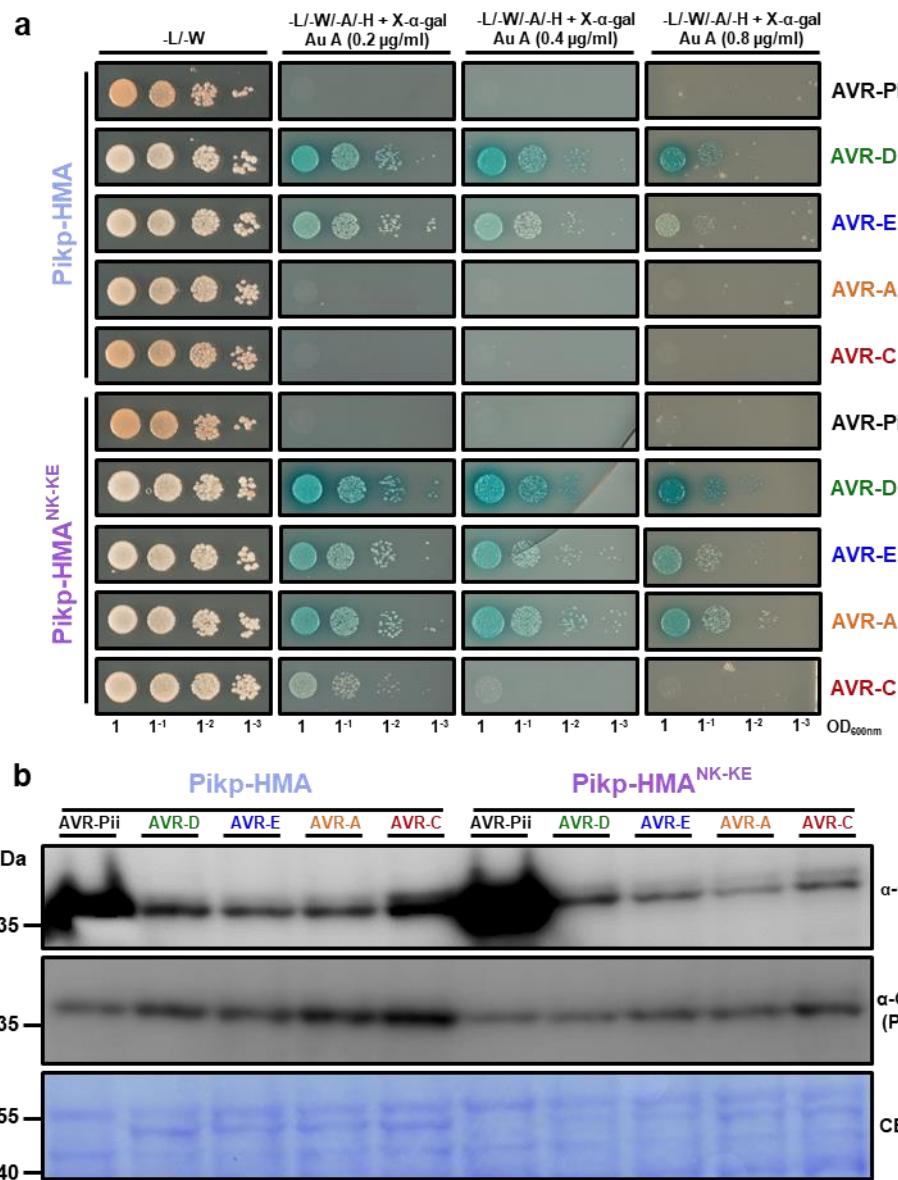
### 4.3 The engineered Pikp-HMA<sup>NK-KE</sup> mutant shows increased binding to effector variants in vivo compared to Pikp-HMA

To determine if the extended cell death responses mediated by Pikp<sup>NK-KE</sup> correlate with an increase in binding to the AVR-Pik effector, I carried out a Y2H assay, comparing the binding of Pikp-HMA and Pikp-HMA<sup>NK-KE</sup> to AVR-PikD, AVR-PikE, AVR-PikA and AVR-PikC (**Figure 4.5**). An unrelated *M. oryzae* effector, AVR-Pii, was included as negative control.

Yeast cells expressing the different Pik-HMA/AVR-Pik combinations were deposited on selective plates with increased stringency compared to Y2H assays presented in chapter three, imposed by increasing concentrations of Aureobasidin A. I observed an increase in growth and blue coloration for Pikp-HMA<sup>NK-KE</sup> with AVR-PikE and AVR-PikA when compared with Pikp-HMA, and this difference was especially evident with more stringent conditions (**Figure 4.5a**). Unexpectedly, there was some yeast growth for Pikp-HMA<sup>NK-KE</sup> with AVR-PikC at the lowest stringency (**Figure 4.5a**). No interaction was detected for the AVR-Pii negative control (**Figure 4.5a**). Accumulation of all proteins in yeast was confirmed by western blot (**Figure 4.5b**).

These results indicate that the engineered HMA domain has higher binding to AVR-PikE and AVR-PikA compared to Pikp-HMA, correlating with the cell death phenotypes in planta. Furthermore, this assay showed a discrete gain of binding of Pikp-HMA<sup>NK-KE</sup> to AVR-PikC, an effector previously not bound by Pikp-HMA nor Pikm-HMA.

## Chapter 4

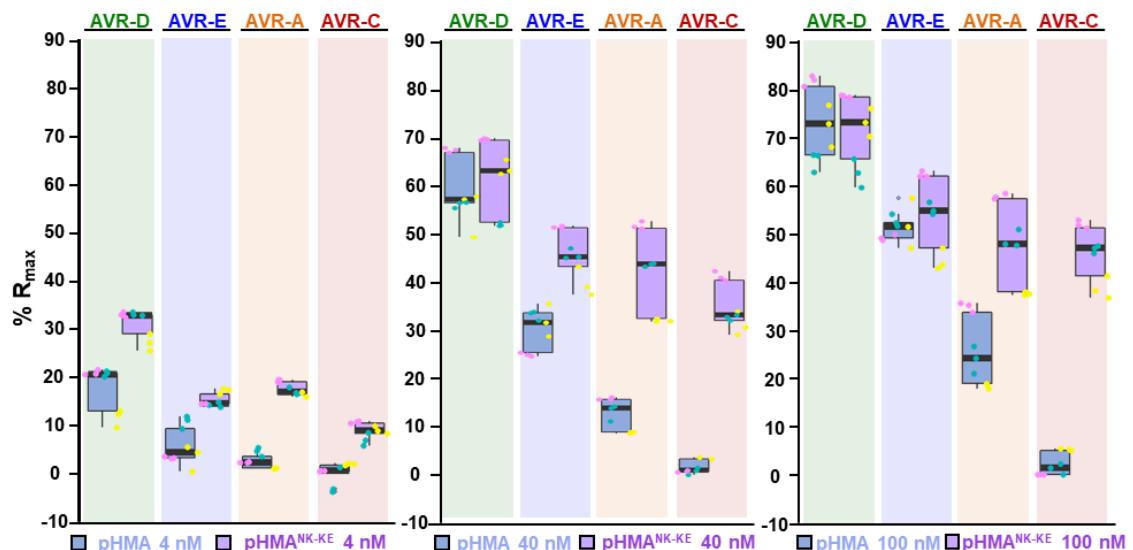


**Figure 4.5. Pikp-HMA<sup>NK-KE</sup> displays increased binding to AVR-Pik variants in vivo. (a)** Yeast-Two-Hybrid assay of Pikp-HMA and Pikp-HMA<sup>NK-KE</sup> with AVR-Pik alleles. Control plate for yeast growth is on the left with quadruple dropout media supplemented with X- $\alpha$ -gal and increasing concentrations of Aureobasidin A on the right. Growth and development of blue colouration in the selection plate are both indicative of protein:protein interactions. For each HMA, an unrelated *M. oryzae* effector, AVR-Pii, serves as negative control. HMA domains were fused to the GAL4 DNA binding domain, and AVR-Pik alleles to the GAL4 activator domain. Each experiment was repeated a minimum of three times, with similar results. **(b)** Western blot analysis confirming accumulation of proteins in yeast. Yeast lysate was probed for the expression of HMA domain with anti-GAL4 DNA binding domain (BD) and AVR-Pik effectors anti-GAL4 activation domain (AD). CBS stands for Coomassie Blue staining.

## 4.4 In vitro binding affinity to AVR-Pik effectors is increased in Pikp-HMA<sup>NK-KE</sup>

I aimed to quantify the increased Pikp-HMA<sup>NK-KE</sup> binding to AVR-Pik effectors observed in the Y2H assay. To do this, I first overexpressed and purified the engineered Pikp-HMA<sup>NK-KE</sup> domain to homogeneity from *E. coli*, as detailed in chapter three for Pikm-HMA (see also materials and methods).

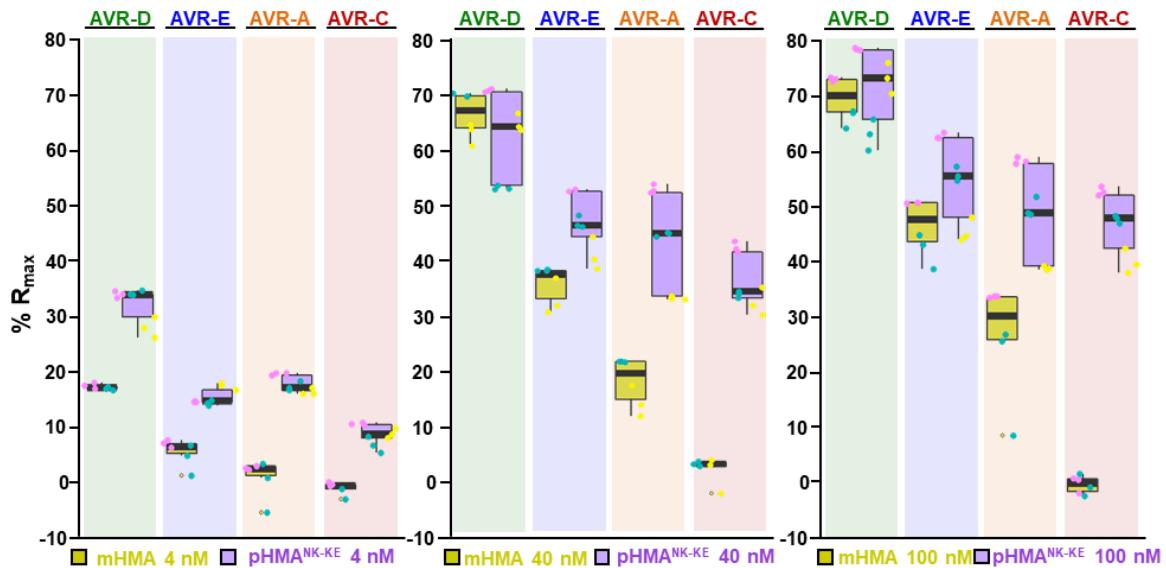
Then, I used SPR to measure the binding of Pikp-HMA<sup>NK-KE</sup> to AVR-Pik effectors, alongside Pikp-HMA and Pikm-HMA for comparison (Figure 4.6 and Figure 4.7). I captured the AVR-Pik effectors onto a Biacore NTA chip, and measured response units (RUs) after injecting the Pik-HMAs at 4 nM, 40 nM and 100 nM concentration. RUs were normalised to the  $R_{max}$  (theoretical maximum response, as explained in materials and methods), assuming a 1:2 AVR-Pik/Pik-HMA interaction model for Pikp-HMA and Pikp-HMA<sup>NK-KE</sup>, and 1:1 for Pikm-HMA.



**Figure 4.6. Binding of Pikp-HMA<sup>NK-KE</sup> to the AVR-Pik effectors is consistently higher compared to Pikp-HMA in vitro.** Box plots showing %Rmax for Pikp-HMA (Blue) and Pikp-HMA<sup>NK-KE</sup> (purple) with the AVR-Pik effectors alleles at HMA concentration of, from left to right, 4 nm, 40 nM and 100 nM as measured by surface plasmon resonance. The centre line represents the median, the box limits are the upper and lower quartiles, the whiskers extend to the largest value within Q1 - 1.5x the interquartile range (IQR) and the smallest value within Q3 + 1.5x IQR. All the data points are represented as dots with distinct colours for each biological replicate. For each experiment, three biological replicates with three internal repeats were performed.

This assay revealed an increased binding of Pikp-HMA<sup>NK-KE</sup> to all AVR-Pik effectors compared to wild-type Pikp-HMA in all the measured Pik-HMA concentrations (Figure 4.6). The differences in binding to AVR-PikE and AVR-PikA correlate with the cell death

response phenotypes observed in *N. benthamiana* (**Figure 4.3**). The binding of Pikp-HMA<sup>NK-KE</sup> to the AVR-Pik effectors was comparable but also consistently higher compared to Pikm-HMA (**Figure 4.7**) and also correlates with the slight increase in cell death observed by ion leakage and UV scoring (**Figure 4.3**).



**Figure 4.7. Binding of the Pikp-HMA<sup>NK-KE</sup> domain to the AVR-Pik effectors is consistently higher compared to Pikm-HMA in vitro.** Box plots showing %R<sub>max</sub> for Pikm-HMA (gold) and Pikp-HMA<sup>NK-KE</sup> (purple) with the AVR-Pik effectors alleles at HMA concentration of, from left to right, 4 nM, 40 nM and 100 nM as measured by surface plasmon resonance. The centre line represents the median, the box limits are the upper and lower quartiles, the whiskers extend to the largest value within Q1 - 1.5x the interquartile range (IQR) and the smallest value within Q3 + 1.5x IQR. All the data points are represented as dots with distinct colours for each biological replicate. For each experiment, at least two biological replicates with three internal repeats were performed.

Interestingly, Pikp-HMA<sup>NK-KE</sup> showed a considerable gain of binding to AVR-PikC in vitro, even though neither Pikp-HMA nor Pikm-HMA shows binding to this effector (**Figure 4.6**, **Figure 4.7**). This was similar to the result from Y2H assays, although binding in vitro was more significant to the binding observed in yeast (**Figure 4.5**).

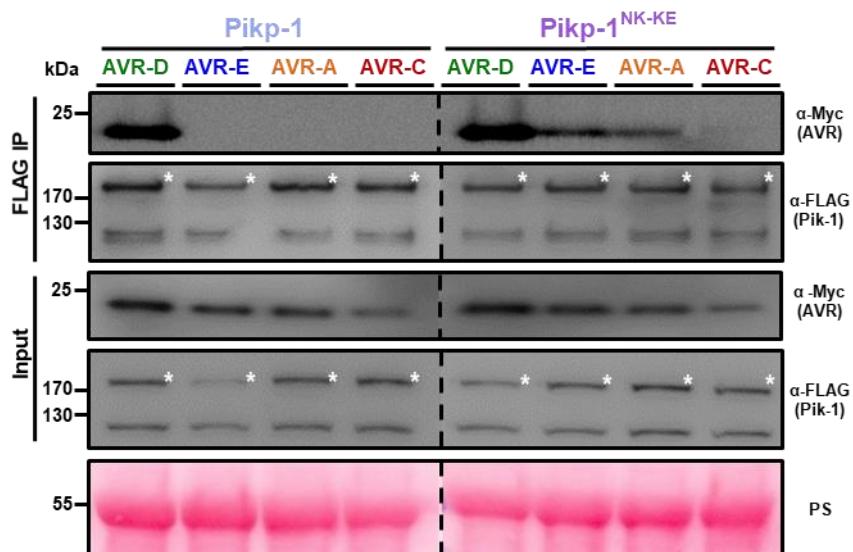
As for Pikp-HMA in the previous chapter, Pikp-HMA<sup>NK-KE</sup> binds non-specifically to the NTA chip at high concentration. Therefore, when I tried to determine the equilibrium dissociation constant,  $K_D$ , by extending the SPR to multicycle kinetics, the Biacore T100 evaluation software reported the assay of insufficient quality.

Altogether, the SPR and Y2H results indicate that Pikp-HMA<sup>NK-KE</sup> has a higher binding affinity for AVR-Pik effectors compared to wild-type Pikp-HMA. This points to an increased binding to AVR-PikE and AVR-PikA as the determinant for the extended cell death response to those effectors.

## 4.5 The engineered sensor NLR $\text{Pikp-1}^{\text{NK-KE}}$ has extended association to AVR-Pik effector variants in planta

Given that the isolated  $\text{Pikp-HMA}^{\text{NK-KE}}$  domain has increased binding to polymorphic AVR-Pik effectors *in vivo* and *in vitro*, I sought to investigate whether this is translated to the context of the full-length protein in planta. For this, I co-expressed either wild-type  $\text{Pikp-1}$  or  $\text{Pikp-1}^{\text{NK-KE}}$  with the AVR-Pik effector variants in *N. benthamiana*, followed by co-immunoprecipitation (co-IP) and western blotting (Figure 4.8).

Under these assay conditions, AVR-PikD co-immunoprecipitates with  $\text{Pikp-1}$  robustly. However, despite AVR-PikE and AVR-PikA interacting with  $\text{Pikp-HMA}$  by Y2H and SPR, there is no detectable association to the full length  $\text{Pikp-1}$  in planta (Figure 4.8), which matches the cell death phenotypes observed in *N. benthamiana*.



**Figure 4.8.  $\text{Pikp-1}^{\text{NK-KE}}$  associates with AVR-Pik effector alleles in planta.** Co-immunoprecipitation of full length  $\text{Pikp-1}$  and  $\text{Pikp-1}^{\text{NK-KE}}$  with AVR-Pik variants. N-terminally 4xMyc tagged AVR-Pik effectors were transiently co-expressed with  $\text{Pikp-1}:6x\text{His}3x\text{FLAG}$  (left) or  $\text{Pikp-1}^{\text{NK-KE}}:6x\text{His}3x\text{FLAG}$  (right) in *N. benthamiana*. Immunoprecipitates obtained with anti-FLAG antiserum, and total protein extracts, were probed with appropriate antisera. A dashed line indicates a crop site on the same blot used to compose the figure. Each experiment was repeated at least three times, with similar results. The asterisks mark the  $\text{Pikp-1}$  band.

By contrast, AVR-PikE and AVR-PikA clearly associate with the engineered NLR receptor  $\text{Pikp-1}^{\text{NK-KE}}$  (Figure 4.8). This association is reduced in AVR-PikE and AVR-PikA compared with AVR-PikD, correlating with the hierarchical cell death response observed in planta (Figure 4.8). Although co-expression of  $\text{Pikp-1}^{\text{NK-KE}}$  and AVR-PikC does not result in macroscopic cell death in *N. benthamiana* (Figure 4.3), I also observed

a very weak association of AVR-PikC to  $\text{Pikp}^{\text{NK-KE}}$ , particularly at high exposure (**Figure 4.8**).

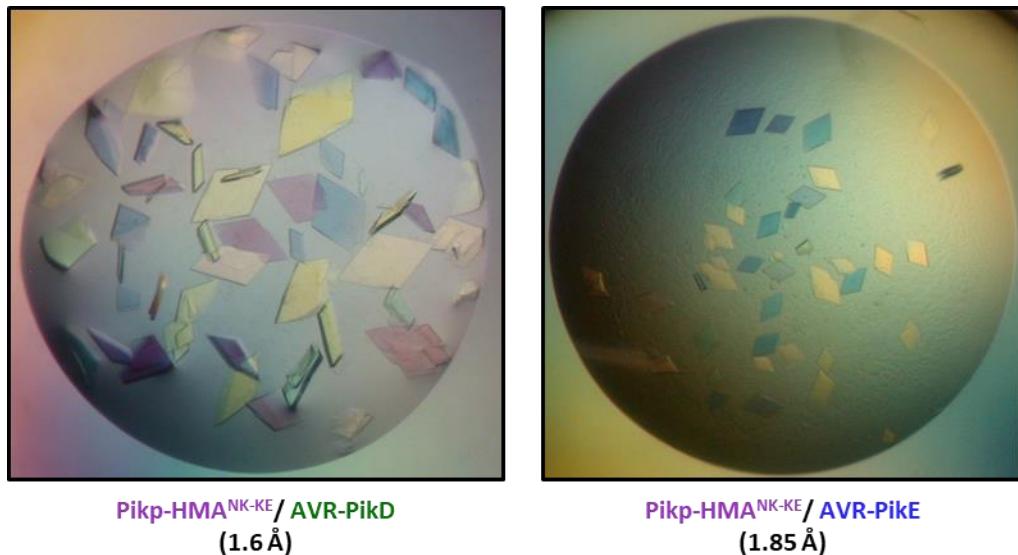
These results show that binding to the engineered HMA domain by AVR-Pik alleles translates to binding to the full-length receptor in planta. This demonstrates that NLR receptors can be successfully engineered to modify their binding affinity to pathogen effectors in planta.

#### 4.6 The engineered $\text{Pikp-HMA}^{\text{NK-KE}}$ adopts a more favourable conformation at the binding interface with the AVR-Pik effectors

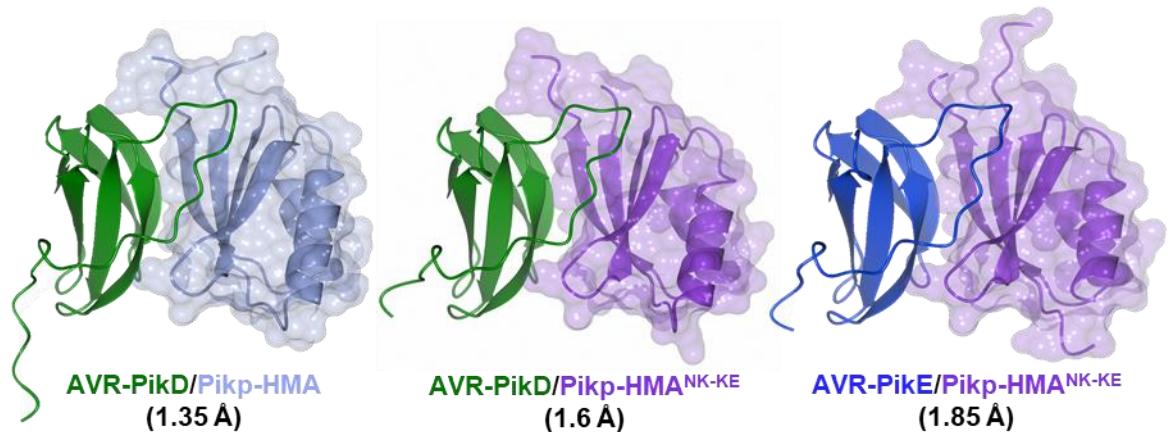
To understand the mechanistic basis of how  $\text{Pikp}^{\text{NK-KE}}$  extends binding and recognition to AVR-Pik effectors, I obtained the crystal structures of  $\text{Pikp-HMA}^{\text{NK-KE}}$  bound to AVR-PikD, and to AVR-PikE. For this, I expressed and co-purified the complexes as described in the previous chapter for the  $\text{Pikm-HMA/AVR-Pik}$  complexes.

Homogeneous samples of  $\text{Pikp-HMA}^{\text{NK-KE}}/\text{AVR-PikD}$  and  $\text{Pikp-HMA}^{\text{NK-KE}}/\text{AVR-PikE}$  were obtained in high concentrations and used for setting up trials with the Morpheus® HT-96 commercial crystallisation screen. Large, diamond-shape crystals appeared overnight (**Figure 4.9**) (see Materials and Methods for conditions and concentrations). Crystals were harvested by Dr. Clare Stevenson (JIC crystallography platform) and sent to Diamond Light Source (UK). X-ray diffraction data were collected by Prof. David Lawson (JIC crystallography platform) to 1.6 Å and 1.85 Å resolution for  $\text{Pikp-HMA}^{\text{NK-KE}}/\text{AVR-PikD}$  and  $\text{Pikp-HMA}^{\text{NK-KE}}/\text{AVR-PikE}$ , respectively. Dr. Marina Franceschetti solved and refined the protein structures for these complexes. Details of X-ray data collection, structure solution, and completion are given in **Table 4.1**.

Overall, the architecture of the  $\text{Pikp-HMA}^{\text{NK-KE}}/\text{AVR-Pik}$  complexes do not present any major structural differences compared with the complexes obtained previously (root mean square deviation (rmsd) for  $\text{Pikp-HMA/AVR-PikD}$  with  $\text{Pikp-HMA}^{\text{NK-KE}}/\text{AVR-PikD}$  was 0.45 Å, based on 972 atoms and 1.3 Å for  $\text{Pikp-HMA/AVR-PikD}$  with  $\text{Pikp-HMA}^{\text{NK-KE}}/\text{AVR-PikE}$ , based on 1185 atoms) (**Figure 4.10**). This points to subtle structural changes as determinants for the extended recognition in the engineered  $\text{Pikp-HMA}^{\text{NK-KE}}$ .



**Figure 4.9. Pik-HMA<sup>NK-KE</sup>/AVR-Pik complexes can be successfully crystallized.** Images of crystals of Pik-HMA<sup>NK-KE</sup> in complex with AVR-PikD and AVR-PikE obtained with the Morpheus® HT-96 crystallisation screen and visualised with polarising light to show birefringence.



**Figure 4.10. NK-KE mutations do not produce major structural changes in the Pik-HMA/AVR-Pik complexes.** Schematic representations of the overall structures of Pikp-HMA<sup>NK-KE</sup> in complex with AVR-PikD and AVR-PikE. Pikp-HMA<sup>NK-KE</sup> are shown in purple cartoon representation with the molecular surface also shown. The effectors are coloured as labelled and shown in cartoon representation. For clarity, only the Pik-HMA molecule making extensive contacts with the effector is shown. Pikp-HMA/AVR-PikD complex presented in chapter 3 is included as a comparison (left).

## Chapter 4

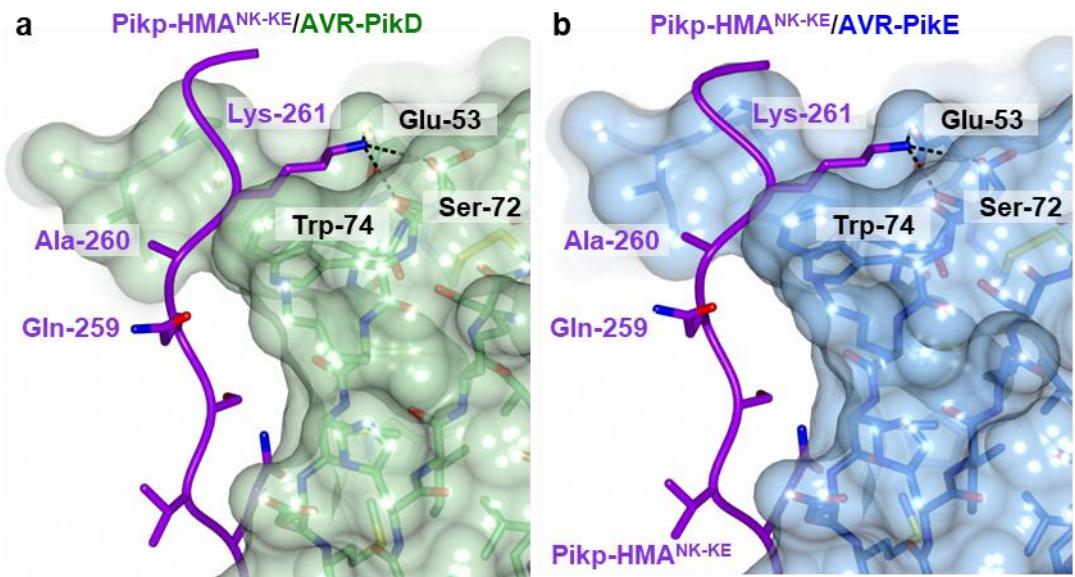
	Pikp-HMA <sup>NK-KE</sup> /AVR-PikD	Pikp-HMA <sup>NK-KE</sup> /AVR-PikE
<b>Data collection statistics</b>		
Wavelength (Å)	0.9763	0.9763
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions		
<i>a, b, c</i> (Å)	29.79, 65.33, 75.86	66.46, 80.70, 105.58
Resolution (Å)*	32.80-1.60 (1.63-1.60)	29.50-1.85 (1.89-1.85)
<i>R</i> <sub>merge</sub> (%) <sup>#</sup>	8.1 (97.1)	5.2 (75.1)
<i>l</i> / <i>σ</i> <sup>#</sup>	16.1 (2.6)	31.0 (4.1)
Completeness (%) <sup>#</sup>	100 (100)	99.8 (97.8)
Unique reflections <sup>#</sup>	20294 (978)	49337 (2963)
Redundancy <sup>#</sup>	12.8 (13.3)	18.3 (17.8)
CC <sup>(1/2)</sup> (%) <sup>#</sup>	99.9 (86.6)	100 (95.2)
<b>Refinement and model statistics</b>		
Resolution (Å)	32.82-1.60 (1.64 – 1.60)	29.52-1.85 (1.90 – 1.85)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%) <sup>^</sup>	19.7/23.2 (25.5/27.3)	18.6/23.0 (29.1/35.0)
No. atoms (Protein)	1277	3604
B-factors (Protein)	25.6	39.7
R.m.s deviations <sup>^</sup>		
Bond lengths (Å)	0.009	0.012
Bond angles (°)	1.5	1.4
Ramachandran plot (%)**		
Favoured	98.1	97.3
Outliers	0	0.2
MolProbity Score	1.41 (93 <sup>th</sup> percentile)	1.59 (91 <sup>st</sup> percentile)

**Table 4.1. X-ray data collection and refinement statistics for Pikp-HMA<sup>NK-KE</sup>/AVR-PikD and Pikp-HMA<sup>NK-KE</sup>/AVR-PikE complexes.** \*The highest resolution shell is shown in parenthesis. <sup>#</sup> As calculated by Aimless. <sup>^</sup> As calculated by Refmac5. <sup>\*\*</sup> As calculated by MolProbity.

As shown for Pikp-HMA and Pikm-HMA, interface three in the Pikp-HMA<sup>NK-KE</sup>/AVR-Pik complexes also involves interactions of a lysine residue with the effector. In this interface, the Lys261 side-chain of Pikp-HMA<sup>NK-KE</sup> forms intimate contacts with the side-chains of Glu53 and Ser72 in the surface pocket of the effector (**Figure 4.11**).

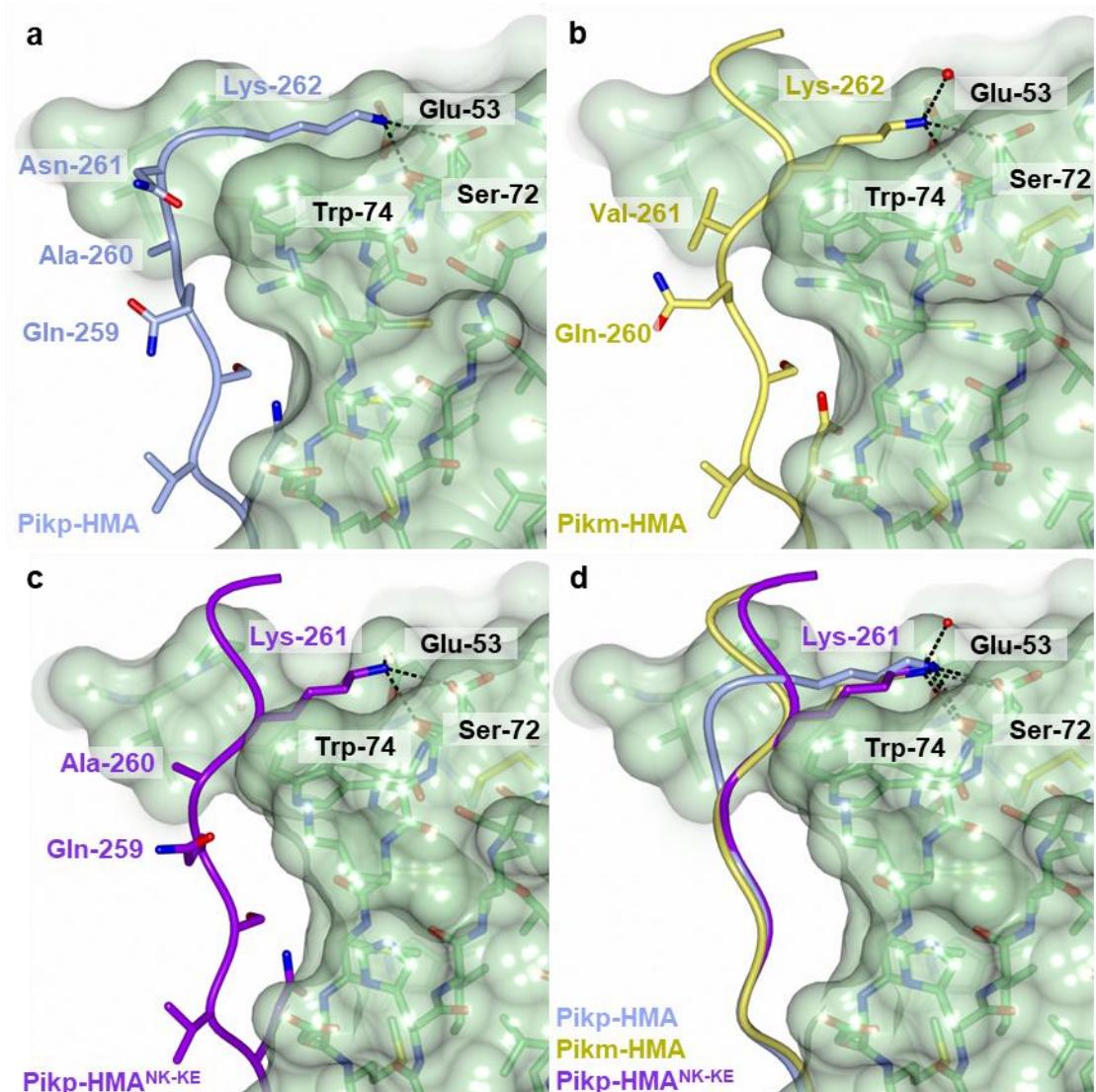
In the crystal structures obtained in the previous chapter, Pikp-HMA adopted an unfavourable conformation in order to place the side-chain of Lys262 in the AVR-Pik effector pocket. However, because Lys262 is placed one position towards the N-terminus in Pikm-HMA, it adopts a more favourable position. The Lys residue in Pikp-HMA<sup>NK-KE</sup> is also displaced one position towards the N-terminus compared to wild-type Pikp-HMA. Comparisons between the structures of Pikp-HMA, Pikm-HMA and Pikp-HMA<sup>NK-KE</sup> bound

to AVR-PikD show that this change enables Pikp-HMA<sup>NK-KE</sup> to adopt the more favourable Pikm-HMA-like conformation (**Figure 4.12**).

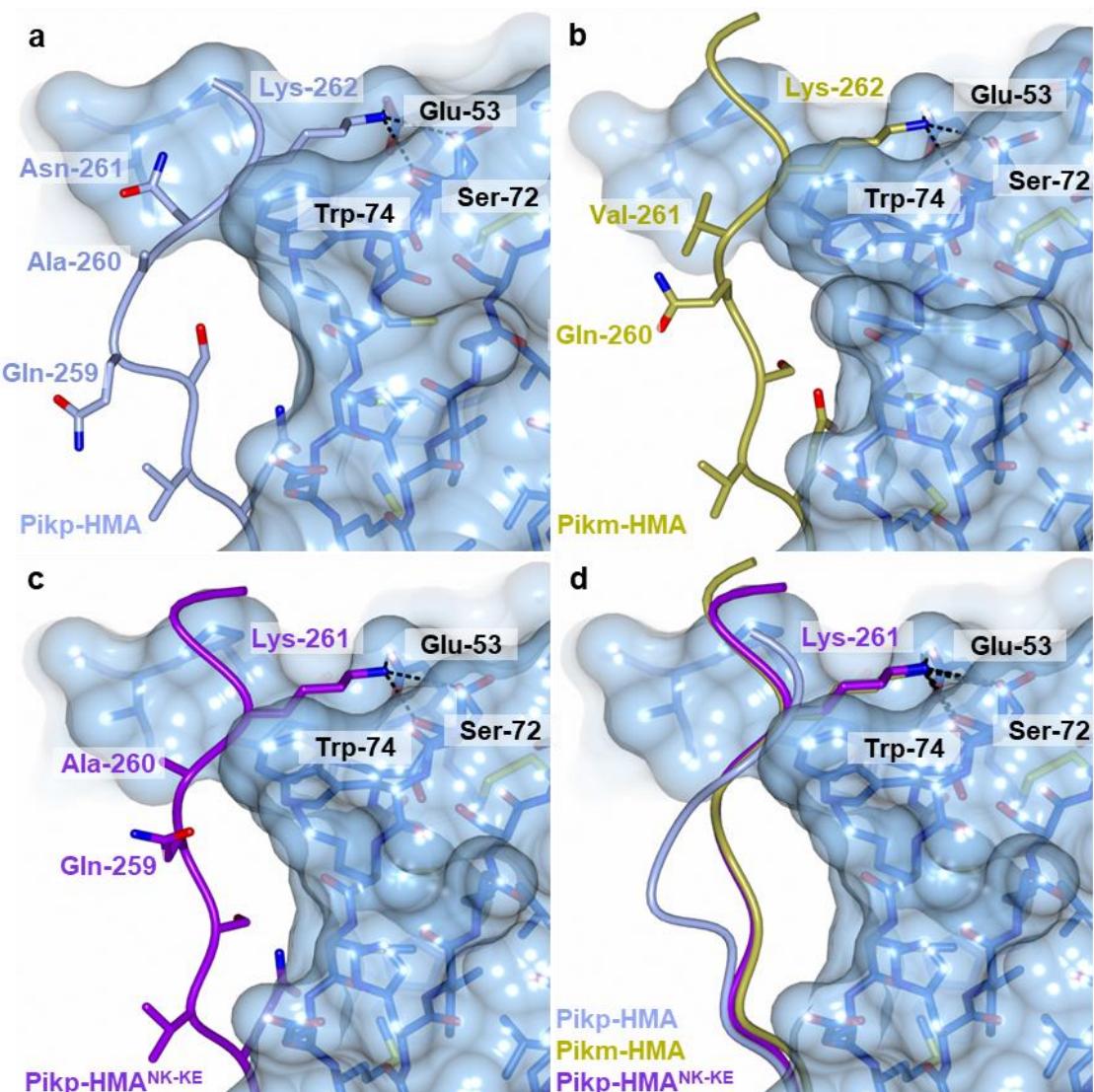


**Figure 4.11. Lys261 in Pikp-HMA<sup>NK-KE</sup> plays a central role in the interaction at interface three.** Close-up view of the interactions across interface three in the (a) Pikp-HMA<sup>NK-KE</sup>/AVR-PikD and (b) Pikp-HMA<sup>NK-KE</sup>/AVR-PikE complexes. AVR-Pik effectors are coloured and shown in cartoon form, with side chains as sticks. The molecular surface of the effector is also shown. The Pikp-HMA<sup>NK-KE</sup> domains are coloured in purple and represented in Ca-worm with sidechain representation.

In the Pikp-HMA/AVR-PikE structure, Pikp-HMA must loop-out regions adjacent to Lys262 to position it in the effector pocket, compromising the packing at the interface and impacting binding affinity. By contrast, Pikm-HMA does not adopt this conformation and maintains a robust interaction with AVR-PikE that translates to immune recognition. In the crystal structure of Pikp-HMA<sup>NK-KE</sup> in complex with AVR-PikE, residues adjacent to the Lys261 are not misplaced and adopt a conformation like Pikm-HMA (**Figure 4.13**). This enables a stronger interaction across interface three compared with the wild-type Pikp-HMA, conferring an increased binding to AVR-Pik effectors and, subsequently, an extended recognition phenotype.



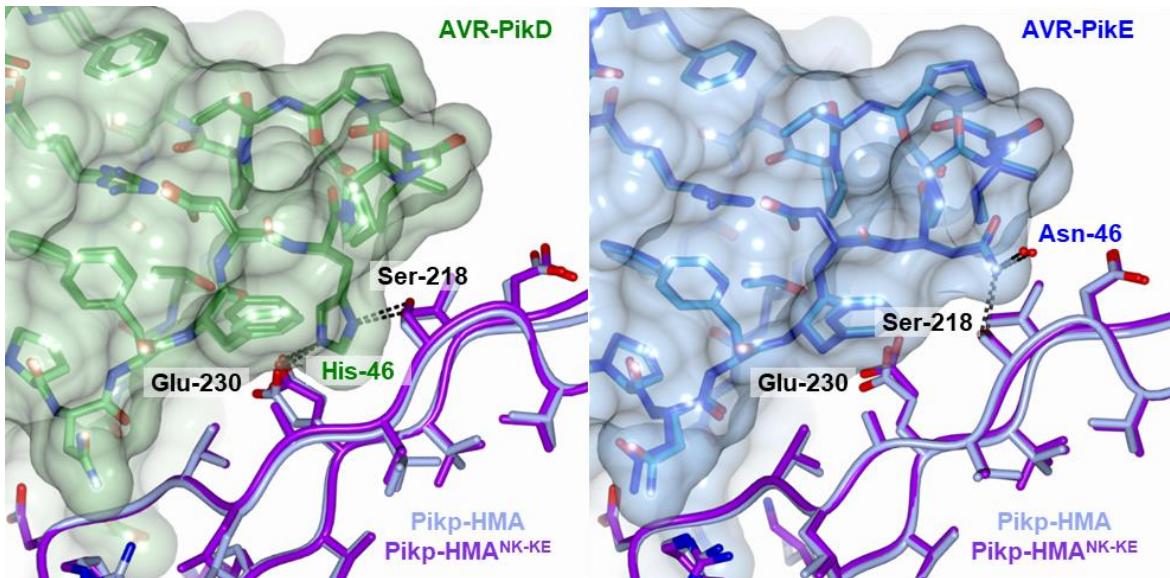
**Figure 4.12.  $\text{Pikp-HMA}^{\text{NK-KE}}$  adopts a different conformation across interface three, compared to  $\text{Pikp-HMA}$ .** Close-up view of the interactions across interface three in the (a)  $\text{Pikp-HMA/AVR-PikD}$ , (b)  $\text{Pikm-HMA/AVR-PikD}$  and (c)  $\text{Pikp-HMA}^{\text{NK-KE}}/\text{AVR-PikD}$ . AVR-PikD is shown in green cartoon, with side chains as sticks, the molecular surface of the effector is also shown. The Pik-HMA domains are coloured as labelled and represented in Ca-worm with sidechain representation. (d) Superposition of  $\text{Pikp-HMA}$ ,  $\text{Pikp-HMA}^{\text{NK-KE}}$  and  $\text{Pikm-HMA}$  chains bound to AVR-PikD. Only side chain of the Lys-261/262 side chain is shown for clarity.  $\text{Pikp-HMA}^{\text{NK-KE}}$  adopts a different conformation at interface three compared to  $\text{Pikp-HMA}$ , more similar to  $\text{Pikm-HMA}$ .



**Figure 4.13.  $\text{Pikp-HMA}^{\text{NK-KE}}$  binds AVR-PikE with a more favourable conformation across interface three, compared to  $\text{Pikp-HMA}$ .** Close-up view of the interactions across interface three in the (a)  $\text{Pikp-HMA/AVR-PikE}$ , (b)  $\text{Pikm-HMA/AVR-PikE}$  and (c)  $\text{Pikp-HMA}^{\text{NK-KE}}/\text{AVR-PikE}$  structures. AVR-PikE is shown in blue cartoon, with side chains as sticks, the molecular surface of the effector is also shown. The Pik-HMA domains are coloured as labelled and represented in Ca-worm with sidechain representation. (d) Superposition of  $\text{Pikp-HMA}$ ,  $\text{Pikp-HMA}^{\text{NK-KE}}$  and  $\text{Pikm-HMA}$  chains bound to AVR-PikE. Only the Lys-261/262 side chain is shown for clarity.  $\text{Pikp-HMA}^{\text{NK-KE}}$  adopts a conformation similar to  $\text{Pikm-HMA}$  at interface three. Compared to  $\text{Pikp-HMA}$ , the HMA chain is not displaced from the effector surface, adopting a more favourable position.

On the other hand, comparisons between Pikp-HMA and  $\text{Pikp-HMA}^{\text{NK-KE}}$  at interface two showed only limited structural perturbations between the AVR-PikD or AVR-PikE effectors bound to Pikp-HMA or  $\text{Pikp-HMA}^{\text{NK-KE}}$  (**Figure 4.14**). This confirms that the effect of the mutations in the engineered NLR are limited to interface three.

Altogether, the crystal structures of the  $\text{Pikp-HMA}^{\text{NK-KE}}$  complexes confirmed that the engineered NLR adopts a more favourable position at the resurfaced interface with the effector.



**Figure 4.14. Pik-HMA has the same conformation across interface two in the  $\text{Pikp-HMA}$  and  $\text{Pikp-HMA}^{\text{NK-KE}}$  complexes with AVR-PikD or AVR-PikE.** Schematic view of the conformations adopted by  $\text{Pikp-HMA}$  and  $\text{Pikp-HMA}^{\text{NK-KE}}$  in complex with AVR-PikD or AVR-PikE at interface two. In each panel, the effector is coloured as labelled and represented as sticks with the molecular surface also shown. Pik-HMA chains are coloured as labelled and shown in the Ca-worm with side-chain representation. The structures were superpose matching the effector molecules.

## 4.7 Mutations in AVR-Pik effectors at the engineered interface impacts binding in vivo

To further investigate the involvement of the engineered binding interface on the increase of binding to AVR-Pik effectors, I performed a Y2H assay for *Pikp-HMA<sup>NK-KE</sup>* using mutants in the effectors at interface two (His46Glu) and interface three (Glu53Arg) described in chapter three (**Figure 4.15**).

Interestingly, Y2H shows that the *AVR-PikD<sup>H46E</sup>* mutant, where interface two is disrupted, interacts with *Pikp-HMA<sup>NK-KE</sup>* (**Figure 4.15a**), similar to the interaction observed between this mutant and *Pikm-HMA* (**Figure 3.3a**). This result suggests that disruption at interface two can be compensated for by the engineered interface three.

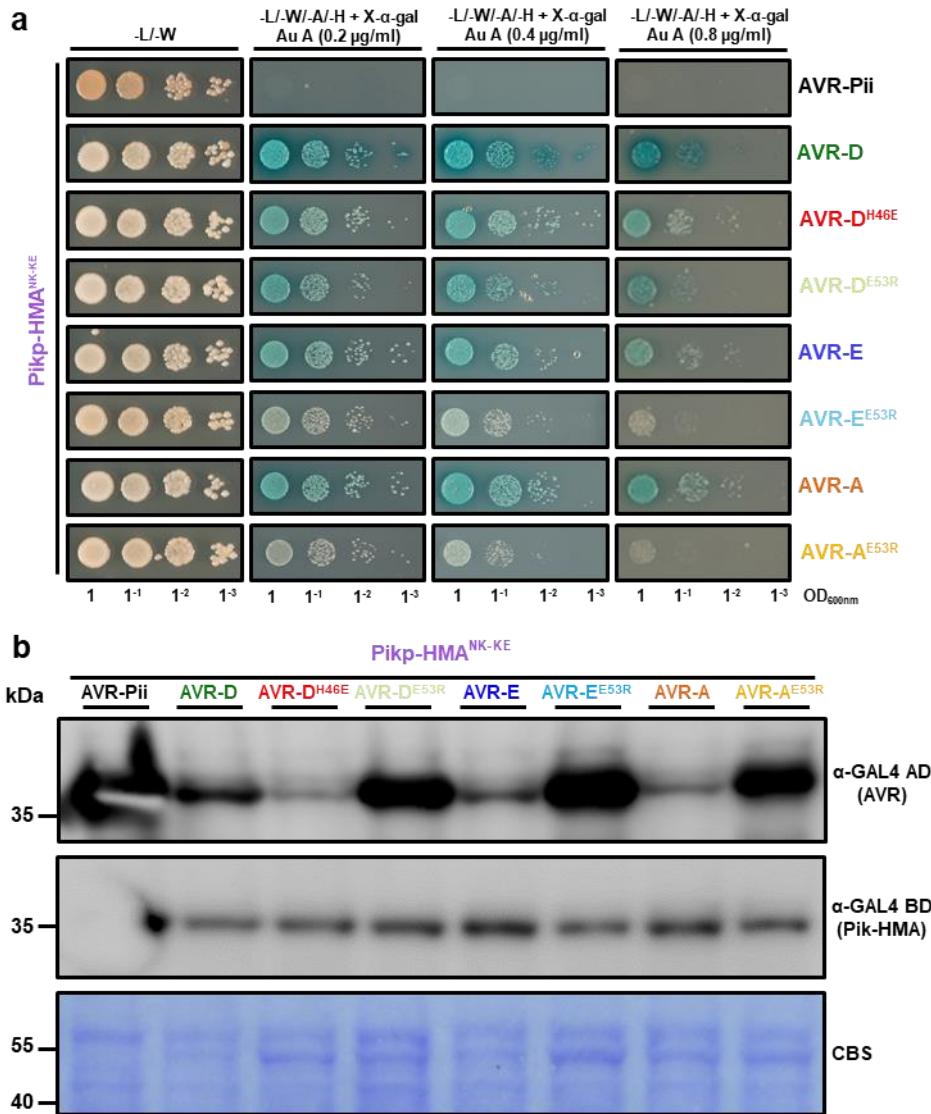
A Glu53Arg mutation at interface three in *AVR-PikD* did not have a substantial effect in the binding to *Pikp-HMA<sup>NK-KE</sup>* (**Figure 4.15a**), probably because the disruption in interface three can be complemented by interactions across interface two. By contrast, the same mutation in the *AVR-PikE* or *AVR-PikA* backgrounds displayed reduced binding, as shown by the reduced growth and blue colouration, particularly at higher stringency (**Figure 4.15a**). This confirms that binding to interface three is important for interaction of *Pikp-HMA<sup>NK-KE</sup>* with *AVR-PikE* and *AVR-PikA*. Expression of all proteins in yeast was confirmed by western blot (**Figure 4.15b**). Glu53Arg mutants were consistently overexpressed in yeast cells compared to their wild-type counterparts, which supports the observations about the lack of binding in these cases.

I also investigated how mutations in interface two and three affect association to the full-length receptor in planta using co-IP. Contrary to the results obtained in the Y2H, a His46Glu mutation in *AVR-PikD* abrogates the association to *Pikp-1<sup>NK-KE</sup>* (**Figure 4.16**). This suggests that the engineered NLR still relies on interface two for an efficient binding to AVR-Pik effectors in planta.

In agreement with the Y2H experiments, the Glu53Arg mutation in *AVR-PikD* has no effect on the association to *Pikp-1<sup>NK-KE</sup>*. However, there is a reduced association (depicted by a less intense band) in the *AVR-PikE* and *AVR-PikA* backgrounds (**Figure 4.16**).

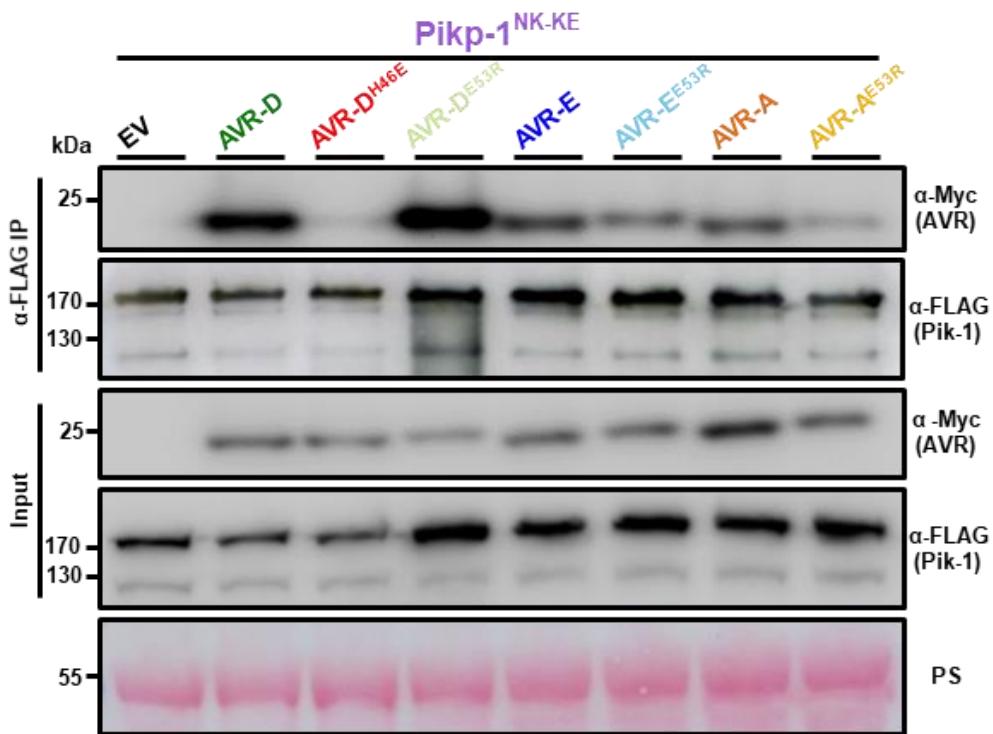
The results from the Y2H and the co-IP experiments confirm that, whilst interface two is still important for the interaction, the engineered interface three compensates for disruptions in interface two by polymorphisms in *AVR-PikE* and *AVR-PikA*.

## Chapter 4



**Figure 4.15. Effector mutations at interface three perturb interactions with Pikp-HMA<sup>NK-KE</sup> in vivo. (a)**

Yeast-Two-Hybrid assay of Pikp-HMA<sup>NK-KE</sup> with AVR-Pik variants and mutants disrupting either interface two (His46Glu) or interface three (Glu53Arg). The control plate for yeast growth is on the left with quadruple dropout media supplemented with X- $\alpha$ -gal and increasing concentrations of Aureobasidin A on the right. Growth and development of blue colouration in the selection plate are both indicative of protein:protein interactions. The unrelated *M. oryzae* effector, AVR-Pii, serves as negative control. Pikp-HMA<sup>NK-KE</sup> was fused to the GAL4 DNA binding domain, and AVR-Pik alleles to the GAL4 activator domain. Each experiment was repeated a minimum of three times, with similar results. **(b)** Western blot analysis confirming accumulation of proteins in yeast for the Y2H assay. Yeast lysate was probed for the expression of HMA domain with anti-GAL4 DNA binding domain (BD) and AVR-Pik effectors anti-GAL4 activation domain (AD). CBS stands for Coomassie Blue staining.



**Figure 4.16. Effector mutations at interface two and three disrupt association to  $\text{Pikp-1}^{\text{NK-KE}}$  in planta.** Co-immunoprecipitation of full length  $\text{Pikp-1}^{\text{NK-KE}}$  with AVR-Pik variants and mutants disrupting either interface two (His46E) or interface three (Glu53Arg). N-terminally 4xMyc tagged AVR-Pik effectors were transiently co-expressed with  $\text{Pikp-1}^{\text{NK-KE}}:6\text{xHis3xFLAG}$  in *N. benthamiana*. Immunoprecipitates obtained with anti-FLAG antiserum, and total protein extracts, were probed with appropriate antisera. Each experiment was repeated at least three times, with similar results.

## 4.8 Reduced binding to the AVR-Pik effectors mutants disrupts cell death responses in planta

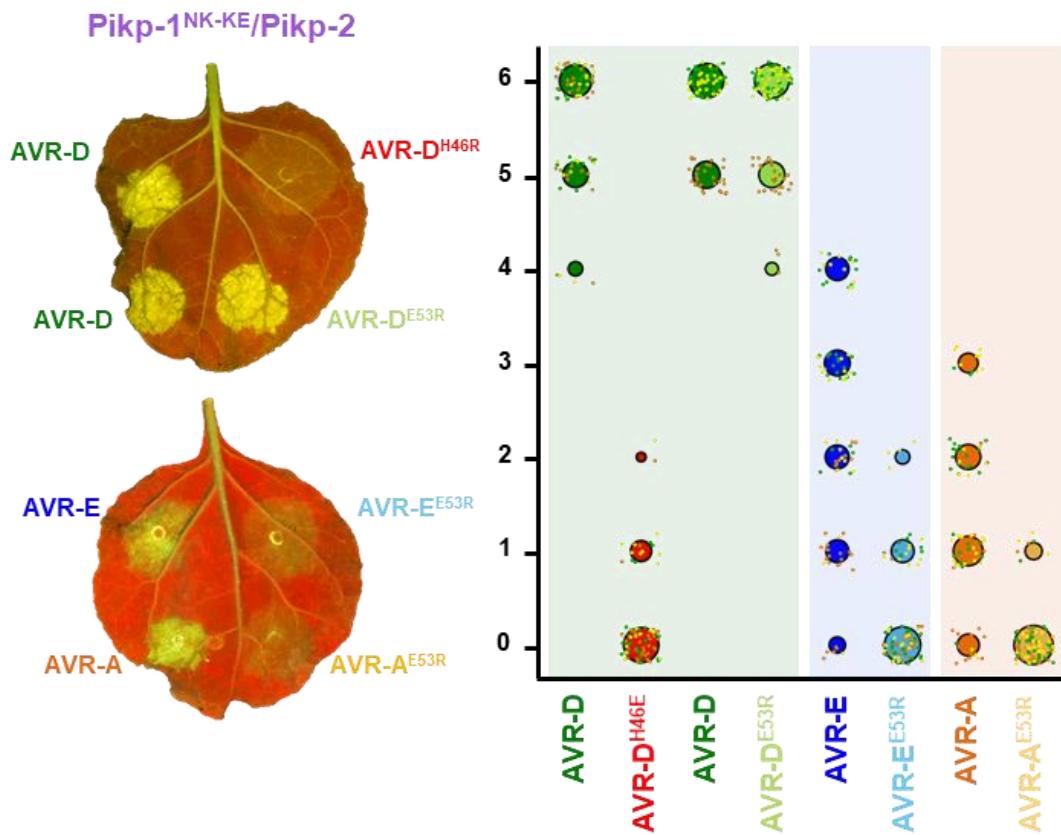
To strengthen the hypothesis that improved binding to the engineered HMA/AVR-Pik interface correlates with an extended cell death phenotype in *N. benthamiana*, I performed cell death assays with the AVR-Pik mutants in the binding interfaces, quantifying the fluorescence under UV light (**Figure 4.17**).

As for  $\text{Pikp}$ , the His46Glu mutation in AVR-PikD abrogates cell death responses when co-expressed with  $\text{Pikp}^{\text{NK-KE}}$  (**Figure 4.17**), correlating with the lack of association to the full-length NLR in planta (**Figure 4.16**). This confirms the importance of interface two for binding and cell death response in the engineered NLRs.

At interface three, the Glu53Arg mutation in AVR-PikD has no effect on the cell death response compared to wild-type (**Figure 4.17**). This correlates with the Y2H and plant co-IP data showing this mutation does not alter the binding to the receptor. However, the equivalent mutation in AVR-PikE and AVR-PikA largely reduced cell death responses

triggered by  $\text{Pikp-1}^{\text{NK-KE}}$  in *N. benthamiana* (Figure 4.17). This is most likely due the reduced binding to the NLR receptor observed by Y2H and co-IP.

The cell death assays further confirm the link between binding of AVR-Pik effectors to the engineered HMA domain and the gain of cell death response in *N. benthamiana*. These results also show that both, interface two and three, are important in effector recognition, with interface three being responsible for extended recognition specificity in the engineered NLR.



**Figure 4.17. Effector mutations at interface two and three abrogate  $\text{Pikp-1}^{\text{NK-KE}}$ -mediated cell death in *N. benthamiana*.** **Left** - Leaf images showing side-by-side cell death in *N. benthamiana* for the wild type AVR-Pik effector alleles and mutants at position 46 and 53 under UV light. **Right** - UV fluorescence scoring represented as dot-plots. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score. The number of repeats was 90.

## 4.9 Discussion

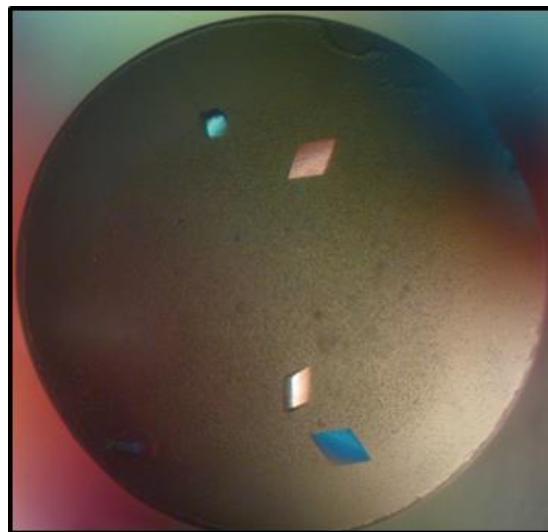
Plant pathogens are a continuous threat to global food security. In the context of a growing world population, tackling plant diseases is a key goal on achieving The Zero Hunger Challenge launched by the United Nations.

Modifying plant immune receptors has a huge potential to achieve genetic resistance to some of the most destructive plant pathogens. To date, engineering plant NLRs to improve resistance has remained largely elusive (Dangl et al., 2013; Monteiro and Nishimura, 2018; Rodriguez-Moreno et al., 2017). Integrated domains from plant NLRs represent an attractive target to engineer pathogen recognition specificities (Ellis, 2016), especially in the cases that can be informed by structural and biochemical studies (De la Concepcion et al., 2018; Guo et al., 2018b; Maqbool et al., 2015; Ortiz et al., 2017; Varden et al., 2019; Zhang et al., 2017b). In this chapter, I used the Pik/AVR-Pik system and the data obtained in chapter three to engineer the binding interface between these two proteins. This resulted in the expansion of the effector recognition profile in Pikp to different AVR-Pik variants, showing that it is possible to use structural information to guide the engineering of NLRs with expanded effector recognition specificities.

Allelic variation in Pik NLRs has given rise to different effector recognition profiles, and subtle differences at the binding interface of the integrated HMA domain and the AVR-Pik effector underpins this phenotype (De la Concepcion et al., 2018). Compared to Pikp, Pik allele Pikm has a broader recognition profile (Kanzaki et al., 2012) that is sustained by a more favourable interaction at the HMA/AVR-Pik interface three described in the previous chapter (De la Concepcion et al., 2018). This interface compensates for the disruption in other binding interfaces caused by polymorphisms in the AVR-Pik effector variants (De la Concepcion et al., 2018).

Through mutations of residues in Pikp, I combined the favourable interfaces from Pikp and Pikm into a single protein. This mutant presented increased binding affinity to AVR-Pik effectors compared with Pikp or Pikm, correlating with the gain of cell death response in planta. The results obtained in this chapter further strengthen the link between immune signalling and tighter binding affinity between effectors and NLR integrated domains established previously (De la Concepcion et al., 2018; Maqbool et al., 2015; Varden et al., 2019). Now I show that this binding affinity, and the subsequent immune responses, can be manipulated by rational design. This can be a general strategy for NLR engineering in cases where the integrated domain directly binds pathogen effectors.

Surprisingly, the combination of two robust interaction interfaces from Pik alleles (Pikp interface 2 and Pikm interface 3) into the engineered Pik, seems to confer the ability to bind the unrecognised AVR-PikC effector, even though Pikp-HMA and Pikm-HMA do not (De la Concepcion et al., 2018; Maqbool et al., 2015). Despite the robust binding in vitro, binding to AVR-PikC in the Y2H assay and plant co-IPs seems rather weak and is not enough to trigger cell death in *N. benthamiana*, questioning its biological relevance. However, the formation Pikp-HMA<sup>NK-KE</sup>/AVR-PikC protein complex was robust enough to allow me to co-purify and crystallise it using the Morpheus® HT-96 crystallisation screen. Crystals with similar morphology to those obtained for Pikp-HMA<sup>NK-KE</sup>/AVR-PikD and Pikp-HMA<sup>NK-KE</sup>/AVR-PikE appeared overnight (**Figure 4.18**) and X-ray diffraction data was collected to 2.15 Å resolution. The structure was subsequently solved and refined by Josephine Maidment; details of X-ray data collection, structure solution, and refinement are given in **Table 4.2**. This structure demonstrated that NLR engineering has the potential to gain binding to previously unbound NLRs, representing a stepping stone to further efforts on engineering NLR recognition to effectors currently not recognized in nature.



Pikp-HMA<sup>NK-KE</sup>/AVR-PikC  
(2.15 Å)

**Figure 4.18. Pik-HMA<sup>NK-KE</sup> and AVR-PikC can be co-crystallized.** Images of crystals under polarised light obtained in the Morpheus® HT-96 screen for Pik-HMA<sup>NK-KE</sup> in complex with AVR-PikC.

Genome-wide analysis has revealed many different types of domains integrated in plant NLRs (Bailey et al., 2018; Kroj et al., 2016; Sarris et al., 2016). Although they are not the most abundant, integrated HMA domains have been more widely studied because they are not only present in the NLR pair Pik (De la Concepcion et al., 2018; Maqbool et al., 2015; Varden et al., 2019), but also in another rice NLR pair named Pia (Cesari et al.,

## Chapter 4

2013; Guo et al., 2018b; Kroj et al., 2016; Ortiz et al., 2017). The integrated HMA domain in the NLR Pia-2 (the sensor of the rice Pia, also known as RGA5), binds two *M. oryzae* effectors, AVR1-CO39 and AVR-Pia. This opens the possibility of applying the approach developed in this chapter to engineer integrated HMAs that bind multiple effectors.

It remains to be seen whether the engineered NLR developed in this chapter confers a broader disease resistance profile of *M. oryzae* strains carrying the different effector variants in rice. Nevertheless, this study represents an important milestone towards achieving a scientific goal that has been long pursued in the field of plant biotechnology.

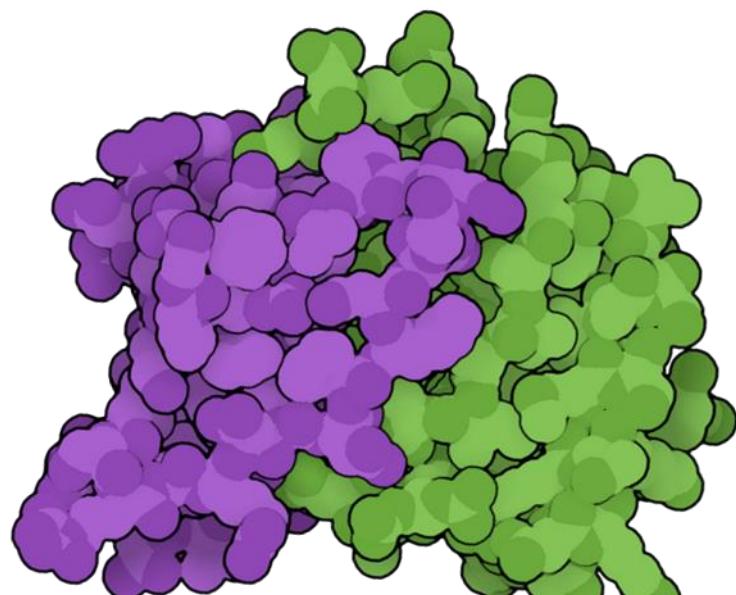
Pikp-HMA <sup>NK-KE</sup> /AVR-PikC	
<b>Data collection statistics</b>	
Wavelength (Å)	0.9763
Space group	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions	
<i>a, b, c</i> (Å)	66.78, 80.21, 105.68
Resolution (Å)*	46.17-2.15 (2.22-2.15)
<i>R</i> <sub>merge</sub> (%)	5.3 (99.6)
<i>l</i> / <i>σl</i>	23.9 (2.3)
Completeness (%)	99.9 (99.8)
Unique reflections	31604 (2694)
Redundancy	13.2 (13.7)
CC <sup>(1/2)</sup> (%)	100.0 (92.5)
<b>Refinement and model statistics</b>	
Resolution (Å)	44.16-2.15 (2.21-2.15)
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub> (%)	22.2/27.1 (36.1/34.1)
No. atoms (Protein)	6959
B-factors (Protein)	64.0
R.m.s deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.479
Ramachandran plot (%)**	
Favoured	97.21
Outliers	0
MolProbity Score	1.87 (89th percentile)

**Table 4.2. X-ray data collection and refinement statistics for Pikp-HMA<sup>NK-KE</sup>/AVR-PikC complex.**  
\*The highest resolution shell is shown in parenthesis. \*\* As calculated by MolProbity.

## Chapter 4

# 5

**A single amino acid  
polymorphism determines  
pairing specificity in the PiK  
receptor pair**



**AVR-PikA/Pikm-HMA**  
**6FUD**



# 5. A single amino acid polymorphism determines pairing specificity in the Pik receptor pair

## Abstract

Pathogen perception inside cells relies on cooperation of immune receptors from the NLR superfamily. Upon recognition, NLR pairs form heterocomplexes to orchestrate immune responses that ultimately lead to cell death and stop pathogen invasion. Natural selection drives specialization of host immune receptors towards an effective activation, whilst keeping a tight regulation of immunity. However, the biochemical basis of adaptation and specialization between immune receptors remains largely unknown. Here, I describe the functional specialization between alleles of the rice NLR pair Pik. Allelic NLRs are better adapted to cooperate with their respective pairs, mounting an effective response against AVR-Pik effectors from the rice blast fungus. Mismatch between members of paired NLR alleles lead to autoimmune phenotypes, a hallmark of diversifying evolution in plant NLRs, also known as hybrid necrosis in plant breeding. A single amino acid polymorphism largely underpins Pik pair co-evolution. These results provide a framework for how immune receptors that act together co-evolve to provide an effective immune response against pathogens. As diversifying NLR evolution can ultimately result in hybrid necrosis, potentially driving reproductive isolation and plant speciation, these results have implications beyond plant immunity.

## Chapter 5

## 5.1 Introduction

Upon activation, NLRs trigger immune signalling by acting as nucleoside-operated switches. Recognition of effectors produces conformational changes that enable exchange of ADP for ATP (Bernoux et al., 2016; Tameling et al., 2002; Wang et al., 2019b; Williams et al., 2011). This triggers NLR oligomerization in supramolecular signalling platforms named inflammasomes or resistosomes, leading to immune responses, including programmed cell death (Hu et al., 2015; Sharif et al., 2019; Tenthorey et al., 2017; Wang et al., 2019a; Zhang et al., 2015a). The assembly of such sophisticated molecular machinery needs to be well coordinated and tightly regulated to ensure an efficient immune response, while avoiding the deleterious effect of constitutive immune activation (Chae et al., 2016; Karasov et al., 2017; Richard and Takken, 2017).

NLRs are one of the most expanded and diversified protein families in plants (Baggs et al., 2017; Cao et al., 2011; Meyers et al., 2003; Van de Weyer et al., 2019). They present signatures of rapid evolution, including long-term balancing selection, high copy number variation between species and are situated at hot-spots for genomic rearrangements (Jacob et al., 2013; Jiao and Schneeberger, 2019; Koenig et al., 2019; Yang et al., 2013). As a consequence, NLRs display not only several nonsynonymous polymorphisms, but also many premature stops and deletions (Clark et al., 2007).

In an additional layer of complexity, plant NLRs can also work in genetically linked pairs (Eitas and Dangl, 2010; Griebel et al., 2014; Narusaka et al., 2009; Sinapidou et al., 2004) or as part of complex immune networks (Castel et al., 2019; Wu et al., 2018; Wu et al., 2017). In such cases, NLRs often optimise their activities towards immune activation and specialize as sensors and helpers (Adachi et al., 2019; Jubic et al., 2019). Paired NLRs are prevalent in plant genomes (Stein et al., 2018; Wang et al., 2019c) and they often have atypical domains integrated into their architecture (Bailey et al., 2018; Kroj et al., 2016; Sarris et al., 2016). These domains are likely derived from pathogen host targets and bind pathogen effectors, acting as sensor domains (Bialas et al., 2018; Cesari, 2018; Cesari et al., 2014a; Wu et al., 2015) which rapidly co-evolve with pathogens (Costanzo and Jia, 2010; Kanzaki et al., 2012).

NLRs with different evolutionary trajectories may drift and become misregulated. When maladapted NLRs occur in the same individual through crossing, it can trigger constitutive activation of immunity, leading to deleterious phenotypes, including dwarfism, necrosis and lethality (Barragan et al., 2019; Bomblies et al., 2007; Chae et al., 2014; Tran et al., 2017; Vaid and Laitinen, 2019). This genetic incompatibility has

## Chapter 5

been long known in agriculture as hybrid necrosis (Bomblies and Weigel, 2007; Hermsen, 1963a, b; Yamamoto et al., 2010). NLR mismatch is one of the bases of this phenomenon and therefore has important implications in agriculture and plant evolution (Bomblies and Weigel, 2007; Ispolatov and Doebeli, 2009).

It can be hypothesised that paired NLRs must balance a trade-off between rapid diversifying evolution to adapt to fast evolving pathogens, whilst maintaining a fine-tuned regulation of complicated protein assemblies that lead to immune responses. However, our understanding of how co-evolution has impacted the activities of paired immune receptors remains limited and the biochemical basis of their adaptive specialization remains largely unknown. To date, we lack a validated framework to explain how plant immune receptors adapt and specialize through evolution, even though this process has important consequences for plant diversification and evolution (Bomblies et al., 2007; Bomblies and Weigel, 2007).

The rice NLR pair Pik represents a good model to study evolutionary processes in plant NLRs. Sensor and helper Pik NLR genes have a head-to-head configuration, creating a coregulated module that probably maintains fine control over immune responses (Ashikawa et al., 2008; Bialas et al., 2018). Co-evolution with pathogen effectors has driven the emergence of Pik NLR alleles with differential recognition specificities to AVR-Pik effector variants (Kanzaki et al., 2012; Li et al., 2019a). To date, five Pik alleles named Pikp, Pikm, Pik, Piks and Pikh have been described according to their ability to recognize different strains of *M. oryzae* harbouring variants of the effector AVR-Pik (Ashikawa et al., 2012; Ashikawa et al., 2008; Kanzaki et al., 2012; Kiyosawa, 1969a, b, 1978). The integrated HMA domain in Pik-1 directly engages with AVR-Pik effectors and underpins recognition specificity (De la Concepcion et al., 2019; De la Concepcion et al., 2018; Maqbool et al., 2015). Consistent with this, polymorphism analysis identified this domain as the most variable in Pik-1 (Costanzo and Jia, 2010). It is still unknown how other domains in the sensor or helper Pik NLRs adapt to rapid changes in this integrated domain (Costanzo and Jia, 2010)

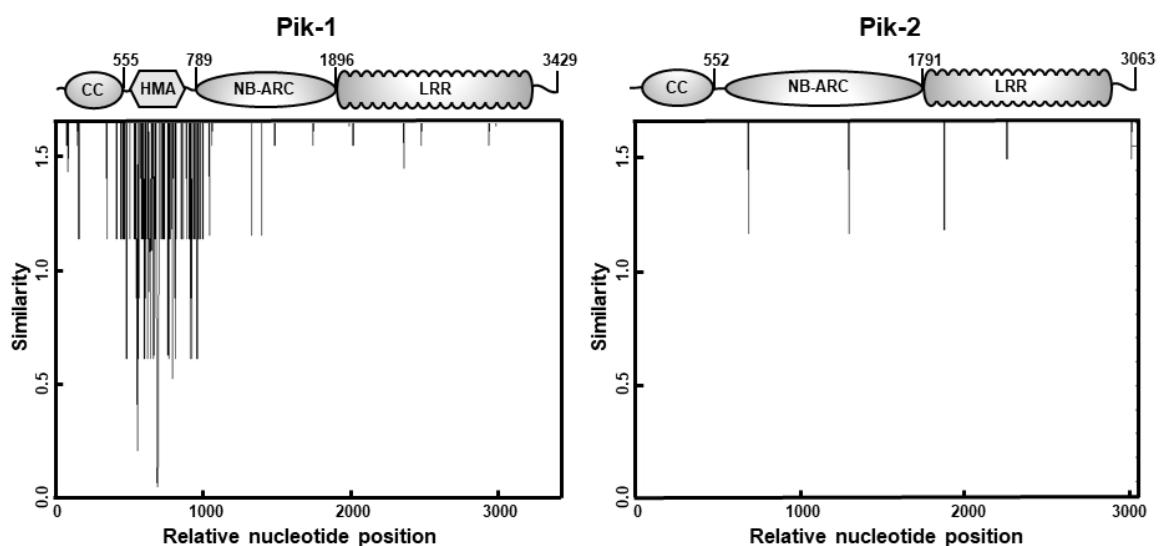
In this chapter, I used the Pik alleles Pikp and Pikm to explore allelic specialization in NLR pairs. I found that allelic Pik sensors and helpers have co-evolved to trigger an efficient immune response to AVR-Pik effectors. Maladapted pairs display constitutive cell death equivalent to autoimmunity or hybrid necrosis phenotypes. Furthermore, I identified a single amino acid polymorphism that underpins both allelic specialization and autoimmunity. This finding allowed me to trace the evolutionary history of this co-adaptation. Altogether, the results presented in this chapter demonstrate that NLR pairs

co-adapt, offering a molecular framework to understand how they evolve to efficiently respond to pathogen effectors while maintaining a tight regulation of immune responses.

## 5.2 Rice Pik alleles forms two phylogenetically distinct groups

To understand the diversity of Pik NLR pairs in rice, I collected sequences from different cultivars spanning indica and japonica that were previously reported to carry functional Pik alleles (Chaipanya et al., 2017; Costanzo and Jia, 2010; Hua et al., 2012; Xu et al., 2008). After alignment using Clustal Omega (Sievers and Higgins, 2014), I performed similarity analysis for both Pik-1 and Pik-2 NLRs (**Figure 5.1**). As reported previously, most of the nucleotide differences fall in the integrated HMA domain of Pik-1 (**Figure 5.1, left**) (Costanzo and Jia, 2010). Helper NLR Pik-2 shows high levels of similarity, consistent with a conserved role in signalling transduction (**Figure 5.1, right**) (Costanzo and Jia, 2010).

Although they are highly conserved, Pik-2 NLRs display three main polymorphic sites (**Figure 5.1, right**). Interestingly, these encode for non-synonymous, subtle amino acid changes Asp/Glu at position 230, Thr/Ser at position 434 and Met/Val at position 627. I analysed the distribution of these polymorphisms in the different cultivars and I observed that, with only two exceptions, Pik-2 polymorphisms tend to cluster in two groups, one containing the sequence Glu230/Ser434/Val627 and the other Asp230/Thr434/Met627. (**Figure 5.2**).



**Figure 5.1. The HMA domain in Pik-1 is the most variable region in the Pik NLR pair.** Similarity plots of the aligned coding sequences of Pik-1 (**left**) and Pik-2 (**right**). The length of the lines represents the dissimilarity in that specific position. The schematic representation of the Pik NLR domains is shown above. Similarity scores were calculated according to Costanzo and Jia, 2010.

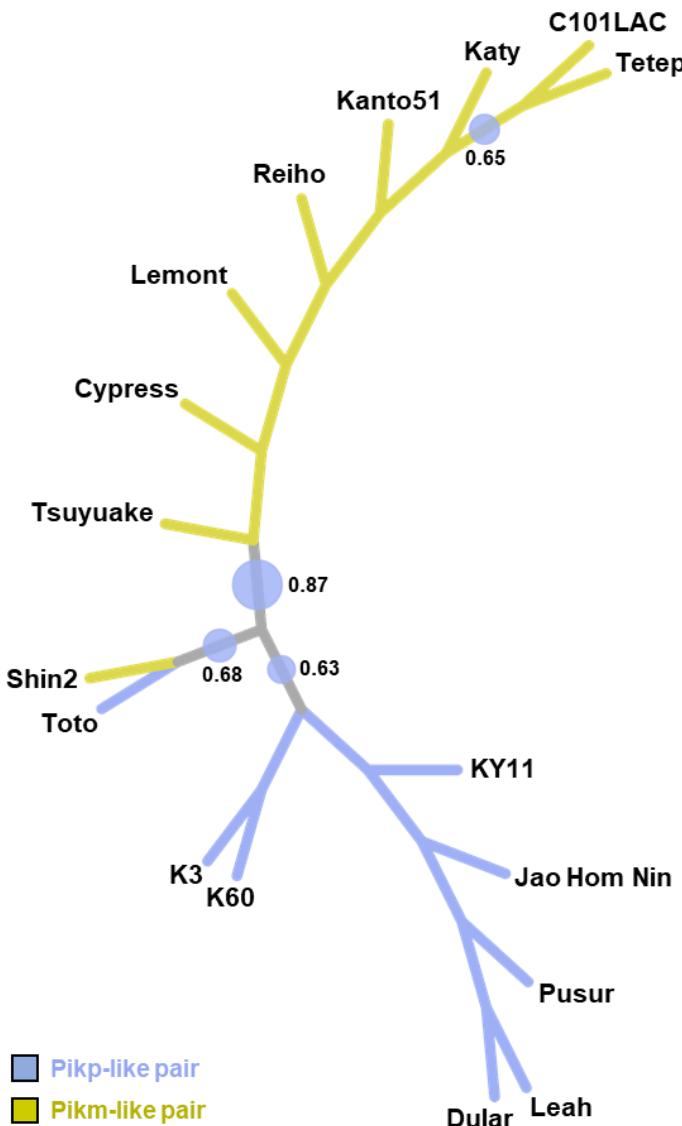
Although the HMA of the Pik-1 NLR is the most polymorphic domain (**Figure 5.1, left**), Pik-1 sequences can be grouped according to the similarity of their HMA domains to the well-studied Pikp-HMA or Pikm-HMA domains. I related the type of HMA domains (Pikp-like or Pikm-like) in the sensor NLR Pik-1 with the polymorphisms observed in the helper Pik-2. This uncovered a trend for cultivars harbouring Pikm-like HMAs to co-exist with Pik-2 NLRs displaying Glu230/Ser434/Val627 polymorphisms, while cultivars with sensor NLRs similar to Pikp-1 mainly carry Asp230/Thr434/Met627 polymorphisms in Pik-2 (**Figure 5.2**). I hypothesised that this could be a signature of co-adaptation between the rapidly changing HMA domain in Pik-1 and the highly conserved helper NLR Pik-2.

<u>Cultivar</u>	<u>Taxon</u>	<u>Origin</u>	<u>HMA-Type</u>	<u>Pik-2 helper polymorphisms</u>			<u>Accession</u>
				230	434	627	
Tsuyuake	Japonica	Japan	M	E	S	V	AB462325
Cypress	Japonica	U.S.	M	E	S	V	GU811861
Lemont	Japonica	U.S.	M	E	S	V	GU811864
Reiho	Indica	Japan	M	E	S	V	GU811865
Kanto51	Japonica	Japan	M	E	S	V	GU811866
Katy	Japonica	U.S.	M	E	S	V	GU811867
Tetep	Indica	Vietnam	M	E	S	V	GU811862
C101LAC	Indica	West Africa	M	E	S	V	HQ606329
Shin2	Japonica	Japan	M	D	T	V	GU811871
Toto	Japonica	China	P	D	T	V	GU811872
K3	Japonica	India	P	D	T	M	HQ662330
K60	Japonica	Japan	P	D	T	M	HM035360
KY11	Indica	China	P	D	T	M	KU365338
Jao Hom Nin	Indica	Thailand	P	D	T	M	HQ660231
Pusur	Indica	Pakistan	P	D	T	M	GU811868
Dular	Indica	Pakistan	P	D	T	M	GU811869
Leah	Indica	U.S.	P	D	T	M	GU811870

**Figure 5.2. Pik-2 polymorphisms are not evenly distributed in the rice cultivars.** Distribution of Pik-2 polymorphisms across the different Pik containing cultivars used in this study. HMAs are divided in two types, P and M, according to their similarity to Pikp-HMA and Pikm-HMA, respectively. Cultivars with a Pikm-like HMA are highlighted in gold, whilst cultivars carrying a Pikp-like HMA are highlighted in ice blue.

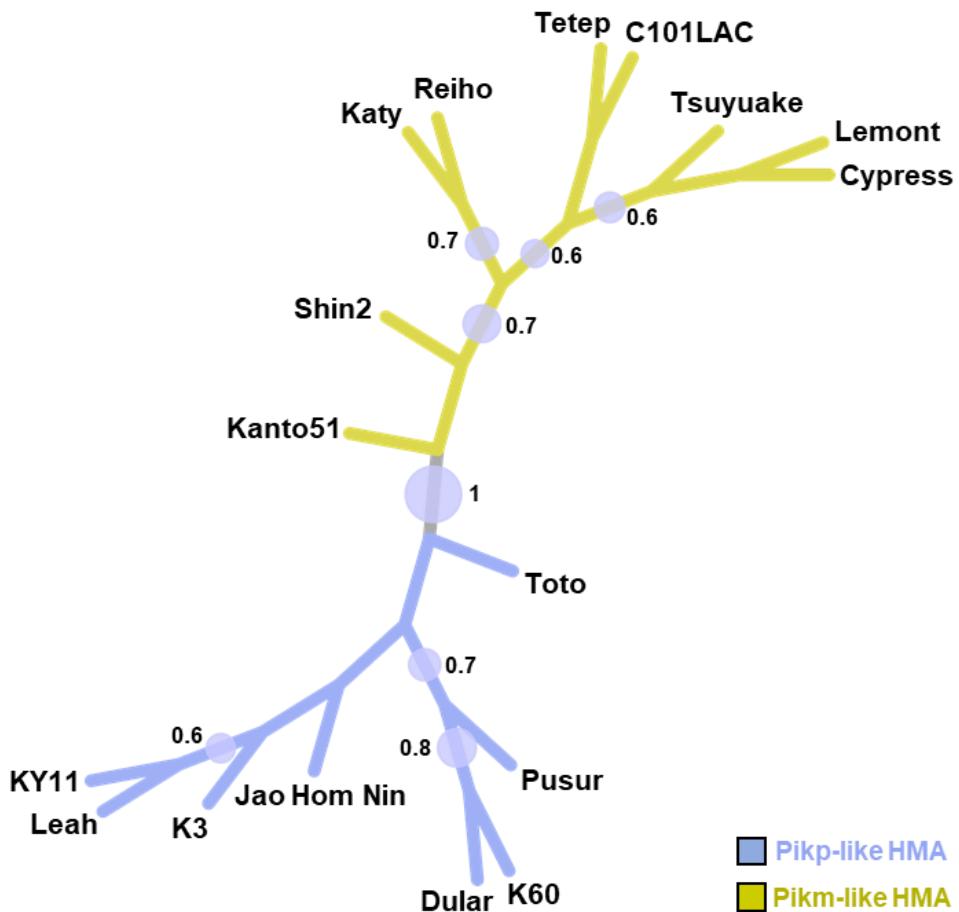
To infer the evolutionary relations of the Pik-2 sequences I constructed a phylogeny, based on the coding sequences, using Mega X (Kumar et al., 2018). Given the high degree of similarity between them, I used the full-length sequences to generate the tree (**Figure 5.3**). Consistent with the polymorphism distribution mentioned above, Pik-2 sequences divide into two distinct phylogenetic groups (bootstrap value = 87). Most of the helper Pik-2 NLRs paired with a sensor Pik-1 similar to Pikm clustered together in a single group, represented by the Pikm cultivar Tsukuaye (**Figure 5.3, highlighted in gold**). By contrast, most of Pik-2 NLRs paired with Pikp-1 like sensor NLRs cluster together in a group represented by the Pikp cultivar K60 (**Figure 5.3, highlighted in ice**

blue). A minor group containing the only two cultivars with mixed Pik-2 polymorphisms can be also identified as a separate clade (bootstrap value = 68) (Figure 5.3).



**Figure 5.3. Pik-2 NLR sequences cluster in two distinct phylogenetic clades.** Maximum likelihood phylogenetic tree of coding sequences of rice Pik-2 NLRs. The tree was calculated using the maximum likelihood method and Tamura-Nei model (Tamura and Nei, 1993) in MEGA X (Kumar et al., 2018). The tree with the highest log likelihood (-4280.68) is shown. Initial trees for the heuristic search were obtained automatically by applying Neighbour-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 200.0000)). This analysis involved 17 nucleotide sequences. There were a total of 3066 positions in the final dataset. The tree was represented using Interactive Tree Of Life (iTOL) v4 (Letunic and Bork, 2019). Cultivar names are placed next to their corresponding branch. Branches corresponding to cultivars whose Pik-1 pair harbours Pikp-like or Pikm-like HMs are highlighted in ice blue or gold, respectively. Bootstrap values higher than 0.60 are indicated next to their corresponding nodes.

For comparison, I also built a phylogeny using the full-length sequences of Pik-1 NLRs (**Figure 5.4**). These sequences clearly split in two very distinct phylogenetic groups (bootstrap value = 100) divided in Pikm-1-like and Pikp-1-like NLRs (**Figure 5.4, highlighted in gold and ice blue, respectively**). As for Pik-2, each clade was formed by members with different geographical origins and belonging to both japonica and indica taxa (**Figure 5.4**).



**Figure 5.4. Pik-1 NLR sequences cluster in two distinct phylogenetic clades.** Maximum Likelihood Phylogenetic tree of coding sequences of rice Pik-1 NLRs. The tree was calculated using the Maximum likelihood method and Tamura-Nei model (Tamura and Nei, 1993) in MEGA X (Kumar et al., 2018). The tree with the highest log likelihood (-5520.10) is shown. Initial trees for the heuristic search were obtained automatically by applying Neighbour-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories +G, parameter = 0.0500)). This analysis involved 17 nucleotide sequences. There were a total of 3441 positions in the final dataset. The tree was represented using Interactive Tree Of Life (iTOL) v4 (Letunic and Bork, 2019). Cultivar names are placed next to their corresponding branch. Branches corresponding to Pik-1 sequences harbouring Pikp-like or Pikm-like HMAs are highlighted in ice blue and gold, respectively. Bootstrap values higher than 0.6 are indicated next to their corresponding nodes.

The phylogenetic analysis revealed that helper NLR Pik-2 sequences follow the distribution of their cognate NLR pairs. Furthermore, despite their high similarity, polymorphic residues in Pik-2 seems to cluster according to the HMA domain contained in their Pik-1 pair. Altogether, these data seed the hypothesis of a co-adaptation between Pik-2 helper NLR and their cognate sensor Pik-1, probably following the diversification of the Pik-1 HMA domain. Therefore, Pik-2 polymorphisms may be adaptive.

### 5.3 Pikm-2 polymorphisms are required for efficient cell death response to AVR-Pik effectors in *N. benthamiana*

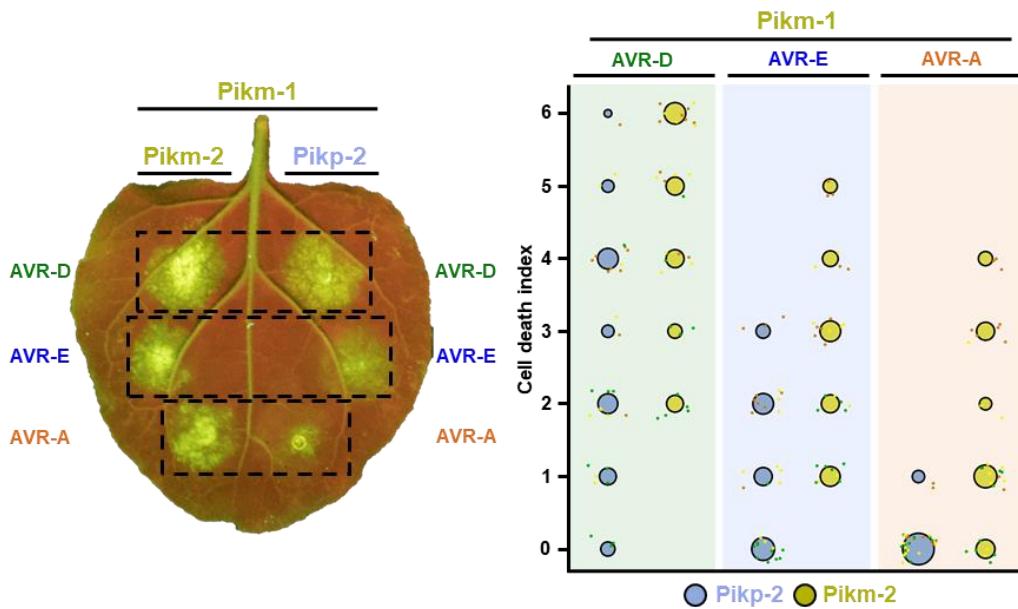
To test whether Pik-2 polymorphisms are adaptive, and to obtain experimental support for co-evolutionary dynamics between paired NLRs, I used our knowledge that two of the most studied Pik alleles, Pikp and Pikm, fall in phylogenetically distinct groups. Pikm originated in the chinese japonica cultivar Hokushi Tami (Kiyosawa, 1978) whereas Pikp originated in the indica cultivar Pusur in Pakistan (Kiyosawa, 1969a). Thus, these alleles have likely been exposed to differential selection pressures, either during domestication of elite cultivars, exposure to different rice blast populations, or a combination or both. These differences have most likely driven Pik specialization for AVR-Pik effector recognition via changes in the HMA domains of the sensor NLR Pik-1 (Costanzo and Jia, 2010; De la Concepcion et al., 2018; Kanzaki et al., 2012).

Pikp and Pikm pathogen recognition specificities can be recapitulated in the model plant *N. benthamiana* (De la Concepcion et al., 2018; Maqbool et al., 2015). This offered an amenable system to experimentally test co-adaptation and specialization in the Pik NLR pair. When co-expressed with Pikm-1, Pikm-2 mediates a cell death response to AVR-PikD, AVR-PikE and AVR-PikA in *N. benthamiana* (De la Concepcion et al., 2018). To test for functional differences between Pikp-2 and Pikm-2, I co-expressed them with Pikm-1 and AVR-PikD, AVR-PikE or AVR-PikA and quantified the cell death outcomes (**Figure 5.5**).

In this assay, the Pikm pair mediated a hierarchical cell death response in the order of AVR-PikD > AVR-PikE > AVR-PikA (**Figure 5.5**). However, the intensity of the cell death outcome was consistently lower when Pikm-1 was co-expressed with Pikp-2 instead of Pikm-2 (**Figure 5.5**). This is particularly noticeable for AVR-PikA, for which the cell death response can only be observed when co-expressed with Pikm-1 and Pikm-2. Protein accumulation of both Pikp-2 and Pikm-2 proteins in planta was very similar (**Figure 5.10**).

This indicates that the Pikm-2 polymorphisms are required for the Pikm extended cell death response to AVR-Pik effectors in *N. benthamiana*, suggesting a possible functional

specialization of the helper NLR Pik-2 towards an effective cell death response to effectors.



**Figure 5.5. Pikm-2 is required for Pikm extended effector recognition in *N. benthamiana*.** Representative leaf depicting Pik mediated cell death as autofluorescence under UV-light. Pikm-1 is co-expressed with either Pikm-2 or Pikp-2 and AVR-Pik effector alleles recognized by Pikm. Side-by-side infiltration are highlighted with dashed boxes. Scoring of cell death triggered by Pikp-2 or Pikm-2 with each AVR-PikD, AVR-PikE and AVR-PikA is represented as dot plots. The number of repeats was 30. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

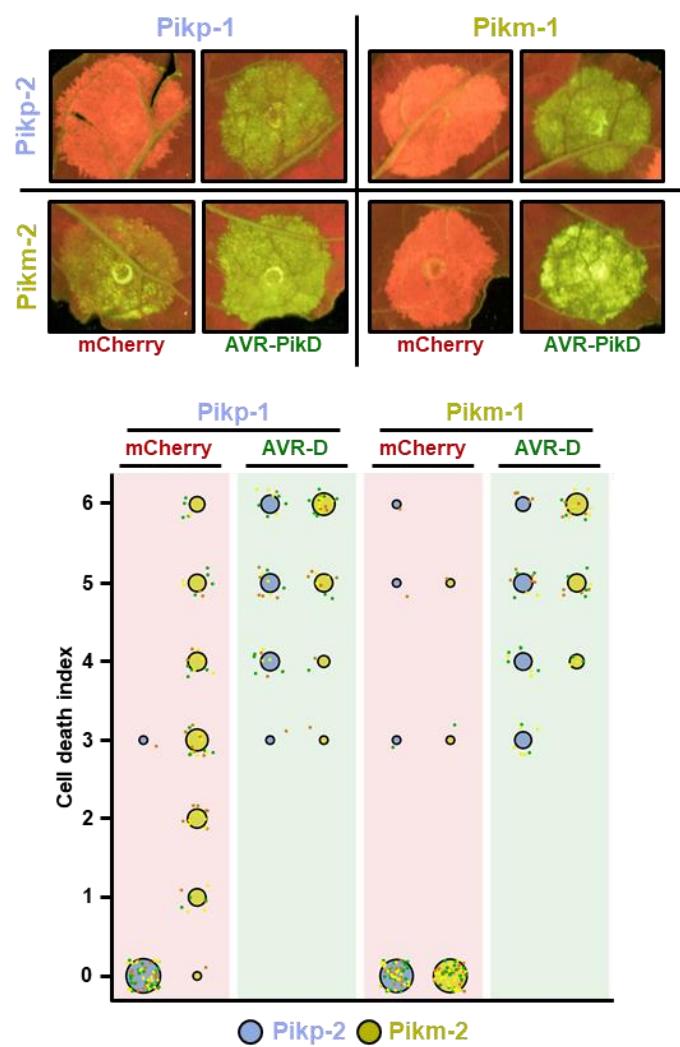
## 5.4 Maladapted Pik pair Pikp-1/Pikm-2 triggers constitutive cell death responses in *N. benthamiana*

When independently evolved NLR receptors meet in the mixed immune background of a hybrid plant, it can lead to misregulation in the form of suppression (Hurni et al., 2014; Stirnweis et al., 2014a), or constitutive activation of immune responses (Chae et al., 2014; Tran et al., 2017). To test for signatures of independent co-evolution between Pikp and Pikm NLR pairs, I performed cell death assays in *N. benthamiana* using different combination of Pik sensor and helpers in the absence and presence of AVR-Pik effectors (Figure 5.6).

As previously reported, Pikp and Pikm allelic pairs triggered a strong cell death when co-expressed with rice blast effector AVR-PikD, but not in the absence of effector (De la Concepcion et al., 2018; Maqbool et al., 2015) (Figure 5.6). However, when Pikp-1 was

co-expressed together with Pikm-2, I observed a clear cell death response in the absence of effector (**Figure 5.6**). Interestingly, I did not observe NLR autoactivation in the reciprocal pair Pikm-1/Pikp-2 (**Figure 5.6**).

These results suggest that Pikp and Pikm pairs display signatures of co-evolution. These pairs may have differentiated enough to trigger a misregulated form of immune response when they are mismatched, leading to constitutive cell death in *N. benthamiana*. Interestingly, only mismatches involving Pikm-2 trigger cell death in the absence of the effector, indicating that this NLR harbours the determinants of this phenotype.



**Figure 5.6. Pikm-2 triggers constitutive cell death in the presence of Pikp-1.** Representative leaf spot images and scoring of Pik mediated cell death as autofluorescence under UV-light in the presence or absence of AVR-PikD. Cell death assay scoring represented as dot plots comparing cell death triggered by Pikp-2 and Pikm-2 when co-expressed with Pikp-1 or Pikm-1. The number of repeats was 60 and 30 for the spots co-infiltrated with mCherry and AVR-PikD, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

## 5.5 Pik autoactivation is linked to immune signalling

To gain knowledge of the constitutive cell death mediated by Pikm-2 and to understand the link with NLR immunity, I mutated Pikm-2 in the well-described NB-ARC P-loop and MHD motifs. I then tested the ability of these mutants to trigger constitutive cell death responses in the absence of the AVR-Pik effector.

The P-loop motif forms one of the most conserved regions in NLR proteins and is required for ADP/ATP binding and NLR activation. Mutations in this position yield inactive NLRs and have been extensively used in NLR research (Dinesh-Kumar et al., 2000; Sukarta et al., 2016; Tameling et al., 2002; Tameling et al., 2006; Tao et al., 2000; Tornero et al., 2002). Previous research carried by Prof. Hiromasa Saitoh (Tokyo Agricultural University) confirmed that Pikp immune signalling triggered by recognition of AVR-PikD requires a functional P-loop to trigger cell death in *N. benthamiana* (personal communication).

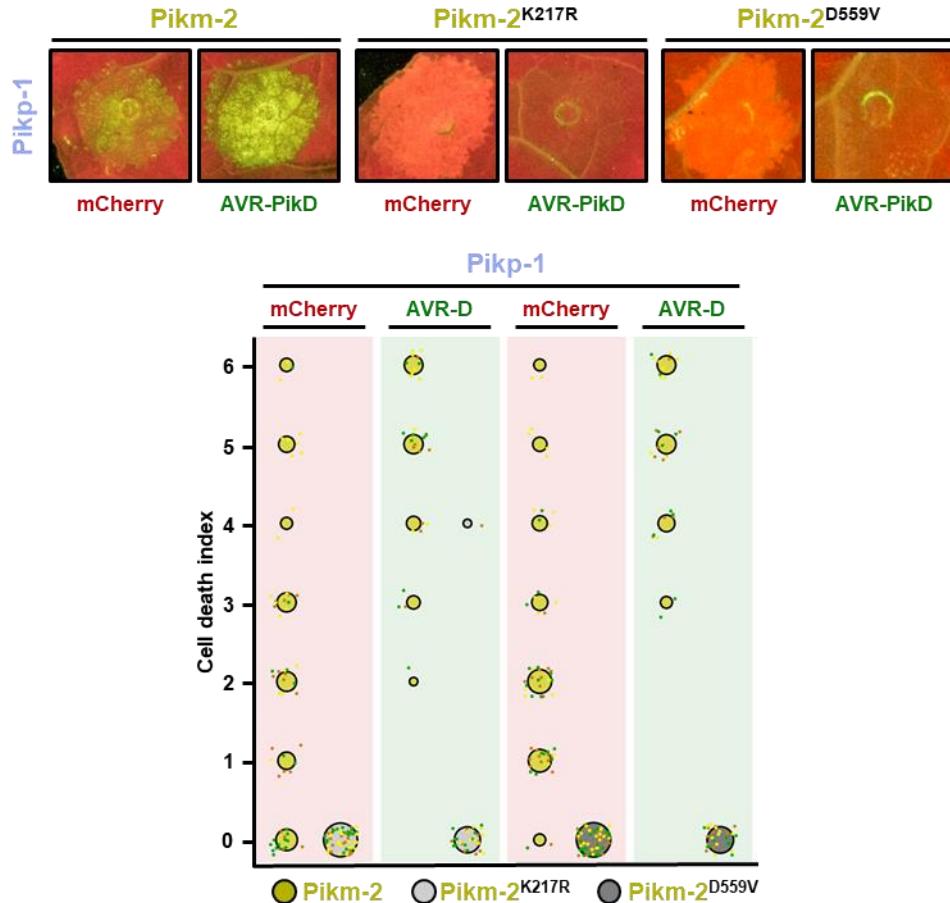
I therefore introduced the loss-of-function mutation Lys217Arg in the Pikm-2 NLR. This mutation greatly reduces ADP/ATP binding in other plant NLRs (Tameling et al., 2002) and, unsurprisingly, abolished Pik-mediated cell death response in the presence of AVR-PikD effector (**Figure 5.7**). This mutation also abrogated the cell death responses in the absence of effector observed in the Pikp-1/Pikm-2 NLR mismatch (**Figure 5.7**), suggesting a link between Pik immune activation and the constitutive cell death mediated by Pikm-2 in the presence of Pikp-1.

NLR activity is also altered by mutations in the MHD motif. Asp to Val mutations in this motif are predicted to modify the local conformation around the ATP-binding pocket. This changes the ATP/ADP binding preference and, contrary to P-loop mutations, renders NLRs constitutively active (Bendahmane et al., 2002; Bernoux et al., 2016; Howles et al., 2005; Tameling et al., 2006; van Ooijen et al., 2008; Williams et al., 2011). Surprisingly, in the case of Pikp-mediated signalling, replacing the MHD residues for alanine in Pikp-2 abolished the cell death response in the presence of AVR-PikD (Prof. Hiromasa Saitoh, personal communication). Although a similar result was observed for the rice NLR Pia-1 (Cesari et al., 2014b), this is different to most of the known NLRs and suggests a different activation mechanism of the Pik NLR pair.

To explore the role of the MHD motif in Pik signalling and constitutive cell death, I introduced the Asp559Val mutation in Pikm-2 and I co-expressed it with Pikp-1 in cell death assays in the presence or absence of AVR-PikD effector. As for the mutants in the P-loop motif, I did not detect cell death mediated by Pikm-2 Asp559Val either in the

presence or the absence of AVR-PikD. This confirmed that Pikm-2 requires an intact MHD motif to trigger cell death in the presence of AVR-PikD. Furthermore, the result strengthens the potential link between constitutive cell death and immune activation (**Figure 5.7**).

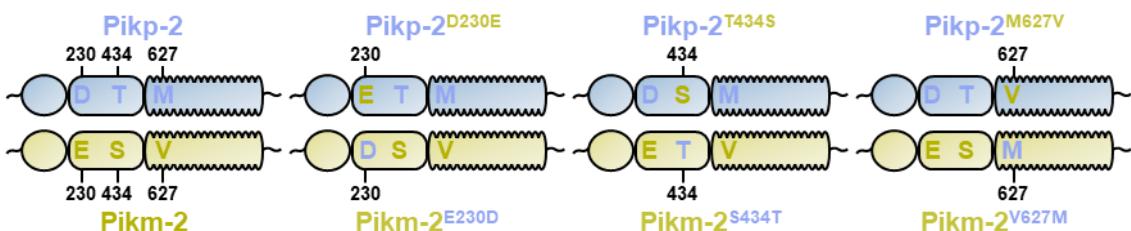
Altogether, these results confirm that cell death phenotypes produced by the mismatch between Pikp-1 and Pikm-2 require intact P-loop and MHD motifs. This established that cell death responses in maladapted Pik pairs are linked to immune signalling.



**Figure 5.7. Pikm-2 autoactivation is dependent on P-loop and MHD motifs.** Representative leaf spot images and scoring of Pikm-2 mediated cell death as autofluorescence under UV-light. Cell death scoring is represented as dot plots comparing cell death triggered by Pikm-2 mutant in P-loop (Lys217Arg) and MHD (Asp559Val) motifs and wild-type Pikm-2. Mutants and wild-type proteins were co-expressed with Pikp-1 and mCherry (red panel) or AVR-PikD (green panel). The number of repeats was 60 and 30 for the spots co-infiltrated with mCherry and AVR-PikD, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

## 5.6 A single amino acid polymorphism in Pik-2 has a major role in cell death responses to AVR-Pik effectors

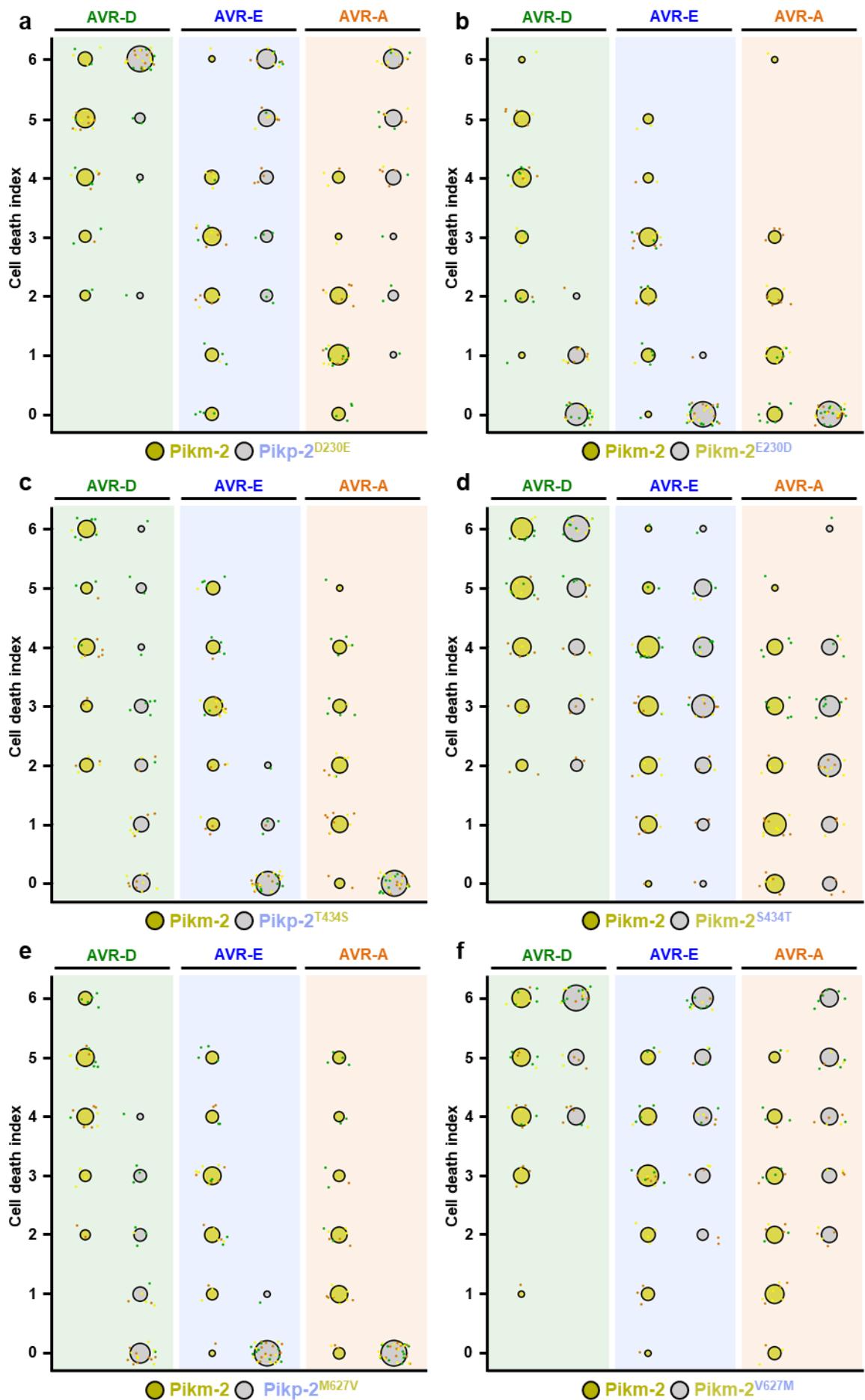
I aimed to dissect the basis of the differential cell death phenotypes displayed by Pikp-2 and Pikm-2 in response to AVR-Pik effectors. To this end, I took advantage of the high similarity between Pikp-2 and Pikm-2 (Figure 5.1), with only three amino acid differences between them (Figure 5.8). Using site-directed mutagenesis, I exchanged Pikp-2 and Pikm-2 polymorphisms at each position (Figure 5.8) and I tested the mutants in cell death assays.



**Figure 5.8. Schematic representations of Pik-2 mutation at polymorphic sites.** Distribution of polymorphisms in Pik-2 allelic NLRs and mutants. Polymorphic sites are numbered and residues are coloured according to whether they are present in Pikp-2 (ice blue) or Pikm-2 (Gold).

I co-expressed Pikm-1 and each Pik-2 mutant with either AVR-PikD, AVR-PikE or AVR-PikA, using infiltration of Pikm-1 and Pikm-2 as control. For each assay I scored the cell death responses and compared the differences with the Pikm control to uncover the contribution of each polymorphism to cell death mediated by AVR-Pik effector recognition (Figure 5.9).

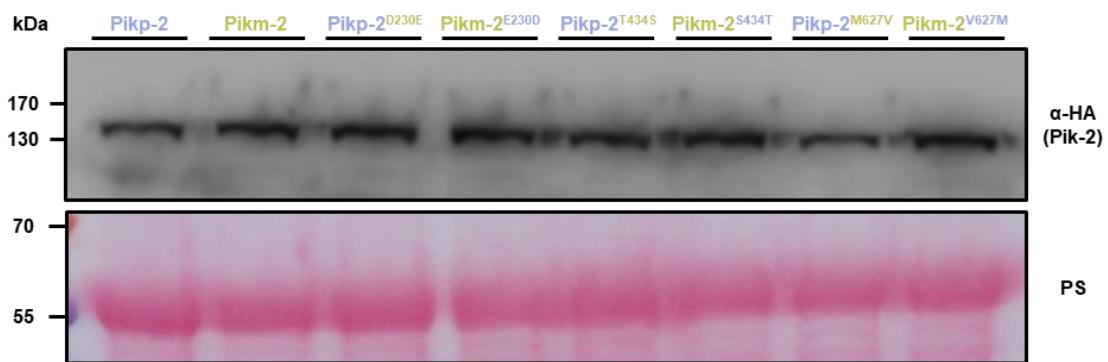
Pikp-2 has an Asp residue at position 230, whilst Pikm-2 has a Glu. Although the properties of these residues are highly similar and only differ in an extra carbon in the Glu lateral chain, the Asp230Glu mutation had a major effect in cell death response to AVR-Pik effectors (Figure 5.9a). When co-expressed with Pikm-1, Pikp-2 Asp230Glu mediated a remarkable increase in cell death response to each AVR-Pik effector. These responses were not only higher than those measured for Pikm-1 with Pikp-2 (Figure 5.5) but were also higher compared to those triggered by Pikm-1 and Pikm-2 (Figure 5.9a). By contrast, Glu230Asp in Pikm-2 reduced the cell death response to each AVR-Pik effector, displaying only a slight response to AVR-PikD (Figure 5.9b). This points to a major involvement of the Pikm-2 Glu230 residue in the Pikm extended responses to AVR-Pik effectors.



**Figure 5.9. A single Pik-2 polymorphism has a major impact in cell death response to AVR-Pik effectors.** Cell death scoring is represented as dot plots comparing cell death triggered by Pik-2 mutants **(a)** Pikp-2 Asp230Glu, **(b)** Pikm-2 Glu230Asp, **(c)** Pikp-2 Thr434Ser, **(d)** Pikm-2 Ser434Thr, **(e)** Pikp-2 Met627Val and **(f)** Pikm-2 Val627Met. The Pik-2 mutants were co-expressed with Pikm-1 and AVR-PikD, AVR-PikE or AVR-PikA. Pikm NLR pair was co-infiltrated for side-by-side comparison. The number of repeats was 30. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

Swapping Pikp-2 and Pikm-2 residues at position 434 (Thr and Ser, respectively), had no major effect on the cell death response. Pikp-2 Thr434Ser had a lower response to AVR-Pik effectors compared to Pikm-2 (**Figure 5.9c**), whilst Pikm-2 Ser434Thr had responses that were similar to the wild-type (**Figure 5.9d**). Mutations at position 627 had a similar effect, with Pikp-2 Met627Val displaying a reduced cell death response to AVR-Pik effectors in comparison to wild-type Pikm-2 (**Figure 5.9e**). Interestingly, the reciprocal mutation Val627Met in Pikm-2 consistently increased cell death responses, particularly to AVR-PikE and AVR-PikA (**Figure 5.9f**). This implies a negative contribution of Pikm-2 polymorphism Val627 towards cell death responses. All the Pik-2 mutants had a similar level of protein accumulation in *N. benthamiana* compared to the wild-type Pikp-2 and Pikm-2, confirming that the cell death phenotypes observed are not due to differences in protein expression (**Figure 5.10**).

Altogether, these results uncovered a single polymorphism, Glu230, as a major determinant of the increased cell death responses to AVR-Pik effectors displayed by the Pikm NLR pair. Interestingly, the other polymorphisms Ser434 and Val627 seem to have a negative contribution to cell death, probably due to a regulative role. In agreement with this, cell death response of Pikp-2 mutants with these residues is lower compared to the measured for the wild-type (**Figure 5.5**, **Figure 5.9c** and **Figure 5.9e**).



**Figure 5.10. Pik-2 alleles and mutants have similar levels of protein accumulation in planta.** Western blots showing accumulation of point mutant proteins in the background of Pikp-2 and Pikm-2. C-terminally 6xHA tagged Pik-2 mutants were transiently expressed *N. benthamiana*. C-terminally 6xHA tagged wild-type Pikp-2 and Pikm-2 are included as controls. Total protein extracts were probed  $\alpha$ -HA antisera. Total protein loading is shown by Ponceau staining (PS).

## 5.7 NLR specialization and autoimmunity are linked to the same polymorphism

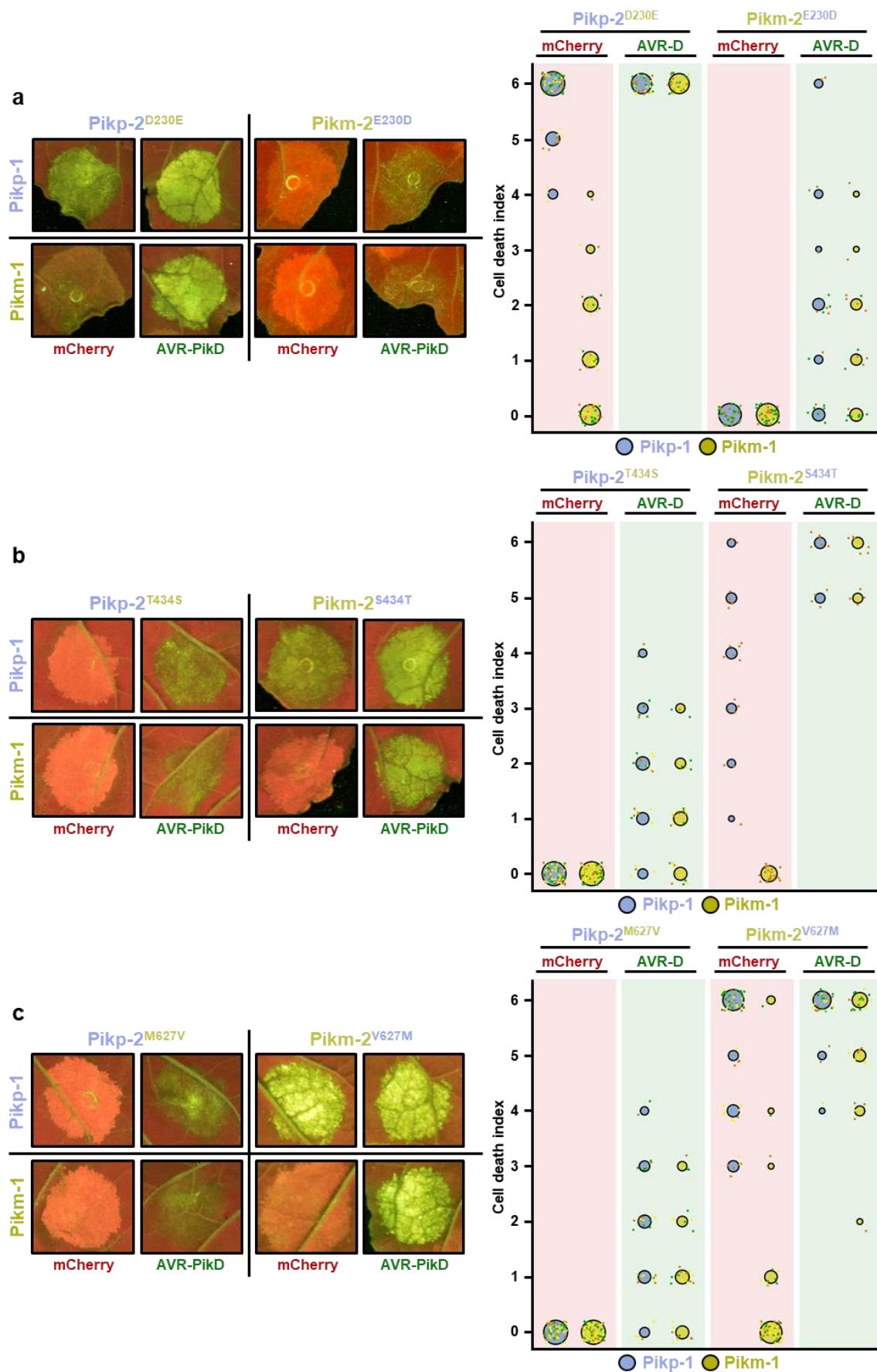
To understand the basis of Pikm-2-mediated autoimmunity, I used the point mutants in Pik-2 polymorphic positions to narrow down the determinant of constitutive cell death. For this, I co-expressed each mutant described above (Figure 5.8) with either Pikp-1 or Pikm-1 in the presence or absence of AVR-PikD effector.

The Asp to Glu mutation at position 230 in Pikp-2 conferred a strong cell death response in the absence of the effector when co-expressed with Pikp-1 (Figure 5.11a), a small constitutive cell death could also be observed with Pikm-1 (Figure 5.11a). By contrast, the reciprocal mutation at the Pikm-2 equivalent position abrogated constitutive cell death in the presence of Pikp-1 and significantly reduced the cell death response mediated by AVR-PikD recognition (Figure 5.11a).

Introducing the mutations Thr434Ser or Met627Val in Pikp-2 did not confer constitutive cell death when co-expressed with Pikp-1 (Figure 5.11b, c). Consequently, the reciprocal mutations in Pikm-2 did not abrogate the cell death triggered in the absence of effectors, nor reduced the cell death response to AVR-PikD (Figure 5.11b, c).

Overall, these results narrowed down the determinant of autoimmunity in Pik maladapted pairs to a single amino acid polymorphism. Interestingly, the same polymorphism seems related to the stronger cell death responses to AVR-Pik effectors mediated by Pikm. This establish a link between immune specialization and gain of autoimmune responses in NLR pairs, two hallmarks of co-evolution.

## Chapter 5



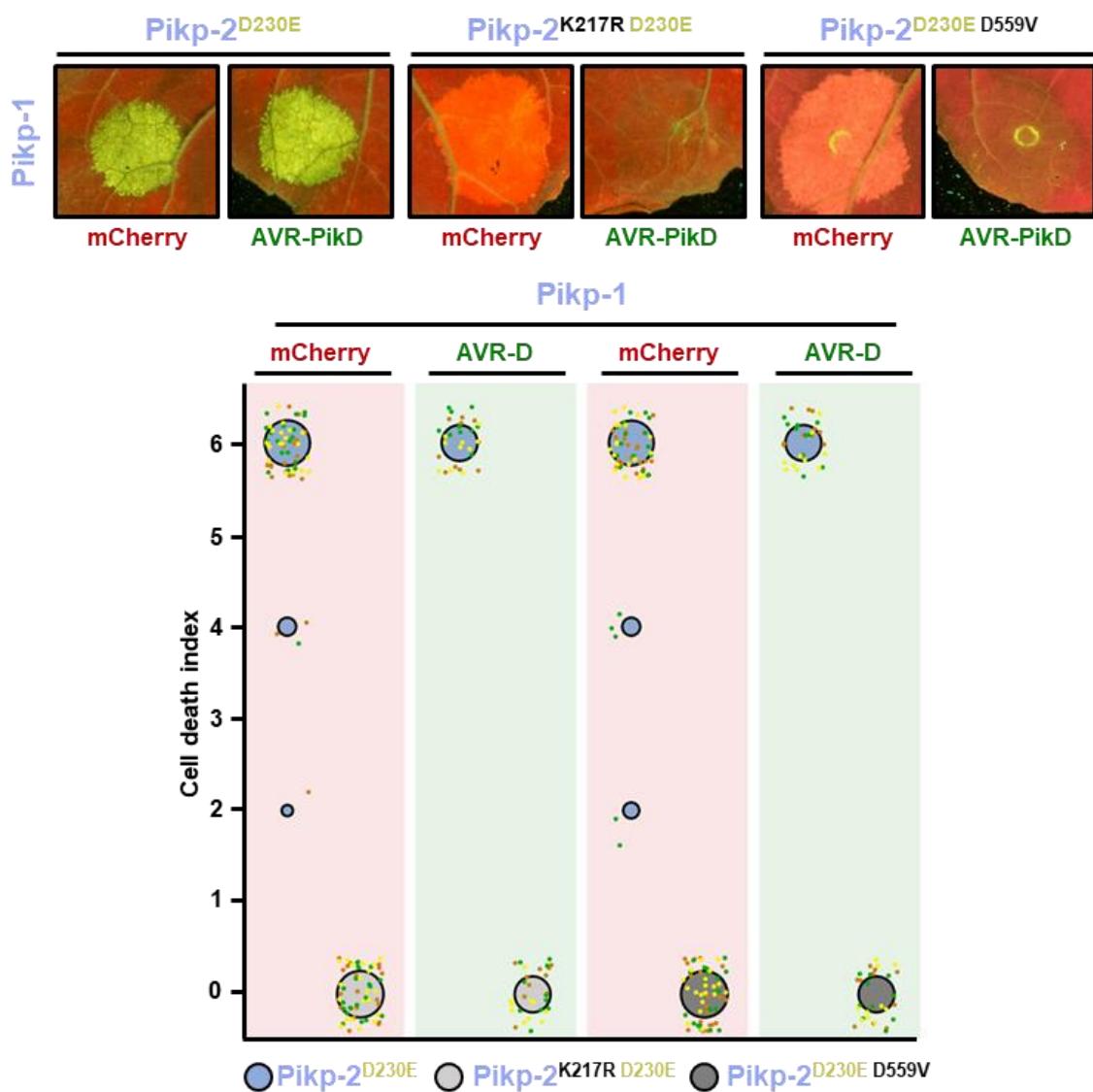
**Figure 5.11. The pik-2 polymorphism at position 230 is the determinant for constitutive cell death.** Representative leaf spot images and scoring of cell death mediated by Pik-2 as autofluorescence under UV-light. Cell death scoring is represented as dot plots comparing cell death triggered by Pik-2 mutants at polymorphic positions **(a)** 230, **(b)** 434 and **(c)** 627. Pik-2 mutants were co-expressed with Pikp-1 and mCherry (red panel) or AVR-PikD (green panel). The number of repeats was 60 and 30 for the spots co-infiltrated with mCherry and AVR-PikD, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score. Assays corresponding to Pikm-2 Glu230Asp and Pikm-2 Ser434Thr had only two and one biological replicates, respectively.

## 5.8 Constitutive cell death responses in autoactive mutants are dependent on the P-loop and MHD motifs

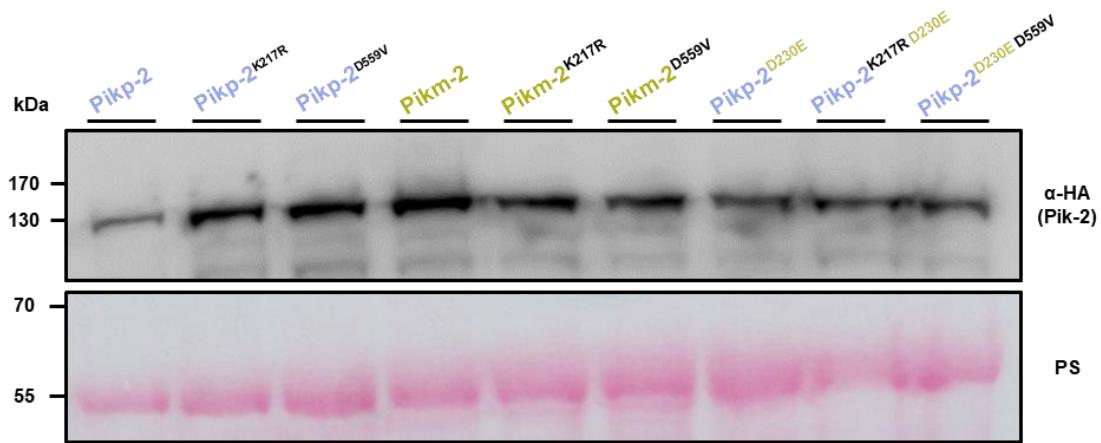
Given that the Asp230Glu mutation in Pikp-2 increases Pikm-1 mediated cell death response to AVR-Pik effectors, and displays a much higher constitutive cell death response in the presence of Pikp-1, I sort to investigate whether the activation mechanism in this mutant is different from its wild type counterpart. To do this, I mutated the P-loop and MHD motifs in the Pikp-2 Asp230Glu background and I tested their ability to trigger constitutive cell death responses.

As for Pikm-2, both a Lys217Arg mutation in the P-loop and an Asp559Val mutation in the MHD motifs abolished constitutive cell death responses in the absence of effectors (**Figure 5.12**), as well as the cell death response triggered by AVR-PikD recognition (**Figure 5.12**). None of these mutations alter the protein accumulation of the Pik-2 alleles or mutants (**Figure 5.13**).

These results confirm that Pik-2 mediated cell death requires an intact P-loop and MHD motifs. This further strengthens the link between constitutive cell death and Pik immune signalling.



**Figure 5.12. *Pikp-2* Asp230Glu autoactivation is dependent on P-loop and MHD motifs.** Representative leaf spot images and scoring of *Pikp-2* mediated cell death as autofluorescence under UV-light. Cell death scoring is represented as dot plots comparing cell death triggered by *Pikp-2* Asp230Glu mutant and its versions mutated in P-loop (Lys217Arg) and MHD (Asp559Val) motifs. *Pikp-2* mutants were co-expressed with *Pikp-1* and mCherry (red panel) or AVR-PikD (green panel). The number of repeats was 60 and 30 for the spots co-infiltrated with mCherry and AVR-PikD, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.



**Figure 5.13. Mutations in Pik-2 P-loop and MHD motifs do not affect protein accumulation.** Western blots showing accumulation of P-loop and MHD mutants in the background of Pikp-2, Pikm-2 and Pikp-2 Asp230Glu. C-terminally 6xHA tagged Pik-2 mutants were transiently expressed *N. benthamiana*. C-terminally 6xHA tagged Pikp-2, Pikm-2 and Pikp-2 Asp230Glu are included as controls in each case. Total protein extracts, were probed  $\alpha$ -HA antisera. Total protein loading is shown by Ponceau staining (PS).

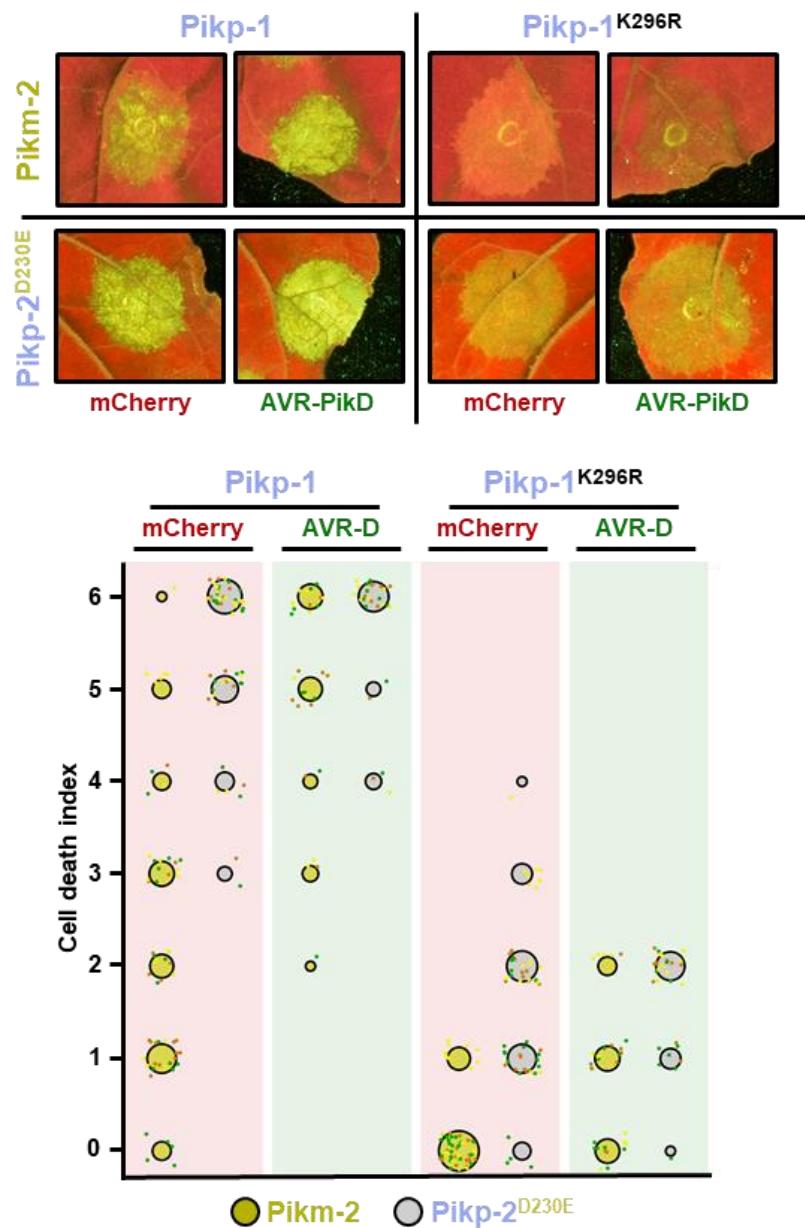
## 5.9 The P-loop motif in the sensor NLR Pik-1 is important, but not essential, for cell death responses in *N. benthamiana*

Constitutively active NLRs represent a useful tool for the study of NLR immune signalling. To get a better understanding of Pik cell death signalling and the requirements for NLR activation, I used the autoimmune combination with Pikm-2 and Pikp-2 Asp230Glu to assess the involvement of the Pik-1 P-loop motif in immune signalling.

Consistent with a cooperation/transactivation mechanism where both sensor and helper P-loop motifs are required for cell death signalling (exemplified by *Arabidopsis* DM1/DM2 (Tran et al., 2017)), previous experiments by Prof. Hiromasa Saitoh (Tokyo Agricultural University) found that mutations in the P-loop motif of the sensor NLR Pikp-1 abrogate the cell death response to AVR-Pik effectors (personal communication). I aimed to clarify whether this phenotype is indeed due to a lack of signalling or is just a reduction in effector recognition. Therefore, I co-expressed a Pikp-1 mutant with a Lys296Arg mutation in the P-loop motif (generated by Josephine Maidment at the Banfield lab) with Pikm-2 and Pikp-2 Asp230Glu and I compared the cell death activity with wild-type Pikp-1 (**Figure 5.14**). No studies were done for the MHD motif as this is already degenerated in Pik-1 (the wild-type harbours IHP instead of the canonical MHD residues).

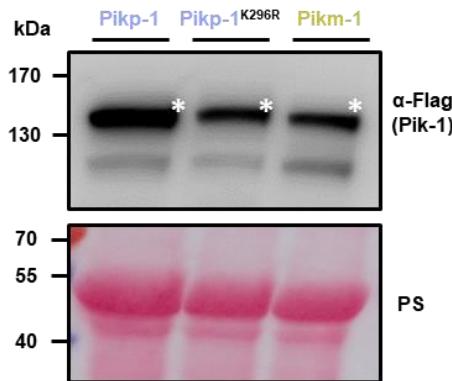
In agreement with Prof. Saitoh's data, Pikp-1 Lys296Arg almost abolishes constitutive cell death signalling triggered by NLR mismatch with Pikm-2 (**Figure 5.14**). However, while largely reduced, there was still a visible cell death response to AVR-PikD (**Figure**

5.14). This response was more accentuated when co-expressed with highly autoactive mutant *Pikp-2* Asp230Glu (Figure 5.14). Protein accumulation of *Pikp-1* Lys296Arg was similar to wild-type versions of allelic *Pikp-1* and *Pikm-1* (Figure 5.15).



**Figure 5.14. The *Pik-1* P-loop motif is important but not essential for *Pik* mediated cell death.** Representative leaf spot images and scoring of *Pik* mediated cell death as autofluorescence under UV-light. Cell death scoring is represented as dot plots comparing cell death triggered by *Pikm-2* or *Pikp-2* Asp230Glu in the presence of wild-type *Pikp-1* or a version mutated in the P-loop motif (Lys296Arg) and MHD (Asp559Val). The different NLR pair combinations were co-infiltrated with *mCherry* (red panel) or *AVR-PikD* (green panel). The number of repeats was 60 and 30 for the spots co-infiltrated with *mCherry* and *AVR-PikD*, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

These unexpected data suggest that the sensor Pik-1 P-loop motif is important for cell death activation, but not completely required. This uncovered an unequal contribution of Pik-1 towards cell death signalling compared to Pik-2, where mutations in P-loop motif completely abolish cell death (**Figure 5.7**, **Figure 5.12**). Therefore, the Pik activation mechanism might differ slightly from the cooperation model of NLR activation previously proposed (Bialas et al., 2018).



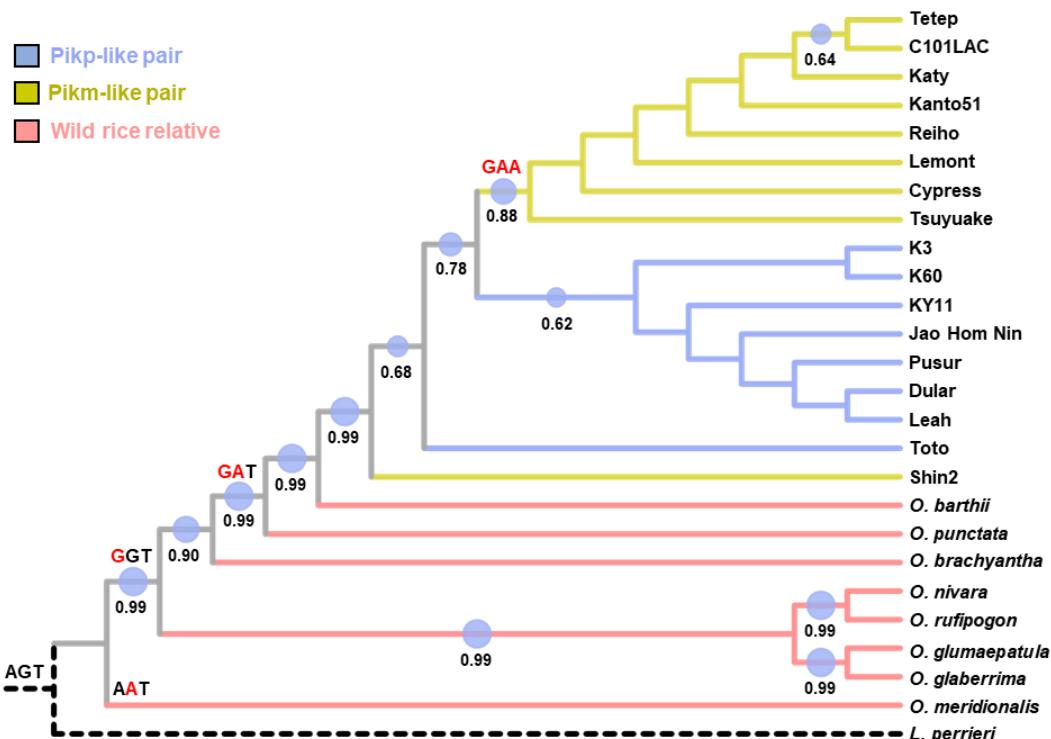
**Figure 5.15. A mutation in the Pik-1 P-loop motif does not affect protein accumulation.** Western blot depicting Pik-1 protein accumulation. C-terminally 6xHis3xFLAG tagged Pikp-1 Lys296Arg mutant was transiently expressed *N. benthamiana*. C-terminally 6xHis3xFLAG tagged wild-type Pikp-1 and Pikm-1 are included as controls (left and right, respectively). Total protein extracts, were probed with α-FLAG antisera. Asterisks mark the band corresponding to Pik-1. Total protein loading is shown by Ponceau staining (PS).

## 5.10 The Glu230 polymorphism has evolved in modern rice

Having narrowed the determinant of Pik NLR pair specialization and compatibility to a single amino acid polymorphism, I set out to investigate the history of this polymorphism to gain an evolutionary perspective of the specialization process in plant NLR pairs. For this, I first combined the Pik-2 coding sequences from rice cultivars described above (**Figure 5.3**) with the Pik-2 orthologs from wild Asian and African relative species reported by Stein et al. 2018. Then I calculated a phylogenetic tree using MEGA X (Kumar et al., 2018) and I rooted it with the African outgroup species *Leersia perrieri* (**Figure 5.16**).

As expected, Pik-2 sequences from wild rice species were phylogenetically distinct from those belonging to modern rice and formed a more basal clade (**Figure 5.16, highlighted in pink**). As shown before, Pik-2 sequences from rice cultivars form two phylogenetically distinct clades grouping Pik-2 sequences paired with sensor NLRs similar to Pikp-1 or Pikm-1 (**Figure 5.16, highlighted in ice blue and gold, respectively**). I noted that the phylogenetic tree did not completely follow the proposed evolutionary history of the rice species (Stein et al., 2018). This is most probably due to

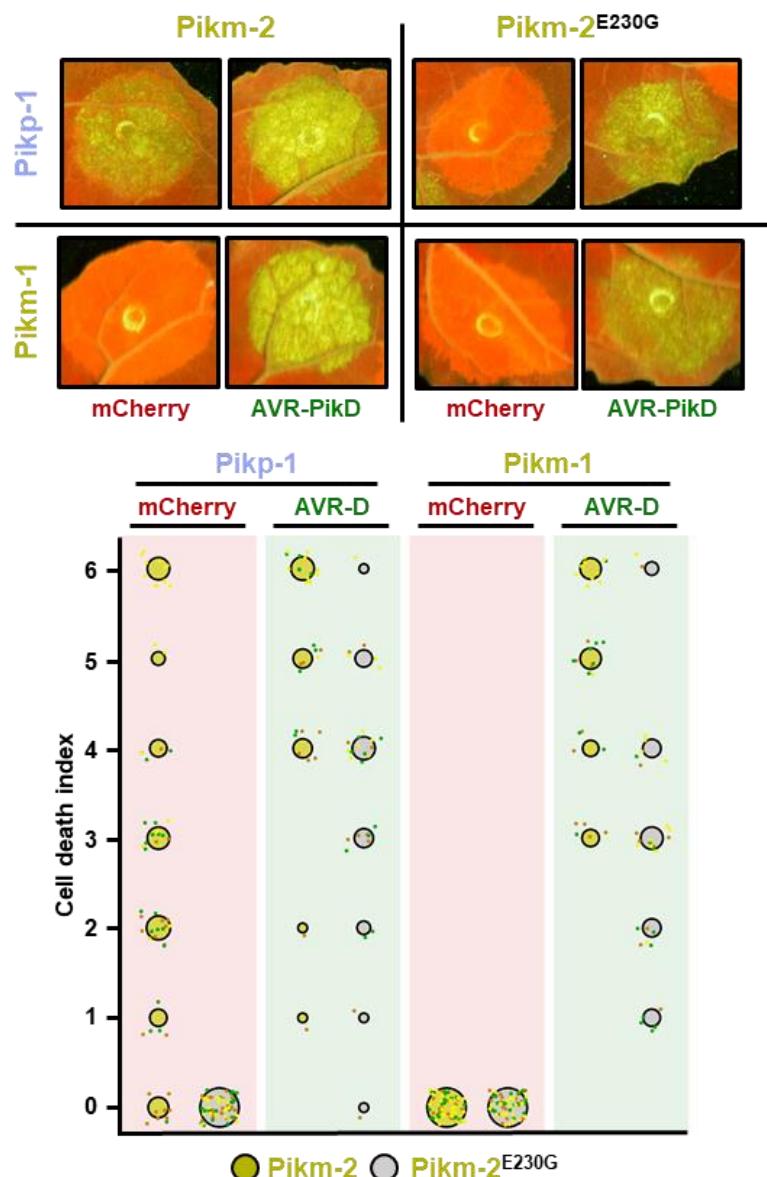
the very complex evolution of the *Oryza* genus (Huang et al., 2012; Wing et al., 2018), which includes different domestication events (Civáň et al., 2015; Molina et al., 2011; Stein et al., 2018). Additionally, *Oryza* show extensive gene flow between species (Stein et al., 2018). Also, multiple genes have been introgressed from Japonica to other species (Choi et al., 2017) and modern rice cultivars have strong signatures of artificial selection and domestication, which can particularly affect resistance genes (Xie et al., 2015).



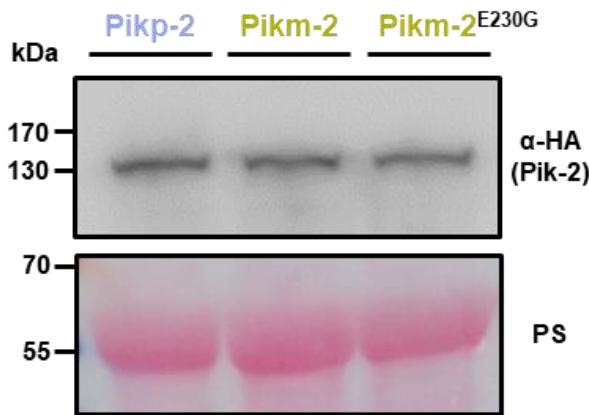
**Figure 5.16. Evolutionary history of Pik-2 polymorphism at position 230.** Maximum likelihood phylogenetic tree of Pik-2 coding sequences from cultivated rice and wild African and Asian relative species. The tree was calculated using the Maximum likelihood method and Tamura-Nei model (Tamura and Nei, 1993) in MEGA X (Kumar et al., 2018). The tree with the highest log likelihood (-12448.30) is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbour-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood values. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories +G, parameter = 0.8217)). This analysis involved 26 nucleotide sequences. There were a total of 3281 positions in the final dataset. The tree was represented using Interactive Tree Of Life (iTOL) v4 (Letunic and Bork, 2019). Asian rice cultivar and wild relative species names are placed next to their corresponding branch. Branches corresponding to Pik-1 sequences harbouring Pikp-like or Pikm-like HMs are highlighted in ice blue and gold, respectively. Wild rice relative species are highlighted in pink and the African outlier species *Leersia perrieri* is depicted by a dash line. Bootstrap values higher than 0.6 are indicated below to their corresponding nodes. Ancestral states were inferred in MEGA X (Kumar et al., 2018) using the Maximum Likelihood method (Nei and Kumar, 2000) and Tamura-Nei model (Tamura and Nei, 1993). The tree shows a set of possible nucleotides (states) at the Pik-2 polymorphic position 230 based on their inferred likelihood at site 742 to 744 of the sequence alignment. Non-synonymous changes at the codon are depicted in red above their corresponding node.

To reconstruct the evolutionary events during Pik specialization, I used MEGA X (Kumar et al., 2018) to infer the ancestral state of the nucleotide sequences codifying for the Pik-2 polymorphism at position 230. Interestingly, the *L. perrieri* outgroup species harbours a Ser residue coded by AGT at the equivalent position to Pik-2 230 (**Figure 5.16**). A first major change at this position seems to be a non-synonymous Ser to Gly mutation, produced by a transition from AGT to GGT in the formation of the wild oryza species clade (**Figure 5.16**). This change allowed another mutation event produced by a transition from GGT to GAT. This mutation seemed to happen in the wild rice clade and codes for an Asp residue in position 230, which is maintained in Pik-2 NLRs from modern rice varieties clustering in the Pikp-like clade (**Figure 5.16**). The last step of the evolution at Pik-2 position 230 seemed to occur in the rise of the Pikm-like clade. Here, a transversion GAT to GAA produced the Asp230Glu mutation that determines specialization in the Pikm NLR pair and makes Pikm-2 incompatible with Pikp-1 (**Figure 5.16**).

To experimentally validate the evolutionary history inferred by the Pik-2 phylogeny, I introduced the Glu230Gly mutation in Pikm-2 using site directed mutagenesis and tested its ability to trigger cell death in *N. benthamiana*. I hypothesised that reverting the Pikm determinant of specialization to a basal state would restore the compatibility with Pikp NLR alleles and, possibly, reduce the cell death response triggered by recognition of AVR-Pik effectors. Indeed, when co-expressed with Pikp-1, the Glu230Gly mutation abolished the constitutive cell death triggered by Pikm-2 in the absence of the effector (**Figure 5.17**). This mutation did not abrogate the cell death response to AVR-PikD effector, although it slightly reduced it compared with the wild type (**Figure 5.17**). Additionally, Pikm-2 Glu230Gly did not affect NLR compatibility with Pikm-1 and was still able to trigger a visible cell death in the presence of AVR-PikD effector, although with lower intensity compared with the wild type (**Figure 5.17**). This further confirms the link between Glu230 and cell death intensity in response to AVR-Pik effectors. I also tested the protein accumulation of Pikm-2 Glu230Gly mutants by western blot confirming that the observed phenotypes are not due to reduction of protein accumulation compared with Pikp-2 or Pikm-2 (**Figure 5.18**).



**Figure 5.17. Reversion to ancestral state of  $Pikm-2$  Glu230 abolish autoimmunity.** Representative leaf spot images depicting  $Pik$  mediated cell death as autofluorescence under UV-light in the presence or absence of AVR-Pik effector. Scoring of the cell death triggered by  $Pikm-2$  or  $Pikm-2$  Glu230Gly mutant when co-expressed with  $Pikp-1$  or  $Pikm-1$  is represented as dot plots. The number of repeats was 60 and 30 for the spots co-infiltrated with mCherry and AVR-PikD, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.



**Figure 5.18. Glu230Gly mutation does not affect Pik-2 protein accumulation.** Western blots showing accumulation of Pikm-2 Glu230Gly. C-terminally 6xHA tagged Pikm-2 Glu230Gly mutant was transiently expressed *N. benthamiana*. C-terminally 6xHA tagged Pikp-2 and Pikm-2 alleles are included as controls. Total protein extracts, were probed  $\alpha$ -HA antisera. Total protein loading is shown by Ponceau staining (PS).

Overall, I reconstructed the evolutionary history of Pik NLR specialization. A multi-step mutation led to the emergence of Glu230 in the Pikm-like clade, which is linked to an efficient cell death response to AVR-Pik effectors. In turn, this polymorphism renders an NLR incompatible with sensor NLRs from the Pikp-like clade. I experimentally validated this observation by reverting the Pikm-2 230 polymorphism to the ancestral state, confirming the phenotypes observed above.

## 5.11 Sensor/helper hetero-pairing alters protein accumulation in Pik NLRs

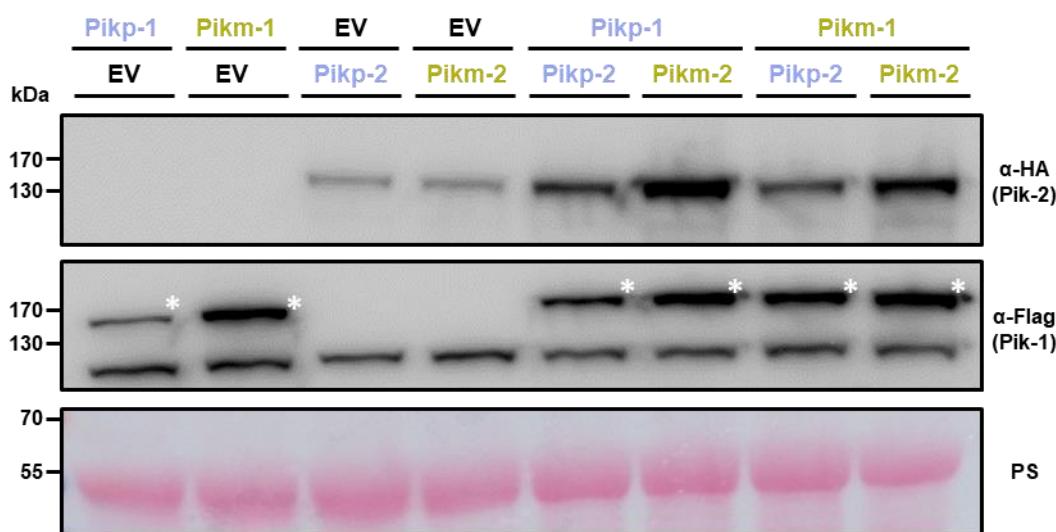
I aimed to obtain mechanistic understanding of Pik NLR pair co-evolution and autoactivation. For this, I first investigated whether accumulation of sensor Pik-1 or helper Pik-2 proteins is altered in the presence of the co-evolved or maladapted pair.

I co-expressed both Pikp-1 and Pikm-1 alleles in *N. benthamiana* in combination with helper Pikp-2 or Pikm-2 alleles. After protein extraction, I performed western blot to detect protein accumulation for both Pik-1 and Pik-2 proteins (**Figure 5.19**). All Pik-1 and Pik-2 alleles were also co-infiltrated with empty vector constructs to evaluate protein accumulation in the absence of the NLR pairs.

Interestingly, protein accumulation of Pik-1 and Pik-2 alleles were consistently increased when they were expressed together compared to co-expression with empty vector (**Figure 5.19**). This is consistent with a model where Pik-1 and Pik-2 associate in sensor/helper NLR heterocomplexes, stabilizing both proteins.

Protein accumulation was increased in both adapted and maladapted combinations but interestingly, accumulation of the helper Pik-2 in the autoimmune pair Pikp-1/Pikm-2 was consistently higher (**Figure 5.19**). This could be due to a different sensor/helper stoichiometry in the consistently activated Pik complex, as observed in some NLR complexes (Hu et al., 2015; Sharif et al., 2019; Tenthorey et al., 2017; Zhang et al., 2015a).

In summary, this differential accumulation of the Pik pair NLRs suggests interaction at the biochemical level between Pik sensor and helper NLRs, possibly involving changes in stoichiometry and association in larger immune signalling complexes.



**Figure 5.19. Pik NLRs display increased protein accumulation when co-expressed together in planta.** Western blots showing protein accumulation of Pik-1 and Pik-2 alleles in different combinations. C-terminally 6xHis3xFLAG tagged Pik-1 alleles were transiently co-expressed with empty vector (EV) or C-terminally 6xHA tagged Pik-2 alleles in *N. benthamiana*. Total protein extracts, were probed with  $\alpha$ -FLAG and  $\alpha$ -HA antisera for Pik-1 and Pik-2, respectively. Asterisks mark the band corresponding to Pik-1. Total protein loading is shown by Ponceau staining (PS).

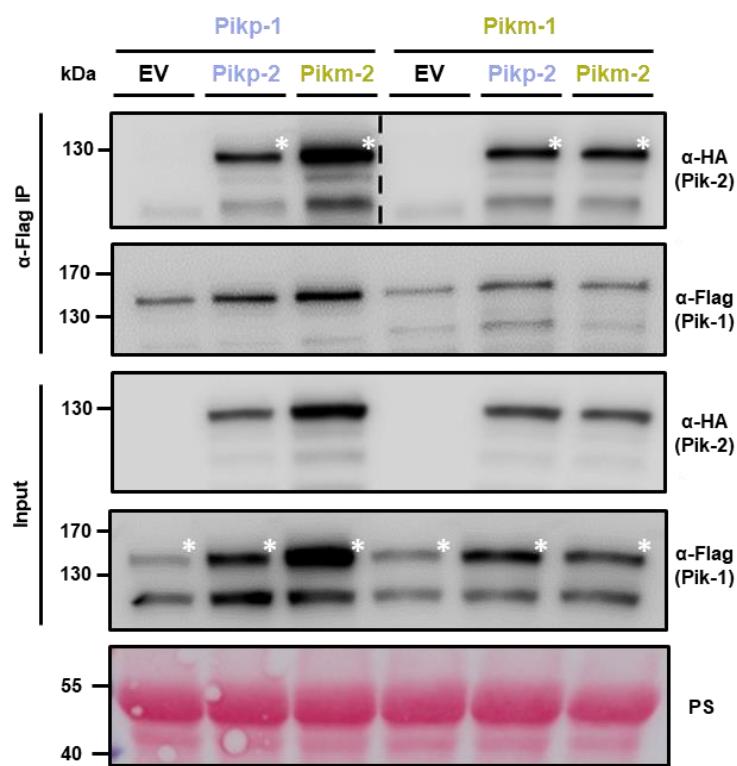
## 5.12 Co-evolved and maladapted Pik pairs form heterocomplexes

Upon activation, plant and animal NLRs associate in large homo and heterooligomers to trigger immune signalling (Hu et al., 2015; Sharif et al., 2019; Tenthorey et al., 2017; Wang et al., 2019a; Zhang et al., 2015a). Prompted by the differences in protein accumulation observed between different combinations of Pik-1 and Pik-2 alleles (**Figure 5.19**), I sought to investigate if differences in cell death phenotypes in maladapted Pik pairs are underpinned by differences in NLR hetero-association.

I co-expressed C-terminally 6xHis3xFLAG tagged Pikp-1 or Pikm-1 with either C-terminally 6xHA tagged Pikp-2 or Pikm-2 in *N. benthamiana*. Following total protein extraction, I performed plant co-immunoprecipitation to test for differences in NLR association (**Figure 5.20**). Pikp-1 and Pikm-1 were co-infiltrated with empty vector as negative controls.

Both helper NLR alleles, Pikp-2 and Pikm-2, could be detected after immunoprecipitation of either Pikp-1 or Pikm-1 sensor NLRs (**Figure 5.20**). Although there is higher accumulation of Pikm-2 in the immunoprecipitate (IP) from the autoactive combination with Pikp-1, I cannot fully conclude is due differences in NLR association as this is also possibly due to differences in protein accumulation in the input (**Figure 5.20**).

Therefore, the cell death phenotypes observed in mismatched pairs are not underpinned by major alterations in association. Interestingly, the association between Pik-1 and Pik-2 in non-activated complexes suggests that Pik sensor and helper NLRs form pre-activation complexes in the resting state.

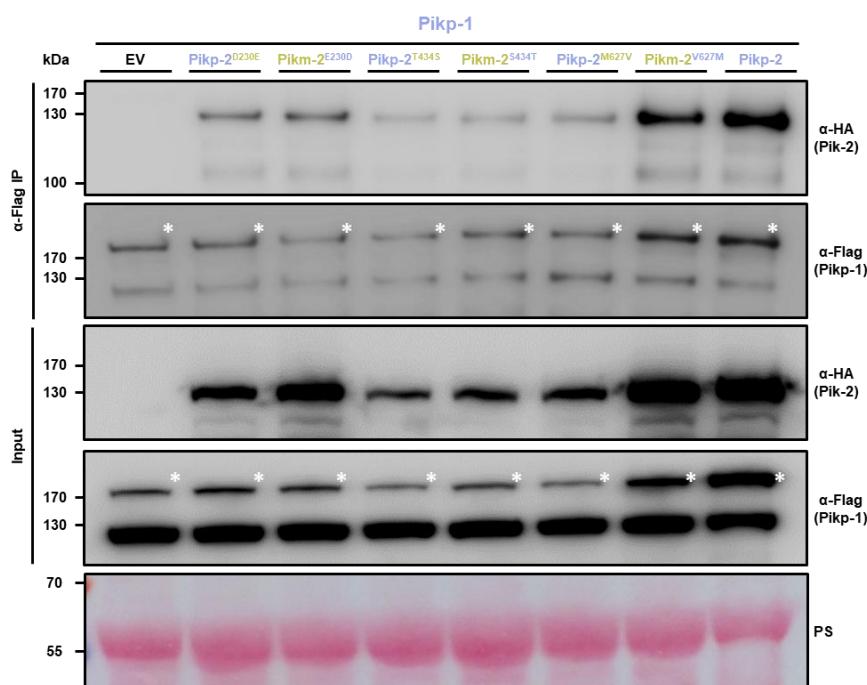


**Figure 5.20. Co-adapted and maladapted Pik NLR pairs associate in planta.** Co-immunoprecipitation of full length Pikp-1 and Pikm-1 alleles in combination with either Pikp-2 or Pikm-2 helper NLRs. C-terminally 6xHA tagged Pikp-2 NLRs were transiently co-expressed with Pikp-1:6xHis3xFLAG (**left**) or Pikm-1:6xHis3xFLAG (**right**) in *N. benthamiana*. Immunoprecipitates obtained with anti-FLAG antiserum, and total protein extracts, were probed with appropriate antisera. A dashed line indicates a crop site on the same blot used to compose the figure. Each experiment was repeated at least three times, with similar results. Asterisks mark the band corresponding to Pik-1. Total protein loading is shown by Ponceau staining (PS).

## 5.13 Point mutations in Pik-2 helper NLR do not alter association with the sensor NLR

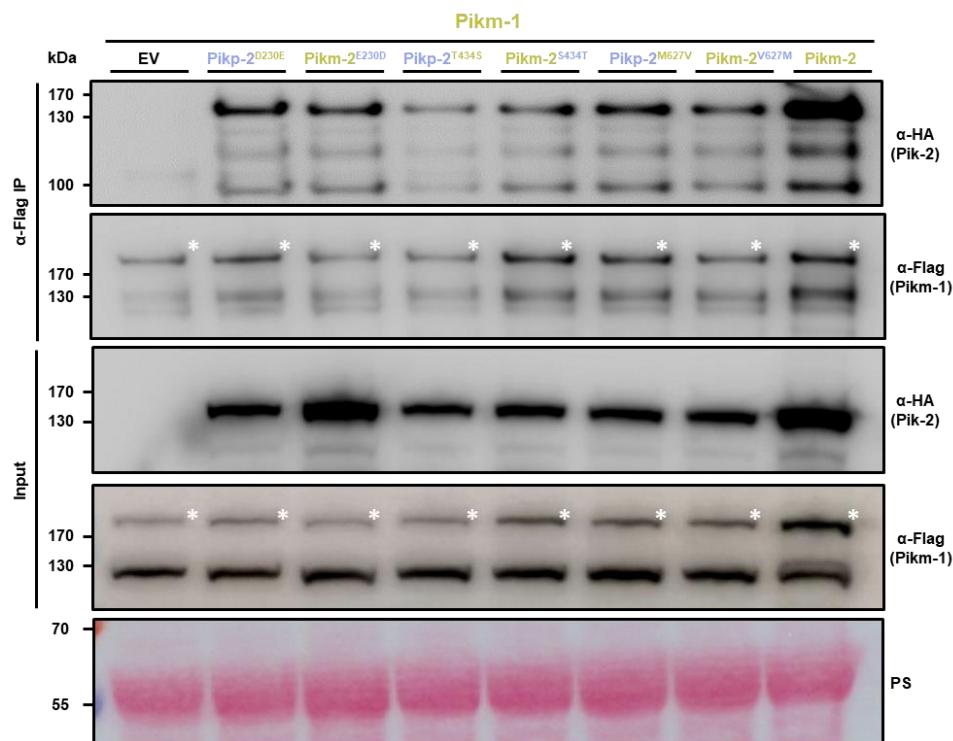
I previously showed that a single amino acid in the Pik-2 helper NLR underpins differences in effector recognition and constitutive immune activation (**Figure 5.9, Figure 5.11**). Given that combination between allelic sensor and helper Pik NLRs do not present dramatic differences in association, I extended the co-IP assays to the point mutants in Pik-2 polymorphic sites.

First, I used the mutants generated and described in section **5.6** (**Figure 5.8**) to test if changes in a specific Pik-2 polymorphic position mediate any change in NLR hetero-association. After co-expression and immunoprecipitation with Pikp-1, I could detect association of this NLR with all the mutants (**Figure 5.21**). Interestingly, all the mutants but one (Pikm-2 Val627Met), displayed lower intensity in the band corresponding to the co-immunoprecipitated Pik-2 proteins compared with the wild-type (**Figure 5.21**). However, differences in protein accumulation make it difficult to claim differences in the strength of helper/sensor association.



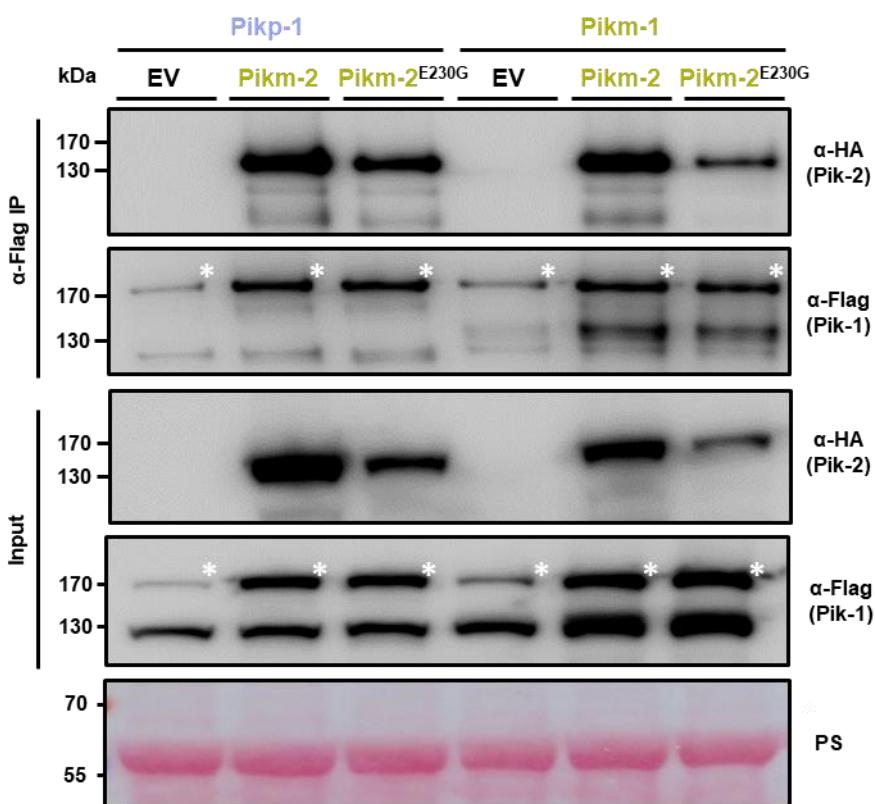
**Figure 5.21. Pik-2 mutants associate with Pikp-1 in planta. Co-immunoprecipitation of full length Pikp-1 with each Pik-2 mutant in polymorphic sites.** C-terminally 6xHA tagged Pik-2 NLR mutants were transiently co-expressed with Pikp-1:6xHis3xFLAG in *N. benthamiana*. Immunoprecipitates obtained with anti-FLAG antiserum, and total protein extracts, were probed with appropriate antisera. Co-expression with empty vector (EV) and wild-type Pikp-2 were included as negative and positive control, respectively. Each experiment was repeated twice, with similar results. Asterisks mark the band corresponding to Pikp-1. Total protein loading is shown by Ponceau staining (PS).

As for Pikp-1, co-immunoprecipitation of Pikm-1 also revealed a clear association to all the Pik-2 mutants (**Figure 5.22**). In this case, I also could observe stronger association of Pikm-1 to the wild-type Pikm-2 compared to the Pik-2 mutants, suggesting the possibility that co-evolved NLRs have stronger association between them. Although protein accumulation in the inputs were more uniform, further repetitions of this experiment are needed to fully support this hypothesis.



**Figure 5.22. Pik-2 mutants associate with Pikm-1 in planta.** Co-immunoprecipitation of full length Pikm-1 with each Pik-2 mutant in polymorphic sites. C-terminally 6xHA tagged Pik-2 NLR mutants were transiently co-expressed with Pikm-1:6xHis3xFLAG in *N. benthamiana*. Immunoprecipitates obtained with anti-FLAG antiserum, and total protein extracts, were probed with appropriate antisera. Co-expression with empty vector (EV) and wild-type Pikm-2 were included as negative and positive control, respectively. Each experiment was repeated twice, with similar results. Asterisks mark the band corresponding to Pikm-1. Total protein loading is shown by Ponceau staining (PS).

Finally, I also tested if reversion of polymorphism 230 to the ancestral state affects to NLR association. I co-expressed both sensor Pikp-1 and Pikm-1 alleles with Pikm-2 Glu230Gly and performed co-immunoprecipitation from plant tissue (**Figure 5.23**).



**Figure 5.23. Reversion to ancestral state in polymorphism 230 does not abrogate association with Pik-1 alleles.** Co-immunoprecipitation of Pikm-2 Glu230Gly mutant with full length Pikp-1 (left) and Pikm-1 (right) alleles. C-terminally 6xHA tagged Pikm-2 Glu230Gly was transiently co-expressed with either Pikp-1:6xHis3xFLAG or Pikm-1:6xHis3xFLAG in *N. benthamiana*. Immunoprecipitates obtained with anti-FLAG antiserum, and total protein extracts, were probed with appropriate antisera. Co-expression with empty vector (EV) and wild-type Pikm-2 were included as negative and positive control, respectively. Each experiment was repeated twice, with similar results. Asterisks mark the band corresponding to Pik-1. Total protein loading is shown by Ponceau staining (PS).

Compared to wild-type Pikm-2, Pikm-2 Glu230Gly accumulates in lower amounts when co-expressed with Pikp-1, which impacts in the final amount of protein detected in the immunoprecipitates (Figure 5.23). This may be because this mutant is not constitutively active in the presence of Pikp-1 (Figure 5.17). Co-expression and immunoprecipitation with Pikm-1 also displayed lower accumulation in comparison with Pikm-2 (Figure 5.23). Given that this mutation does not affect the accumulation of the proteins per se (Figure 5.18), this may reflect a phenotype of the reversion of Glu230 to Gly, correlating with the lower cell death displayed by this mutant (Figure 5.17).

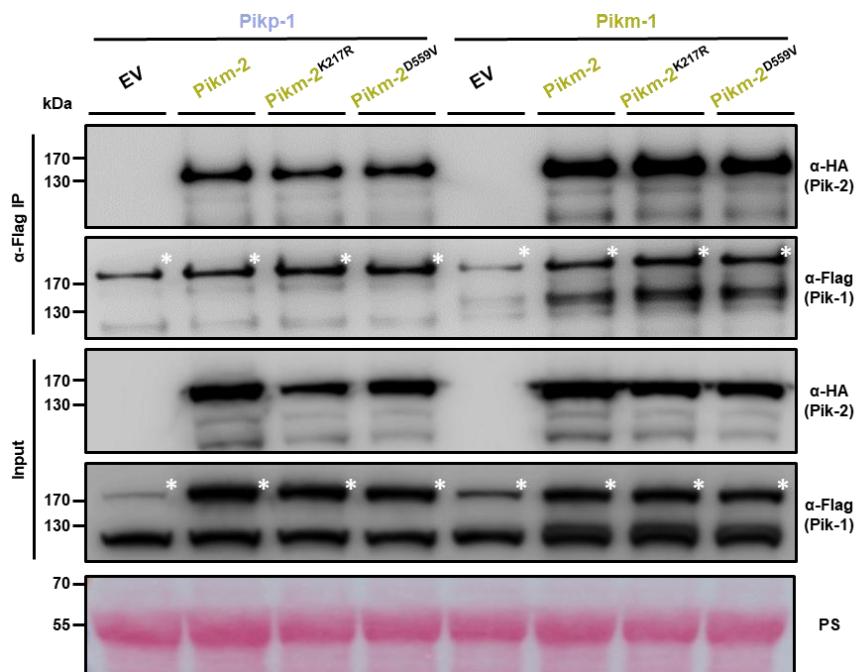
Combined with the data obtained in 5.12, these results suggest that no dramatic changes in association between sensor and helper NLRs govern the cell death and autoimmune phenotypes described previously. On the contrary, it points to subtle differences, perhaps both in association and stoichiometry between Pik NLRs.

## 5.14 Sensor/helper association of Pik NLR pairs is independent of NLR activation

NLR activation and oligomerization are dependent on ADP/ATP exchange (Wang et al., 2019a; Wang et al., 2019b; Zhang et al., 2015a). Mutations at the conserved P-loop and MHD motifs that impair ADP/ATP usage abrogate Pik mediated immune signalling, even in combinations where signalling is constitutively active (Figure 5.7, Figure 5.12).

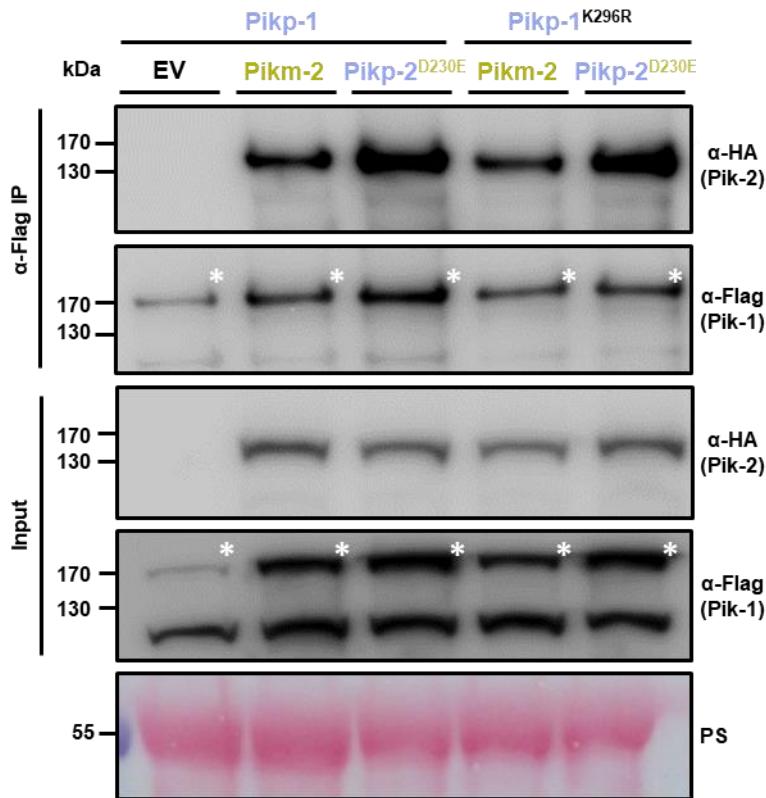
As Pik NLR pairs associate in pre-activation complexes (Figure 5.20), I investigated whether an active NLR is required for this process. Therefore, I co-expressed Pikm-2 P-loop and MHD mutants with either Pikp-1 or Pikm-1 in *N. benthamiana*. Following protein extraction, I performed co-immunoprecipitation of Pik-1 and tested the presence of Pik-2 in the immunoprecipitates by Western blot (Figure 5.24).

The results show that, although mutations at the conserved P-loop and MHD motifs of Pik-2 completely abolish Pik-mediated immune signalling, they do not affect the ability to associate with the sensor NLR Pik-1 compared to wild type Pikm-2 (Figure 5.24).



**Figure 5.24. P-loop and MHD mutations do not affect Pik-2 association to Pik-1.** Co-immunoprecipitation of Pikm-2 P-loop and MHD mutants with full length Pikp-1 (left) and Pikm-1 (right) alleles. C-terminally 6xHA tagged Pikm-2 mutants in P-loop (Lys217Arg) and MHD (Asp559Val) motifs were transiently co-expressed with either Pikp-1:6xHis3xFLAG or Pikm-1:6xHis3xFLAG in *N. benthamiana*. Immunoprecipitates obtained with anti-FLAG antiserum, and total protein extracts, were probed with appropriate antisera. Co-expression with empty vector (EV) and wild-type Pikm-2 were included as negative and positive control, respectively. Each experiment was repeated twice, with similar results. Asterisks mark the band corresponding to Pik-1. Total protein loading is shown by Ponceau staining (PS).

I obtained similar results for the P-loop motif of Pik-1. Although this mutation largely reduces cell death in autoimmune combinations (**Figure 5.14**), both Pikp-1 and Pikp-1 Lys296Arg associate with Pikm-2 and Pikp-2 Asp230Glu in a similar manner (**Figure 5.25**).



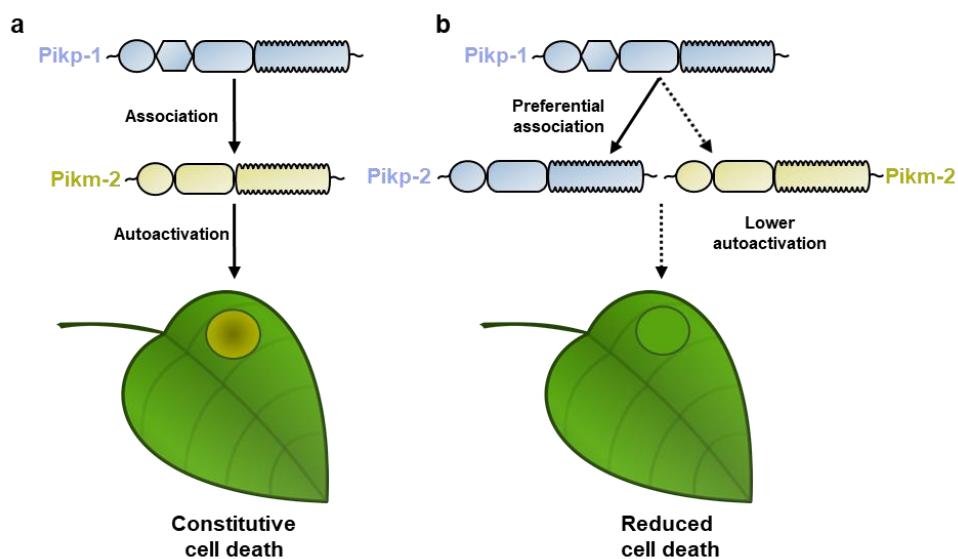
**Figure 5.25. P-loop mutations does not affect Pik-1 association to Pik-2.** Co-immunoprecipitation of Pikm-2 and Pikp-2 Asp230Glu with wild-type Pikp-1 (**left**) and Pikp-1 P-loop mutant (**right**). C-terminally 6xHA tagged Pikm-2 and Pikp-2 Asp230Glu were transiently co-expressed with C-terminally 6xHis3xFLAG tagged wild-type Pikp-1 or Pikp-1 Lys296Arg in *N. benthamiana*. Immunoprecipitates obtained with anti-FLAG antiserum, and total protein extracts, were probed with appropriate antisera. Co-expression with empty vector (EV) and wild-type Pikp-1 were included as negative and positive control, respectively. Each experiment was repeated twice, with similar results. Asterisks mark the band corresponding to Pik-1. Total protein loading is shown by Ponceau staining (PS).

Overall, these results imply that Pik NLR pair association in pre-activation complexes does not require active NLRs and is independent of nucleoside usage. Altogether, with the previous sections, these results point to a scenario where sensor and helper Pik NLRs are forming an inactive complex that, upon activation and possible ADP/ATP exchange, undergo changes in conformation and/or stoichiometry to trigger immune signalling.

## 5.15 Sensor and helper NLRs preferentially associate to their co-evolved pair

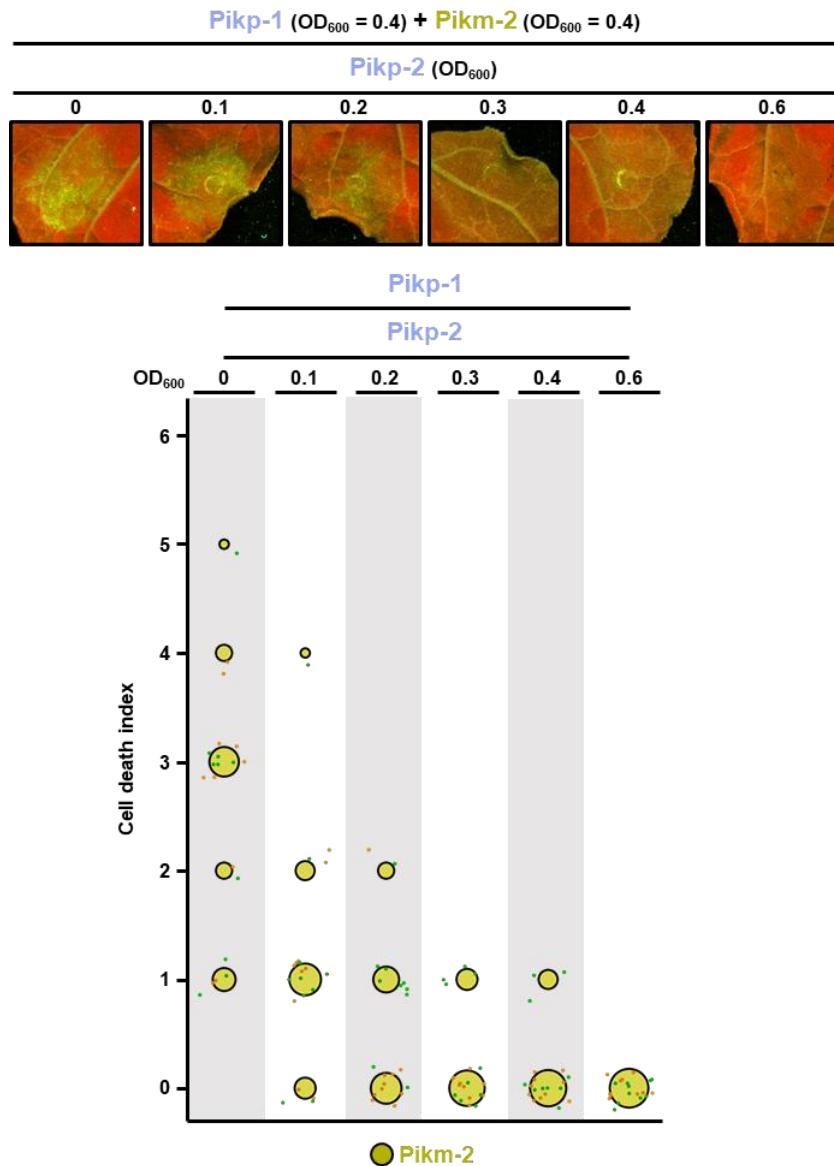
The findings that mutations and polymorphisms that alter Pik cell death and autoimmune phenotypes do not dramatically alter association between helper and sensor suggests that these phenotypes are linked to small changes or rearrangements of pre-activation complexes. To gain a deeper knowledge of Pik pair association, I investigated whether allelic Pik NLRs display any preference to associate to their co-evolved NLR pair.

Taking advantage of the ability of both autoactive and non-autoactive pairs to associate, I designed a NLR competition assay to test for differential binding affinities between allelic NLRs. When co-expressed together, *Pikp-1* and *Pikm-2* associate and trigger autoactivation, leading to a constitutive cell death phenotype (**Figure 5.26a**). In a scenario where co-evolved Pik NLR pairs display preferences in helper/sensor association, if both *Pikp-2* and *Pikm-2* are present, the majority of sensor *Pikp-1* NLRs would associate with the non-autoactive helper *Pikp-2*, reducing the levels of constitutive cell death (**Figure 5.26b**). To test this hypothesis, I co-infiltrated *Pikp-1* and *Pikm-2* NLRs in *N. benthamiana* with a fixed concentration ( $OD_{600} = 0.4$ ) for each construct. I also co-infiltrated increasing concentrations of *Pikp-2* (spanning from  $OD_{600} = 0.1$  to  $OD_{600} = 0.6$ ) and I scored the cell death phenotype under UV light.



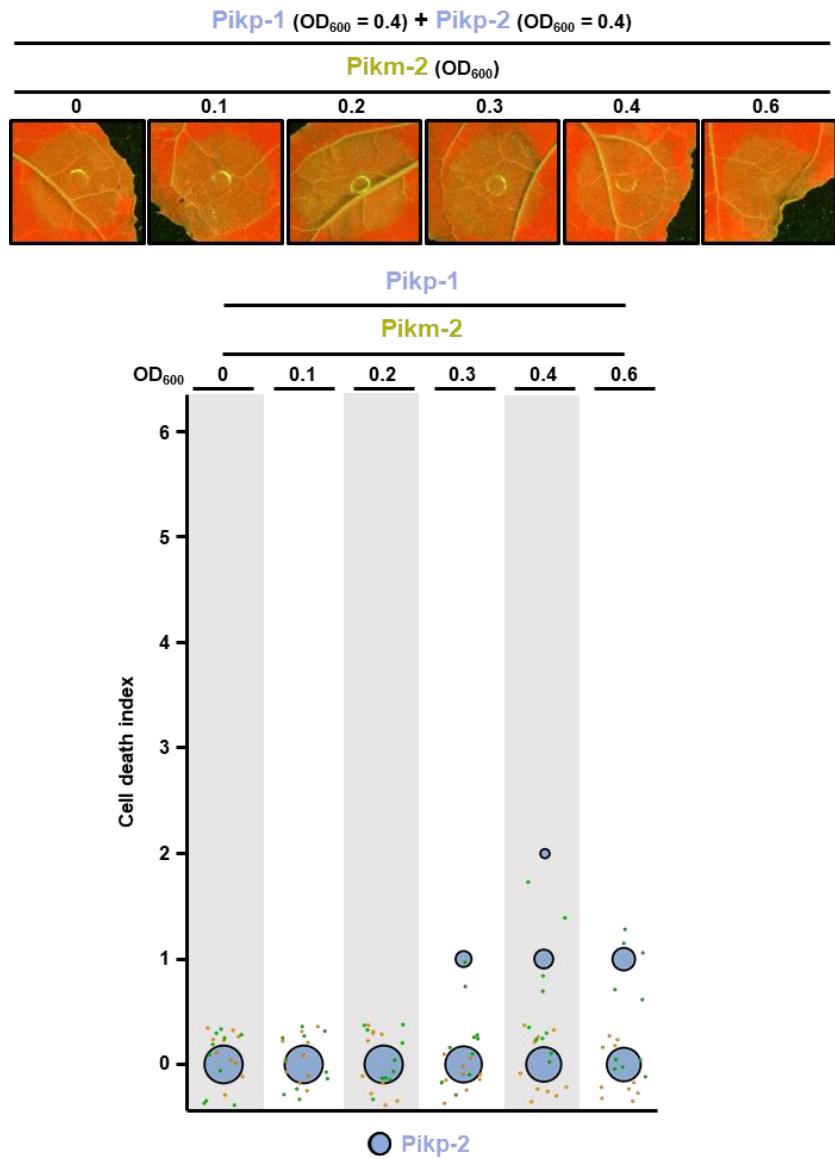
**Figure 5.26. Schematic representations of Pik NLR competition assays.** (a) When *Pikp-1* (coloured in Ice blue) is co-expressed with *Pikm-2* (coloured in gold), both NLRs associate and trigger NLR activation that leads to constitutive cell death in *N. benthamiana*, depicted by the development of chlorotic and necrotic leaf tissue. (b) In a preferential association scenario, with both *Pikp-2* and *Pikm-2* present, *Pikp-1* would associate more with co-evolved *Pikp-2* instead of to *Pikm-2* (depicted by the solid and dashed lines, respectively). This would reduce constitutive immune signalling and cell death.

Interestingly, Pikp-2 acted as a suppressor of autoimmune phenotypes triggered by Pikp-1/Pikm-2 as increasing concentrations of Pikp-2 lowered the constitutive cell death phenotype (**Figure 5.27**). I could observe a large reduction of cell death even in the lowest concentration of Pikp-2 (**Figure 5.27**), indicating that Pikp-1 displays preference to signal through co-evolved Pikp-2 rather than Pikm-2.



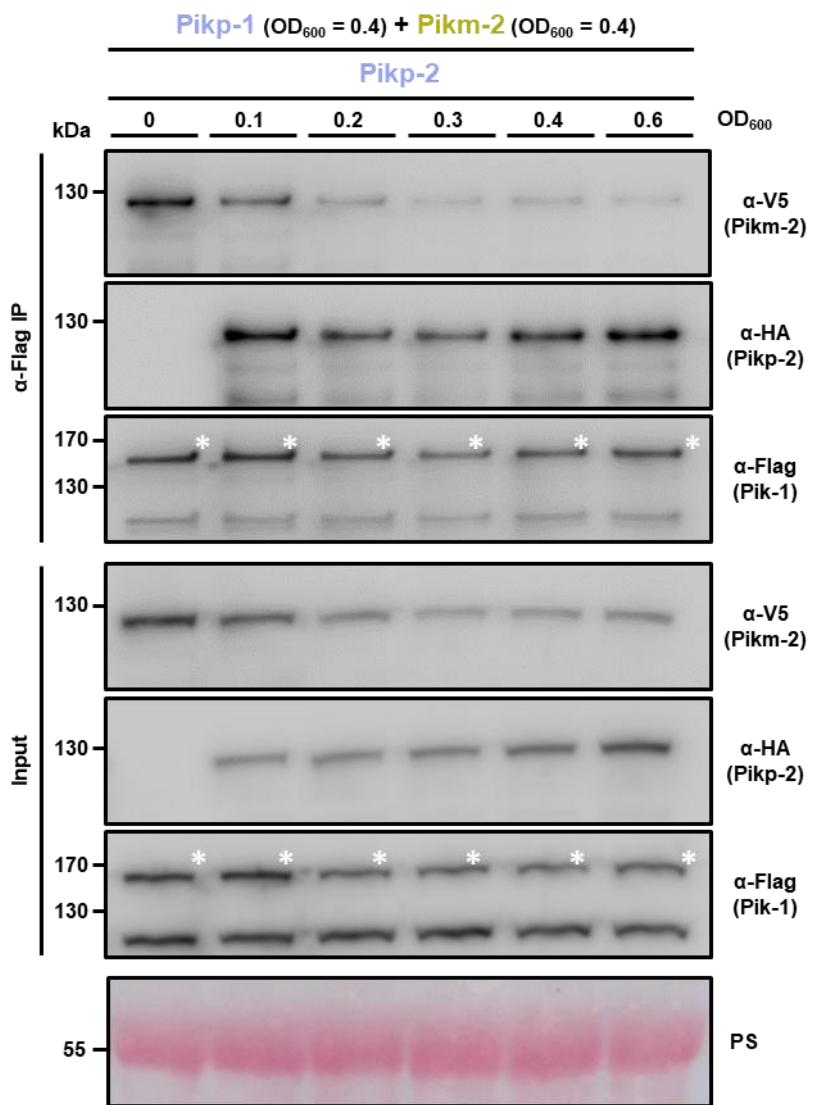
**Figure 5.27. Pikp-2 suppress Pikm-2 mediated cell death.** Representative leaf spot images depicting Pikm-2 mediated cell death in the presence of Pikp-1 and increasing concentration of Pikp-2 as autofluorescence under UV-light. Scoring of the cell death assay is represented as dot plots. For each experiment, Pikp-1 and Pikm-2 were co-infiltrated at OD<sub>600</sub> = 0.4 each. Increasing concentrations of Pikp-2 were added to each experiment (from left to right: 0, 0.1, 0.2, 0.3, 0.4 and 0.6 OD<sub>600</sub>). A total of two biological replicates with 10 internal repeats each were performed for each experiment. For each sample, all the data points are represented as dots with a distinct colour for each of the two biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

To validate the assay, I also co-infiltrated a fixed concentration of Pikp-1 and Pikp-2, this time increasing the amount of Pikm-2. In this case, Pikm-2 could not overcome the suppression by the presence of Pikp-2, even at the higher concentrations (**Figure 5.28**).



**Figure 5.28. Pikm-2 cannot overcome suppression by Pikp-2.** Representative leaf spot images depicting Pikm-2 mediated cell death in the presence of Pikp-1 and Pikp-2 and increasing concentration of Pikm-2 as autofluorescence under UV-light. Scoring of the cell death assay is represented as dot plots. For each experiment, Pikp-1 and Pikp-2 were co-infiltrated at OD<sub>600</sub> = 0.4 each. Increasing concentrations of Pikm-2 were added to each experiment (from left to right: 0, 0.1, 0.2, 0.3, 0.4 and 0.6 OD<sub>600</sub>). A total of two biological replicates with 10 internal repeats each were performed for each experiment. For each sample, all the data points are represented as dots with a distinct colour for each of the two biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

Finally, I asked if the suppression of the constitutive cell death by Pikp-2 is due to a higher association affinity to the sensor NLR Pikp-1, than to Pikm-2. To test this, I co-infiltrated Pikp-1 and Pikm-2 with increasing amounts of Pikp-2. After total protein extraction, I performed co-immunoprecipitation of Pikp-1 and I tested for the presence of Pikp-2 or Pikm-2 in the immunoprecipitates (**Figure 5.29**).



**Figure 5.29. Pikp-2 outcompetes Pikm-2 association to Pikp-1.** Co-immunoprecipitation of Pikm-2 and Pikp-1 in the presence of increasing concentrations of Pikp-2. C-terminally V5 tagged Pikm-2 and C-terminally 6xHis3xFLAG tagged Pikp-1 were transiently co-expressed in *N. benthamiana* alongside with increasing concentrations of C-terminally 6XHA tagged Pikp-2 (from left to right: 0, 0.1, 0.2, 0.3, 0.4 and 0.6 OD<sub>600</sub>). Immunoprecipitates obtained with anti-FLAG antiserum, and total protein extracts, were probed with appropriate antisera. Each experiment was repeated twice, with similar results. Asterisks mark the band corresponding to Pikp-1. Total protein loading is shown by Ponceau staining (PS).

Differences in protein accumulation observed in the different combinations of Pik pairs makes it particularly challenging to obtain even inputs for this experiment. For example, Pik-2 proteins seem more stable in association with Pik-1 (**Figure 5.19**), therefore if a Pik-2 protein is outcompeted from a hypothetical complex, this will be perceived as a reduced protein accumulation in the input. On the contrary, if Pik-2 proteins seem stabilized in autoactive combinations (**Figure 5.19**), then the amount of protein in concentrations where Pikp-2 suppresses constitutive cell death may seem lowered.

Nevertheless, co-immunoprecipitation results depicted a preference in association of Pikp-1 to Pikp-2 over Pikm-2. Increasing concentrations of Pikp-2 reduced the association of Pikp-1 to Pikm-2, outcompeting Pikm-2 from a heterocomplex with Pikp-1 (**Figure 5.29**). This correlates with the reduction of the constitutive cell death assay observed in the NLR competition experiments (**Figure 5.27**).

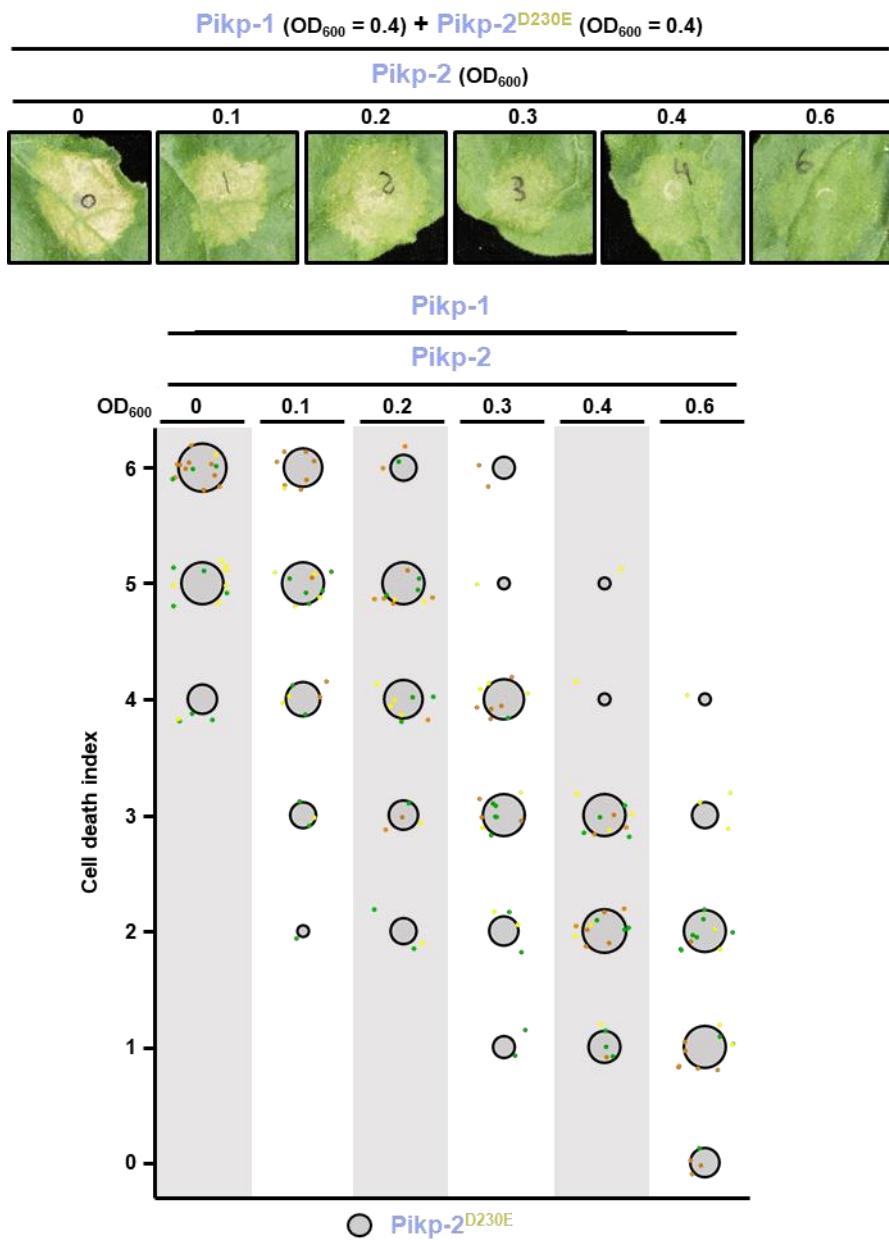
Altogether, these data reveal that co-evolved Pik NLRs display preference in association over non-co-evolved NLRs. This represents another example of NLR pair co-adaptation. These differences may underpin the observed phenotypes in cell death and autoimmunity.

## **5.16 Pik-2 polymorphisms underpin helper/sensor association preference**

I aimed to investigate if the preferential binding between Pikp-1 and Pikp-2 observed before is underpinned by the same polymorphism involved in cell death response to AVR-Pik effectors and autoimmunity. For this, I repeated the NLR competition assay co-infiltrating a fixed concentration of Pikp-1 and the autoactive mutant Pikp-2 Asp230Glu, with increased amounts of Pikp-2.

Constitutive cell death phenotypes triggered by the combination of Pikp-1 and Pikp-2 Asp230Glu is strong (**Figure 5.6**, **Figure 5.11**). However, as for Pikm-2, increasing concentrations of Pikp-2 significantly reduced the cell death phenotypes (**Figure 5.30**).

This result shows that Pik-2 polymorphisms involved in specialization towards AVR-Pik effector response and autoimmunity, may also be related to association between sensor and helper NLRs.



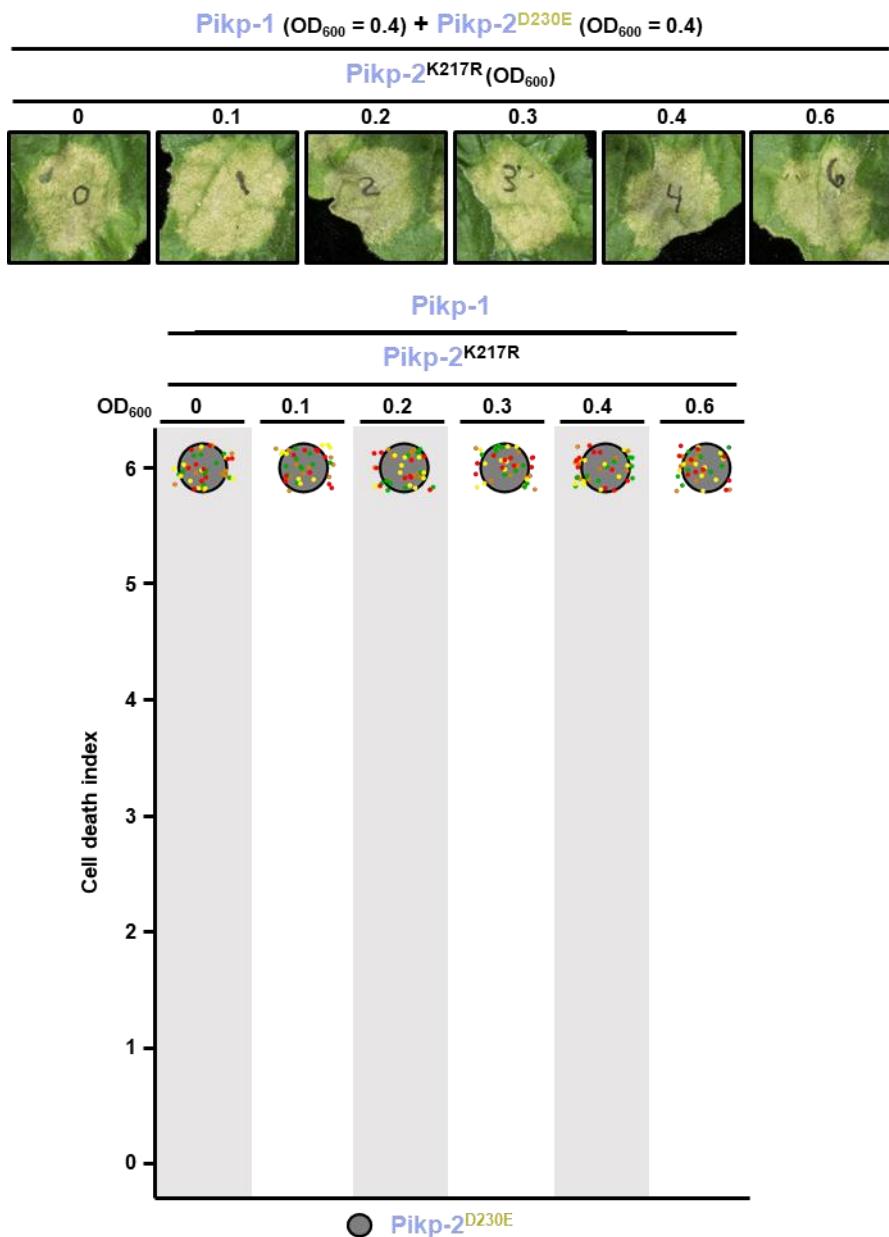
**Figure 5.30. Pikp-2 partially suppress cell death mediated by Pikp-2 Asp230Glu.** Representative leaf spot images depicting Pikp-2 Asp230Glu mediated cell death in the presence of Pikp-1 and increasing concentration of Pikp-2. Scoring of the cell death assay is represented as dot plots. For each experiment, Pikp-1 and Pikp-2 Asp230Glu were co-infiltrated at  $OD_{600} = 0.4$  each. Increased concentration of Pikp-2 were added to each experiment (from left to right: 0, 0.1, 0.2, 0.3, 0.4 and 0.6  $OD_{600}$ ). A total of three biological replicates with 10 internal repeats each were performed for each experiment. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

## 5.17 P-loop and MHD motifs are essential for Pikp-2 suppression of constitutive cell death

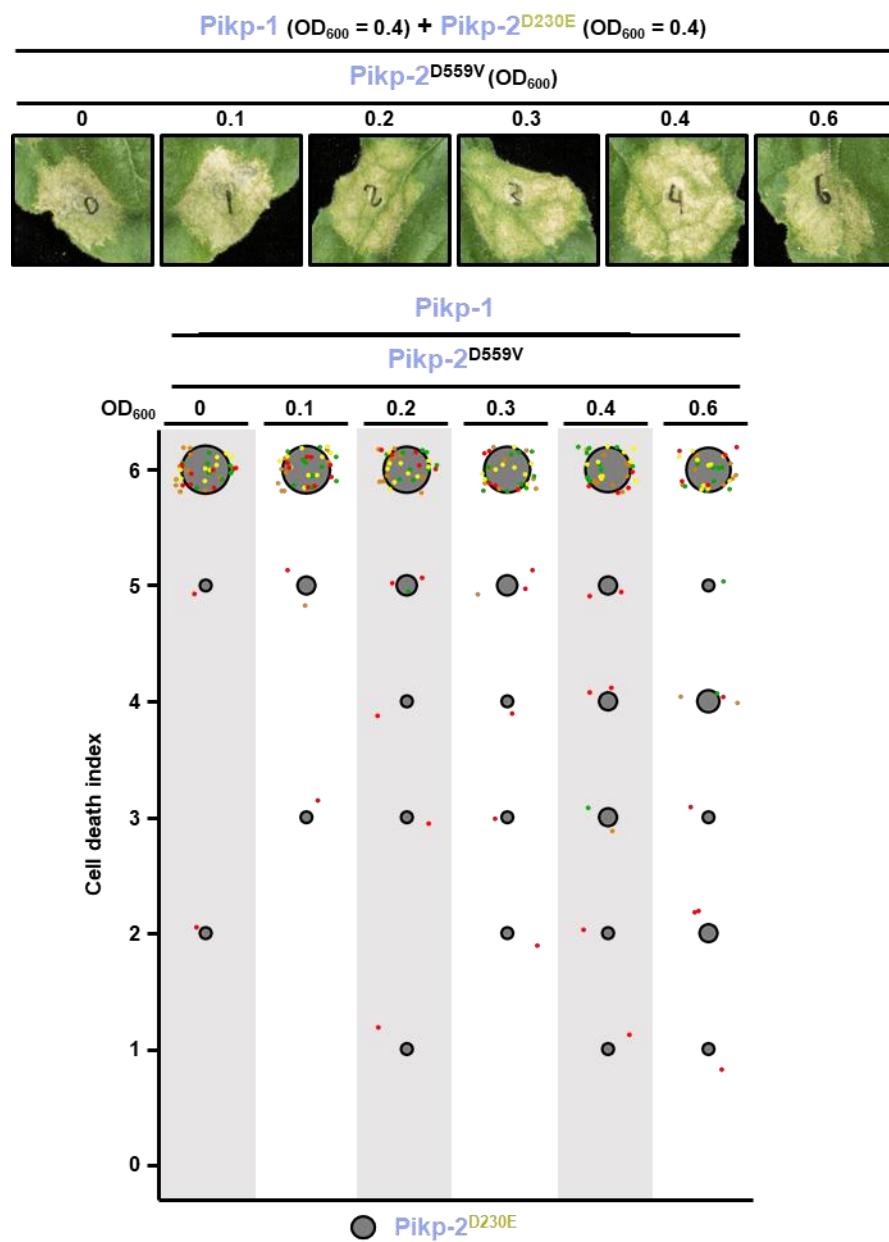
Pik NLRs do not require a functional P-loop or MHD motif to form heterocomplexes (**Figure 5.24**, **Figure 5.25**). To obtain a better understanding of the requirement of active NLRs in paired NLR association, I used the Pik competition assay with a fixed amount of Pikp-1 and Pikp-2 Asp230Glu and, instead of wild-type Pikp-2, I added increased amounts of Pikp-2 P-loop or MHD mutant.

Opposite to wild type Pikp-2, where increasing amounts of this NLR suppress constitutive immune responses, Pikp-2 Lys217Arg did not reduce the cell death phenotypes even in the higher concentrations (**Figure 5.31**). Similarly, increasing concentrations of the MHD motif mutant Pikp-2 Asp559Val did not display a significant reduction of the cell death outcome (**Figure 5.32**).

These results show that, although they are not required for association in pre-activation complexes, P-loop and MHD motifs are required for the preferential sensor/helper association observed in Pikp-1 and Pikp-2. Therefore, association from a resting state into an activated complex requires active NLRs, possibly involves ADP/ATP exchange, oligomerization and changes in helper/sensor stoichiometry. This is consistent with studies in *Arabidopsis* NLR RPP-7 where P-loop mutants retain the ability to associate with autoactive forms of HR4 but are not capable to form high order assemblies (Li et al., 2019b).



**Figure 5.31. Cell death suppression by Pikp-2 requires an intact P-loop motif.** Representative leaf spot images depicting Pikp-2 Asp230Glu mediated cell death in the presence of Pikp-1 and increasing concentration of Pikp-2 Lys217Arg. Scoring of the cell death assay is represented as dot plots. For each experiment, Pikp-1 and Pikp-2 Asp230Glu were co-infiltrated at  $OD_{600} = 0.4$  each. Increased concentration of Pikp-2 Lys217Arg were added to each experiment (from left to right: 0, 0.1, 0.2, 0.3, 0.4 and 0.6  $OD_{600}$ ). A total of four biological replicates with 10 internal repeats each were performed for each experiment. For each sample, all the data points are represented as dots with a distinct colour for each of the four biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.



**Figure 5.32. Cell death suppression by Pikp-2 requires an intact MHD motif.** Representative leaf spot images depicting Pikp-2 Asp230Glu mediated cell death in the presence of Pikp-1 and increasing concentration of Pikp-2 Asp559Val. Scoring of the cell death assay is represented as dot plots. For each experiment, Pikp-1 and Pikp-2 Asp230Glu were co-infiltrated at  $OD_{600} = 0.4$  each. Increased concentration of Pikp-2 Asp559Val were added to each experiment (from left to right: 0, 0.1, 0.2, 0.3, 0.4 and 0.6  $OD_{600}$ ). A total of four biological replicates with 10 internal repeats each were performed for each experiment. For each sample, all the data points are represented as dots with a distinct colour for each of the four biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

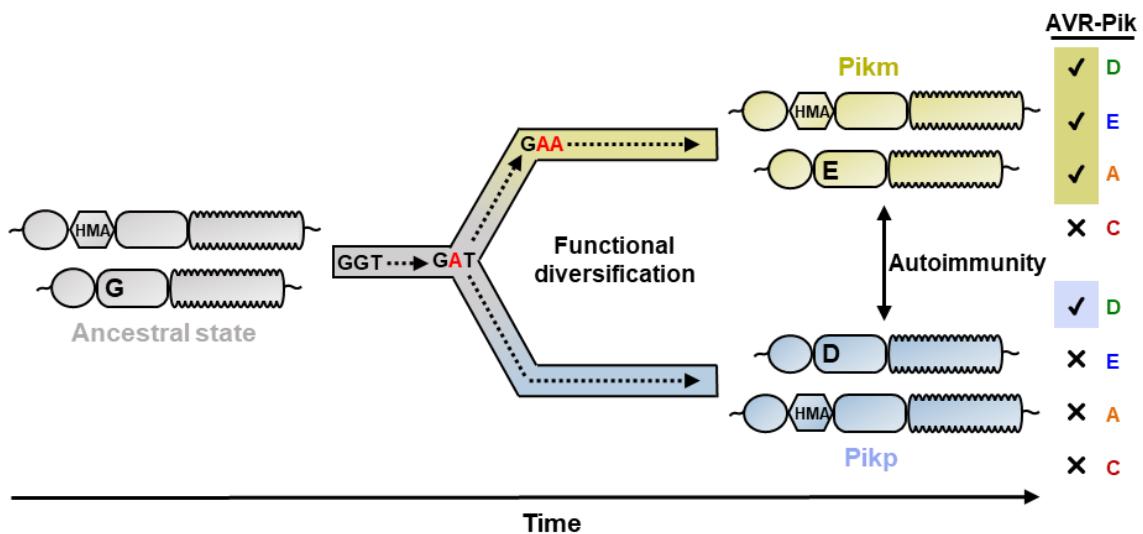
## 5.18 Discussion

Genetic incompatibility caused by subfunctionalization of genes between isolated populations is described as Bateson–Dobzhansky–Muller (DBM) hybrid incompatibility (Dobzhansky, 1937), and has been suggested as a cause of speciation (Lynch and Force, 2000; Taylor et al., 2001; Werth and Windham, 1991). Pioneering studies in the model system *Arabidopsis thaliana* have linked hybrid necrosis and DBM evolution to constitutive activation of plant NLRs (Bomblies et al., 2007; Bomblies and Weigel, 2007; Li et al., 2019b; Tran et al., 2017). Indeed, NLRs have been found to be hot spots for deleterious combinations that lead to hybrid necrosis (Chae et al., 2014).

Rapid co-evolution with pathogen effectors can drive functional specialization of immune receptors. These changes can therefore give rise to autoimmune phenotypes and lead to speciation in the evolutionary timescale. This places NLR functions beyond immune resistance and implies an important role for these proteins as determinants of plant speciation and evolution.

The notion that NLRs can work in pairs is relatively new in the plant-microbe interaction field (Eitas and Dangl, 2010; Jones et al., 2016). Under this emerging framework, it becomes evident that paired NLR must co-adapt to optimise and maintain a tight control of immune responses. The extent to which NLR pairs co-evolve to efficiently respond to pathogen effectors while keeping a fine-tuned regulation of immune responses is not well understood. Particularly, how major evolutionary events, such as the integration of an unconventional domain, affect the paired receptors remain obscure. To date, we have little evidence of co-evolution between NLR pairs in the molecular detail, particularly in systems other than *Arabidopsis*.

In this chapter, I described how allelic rice NLR *Pik* pairs with different effector recognition specificities have co-evolved, functionally diversifying and triggering autoimmune phenotypes when the members of the pair are mismatched (**Figure 5.33**). A single amino acid polymorphism largely underpins specialization of the helper NLR to its corresponding sensor NLR. Changes in this receptor affect cell death outcomes in effector recognition and autoimmune phenotypes. By narrowing down the determinants of NLR specialization to a single amino acid, I traced the evolutionary history of this polymorphism (**Figure 5.33**).



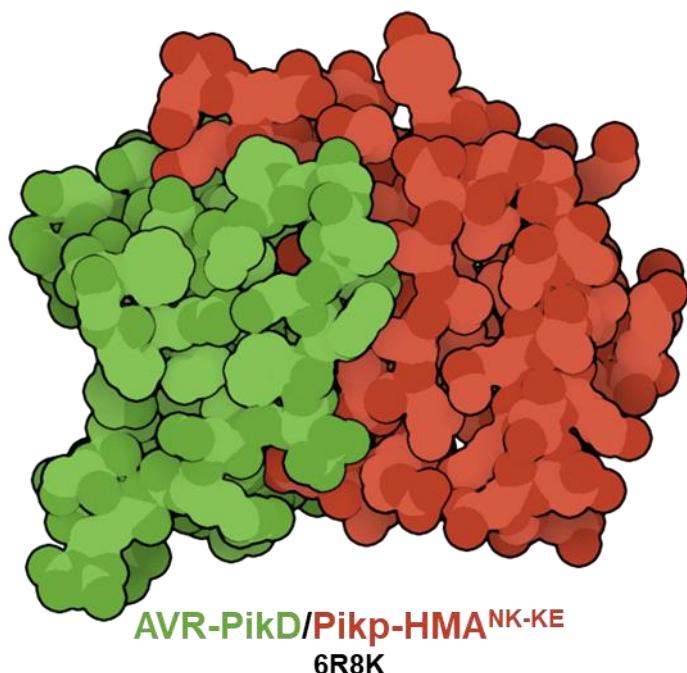
**Figure 5.33. Schematic representation of the evolutionary model proposed in this chapter.** Both Pikp (coloured in ice blue) and Pikm (coloured in gold), have evolved and specialized from an ancestral NLR pair (coloured in grey). They functionally diversified to gain recognition to a different subset of AVR-Pik effectors (represented on the right). Residues at Pik-2 polymorphic position 230 are indicated and mutations predicted to have occurred during this transition are indicated in red. The combination of modern Pikp and Pikm causes NLR autoactivation and constitutive cell death in *N. benthamiana*.

This work provides an evolutionary framework for how differential selective pressures, such as recognition of pathogen strains via effector binding, impact NLRs pairs. This uncovers the potential of paired NLRs to give rise to autoimmune phenotypes during evolution and importantly, links pathogen perception and autoimmunity.

## Chapter 5

# 6

## The integrated HMA domain determines allelic compatibility in the rice NLR pair Pik





## 6. The integrated HMA domain determines allelic compatibility in the rice NLR pair Pik

### Abstract

Throughout evolution, NLR receptors must achieve fine-tuned pathogen recognition whilst maintaining tight control over immune activation. Diversifying evolution renders NLR alleles that trigger autoimmune responses when combined in the same genetic background. Several NLR pairs harbour atypical domains integrated into their architecture, which are likely derived from pathogen targets. These domains can directly bind effectors and are under rapid co-evolution with pathogens and hence can be very variable. Acquisition of these domains is a major event that will likely alter NLR activation, shaping their evolution. How plant NLRs and their rapidly evolving integrated domains co-adapt to enable an efficient and regulated response to pathogen effectors is unknown. The integrated HMA domain in the NLR pair Pik mediates recognition to rice blast pathogen effector AVR-Pik. This domain is highly polymorphic and determines differential effector recognition between NLR alleles. Allelic Pikp and Pikm pairs display autoimmune responses when sensor and helper NLRs are mismatched. Here, I uncovered signatures of co-adaptation between Pik NLRs and their integrated HMA domain. Pikm-HMA, but not Pikp-HMA, suppresses the constitutive cell death triggered by Pikm-2, underpinning allelic compatibility. Structural analysis and mutagenesis identified the amino acids determining Pikm-HMA repression. These results exemplified how, upon integration, integrated domains co-adapt with their host NLRs. Diversifying selection in the integrated domain imposed by differential co-evolution with pathogen effectors can lead to distinct activation mechanisms and incompatibility between allelic NLRs.

## Chapter 6

## 6.1 Introduction

Intracellular NLR immune receptors form the most diverse plant protein family (Baggs et al., 2017; Meyers et al., 2003), largely shaped by their interaction with rapidly evolving pathogen effectors (Bergelson et al., 2001; Upson et al., 2018). Rapid diversification of plant NLRs can lead to deleterious effects produced by constitutive immune activation (Bomblies et al., 2007; Chae et al., 2014; Tran et al., 2017). These autoimmune phenotypes restrain the evolution of plant NLRs (Chae et al., 2016; MacQueen et al., 2016) and can lead to the reproductive isolation of plant populations with allelic immune receptors (Bomblies et al., 2007; Bomblies and Weigel, 2007).

Extensive structure-function studies have revealed that NLR protein domains need to be finely matched for optimal specificity and robustness of immune responses. Minor changes can drastically alter the activity of NLR variants (Bernoux et al., 2016; Stirnweis et al., 2014b). Even swapping small protein domains between allelic NLRs can trigger severe autoimmune responses (Barragan et al., 2019; Slootweg et al., 2017; Slootweg et al., 2013; Wang et al., 2015b). The mechanistic causes of autoimmunity could be due incompatible intramolecular interactions between diversified NLRs and host proteins, including other NLRs (Tran et al., 2017).

One of the most striking cases of immune receptor evolution driven by pathogen effectors is the integration of unconventional domains derived from effector host targets into the architecture of some plant NLRs (Kroj et al., 2016; Sarris et al., 2016). Although these domains were detected in early analysis of plant NLR repertoires (Meyers et al., 2003), it was not until recently that functional studies uncovered their role in mediating pathogen recognition by, in most cases, directly binding pathogen effectors (Cesari et al., 2013; De la Concepcion et al., 2018; Le Roux et al., 2015; Maqbool et al., 2015; Ortiz et al., 2017; Sarris et al., 2015). This led to one of the most recent paradigm shifts in molecular plant-microbe interactions (Cesari et al., 2014a).

Overall, it remains to be understood how paired NLR receptors convert effector detection into defence activation, and what is the role of integrated domains in activating immune signalling. To date, studies in different NLR pairs harbouring integrated domains have suggested distinct contributions of these domains lead to cell death signalling.

In the *Arabidopsis* NLR pair RRS1/RPS4, the sensor NLR RRS1 harbours an integrated WRKY domain that binds pathogen effectors and triggers immune responses (Le Roux et al., 2015; Sarris et al., 2015). The helper NLR RPS4 is constitutively active and triggers spontaneous cell death. In the absence of disease, RRS1 suppresses RPS4 activation,

but detection of pathogen effectors releases this repression, leading to activation (Le Roux et al., 2015; Sarris et al., 2015). The integrated WRKY domain in RRS1 has evolved to suppress cell death signalling mediated by RPS4. Thus, deletion of this domain and domain swaps with independently evolved WRKY domains render constitutive activation of immune responses (Ma et al., 2018). Certain mutations in the WRKY domain also yield constitutively active NLRs (Noutoshi et al., 2005), possibly by affecting the ability of this domain to repress RPS4 activity.

In the rice NLR pair Pia, helper NLR Pia-1 is also constitutively active and is repressed by sensor NLR Pia-2 (Cesari et al., 2014b). Binding of rice blast effector AVR-Pia or AVR1-CO39 to the integrated HMA domain in Pia-2 releases Pia-1 signalling and leads to disease resistance (Cesari et al., 2013; Guo et al., 2018b; Ortiz et al., 2017). However, contrary to the *Arabidopsis* RRS1/RPS4 pair, deletion of the Pia-2 integrated HMA does not trigger constitutively active responses (Cesari et al., 2014b), suggesting this domain is not responsible for the repression of Pia-1 immune responses.

Although it has not yet been elucidated, the activation mechanism of the rice NLR pair Pik seems different from the RRS1/RPS4 and Pia-1/Pia-2 pairs. Neither Pik-1 nor Pik-2 trigger immune signalling in the absence of the matching NLR pair (De la Concepcion et al., 2018; Maqbool et al., 2015), suggesting a transactivation or cooperation mechanism between sensor and helper NLRs (Bialas et al., 2018). As shown in the previous chapter, Pik-1 and Pik-2 associate in pre-activation complexes and, upon binding of AVR-Pik effectors to the integrated HMA domain in Pik-1, the complex undergoes ATP/ADP-dependent changes in association, leading to immune activation. The role of the integrated HMA domain beyond binding to AVR-Pik effectors is currently unknown.

Diverse domains have been integrated in NLRs, particularly in the context of genetically linked pairs (Bailey et al., 2018; Kroj et al., 2016; Sarris et al., 2016; Wang et al., 2019c), indicating that this is a successful strategy of receptor diversification. As these domains directly engage with pathogen effectors, they are often the most diverse regions in NLRs (Costanzo and Jia, 2010). Integration of a domain acting as effector bait will dramatically shape the evolution and the activation mechanism of the acceptor NLR. However, as stated before, NLRs are particularly prone to functional alterations produced by even small modifications, causing loss-of-function and autoimmune phenotypes. Therefore, how can NLRs adapt to a major event, such as the integration of a rapidly evolving domain into their architecture?

In the previous chapter, I showed how sensor and helper Pik NLRs have co-evolved, possibly following the diversifying evolution of this NLR to achieve enhanced effector

recognition. In turn, this leads to constitutive cell death phenotypes when allelic helper NLR *Pikm-2* is mismatched with the sensor NLR allele *Pikp-1*. Both specialization and autoimmune phenotypes are largely underpinned by a single polymorphism in *Pik-2*.

Interestingly, NLR pairing involving allelic sensor *Pikm-1* do not trigger any autoimmune phenotype, suggesting that this NLR may have co-evolved to attenuate these phenotypes. The *Pik-1* HMA is the most variable domain between *Pik* NLR alleles (Costanzo and Jia, 2010), and polymorphisms at the binding interface with AVR-*Pik* effectors underpin allelic specificity in effector recognition (De la Concepcion et al., 2019; De la Concepcion et al., 2018). However, many of the polymorphisms between *Pikp*-HMA and *Pikm*-HMA fall outside the binding interfaces, suggesting these integrated domains may have undergone a functional diversification separate from effector binding. This may underpin sensor/helper specificity.

In this chapter, I use autoactive NLR combinations and mutagenesis to investigate the role of the integrated HMA domain in allelic compatibility. Using deletions and chimeric proteins, I found that the *Pikm*-HMA domain has co-evolved to repress activation of the *Pikm-2* helper NLR in the absence of effectors. By contrast, the *Pikp*-HMA does not act as a negative regulator, suggesting a different activation mechanism may exist between diversified *Pik* alleles. These differences in regulation underpin allelic compatibility between *Pik* alleles and are largely underpinned by individual residues in *Pikm*-HMA, located away from the interaction interface with AVR-*Pik* effectors.

This work sheds light on the activation mechanisms of *Pik* receptors. Furthermore, the results exemplify how integrated domains and NLRs co-adapt, and how NLR diversification can lead to differential regulation of allelic NLRs, ultimately underpinning autoimmune phenotypes. Knowledge of NLR allelic incompatibility is required for the rational design of engineered receptors with increased pathogen detection capabilities. Therefore, understanding how NLRs evolve and how amenable these proteins are to changes will impact plant breeding.

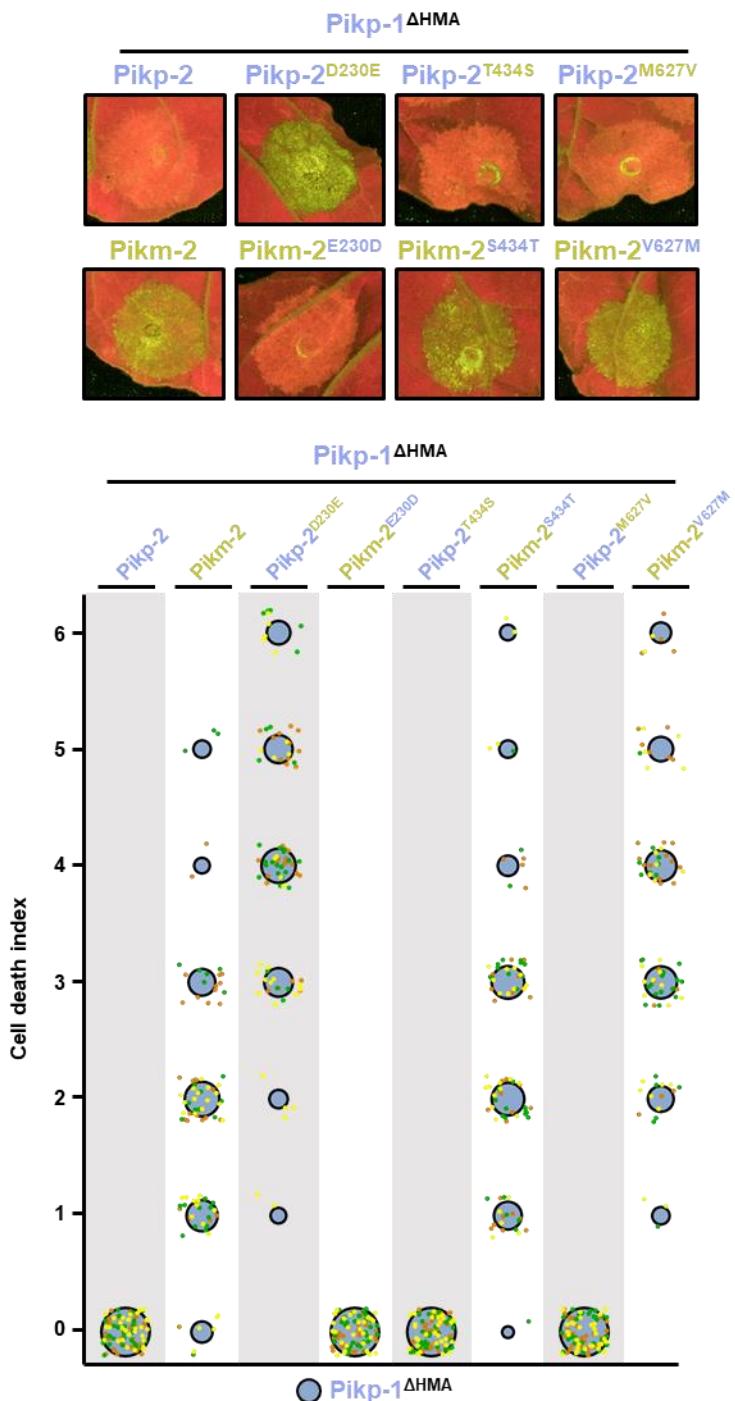
## 6.2 **Pik-HMA is not required for cell death signalling**

Although there is a good understanding of the molecular and structural mechanisms of effector recognition by *Pik*-HMA domains (De la Concepcion et al., 2019; De la Concepcion et al., 2018; Maqbool et al., 2015; Varden et al., 2019), how this domain translates recognition to response is unknown.

The discovery of Pik NLR pair combinations and mutations that trigger constitutive cell death (**Figure 5.6, Figure 5.11**) represents a useful tool for the characterization of Pik mediated immune responses in the absence of the effector. I took advantage of these findings to investigate whether the integrated HMA domain in Pik-1 is required for Pik mediated cell death responses. For this, I deleted this domain from Pikp-1 and co-expressed this NLR in *N. benthamiana* with wild-type Pikp-2 and Pikm-2, as well as the point mutants described in the previous chapter, measuring cell death responses under UV light.

Interestingly, the combination of Pikp-1  $\Delta$ HMA with Pikm-2, but not Pikp-2, triggered constitutive cell death as for wild-type Pikp-1 (**Figure 6.1**). Furthermore, co-expression with Pik-2 mutants showed that this phenotype is also determined by the Glu residue in Pik-2 polymorphic position 230 (**Figure 6.1**).

These results indicate that the Pik-1 HMA domain is not required for the activation of cell death signalling by Pik-2. Another interesting observation is that only Pikm-2 (and mutants with same residue at polymorphic site 230) triggered cell death. This confirms that the Asp230Glu confers a different activation mechanism to Pikm-2 and this activation is not dependent on HMA domains.

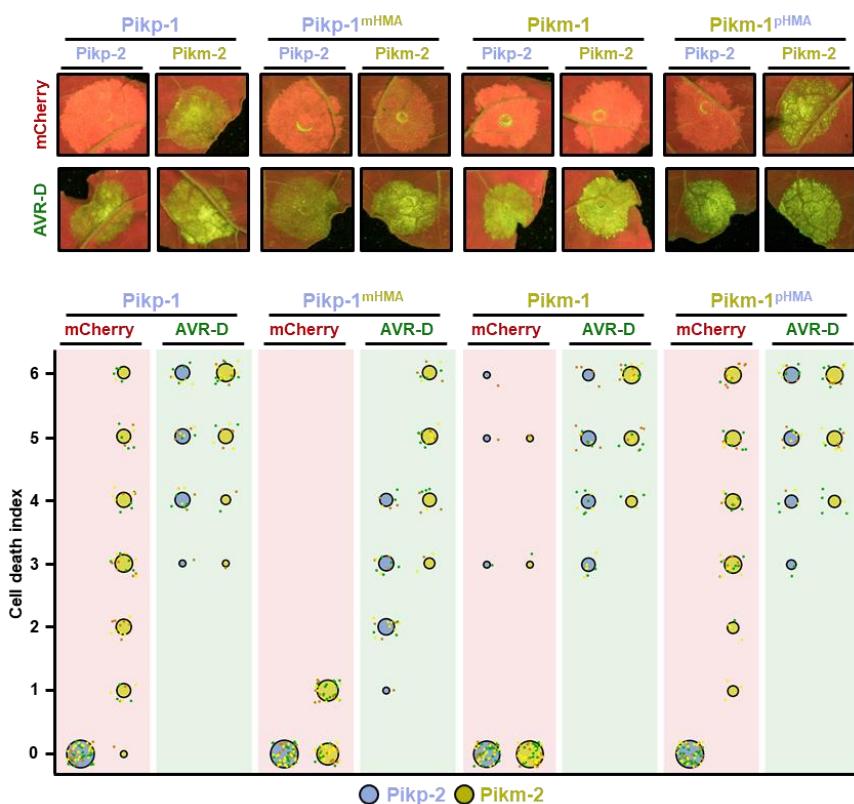


**Figure 6.1. Pik-HMA is not required for cell death signalling.** Representative leaf spot images and scoring of Pik mediated cell death as autofluorescence under UV-light. Cell death scoring is represented as dot plots comparing cell death response triggered by Pik-2 alleles and mutants in the presence of Pikp-1 $\Delta$ HMA in the absence of AVR-Pik effector. For every experiment the total number of repeats was 90. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates in the sample with that score.

### 6.3 Pikm-HMA acts as a suppressor of Pikm-2 constitutive cell death

The result described above suggests that domains in Pik-1 other than the HMA are responsible for the activation of the Pik-2 helper NLR. As I did not detect autoimmune phenotypes in the presence of Pikm-1, I hypothesised that the integrated HMA domain may be acting as a repressor of the cell death triggered by helper NLR Pikm-2.

To test this hypothesis, I performed cell death assays in *N. benthamiana* using HMA domain swap constructs between Pikp-1 and Pikm-1 made by Dr. Marina Franceschetti. I co-expressed these constructs with either Pikp-2 and Pikm-2, both in the presence and absence of AVR-Pik effectors (Figure 6.2).



**Figure 6.2. Pikm-HMA suppresses Pik cell death signalling in the absence of effector.** Representative leaf spot images and scoring of Pik mediated cell death as autofluorescence under UV-light. Cell death scoring is represented as dot plots comparing cell death triggered by Pikp-2 (ice blue) and Pikm-2 (gold). Pik-2 proteins were co-expressed with Pikp-1 harbouring the HMA domain of Pikm-1 and vice versa in the presence of mCherry (red panels) or AVR-PikD (green panels). Pikp-1 and Pikm-2 data from previous experiment (Figure 5.6) were included for comparison. The number of repeats was 60 and 30 for the spots co-infiltrated with mCherry and AVR-PikD, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score

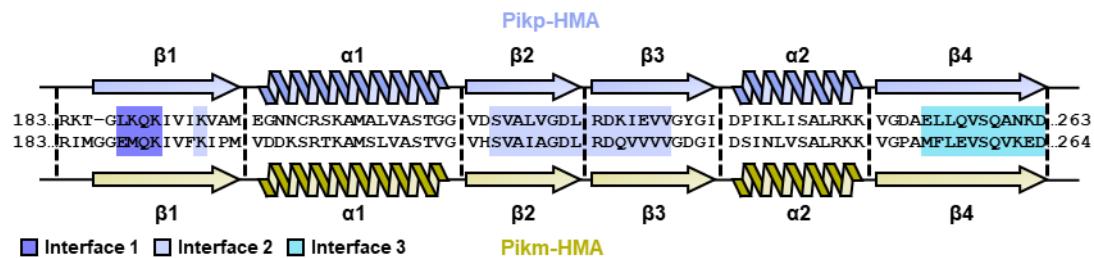
This assay revealed that introducing the Pikm-HMA domain in Pikp-1 suppresses the constitutive cell death triggered by Pikm-2 in the absence of the effector (**Figure 6.2**). This NLR also triggered cell death responses to AVR-PikD and, as in the previous chapter, these were consistently increased when co-infiltrated with Pikm-2 rather than Pikp-2 (**Figure 6.2**). By contrast, introducing the Pikp-HMA domain into Pikm-1 resulted in a gain of constitutive cell death response in the presence of Pikm-2 (**Figure 6.2**).

Together, these results suggest that whilst other domains in the Pik-1 NLR are responsible for NLR activation, Pikm-HMA actively represses Pikm-2 in the absence of effector. However, Pikp-HMA does not act as a suppressor of Pikm-2 activity and this may underpin the Pik allelic compatibility described in chapter 5.

## 6.4 Chimeric Pik-HMAs uncover the regions involved in cell death repression and allelic incompatibility

The fact that only Pikm-HMA, and not Pikp-HMA, represses the constitutive cell death triggered by Pikm-2 suggests that this HMA domain may have co-evolved with the Pikm-2 helper NLR. To gain insights into this co-adaptation, I used the crystal structure of Pikm-HMA (De la Concepcion et al., 2018) to design a suite of chimeric HMA domains with secondary structure elements of Pikm-HMA replaced by those from Pikp-HMA.

The Pik-HMA domain is comprised of a four-stranded antiparallel  $\beta$ -sheet and two  $\alpha$ -helices packed in an  $\alpha/\beta$  sandwich (De la Concepcion et al., 2018; Maqbool et al., 2015). The first  $\beta$ -strand is formed by the Pikm-HMA residues Gly187 to Met198. Although they are not present in the crystal structure, I also included the N-terminal linker residues Arg183 to Gly186 as they are polymorphic with Pikp-HMA (**Figure 6.3**). This is followed by the  $\alpha$ -helix encompassing Pikm-1 residues Val199 to Gly216.  $\beta$ -3 and  $\beta$ -4 together form the extensive effector binding interface two and are comprised by residues spanning Val217 to Leu226 and Arg227 to Ile237, respectively (**Figure 6.3**).  $\alpha$ -2 is the most conserved region between Pikp and Pikm HMA domains and encompasses amino acids Asp238 to Lys249. Finally, the C-terminal  $\beta$ -4 forms HMA/AVR-Pik interface three, which is the determinant for Pikm extended recognition specificity (De la Concepcion et al., 2018) and is comprised of residues Val250 to Asp264 (**Figure 6.3**).



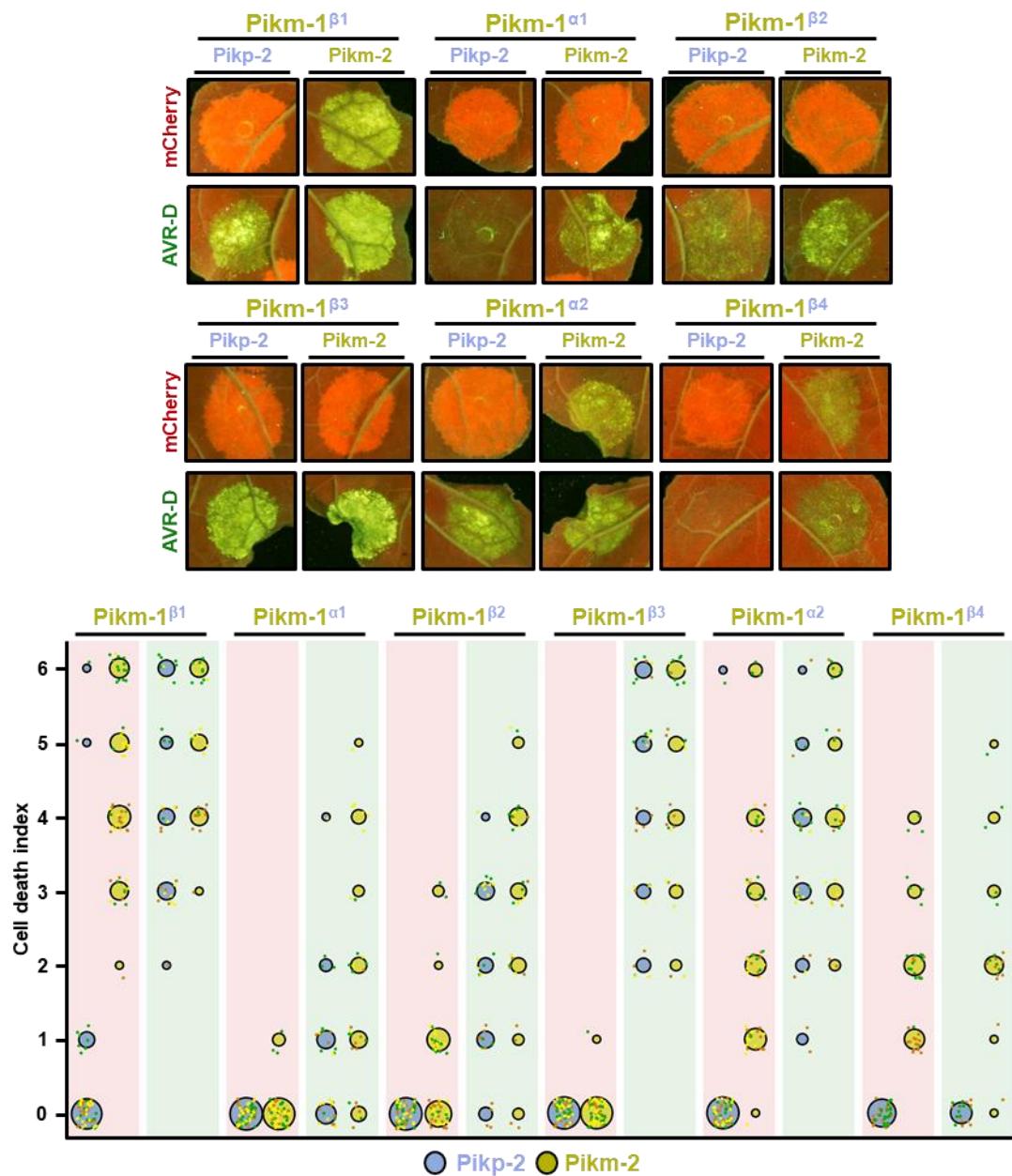
**Figure 6.3. Schematic representation of secondary structure domain swaps between Pikm-HMA and Pikp-HMA regions.** Sequence alignment of Pikp-1 and Pikm-1 HMA domains. Secondary structure features of the HMA fold are shown above and below. Residues located at binding interfaces are coloured in purple, ice blue or cyan for interfaces one, two or three, respectively. Secondary structure regions swapped from Pikp-HMA into Pikm-HMA are delimited by dashed lines.

To uncover the HMA secondary structure elements harbouring the determinants of Pikm-2 suppression and allelic incompatibility, I introduced each of the chimeric HMAs in the full-length Pikm-1 NLR. After co-expression with either Pikp-2 or Pikm-2, in the presence or absence of AVR-Pik effectors, I scored cell death responses in *N. benthamiana* (Figure 6.4). This assay revealed that replacing either Pikm-HMA β-1, α-2 or β-4 for the structurally equivalent regions in Pikp-HMA leads to loss of Pikm-2 repression and constitutive cell death in the absence of AVR-Pik (Figure 6.4).

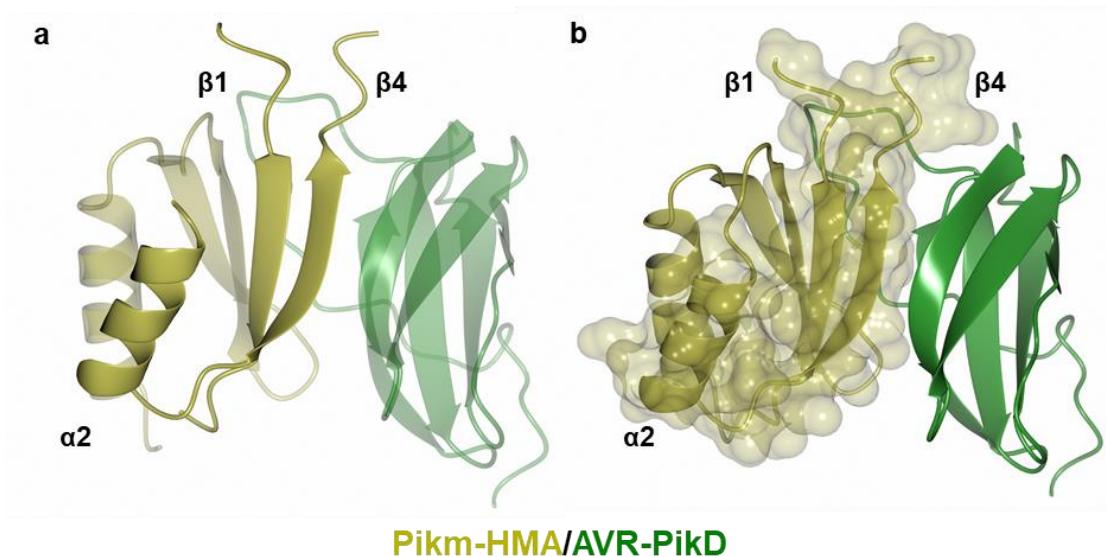
On a side note, replacing Pikm-HMA β-4 strand containing interface three for Pikp-HMA, largely reduced recognition of the AVR-PikD effector. By contrast, introducing Pikp-HMA β-3 containing interface two led to an increase of cell death responses to AVR-PikD (Figure 6.4). This further confirms the results described in chapter three and chapter four related to how Pikp and Pikm HMAs have different specialized interfaces for binding AVR-Pik effectors. Western blots analyses to ensure Pik-1 chimeras are accumulating in the same levels are ongoing.

The Pikm-HMA β-1, α-2 or β-4 regions fold together in the protein structure (Figure 6.5), suggesting that they carry out a common action to repress Pikm activation in the absence of effector, possibly by participating in intramolecular interactions with other domains in Pik-1 or Pik-2.

In summary, these results shed light on the structural features underpinning Pikm-HMA repression of constitutive NLR responses and autoimmunity.



**Figure 6.4. Pikm-HMA secondary structure regions  $\beta 1$ ,  $\alpha 2$  and  $\beta 4$  are involved in allelic incompatibility.** Representative leaf spot images and scoring of Pik mediated cell death as autofluorescence under UV-light. Cell death scoring is represented as dot plots comparing cell death triggered by Pikp-2 (ice blue) and Pikm-2 (gold). Pik-2 proteins were co-expressed with Pikm-1 harbouring chimeric HMA domains swapping secondary structure element of Pikp-HMA into Pikm-HMA in the presence of mCherry (red panel) or AVR-PikD (green panel). The number of repeats was 60 and 30 for the spots co-infiltrated with mCherry and AVR-PikD, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score. For Pikm-1  $\beta 4$  chimera, only two biological replicates were collected.

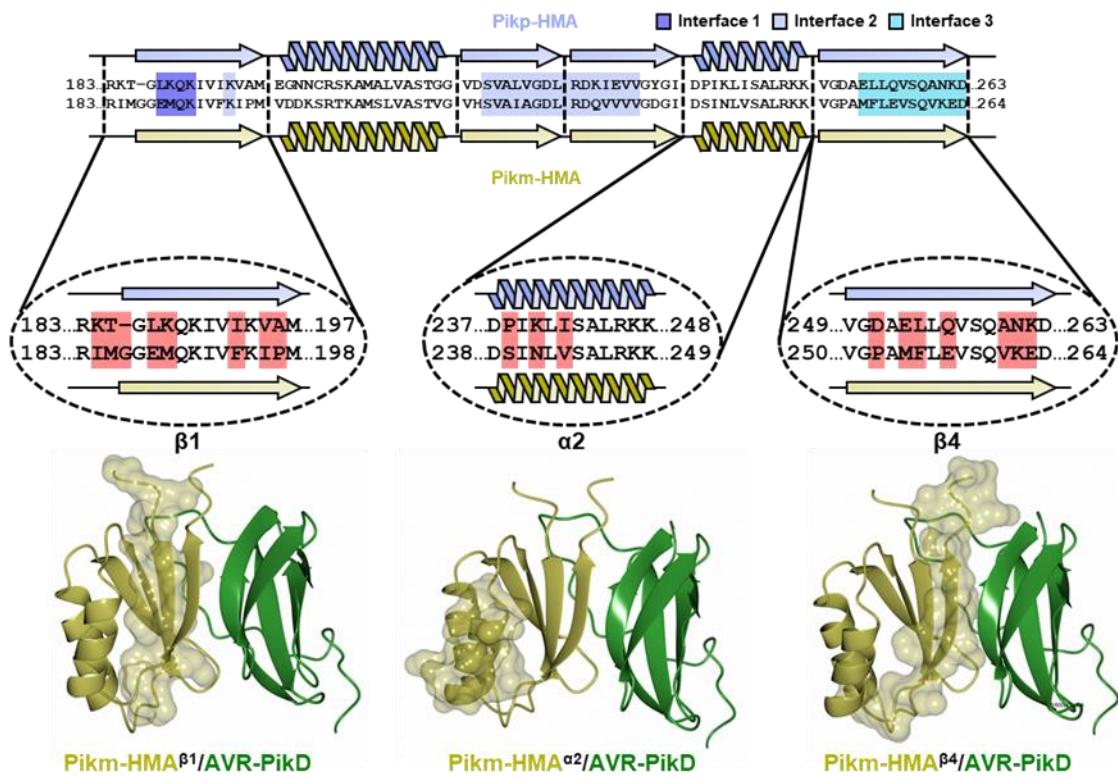


**Figure 6.5. Secondary regions involved in allelic incompatibility are situated together in the Pik-HMA structure.** Schematic representations of the structures of Pikm-HMA in complex with AVR-PikD effector. Pikm-HMA and AVR-PikD are shown in gold and green cartoon representation, respectively. **(a)** Pikm-HMA Secondary regions  $\beta$ 1,  $\alpha$ 2 and  $\beta$ 4 are highlighted. **(b)** Representation with the molecular surface of these regions also shown.

## 6.5 Two amino acid polymorphisms in Pik-HMA underpin de-repression of constitutive cell death and allelic incompatibility

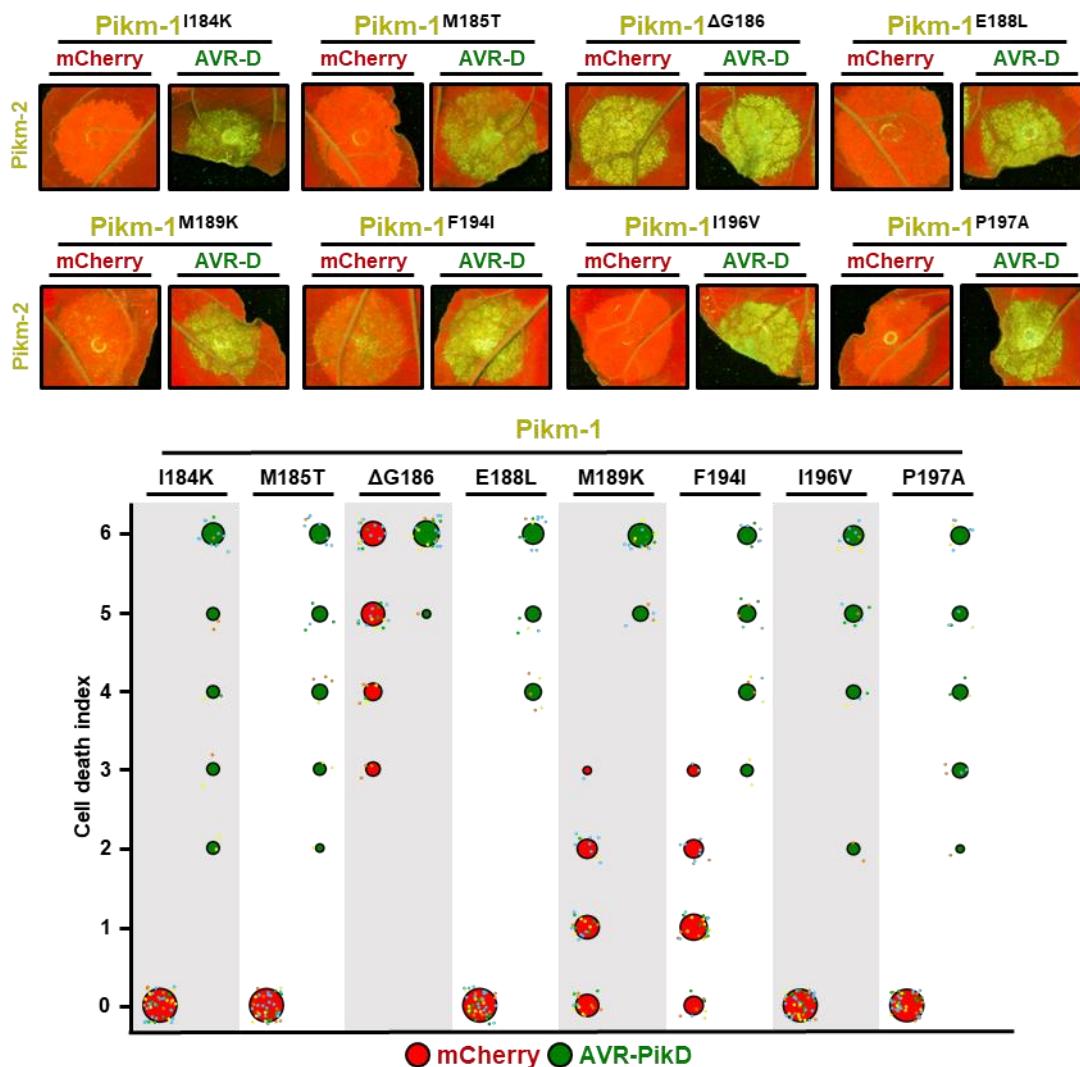
Having identified the secondary structure elements involved in suppression of Pikm-2 constitutive activation and allelic compatibility, I aimed to narrow down the amino acid residues underpinning these phenotypes. For this, I used site directed mutagenesis to scan the Pikm-HMA  $\beta$ -1,  $\alpha$ -2 or  $\beta$ -4 regions by replacing Pikm-HMA polymorphic residues with those in Pikp-HMA.

The Pik-HMA  $\beta$ -1,  $\alpha$ -2 and  $\beta$ -4 regions harbour eight, three and seven polymorphisms between Pikm and Pikp, respectively (Figure 6.6). In the Pik-HMA/AVR-PikD crystal structure, the first  $\beta$ -strand encompassing residues Arg183 to Met198 is located adjacent to the effector, with a small contribution to overall binding (Figure 6.6). This region contains three polymorphisms at the N-terminus that are not present in the structure (Ile184Lys, Met185Thr and a deletion of Gly186). Two other polymorphisms (Glu188Leu and Met189Lys) are situated at the binding interface one, whilst three polymorphisms (Phe194Ile, Ile196Val and Pro197Ala) are situated around the only residue in Pik-HMA  $\beta$ -1 contributing to the formation of the binding interface two with the AVR-Pik effector (Figure 6.6).



**Figure 6.6. Pik-HMA regions involved in allelic incompatibility are polymorphic.** Sequence alignment of Pikp-1 and Pikm-1 HMA domains. Secondary structure features of the HMA fold are shown above and below. Residues located at binding interfaces are coloured in purple, ice blue or cyan for interfaces one, two or three, respectively. Residues belonging to Pik-HMA regions  $\beta 1$ ,  $\alpha 2$  and  $\beta 4$  are zoomed in. Polymorphic residues at these regions are highlighted in red. Each region is accompanied by a schematic representation of the Pikm-HMA (gold) in complex with AVR-PikD effector (green) with the molecular surface of the given Pikm-HMA secondary structural element also shown.

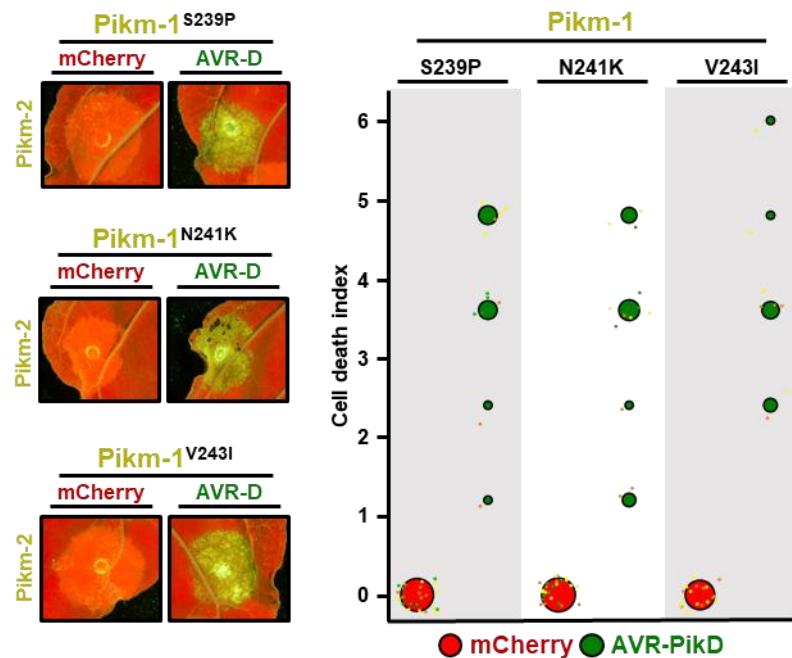
When I co-expressed these mutants with Pikm-2 and either mCherry or AVR-PikD, I observed that a single mutant Pikm-1  $\Delta$ Gly186 triggered strong cell death responses in the absence of AVR-Pik effector (Figure 6.7). Although in some cases Met189Lys and Phe194Ile mutations also displayed a slight constitutive cell death, the deletion of Gly186 residue was the only mutation triggering a robust autoimmune response (Figure 6.7).



**Figure 6.7. A single polymorphism is the major determinant of allelic compatibility mediated by *Pikm-HMA*  $\beta$ 1.** Representative leaf spot images and scoring of Pik mediated cell death as autofluorescence under UV-light. Cell death scoring is represented as dot plots comparing cell death triggered by *Pikm-2* in the presence of mCherry (red) or AVR-PikD (Green). *Pikm-2* protein was co-expressed with *Pikm-1* harbouring point mutations in the HMA domain swapping polymorphisms of *Pikp-HMA* into *Pikm-HMA* across  $\beta$ 2 region. The number of repeats was 40 and 20 for the spots co-infiltrated with mCherry and AVR-PikD, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the four biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

*Pik-HMA*  $\alpha$ -2 region includes residues Asp238 to Lys249 and is located distal from the binding interface with the AVR-Pik effector in the crystal structure (Figure 6.6). This region is the most similar between *Pikm-HMA* and *Pikp-HMA* with only three polymorphisms (Ser239Pro, Asn241Lys and Val243Ile) (Figure 6.6).

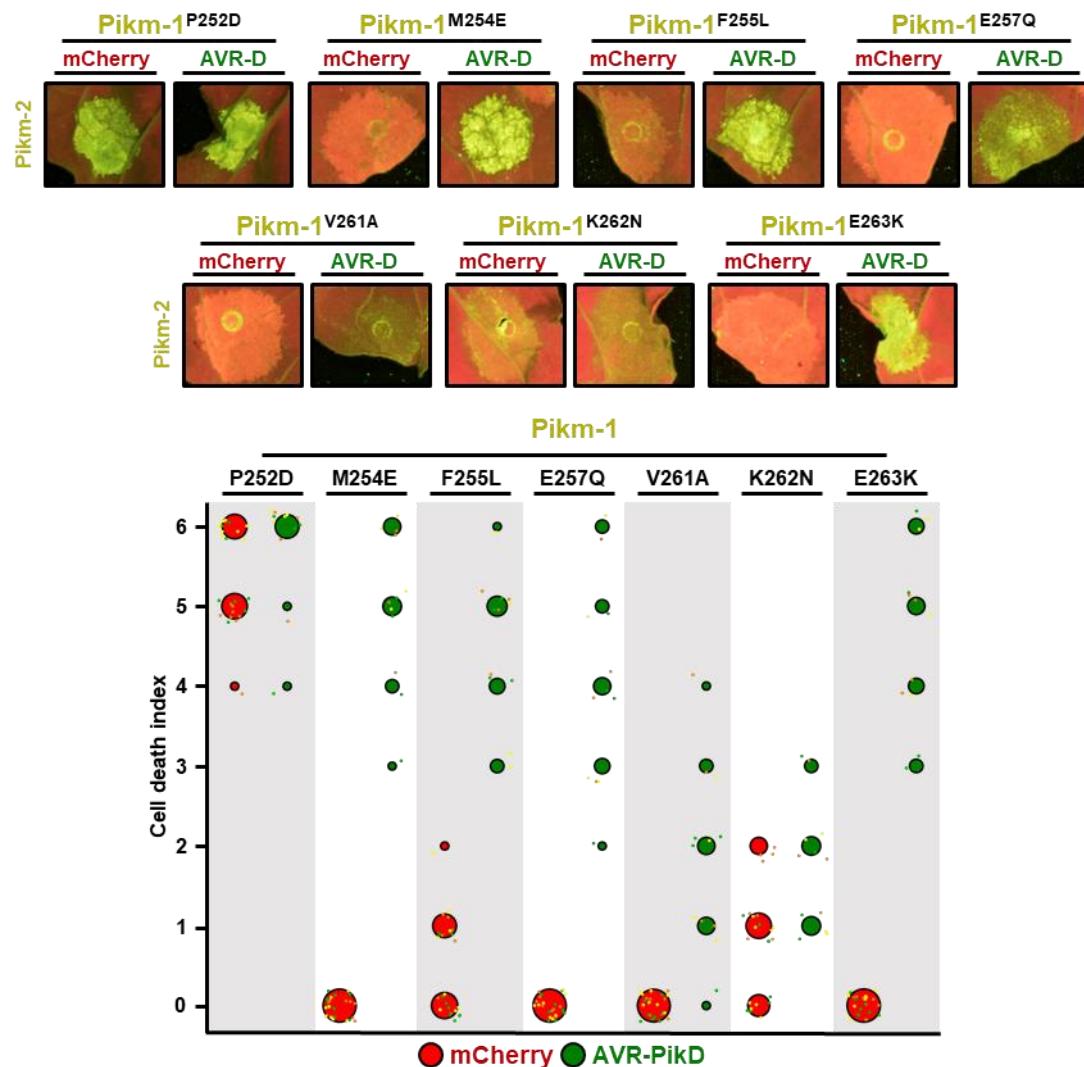
Surprisingly, none of the single mutations of these polymorphisms were sufficient to trigger constitutive cell death responses in the presence of *Pikm-2* (Figure 6.8). This may indicate that this phenotype is underpinned by a combination of the residues in this region, possibly with certain degree of redundancy.



**Figure 6.8. Allelic incompatibility mediated by *Pikm*-HMA  $\alpha 2$  is not determined by a single residue.**  
 Representative leaf spot images and scoring of *Pik* mediated cell death as autofluorescence under UV-light. Cell death scoring is represented as dot plots comparing cell death triggered by *Pikm-2* in the presence of mCherry (red) or AVR-PikD (Green). *Pikm-2* protein was co-expressed with *Pikm-1* harbouring point mutations in the HMA domain swapping polymorphisms of *Pikp*-HMA into *Pikm*-HMA across  $\alpha 2$  region. The number of repeats was 24 and 12 for the spots co-infiltrated with mCherry and AVR-PikD, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

The C-terminal  $\beta$ -strand,  $\beta$ -4, is situated adjacent to the effector in the crystal structure, making intimate contacts and forming the extensive binding interface three that confers extended recognition specificity to *Pikm* (Figure 6.6) (De la Concepcion et al., 2019; De la Concepcion et al., 2018). Most polymorphisms in this region are in the binding interface (Met254Glu, Phe255Leu, Glu257Gln, Val261Ala, Lys262Asn and Glu263Lys) (Figure 6.6), and previously, polymorphisms 262 and 263 were shown to be determinants in extended recognition responses to AVR-Pik effectors (De la Concepcion et al., 2019; De la Concepcion et al., 2018). Only one polymorphism Pro252Asp is located outside the binding interface (Figure 6.6).

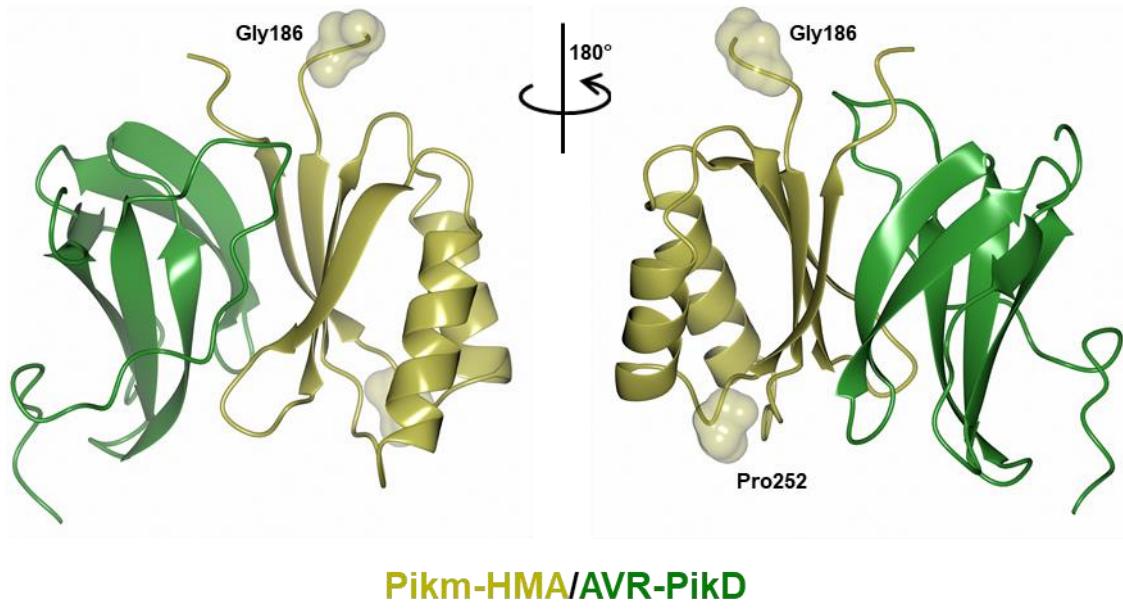
As for  $\beta$ -1, cell death assays using mutants in the *Pikm-HMA*  $\beta$ -4 revealed a major contribution from a single polymorphism, Pro252Asp, towards repression of cell death and allelic incompatibility (Figure 6.9). Interestingly, this polymorphic residue is the only one situated outside of the effector binding interface.



**Figure 6.9. A single polymorphism is the major determinant of allelic compatibility mediated by *Pikm-HMA*  $\beta$ 4.** Representative leaf spot images and scoring of *Pik* mediated cell death as autofluorescence under UV-light. Cell death scoring is represented as dot plots comparing cell death triggered by *Pikm-2* in the presence of *mCherry* (red) or *AVR-PikD* (Green). *Pikm-2* protein was co-expressed with *Pikm-1* harbouring point mutations in the HMA domain swapping polymorphisms of *Pikp-HMA* into *Pikm-HMA* across  $\beta$ 4 region. The number of repeats was 24 and 12 for the spots co-infiltrated with *mCherry* and *AVR-PikD*, respectively. For each sample, all the data points are represented as dots with a distinct colour for each of the three biological replicates; these dots are jittered about the cell death score for visualisation purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score.

In agreement with the results from previous chapters, the Lys262Asn mutation (opposite to the mutation introduced to extend binding and recognition to AVR-Pik effectors (De la Concepcion et al., 2019)) largely reduced the recognition of AVR-PikD (**Figure 6.9**), confirming again the important role of Lys262 in Pikm-HMA binding to AVR-Pik effectors. The Glu263Lys mutation in the second residue modified in the engineered Pik receptor did not affect AVR-PikD recognition (**Figure 6.9**). Interestingly, the Val261Ala mutation in the position preceding Lys262 significantly reduced the cell death response to AVR-PikD (**Figure 6.9**), indicating an important role of Val261 in responses to AVR-Pik effectors that has, to date, remained overlooked. This position was not modified in the engineering of extended recognition to AVR-Pik effectors (De la Concepcion et al., 2019). Therefore, this finding will inform further engineering efforts to expand Pik recognition specificities.

In summary, the mutation scan of the Pik-HMA secondary structure regions uncovered two main polymorphisms underpinning Pikm-2 constitutive cell death suppression and allelic incompatibility. These two residues, Gly186 and Pro252, are situated away from the binding interface with the AVR-Pik effector (**Figure 6.10**), pointing to HMA interactions other than with the effector (possibly with other domains in Pikm-1 or Pikm-2) as determinants for Pikm-2 suppression and allelic compatibility.



**Figure 6.10. Residues involved in allelic incompatibility are situated away from the interaction interface with AVR-Pik effectors.** Schematic representations of the structures of Pikm-HMA in complex with AVR-PikD effector. Pikm-HMA and AVR-PikD are shown in gold and green cartoon representation, respectively. Single residues underpinning allelic incompatibility are labelled and represented with the molecular surface.

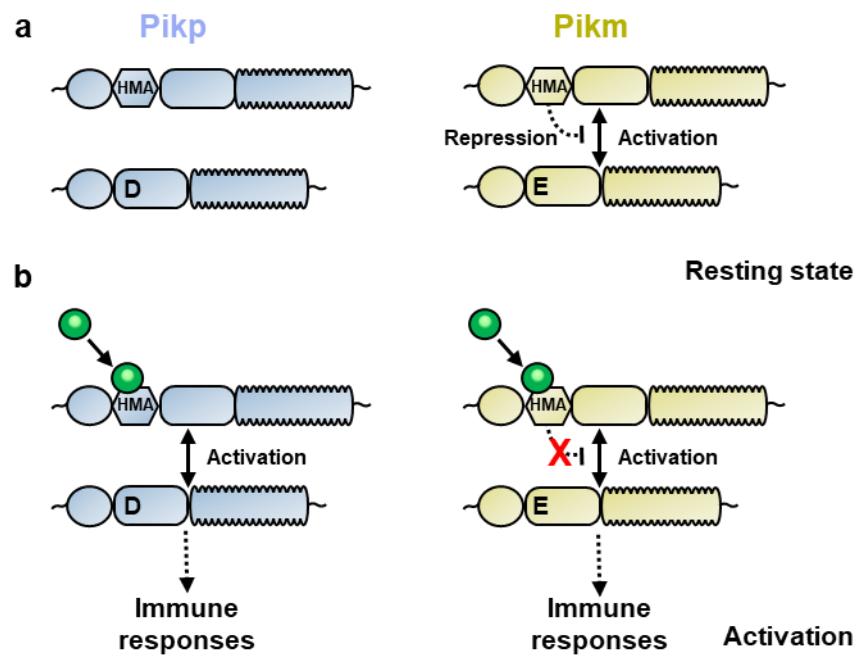
## 6.6 Discussion

The results presented in the previous chapter showed that the Pik-2 helper NLR has co-evolved with the sensor NLR Pik-1, most likely following the rapid evolution of their cognate HMA domain. A single amino acid polymorphism in Pikm-2 conferred increased response to AVR-Pik effectors, but also underpinned allelic incompatibility when mismatched with the maladapted sensor NLR allele Pikp-1.

In this chapter, I took advantage of mutants and Pik NLR combinations that trigger a constitutive immune response to uncover a regulatory role for Pikm-HMA. The results obtained suggest two distinct activation mechanisms in the Pik NLR alleles Pikp and Pikm, likely evolved through the differential co-evolution between these allelic NLR pairs.

Both Pikp and Pikm NLR pairs form a pre-activation complex in the absence of effector (**Figure 6.11a**). However, contrary to Pikp, Pikm complex is constitutively active, which is underpinned by the single amino acid polymorphism, Asp230Glu, in the NB-ARC domain of Pikm-2. The Pikm-HMA domain in Pikm-1 act as a repressor of this constitutive activation (**Figure 6.11a**). Binding of AVR-Pik effectors to the HMA domain in Pikp activates the NLR, possibly triggering a rearrangement of the association within the Pik complex, dependent on ATP/ADP usage. By contrast, binding of AVR-Pik effectors to Pikm-HMA releases repression of the Pikm complex, leading to cell death responses (**Figure 6.11b**).

These data exemplify how small modifications in allelic NLRs can lead to mechanistic diversification of the immune responses, a finding that may hold true for other NLR pairs. For example, in *Arabidopsis* RRS1/RPS4, binding of bacterial pathogen effectors AVR-Rps4 and PopP2 to RRS1 WRKY domain de-represses RPS4 constitutive activation (Le Roux et al., 2015; Sarris et al., 2016). Interestingly, some mutations in RPS4 and RRS1 have been found to compromise PopP2 but not AvrRps4 recognition (Ma et al., 2018), which may suggest that recognition of these effectors results in subtly different mechanisms of activation.



**Figure 6.11. Schematic representations of the differential regulation mechanisms proposed for Pikp and Pikm alleles.** Pik NLRs cooperate to activate immune signalling. **(a)** Resting state. In the absence of effectors, the Pikp NLR pair (coloured in ice blue) associates in a resting complex. In Pikm (coloured in gold), helper NLR Pikm-2 is constitutively activated by Pik-1 domains. The Glu230 residue in Pikm-2 largely underpins activation. Pikm-HMA represses Pikm-2 to maintain an inactive state. **(b)** Activation. The presence of AVR-Pik effector is detected through direct binding to the Pik-1 HMA domain. Upon recognition, Pikp-1 and Pikp-2 cooperate to activate immune responses. In Pikm, binding of the HMA domain to AVR-Pik effectors releases Pikm-2 suppression.

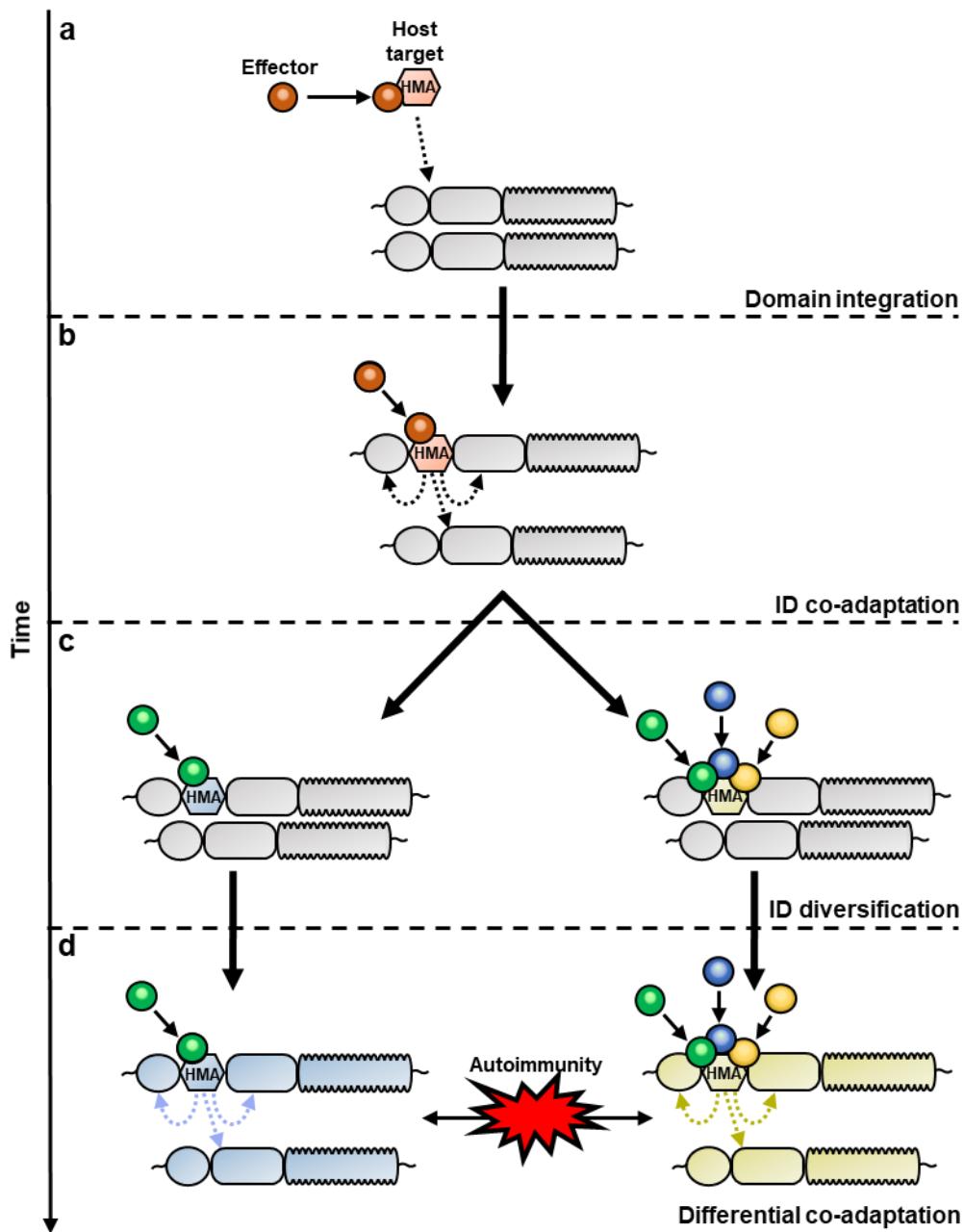
The differential role in immune regulation between Pikp-HMA and Pikm-HMA domains is an interesting finding that develops understanding of the activation mechanisms by AVR-Pik effectors. The data presented here show that several residues are involved in repression of Pikm constitutive responses. Interestingly, these are located away from the binding interface with the AVR-Pik effector, suggesting that the Pikm-HMA domain could associate with other Pik NLR domains to keep the Pikm complex inactive. This finding confirms that effector recognition and autoimmune responses are not necessarily linked, which is supported by similar findings in other integrated domains (Newman et al., 2019).

Furthermore, the polymorphisms I found to have a major effect on Pikm allelic compatibility are introducing major steric changes. This suggests that effector binding to Pikm-HMA produces structural changes that lead to the release of repression. In autoimmune combinations and mutants, polymorphisms in the Pik-1 HMA domain may mimic these changes, leading to constitutive NLR activation. Further experiments will unravel the biochemical basis of Pikm-HMA regulation of immune responses.

The results obtained here also demonstrate that integrated domains must co-adapt with other domains within the sensor and helper NLRs. Diversifying evolution of integrated domains can therefore drive differential co-adaptation between NLR alleles, leading to differential activation and regulation of immune responses between allelic NLRs. This may represent the basis of many autoimmune phenotypes reported for several NLR combinations. As the process leads to reproductive isolation between plant populations, and can drive plant speciation, these findings have major implication beyond plant immunity.

An evolutionary model suggested by these findings hypothesises that the HMA domain of a host protein targeted by a pathogen effector was initially fused into the ancestor of the *Pik* NLR pair (**Figure 6.12a**). To ensure an optimal balance between effector recognition and regulation of immune responses, the newly integrated HMA domain had to co-adapt with other domains in the sensor and helper NLRs (**Figure 6.12b**). *Pikp* and *Pikm* alleles diversified, driven by a differential co-evolution of their HMA domains with a different subset of AVR-*Pik* effectors. Different changes in the sensor NLR *Pik-1* HMA domains and helper NLR *Pik-2* conferred a distinct effector recognition specificity and modified the activation mechanisms of the *Pikp* and *Pikm* NLR pairs (**Figure 6.12c**). *Pikp* and *Pikm* NLRs maintained different co-evolutionary pathways with their cognate integrated HMA. *Pikm*-HMA gradually gained a pivotal role in limiting constitutive activation of the *Pikm* complex in the absence of pathogen effectors. Consistent with this, appropriate pairing is required for inhibition and *Pikm-1* HMA domain deletion or exchange with the independently evolved *Pikp*-HMA results in autoactivity (**Figure 6.12d**). Overall, this model provides a mechanistic understanding of how allelic plant NLRs that cooperate in immune activation, can become incompatible, potentially leading to hybrid necrosis between plant varieties.

The discovery of NLR integrated domains opened the possibility of engineering synthetic resistance. In previous chapters, I have shown that engineering integrated NLR domains to achieve extended effector recognition is achievable (De la Concepcion et al., 2019). However, as modifications that render autoactive NLRs represent a bottleneck for engineering efforts, the regulatory role of integrated domains in NLR activation needs to be fully understood. By scanning the mutational landscape of the *Pikm*-HMA domain using chimeras and point mutants, I have uncovered positions in this domain outside of the binding interface with pathogen effectors that need to remain unaltered to ensure a proper regulation of NLR responses. These will inform future engineering efforts to use the *Pik*-HMA domain as a chassis to achieve increased disease resistance.

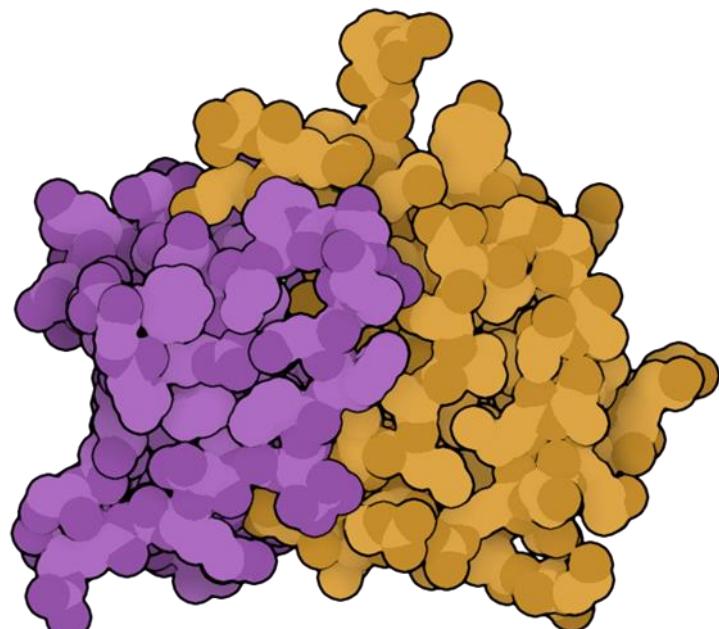


**Figure 6.12. Schematic representations of the events proposed during adaptive evolution of the Pik NLR pair.** (a) Domain integration. An unconventional HMA domain, possibly derived from pathogen host targets, gets integrated into the architecture of the ancestor of Pik-1 NLR. (b) Integrated domain (ID) co-adaptation. Following integration, the HMA and other Pik NLR domains undergo adaptive changes to fine-tune activation and regulation of immune responses. (c) ID diversification. Integrated domains rapidly co-evolved with pathogen effector and diversify to mediate recognition to different effector subsets. (d) Differential co-adaptation. Following the rapid diversification, integrated domains and other Pik NLR domains co-adapt giving raise to different mechanisms and regulation of immune responses. NLR mismatches between diversified NLRs trigger autoimmune related phenotypes.

## Chapter 6

# 7

## **The structure of rice blast effector AVR-Pii bound to host Exo70 uncovers a new fungal effector fold**



AVR-PiKE/Pikp-HMA<sup>NK-KE</sup>  
6R8M



## 7. The structure of rice blast effector AVR-Pii bound to host Exo70 uncovers a new fungal effector fold

### Abstract

The cloning and functional characterization of effectors from the rice blast fungus has provided invaluable insights into the mechanisms of pathogenesis and how the plant immune system recognizes and respond to pathogens. A decade ago, together with AVR-Pik and AVR-Pia, an important *M. oryzae* effector named AVR-Pii was identified. This effector binds rice exocyst components OsExo70F-2 and OsExo70F-3. Association with OsExo70F-3 is detected by the immune receptor pair Pii, possibly via an unconventional RIN4-NOI domain integrated in the NLR. However, the knowledge of the molecular basis of AVR-Pii function and recognition by the plant immune system was still very limited. Here, I dissected the molecular basis of AVR-Pii interaction with rice Exo70 proteins. I determined the crystal structure of AVR-Pii in complex with rice OsExo70F-2, which defined a new effector/target binding interface. I used structure-informed mutations to gain a better understanding of how the AVR-Pii/Exo70 interface underpins protein function. Furthermore, the crystal structure of AVR-Pii revealed a new type of protein fold in fungal effectors, based on a Zinc-finger motif, coined Zinc-finger Fold (ZiF). As the exocytosis complex is increasingly recognized as having an important role in plant-microbe interactions, both in pathogenesis and symbiosis, the data obtained in this chapter can open new research avenues in plant-microbe interactions. Additionally, as Exo70 protein domains have been found as integrations in NLRs, this data has the exciting potential of contributing to engineering disease resistance.

## Chapter 7

## 7.1 Introduction

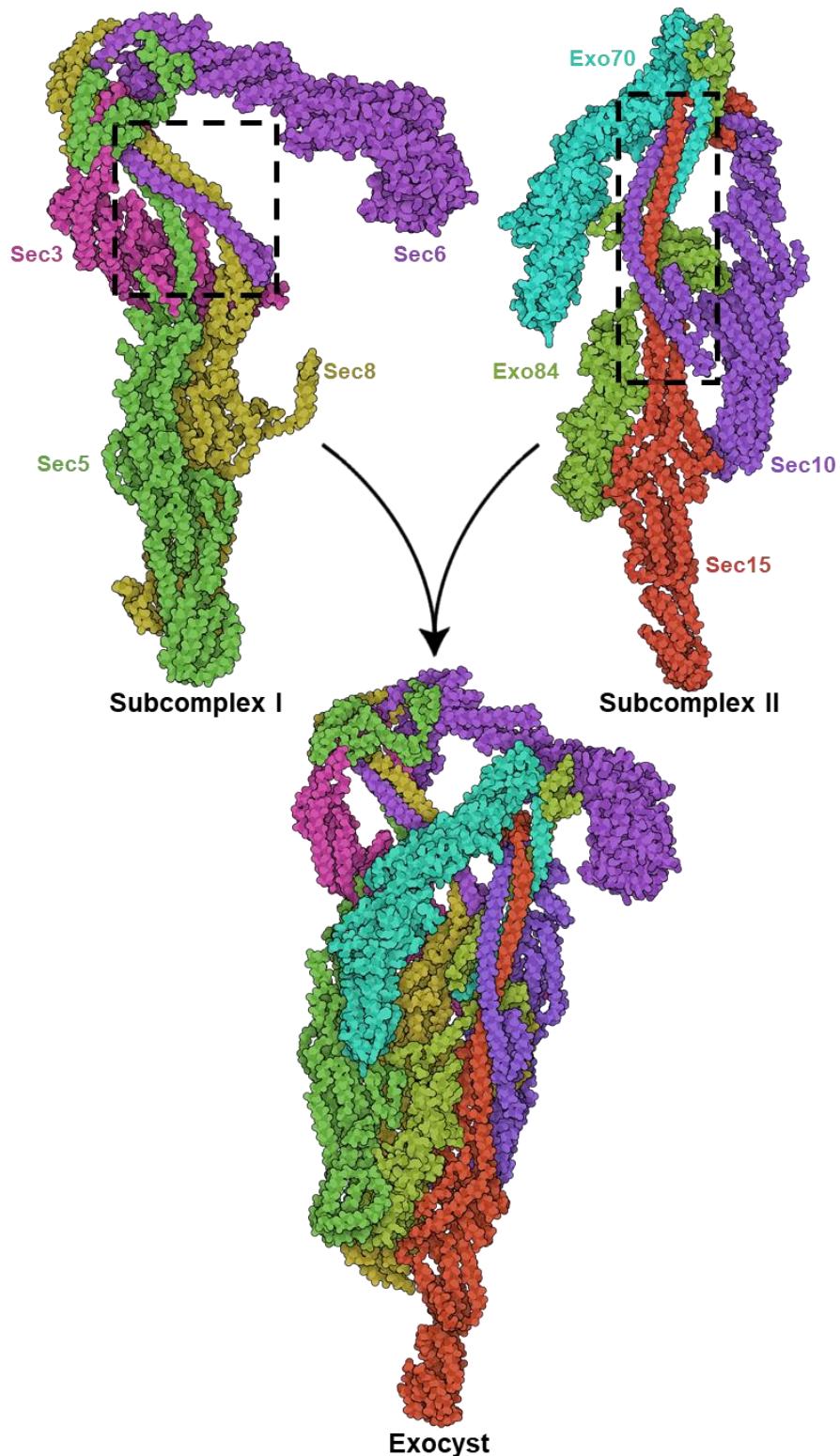
The rice blast fungus secretes effectors to successfully colonize plants. Genome sequencing of this pathogen allowed identification of many putative effectors (Dean et al., 2005). However, only a few of these proteins have been functionally characterized to date.

In 2009 the rice blast fungal effectors AVR-Pik, AVR-Pia and AVR-Pii were cloned using association genetics (Yoshida et al., 2009). Since then, research on AVR-Pik and AVR-Pia has driven fundamental discoveries in the field of effector biology and plant immunity (Bialas et al., 2018). For example, these two effectors were the founding members of a new fungal effector family very prevalent in *M. oryzae* (de Guillen et al., 2015) and have been critical in unravelling the function of unconventional domains integrated in plant NLRs (Cesari et al., 2014a; Maqbool et al., 2015). Characterisation of the interaction of these effectors with their cognate receptors has provided unprecedented information on effector recognition by plant NLRs (De la Concepcion et al., 2018; Guo et al., 2018b; Maqbool et al., 2015; Ortiz et al., 2017; Varden et al., 2019) and the arms-race co-evolution between plants and pathogens (De la Concepcion et al., 2018; Kanzaki et al., 2012), paving the way towards engineering disease resistance (De la Concepcion et al., 2019).

In comparison, the *M. oryzae* AVR-Pii effector has remained under-investigated. With only 51 residues, AVR-Pii is substantially smaller than AVR-Pik or AVR-Pia, and is not predicted to be a member of the MAX effector family (de Guillen et al., 2015). Interestingly, AVR-Pii has been shown to specifically associate with protein components of the rice exocyst complex (Fujisaki et al., 2015). The exocyst is an evolutionary conserved complex that mediates tethering of post-Golgi vesicles to the plasma membrane for secretion (Wu and Guo, 2015). This complex orchestrates cargo delivery and controls the spatiotemporal regulation of exocytosis; it is essential for vesicle secretion of different cargoes (Heider and Munson, 2012; Saeed et al., 2019; Žárský et al., 2013).

The oligomeric nature of the exocyst complex was first identified in yeast (TerBush et al., 1996) where it was shown to be an octamer formed by the proteins Sec3, Sec5, Sec6, Sec8, Sec10, Sec15, Exo70, and Exo84 (TerBush et al., 1996). To perform their function in secretion, these subunits hierarchically assemble in a holocomplex (Heider et al., 2015; Picco et al., 2017) (**Figure 7.1**). Two subcomplexes, subcomplex I (containing Sec3, Sec5, Sec6, and Sec8) and subcomplex II (formed by Sec10, Sec15, Exo70, and

Exo84), are formed by the association of each subunit via a four-helix bundle in the core exocyst assembly region (CorEx) (Mei and Guo, 2018; Mei et al., 2018) (**Figure 7.1**). Both subcomplexes then dynamically merge to assemble the holocomplex in the plasma membrane (**Figure 7.1**) (Ahmed et al., 2018).

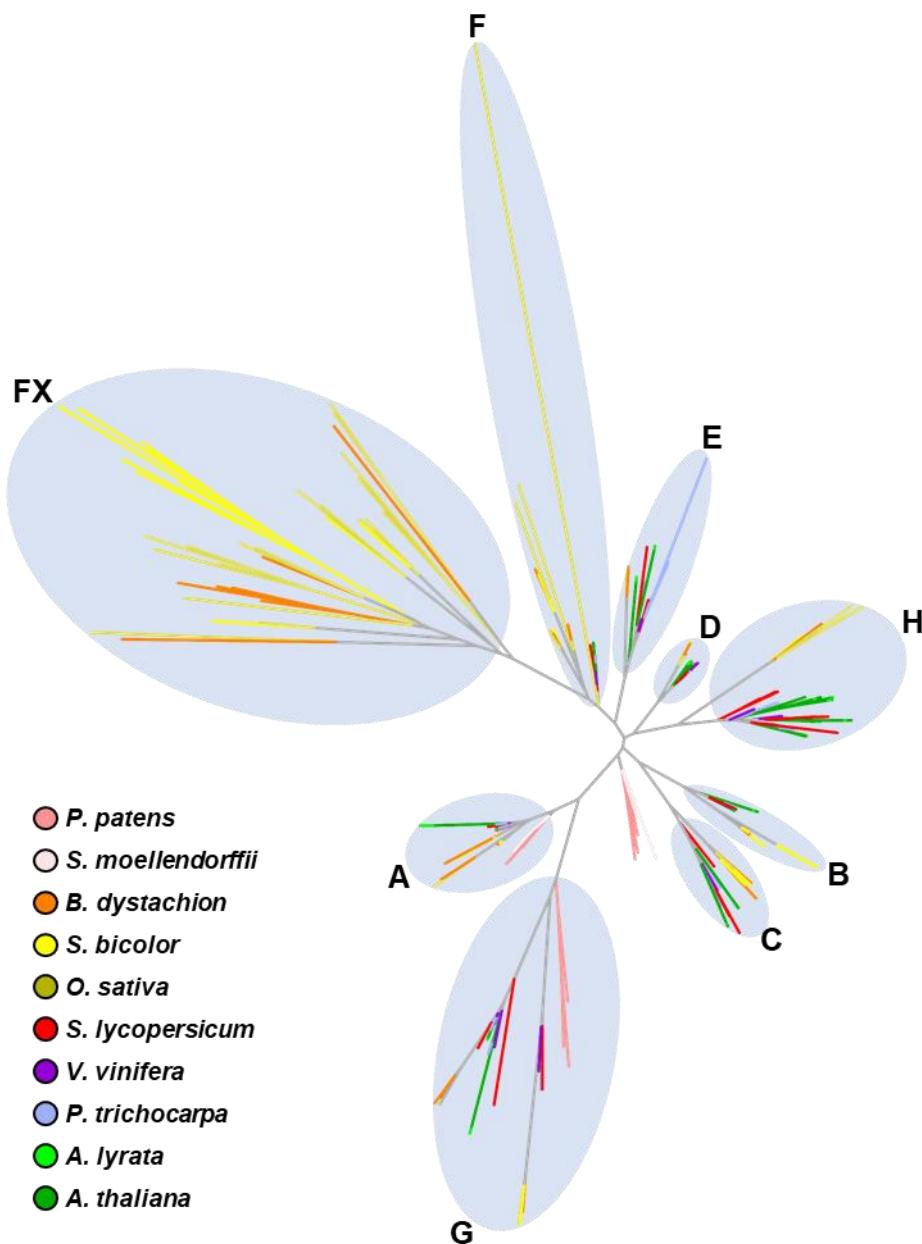


**Figure 7.1. The exocyst is a multimeric complex.** Schematic representation of the subunits and concerted assembly of the exocyst complex. Sec3, Sec5, Sec6 and Sec8 assemble into subcomplex I (**left**). In a similar fashion, Sec10, Sec15, Exo70 and Exo84 assemble into subcomplex II (**right**). Exocyst subunits assemble in subcomplexes through formation of the four-helix bundle in the core exocyst assembly region (CorEx) indicated with dashed black lines. This representation is based on the yeast exocyst Cryo-EM structure (5YFP) (Mei et al., 2018) and generated using Illustrate software for biomolecular illustration (Goodsell et al., 2019).

Exocyst subunits are conserved across all eukaryotes. However, while in yeast and mammals Exo70 subunits are encoded by a single gene, in plants the Exo70 subunits have undergone a dramatic expansion in diversity and copy number (Brabham et al., 2018; Cvrckova et al., 2012; Elias et al., 2003) (**Figure 7. 2**). Plant Exo70 proteins cluster in eight phylogenetically distinct groups (A-H) (**Figure 7. 2**). This expansion is particularly dramatic in grasses, with the emergence and diversification of a grass-specific subgroup named Exo70 FX (Brabham et al., 2018; Cvrckova et al., 2012) (**Figure 7.2**). Although they are very diverse in sequence, structural biology studies have revealed that Exo70 proteins from very divergent groups such as plants, yeast and mammals share a similar fold (Dong et al., 2005; Hamburger et al., 2006; Mei et al., 2018; Moore et al., 2007; Zhang et al., 2016). This suggests that the overall function of Exo70 protein in the exocyst, as an integral element of the secretory machinery of the cell, is conserved.

The large expansion of plant Exo70 proteins suggests that that plant Exo70 subunits may have functionally diversified and exocyst complexes equipped with specific Exo70 subunits may have adopted specialized functions (Saeed et al., 2019). Indeed, plant exocyst complexes harbouring different Exo70 subunits have been found to be involved in different processes such as root development (Kalmbach et al., 2017; Ogura et al., 2019), cell wall deposition (Kulich et al., 2015; Kulich et al., 2018) and symbiosis with arbuscular mycorrhiza (Zhang et al., 2015b). The exocyst has been also linked to unconventional secretory pathways (Saeed et al., 2019) and there is increasing evidence for the role of Exo70 subunits in trafficking processes distinct from exocytosis, such as autophagy (Bodemann et al., 2011; Farré and Subramani, 2011; Kulich et al., 2013; Lin et al., 2015; Teh et al., 2019).

Recently, an increasing number of studies have identified links between exocytosis and pathogenesis (Arasaki et al., 2018). In plants, several Exo70 subunits are related to plant resistance to pathogens and pests (Du et al., 2018; Guo et al., 2018a; Pečenková et al., 2011; Stegmann et al., 2013), particularly those belonging to the group F, which is expanded in grasses (Cvrckova et al., 2012; Ostertag et al., 2013).



**Figure 7.2. Exo70 proteins have diversified in plants.** Phylogenetic tree of amino acid sequences of plant Exo70 proteins based on alignments from Cvrckova et al., 2012. The phylogeny was built using FastTreeMP version 2.1.7 (Price et al., 2009) and the tree was represented using Interactive Tree Of Life (iTOL) v4 (Letunic and Bork, 2019).

In *Arabidopsis*, the subunit AtExo70B-1 has been proposed to be guarded by NLRs, leading to disease resistance when this subunit is targeted by pathogens (Liu et al., 2017; Wang et al., 2019d; Zhao et al., 2015). Importantly, some Exo70 subunits including AtExo70B-1 have been shown to interact with *Arabidopsis* RIN4 (Redditt et al., 2019; Sabol et al., 2017), a well-known regulator of plant immunity (Afzal et al., 2011; Afzal et al., 2013; Belkhadir et al., 2004). RIN4 plays a central role in plant immunity and is targeted by effectors from many different pathogens (Mukhtar et al., 2011; Weßling et

al., 2014), and, as a consequence, this protein is also monitored by different plant NLRs (Axtell and Staskawicz, 2003; Boyes et al., 1998; Mackey et al., 2002). Both Exo70 and RIN4 domains are consistently found as integrated domains in plant NLRs (Bailey et al., 2018; Brabham et al., 2018; Koj et al., 2016; Periyannan et al., 2013; Sarris et al., 2016; Steuernagel et al., 2018), which reflects the importance of these two proteins in plant immunity (Mukhtar et al., 2011; Weßling et al., 2014).

The rice blast effector AVR-Pii targets rice Exo70 subunits in a remarkably specific manner. Although the rice Exo70 family is comprised of 47 homologs (Cvrckova et al., 2012), AVR-Pii associates with only two members of the Exo70 group F, OsExo70F-2 and OsExo70F-3 (Fujisaki et al., 2015). As Exo70 proteins from phylogenetically distinct organisms share a common fold (Dong et al., 2005; Hamburger et al., 2006; Moore et al., 2007), AVR-Pii must exploit subtle differences to achieve this high interaction specificity. This suggests that AVR-Pii is targeting specific processes carried out by exocyst complexes harbouring OsExo70F-2/OsExo70F-3. How AVR-Pii specifically targets OsExo70F-2 and OsExo70F-3 and their function during rice blast pathogenesis is still unknown.

Like AVR-Pik and AVR-Pia, AVR-Pii is recognized by a paired NLR named Pii (Takagi et al., 2017; Takagi et al., 2013). Pii is an allelic version of the NLR pair Pi-5 described previously (Lee et al., 2009). Interestingly, this recognition requires OsExo70F-3 but not OsExo70F-2 (Fujisaki et al., 2015), suggesting that the association of AVR-Pii to OsExo70F-3 is monitored by the Pii NLR pair, leading to disease resistance (Fujisaki et al., 2015). A member of this NLR pair, Pii-2, has been recently found to harbour an unconventional RIN4-NOI motif (Afzal et al., 2013) integrated in its architecture (Fujisaki et al., 2017). It has been proposed that, instead of directly binding AVR-Pii, Pii-2 monitors OsExo70F-3 by association with its integrated NOI domain, which represents a new mechanism of pathogen detection by NLR integrated domains (Fujisaki et al., 2017).

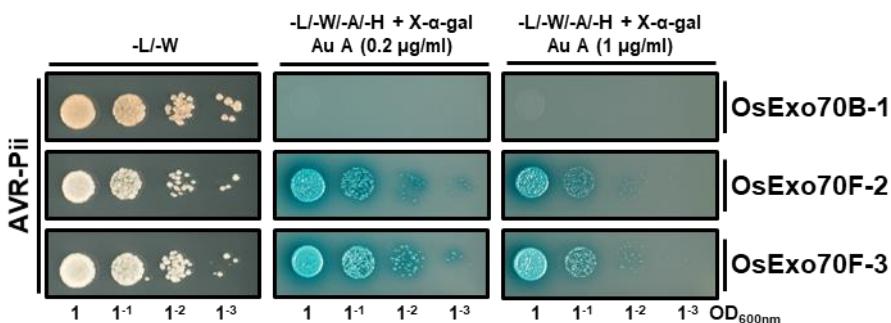
In this chapter, I investigate the structural and biophysical basis of AVR-Pii association to rice Exo70 proteins. I successfully produced and purified the *M. oryzae* effector AVR-Pii and diverse plant Exo70 proteins in *E. coli*, which allowed me to recapitulate the binding between effector and targets in vitro. Furthermore, I used X-ray crystallography to obtain the structure of AVR-Pii in complex with rice OsExo70F-2. With this, I described a new effector/target interface and I dissected the determinants of AVR-Pii binding to host targets using structure-informed mutations. Due to time constraints, some data presented in this chapter are still preliminary. However, they shed light on the structure/function relationship of AVR-Pii and the molecular basis of Pii-mediated

recognition. In addition, AVR-Pii represents a new type of fungal effector fold based on a zinc finger motif. The structural information gained in this study can be used to predict and describe new fungal effectors. As an increasing number of plant Exo70 proteins are shown to be involved in pathogenesis, and are targeted by pathogen effectors or found as integrated domains, the results obtained in this chapter may aid the development of new disease resistance strategies.

## 7.2 AVR-Pii specifically associates with OsExo70F-2 and OsExo70F-3 in vivo

Previous studies reported that *M. oryzae* AVR-Pii specifically associates with OsExo70F-2 and OsExo70F-3 in Y2H and plant co-IP (Fujisaki et al., 2015). As a starting point for this project, I set up a Y2H screening system to study AVR-Pii/Exo70 interaction at JIC. For this, I replicated the Y2H results from Fujisaki et al. 2015 with the same system as described for AVR-Pik/HMA interactions in previous chapters.

In agreement with previous research, there was significant yeast growth and the development of blue colouration with X- $\alpha$ -gal, both readouts of protein interactions, in the selective media for AVR-Pii and OsExo70F-2 or OsExo70F-3, but not with OsExo70B-1 (Figure 7.3). This confirmed that the AVR-Pii effector specifically interacts with these targets. There was clear growth also at higher concentrations of Aureobasidin A (Figure 7.3), suggesting that the association of AVR-Pii with the OsExo70 targets is robust.



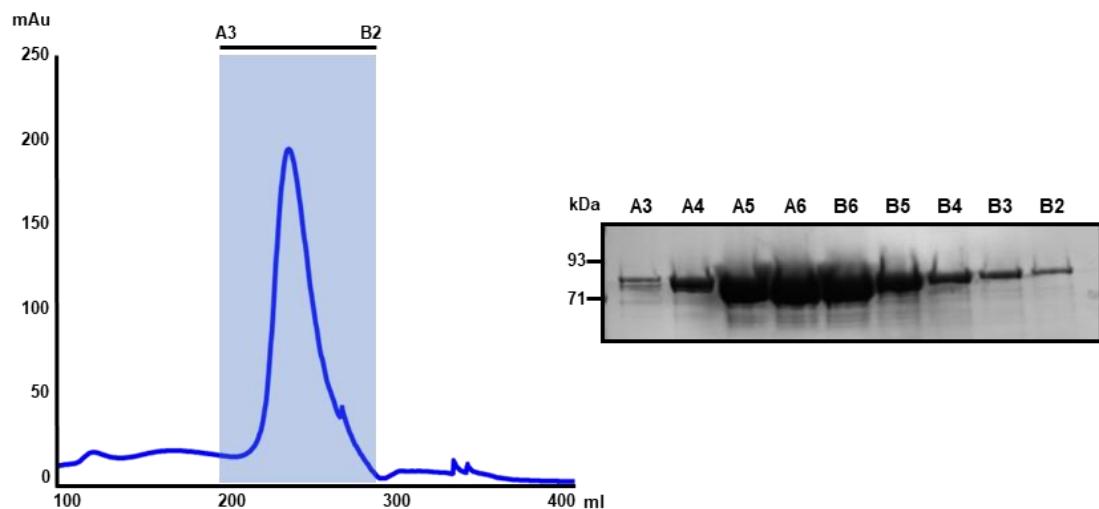
**Figure 7.3. AVR-Pii binds specifically to rice Exo70F-2 and Exo70F-3 in vivo.** Yeast-Two-Hybrid assay of AVR-Pii with rice Exo70 homologs Exo70B-1, Exo70F-2 and Exo70F-3. The control plate for yeast growth is on the left, with quadruple dropout media supplemented with X- $\alpha$ -gal and increasing concentrations of Aureobasidin A (Au A) on the right. Growth and development of blue colouration in the selection plate are both indicative of protein:protein interactions. Exo70 proteins were fused to the GAL4 DNA binding domain, and AVR-Pii to the GAL4 activator domain. Each experiment was repeated a minimum of three times, with similar results.

### 7.3 Plant Exo70 proteins can be successfully produced and purified in *E. coli*

Exocyst subunits, including Exo70, have been challenging to produce and purify for structural studies as they are generally poorly folded, aggregation prone, and refractory to crystallization (Lepore et al., 2018). After extensive research of the genetics and cell biology of exocytosis (Heider and Munson, 2012; Saeed et al., 2019; Wu and Guo, 2015; Žárský et al., 2013), there remain just a few structural and biophysical studies involving these proteins (Dong et al., 2005; Hamburger et al., 2006; Huang et al., 2019; Moore et al., 2007; Zhang et al., 2016). The recent cryo-EM model of the yeast exocyst revealed an intimate interaction between the four subunits of each exocyst subcomplex through an  $\alpha$ -helical CorEx motif (Mei and Guo, 2018; Mei et al., 2018) (Figure 7.1), partly explaining why the production and purification of isolated exocyst subunits remained challenging to date (Lepore et al., 2018). Consequently, the Exo70 proteins successfully expressed and purified had truncations in the first  $\alpha$ -helix, which contributes to the interactions in the CorEx (Dong et al., 2005; Hamburger et al., 2006; Moore et al., 2007; Zhang et al., 2016).

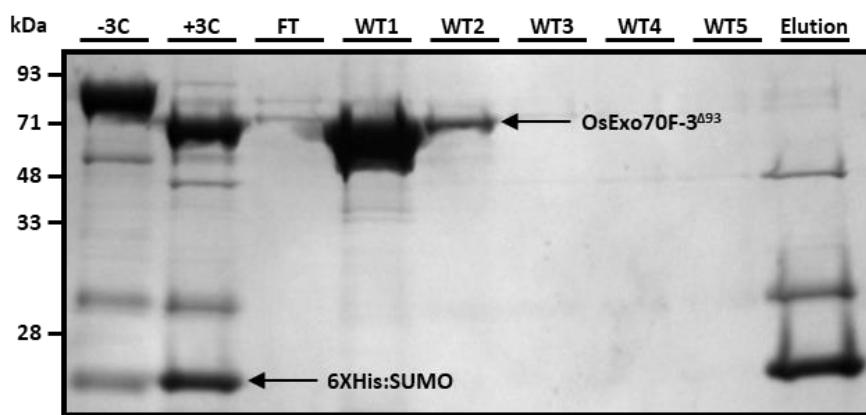
To enable the study of the AVR-Pii/Exo70 interactions in vitro, I first aimed to produce stable plant Exo70 proteins in *E. coli* (see details in materials and methods). For this, I predicted the secondary and tertiary characteristics of the Exo70 proteins of interest using the Raptor X server (Källberg et al., 2012). After this, I assembled constructs truncating the first predicted  $\alpha$ -helix in the pOPINS3C vector, enabling the heterologous production of soluble Exo70 protein lacking the first  $\alpha$ -helix with a N-terminal hexa-histidine (6xHis) tag adjacent to a SUMO tag.

This fusion protein can be separated from the *E. coli* cell lysate by immobilised metal affinity chromatography (IMAC) coupled to gel filtration in a size exclusion chromatography. Elution fractions were collected and evaluated by SDS-PAGE, revealing a band close to 70 kDa (Figure 7.4). These fractions were combined and concentrated, before the cleavage of the 6xHistidine:SUMO tag with 3C protease.

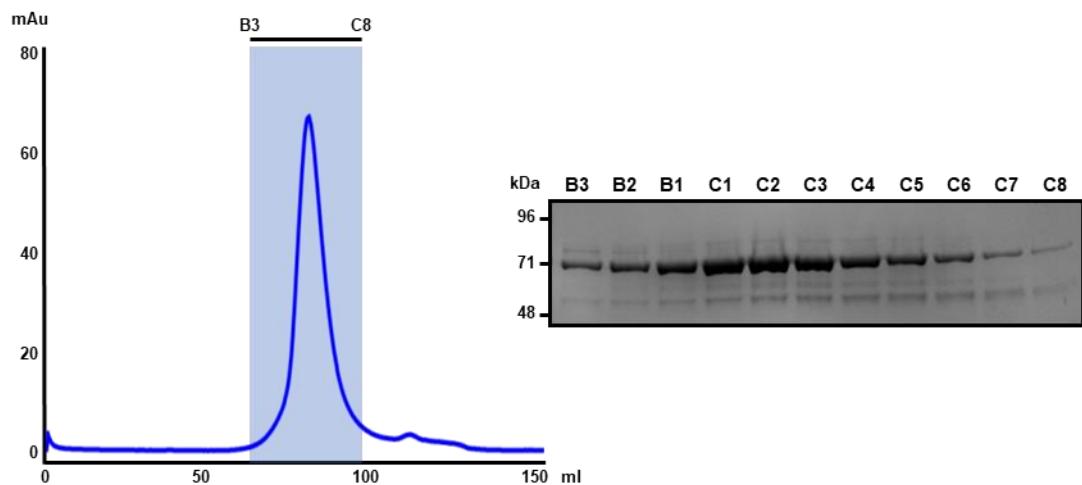


**Figure 7.4. Plant Exo70 proteins can be efficiently purified by IMAC and gel filtration.** Elution trace of SUMO:OsExo70F-3 $\Delta$ 93 after IMAC and gel filtration. Fractions from A3 to B2 were collected and analysed by SDS-PAGE to assess the presence of the purified protein.

Plant Exo70 proteins were then separated from the SUMO tag by passing the protein mixture solution through a HisTrap™ column (Figure 7.5). Exo70 proteins were mainly present in the first and second wash-through (WT) from the column, whilst the SUMO tag was mainly in the final elution (Figure 7.5). Fractions containing Exo70 proteins were pooled together and concentrated, before final purification with gel filtration (Figure 7.6). Fractions containing purified Exo70 proteins were combined and concentrated for biophysical studies.



**Figure 7.5. The 6xHis:SUMO tag can be efficiently cleaved to yield Exo70 protein.** SDS-PAGE analysis of fractions collected in the OsExo70F-3 $\Delta$ 93 HisTrap™ purification before and after 3C protease treatment. OsExo70F-3 $\Delta$ 93 was successfully eluted in the first and second wash-through (WT) steps. The 6xHis:SUMO tag eluted upon treatment with buffer containing imidazole.



**Figure 7.6. The final gel filtration step yields pure OsExo70F-3 $\Delta^{93}$  protein.** Elution trace of OsExo70F-3 $\Delta^{93}$  in the gel filtration purification step. SDS-PAGE analysis of the collected fractions showed the purified protein in fractions B3 – C8.

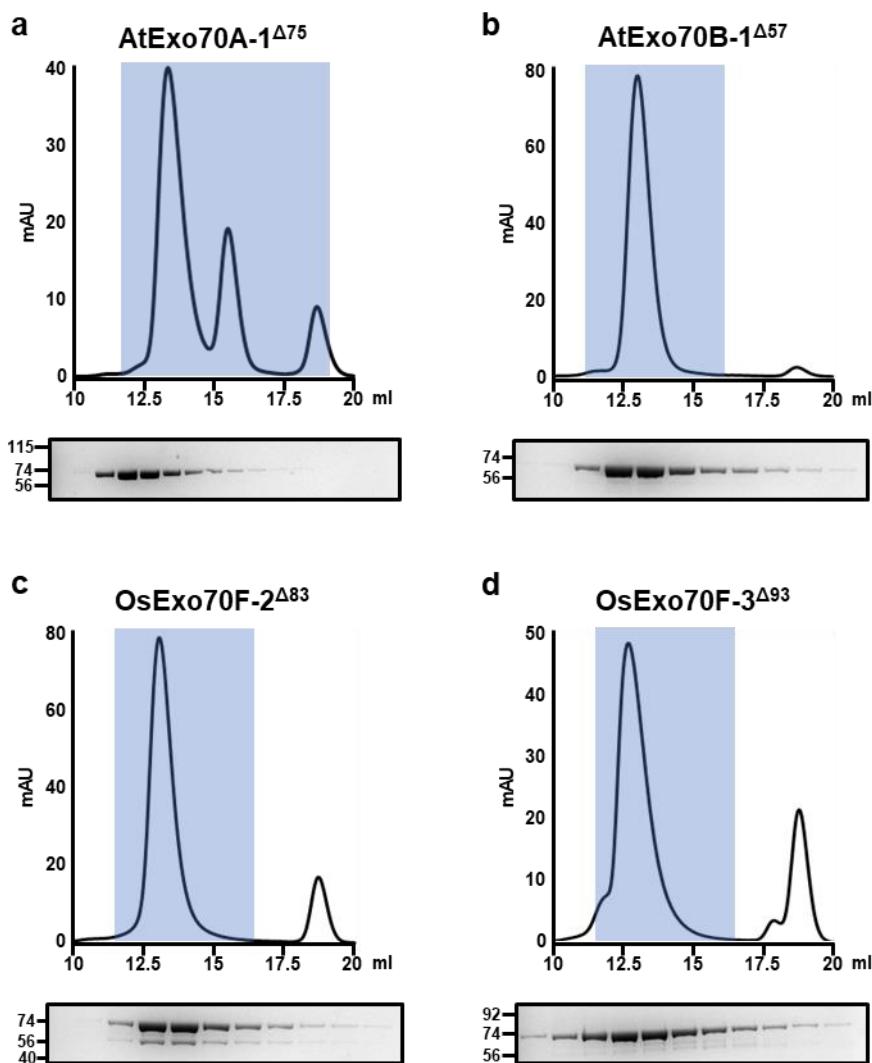
Therefore, I set up a pipeline for the production and purification of plant Exo70 proteins in sufficient concentrations to enable biophysical and structural studies. Using this approach, I successfully purified several plant Exo70 proteins, which can potentially also be used in studies of their role in diverse organisms and processes.

## 7.4 Purified proteins conserve the characteristic Exo70 protein fold

Once purified, I investigated whether the plant Exo70 proteins were correctly folded and therefore suitable for structural and biophysical studies. As a first approach, I used analytical gel filtration (AGF) to infer the oligomeric or aggregation state of the purified proteins (Figure 7.7).

As a reference, I performed this assay using *Arabidopsis* AtExo70A-1 $\Delta^{75}$  as this protein is the only plant Exo70 previously produced and crystallized (Zhang et al., 2016). Although the elution trace showed multiple peaks, SDS-PAGE analysis of the fractions confirmed that AtExo70A-1 $\Delta^{75}$  elutes as a single peak, while the others may correspond to impurities (Figure 7.7a).

This result was similar for other plant Exo70 proteins including *Arabidopsis* AtExo70B-1 $\Delta^{57}$  (Figure 7.7b), rice OsExo70F-2 $\Delta^{83}$  (Figure 7.7c) and rice OsExo70F-3 $\Delta^{93}$  (Figure 7.7d). All these proteins eluted as a single peak with a very similar elution volume, suggesting a similar conformation.

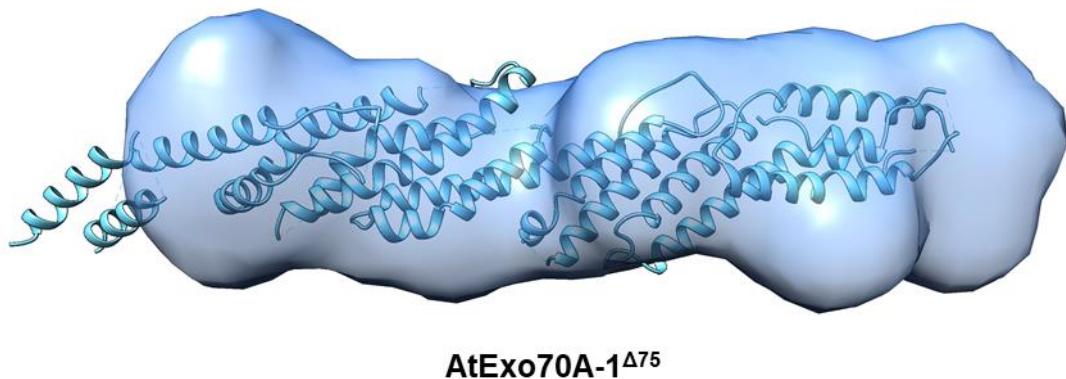


**Figure 7.7. Purified plant Exo70 proteins elute mainly as single peaks in analytical gel filtration analysis.** Elution traces of purified (a) AtExo70A-1 $\Delta^{75}$ , (b) AtExo70B-1 $\Delta^{57}$ , (c) OsExo70F-2 $\Delta^{83}$  and (d) OsExo70F-3 $\Delta^{93}$  in analytical gel filtration. SDS-PAGE analysis of the fractions highlighted in ice blue are shown below each trace.

Although AGF is useful as a first qualitative approach, the resolution of this analysis is not enough to precisely describe the oligomeric state of the sample. Therefore, I collected small angle X-ray scattering (SAXS) data for AtExo70A-1 $\Delta^{75}$ , Exo70B-1 $\Delta^{57}$ , Exo70F-2 $\Delta^{83}$  and Exo70F-3 $\Delta^{93}$ . Due to time constraints, a full analysis of the data could not be completed for each protein. However, preliminary data using AtExo70A-1 $\Delta^{75}$  confirmed the purified protein adopts a globular, rod-like, conformation in solution. The estimated molecular mass was 62.35 kDa, similar to the predicted mass of 67.33 kDa according to the protein sequence. SAXS data also indicated a slight aggregation of the particles.

With the collected SAXS data, I generated the shape reconstruction of the AtExo70A-1<sup>Δ75</sup> particles in solution, into which I docked the published AtExo70A-1<sup>Δ75</sup> crystal structure (Zhang et al., 2016). This showed that the purified protein has a conformation similar to the monomeric Exo70 fold (**Figure 7.8**).

Altogether, the data indicate that the purified plant Exo70 proteins are soluble and correctly folded. Although preliminary data also showed that the samples might be prone to some aggregation, they are mainly monomeric and uniform enough to perform biophysical experiments.



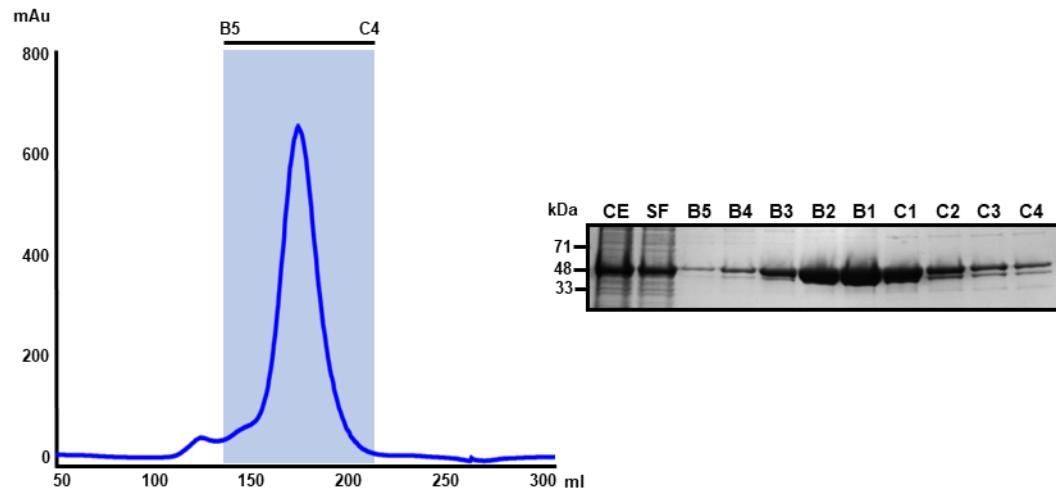
**Figure 7.8. Purified proteins adopt a folding shape similar to a monomeric AtExo70A-1 protein in solution.** Analysis of AtExo70A-1<sup>Δ75</sup> by small angle X-ray scattering. Superposition of the crystal structure of AtExo70A-1<sup>Δ75</sup> (Zhang et al., 2016) with the most probable ab initio envelope of the purified AtExo70A-1<sup>Δ75</sup> in solution.

## 7.5 Rice blast effector AVR-Pii can be produced and purified from *E. coli*

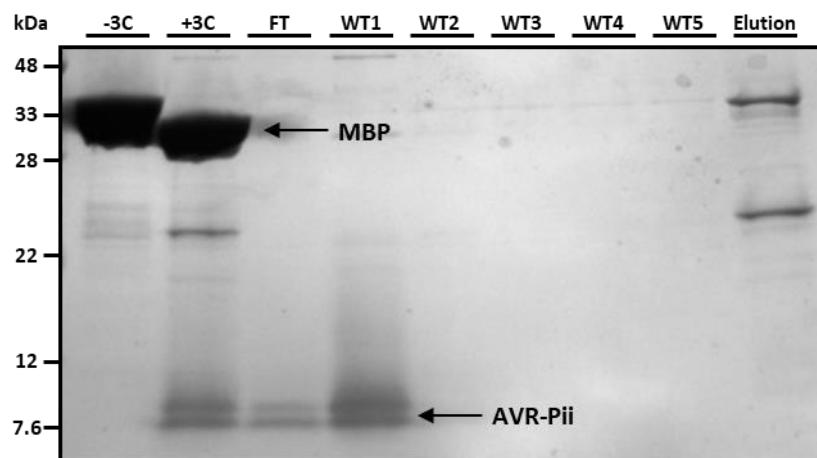
To enable biophysical and structural studies on the interaction between *M. oryzae* AVR-Pii and rice Exo70 proteins, I also developed a purification protocol for the effector domain of AVR-Pii, encompassing residues 20 to 70. Preliminary expression trials with only a 6xHis tag or with a 6xHis:SUMO tag did not yield soluble protein. However, using a AVR-Pii construct in the pOPINM vector produced a soluble protein.

The soluble fusion protein 6xHis:MBP:AVR-Pii could be purified from *E. coli* cell lysates by tandem IMAC and gel filtration (**Figure 7.9**). The fractions containing the eluted protein were subsequently treated with 3C protease as before to remove the MBP tag. After this, AVR-Pii was purified using HisTrap™ and MBPTrap™ columns attached in

tandem (**Figure 7.10**). Purified AVR-Pii was mainly present as a double band in the flow-through (FT) and wash-through (WT) from the columns (**Figure 7.10**).

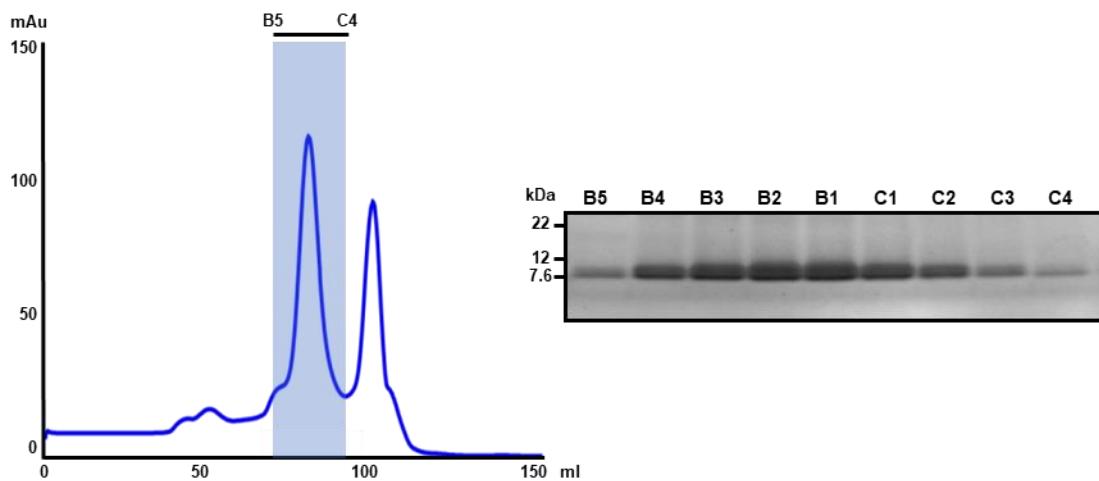


**Figure 7.9. Rice blast effector AVR-Pii can be efficiently purified by IMAC and gel filtration.** Elution trace of 6xHis:MBP:AVR-Pii after IMAC and gel filtration. Fractions from B5 to C4 were collected and analysed by SDS-PAGE, together with bacterial crude extract (CE) and the soluble fraction (SF) for comparison.



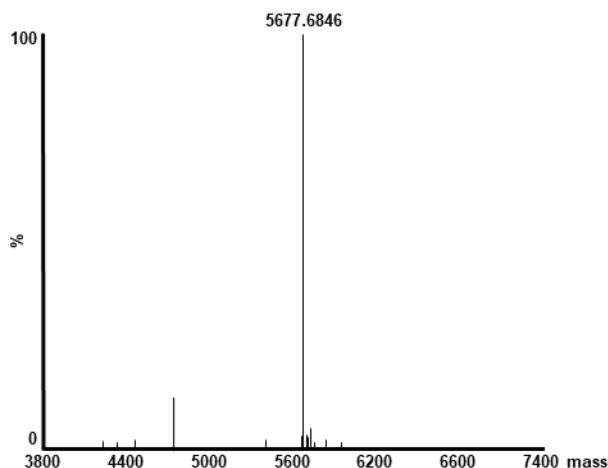
**Figure 7.10. 6xHis:MBP tag can be efficiently cleaved.** SDS-PAGE analysis of fractions collected in AVR-Pii HisTrap™ and MBPTrap™ tandem purification, following 3C protease treatment. Protein fractions before and after treatment with 3C protease are shown as inputs. AVR-Pii was purified forming a double band in the flow-through (FT) and the first wash-through (WT1). The 6xHis:MBP tag eluted in the upon treatment with buffer containing imidazole.

Because the AVR-Pii effector eluted as a double band, fractions were collected and subjected to a final gel filtration step using a GE Healthcare S75 16/60 column (**Figure 7.11**). This extra step showed an elution of two peaks, with pure AVR-Pii protein eluting as a single band in one of these (**Figure 7.11**).



**Figure 7.11. A final gel filtration step yields a pure AVR-Pii protein.** Gel filtration elution trace of AVR-Pii, showing two protein peaks in this purification step. SDS-PAGE analysis showed purified AVR-Pii elutes in fractions B5 – C4.

Fractions containing pure AVR-Pii effector were collected and concentrated for biophysical analysis. The state of the protein was assessed by intact mass spectrometry, revealing a main peak with a molecular weight of 5677.68 Da, identical to that calculated for AVR-Pii (Figure 7.12). In summary, I demonstrated that AVR-Pii effector can be produced and purified from *E. coli* in sufficient amounts to permit biophysical and structural studies.

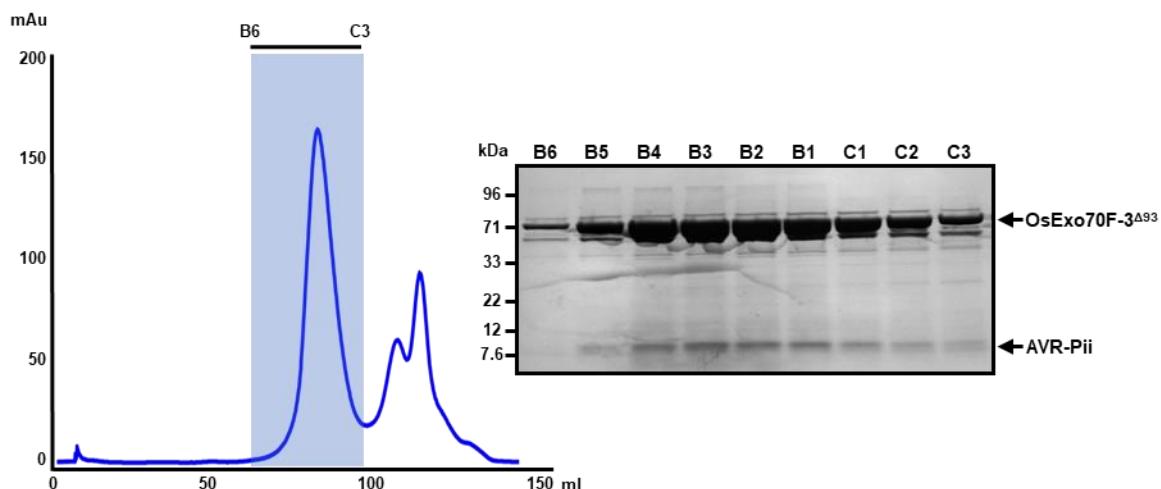


**Figure 7.12. Mass spectrometry analysis of AVR-Pii confirms the expected mass.** Intact mass analysis of purified AVR-Pii protein shows a main peak close to the calculated protein mass (5,675.0 Da).

## 7.6 AVR-Pii binds OsExo70F-2 and OsExo70F-3 with high affinity in vitro

Following purification of both *M. oryzae* effector AVR-Pii and the rice Exo70 proteins, I investigated whether these proteins interact in vitro. As a first approach, I used a co-expression and co-purification strategy to see if AVR-Pii binds its Exo70 proteins on expression in *E. coli*.

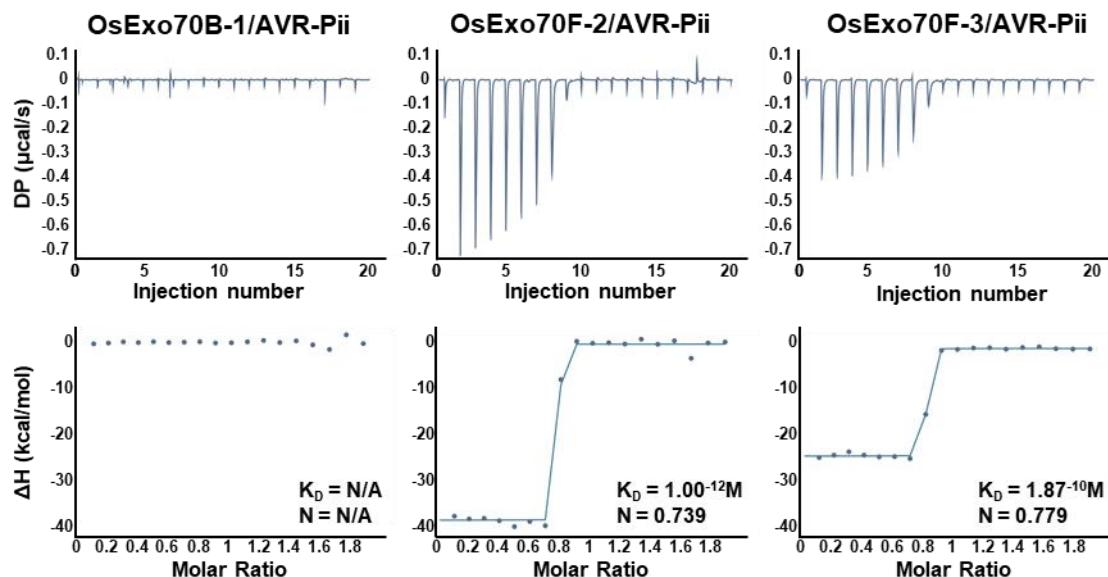
After co-expression and co-purification of 6xHis:MBP:AVR-Pii and untagged SUMO:OsExo70F-3 from *E. coli* cell lysates using IMAC and gel filtration I observed a protein band corresponding to the size of OsExo70F-3 co-purifying with a protein corresponding to the size of AVR-Pii (**Figure 7.13**). This confirms that AVR-Pii interacts with a rice Exo70 target in *E. coli*.



**Figure 7.13. AVR-Pii forms a complex with OsExo70F-3 in vitro.** Elution trace of co-purification of OsExo70F-3 $\Delta$ 93 and AVR-Pii after gel filtration. Fractions from B6 to C3 were collected and analysed by SDS-PAGE showing the presence of both proteins.

To quantitatively investigate the strength of the binding between AVR-Pii effector and the host targets I used Isothermal Titration Calorimetry (ITC). With this, I measured heat differences (indicative of protein/protein interactions) after titration of a solution containing purified rice Exo70 proteins with AVR-Pii effector.

These experiments showed that AVR-Pii binds to both OsExo70F-2 and OsExo70F-3 with a very high affinity ( $K_D$  ranging from  $1.00^{-9}$  M to  $1.00^{-12}$  M) (**Figure 7.14**). No binding was detected between AVR-Pii and OsExo70B-1 (**Figure 7.14**), confirming the specificity of the binding.



**Figure 7.14. AVR-Pii binds OsExo70F-2 and OsExo70F-3 with high affinity.** The binding affinity of AVR-Pii to OsExo70F-2 and OsExo70F-3 was determined using isothermal titration calorimetry (ITC). Upper panel shows heat differences upon injection of AVR-Pii into Exo70 and the bottom panel show integrated heats of injection (blue dots) and the best fit (solid line) using MicroCal PEAQ-ITC Analysis Software. When binding was detected  $K_D(M)$  and molar ratios are indicated. Panels are representative of three replicates.

The high binding affinity of AVR-Pii for OsExo70F-2 and OsExo70F-3 generates additional considerations when analysing the binding. Firstly, the saturation of the Exo70 titration with AVR-Pii is very fast, which creates a steep slope. The error calculated for such a small  $K_D$  value can be as much as two orders of magnitude. Secondly, the number of binding sites calculated is also affected by the sharp saturation, most likely explaining why the values for the number of binding sites range from 0.7 to 1.2 (Figure 7.14). This makes it challenging to provide a specific value for the binding affinity with high confidence. To overcome this, I will use AFFINImeter™ software to analyse all the binding data together (Piñeiro et al., 2019). This software derives a calorimetric coefficient for each ITC experiment, and will confidently calculate the apparent  $K_D$  for each reaction, for a given number of sites and given concentrations of each proteins.

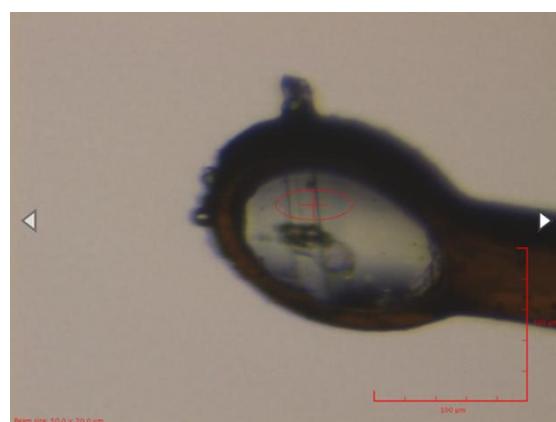
In summary, the ITC data clearly demonstrates that AVR-Pii specifically binds to OsExo70F-2 and OsExo70F-3 with nanomolar affinity.

## 7.7 Crystallization and preliminary structure of AVR-Pii bound to rice OsExo70F-2

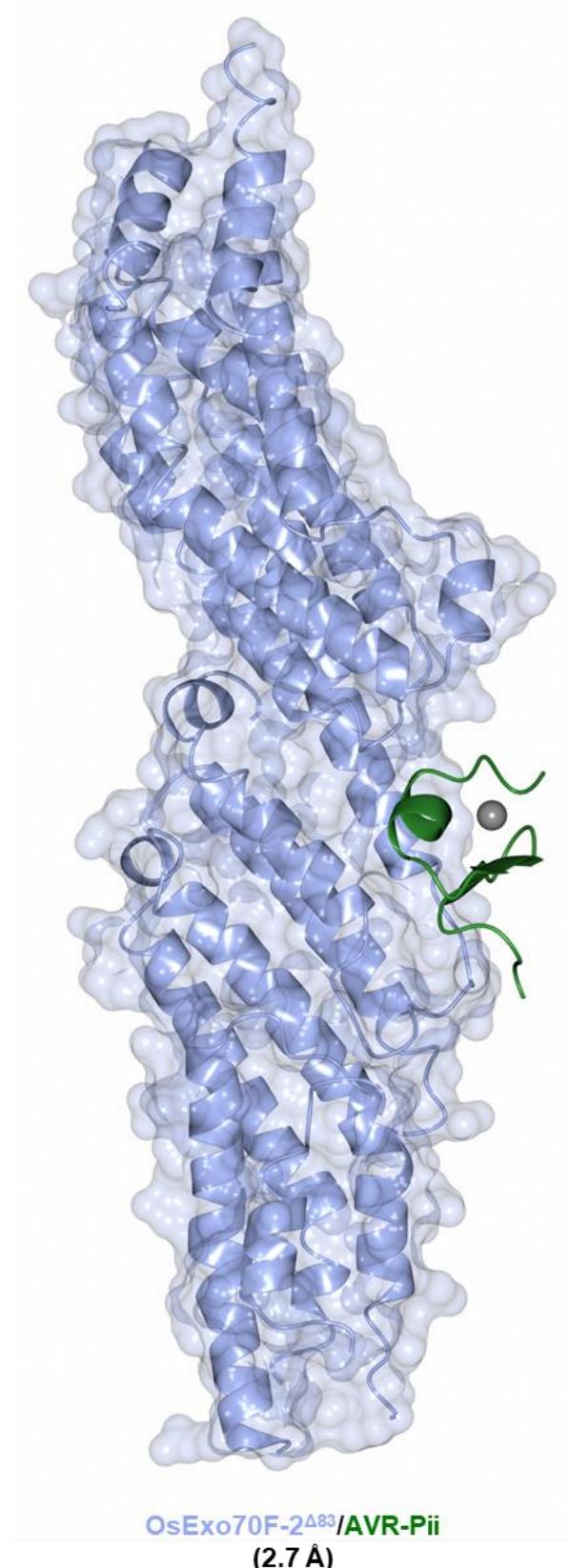
To investigate the structural basis of AVR-Pii binding to rice Exo70 proteins, used X-ray crystallography to obtain the structure of AVR-Pii in complex with one of its targets.

First attempts using a co-expression strategy, as for AVR-Pik/Pik-HMA, did not yield a complex with sufficient purity for crystallography. As an alternative, I mixed separately purified AVR-Pii effector and rice Exo70 proteins. After incubation, I concentrated the sample using a 30 kDa cut off concentrator, to remove any unbound AVR-Pii from the solution.

Using these samples, several commercial crystallisation screens were set up. After four to six days, protein crystals for the complex between OsExo70F-2 and AVR-Pii appeared from conditions in the PEG Suite screen (Qiagen). The conditions were optimised until long rod-shaped crystals, suitable for X-ray data collection, were obtained (**Figure 7.15**) (see Materials and methods for conditions and concentrations). Protein crystals were harvested by Dr. Clare Stevenson (JIC crystallography platform) and sent to Diamond Light Source (UK). Together with Professor David Lawson (JIC crystallography platform) we collected X-ray diffraction data and a model of Exo70-F2 was positioned in the electron density by molecular replacement. Subsequent phasing calculations generated new electron density maps that revealed unaccounted for electron density. Careful interpretation of these maps showed that the sequence of AVR-Pii could be confidently modelled into the density, revealing the structure of the complex between these proteins with a 1:1 stoichiometry (**Figure 7.16**). At the time of thesis submission, this structure is still undergoing rounds of refinement and rebuilding, but is of sufficient quality to inform further experiments to test protein function.



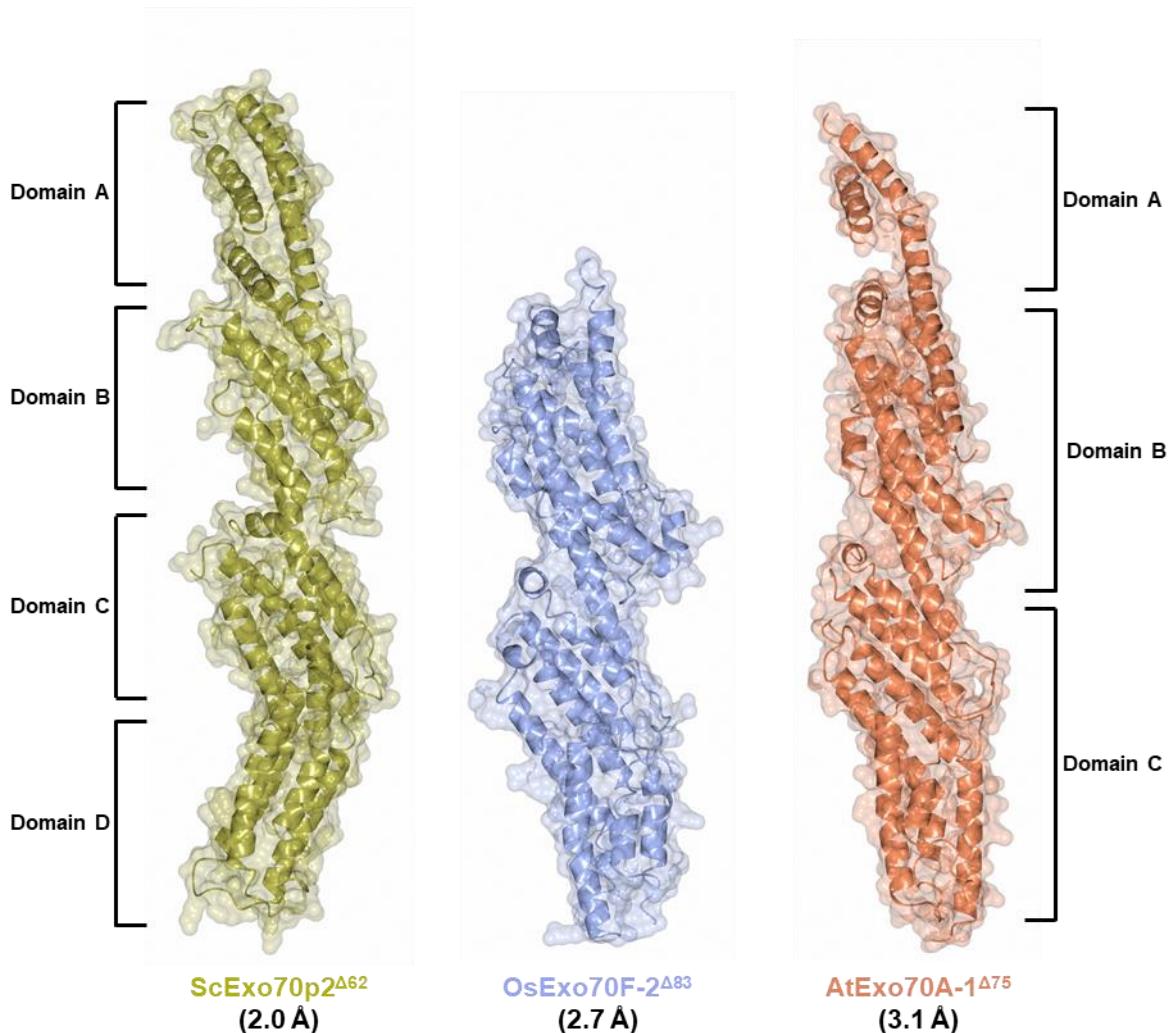
**Figure 7.15. OsExo70F-2/AVR-Pii complexes form rod-like crystals.** Image of the OsExo70F-2/AVR-Pii crystal mounted in the loop and centred in the beam during X-ray data collection.



**Figure 7.16. Crystal structure of AVR-Pii bound to rice Exo70F-2.** Schematic representations of the structures of OsExo70F-2 in complex with AVR-Pii effector. OsExo70F-2 is shown in ice blue cartoon representation with the molecular surface also shown. The AVR-Pii effector is coloured in green and shown in cartoon representation. Zinc atom is also represented as a grey sphere.

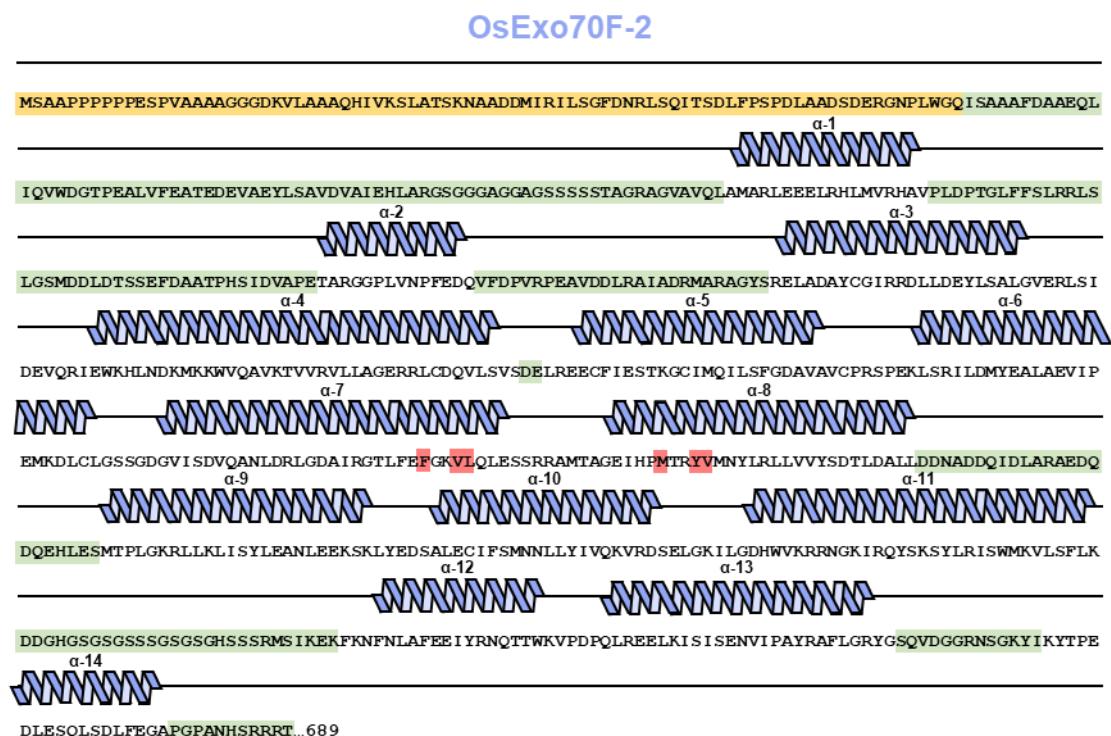
## 7.8 OsExo70F-2 adopts a conserved Exo70 fold

The crystal structure of rice OsExo70F-2 confirmed that this protein forms a long rod-shaped fold composed of contiguous  $\alpha$ -helical bundles, as previously shown for yeast ScExo70p (Dong et al., 2005) and Arabidopsis AtExo70A-1 (Zhang et al., 2016) (Figure 7.17). Previous research has shown that Exo70 proteins from distinct organisms share a similar fold (Dong et al., 2005; Hamburger et al., 2006; Moore et al., 2007; Zhang et al., 2016). The Exo70 fold consists of sequentially arranged  $\alpha$ -helical bundles divided into four domains (A-D) in the case of the yeast ScExo70p (Dong et al., 2005), or three domains in the Arabidopsis AtExo70A-1 (Zhang et al., 2016) (Figure 7.17).



**Figure 7.17. Rice Exo70F-2 adopt a similar fold to other Exo70 proteins.** Schematic representations of the overall structures of yeast Exo70p, rice Exo70F-2 and Arabidopsis Exo70A-1. Each Exo70 structure is coloured as labelled and is shown in cartoon representation with the molecular surface.

The OsExo70F-2 protein used for crystallization contained all but the first 83 N-terminal residues, however the electron density observed for the region defined as “Domain A” is of poor quality, indicating that this region is disordered. This was previously observed for AtExo70A-1 (Zhang et al., 2016). Therefore, the preliminary OsExo70F-2 structure encompasses residues 158 to 678 of the full-length protein. However, additional residues, including 176 to 216, 231 to 256, 460 to 482, 571 to 598 and 648 to 660, are not modelled in the structure due to insufficient quality of the electron density (**Figure 7.18**).



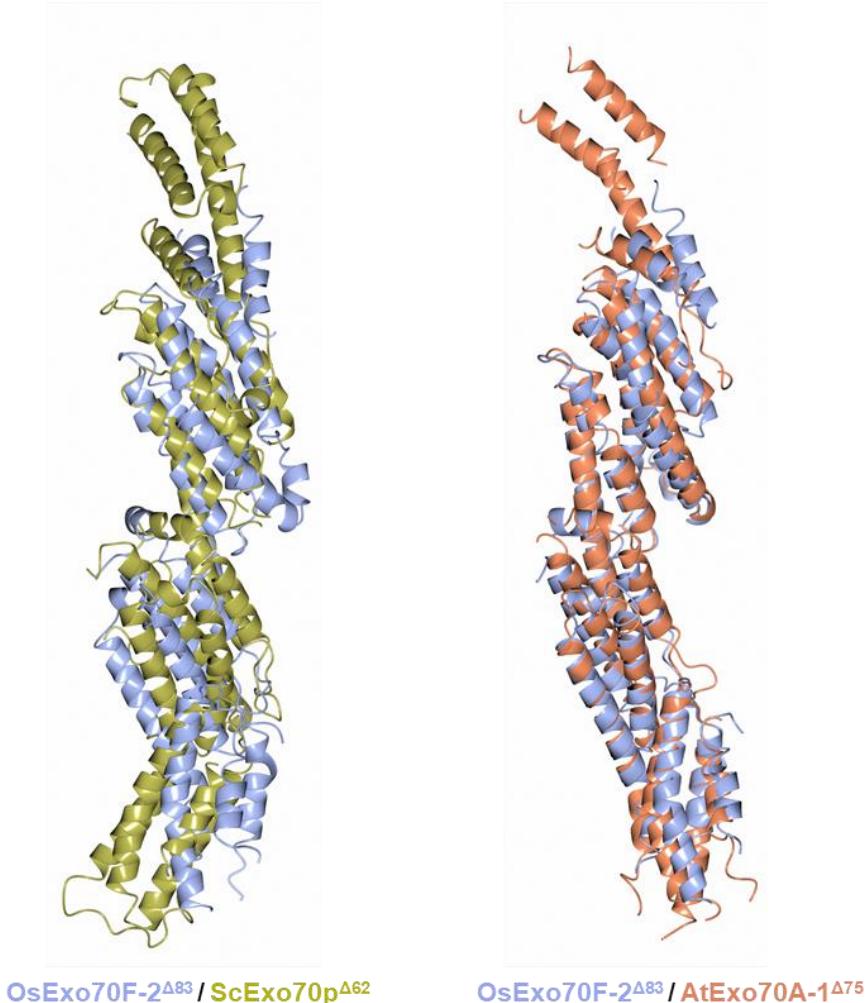
**Figure 7.18. Distribution of  $\alpha$ -helices in the sequence of OsExo70F-2.** Amino acid sequence of OsExo70F-2. N-terminal residues truncated in the construct for heterologous expression are highlighted in orange and residues that are not present in the electron density are highlighted in green. Residues contributing to the AVR-Pii binding pocket are highlighted in red. Secondary structure features of the Exo70 fold are shown above the respective sequence.

I could trace 14  $\alpha$ -helices in the structure of OsExo70F-2, named  $\alpha$ -1 to  $\alpha$ -14. Only the N-terminal  $\alpha$ -helices  $\alpha$ -1 and  $\alpha$ -2 (encompassing residues 158 to 175 and 217 to 230, respectively) can be modelled in Domain A, although these are of poor confidence due to quality of the electron density in this region.

Domain B in OsExo70F-2 is comprised of  $\alpha$ -3 to  $\alpha$ -7, encompassing residues 257 to 427. This domain is better defined in the electron density, and amino acid side chains can be placed in the electron density map with confidence (**Figure 7.18**). The C-terminal region

(Domains C and D) spans residues 428 to 678, forming  $\alpha$ -8 to  $\alpha$ -14 (Figure 7.18). In OsExo70F-2, this region is more compact compared to the equivalent domains in the yeast and Arabidopsis proteins (Figure 7.17). The overall structure of the rod narrows substantially in the junction between domain B and domain C (Figure 7.17). Interestingly, this is the position where the AVR-Pii effector associates to OsExo70F-2, between OsExo70F-2  $\alpha$ -7 and  $\alpha$ -8 (Figure 7.16).

To better understand conservation between rice OsExo70F-2, yeast ScExo70p and Arabidopsis AtExo70A-1, I superposed their structures. Although the overall fold is very similar, the root mean square deviation (rmsd) of the OsExo70F2 and AtExo70A-1 superposition was 2.9 Å, based on 287  $\text{C}\alpha$  atoms. This difference was much higher in the superposition of OsExo70F2 and yeast ScExo70p, with an rmsd value of 6.5 Å, based on 322  $\text{C}\alpha$  atoms (Figure 7.19).



**Figure 7.19. Rice OsExo70F-2 is structurally similar to yeast and Arabidopsis Exo70 proteins.** Superposition of the overall structures of yeast Exo70p with rice Exo70F-2 (left) and Arabidopsis Exo70A-1 with rice Exo70F-2 (right). Each Exo70 structure is coloured as labelled and is shown in cartoon representation.

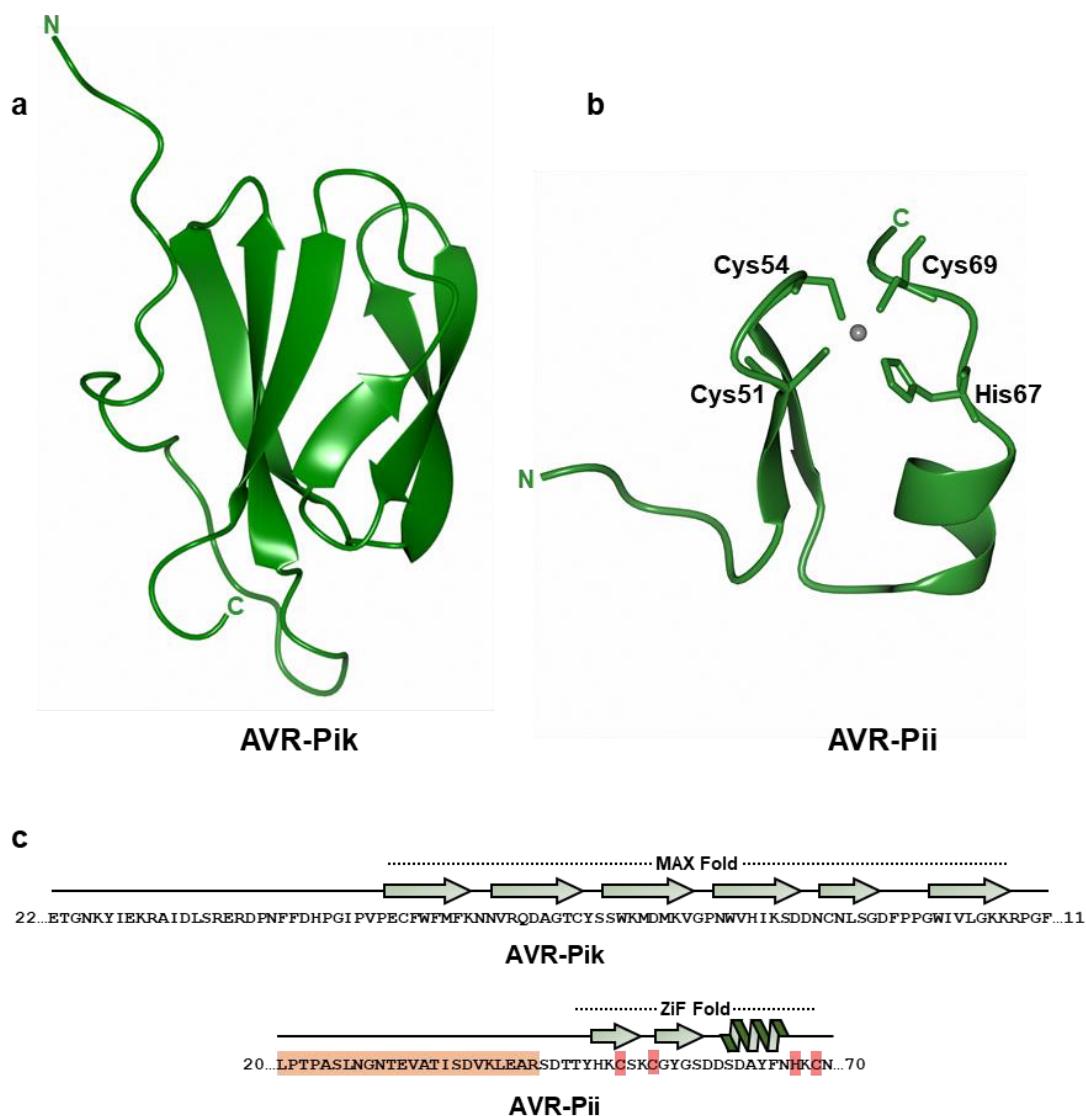
Overall, the structure of rice OsExo70F-2 shows strong conservation with other Exo70 proteins, despite sharing a low sequence identity. This indicates that the core biochemical functions of these domains are most likely conserved. However, the structures are subtly different, and this may underpin specialized roles. Differential functions between different Exo70 proteins may explain why AVR-Pii specifically targets a small subset of rice Exo70 proteins.

## 7.9 AVR-Pii adopts a new fungal effector fold, based on a Zinc-Finger motif

To date, all the rice blast effectors structurally characterised share the same protein fold. This conserved fold (a six-stranded  $\beta$ -sandwich) defines the large MAX effector family from *M. oryzae*, and other ascomycete fungi, including the AVR-Pik and AVR-Pia effectors, despite low sequence identity (Bialas et al., 2018; de Guillen et al., 2015) (Figure 7.20a, c).

In contrast to the MAX effectors, the crystal structure of AVR-Pii (in complex with OsExo70F-2) revealed a different effector conformation (Figure 7.20b). This fold is defined by a Zinc-finger motif (two short  $\beta$ -sheets and an  $\alpha$ -helix), with AVR-Pii residues Cys51, Cys54, His67 and Cys69 chelating a zinc atom (Figure 7.20b, c). This protein fold has not been observed before in fungal effectors and, given the central role of the Zinc-finger motif, I named AVR-Pii as a Zinc-finger Fold (ZiF) effector.

Interestingly, only the C-terminal 26 residues of AVR-Pii were observed in the electron density, suggesting that the 25 residues of the N-terminal region are disordered. As they do not appear to contribute to binding of OsExo70F-2, these residues may be involved in a different process, or are not of functional significance.

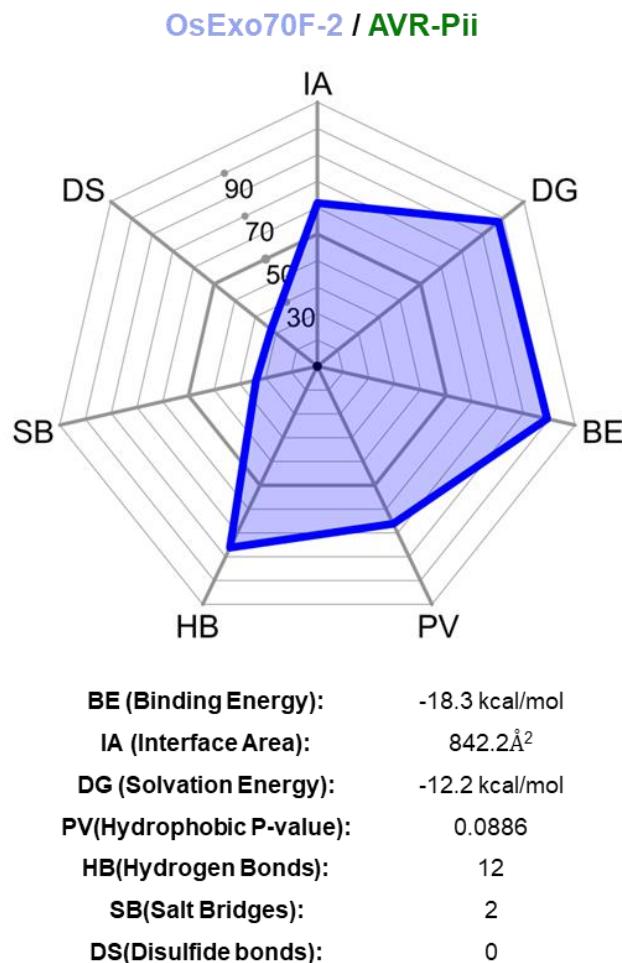


**Figure 7.20. AVR-Pii adopts a new fungal effector fold.** Overall structures of *M. oryzae* effectors AVR-Pik (a) and AVR-Pii (b). Residues contributing to the formation of the Zinc Finger Fold (ZiF) are indicated. (c) Amino acid sequences of AVR-Pik and AVR-Pii. Secondary structure features of the MAX fold and the ZiF fold are shown above. AVR-Pii residues missing in the crystal structure are highlighted in orange and residues contributing to the ZiF fold are highlighted in red.

## 7.10 AVR-Pii binds OsExo70s in a hydrophobic pocket

To understand the structural basis of AVR-Pii binding to rice Exo70 proteins, I used the crystal structure of AVR-Pii bound to OsExo70F-2 to dissect the determinants of effector binding and specificity.

Analysis of the binding interface between OsExo70F-2 and AVR-Pii using QtPISA (Krissinel, 2015) demonstrated significant differences in the area involved in the interaction between the effector and the target. Only a 3.1% of the OsExo70F-2 surface is contributing to the interaction, while a 43.4% of the AVR-Pii surface is involved. This reflects the disproportion of protein sizes between the AVR-Pii effector (5.5 kDa) and host target OsExo70F-2 (67.2 kDa). The key interface components of the interaction are shown using QtPISA interaction radars (Krissinel, 2015) in **Figure 7.21**.



**Figure 7.21. Analysis of OsExo70F-2 binding interface with AVR-Pii effectors using QtPISA.** QtPISA interaction radar generated for the OsExo70F-2/AVR-Pii complex. Values obtained for each key interface parameter in the analysis are also shown. The QtPISA interaction radars score macromolecular interfaces based on statistical analysis of all interfaces found in the Protein DataBank (see Methods). The abbreviations in the radars are defined in the tables underneath each panel.

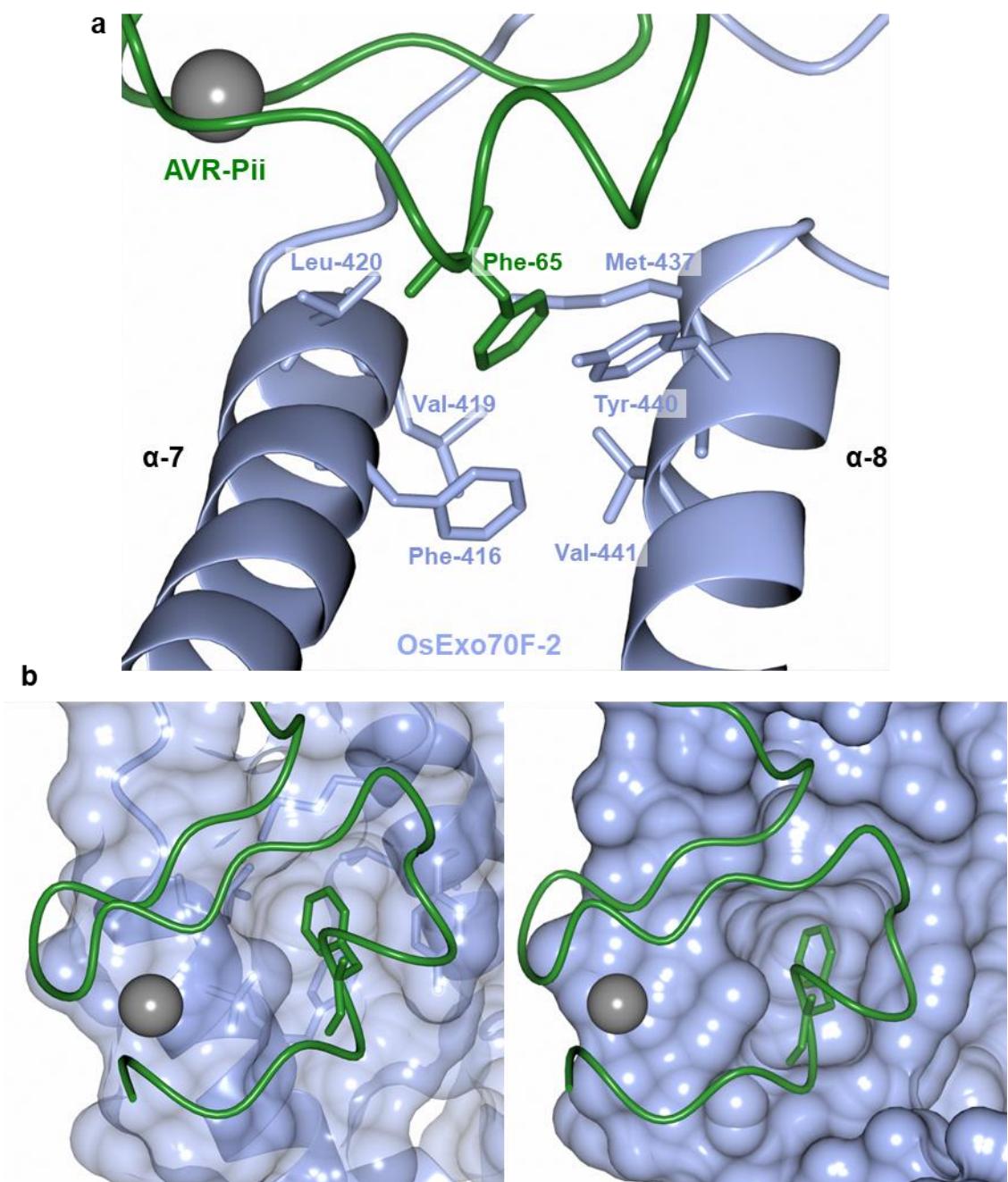
## Chapter 7

As mentioned above, only the C-terminal 26 AVR-Pii residues are defined in the structure of the complex. For the purpose of comparison, while the interface area of the OsExo70F-2/AVR-Pii complex is smaller than the area of Pikp-HMA/AVR-PikD (842.2 Å<sup>2</sup> compared to 986.3 Å<sup>2</sup>), the total binding energy and the solvation energy are higher (-18.3 kcal/mol and -12.2 kcal/mol compared with -10.9 kcal/mol and -2.4 kcal/mol respectively) (**Figure 7.21**).

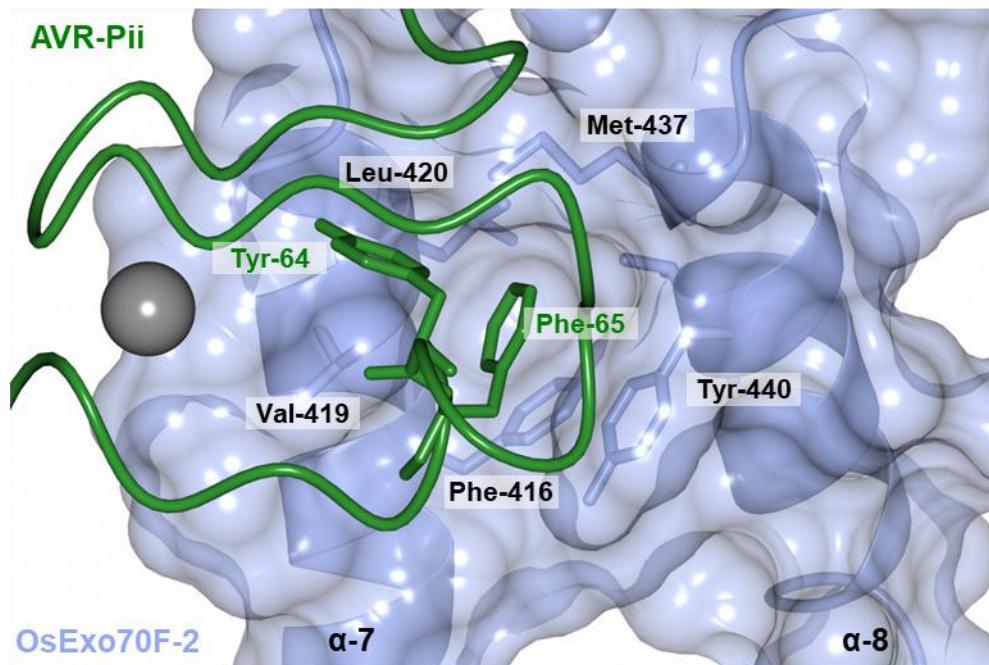
Inspection of the interface in the OsExo70F-2/AVR-Pii complex revealed a major role of the side chain of residue Phe65 in the binding. This residue inserted in a hydrophobic pocket lined by OsExo70F-2 residues Phe416, Val419 and Leu420 from  $\alpha$ -7 and Met437, Tyr440 and Val441 from  $\alpha$ -8 (**Figure 7.22**). An adjacent residue in AVR-Pii, Tyr45, also forms a significant contribution to the interaction. The phenyl ring of Tyr45 interacts with OsExo70F-2 Val419, positioning both Val419 and Leu420 residues in a favourable position for interacting with Phe46 in the aforementioned hydrophobic pocket (**Figure 7.23**).

Further, a region of AVR-Pii encompassing residues His49, Lys50, Cys51 and Ser52, forms main chain hydrogen bond interactions with side chains of OsExo70F-2, contributing to binding.

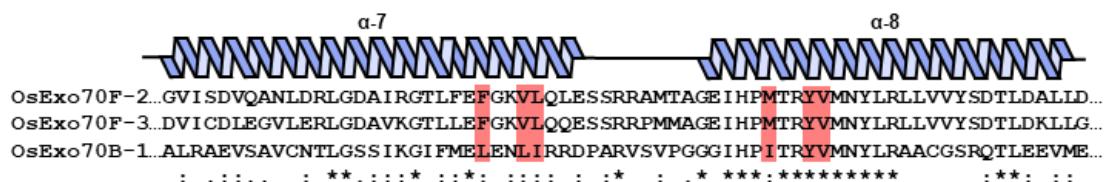
In summary, the crystal structure of AVR-Pii bound to OsExo70F-2 revealed the structural basis of how this effector has adapted to target rice Exo70. Although only formed by a limited interface, the AVR-Pii effector binds strongly to a hydrophobic pocket in OsExo70F-2, mainly using the side chains of residues Tyr64 and Phe65. As the residues at  $\alpha$ -7 contributing to the binding pocket are less conserved in OsExo70B1, subtle variations in this pocket may be underpinning AVR-Pii effector/target specificity (**Figure 7.24**).



**Figure 7.22. AVR-Pii residue Phe65 mediates the effector interaction with the host target.** Schematic view of the interaction of AVR-Pii Phe65 at hydrophobic pocket of OsExo70F-2. **(a)** zoom in view of the residues involved in the interaction **(b)** surface representation of AVR-Pii Phe 65 position at the binding pocket. In each panel, the effector is shown as Ca-worm with side-chain of Phe65 represented as cylinders and coloured in green. OsExo70F-2 is represented as ribbons and coloured in ice blue.



**Figure 7.23. AVR-Pii residue Tyr64 makes important contributions to AVR-Pii effector/target interaction.** Zoom in view of the interaction of AVR-Pii Tyr64 and Phe65 at hydrophobic pocket of OsExo70F-2. The effector is shown as Ca-worm with side-chain of Tyr64 and Phe65 represented as cylinders and coloured in green. OsExo70F-2 is represented as ribbons and coloured in ice blue with molecular surface also shown.

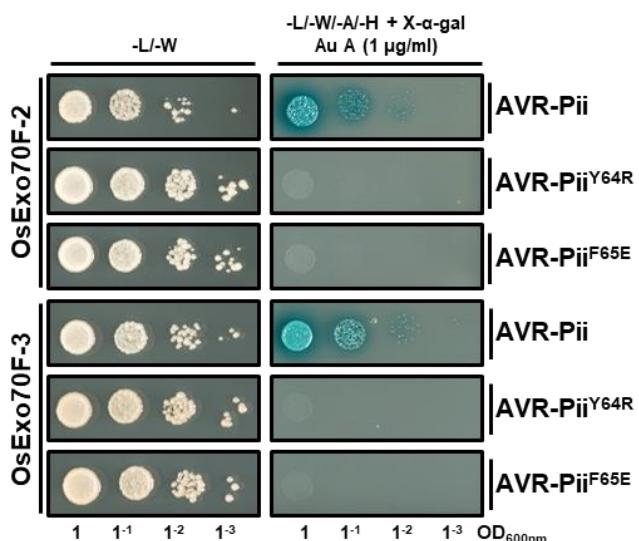


**Figure 7.24. Variation in OsExo70 hydrophobic pocket may underpin differences in effector binding.** Amino acid sequence alignment of OsExo70F-2, OsExo70F-3 and OsExo70B-1 α-7 and α-8. Residues contributing to the AVR-Pii binding pocket are highlighted in red. Secondary structure features of the Exo70 fold are shown above the respective sequence.

## 7.11 Mutations at the effector interface abrogate binding to rice Exo70s

To validate structure and explore the role of AVR-Pii residues Tyr64 and Phe65 in binding to rice Exo70 proteins, I produced mutations in these positions designed to disrupt the binding interface. This involved mutating Tyr64 to an Arg residue and Phe65 to a Glu residue. I then tested the ability of these mutants to associate with OsExo70F-2 and OsExo70F-3.

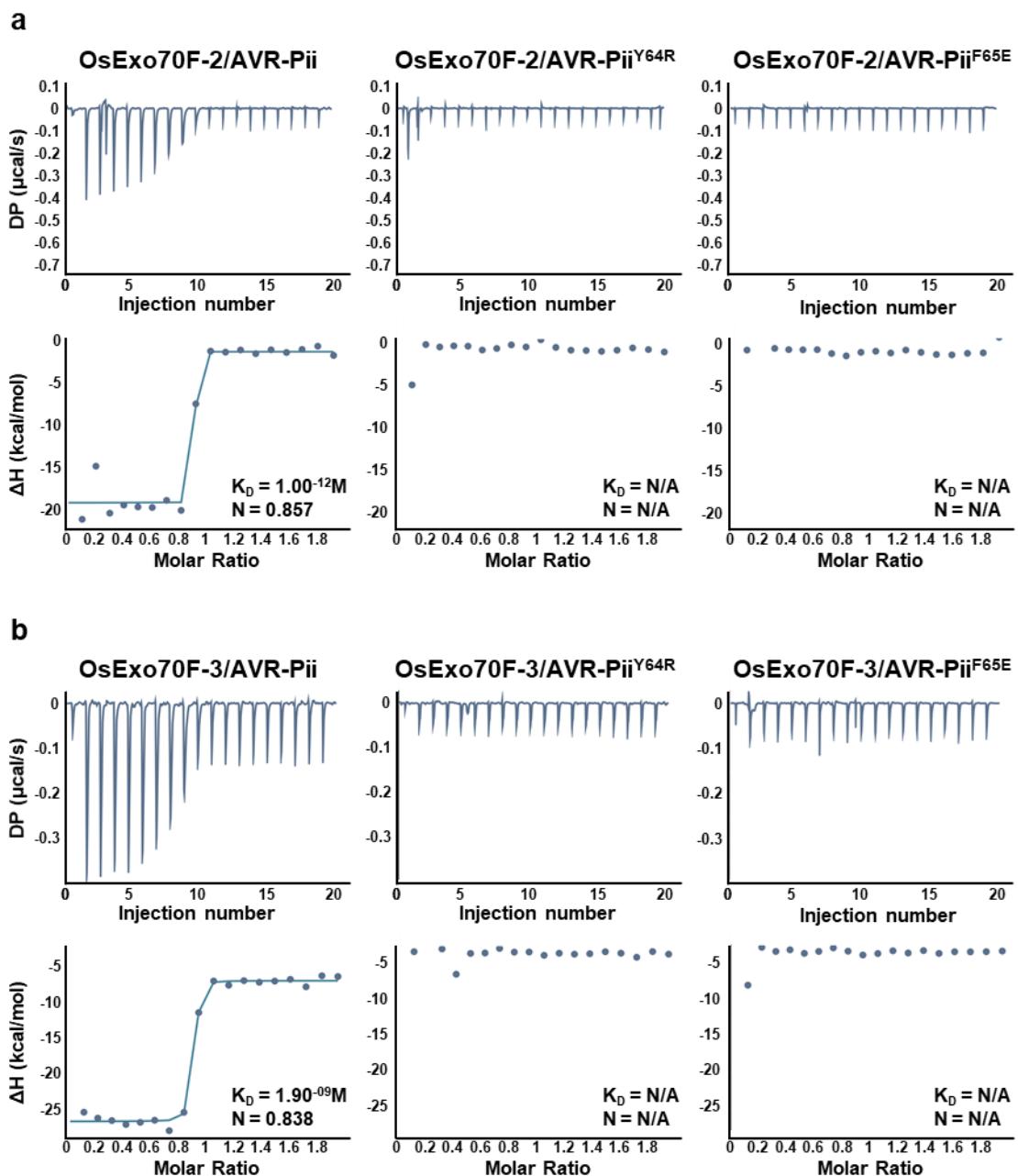
First, I performed a Y2H analysis. In this assay, no yeast growth nor development of blue colouration was observed for either of the AVR-Pii mutants with either of OsExo70F-2 or OsExo70F-3, in contrast to what is observed with the wild type effector (Figure 7.25). This is consistent with the central role of residues Tyr64 and Phe65 in the binding of the effector targets.



**Figure 7.25. Effector mutations at AVR-Pii binding interface perturb interactions with rice Exo70s in vivo.** Yeast-Two-Hybrid assay of AVR-Pii mutants Tyr64Arg and Phe65Glu with rice Exo70F-2 and Exo70F-3. The control plate for yeast growth is on the left with quadruple dropout media supplemented with X- $\alpha$ -gal and Aureobasidin A (Au A) on the right. Growth and development of blue colouration in the selection plate are both indicative of protein:protein interactions. Wild-type AVR-Pii is included as positive control. Rice Exo70s were fused to the GAL4 DNA binding domain, and AVR-Pii effectors to the GAL4 activator domain. Each experiment was repeated a minimum of three times, with similar results.

Secondly, I used the in vitro ITC assay to observe any binding for the effector mutants with OsExo70F-2 or OsExo70F-3. For this I expressed and purified both AVR-Pii Tyr64Arg and Phe65Glu mutants from *E. coli*. There were no observable differences in protein stability compared with the wild type effector, suggesting that these mutations do not affect the overall protein conformation. I then tested the binding of these mutants with

OsExo70F-2 and OsExo70F-3 proteins. This experiment revealed that neither AVR-Pii Tyr64Arg nor AVR-Pii Phe65Glu interact with OsExo70F-2 (**Figure 7.26a**) or OsExo70F-3 (**Figure 7.26b**).



**Figure 7.26. Effector mutations at AVR-Pii binding interface abrogate interactions with rice Exo70s in vitro.** The binding affinity of AVR-Pii mutants Tyr64Arg and Phe65Glu to (a) OsExo70F-2 and (b) OsExo70F-3 was determined using isothermal titration calorimetry (ITC). Wild type AVR-Pii was included as positive control. Upper panel shows heat differences upon injection of the effector into the Exo70 proteins and the bottom panel show integrated heats of injection (blue dots) and the best fit (solid line) using MicroCal PEAQ-ITC Analysis Software. When binding was detected  $K_D(M)$  and molar ratios are indicated. Panels are representative of three replicates.

Altogether, these results confirm that both of the residues identified as major components of the AVR-Pii effector/target interaction in the crystal structure are essential for binding. These mutants offer the exiting possibility of obtaining a better understanding of the mechanism of AVR-Pii virulence activity and mechanism of recognition. To address these questions, I have generated AVR-Pii constructs harbouring the wild type and the mutated sequences at residues Tyr64 and Phe65, under control of the rice blast effector AVR-Pik promotor. These constructs will be transformed into *M. oryzae* strain Sasa2 lacking AVR-Pii effector by Dr. Koki Fujisaki (Iwate Biotechnology Research Center). Transgenic Sasa2 lines containing the different AVR-Pii effector mutants will be inoculated into rice plants from cultivar Hitomebore harbouring the Pii resistance and control plants. This will help to further establish a causal link between AVR-Pii binding to host rice OsExo70F-3 and Pii mediated immunity.

## 7.12 OsExo70F-2 and OsExo70F-3 binds to rice RIN4 proteins in vivo

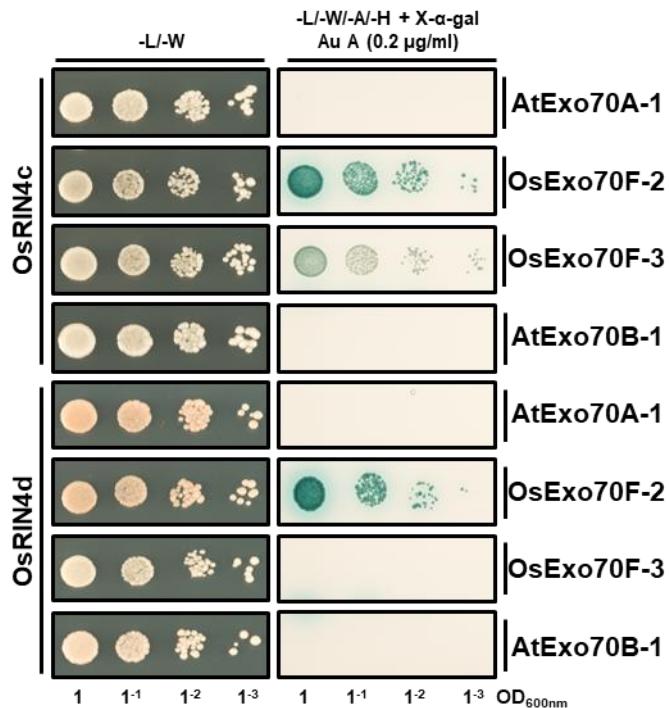
AVR-Pii selectively binds to only two out of 47 members of the rice Exo70 protein family (Fujisaki et al., 2015). This remarkable specificity may indicate that the effector is targeting a specific function carried out by exocyst complexes containing OsExo70F-2 or OsExo70F-3. To gain insight into possible functions of OsExo70F-2 and OsExo70F-3 targeted by AVR-Pii, I investigated possible interactions of rice Exo70s with other host proteins.

Arabidopsis AtExo70B-1 has been suggested to have a role in plant immunity by several studies (Liu et al., 2017; Stegmann et al., 2013; Wang et al., 2019d; Zhao et al., 2015). Interestingly, this protein has been shown to interact with Arabidopsis RIN4 (Sabol et al., 2017), linking the function of the plant exocyst and plant immunity.

Dr. Koki Fujisaki (Iwate Biotechnology Research Center) previously identified two rice RIN4 paralogs that were expressed during rice blast infection, named OsRIN4c and OsRIN4d (personal communication). To test whether OsExo70F-2 and OsExo70F-3 associate with rice RIN4 proteins, Neftaly Cruz and I performed Y2H assays with rice Exo70 proteins and OsRIN4c or OsRIN4d (**Figure 7.27**).

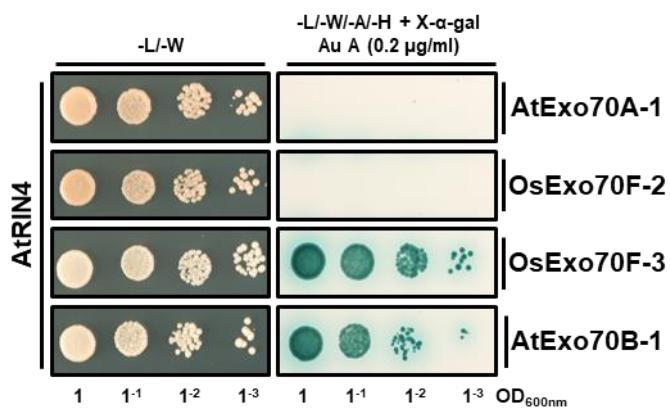
The results revealed that OsExo70F-2 binds to OsRIN4c and OsRIN4d in vivo, whilst OsExo70F-3 clearly associated with OsRIN4c (**Figure 7.27**). Arabidopsis AtExo70A-1 and AtExo70B-1 were included as negative controls, and did not show binding to OsRIN4c or OsRIN4d. These findings establish a putative link between the rice Exo70 proteins targeted by AVR-Pii and the plant immune system, providing clues as to the

functions that may be altered by AVR-Pii, which requires further study, much of which is beyond the scope of this thesis.



**Figure 7.27. Rice Exo70F-2 and Exo70F-3 bind to Rice RIN4 in vivo.** Yeast-Two-Hybrid assay of Rice RIN4c and RIN4d with rice Exo70 homologs OsExo70F-2 and OsExo70F-3. Control plate for yeast growth is on the left with quadruple dropout media supplemented with X- $\alpha$ -gal and Aureobasidin A (Au A) on the right. Growth and development of blue colouration in the selection plate are both indicative of protein:protein interactions. Arabidopsis Exo70 homologs AtExo70A-1 and AtExo70B-1 were included as negative controls. Exo70 proteins were fused to the GAL4 DNA binding domain, and RIN4 proteins to the GAL4 activator domain. Each experiment was repeated twice, with similar results.

However, in preliminary experiments to investigate the specificity of the Exo70/RIN4 binding, we repeated the Y2H assay to include Arabidopsis RIN4. Arabidopsis AtExo70A-1 and AtExo70B-1 were included as negative and positive controls, respectively. Interestingly, OsExo70F-2 seems to specifically bind to rice RIN4 but, in contrast, association of OsExo70F-3 to RIN4 proteins seems more promiscuous as we observed growth and development of blue coloration in the Y2H assay between OsExo70F-3 and AtRIN4 (Figure 7.28).



**Figure 7.28. Rice Exo70F-3 but not rice Exo70F-2 bind to Arabidopsis RIN4 in vivo.** Yeast-Two-Hybrid assay of Arabidopsis RIN4 with rice Exo70 homologs OsExo70F-2 and OsExo70F-3. Control plate for yeast growth is on the left with quadruple dropout media supplemented with X- $\alpha$ -gal and Aureobasidine A (Au A) on the right. Growth and development of blue colouration in the selection plate are both indicative of protein:protein interactions. Arabidopsis Exo70 homologs AtExo70A-1 and AtExo70B-1 were included as negative and positive controls, respectively. Exo70 proteins were fused to the GAL4 DNA binding domain, and RIN4 proteins to the GAL4 activator domain. Each experiment was repeated twice, with similar results.

These results suggest a possible role of OsExo70F-2 and OsExo70F-3 in plant immunity via association to the central regulator RIN4. This role might be altered by the binding of the AVR-Pii effector and will be investigated in the future.

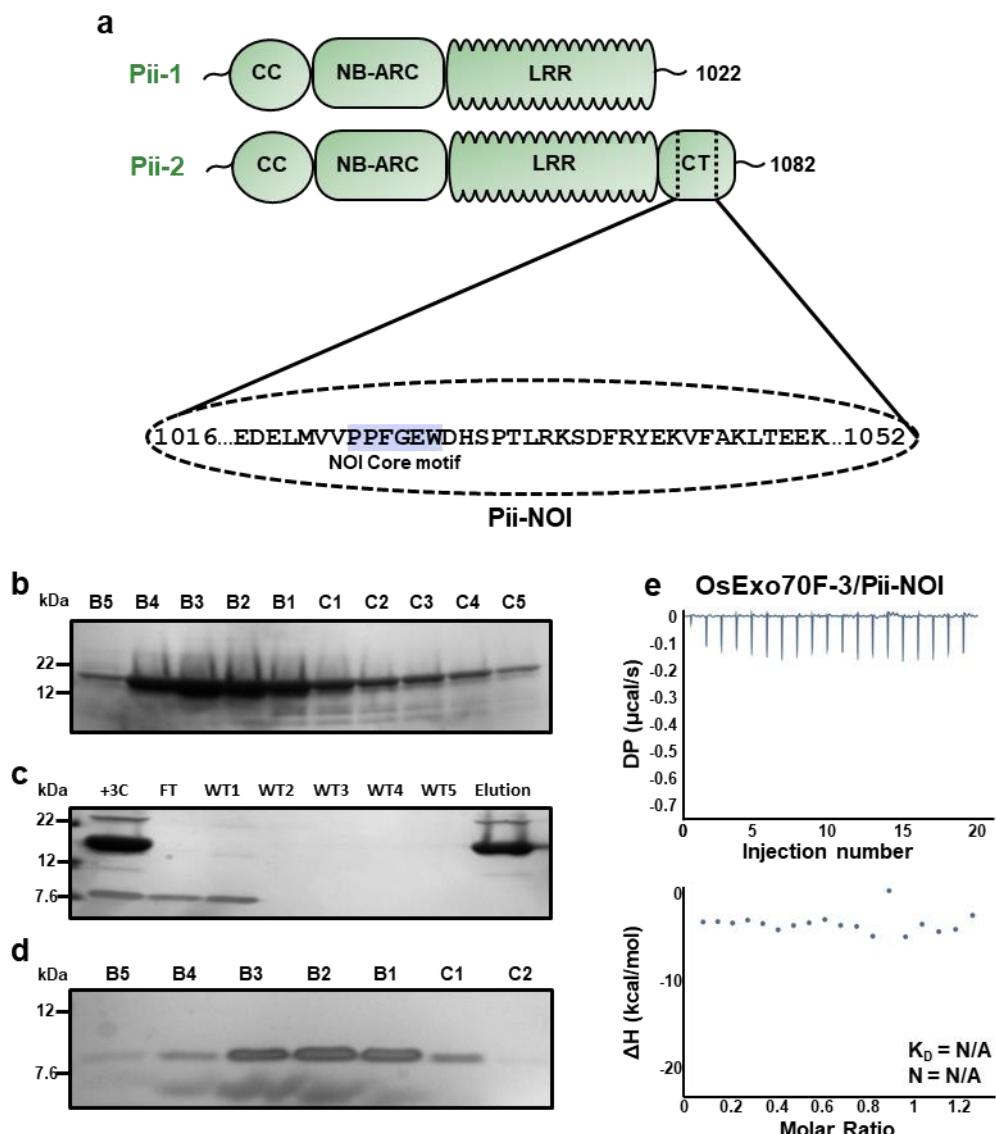
### 7.13 The Pii-NOI integrated domain does not bind OsExo70F-3 in vitro

Rice varieties with the Pii resistance pair recognize and trigger immune responses to rice blast strains carrying AVR-Pii (Takagi et al., 2017; Takagi et al., 2013). This recognition is dependent on OsExo70F-3, but not on OsExo70F-2 (Fujisaki et al., 2015). Interestingly, a member of the Pii NLR pair, Pii-2, has been reported to harbour an unconventional domain with sequence similarities to the RIN4-NOI domain at its C-terminus (Bailey et al., 2018; Kroj et al., 2016; Sarris et al., 2016).

Using Y2H assays, Dr. Koki Fujisaki reported that OsExo70F-3 binds to the NOI domain integrated in Pii-2, suggesting that this interaction may be involved in pathogen recognition (Fujisaki et al., 2017). I sort to explore the structural and biophysical basis of this interaction to better understand its putative role in recognition.

First, I used Raptor X to attempt to define the boundaries of the integrated NOI domain in Pii-2 (Källberg et al., 2012). This analysis suggested that the unconventional NOI domain in Pii-2 spans residues 1016 to 1052 (Figure 7.29a). I subsequently cloned this

region into the pOPINS3C expression vector to test for the production of Pii-NOI with an N-terminal SUMO tag.



**Figure 7.29. The NOI domain integrated in Pii-2 NLR does not interact with OsExo70F-3 in vitro.** (a) schematic representation of rice Pii-1 and Pii-2 NLRs. C-terminal region of Pii-2 (CT) contains an integrated domain containing a RIN4-NOI region (Pii-NOI). Amino acid sequence of Pii-NOI is shown in a dashed circle with the residues of the core NOI motif (PxFGxW) highlighted in ice blue. (b) Purification of SUMO:Pii-NOI after IMAC + gel filtration. Fractions from B5 to C5 were collected and treated with 3C protease. (c) HisTrap™ purification of Pii-NOI following 3C protease treatment. Protein fraction after treatment with 3C protease is shown as input. Pii-NOI was purified in the flow-through (FT) and the first wash-through with (WT1). SUMO tag eluted in the elution treatment with buffer containing imidazole. (d) Protein fraction collected after final gel filtration step were analysed by SDS-PAGE to assess the presence of the purified protein. (e) The binding of Pii-NOI to OsExo70F-3 was determined using isothermal titration calorimetry (ITC). Upper panel shows heat differences upon injection of Pii-NOI into OsExo70F-3 and the bottom panel shows integrated heats of injection (blue dots). Using MicroCal PEAQ-ITC Analysis Software no binding was detected. Panel is representative of two replicates.

Using standard protocols developed for the production of AVR-Pik proteins, I could purify 6xHisSUMO:Pii-NOI by tandem IMAC and gel filtration (**Figure 7.29b**). The SUMO tag was subsequently removed by 3C protease cleavage followed by HisTrap™ affinity purification. The Pii-NOI domain was recovered in the flow-through (FT) and wash-through (WT) from the column (**Figure 7.29c**). After a final gel filtration step, I recovered pure Pii-NOI that could be concentrated for biophysical analysis (**Figure 7.29d**).

To confirm whether OsExo70F-3 binds to Pii-NOI in vitro, I used purified proteins to test the binding by ITC. This analysis shows no binding between these two proteins (**Figure 7.29d**). As this domain seems to bind to both OsExo70F-2 and OsExo70F-3 in Y2H assays (data not shown), this discrepancy could be due to technical issues such as false positives result in the Y2H assay, or improper protein folding in solution. Alternatively, this could also indicate that the binding event underpinning AVR-Pii recognition by Pii NLRs may be more complicated than previously thought, possibly involving additional components. This requires further analysis.

## 7.14 Discussion and future experiments

To carry out their function, many pathogen effector proteins are translocated into host cells where they interact with specific factors to manipulate cellular processes and promote colonization (Win et al., 2012a). When the host targets belong to large expanded protein families, effectors need to discriminate their targets from many similar proteins. Understanding how effectors have evolved this fine specificity is an exciting topic in effector biology.

The rice blast effector AVR-Pii has a remarkable specificity, binding to only two members of the rice Exo70 family (Fujisaki et al., 2015). In this chapter, I unravelled the biophysical and structural basis of AVR-Pii binding to rice OsExo70F-2 and OsExo70F-3. The crystal structure of AVR-Pii bound to OsExo70F-2 revealed the determinants of this effector specificity at the atomic level.

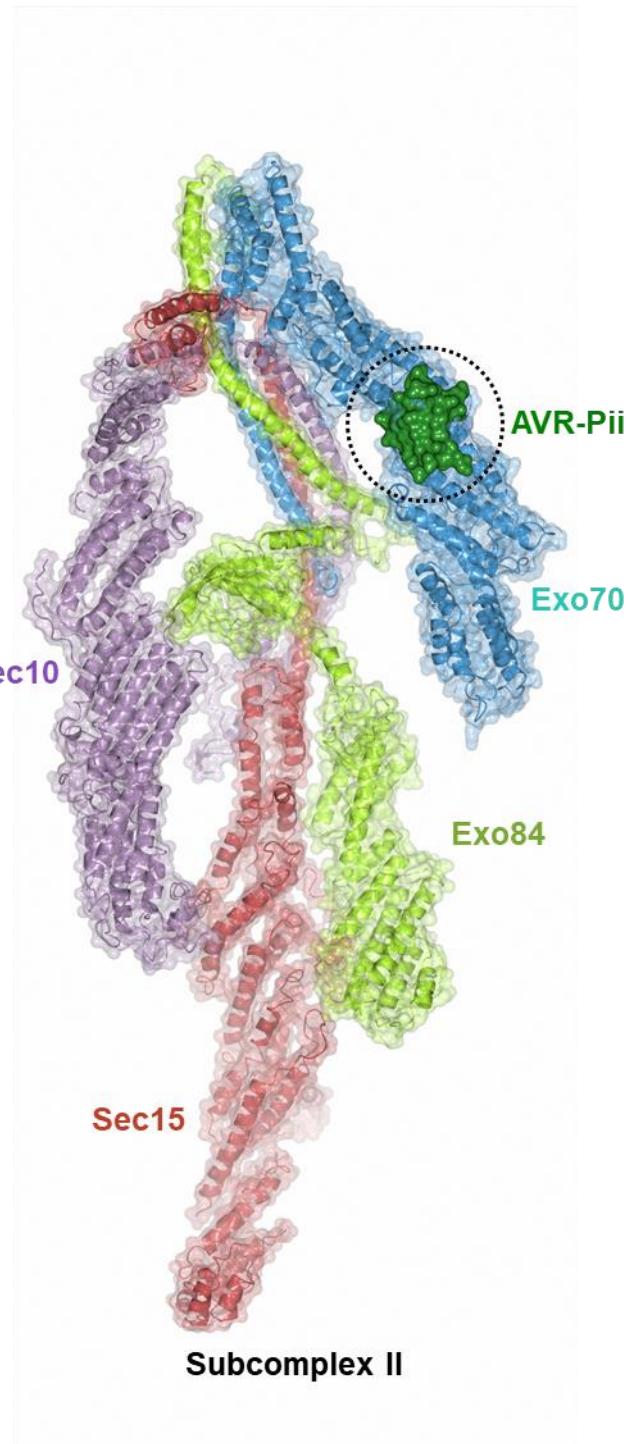
The remarkable specificity of AVR-Pii points towards a specific function of this effector. However, to date, this function remains unknown. As exocytosis is increasingly recognized as having an important role in plant-microbe interactions (Pečenková et al., 2011; Sabol et al., 2017; Stegmann et al., 2013), the interaction of AVR-Pii with exocyst components OSExo70F-2 and OsExo70F-3 may have multiple contributions to pathogenesis. In this chapter, I showed the association of both OsExo70F-2 and OsExo70F-3 with rice RIN4, linking these effector targets with a well-established component of plant immune pathways.

The exocyst is a well-known, essential component of secretion. However, there is increasing evidence to link this complex to unconventional secretion pathways and processes distinct from exocytosis, particularly in plants (Saeed et al., 2019). One interesting area is a putative role for the exocyst in the formation and release of extracellular vesicles (EVs). EVs have been shown to contain proteins and small RNAs, and are translocated to the apoplast in the presence of the pathogen to release components that may contribute to limiting infection (Baldrich et al., 2019; Cai et al., 2018). Interestingly, the exocyst has been linked to the formation of EVs in both plants and animals (Boevink, 2017; Chacon-Heszele et al., 2014; Lin et al., 2015; Regente et al., 2017; Rutter and Innes, 2017; Wang et al., 2010). Therefore, given the emerging importance of EVs in plant-microbe interactions, it is tempting to speculate that pathogen effectors might be altering EV delivery by targeting the exocyst.

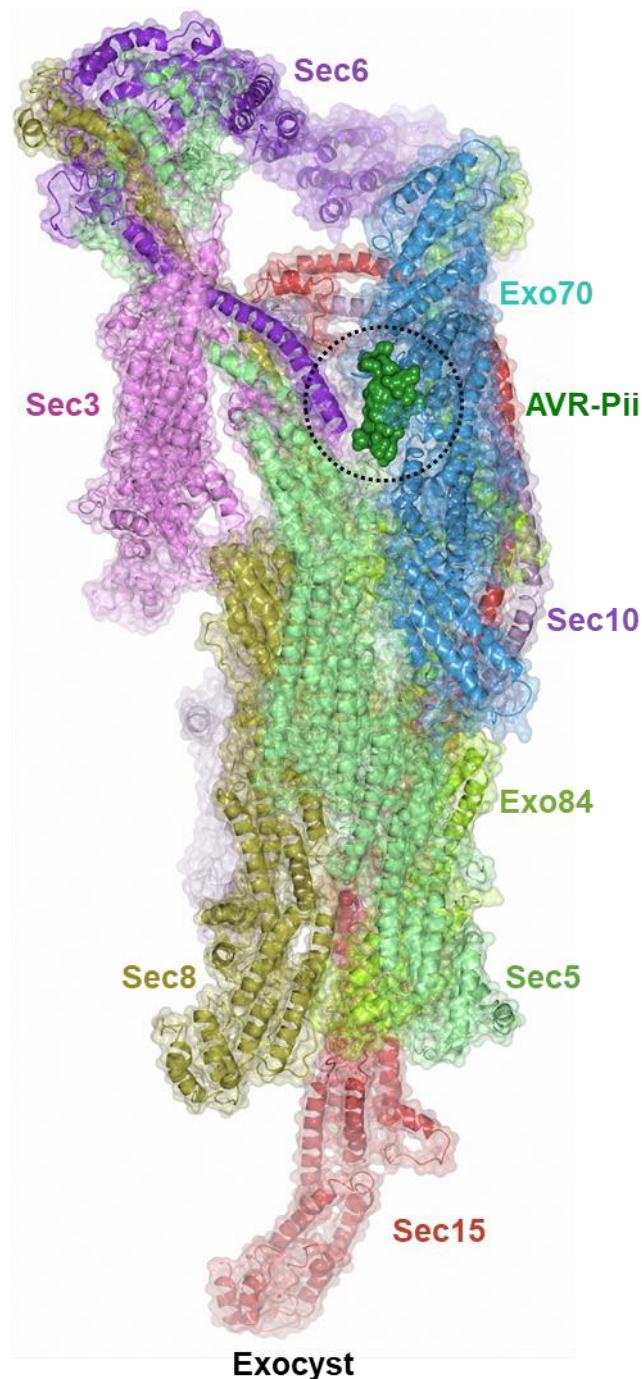
Whether AVR-Pii manipulates exocyst components equipped with OsExo70F-2 or OsExo70F-3, and the mechanism of action of the effector, will be a topic of future studies. The structural information obtained here provides interesting preliminary data to explore this mechanism. For example, by superposing the complex of OsExo70F-2/AVR-Pii on the intact exocyst, suggests that AVR-Pii would not affect the association of Subcomplex II subunits (**Figure 7.30**). Therefore, it is more likely that binding of the effector to specific Exo70 proteins would alter the interaction of this exocyst subunit with one (or more) of the many regulatory proteins that modulate exocytosis.

Interestingly, AVR-Pii could occupy a central position in the exocyst holocomplex (**Figure 7.31**), raising the possibility that the effector could alter the association of Subcomplex I and Subcomplex II in the plasma membrane, affecting the fusion of the vesicles transporting cargoes for exocytosis (Ahmed et al., 2018). Biochemical and cell biology experiments to test whether AVR-Pii affects the formation and location of exocyst complexes containing OsExo70F-2 or OsExo70F-3 subunits will be critical to explore this possibility.

Finally, the resources developed here will also advance research on the AVR-Pii recognition mechanism by the Pii NLR pair. Although the preliminary data showed no interaction between Pii-NOI domain and OsExo70F-3, the crystal structure of NOI would be a useful tool to map the determinants of recognition specificity.



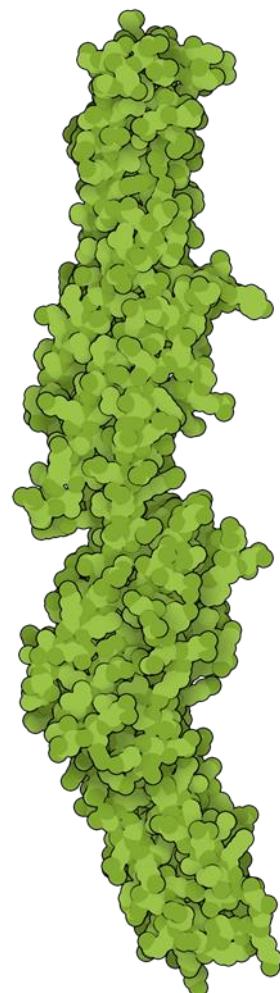
**Figure 7.30. AVR-Pii effector is predicted to sit in the outside of exocyst subcomplex II.** Schematic representations of the position of AVR-Pii in the exocyst subcomplex II. Overall structures of yeast Exo70, Exo84, Sec10 and Sec15 are shown in cartoon representation with the molecular surface and coloured as labelled. Rice blast AVR-Pii effector is superimposed in the structure and represented as solid surface coloured in dark green. Location of AVR-Pii in the overall structure is highlighted in a dashed circle.



**Figure 7.31. Structure superposition of AVR-Pii effector in the exocyst holocomplex.** Schematic representations of the position of AVR-Pii in the exocyst holocomplex. Overall structures of yeast Sec3, Sec5, Sec6, Sec8, Sec10, Sec15, Exo70 and Exo84 are shown in cartoon representation with the molecular surface and coloured as labelled. Rice blast AVR-Pii effector is superimpose in the structure and represented as solid surface coloured in dark green. Location of AVR-Pii in the overall structure is highlighted in a dashed circle.

# 8

## Discussion



**OsExo70F-2**



# 8. Discussion

## 8.1 The quest for engineered disease resistance

The development of synthetic disease resistance in crops has been a major goal since modern molecular biology techniques were first employed in agriculture. To this end, a major focus over the last two decades has been the engineering of intracellular plant immune receptors, however this has proved very challenging. The recent discovery of effector-binding domains derived from pathogen targets that are integrated into NLR receptors represented a new paradigm in plant immunity (Cesari et al., 2014a). These domains have rapidly become central to the understanding and manipulation of pathogen recognition, representing some of the most promising elements in the toolbox for achieving extended disease resistance (Ellis, 2016).

In this thesis, I used the integrated HMA domain of the rice Pik NLR pair to advance the understanding of several aspects of NLR biology. In chapter three I uncovered the basis of the differential effector recognition specificities in allelic Pik NLRs. This work mechanistically linked the binding affinities of Pik-HMA alleles to AVR-Pik effector variants and immune recognition (De la Concepcion et al., 2018). Importantly, this enabled the identification of the structural determinants underpinning extended recognition in allelic NLRs. In chapter four this was used to successfully engineer the integrated HMA domain in the NLR allele *Pikp* to expand effector recognition (De la Concepcion et al., 2019). This represents a novel and promising approach to engineer disease resistance.

Furthermore, the results of chapters five and six demonstrate how NLR pairs specialize during evolution to enhance immune responses to effectors. These results revealed a set of subtle changes in the Pik NLR that greatly increase the cell death outcomes after effector recognition. Moreover, the specific determinants of NLR pair regulation of immune responses by the integrated HMA domain were also identified.

Altogether, these findings have important implications for future engineering efforts as they can guide the design of receptors with improved capabilities for effector binding and immune activation.

### 8.1.1 Multiple routes for NLR engineering

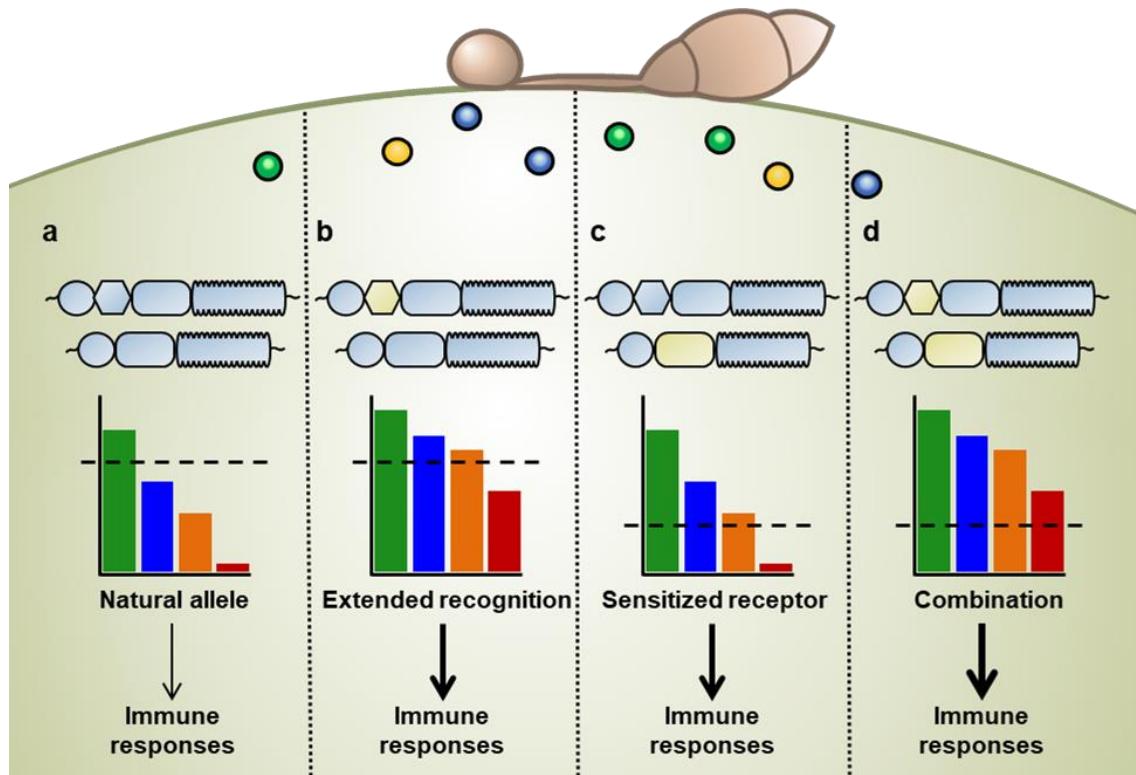
The structural information generated in this thesis provides unprecedent insights into how effector binding affinities underpin recognition specificities in allelic NLRs (**Figure**

**8.1a).** This work also showed that binding of effectors to integrated domains is not in itself sufficient to activate immune signalling but requires a certain binding affinity to trigger immunity. Therefore, an NLR acts as a switch that integrates different signals but only triggers immune signalling when a certain activation threshold is reached (**Figure 8.1a**).

This work showed how subtle structural changes can alter binding affinities between NLRs and effectors and therefore the activation of immune responses. Indeed, this enabled the manipulation of the *Pikp* NLR to increase binding to AVR-Pik effectors (**Figure 8.1b**). The increase in binding affinities translated to a gain of effector recognition in *N. benthamiana* (**Figure 8.1b**). This proof-of-concept represents an important stepping stone towards the rational engineering of NLR receptors.

The study of the relationship between allelic NLR pairs uncovered how very subtle changes in a receptor can lower the activation threshold (**Figure 8.1c**). These sensitized or ‘trigger happy’ receptors have been previously found in mutagenesis studies but rarely led to disease resistance (Farnham and Baulcombe, 2006; Giannakopoulou et al., 2015; Harris et al., 2013; Segretin et al., 2014; Stirnweis et al., 2014b). Here, I report how this process of NLR sensitization has occurred during evolution, offering the possibility to study it in the natural, regulated context.

I further explored this natural context to understand the role that the integrated HMA domains play in the activation of immune responses. By dissecting the mutational landscape of the *Pikm*-HMA domain using chimeras and point mutations, I uncovered residues and positions that maintain a controlled regulation of the sensitized *Pikm*-2 receptor. This opens the door to combining mutations to enhance binding affinities with mutations that lower the activation threshold (**Figure 8.1d**). Altogether, the features revealed by this work will inform further engineering efforts using HMA domains and Pik NLRs as a chassis to achieve disease resistance.



**Figure 8.1. Different strategies for engineering Pik NLRs.** (a) Pik-HMA binds *M. oryzae* effectors with different affinities (represented by the height of the coloured bars). Only some of them reach the activation threshold (represented by the dashed line) that leads to immune responses. (b) Mutations in the HMA domain extended NLR recognition by increasing binding to *M. oryzae* effectors, which in turn increased the strength of immune responses (depicted by the thickness of the arrow). (c) By natural variation or mutation, some Pik receptors have a lower threshold of activation. (d) The combination of extended recognition mutations and the use of sensitized receptors can lead to wider and increased immune responses to effectors. Figure has been adapted from (Bialas et al., 2018; Giannakopoulou et al., 2016).

### 8.1.2 Extending effector recognition beyond nature

Whilst the data in this thesis are an important step forward in NLR engineering, the extended recognition achieved by this proof-of-concept was restricted to reproducing that of a known natural allele. The next step would be to expand NLR recognition to include effectors beyond the natural repertoire.

Surprisingly, the HMA domain engineered in chapter four showed a gain of binding to an AVR-Pik variant that is not recognized by any Pik allele, AVR-PikC. Neither the Pikp or Pikm HMA domains investigated in chapter three display any binding to this effector, however the combination of favourable interfaces of both alleles into Pikp-HMA<sup>NK-KE</sup> resulted in a substantial gain of binding to this effector in vitro (De la Concepcion et al., 2019).

This binding did not translate into cell death responses in *N. benthamiana*, maybe because did not reach a threshold for activation. However, the association was strong enough to obtain the structure of the complex between these two proteins, enabling future engineering of the NLR to increase the binding and gain immune responses to an effector currently not recognized in nature.

Another future approach might be the combination of this extended recognition mutant with NLR mutations that reduce the threshold of activation, as discussed above.

A further possibility for engineering could be utilising the Pik combinations and mutants that triggered immune responses in the absence of effectors, which were discovered during investigation of the Pik pair co-evolution. If it is possible to produce these proteins in response to pathogens and pests, this will trigger efficient immune-like cell death responses that will not be dependent on effectors. A similar approach has been previously used to achieve disease resistance to several pathogens (Xu et al., 2017a; Xu et al., 2017b). Pik pair NLRs can trigger HR responses in rice and *N. benthamiana*, and possibly more plant species. Using such strategies for controlled expression of this pair only under the attack of pathogens or pests could be a translatable way of providing synthetic disease resistance to different crops, independently of effector recognition. This approach is certainly worth investigating in the future.

### 8.1.3 Future experiments and limitations of structure-guided engineering

The work discussed above still carries certain limitations that need to be overcome in the future, given that the extended effector recognition and subsequent increase in cell death responses were solely assessed in the heterologous *N. benthamiana* system.

Although Pik-mediated cell death assays in *N. benthamiana* seem to correlate with rice disease resistance specificities (De la Concepcion et al., 2018), both fungal effector and rice NLRs are overexpressed in these assays. In naturally occurring infections, effectors are locally delivered by the fungus at the fungal/host interface in much smaller amounts. Likewise, plant NLRs often present very low basal levels of expression. Therefore, to achieve the ultimate goal of application in the field, it is essential to demonstrate that the structure-guided extended recognition developed in this thesis translates into disease resistance in rice plants. At the time of submission of this thesis, our collaborators at Iwate Biotechnology Research Centre (Japan) had begun the generation of transgenic rice carrying Pik receptors with extended binding and recognition for AVR-Pik. These plants will be further tested in susceptibility assays against rice blast strains harbouring different AVR-Pik variants.

When considering future applications, it is also important to note that the structure-guided engineering approach described here may be limited to cases where effectors directly bind to the receptor and structural information is available.

## **8.2 The study of integrated domains: mechanistic research with evolutionary perspective**

The study of biological processes often results in a conceptual gap between mechanistic and evolutionary research. While mechanistic research often lacks an evolutionary perspective of the biological processes being studied, evolutionary research tends not to include the diversity of concepts that emerge from mechanistic approaches (Upson et al., 2018).

The field of molecular plant-microbe interactions is particularly suitable for both mechanistic and evolutionary research. Recognition of pathogen effectors by the plant immune system has vast evolutionary implications, as this interface between plants and microbes presents rapid co-evolutionary dynamics (Allen et al., 2004; Kanzaki et al., 2012). At the same time, the interaction between plants and microbes has largely benefitted from mechanistic research to understand the molecular processes underpinning effector recognition (Guo et al., 2018b; Maqbool et al., 2015; Ortiz et al., 2017) and immune responses (Wang et al., 2019a; Wang et al., 2019b).

During this thesis, I worked extensively at the interface between mechanistic and evolutionary research to obtain a holistic understanding of NLR integrated domains. This approach led to the most extensive mechanistic investigation of the arms race co-evolution between a pathogen effector and a cognate resistance gene reported to date. By using structural biology and biophysical approaches I provided understanding of the events that lead to loss of pathogen resistance, underpinned by reduction of binding to polymorphic effectors (De la Concepcion et al., 2018).

### **8.2.1 Co-evolution of NLRs as a Bateson-Dobzhansky-Muller evolution**

A recently established model in plant immunity is that NLRs often works together in pairs or networks and oligomerize in sophisticated signalling structures upon activation. Another recent paradigm is that combinations of NLRs that have not evolved together can lead to constitutive immune responses, causing a fitness cost in hybrids, known as hybrid necrosis (Bomblies and Weigel, 2007).

These concepts suggest NLRs are probably in a constant co-evolution to optimise and maintain immune regulation. However, not many cases of this hypothesis have been proved to date. Pik NLR combinations that trigger constitutive immune responses represent an excellent system to shed light on the evolution of plant NLRs. This enabled me to understand how paired sensor and helper NLR pairs specialize, as detailed in chapter five. Furthermore, this revealed how very subtle changes that occur during evolution can have major effects on how NLRs relate to each other.

Overall, these findings linked plant immunity to interesting evolutionary concepts. In the resulting model that emerges from this research, differential recognition specificities will firstly drive NLR evolution in different directions. These will then become mismatched and the gene flow between plant populations with a different NLR sets becomes restricted (Bomblies et al., 2007; Bomblies and Weigel, 2007; Yamamoto et al., 2010). This supposes a new addendum to the Bateson-Dobzhansky-Muller model of speciation (Dobzhansky, 1937).

### **8.2.2 Integrated domains shape the evolution and regulation of NLRs**

A conceptual paradox also emerges from the findings of chapters five. If NLRs can easily become autoactive as a result of small changes, how is possible that the integration of an atypical and rapidly-evolving domain into their architecture is evolutionarily favoured? Do integrated domains co-adapt to their acceptor NLRs to compensate for mismatches?

To date, it is not clear how these domains integrate in NLRs and what their relationship with other domains is (Baggs et al., 2017). However, by investigating the integrated HMA domain of Pik I could show how this domain co-evolved with the acceptor NLR pair, acquiring regulatory functions distinct from the binding of pathogen effectors.

Interestingly, the differential regulatory effect between integrated HMA domains turned out to be the basis of the autoimmune phenotypes observed in NLR mismatches. This shows how domain integration influences the evolution of NLR receptors, bringing not only new capabilities to the receptor, but also negative consequences to fitness that are reduced by the acquisition of regulatory roles.

The model described in this thesis further explains the interesting observation that all validated integrated domains have been found in paired NLRs that are placed in co-regulatory modules which share a promotor region (Cesari et al., 2014a). By spatially regulating these NLRs with integrated domains in the genome, there is a reduction in the possibility of misregulation. This would solve the paradox between the selection of the integration domain and the associated fitness cost. Furthermore, this would offer an

explanation as to why paired NLRs are overrepresented in some plant genomes, such as rice (Stein et al., 2018; Wang et al., 2019c).

### 8.2.3 Autoimmunity as a research tool

The use of constitutively active immune receptors as a research tool is starting to be explored in the field of NLR biology. For example, mutational screenings for suppressors of constitutive immune responses have been used to find genes involved in immune signalling (Johnson et al., 2012; Liu et al., 2017), and to inform the features of NLR pair interactions (Newman et al., 2019). NLR research using autoactive NLRs has the advantage of simplifying the complex requirements of immune activation by removing the variable of the effector. This approach also has the advantage of rendering full receptor activation, whilst relying solely on effector recognition can provide a mixture of active and inactive receptors.

In chapter five, I discovered some Pik NLR combinations and mutants that trigger constitutive cell death in the absence of the effector. Some of these have significantly higher cell death responses than those triggered after recognition of AVR-Pik, which is possibly because of a higher number of active receptors.

I use these mutants as a tool to demonstrate that effector recognition and Pik NLR activation are separable traits, as shown for other NLR pairs (Newman et al., 2019). Furthermore, these mutants enabled me to investigate the role of integrated HMA domains beyond effector binding and how these functions are underpinned by structural changes. This domain represses the activation of immune responses mediated by Pik-2 until the effector is bound, a similar role has been also found for the WRKY domain of RRS1/RPS4 NLR pair (Ma et al., 2018).

Taking advantage of constitutively active Pik, I also expanded the knowledge of the Pik NLR activation. Sensor and Helper Pik NLRs form a pre-activation complex without the need for ADP/ATP exchange. Activation of immune responses seems to rearrange the formation of this complex, likely implicating a change of sensor/helper stoichiometry, as for NAIP/NLRC4 inflammasomes (Hu et al., 2015; Zhang et al., 2015a). This rearrangement is dependent on ADP/ATP exchange and has been fine-tuned during the evolutionary process, as depicted in the competition assays in the Pik pair. Future native PAGE analyses of autoactive Pik may help to reveal these stoichiometric changes.

Experiments with autoactive Pik will probably help to unravel the requirements of NLR activation. For example, combining these mutations with different domain truncations can shed light on the minimal requirements for Pik NLR activation.

## 8.3 Effectors as tools for biological research

Effector biology is a field of research that is fundamental to the understanding of pathogen virulence and engineering extended pathogen recognition (Bialas et al., 2018). Hundreds of putative effectors are contained in dynamic genome regions of the most destructive plant pathogens (Dean et al., 2005; Haas et al., 2009; Kämper et al., 2006). These are fast-evolving proteins, which share little or no similarity with most of the other known functional protein domains (Franceschetti et al., 2017).

In chapter 7, I uncovered how the *M. oryzae* effector AVR-Pii binds to its host target. To date, this is one of the first structures of a rice blast effectors bound to a host protein and it revealed the structural basis of binding selectivity, which might be used as a research tool in plant immunity and other fields.

### 8.3.1 The new protein fold of *M. oryzae* AVR-Pii enables the discovery of new fungal effectors and integrated domains

The crystal structure of the AVR-Pii effector revealed a new protein fold for fungal effectors. Four residues forming a zinc finger are largely sustaining this fold, which I coined Zinc finger fold, or ZiF. As the crystal structure provided the structural elements sustaining the structure of the effector, this finding can easily inform HMM searches in the *M. oryzae* (and possibly other fungal pathogens) candidate effector list to investigate a putative effector family sharing the ZiF fold (de Guillen et al., 2015; Win et al., 2012b).

Identifying other effectors with this fold will open new research avenues. It would be interesting to investigate whether all effectors with the ZiF fold bind to plant Exo70 proteins or have a diverse set of targets. As the crystal structure of the complex between OsExoF-2 and AVR-Pii identified the two amino acids underpinning this binding, this could be confidently inferred from the effector sequence. Finding ZiF effector variants across multiple *Magnaporthe* genomes with different hosts may provide new specificities to different plant Exo70 proteins, which eventually can inform engineering approaches to extend Pii resistance to other members of a potential ZiF effector family.

Furthermore, by creating a ZiF effector library it will be possible to test their interaction with any of the multiple Exo70 domains that have been found integrated in plant NLRs (Bailey et al., 2018; Brabham et al., 2018; Kroj et al., 2016; Sarris et al., 2016; Steuernagel et al., 2018). This would rapidly identify resistance gene candidates and their cognate effectors. As I have already adapted the techniques developed during research on Pik resistance to the study of AVR-Pii interactions with Exo70 proteins, they

could be quickly applied to dissect the interaction between ZiF effectors and Exo70 domains integrated in NLR receptors.

Yet another interesting finding obtained from the structure of AVR-Pii bound to its host Exo70 target is that only the C-terminal half of the effector is present, indicating that the N-terminus does not participate in the binding and might be disordered. Interestingly, the amino acids right before those visible in the structure (LEAR) were found to be conserved in preliminary sequence alignments during the cloning of AVR-Pii (Yoshida et al., 2009). Does the LEAR motif of AVR-Pii reassemble the RXLR motif of oomycete effectors (Jiang et al., 2008; Whisson et al., 2007)? The intact mass results of the effector confirmed the purified protein matched the expected size and would not support that a cleavage event is occurring, as in the case of the RXLR motif in oomycete effectors (although this occurs during the delivery by the pathogen). Either way, the N-terminal region might be involved in effector delivery by the fungus, a process that is still not very well understood (Zhang and Xu, 2014). This hypothesis could be easily tested by screening for N-terminal region mutants (which should not affect binding to host OsExo70F-3) that become virulent in rice plants harbouring Pii resistance.

### 8.3.2 Applications of effector tools

Conceptually, intracellular effectors are molecules produced by very different pathogens (viruses, bacteria, oomycete, fungi or even animals) that are designed to carry out their functions inside of a different organism (Franceschetti et al., 2017; Presti et al., 2015). This makes them some of the most fascinating proteins in biology. Moreover, as they display an outstanding functional diversity (Franceschetti et al., 2017), research in effector biology can potentially provide major findings and tools with wide applications. A representative example is bacterial Transcriptional Activator-Like (TAL) effectors from *Xanthomonas* (Bogdanove et al., 2010). As they bind specific DNA sequences and the specificity is tuneable, this is a perfect system to guide nucleases for gene editing purposes (Bogdanove and Voytas, 2011). Nowadays, they represent major tools for genome editing in human cells, with promising applications in the biomedical field (Ding et al., 2013).

Some pathogen effectors have proven to be very specific in their interaction with the host proteome, associating only with a small subset of targets within large protein families (Dagdas et al., 2016; Win et al., 2012a). The discriminatory power of pathogen effectors is an exciting topic for effector biology research that can lead to the development of new

molecular tools. AVR-Pii is a striking case of specificity as IP-MS showed association with only 2 out of 47 Exo70 proteins of the rice proteome (Fujisaki et al., 2015).

By obtaining the crystal structure with its host target, I revealed the basis of the strong target specificity of AVR-Pii. With this information, this effector may be re-tooled for use as a molecular probe for understanding the exocyst functions. For example, if the function of AVR-Pii involves inhibiting the exocyst by binding specifically to the Exo70 subunit, altering effector binding specificities in AVR-Pii could potentially create a tuneable inhibitor with allelic specificity. Contrary to the majority of inhibitors exhibiting broad spectrum activity (Huang et al., 2019; Zhang et al., 2016), this would circumvent pleiotropic effects and enable the study of the functional diversity of exocysts.

Moreover, Exo70 proteins are involved in an increasing numbers of plant processes distinct from exocytosis, with each one involving a particular set of interactions with other proteins (Saeed et al., 2019). However, mechanistic approaches to the study of the plant exocyst have been largely neglected. In this context, the purification protocols and biophysical assays developed in this thesis may help to fill this research niche.

### 8.4 Closing remarks

This study investigating the structural and evolutionary biology of NLR integrated domains has led to the identification of the mechanistic basis of pathogen recognition and the development of a structure-guided approach to engineer disease resistance. This has been one of the first times that an immune receptor is successfully re-tooled to have an expanded recognition to pathogen effectors, having great potential to inform other engineering efforts to obtain disease resistance. The characterisation of the Pik NLR pair evolutionary relationship has led to the identification of a mechanism of NLR incompatibility and the identification of a regulatory role for the Pik-HMA integrated domain. This can potentially explain cases of hybrid necrosis, serving as an example of how disease resistance maybe linked with plant speciation. The extensive set of mutations generated here will be of further use for research into the mechanism of NLR activation and will also inform NLR engineering. Finally, the extension of my research to a different effector recognized by rice NLRs led to the discovery of a potentially new class of effectors and the structural basis of a new effector/target interface that will be focus of future studies.

# References

**Adachi, H., Derevnina, L., and Kamoun, S.** (2019). NLR singletons, pairs, and networks: evolution, assembly, and regulation of the intracellular immunoreceptor circuitry of plants. *Curr Opin Plant Biol* 50, 121-131.

**Ade, J., DeYoung, B.J., Golstein, C., and Innes, R.W.** (2007). Indirect activation of a plant nucleotide binding site-leucine-rich repeat protein by a bacterial protease. *Proc Natl Acad Sci USA* 104, 2531-2536.

**Afzal, A.J., da Cunha, L., and Mackey, D.** (2011). Separable fragments and membrane tethering of *Arabidopsis* RIN4 regulate its suppression of PAMP-triggered immunity. *Plant Cell* 23, 3798-3811.

**Afzal, A.J., Kim, J.H., and Mackey, D.** (2013). The role of NOI-domain containing proteins in plant immune signaling. *BMC Genomics* 14, 327.

**Ahmed, S.M., Nishida-Fukuda, H., Li, Y., McDonald, W.H., Grdinaru, C.C., and Macara, I.G.** (2018). Exocyst dynamics during vesicle tethering and fusion. *Nat Commun* 9, 5140.

**Allen, R.L., Bittner-Eddy, P.D., Grenville-Briggs, L.J., Meitz, J.C., Rehmany, A.P., Rose, L.E., and Beynon, J.L.** (2004). Host-parasite coevolutionary conflict between *Arabidopsis* and downy mildew. *Science* 306, 1957-1960.

**Arasaki, K., Kimura, H., Tagaya, M., and Roy, C.R.** (2018). *Legionella* remodels the plasma membrane-derived vacuole by utilizing exocyst components as tethers. *J Cell Biol* 217, 3863-3872.

**Arora, S., Steuernagel, B., Gaurav, K., Chandramohan, S., Long, Y., Matny, O., Johnson, R., Enk, J., Periyannan, S., Singh, N., et al.** (2019). Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nat Biotechnol* 37, 139-143.

**Ashikawa, I., Hayashi, N., Abe, F., Wu, J., and Matsumoto, T.** (2012). Characterization of the rice blast resistance gene Pik cloned from Kanto51. *Mol Breed* 30, 485-494.

**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.

**Axtell, M.J., and Staskawicz, B.J.** (2003). Initiation of RPS2-specified disease resistance in *Arabidopsis* is coupled to the AvrRpt2-directed elimination of RIN4. *Cell* 112, 369-377.

**Baggs, E., Dagdas, G., and Krasileva, K.V.** (2017). NLR diversity, helpers and integrated domains: making sense of the NLR IDentity. *Curr Opin Plant Biol* 38, 59-67.

**Bailey, P.C., Schudoma, C., Jackson, W., Baggs, E., Dagdas, G., Haerty, W., Moscou, M., and Krasileva, K.V.** (2018). Dominant integration locus drives continuous diversification of plant immune receptors with exogenous domain fusions. *Genome Biol* 19, 23.

**Baldrich, P., Rutter, B.D., Karimi, H.Z., Podicheti, R., Meyers, B.C., and Innes, R.W.** (2019). Plant extracellular vesicles contain diverse small rna species and are enriched in 10- to 17-nucleotide “tiny” RNAs. *Plant Cell* 31, 315-324.

**Barragan, C.A., Wu, R., Kim, S.T., Xi, W., Habring, A., Hagmann, J., Van de Weyer, A.L., Zaidem, M., Ho, W.W.H., Wang, G., et al.** (2019). RPW8/HR repeats control NLR activation in *Arabidopsis thaliana*. *PLoS Genet* 15, e1008313.

**Bebber, D.P., and Gurr, S.J.** (2015). Crop-destroying fungal and oomycete pathogens challenge food security. *Fungal Genet Biol* 74, 62-64.

**Belkhadir, Y., Nimchuk, Z., Hubert, D.A., Mackey, D., and Dangl, J.L.** (2004). *Arabidopsis* RIN4 negatively regulates disease resistance mediated by RPS2 and RPM1 downstream or independent of the NDR1 signal modulator and is not required for the virulence functions of bacterial Type III effectors AvrRpt2 or AvrRpm1. *Plant Cell* 16, 2822-2835.

**Bendahmane, A., Farnham, G., Moffett, P., and Baulcombe, D.C.** (2002). Constitutive gain-of-function mutants in a nucleotide binding site-leucine rich repeat protein encoded at the Rx locus of potato. *Plant J* 32, 195-204.

**Bent, A.F., Kunkel, B.N., Dahlbeck, D., Brown, K.L., Schmidt, R., Giraudat, J., Leung, J., and Staskawicz, B.J.** (1994). RPS2 of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science* 265, 1856-1860.

**Bentham, A.R., Zdrzalek, R., De la Concepcion, J.C., and Banfield, M.J.** (2018). Uncoiling CNLs: Structure/function approaches to understanding CC domain function in plant NLRs. *Plant and Cell Physiology* 59, 2398-2408.

**Bergelson, J., Kreitman, M., Stahl, E.A., and Tian, D.** (2001). Evolutionary dynamics of plant R-Genes. *Science* 292, 2281-2285.

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

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

**Bialas, A., Zess, E.K., De la Concepcion, J.C., Franceschetti, M., Pennington, H.G., Yoshida, K., Upson, J.L., Chanclud, E., Wu, C.H., Langner, T., et al.** (2018). Lessons in effector and NLR biology of plant-microbe systems. *Mol Plant-Microbe Interact* 31, 34-45.

**Bodemann, B.O., Orvedahl, A., Cheng, T., Ram, R.R., Ou, Y.H., Formstecher, E., Maiti, M., Hazelett, C.C., Wauson, E.M., Balakireva, M., et al.** (2011). RalB and the exocyst mediate the cellular starvation response by direct activation of autophagosome assembly. *Cell* 144, 253-267.

**Boevink, P.C.** (2017). Exchanging missives and missiles: the roles of extracellular vesicles in plant-pathogen interactions. *J Exp Bot* 68, 5411-5414.

**Bogdanove, A.J., Schornack, S., and Lahaye, T.** (2010). TAL effectors: finding plant genes for disease and defense. *Curr Opin Plant Biol* 13, 394-401.

**Bogdanove, A.J., and Voytas, D.F.** (2011). TAL effectors: Customizable proteins for DNA targeting. *Science* 333, 1843-1846.

**Bomblies, K., Lempe, J., Epple, P., Warthmann, N., Lanz, C., Dangl, J.L., and Weigel, D.** (2007). Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol* 5, e236.

**Bomblies, K., and Weigel, D.** (2007). Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat Rev Genet* 8, 382-393.

**Bonardi, V., and Dangl, J.L.** (2012). How complex are intracellular immune receptor signaling complexes? *Front Plant Sci* 3, 237.

**Bos, J.I., Kanneganti, T.D., Young, C., Cakir, C., Huitema, E., Win, J., Armstrong, M.R., Birch, P.R., and Kamoun, S.** (2006). The C-terminal half of *Phytophthora infestans* RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in *Nicotiana benthamiana*. *Plant J* 48, 165-176.

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

**Boyes, D.C., Nam, J., and Dangl, J.L.** (1998). The *Arabidopsis thaliana* RPM1 disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the hypersensitive response. *Proc Natl Acad Sci USA* 95, 15849-15854.

**Brabham, H.J., Hernández-Pinzón, I., Holden, S., Lorang, J., and Moscou, M.J.** (2018). An ancient integration in a plant NLR is maintained as a *trans*-species polymorphism. *bioRxiv*, 239541.

**Broz, P., and Dixit, V.M.** (2016). Inflammasomes: mechanism of assembly, regulation and signalling. *Nat Rev Immunol* 16, 407.

**Burdett, H., Bentham, A.R., Williams, S.J., Dodds, P.N., Anderson, P.A., Banfield, M.J., and Kobe, B.** (2019). The plant "Resistosome": Structural insights into immune signaling. *Cell Host Microbe* 26, 193-201.

**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.

**Cao, J., Schneeberger, K., Ossowski, S., Günther, T., Bender, S., Fitz, J., Koenig, D., Lanz, C., Stegle, O., Lippert, C., et al.** (2011). Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* 43, 956.

**Carter, M.E., Helm, M., Chapman, A.V.E., Wan, E., Restrepo Sierra, A.M., Innes, R.W., Bogdanove, A.J., and Wise, R.P.** (2019). Convergent evolution of effector protease recognition by *Arabidopsis* and barley. *Mol Plant-Microbe Interact* 32, 550-565.

**Casey, L.W., Lavrencic, P., Bentham, A.R., Cesari, S., Ericsson, D.J., Croll, T., Turk, D., Anderson, P.A., Mark, A.E., Dodds, P.N., et al.** (2016). The CC domain structure

from the wheat stem rust resistance protein Sr33 challenges paradigms for dimerization in plant NLR proteins. Proc Natl Acad Sci USA 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., Bernoux, M., Moncquet, P., Kroj, T., and Dodds, P.N.** (2014a). A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. Front Plant Sci 5, 606.

**Cesari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y., Shimamoto, K., Dodds, P., Terauchi, R., and Kroj, T.** (2014b). The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. EMBO J 33, 1941-1959.

**Cesari, S., Moore, J., Chen, C., Webb, D., Periyannan, S., Mago, R., Bernoux, M., Lagudah, E.S., and Dodds, P.N.** (2016). Cytosolic activation of cell death and stem rust resistance by cereal MLA-family CC–NLR proteins. Proc Natl Acad Sci USA 113, 10204-10209.

**Cesari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A., Rivas, S., Alaux, L., Kanzaki, H., Okuyama, Y., et al.** (2013). The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. Plant Cell 25, 1463-1481.

**Chacon-Heszele, M.F., Choi, S.Y., Zuo, X., Baek, J.I., Ward, C., and Lipschutz, J.H.** (2014). The exocyst and regulatory GTPases in urinary exosomes. Physiol Rep 2.

**Chae, E., Bomblies, K., Kim, S.T., Karelina, D., Zaidem, M., Ossowski, S., Martin-Pizarro, C., Laitinen, R.A., Rowan, B.A., Tenenboim, H., et al.** (2014). Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. Cell 159, 1341-1351.

**Chae, E., Tran, D.T., and Weigel, D.** (2016). Cooperation and conflict in the plant immune system. PLoS Path 12, e1005452.

**Chaipanya, C., Telebancio-Yanoria, M.J., Quime, B., Longya, A., Korinsak, S., Korinsak, S., Toojinda, T., Vanavichit, A., Jantasuriyarat, C., and Zhou, B.** (2017). Dissection of broad-spectrum resistance of the Thai rice variety Jao Hom Nin conferred by two resistance genes against rice blast. Rice 10, 18-18.

**Chapman, A.M., and McNaughton, B.R.** (2016). Scratching the surface: Resurfacing proteins to endow new properties and function. Cell Chem Biol 23, 543-553.

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

**Cheng, T.C., Hong, C., Akey, I.V., Yuan, S., and Akey, C.W.** (2016). A near atomic structure of the active human apoptosome. eLife 5, e17755.

**Choi, J.Y., Platts, A.E., Fuller, D.Q., Hsing , Y.-I., Wing, R.A., and Purugganan, M.D.** (2017). The rice paradox: Multiple origins but single domestication in Asian rice. *Mol Biol Evol* 34, 969-979.

**Chou, S., Krasileva, K.V., Holton, J.M., Steinbrenner, A.D., Alber, T., and Staskawicz, B.J.** (2011). *Hyaloperonospora arabidopsis* ATR1 effector is a repeat protein with distributed recognition surfaces. *Proc Natl Acad Sci USA* 108, 13323-13328.

**Chui, A.J., Okondo, M.C., Rao, S.D., Gai, K., Griswold, A.R., Johnson, D.C., Ball, D.P., Taabazuing, C.Y., Orth, E.L., Vittimberga, B.A., et al.** (2019). N-terminal degradation activates the NLRP1B inflammasome. *Science* 364, 82-85.

**Chung, E.-H., da Cunha, L., Wu, A.-J., Gao, Z., Cherkis, K., Afzal, A.J., Mackey, D., and Dangl, J.L.** (2011). Specific threonine phosphorylation of a host target by two unrelated Type III effectors activates a host innate immune receptor in plants. *Cell Host Microbe* 9, 125-136.

**Chung, E.-H., El-Kasmi, F., He, Y., Loehr, A., and Dangl, Jeffery L.** (2014). A plant phosphoswitch platform repeatedly targeted by Type III effector proteins regulates the output of both tiers of plant immune receptors. *Cell Host Microbe* 16, 484-494.

**Civáň, P., Craig, H., Cox, C.J., and Brown, T.A.** (2015). Three geographically separate domestications of Asian rice. *Nat Plants* 1, 15164.

**Clark, R.M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., Hu, T.T., Fu, G., Hinds, D.A., et al.** (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* 317, 338-342.

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

**Couto, D., and Zipfel, C.** (2016). Regulation of pattern recognition receptor signalling in plants. *Nat Rev Immunol* 16, 537.

**Cruz, C.D., and Valent, B.** (2017). Wheat blast disease: danger on the move. *Trop Plant Pathol* 42, 210-222.

**Cvrckova, F., Grunt, M., Bezdova, R., Hala, M., Kulich, I., Rawat, A., and Zarsky, V.** (2012). Evolution of the land plant exocyst complexes. *Front Plant Sci* 3.

**Dagdas, Y.F., Belhaj, K., Maqbool, A., Chaparro-Garcia, A., Pandey, P., Petre, B., Tabassum, N., Cruz-Mireles, N., Hughes, R.K., Sklenar, J., et al.** (2016). An effector of the Irish potato famine pathogen antagonizes a host autophagy cargo receptor. *eLife* 5, e10856.

**Dangl, J.L., Horvath, D.M., and Staskawicz, B.J.** (2013). Pivoting the plant immune system from dissection to deployment. *Science* 341, 746-751.

**Dangl, J.L., and Jones, J.D.G.** (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411, 826-833.

**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.

**De la Concepcion, J.C., Franceschetti, M., MacLean, D., Terauchi, R., Kamoun, S., and Banfield, M.J.** (2019). Protein engineering expands the effector recognition profile of a rice NLR immune receptor. *eLife* 8, e47713.

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

**de Torre-Minguela, C., Mesa del Castillo, P., and Pelegrín, P.** (2017). The NLRP3 and pyrin inflammasomes: Implications in the pathophysiology of autoinflammatory diseases. *Front Immunol* 8.

**Dean, R., Van Kan, J.A., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D., Rudd, J.J., Dickman, M., Kahmann, R., Ellis, J., et al.** (2012). The Top 10 fungal pathogens in molecular plant pathology. *Mol Plant-Microbe Interact* 13, 414-430.

**Dean, R.A., Talbot, N.J., Ebbot, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.-R., Pan, H., et al.** (2005). The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434, 980-986.

**Deng, Y., Zhai, K., Xie, Z., Yang, D., Zhu, X., Liu, J., Wang, X., Qin, P., Yang, Y., Zhang, G., et al.** (2017). Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. *Science* 355, 962-965.

**Derevnina, L., Petre, B., Kellner, R., Dagdas, Y.F., Sarowar, M.N., Giannakopoulou, A., De la Concepcion, J.C., Chaparro-Garcia, A., Pennington, H.G., van West, P., et al.** (2016). Emerging oomycete threats to plants and animals. *Philos Trans R Soc, B* 371.

**Diebolder, C.A., Half, E.F., Koster, A.J., Huizinga, E.G., and Koning, R.I.** (2015). Cryoelectron tomography of the NAIP5/NLRC4 inflammasome: Implications for NLR activation. *Structure* 23, 2349-2357.

**Dinesh-Kumar, S.P., Tham, W.-H., and Baker, B.J.** (2000). Structure-function analysis of the tobacco mosaic virus resistance gene *N*. *Proc Natl Acad Sci USA* 97, 14789-14794.

**Ding, Q., Lee, Y.-K., Schaefer, Esperance A.K., Peters, Derek T., Veres, A., Kim, K., Kuperwasser, N., Motola, Daniel L., Meissner, Torsten B., Hendriks, William T., et al.** (2013). A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell* 12, 238-251.

**Dobzhansky, T.** (1937). *Genetics and the origin of species* (Columbia University Press).

**Dodds, P.N., Lawrence, G.J., Catanzariti, A.M., Teh, T., Wang, C.I., Ayliffe, M.A., Kobe, B., and Ellis, J.G.** (2006). Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc Natl Acad Sci USA* 103, 8888-8893.

**Dodds, P.N., Rafiqi, M., Gan, P.H.P., Hardham, A.R., Jones, D.A., and Ellis, J.G.** (2009). Effectors of biotrophic fungi and oomycetes: pathogenicity factors and triggers of host resistance. *New Phytol* 183, 993-1000.

**Dodds, P.N., and Rathjen, J.P.** (2010). Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat Rev Genet* 11, 539.

**Dong, G., Hutagalung, A.H., Fu, C., Novick, P., and Reinisch, K.M.** (2005). The structures of exocyst subunit Exo70p and the Exo84p C-terminal domains reveal a common motif. *Nat Struct Mol Biol* 12, 1094-1100.

**Dong, S., Raffaele, S., and Kamoun, S.** (2015). The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev* 35, 57-65.

**Du, Y., Overdijk, E.J.R., Berg, J.A., Govers, F., and Bouwmeester, K.** (2018). Solanaceous exocyst subunits are involved in immunity to diverse plant pathogens. *J Exp Bot* 69, 655-666.

**Duplessis, S., Cuomo, C.A., Lin, Y.-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D.L., Hacquard, S., Amselem, J., Cantarel, B.L., et al.** (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci USA* 108, 9166-9171.

**Ebbole, D.J.** (2007). *Magnaporthe* as a model for understanding host-pathogen interactions. *Annu Rev Phytopathol* 45, 437-456.

**Eitas, T.K., and Dangl, J.L.** (2010). NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Curr Opin Plant Biol* 13, 472-477.

**El Kasmi, F., Chung, E.-H., Anderson, R.G., Li, J., Wan, L., Eitas, T.K., Gao, Z., and Dangl, J.L.** (2017). Signaling from the plasma-membrane localized plant immune receptor RPM1 requires self-association of the full-length protein. *Proc Natl Acad Sci USA* 114, E7385-E7394.

**Elias, M., Drdova, E., Ziak, D., Bavlka, B., Hala, M., Cvrckova, F., Soukupova, H., and Zarsky, V.** (2003). The exocyst complex in plants. *Cell Biol Int* 27, 199-201.

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

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

**Engler, C., Youles, M., Gruetzner, R., Ehnert, T.M., Werner, S., Jones, J.D., Patron, N.J., and Marillonnet, S.** (2014). A golden gate modular cloning toolbox for plants. *ACS Synth Biol* 3, 839-843.

**Essuman, K., Summers, D.W., Sasaki, Y., Mao, X., DiAntonio, A., and Milbrandt, J.** (2017). The SARM1 Toll/Interleukin-1 receptor domain possesses intrinsic NAD<sup>+</sup> cleavage activity that promotes pathological axonal degeneration. *Neuron* 93, 1334-1343.e1335.

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

**Farnham, G., and Baulcombe, D.C.** (2006). Artificial evolution extends the spectrum of viruses that are targeted by a disease-resistance gene from potato. *Proc Natl Acad Sci USA* 103, 18828-18833.

**Farré, J.-C., and Subramani, S.** (2011). Rallying the exocyst as an autophagy scaffold. *Cell* 144, 172-174.

**Faustin, B., Lartigue, L., Bruey, J.-M., Luciano, F., Sergienko, E., Bailly-Maitre, B., Volkmann, N., Hanein, D., Rouiller, I., and Reed, J.C.** (2007). Reconstituted NALP1 inflammasome reveals two-step mechanism of caspase-1 activation. *Mol Cell* 25, 713-724.

**Fisher, M.C., Henk, D.A., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw, S.L., and Gurr, S.J.** (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484, 186-194.

**Flor, H.H.** (1971). Current status of the gene-for-gene concept. *Annu Rev Phytopathol* 9, 275-296.

**Franceschetti, M., Maqbool, A., Jiménez-Dalmaroni, M.J., Pennington, H.G., Kamoun, S., and Banfield, M.J.** (2017). Effectors of filamentous plant pathogens: Commonalities amid diversity. *Microbiol Mol Biol Rev* 81, e00066-00016.

**Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., et al.** (2017). ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* 50, 1212-1225.

**Fujisaki, K., Abe, Y., Ito, A., Saitoh, H., Yoshida, K., Kanzaki, H., Kanzaki, E., Utsushi, H., Yamashita, T., Kamoun, S., et al.** (2015). Rice Exo70 interacts with a fungal effector, AVR-Pii, and is required for AVR-Pii-triggered immunity. *Plant J* 83, 875-887.

**Fujisaki, K., Abe, Y., Kanzaki, E., Ito, K., Utsushi, H., Saitoh, H., Białas, A., Banfield, M.J., Kamoun, S., and Terauchi, R.** (2017). An unconventional NOI/RIN4 domain of a rice NLR protein binds host EXO70 protein to confer fungal immunity. *bioRxiv*, 239400.

**Giannakopoulou, A., Bialas, A., Kamoun, S., and Vleeshouwers, V.G.A.A.** (2016). Plant immunity switched from bacteria to virus. *Nat Biotechnol* 34, 391.

**Giannakopoulou, A., Steele, J.F., Segretin, M.E., Bozkurt, T.O., Zhou, J., Robatzek, S., Banfield, M.J., Pais, M., and Kamoun, S.** (2015). Tomato I2 immune receptor can be engineered to confer partial resistance to the oomycete *Phytophthora infestans* in addition to the fungus *Fusarium oxysporum*. *Mol Plant-Microbe Interact* 28, 1316-1329.

**Giraldo, M.C., and Valent, B.** (2013). Filamentous plant pathogen effectors in action. *Nature Reviews Microbiology* 11, 800.

**Gladieux, P., Condon, B., Ravel, S., Soanes, D., Maciel, J.L.N., Nhani, A., Chen, L., Terauchi, R., Lebrun, M.-H., Tharreau, D., et al.** (2018). Gene flow between divergent cereal- and grass-specific lineages of the rice blast fungus *Magnaporthe oryzae*. *mBio* 9, e01219-01217.

**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.

**Goodsell, D.S., Autin, L., and Olson, A.J.** (2019). Illustrate: Software for Biomolecular Illustration. *Structure*.

**Griebel, T., Maekawa, T., and Parker, J.E.** (2014). NOD-like receptor cooperativity in effector-triggered immunity. *Trends Immunol* 35, 562-570.

**Guo, J., Xu, C., Wu, D., Zhao, Y., Qiu, Y., Wang, X., Ouyang, Y., Cai, B., Liu, X., Jing, S., et al.** (2018a). Bph6 encodes an exocyst-localized protein and confers broad resistance to planthoppers in rice. *Nat Genet* 50, 297-306.

**Guo, L., Cesari, S., de Guillen, K., Chalvon, V., Mammri, L., Ma, M., Meusnier, I., Bonnot, F., Padilla, A., Peng, Y.L., et al.** (2018b). Specific recognition of two MAX effectors by integrated HMA domains in plant immune receptors involves distinct binding surfaces. *Proc Natl Acad Sci USA* 115, 11637-11642.

**Haas, B.J., Kamoun, S., Zody, M.C., Jiang, R.H.Y., Handsaker, R.E., Cano, L.M., Grabherr, M., Kodira, C.D., Raffaele, S., Torto-Alalibo, T., et al.** (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461, 393-398.

**Hamburger, Z.A., Hamburger, A.E., West, A.P., Jr., and Weis, W.I.** (2006). Crystal structure of the *S.cerevisiae* exocyst component Exo70p. *J Mol Biol* 356, 9-21.

**Harris, C.J., Slootweg, E.J., Goverse, A., and Baulcombe, D.C.** (2013). Stepwise artificial evolution of a plant disease resistance gene. *Proc Natl Acad Sci USA* 110, 21189-21194.

**He, J., Ye, W., Choi, D.S., Wu, B., Zhai, Y., Guo, B., Duan, S., Wang, Y., Gan, J., Ma, W., et al.** (2019). Structural analysis of *Phytophthora* suppressor of RNA silencing 2 (PSR2) reveals a conserved modular fold contributing to virulence. *Proc Natl Acad Sci USA* 116, 8054-8059.

**Heider, M.R., Gu, M., Duffy, C.M., Mirza, A.M., Marcotte, L.L., Walls, A.C., Farrall, N., Hakhverdyan, Z., Field, M.C., Rout, M.P., et al.** (2015). Subunit connectivity, assembly determinants and architecture of the yeast exocyst complex. *Nat Struct Mol Biol* 23, 59.

**Heider, M.R., and Munson, M.** (2012). Exorcising the exocyst complex. *Traffic* 13, 898-907.

**Helm, M., Qi, M., Sarkar, S., Yu, H., Whitham, S.A., and Innes, R.W.** (2019). Engineering a decoy substrate in soybean to enable recognition of the soybean mosaic virus Nla protease. *Mol Plant-Microbe Interact* 32, 760-769.

**Hermsen, J.G.T.** (1963a). The genetic basis of hybrid necrosis in wheat. *Genetica* 33, 245-287.

**Hermsen, J.G.T.** (1963b). Hybrid necrosis as a problem for the wheat breeder. *Euphytica* 12, 1-16.

**Horsefield, S., Burdett, H., Zhang, X., Manik, M.K., Shi, Y., Chen, J., Qi, T., Gilley, J., Lai, J.-S., Rank, M.X., et al.** (2019). NAD<sup>+</sup> cleavage activity by animal and plant TIR domains in cell death pathways. *Science* 365, 793-799.

**Howles, P., Lawrence, G., Finnegan, J., McFadden, H., Ayliffe, M., Dodds, P., and Ellis, J.** (2005). Autoactive alleles of the flax L6 rust resistance gene induce non-race-specific rust resistance associated with the hypersensitive response. *Mol Plant-Microbe Interact* 18, 570-582.

**Hu, Z., Yan, C., Liu, P., Huang, Z., Ma, R., Zhang, C., Wang, R., Zhang, Y., Martinon, F., Miao, D., et al.** (2013). Crystal structure of NLRC4 reveals its autoinhibition mechanism. *Science* 341, 172-175.

**Hu, Z., Zhou, Q., Zhang, C., Fan, S., Cheng, W., Zhao, Y., Shao, F., Wang, H.-W., Sui, S.-F., and Chai, J.** (2015). Structural and biochemical basis for induced self-propagation of NLRC4. *Science* 350, 399-404.

**Hua, L., Wu, J., Chen, C., Wu, W., He, X., Lin, F., Wang, L., Ashikawa, I., Matsumoto, T., Wang, L., et al.** (2012). The isolation of Pi1, an allele at the Pik locus which confers broad spectrum resistance to rice blast. *Theor Appl Genet* 125, 1047-1055.

**Huang, J., Si, W., Deng, Q., Li, P., and Yang, S.** (2014). Rapid evolution of avirulence genes in rice blast fungus *Magnaporthe oryzae*. *BMC Genet* 15, 45.

**Huang, L., Li, X., Li, Y., Yin, X., Li, Y., Wu, B., Mo, H., Liao, C.-J., Mengiste, T., Guo, W., et al.** (2019). Endosidin2-14 targets the exocyst complex in plants and fungal pathogens to inhibit exocytosis. *Plant Physiol* 180, 1756-1770.

**Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., et al.** (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497.

**Huh, S.U., Cevik, V., Ding, P., Duxbury, Z., Ma, Y., Tomlinson, L., Sarris, P.F., and Jones, J.D.G.** (2017). Protein-protein interactions in the RPS4/RRS1 immune receptor complex. *PLoS Path* 13, e1006376.

**Hurni, S., Brunner, S., Stirnweis, D., Herren, G., Peditto, D., McIntosh, R.A., and Keller, B.** (2014). The powdery mildew resistance gene Pm8 derived from rye is suppressed by its wheat ortholog Pm3. *Plant J* 79, 904-913.

**Illergård, K., Ardell, D.H., and Elofsson, A.** (2009). Structure is three to ten times more conserved than sequence—A study of structural response in protein cores. *Proteins: Struct, Funct, Bioinf* 77, 499-508.

**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.

**Islam, M.T., Kim, K.-H., and Choi, J.** (2019). Wheat blast in Bangladesh: The current situation and future impacts. *Plant Pathol J* 35, 1-10.

**Ispolatov, I., and Doebeli, M.** (2009). Speciation due to hybrid necrosis in plant-pathogen models. *Evolution* 63, 3076-3084.

**Jacob, F., Vernaldi, S., and Maekawa, T.** (2013). Evolution and conservation of plant NLR functions. *Front Immunol* 4, 297.

**Janjusevic, R., Abramovitch, R.B., Martin, G.B., and Stebbins, C.E.** (2006). A bacterial inhibitor of host programmed cell death defenses is an E3 ubiquitin ligase. *Science* 311, 222-226.

**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.

**Jiang, R.H.Y., Tripathy, S., Govers, F., and Tyler, B.M.** (2008). RXLR effector reservoir in two *Phytophthora* species is dominated by a single rapidly evolving superfamily with more than 700 members. *Proc Natl Acad Sci USA* 105, 4874-4879.

**Jiao, W.-B., and Schneeberger, K.** (2019). Chromosome-level assemblies of multiple *Arabidopsis thaliana* accessions reveal hotspots of genomic rearrangements. *bioRxiv*, 738880.

**Johnson, K.C., Dong, O.X., Huang, Y., and Li, X.** (2012). A rolling stone gathers no moss, but resistant plants must gather their moses. *Cold Spring Harbor symposia on quantitative biology* 77, 259-268.

**Jones, J.D., and Dangl, J.L.** (2006). The plant immune system. *Nature* 444, 323-329.

**Jones, J.D.G., Vance, R.E., and Dangl, J.L.** (2016). Intracellular innate immune surveillance devices in plants and animals. *Science* 354, aaf6395.

**Jubic, L.M., Saile, S., Furzer, O.J., El Kasmi, F., and Dangl, J.L.** (2019). Help wanted: helper NLRs and plant immune responses. *Curr Opin Plant Biol* 50, 82-94.

**Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., and Xu, J.** (2012). Template-based protein structure modeling using the RaptorX web server. *Nat Protoc* 7, 1511-1522.

**Kalmbach, L., Hématy, K., De Bellis, D., Barberon, M., Fujita, S., Ursache, R., Daraspe, J., and Geldner, N.** (2017). Transient cell-specific EXO70A1 activity in the CASP domain and Casparian strip localization. *Nat Plants* 3, 17058.

**Kamoun, S., Furzer, O., Jones, J.D., Judelson, H.S., Ali, G.S., Dalio, R.J., Roy, S.G., Schena, L., Zamounis, A., Panabieres, F., et al.** (2015). The Top 10 oomycete pathogens in molecular plant pathology. *Mol Plant Pathol* 16, 413-434.

**Kämper, J., Kahmann, R., Böker, M., Ma, L.-J., Brefort, T., Saville, B.J., Banuett, F., Kronstad, J.W., Gold, S.E., Müller, O., et al.** (2006). Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444, 97-101.

**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. *Plant J* 72, 894-907.

**Karasov, T.L., Chae, E., Herman, J.J., and Bergelson, J.** (2017). Mechanisms to mitigate the trade-off between growth and defense. *Plant Cell* 29, 666-680.

**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.

**Kiyosawa, S.** (1969a). Inheritance of blast-resistance in west pakistani rice variety, Pusur. *Jpn J Breed* 19, 121-128.

**Kiyosawa, S.** (1969b). Inheritance of resistance of rice varieties to a philippine fungus strain of *Pyricularia oryzae*. *Jpn J Breed* 19, 61-73.

**Kiyosawa, S.** (1978). Identification of blast-resistance genes in some rice varieties. *Jpn J Breed* 28, 287-296.

**Kiyosawa, S., and Murty, V.V.S.** (1969). The inheritance of blast-resistance in indian rice variety, HR-22. *Jpn J Breed* 19, 269-276.

**Koenig, D., Hagmann, J., Li, R., Bemm, F., Slotte, T., Neuffer, B., Wright, S.I., and Weigel, D.** (2019). Long-term balancing selection drives evolution of immunity genes in *Capsella*. *eLife* 8.

**Kofoed, E.M., and Vance, R.E.** (2011). Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity. *Nature* 477, 592-595.

**Koonin, E.V., and Aravind, L.** (2002). Origin and evolution of eukaryotic apoptosis: the bacterial connection. *Cell Death & Differentiation* 9, 394-404.

**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. *Plant Cell* 30, 285-299.

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

**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.

**Kulich, I., Pečenková, T., Sekereš, J., Smetana, O., Fendrych, M., Foissner, I., Höftberger, M., and Žárský, V.** (2013). *Arabidopsis* exocyst subcomplex containing subunit EXO70B1 is involved in autophagy-related transport to the vacuole. *Traffic* 14, 1155-1165.

**Kulich, I., Vojtíková, Z., Glanc, M., Ortmannová, J., Rasmann, S., and Žárský, V.** (2015). Cell wall maturation of *Arabidopsis* trichomes is dependent on exocyst subunit EXO70H4 and involves callose deposition. *Plant Physiol* 168, 120-131.

**Kulich, I., Vojtíková, Z., Sabol, P., Ortmannová, J., Neděla, V., Tihlaříková, E., and Žárský, V.** (2018). Exocyst subunit EXO70H4 has a specific role in callose synthase secretion and silica accumulation. *Plant Physiol* 176, 2040-2051.

**Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K.** (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35, 1547-1549.

**Le Roux, C., Huet, G., Jauneau, A., Camborde, L., Tremousaygue, D., Kraut, A., Zhou, B., Levaillant, M., Adachi, H., Yoshioka, H., et al.** (2015). A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell* 161, 1074-1088.

**Lee, D., Bourdais, G., Yu, G., Robatzek, S., and Coaker, G.** (2015). Phosphorylation of the plant immune regulator RPM1-interacting protein 4 enhances plant plasma membrane H<sup>+</sup>-ATPase activity and inhibits flagellin-triggered immune responses in *Arabidopsis*. *Plant Cell* 27, 2042-2056.

**Lee, S.-K., Song, M.-Y., Seo, Y.-S., Kim, H.-K., Ko, S., Cao, P.-J., Suh, J.-P., Yi, G., Roh, J.-H., Lee, S., et al.** (2009). Rice Pi5-mediated resistance to *Magnaporthe oryzae* requires the presence of two coiled-coil-nucleotide-binding-leucine-rich repeat genes. *Genetics* 181, 1627-1638.

**Leipe, D.D., Koonin, E.V., and Aravind, L.** (2004). STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: Multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J Mol Biol* 343, 1-28.

**Lepore, D.M., Martinez-Nunez, L., and Munson, M.** (2018). Exposing the elusive exocyst structure. *Trends Biochem Sci* 43, 714-725.

**Letunic, I., and Bork, P.** (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 47, W256-W259.

**Levesque, C.A., Brouwer, H., Cano, L., Hamilton, J.P., Holt, C., Huitema, E., Raffaele, S., Robideau, G.P., Thines, M., Win, J., et al.** (2010). Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biol* 11, R73.

**Li, J., Wang, Q., Li, C., Bi, Y., Fu, X., and Wang, R.** (2019a). Novel haplotypes and networks of AVR-Pik alleles in *Magnaporthe oryzae*. *BMC Plant Biol* 19, 204.

**Li, L., Habring, A., Wang, K., and Weigel, D.** (2019b). Oligomerization of NLR immune receptor RPP7 triggered by atypical resistance protein RPW8/HR as ligand. *bioRxiv*, 682807.

**Lin, Y., Ding, Y., Wang, J., Shen, J., Kung, C.H., Zhuang, X., Cui, Y., Yin, Z., Xia, Y., Lin, H., et al.** (2015). Exocyst-positive organelles and autophagosomes are distinct organelles in plants. *Plant Physiol* 169, 1917-1932.

**Liu, J., Elmore, J.M., Lin, Z.-J.D., and Coaker, G.** (2011). A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. *Cell Host Microbe* 9, 137-146.

**Liu, N., Hake, K., Wang, W., Zhao, T., Romeis, T., and Tang, D.** (2017). Calcium-dependent protein kinase 5 associates with the truncated NLR protein TIR-NBS2 to contribute to *exo70B1*-mediated immunity. *Plant Cell* 29, 746-759.

**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, 753.

**Longya, A., Chaipanya, C., Franceschetti, M., Maidment, J.H.R., Banfield, M.J., and Jantasuriyarat, C.** (2019). Gene duplication and mutation in the emergence of a novel aggressive allele of the AVR-Pik effector in the rice blast fungus. *Mol Plant-Microbe Interact* 32, 740-749.

**Lopez, V.A., Park, B.C., Nowak, D., Sreelatha, A., Zembek, P., Fernandez, J., Servage, K.A., Gradowski, M., Hennig, J., Tomchick, D.R., et al.** (2019). A bacterial effector mimics a host HSP90 client to undermine immunity. *Cell* 179, 205-218.e221.

**Lynch, M., and Force, A.G.** (2000). The origin of interspecific genomic incompatibility via gene duplication. *Am Nat* 156, 590-605.

**Ma, Y., Guo, H., Hu, L., Martinez, P.P., Moschou, P.N., Cevik, V., Ding, P., Duxbury, Z., Sarris, P.F., and Jones, J.D.G.** (2018). Distinct modes of derepression of an *Arabidopsis* immune receptor complex by two different bacterial effectors. *Proc Natl Acad Sci USA* 115, 10218.

**Mackey, D., Holt, B.F., Wiig, A., and Dangl, J.L.** (2002). RIN4 interacts with *Pseudomonas syringae* Type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* 108, 743-754.

**MacQueen, A., Sun, X., and Bergelson, J.** (2016). Genetic architecture and pleiotropy shape costs of Rps2-mediated resistance in *Arabidopsis thaliana*. *Nat Plants* 2, 16110.

**Maekawa, T., Cheng, W., Spiridon, L.N., Töller, A., Lukasik, E., Saijo, Y., Liu, P., Shen, Q.-H., Micluta, 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.

**Malik, S., and Van der Hoorn, R.A.** (2016). Inspirational decoys: a new hunt for effector targets. *New Phytol* 210, 371-373.

**Maqbool, A., Hughes, R.K., Dagdas, Y.F., Tregidgo, N., Zess, E., Belhaj, K., Round, A., Bozkurt, T.O., Kamoun, S., and Banfield, M.J.** (2016). Structural basis of host autophagy-related protein 8 (ATG8) binding by the Irish potato famine pathogen effector protein PexRD54. *J Biol Chem* 291, 20270-20282.

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

**McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J.** (2007). Phaser crystallographic software. *J Appl Crystallogr* 40, 658-674.

**McDonald, B.A., and Stukenbrock, E.H.** (2016). Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philos Trans R Soc, B* 371.

**Mei, K., and Guo, W.** (2018). The exocyst complex. *Curr Biol* 28, R922-R925.

**Mei, K., Li, Y., Wang, S., Shao, G., Wang, J., Ding, Y., Luo, G., Yue, P., Liu, J.J., Wang, X., et al.** (2018). Cryo-EM structure of the exocyst complex. *Nat Struct Mol Biol* 25, 139-146.

**Mestre, P., and Baulcombe, D.C.** (2006). Elicitor-mediated oligomerization of the tobacco N disease resistance protein. *Plant Cell* 18, 491-501.

**Meunier, E., and Broz, P.** (2017). Evolutionary convergence and divergence in NLR function and structure. *Trends Immunol* 38, 744-757.

**Meyers, B.C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R.W.** (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15, 809-834.

**Mindrinos, M., Katagiri, F., Yu, G.L., and Ausubel, F.M.** (1994). The *A. thaliana* disease resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* 78, 1089-1099.

**Molina, J., Sikora, M., Garud, N., Flowers, J.M., Rubinstein, S., Reynolds, A., Huang, P., Jackson, S., Schaal, B.A., Bustamante, C.D., et al.** (2011). Molecular evidence for a single evolutionary origin of domesticated rice. *Proc Natl Acad Sci USA* 108, 8351-8356.

**Monteiro, F., and Nishimura, M.T.** (2018). Structural, functional, and genomic diversity of plant NLR proteins: An evolved resource for rational engineering of plant immunity. *Annu Rev Phytopathol* 56, 243-267.

**Moore, B.A., Robinson, H.H., and Xu, Z.** (2007). The crystal structure of mouse Exo70 reveals unique features of the mammalian exocyst. *J Mol Biol* 371, 410-421.

**Morel, J.B., and Dangl, J.L.** (1997). The hypersensitive response and the induction of cell death in plants. *Cell Death & Differentiation* 4, 671-683.

**Mukhtar, M.S., Carvunis, A.-R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T., Nishimura, M.T., et al.** (2011). Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* 333, 596-601.

**Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A.** (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr Sect D Biol Crystallogr* 67, 355-367.

**Nandety, R.S., Caplan, J.L., Cavanaugh, K., Perroud, B., Wroblewski, T., Michelmore, R.W., and Meyers, B.C.** (2013). The role of TIR-NBS and TIR-X proteins in plant basal defense responses. *Plant Physiol* 162, 1459-1472.

**Narusaka, M., Shirasu, K., Noutoshi, Y., Kubo, Y., Shiraishi, T., Iwabuchi, M., and Narusaka, Y.** (2009). RRS1 and RPS4 provide a dual Resistance-gene system against fungal and bacterial pathogens. *Plant J* 60, 218-226.

**Nei, M., and Kumar, S.** (2000). Molecular evolution and phylogenetics (Oxford University Press, New York).

**Newman, T.E., Lee, J., Williams, S.J., Choi, S., Halane, M.K., Zhou, J., Solomon, P., Kobe, B., Jones, J.D.G., Segonzac, C., et al.** (2019). Autoimmunity and effector recognition in *Arabidopsis thaliana* can be uncoupled by mutations in the RRS1-R immune receptor. *New Phytol* 222, 954-965.

**Nishimura, M.T., Anderson, R.G., Cherkis, K.A., Law, T.F., Liu, Q.L., Machius, M., Nimchuk, Z.L., Yang, L., Chung, E.-H., El Kasmi, F., et al.** (2017). TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in *Arabidopsis*. *Proc Natl Acad Sci USA* 114, E2053-E2062.

**Noutoshi, Y., Ito, T., Seki, M., Nakashita, H., Yoshida, S., Marco, Y., Shirasu, K., and Shinozaki, K.** (2005). A single amino acid insertion in the WRKY domain of the *Arabidopsis* TIR-NBS-LRR-WRKY-type disease resistance protein SLH1 (sensitive to low humidity 1) causes activation of defense responses and hypersensitive cell death. *Plant J* 43, 873-888.

**Nyarko, A., Singarapu, K.K., Figueroa, M., Manning, V.A., Pandelova, I., Wolpert, T.J., Ciuffetti, L.M., and Barbar, E.** (2014). Solution NMR structures of *Pyrenophora tritici-repentis* ToxB and its inactive homolog reveal potential determinants of toxin activity. *J Biol Chem* 289, 25946-25956.

**Ogura, T., Goeschl, C., Filiault, D., Mirea, M., Slovak, R., Wolrab, B., Satbhai, S.B., and Busch, W.** (2019). Root system depth in *Arabidopsis* is shaped by EXOCYST70A3 via the dynamic modulation of auxin transport. *Cell* 178, 400-412.e416.

**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. *Plant Cell* 29, 156-168.

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

**Ostertag, M., Stammler, J., Douchkov, D., Eichmann, R., and Huckelhoven, R.** (2013). The conserved oligomeric Golgi complex is involved in penetration resistance of barley to the barley powdery mildew fungus. *Mol Plant Pathol* 14, 230-240.

**Park, C.H., Chen, S., Shirsekar, G., Zhou, B., Khang, C.H., Songkumarn, P., Afzal, A.J., Ning, Y., Wang, R., Bellizzi, M., et al.** (2012). The *Magnaporthe oryzae* effector AvrPiz-t targets the RING E3 ubiquitin ligase APIP6 to suppress pathogen-associated molecular pattern-triggered immunity in rice. *Plant Cell* 24, 4748-4762.

**Park, C.H., Shirsekar, G., Bellizzi, M., Chen, S., Songkumarn, P., Xie, X., Shi, X., Ning, Y., Zhou, B., Suttiviriya, P., et al.** (2016). The E3 Ligase APIP10 connects the effector AvrPiz-t to the NLR receptor Piz-t in rice. *PLoS Path* 12, e1005529.

**Pečenková, T., Hála, M., Kulich, I., Kocourková, D., Drdová, E., Fendrych, M., Toupalová, H., and Žárský, V.** (2011). The role for the exocyst complex subunits Exo70B2 and Exo70H1 in the plant-pathogen interaction. *J Exp Bot* 62, 2107-2116.

**Peng, Z., Oliveira-Garcia, E., Lin, G., Hu, Y., Dalby, M., Migeon, P., Tang, H., Farman, M., Cook, D., White, F.F., et al.** (2019). Effector gene reshuffling involves dispensable mini-chromosomes in the wheat blast fungus. *PLoS Genet* 15, e1008272.

**Pennisi, E.** (2010). Armed and dangerous. *Science* 327, 804-805.

**Periyannan, S., Moore, J., Ayliffe, M., Bansal, U., Wang, X., Huang, L., Deal, K., Luo, M., Kong, X., Bariana, H., et al.** (2013). The gene *Sr33*, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. *Science* 341, 786-788.

**Picco, A., Irastorza-Azcarate, I., Specht, T., Böke, D., Pazos, I., Rivier-Cordey, A.-S., Devos, D.P., Kaksonen, M., and Gallego, O.** (2017). The *in vivo* architecture of the exocyst provides structural basis for exocytosis. *Cell* 168, 400-412.e418.

**Piñeiro, Á., Muñoz, E., Sabín, J., Costas, M., Bastos, M., Velázquez-Campoy, A., Garrido, P.F., Dumas, P., Ennifar, E., García-Río, L., et al.** (2019). AFFINImeter: A software to analyze molecular recognition processes from experimental data. *Anal Biochem* 577, 117-134.

**Pooja, K., and Katoch, A.** (2014). Past, present and future of rice blast management. *Plant Sci Today* 1, 165-173.

**Presti, L.L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., Zuccaro, A., Reissmann, S., and Kahmann, R.** (2015). Fungal effectors and plant susceptibility. *Annu Rev Plant Biol* 66, 513-545.

**Price, M.N., Dehal, P.S., and Arkin, A.P.** (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26, 1641-1650.

**Raffaele, S., Farrer, R.A., Cano, L.M., Studholme, D.J., MacLean, D., Thines, M., Jiang, R.H., Zody, M.C., Kunjeti, S.G., Donofrio, N.M., et al.** (2010). Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* 330, 1540-1543.

**Raffaele, S., and Kamoun, S.** (2012). Genome evolution in filamentous plant pathogens: why bigger can be better. *Nature Reviews Microbiology* 10, 417-430.

**Redditt, T.J., Chung, E.H., Karimi, H.Z., Rodibaugh, N., Zhang, Y., Trinidad, J.C., Kim, J.H., Zhou, Q., Shen, M., Dangl, J.L., et al.** (2019). AvrRpm1 Functions as an ADP-Ribosyl Transferase to Modify NOI Domain-Containing Proteins, Including Arabidopsis and Soybean RPM1-Interacting Protein4. *Plant Cell* 31, 2664-2681.

**Regente, M., Pinedo, M., San Clemente, H., Balliau, T., Jamet, E., and de la Canal, L.** (2017). Plant extracellular vesicles are incorporated by a fungal pathogen and inhibit its growth. *J Exp Bot* 68, 5485-5495.

**Reubold, T.F., Wohlgemuth, S., and Eschenburg, S.** (2011). Crystal structure of full-length Apaf-1: how the death signal is relayed in the mitochondrial pathway of apoptosis. *Structure* 19, 1074-1083.

**Richard, M.M.S., and Takken, F.L.W.** (2017). Plant autoimmunity: When good things go bad. *Curr Biol* 27, R361-R363.

**Rodriguez-Moreno, L., Song, Y., and Thomma, B.P.** (2017). Transfer and engineering of immune receptors to improve recognition capacities in crops. *Curr Opin Plant Biol* 38, 42-49.

**Ronald, P.C., and Beutler, B.** (2010). Plant and animal sensors of conserved microbial signatures. *Science* 330, 1061-1064.

**Rutter, B.D., and Innes, R.W.** (2017). Extracellular vesicles isolated from the leaf apoplast carry stress-response proteins. *Plant Physiol* 173, 728-741.

**Sabol, P., Kulich, I., and Zarsky, V.** (2017). RIN4 recruits the exocyst subunit EXO70B1 to the plasma membrane. *J Exp Bot* 68, 3253-3265.

**Saeed, B., Brillada, C., and Trujillo, M.** (2019). Dissecting the plant exocyst. *Curr Opin Plant Biol* 52, 69-76.

**Sandstrom, A., Mitchell, P.S., Goers, L., Mu, E.W., Lesser, C.F., and Vance, R.E.** (2019). Functional degradation: A mechanism of NLRP1 inflammasome activation by diverse pathogen enzymes. *Science* 364, eaau1330.

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

**Sarris, P.F., Duxbury, Z., Huh, S.U., Ma, Y., Segonzac, C., Sklenar, J., Derbyshire, P., Cevik, V., Rallapalli, G., Saucet, S.B., et al.** (2015). A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell* 161, 1089-1100.

**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. *Nat Ecol Evol* 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.

**Seck, P.A., Diagne, A., Mohanty, S., and Wopereis, M.C.S.** (2012). Crops that feed the world 7: Rice. *Food Security* 4, 7-24.

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

**Seto, D., Koulena, N., Lo, T., Menna, A., Guttman, D.S., and Desveaux, D.** (2017). Expanded Type III effector recognition by the ZAR1 NLR protein using ZED1-related kinases. *Nat Plants* 3, 17027.

**Shao, F., Golstein, C., Ade, J., Stoutemyer, M., Dixon, J.E., and Innes, R.W.** (2003). Cleavage of *Arabidopsis* PBS1 by a bacterial Type III effector. *Science* 301, 1230-1233.

**Sharif, H., Wang, L., Wang, W.L., Magupalli, V.G., Andreeva, L., Qiao, Q., Hauenstein, A.V., Wu, Z., Nunez, G., Mao, Y., et al.** (2019). Structural mechanism for NEK7-licensed activation of NLRP3 inflammasome. *Nature* 570, 338-343.

**Sievers, F., and Higgins, D.G.** (2014). Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* 1079, 105-116.

**Sinapidou, E., Williams, K., Nott, L., Bahkt, S., Tor, M., Crute, I., Bittner-Eddy, P., and Beynon, J.** (2004). Two TIR:NB:LRR genes are required to specify resistance to *Peronospora parasitica* isolate Cala2 in *Arabidopsis*. *Plant J* 38, 898-909.

**Skamnioti, P., and Gurr, S.J.** (2009). Against the grain: safeguarding rice from rice blast disease. *Trends Biotechnol* 27, 141-150.

**Slootweg, E., Koropacka, K., Roosien, J., Dees, R., Overmars, H., Lankhorst, R.K., van Schaik, C., Pomp, R., Bouwman, L., Helder, J., et al.** (2017). Sequence exchange between homologous NB-LRR genes converts virus resistance into nematode resistance, and vice versa. *Plant Physiol* 175, 498-510.

**Slootweg, E.J., Spiridon, L.N., Roosien, J., Butterbach, P., Pomp, R., Westerhof, L., Wilbers, R., Bakker, E., Bakker, J., Petrescu, A.J., et al.** (2013). Structural determinants at the interface of the ARC2 and leucine-rich repeat domains control the activation of the plant immune receptors Rx1 and Gpa2. *Plant Physiol* 162, 1510-1528.

**Sperschneider, J., Dodds, P.N., Gardiner, D.M., Manners, J.M., Singh, K.B., and Taylor, J.M.** (2015). Advances and challenges in computational prediction of effectors from plant pathogenic fungi. *PLoS Path* 11, e1004806.

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

**Stebbins, C.E.** (2005). Structural microbiology at the pathogen–host interface. *Cell Microbiol* 7, 1227-1236.

**Stegmann, M., Anderson, R.G., Ichimura, K., Pecenkova, T., Reuter, P., Žárský, V., McDowell, J.M., Shirasu, K., and Trujillo, M.** (2012). The ubiquitin ligase PUB22 targets a subunit of the exocyst complex required for PAMP-triggered responses in *Arabidopsis*. *Plant Cell* 24, 4703-4716.

**Stegmann, M., Anderson, R.G., Westphal, L., Rosahl, S., McDowell, J.M., and Trujillo, M.** (2013). The exocyst subunit Exo70B1 is involved in the immune response of *Arabidopsis thaliana* to different pathogens and cell death. *Plant Signal Behav* 8, e27421.

**Stein, J.C., Yu, Y., Copetti, D., Zwickl, D.J., Zhang, L., Zhang, C., Chougule, K., Gao, D., Iwata, A., Goicoechea, J.L., et al.** (2018). Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat Genet* 50, 285-296.

**Steuernagel, B., Periyannan, S.K., Hernández-Pinzón, I., Witek, K., Rouse, M.N., Yu, G., Hatta, A., Ayliffe, M., Bariana, H., Jones, J.D.G., et al.** (2016). Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat Biotechnol* 34, 652.

**Steuernagel, B., Witek, K., Krattinger, S.G., Ramirez-Gonzalez, R.H., Schoonbeek, H.-j., Yu, G., Baggs, E., Witek, A.I., Yadav, I., Krasileva, K.V., et al.** (2018). Physical and transcriptional organisation of the bread wheat intracellular immune receptor repertoire. *bioRxiv*, 339424.

**Stirnweis, D., Milani, S.D., Brunner, S., Herren, G., Buchmann, G., Peditto, D., Jordan, T., and Keller, B.** (2014a). Suppression among alleles encoding nucleotide-binding–leucine-rich repeat resistance proteins interferes with resistance in F1 hybrid and allele-pyramided wheat plants. *Plant J* 79, 893-903.

**Stirnweis, D., Milani, S.D., Jordan, T., Keller, B., and Brunner, S.** (2014b). Substitutions of two amino acids in the nucleotide-binding site domain of a resistance protein enhance the hypersensitive response and enlarge the PM3F resistance spectrum in wheat. *Mol Plant-Microbe Interact* 27, 265-276.

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

**Sukarta, O.C.A., Slootweg, E.J., and Goverse, A.** (2016). Structure-informed insights for NLR functioning in plant immunity. *Semin Cell Dev Biol* 56, 134-149.

**Summers, D.W., Gibson, D.A., DiAntonio, A., and Milbrandt, J.** (2016). SARM1-specific motifs in the TIR domain enable NAD<sup>+</sup> loss and regulate injury-induced SARM1 activation. *Proc Natl Acad Sci USA* 113, E6271-E6280.

**Takagi, H., Abe, A., Uemura, A., Oikawa, K., Utsushi, H., Yaegashi, H., Kikuchi, H., Shimizu, M., Abe, Y., Kanzaki, H., et al.** (2017). Rice blast resistance gene *Pii* is controlled by a pair of NBS-LRR genes *Pii-1* and *Pii-2*. *bioRxiv*, 227132.

**Takagi, H., Uemura, A., Yaegashi, H., Tamiru, M., Abe, A., Mitsuoka, C., Utsushi, H., Natsume, S., Kanzaki, H., Matsumura, H., et al.** (2013). MutMap-Gap: whole-genome resequencing of mutant F2 progeny bulk combined with de novo assembly of gap regions identifies the rice blast resistance gene *Pii*. *New Phytol* 200, 276-283.

**Takken, F.L., and Goverse, A.** (2012). How to build a pathogen detector: structural basis of NB-LRR function. *Curr Opin Plant Biol* 15, 375-384.

**Takken, F.L.W., Albrecht, M., and Tameling, W.I.L.** (2006). Resistance proteins: molecular switches of plant defence. *Curr Opin Plant Biol* 9, 383-390.

**Talbot, N.J.** (2003). On the trail of a cereal killer: Exploring the biology of *Magnaporthe grisea*. *Annu Rev Microbiol* 57, 177-202.

**Tameling, W.I., Elzinga, S.D., Darmin, P.S., Vossen, J.H., Takken, F.L., Haring, M.A., and Cornelissen, B.J.** (2002). The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. *Plant Cell* 14, 2929-2939.

**Tameling, W.I.L., Vossen, J.H., Albrecht, M., Lengauer, T., Berden, J.A., Haring, M.A., Cornelissen, B.J.C., and Takken, F.L.W.** (2006). Mutations in the NB-ARC domain of I-2 that impair ATP hydrolysis cause autoactivation. *Plant Physiol* 140, 1233-1245.

**Tamura, K., and Nei, M.** (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10, 512-526.

**Tao, Y., Yuan, F., Leister, R.T., Ausubel, F.M., and Katagiri, F.** (2000). Mutational analysis of the *Arabidopsis* nucleotide binding site-leucine-rich repeat resistance gene *RPS2*. *Plant Cell* 12, 2541-2554.

**Taylor, J.S., Van de Peer, Y., and Meyer, A.** (2001). Genome duplication, divergent resolution and speciation. *Trends Genet* 17, 299-301.

**Teh, O.-K., Lee, C.-W., Ditengou, F.A., Klecker, T., Furlan, G., Zietz, M., Hause, G., Eschen-Lippol, L., Hoehenwarter, W., Lee, J., et al.** (2019). Phosphorylation of the exocyst subunit Exo70B2 contributes to the regulation of its function. *bioRxiv*, 266171.

**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.

**Terauchi, R., and Yoshida, K.** (2010). Towards population genomics of effector-effector target interactions. *New Phytol* 187, 929-939.

**TerBush, D.R., Maurice, T., Roth, D., and Novick, P.** (1996). The exocyst is a multiprotein complex required for exocytosis in *Saccharomyces cerevisiae*. *EMBO J* 15, 6483-6494.

**Tian, D., Traw, M.B., Chen, J.Q., Kreitman, M., and Bergelson, J.** (2003). Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* 423, 74-77.

**Tornero, P., Chao, R.A., Luthin, W.N., Goff, S.A., and Dangl, J.L.** (2002). Large-scale structure –function analysis of the *Arabidopsis* RPM1 disease resistance protein. *Plant Cell* 14, 435-450.

**Tran, D.T.N., Chung, E.H., Habring-Muller, A., Demar, M., Schwab, R., Dangl, J.L., Weigel, D., and Chae, E.** (2017). Activation of a plant NLR complex through heteromeric association with an autoimmune risk variant of another NLR. *Curr Biol* 27, 1148-1160.

**Tyler, B.M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R.H., Aerts, A., Arredondo, F.D., Baxter, L., Bensasson, D., Beynon, J.L., et al.** (2006). *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313, 1261-1266.

**Uma, B., Rani, T.S., and Podile, A.R.** (2011). Warriors at the gate that never sleep: non-host resistance in plants. *J Plant Physiol* 168, 2141-2152.

**Upson, J.L., Zess, E.K., Bialas, A., Wu, C.-h., and Kamoun, S.** (2018). The coming of age of EvoMPMI: evolutionary molecular plant–microbe interactions across multiple timescales. *Curr Opin Plant Biol* 44, 108-116.

**Vaid, N., and Laitinen, R.A.E.** (2019). Diverse paths to hybrid incompatibility in *Arabidopsis*. *Plant J* 97, 199-213.

**Van de Weyer, A.-L., Monteiro, F., Furzer, O.J., Nishimura, M.T., Cevik, V., Witek, K., Jones, J.D.G., Dangl, J.L., Weigel, D., and Bemm, F.** (2019). A species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. *Cell* 178, 1260-1272.e1214.

**van der Hoorn, R.A.L., and Kamoun, S.** (2008). From guard to decoy: A new model for perception of plant pathogen effectors. *Plant Cell* 20, 2009-2017.

**van Ooijen, G., Mayr, G., Kasiem, M.M., Albrecht, M., Cornelissen, B.J., and Takken, F.L.** (2008). Structure-function analysis of the NB-ARC domain of plant disease resistance proteins. *J Exp Bot* 59, 1383-1397.

**van Wersch, S., and Li, X.** (2019). Stronger when together: Clustering of plant NLR disease resistance genes. *Trends Plant Sci* 24, 688-699.

**Varden, F.A., De la Concepcion, J.C., Maidment, J.H., and Banfield, M.J.** (2017). Taking the stage: effectors in the spotlight. *Curr Opin Plant Biol* 38, 25-33.

**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*.

**Velasquez, A.C., Castroverde, C.D.M., and He, S.Y.** (2018). Plant-pathogen warfare under changing climate conditions. *Curr Biol* 28, R619-r634.

**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<sup>+</sup>-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.** (2015a). 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, G.F., Ji, J., El-Kasmi, F., Dangl, J.L., Johal, G., and Balint-Kurti, P.J.** (2015b). Molecular and functional analyses of a maize autoactive NB-LRR protein identify precise structural requirements for activity. *PLoS Path* 11, e1004674.

**Wang, J., Ding, Y., Wang, J., Hillmer, S., Miao, Y., Lo, S.W., Wang, X., Robinson, D.G., and Jiang, L.** (2010). EXPO, an exocyst-positive organelle distinct from multivesicular endosomes and autophagosomes, mediates cytosol to cell wall exocytosis in *Arabidopsis* and tobacco cells. *Plant Cell* 22, 4009-4030.

**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, L., Zhao, L., Zhang, X., Zhang, Q., Jia, Y., Wang, G., Li, S., Tian, D., Li, W.-H., and Yang, S.** (2019c). Large-scale identification and functional analysis of *NLR* genes in blast resistance in the Tetep rice genome sequence. *Proc Natl Acad Sci USA*, 201910229.

**Wang, W., Liu, N., Gao, C., Rui, L., and Tang, D.** (2019d). The *Pseudomonas syringae* effector AvrPtoB associates with and ubiquitinates *Arabidopsis* exocyst subunit EXO70B1. *Front Plant Sci* 10.

**Wang, Y., Zhang, Q., Sun, M.A., and Guo, D.** (2011). High-accuracy prediction of bacterial type III secreted effectors based on position-specific amino acid composition profiles. *Bioinformatics* 27, 777-784.

**Werth, C.R., and Windham, M.D.** (1991). A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *Am Nat* 137, 515-526.

**Weßling, R., Epple, P., Altmann, S., He, Y., Yang, L., Henz, Stefan R., McDonald, N., Wiley, K., Bader, Kai C., Gläßer, C., et al.** (2014). Convergent targeting of a common host protein-network by pathogen effectors from three kingdoms of life. *Cell Host Microbe* 16, 364-375.

**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.

**Whitham, S., Dinesh-Kumar, S.P., Choi, D., Hehl, R., Corr, C., and Baker, B.** (1994). The product of the tobacco mosaic virus resistance gene N: similarity to toll and the interleukin-1 receptor. *Cell* 78, 1101-1115.

**Wickham, H.** (2009). *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.

**Williams, S.J., Sornaraj, P., deCourcy-Ireland, E., Menz, R.I., Kobe, B., Ellis, J.G., Dodds, P.N., and Anderson, P.A.** (2011). An autoactive mutant of the M flax rust resistance protein has a preference for binding ATP, whereas wild-type M protein binds ADP. *Mol Plant-Microbe Interact* 24, 897-906.

**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., Chaparro-Garcia, A., Belhaj, K., Saunders, D.G., Yoshida, K., Dong, S., Schornack, S., Zipfel, C., Robatzek, S., Hogenhout, S.A., et al.** (2012a). Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harbor Symp Quant Biol* 77, 235-247.

**Win, J., Krasileva, K.V., Kamoun, S., Shirasu, K., Staskawicz, B.J., and Banfield, M.J.** (2012b). Sequence divergent RXLR effectors share a structural fold conserved across plant pathogenic oomycete species. *PLoS Path* 8, e1002400.

**Win, J., Morgan, W., Bos, J., Krasileva, K.V., Cano, L.M., Chaparro-Garcia, A., Ammar, R., Staskawicz, B.J., and Kamoun, S.** (2007). Adaptive evolution has targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. *Plant Cell* 19, 2349-2369.

**Wing, R.A., Purugganan, M.D., and Zhang, Q.** (2018). The rice genome revolution: from an ancient grain to Green Super Rice. *Nat Rev Genet* 19, 505-517.

**Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al.** (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* 67, 235-242.

**Winter, G.** (2010). xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* 43, 186-190.

**Wirthmueller, L., Maqbool, A., and Banfield, M.J.** (2013). On the front line: structural insights into plant-pathogen interactions. *Nature Reviews Microbiology* 11, 761.

**Wood, K., Nur, M., Gil, J., Fletcher, K., Lakeman, K., Gothberg, A., Khuu, T., Kopetzky, J., Pandya, A., Pel, M., et al.** (2019). Effector prediction and characterization in the oomycete pathogen *Bremia lactucae* reveal host-recognized WY domain proteins that lack the canonical RXLR motif. *bioRxiv*, 679787.

**Wu, B., and Guo, W.** (2015). The exocyst at a glance. *J Cell Sci* 128, 2957-2964.

**Wu, C.-H., Derevnina, L., and Kamoun, S.** (2018). Receptor networks underpin plant immunity. *Science* 360, 1300-1301.

**Wu, C.-H., Krasileva, K., Banfield, M., Terauchi, R., and Kamoun, S.** (2015). The “sensor domains” of plant NLR proteins: more than decoys? *Front Plant Sci* 6.

**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. *Proc Natl Acad Sci USA* 114, 8113-8118.

**Xie, W., Wang, G., Yuan, M., Yao, W., Lyu, K., Zhao, H., Yang, M., Li, P., Zhang, X., Yuan, J., et al.** (2015). Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. *Proc Natl Acad Sci USA* 112, E5411-E5419.

**Xu, G., Greene, G.H., Yoo, H., Liu, L., Marqués, J., Motley, J., and Dong, X.** (2017a). Global translational reprogramming is a fundamental layer of immune regulation in plants. *Nature* 545, 487.

**Xu, G., Yuan, M., Ai, C., Liu, L., Zhuang, E., Karapetyan, S., Wang, S., and Dong, X.** (2017b). uORF-mediated translation allows engineered plant disease resistance without fitness costs. *Nature* 545, 491.

**Xu, X., Hayashi, N., Wang, C.T., Kato, H., Fujimura, T., and Kawasaki, S.** (2008). Efficient authentic fine mapping of the rice blast resistance gene *Pik-h* in the *Pik* cluster, using new *Pik-h*-differentiating isolates. *Mol Breed* 22, 289-299.

**Yaeno, T., Li, H., Chaparro-Garcia, A., Schornack, S., Koshiba, S., Watanabe, S., Kigawa, T., Kamoun, S., and Shirasu, K.** (2011). Phosphatidylinositol monophosphate-binding interface in the oomycete RXLR effector AVR3a is required for its stability in host cells to modulate plant immunity. *Proc Natl Acad Sci USA* 108, 14682-14687.

**Yamamoto, E., Takashi, T., Morinaka, Y., Lin, S., Wu, J., Matsumoto, T., Kitano, H., Matsuoka, M., and Ashikari, M.** (2010). Gain of deleterious function causes an autoimmune response and Bateson-Dobzhansky-Muller incompatibility in rice. *Mol Genet Genomics* 283, 305-315.

**Yang, S., Li, J., Zhang, X., Zhang, Q., Huang, J., Chen, J.-Q., Hartl, D.L., and Tian, D.** (2013). Rapidly evolving *R* genes in diverse grass species confer resistance to rice blast disease. *Proc Natl Acad Sci USA* 110, 18572-18577.

**Yoshida, K., Saitoh, H., Fujisawa, S., Kanzaki, H., Matsumura, H., Yoshida, K., Tosa, Y., Chuma, I., Takano, Y., Win, J., et al.** (2009). Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* 21, 1573-1591.

**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, S., and Akey, C.W.** (2013). Apoptosome structure, assembly, and procaspase activation. *Structure* 21, 501-515.

**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.

**Žárský, V., Kulich, I., Fendrych, M., and Pečenková, T.** (2013). Exocyst complexes multiple functions in plant cells secretory pathways. *Curr Opin Plant Biol* 16, 726-733.

**Zhang, C., Brown, M.Q., van de Ven, W., Zhang, Z.-M., Wu, B., Young, M.C., Synek, L., Borchardt, D., Harrison, R., Pan, S., et al.** (2016). Endosidin2 targets conserved exocyst complex subunit EXO70 to inhibit exocytosis. *Proc Natl Acad Sci USA* 113, E41-E50.

**Zhang, L., Chen, S., Ruan, J., Wu, J., Tong, A.B., Yin, Q., Li, Y., David, L., Lu, A., Wang, W.L., et al.** (2015a). Cryo-EM structure of the activated NAIP2-NLRc4 inflammasome reveals nucleated polymerization. *Science* 350, 404-409.

**Zhang, S., and Xu, J.-R.** (2014). Effectors and effector delivery in *Magnaporthe oryzae*. PLoS Path 10, e1003826.

**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.** (2017a). Multiple functional self-association interfaces in plant TIR domains. Proc Natl Acad Sci USA 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., Pumplin, N., Ivanov, S., and Harrison, Maria J.** (2015b). EXO70I is required for development of a sub-domain of the periarbuscular membrane during arbuscular mycorrhizal symbiosis. Curr Biol 25, 2189-2195.

**Zhang, Z.-M., Zhang, X., Zhou, Z.-R., Hu, H.-Y., Liu, M., Zhou, B., and Zhou, J.** (2013). Solution structure of the *Magnaporthe oryzae* avirulence protein AvrPiz-t. J Biomol NMR 55, 219-223.

**Zhang, Z.M., Ma, K.W., Gao, L., Hu, Z., Schwizer, S., Ma, W., and Song, J.** (2017b). Mechanism of host substrate acetylation by a YopJ family effector. Nat Plants 3, 17115.

**Zhao, T., Rui, L., Li, J., Nishimura, M.T., Vogel, J.P., Liu, N., Liu, S., Zhao, Y., Dangl, J.L., and Tang, D.** (2015). A truncated NLR protein, TIR-NBS2, is required for activated defense responses in the exo70B1 mutant. PLoS Genet 11, e1004945-e1004945.

**Zhao, Y., Yang, J., Shi, J., Gong, Y.N., Lu, Q., Xu, H., Liu, L., and Shao, F.** (2011). The NLRC4 inflammasome receptors for bacterial flagellin and Type III secretion apparatus. Nature 477, 596-600.

**Zipfel, C.** (2014). Plant pattern-recognition receptors. Trends Immunol 35, 345-351.



# **Appendices**



# Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen

Juan Carlos De la Concepcion<sup>1,6</sup>, Marina Franceschetti<sup>1,6</sup>, Abbas Maqbool<sup>1</sup>,  
Hiromasa Saitoh<sup>2</sup>, Ryohei Terauchi<sup>3,4</sup>, Sophien Kamoun<sup>1,5</sup> and Mark J. Banfield<sup>1\*</sup>

**Accelerated adaptive evolution is a hallmark of plant-pathogen interactions. Plant intracellular immune receptors (NLRs) often occur as allelic series with differential pathogen specificities. The determinants of this specificity remain largely unknown. Here, we unravelled the biophysical and structural basis of expanded specificity in the allelic rice NLR *Pik*, which responds to the effector AVR-Pik from the rice blast pathogen *Magnaporthe oryzae*. Rice plants expressing the *Pikm* allele resist infection by blast strains expressing any of three AVR-Pik effector variants, whereas those expressing *Pikp* only respond to one. Unlike *Pikp*, the integrated heavy metal-associated (HMA) domain of *Pikm* binds with high affinity to each of the three recognized effector variants, and variation at binding interfaces between effectors and *Pikp*-HMA or *Pikm*-HMA domains encodes specificity. By understanding how co-evolution has shaped the response profile of an allelic NLR, we highlight how natural selection drove the emergence of new receptor specificities. This work has implications for the engineering of NLRs with improved utility in agriculture.**

The innate immune systems of plants and animals monitor the extracellular space and the intracellular environment for the presence and activities of microbial pathogens<sup>1,2</sup>. In plants, immune receptors of the NLR (nucleotide-binding, leucine-rich repeat (LRR)) superfamily monitor the intracellular space for signatures of non-self, typically detecting translocated pathogen effector proteins either by direct binding or indirectly via monitoring their activity on host targets<sup>3,4</sup>. Co-evolution between pathogens and hosts has driven the diversification of plant NLRs, with many NLR genes present in allelic series, with distinct effector recognition profiles<sup>5–15</sup>. Pathogen effectors can show strong signatures of positive selection, including high levels of non-synonymous (resulting in amino acid changes) over synonymous polymorphisms<sup>5,7,12,16–18</sup>. How NLR and effector diversification contributes to gene-for-gene immunity in plants is poorly understood. Defining how allelic NLRs recognize and respond to specific pathogen effectors offers new opportunities to engineer the control of plant diseases<sup>19,20</sup>, leading to improved global food security.

Many NLRs function synergistically, with some acting as 'sensors', to detect pathogens, and others as 'helpers', which are required for the initiation of immunity<sup>1,21,22</sup>. These NLRs can be genetically linked in pairs, with a shared promoter<sup>21,23–26</sup>, or unlinked but part of a complex genetic network<sup>27</sup>. One mechanism of effector recognition by sensor NLRs is via unconventional integrated domains that probably have their evolutionary origin as host effector targets<sup>28–31</sup>. Such integrated domains can act as 'baits' to target effectors by direct binding or act as substrates of an effector's enzymatic activity<sup>28,31</sup>. Genetically paired NLRs with integrated domains have repeatedly evolved in rice<sup>29,30</sup> and can detect effectors from the rice blast pathogen *Magnaporthe oryzae* (also known as *Pyricularia oryzae*), the causative agent of the most devastating disease of rice, which is the staple crop that feeds more than half of the world population<sup>5,25,26,32</sup>.

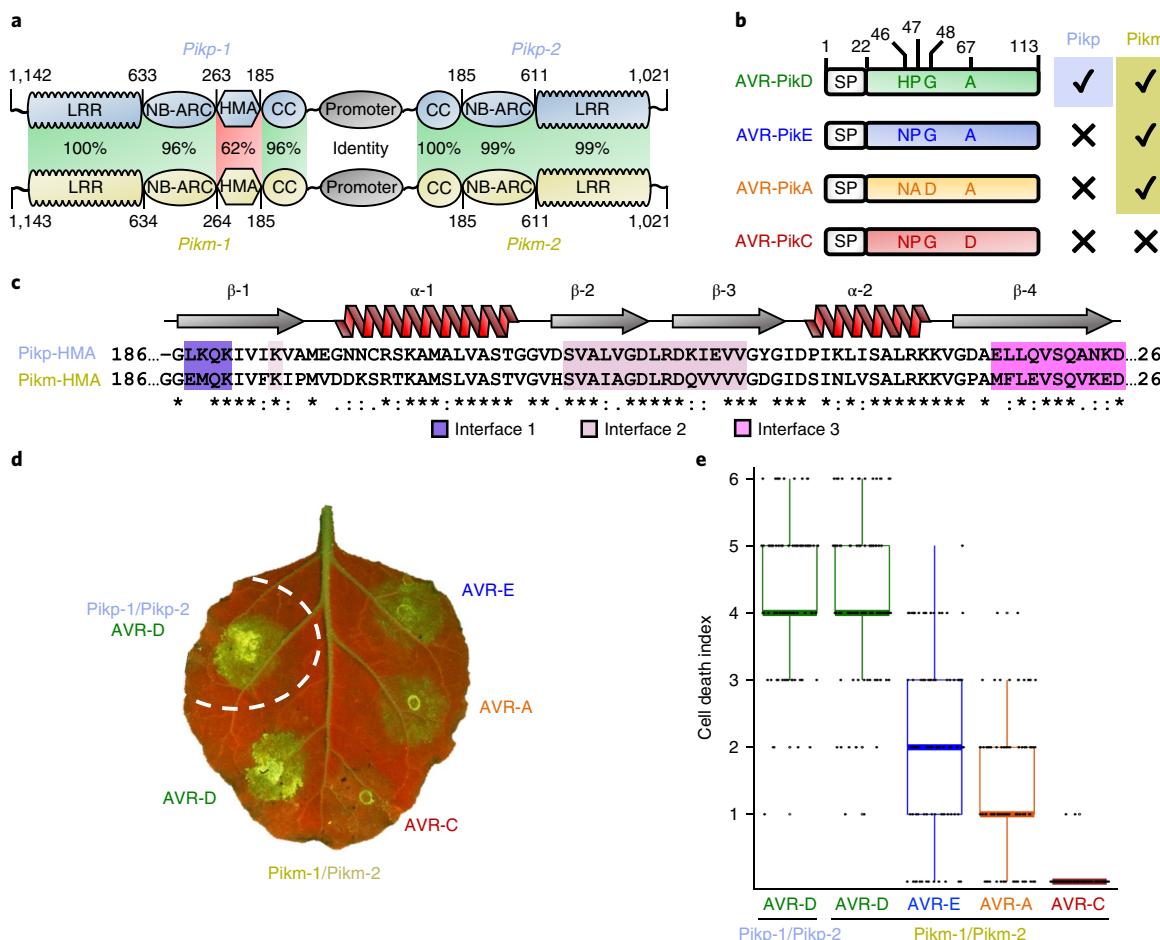
The rice NLR pair *Pik* comprises *Pik-1* (the sensor) and *Pik-2* (the helper). This receptor pair responds to the *M. oryzae* effector AVR-Pik by direct binding to an integrated heavy metal-associated (HMA) domain, positioned between the coiled-coil and the nucleotide-binding domains of *Pik-1*<sup>33</sup> (Fig. 1a). Both the AVR-Pik effectors and the *Pik* NLRs exist as an allelic series in *M. oryzae* and rice, respectively, that most likely arose through co-evolutionary dynamics between pathogen and host<sup>5,34,35</sup>. As such, they represent an excellent system for understanding the mechanistic basis of recognition in plant immunity. A comparison of amino acid sequence identity between the domains of paired *Pik* NLR alleles shows that the integrated HMA domain is the most polymorphic region<sup>35</sup> (Fig. 1a,c), which is consistent with this being the direct binding region for the AVR-Pik effectors. The HMA domain also contains variable amino acids that have been used as markers for *Pik* allele identification in rice<sup>35</sup>. In addition, AVR-Pik is a remarkable example of an effector with an extreme signature of positive selection, as all known AVR-Pik nucleotide polymorphisms are non-synonymous, resulting in amino acid changes<sup>16,18</sup> (Fig. 1b). Furthermore, these polymorphisms map to interface residues identified in the crystal structure of the effector variant AVR-PikD bound to the HMA domain of the NLR allele *Pikp*<sup>33</sup>, suggesting that they are adaptive.

Although rice plants expressing the NLR allele *Pikp* are resistant to *M. oryzae* strains expressing the effector variant AVR-PikD, rice plants expressing the allele *Pikm* respond to strains expressing AVR-PikD, AVR-PikE or AVR-PikA<sup>34</sup> (Fig. 1b). Importantly, neither *Pikp* nor *Pikm* respond to the stealthy effector variant AVR-PikC, which evades detection by any known *Pik* NLR<sup>34</sup>. The molecular mechanism by which *Pik* NLR variation acts to expand effector recognition remains unclear.

Previous work established the structural basis of AVR-PikD recognition by the *Pikp-1* NLR<sup>33</sup>. Here, we reveal how co-evolutionary

<sup>1</sup>Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich, UK. <sup>2</sup>Laboratory of Plant Symbiotic and Parasitic Microbes, Department of Molecular Microbiology, Faculty of Life Sciences, Tokyo University of Agriculture, Tokyo, Japan. <sup>3</sup>Division of Genomics and Breeding, Iwate Biotechnology Research Center, Iwate, Japan. <sup>4</sup>Laboratory of Crop Evolution, Graduate School of Agriculture, Kyoto University, Kyoto, Japan. <sup>5</sup>The Sainsbury Laboratory, Norwich Research Park, Norwich, UK. <sup>6</sup>These authors contributed equally: De la Concepcion, J. C. and Franceschetti, M.

\*e-mail: [mark.banfield@jic.ac.uk](mailto:mark.banfield@jic.ac.uk)



**Fig. 1 |** The *Pikm*-mediated cell death response to AVR-Pik effector variants in *N. benthamiana* phenocopies the *Pikm* resistance profile in rice.

**a**, Schematic representations of Pik NLR alleles. The sensor NLR (*Pikp*-1) and the helper NLR (*Pikp*-2) share a common promoter and the same overall domain architecture. *Pikp*-1 and *Pikp*-2 are shown on the top and *Pikm*-1 and *Pikm*-2 are shown on the bottom. Pairwise protein sequence identity between each domain is indicated, highlighting the diversification of the integrated HMA domain. CC, coiled coil domain; NB-ARC, nucleotide-binding domain.

**b**, Schematic representations of AVR-Pik variants with amino acid polymorphisms shown (single-letter code), along with their *Pikp*-mediated or *Pikm*-mediated response profiles in rice (right)<sup>34</sup>. SP, signal peptide.

**c**, Amino acid sequence alignment of *Pikp*-1 and *Pikm*-1 HMA domains. Secondary structure features of the HMA fold are shown above, and the residues located to the interfaces described in the text and in Fig. 3 are highlighted.

**d**, A representative leaf image showing *Pikm*-mediated cell death to AVR-Pik variants as autofluorescence under UV light; *Pikp*-mediated cell death with AVR-PikD is included as a positive control (surrounded by the dashed circle; no *Pikm*-1/*Pikm*-2 was in this spot).

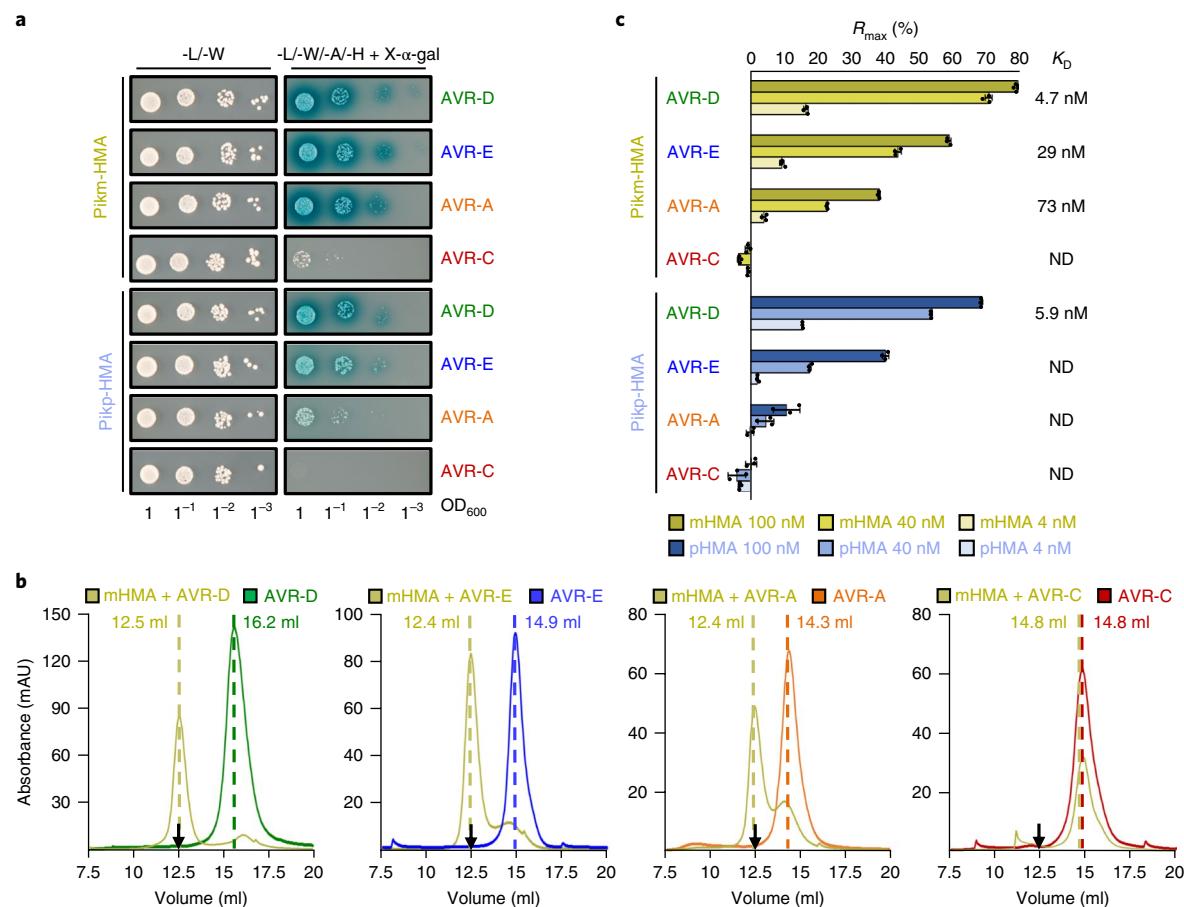
**e**, Box plots showing repeats of the cell death assay. For each sample, the number of repeats was 90. The centre line represents the median, the box limits are the upper and lower quartiles, the whiskers are the 1.5× interquartile range and all of the data points are represented as dots. The cell death scoring scale used is shown in Supplementary Fig. 1d. For brevity, the effectors are labelled without the 'Pik' designation in panels **d** and **e** and, where appropriate, in Figs. 2–6.

dynamics between a pathogen and a host has driven the emergence of new receptor specificities. By taking advantage of our ability to reconstruct complexes between Pik-HMA domains and AVR-Pik effectors, and to recapitulate cell death responses (indicative of immunity) in the model plant *Nicotiana benthamiana*, we show a correlation between protein-binding affinities and the activation of immunity. By obtaining crystal structures of the *Pikm*-HMA domain (henceforth *Pikm*-HMA) in complex with three different AVR-Pik variants, we define the interfaces that support expanded effector recognition. We also obtained new structures of the *Pikp*-HMA domain (henceforth *Pikp*-HMA) in complex with the recognized effector AVR-PikD, but also with the unrecognized AVR-PikE. Together, these structures establish a previously unappreciated role for the carboxy terminus of the HMA domain in mediating effector interaction. Understanding how host NLRs have evolved new specificities in response to pathogen effectors highlights the potential to engineer new-to-nature receptors with improved functions, such as

recognition of stealthy effector variants, and has broad implications for rational design of plant NLRs.

## Results

**Pikm-mediated cell death in *N. benthamiana* recapitulates allele-specific effector responses in rice.** *Pikp*-mediated cell death in *N. benthamiana* phenocopies effector variant-specific resistance in rice, with *Pikp* responding to AVR-PikD, but not to AVR-PikE, AVR-PikA or AVR-PikC<sup>33</sup>. Here, we show that *Pikm* responds to each of AVR-PikD, AVR-PikE or AVR-PikA, but not to AVR-PikC, in this assay (Fig. 1d,e and Table 1). These results match the response of rice cultivars expressing *Pikm* to *M. oryzae* strains encoding the effectors<sup>34</sup>. Interestingly, we observe a qualitative hierarchy in the level of *Pikm*-mediated cell death in response to the effectors in the order AVR-PikD > AVR-PikE > AVR-PikA (Fig. 1d,e). To allow for direct comparison, we repeated this assay using the *Pikp* NLRs and the effector variants in the same expression vectors. We obtained equivalent results to those



**Fig. 2 | Different affinities underpin the recognition and response of Pik NLR alleles to AVR-Pik effector variants.** **a**, Y2H demonstrates the binding of effector variants to both Pikm-HMA and Pikp-HMA. The control plate for yeast growth is on the left, with the selective plate on the right. Each experiment was repeated a minimum of three times, with similar results. **b**, Analytical gel filtration confirms that Pikm-HMA (mHMA) forms complexes with AVR-PikD, AVR-PikE and AVR-PikA in vitro, but not with AVR-PikC. Note that earlier elution correlates with increased molecular mass. The retention volumes for peaks are labelled (the black arrow indicates the Pikm-HMA elution volume; Pikm-HMA does not absorb light at 280 nm). SDS-PAGE with relevant fractions are shown in Supplementary Fig. 2b. Each experiment was repeated a minimum of three times, with similar results. **c**, SPR reveals that the in vitro binding affinity between Pik-HMA and effectors correlates with in planta responses.  $R_{\max}$  (%) is the percentage of the theoretical maximum response, assuming a 1:1 binding model for Pikm (effector:HMA) and a 1:2 binding model for Pikp, at the HMA concentrations shown. Bars represent the average of three measurements, and the error bars represent the standard deviation. Where  $K_D$  values are given, a wider range of HMA concentrations were used for this calculation (see Supplementary Fig. 2c–e,g). pHMA, Pikp-HMA; ND, not determined.

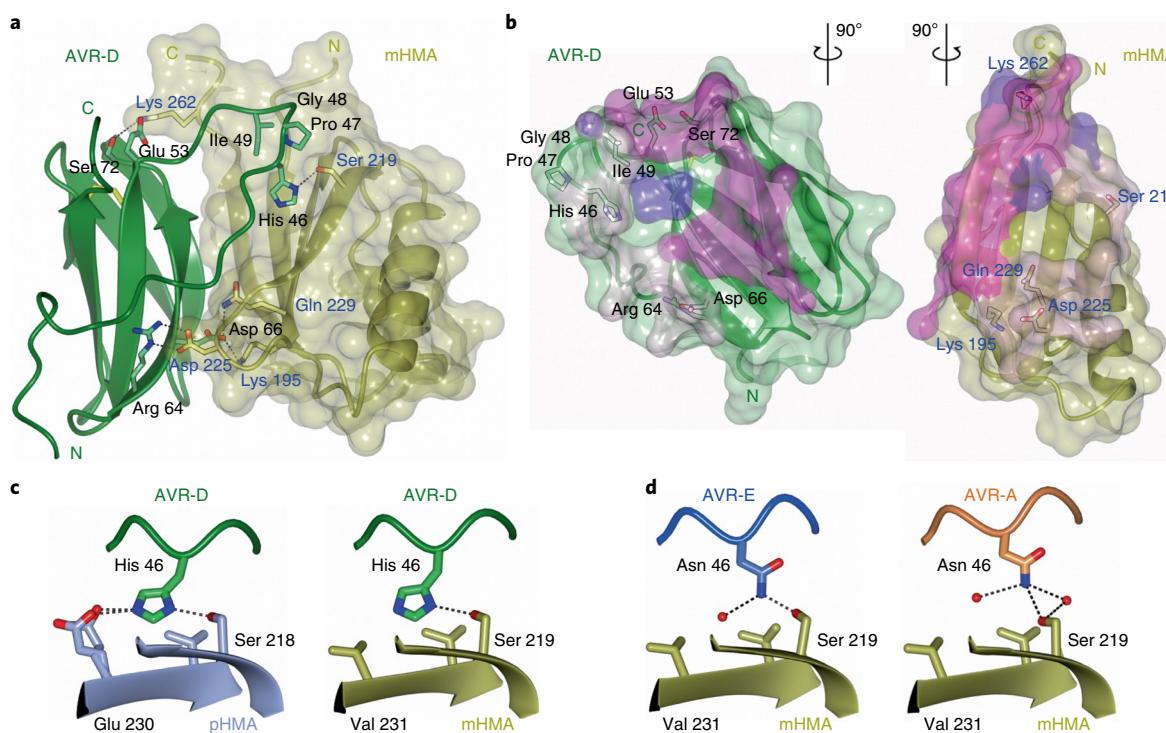
shown previously<sup>33</sup> (Supplementary Fig. 1a,b). The expression of each protein was confirmed by western blot (Supplementary Fig. 1c).

**Allele-specific effector responses in planta correlates with direct Pik-HMA interactions.** We used yeast-2-hybrid (Y2H) to investigate whether the binding of effectors to the Pikp-HMA or Pikm-HMA correlates with in planta response profiles. We observed comparable growth of yeast on selective plates and the development of blue colouration with X- $\alpha$ -gal (which are both indicative of protein–protein interactions) with Pikm-HMA and AVR-PikD, AVR-PikE and AVR-PikA, but not with AVR-PikC (Fig. 2a). Although the Y2H assay with Pikm-HMA or Pikp-HMA showed comparable interaction with AVR-PikD, Pikm-HMA showed increased interaction with AVR-PikE and markedly stronger interaction with AVR-PikA (Fig. 2a). No growth was observed with Pikp-HMA and AVR-PikC. All proteins were confirmed to be expressed in yeast (Supplementary Fig. 2a).

**Pikm-HMA has tighter binding affinities for AVR-Pik effectors than Pikp-HMA in vitro.** To produce stable Pikm-HMA protein for in vitro studies, we cloned a construct with a 5-amino

acid extension at the C terminus (encompassing residues Gly 186–Asp 264 of the full-length protein) compared to the previously studied Pikp-HMA<sup>33</sup>. Using gel filtration with separately purified proteins, Pikm-HMA forms complexes with the effectors AVR-PikD, AVR-PikE or AVR-PikA, but not with AVR-PikC (Fig. 2b and Supplementary Fig. 2b).

To determine the extent to which the expanded response of Pikm to AVR-Pik effectors in *N. benthamiana* is related to the strength of binding to the Pikm-HMA, we determined binding affinities by surface plasmon resonance (SPR). We monitored the response units following Pikm-HMA injection after capturing effectors on the chip surface. Binding of Pikm-HMA to the different effectors was measured at three different concentrations, and the response units were normalized to  $R_{\max}$  (a theoretical maximum response, assuming a 1:1 interaction model). From this, we ranked the order of apparent affinity from highest to lowest (Fig. 2c). We then extended the Pikm-HMA concentration range to enable the estimation of the equilibrium dissociation constant ( $K_D$ ). Using a 1:1 kinetics interaction model, we found that Pikm-HMA bound to AVR-PikD with the highest affinity (lowest  $K_D$ ), followed by AVR-PikE and AVR-PikA (Fig. 2c, Supplementary Fig. 2c–e and Supplementary Table 1).



**Fig. 3 | Structures of Pikm-HMA in complex with AVR-Pik effectors.** **a**, Schematic representation of the structure of Pikm-HMA (mHMA) in complex with AVR-PikD. Pikm-HMA is shown with the selected side chains as sticks; the molecular surface of this domain is also shown. AVR-PikD is shown with the selected side chains as sticks. Hydrogen bonds or salt bridges are shown as dashed lines and the disulfide bond as a yellow bar. Amino acids labelled in black are from AVR-D, those labelled in blue are from mHMA. **b**, The buried surface area of AVR-PikD and Pikm-HMA shown from the perspective of the partner (the change in orientation from panel **a** is indicated). The buried surfaces are coloured according to interfaces described in the text (interface 1 is in purple, interface 2 is in pink and interface 3 is magenta). **c**, Close-up views (part of interface 2) of the orientation and interactions of AVR-PikD (His 46) in the Pikp-HMA (pHMA) and Pikm-HMA complexes. **d**, Close-up views (part of interface 2) of the orientation and interactions of AVR-PikE (Asn 46) (left panel) and AVR-PikA (Asn 46) (right panel) in complex with Pikm-HMA. Water molecules are shown as red spheres.

We observed no significant binding of Pikm-HMA to AVR-PikC (Fig. 2c, Supplementary Fig. 2f and Supplementary Table 1).

We also produced Pikp-HMA with its equivalent 5-amino acid C-terminal extension (including residues Gly 186–Asp 263 of the full-length protein) and analysed effector binding by SPR (Fig. 2c). We ranked effector-binding affinities in the order AVR-PikD > AVR-PikE > AVR-PikA (with no significant binding to AVR-PikC and assuming a 1:2 (effector:Pikp-HMA) interaction model, as previously observed<sup>33</sup>). However, we were only able to reliably determine the  $K_D$  for Pikp-HMA bound to AVR-PikD (Fig. 2c and Supplementary Fig. 2g), as the binding of AVR-PikE and AVR-PikA were of insufficient quality under our assay conditions to allow the  $K_D$  to be determined (Supplementary Fig. 2h,i).

Based on these results and the interactions monitored by Y2H, we conclude that the differential binding affinity to the HMA domains is the source of the allele-specific response profile in *N. benthamiana* and of rice cultivars to *M. oryzae* strains expressing AVR-Pik variants<sup>34</sup>.

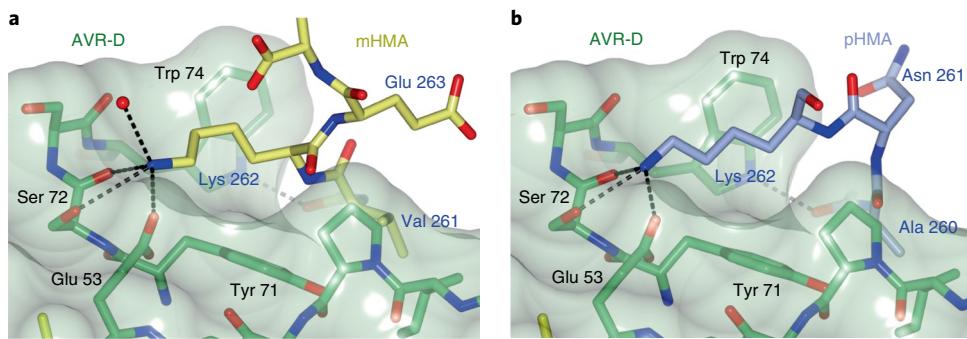
**Structures of Pik-HMAs in complex with AVR-Pik effectors reveal multiple interaction surfaces.** Using a co-expression strategy, we obtained complexes of Pikm-HMA bound to AVR-PikD, AVR-PikE or AVR-PikA. Each of these were crystallized and X-ray diffraction data were collected at the Diamond Light Source (Oxford, UK) to 1.2-Å, 1.3-Å and 1.3-Å resolution, respectively. Details of the X-ray data collection, structure solution and structure completion are given in the Methods section and Supplementary Table 2. The overall orientations of each component in the Pikm-HMA–effector complexes are similar to each other and to the

previously determined Pikp-HMA–AVR-PikD structure<sup>33</sup> (Fig. 3a, Supplementary Fig. 3a,b and Supplementary Table 3). Interestingly, the Pikm-HMA–effector structures form a 1:1 complex, in contrast to Pikp-HMA–AVR-PikD, which formed a 2:1 complex<sup>33</sup>. Pikp-HMA dimerization is most likely an artefact of in vitro protein expression and purification.

Analysis of the interfaces formed between Pikm-HMA and the effectors using QtPISA<sup>36</sup> (Supplementary Table 4 and Supplementary Fig. 4) reveals they are broadly similar to each other, although there is a trend of reducing the total interface area in the order AVR-PikD > AVR-PikE > AVR-PikA. Graphical representation of key interface components (using QtPISA interaction radars<sup>36</sup>; Supplementary Fig. 4) reveals a high likelihood that each interface is biologically relevant: each key component value lies well above the 50% threshold when considered against statistical distributions derived from the Protein Data Bank (PDB) (see Methods and ref. <sup>36</sup>).

Three predominant regions can be identified within each Pikm-HMA–effector interface (Figs. 1c and 3b). These regions (interfaces) are defined here from the HMA side as: interface 1, amino-terminal residues Glu 188–Lys 191; interface 2, residues from β-2 to β-3 (Ser 219–Val 233) and Lys 195 from β-1; and interface 3, residues from β-4 to the C terminus (Met 254–Asp 264) (Figs. 1c and 3b).

Interface 1 is a minor component of the Pikm-HMA–effector interaction, with a single, weak hydrogen bond formed by the side chain of Lys 191 (to the main-chain carbonyl group of Thr 69 of the effector) and a hydrophobic interface contributed by the side chain of Met 189 (to the side chain of Ile 49 of the effector). Interface 2 is more extensive and predominately interacts with AVR-Pik residues from the N-terminal extension of the conserved MAX effector fold<sup>37</sup>,



**Fig. 4 | Different interactions at interface 3 in the complexes of Pikm-HMA and Pikp-HMA with AVR-PikD support recognition and response.**

A close-up view of the interactions across interface 3 in the Pikm-HMA (mHMA; **a**) and Pikp-HMA (pHMA; **b**) complexes with AVR-PikD, showing different conformations for the C-terminal regions of the HMA domains. In particular, note the looping out of Asn 261 of Pikp-HMA (**b**) and the different orientation of the Lys 262 side chain (**a,b**). In each panel, AVR-PikD is shown with the side chains as sticks; the molecular surface of the effector is also shown. The Pik-HMA domains are coloured as labelled.

including Arg 39–Phe 44 and His 46–Ile 49. This interface includes the polymorphic residues at positions 46, 47 and 48 of the effector variants<sup>34</sup> (Figs. 1b and 3). Interface 2 also includes salt-bridge or hydrogen-bond interactions via the side chains of Asp 225 (to Arg 64 of the effectors) and Lys 195 (to Asp 66 of the effectors; Fig. 3a). Finally, interface 3 includes both main-chain hydrogen-bonding interactions between  $\beta$ -4 of the HMA and  $\beta$ -3 of the effectors, and inserts the side chain of Lys 262 into a surface pocket on the effector lined by residues Glu 53, Tyr 71, Ser 72 and Trp 74. Lys 262 makes several interactions in this pocket, including salt bridges or hydrogen bonds with the side chains of Glu 53 and Ser 72 (Figs. 3a and 4a).

We also obtained crystal structures of Pikp-HMA, with the 5-amino acid extension at the C terminus of the HMA, bound to AVR-PikD or AVR-PikE at 1.35-Å and 1.9-Å resolution, respectively (see Methods, Supplementary Table 2 and Supplementary Fig. 3c,d). The Pikp-HMA–AVR-PikE combination does not give rise to responses in planta, but we were able to obtain the complex in solution. The new structure of the Pikp-HMA–AVR-PikD complex is essentially identical to that previously determined<sup>33</sup>, except for the 5-amino acid extension. Interface analysis with QtPISA (Supplementary Table 4 and Supplementary Fig. 4) reveals that the Pikp-HMA–AVR-PikD complex has broadly similar properties to those of Pikm-HMA–effectors (the total interface area and the key component values are well above the 50% threshold in the interaction radars). By contrast, although the Pikp-HMA–AVR-PikE interface shows a broadly similar total interface area to the other complexes, the total calculated binding energy is reduced (the area of the polygon in Supplementary Fig. 4) and five out of six key interface components fall below the 50% threshold, questioning the biological relevance of this interface.

**Structural changes at interface 2 underpin differential effector recognition by Pikm.** Effector variants AVR-PikD, AVR-PikE and AVR-PikA differ at amino acid positions 46, 47 and 48, which localize to interface 2 (Figs. 1b and 3b). Pikp-HMA binds to AVR-PikD (His 46) via hydrogen bonds with residues Ser 218 and Glu 230<sup>33</sup>. In Pikm, the Ser is conserved, but Glu 230 is replaced by Val 231 at the structurally equivalent position, resulting in the loss of a direct hydrogen bond. Despite this, AVR-PikD (His 46) occupies the same position in both complexes (Fig. 3c). Surprisingly, in the Pikm-HMA–AVR-PikE complex, AVR-PikE (Asn 46) is rotated out of the binding pocket, well away from Val 231 (Fig. 3d), and a water molecule occupies the resulting space. Hydrogen bonds are formed between AVR-PikE (Asn 46:N82) and both Pikm-HMA (Ser 219:OH) and the new water molecule. This configuration affects the position of effector residues Phe 44–Gly 48, pushing them away from the HMA domain, further altering interactions across

interface 2. These structural changes correlate with a reduced binding affinity of AVR-PikE with Pikm-HMA compared to AVR-PikD. In the Pikm-HMA–AVR-PikA complex, Asn 46 is rotated even further out of the HMA pocket, and, although a hydrogen bond is still formed with Pikm-HMA (Ser 219:OH), this is substantially different in orientation (Fig. 3d). These changes serve to move residues Asn 46–Pro 50 of AVR-PikA further away from the HMA domain, and again, these structural observations correlate with a reduced effector binding affinity. Interestingly, the polymorphic residues in AVR-PikA (Ala 47 and Asp 48) have no direct role in Pikm-HMA interaction. The polymorphisms in AVR-Pik do not significantly alter protein–protein interactions across interfaces 1 and 3, and these regions seem to stabilize the complexes.

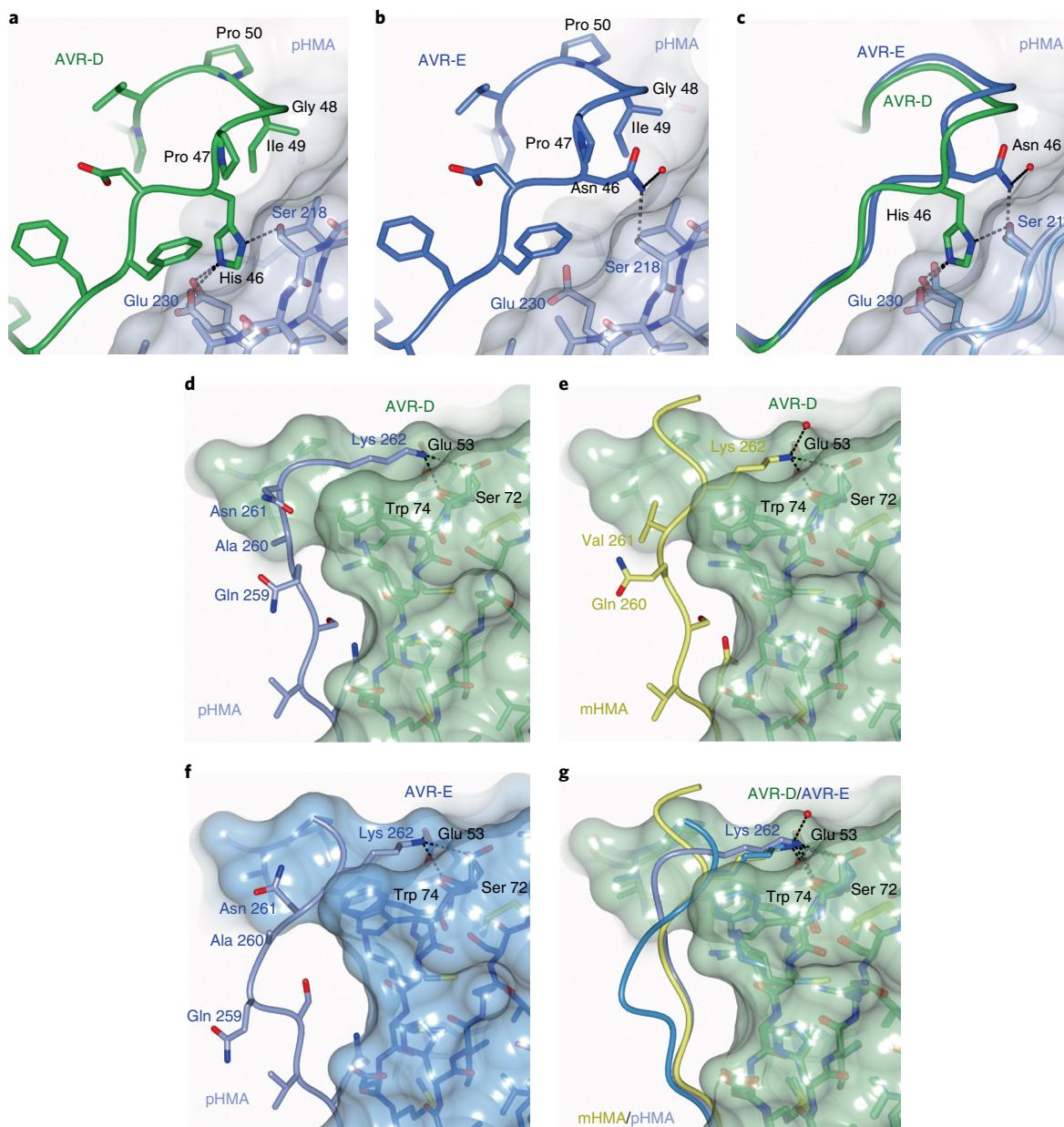
We conclude that the structural changes at interface 2 underlie the weaker binding affinities of Pikm-HMA for AVR-PikE and AVR-PikA than for AVR-PikD.

**Interactions across interface 3 contribute more to Pikm-HMA than to Pikp-HMA binding to AVR-PikD.** As observed at interface 3 for the Pikm-HMA–effector complexes (Fig. 4a), a Lys residue from Pikp-HMA (Lys 262) locates to the binding pocket on the effector containing Glu 53 and Ser 72 (Fig. 4b). However, this Lys is shifted by one residue to the C terminus in the sequence of Pikp-1 (Fig. 1c). This results in a different conformation of the Pikp-HMA residues Ala 260 and Asn 261 when compared to Pikm-HMA (Val 261 and Lys 262), changing the interactions across interface 3. The most dramatic difference is the ‘looping out’ of Pikp-HMA (Asn 261) to retain Lys 262 in the effector-binding pocket (Figs. 4b and 5d,e), which affects the packing of Pikp-HMA (Ala 260) (Val 261 in Pikm-HMA) and the hydrophobic packing of the side chain of Lys 262.

Pik alleles also differ in the composition of residues at interfaces 1 and 2. Of most importance are the changes at interface 2 that contact AVR-PikD (His 46), as discussed above and in Fig. 3c.

We propose that Pikm has evolved more-robust interactions across interface 3 than Pikp to compensate for loss of binding, such as direct hydrogen bonds, at interface 2.

**Interactions across interfaces 2 and 3 underpin the specificity of Pikp to AVR-PikD over AVR-PikE.** Underpinning the global analyses of the Pikp-HMA–AVR-PikD and Pikp-HMA–AVR-PikE complexes are extensive differences at interfaces 2 and 3. At interface 2, AVR-PikE (Asn 46) is fully rotated out of the AVR-PikD (His 46) binding pocket (Fig. 5a–c). A hydrogen bond is still formed between AVR-PikE (Asn 46) and Pikp-HMA (Ser 218), but in a very different orientation (Fig. 5a–c). This results in residues Asn 46–Pro 50 moving away from the HMA domain. This re-configuration

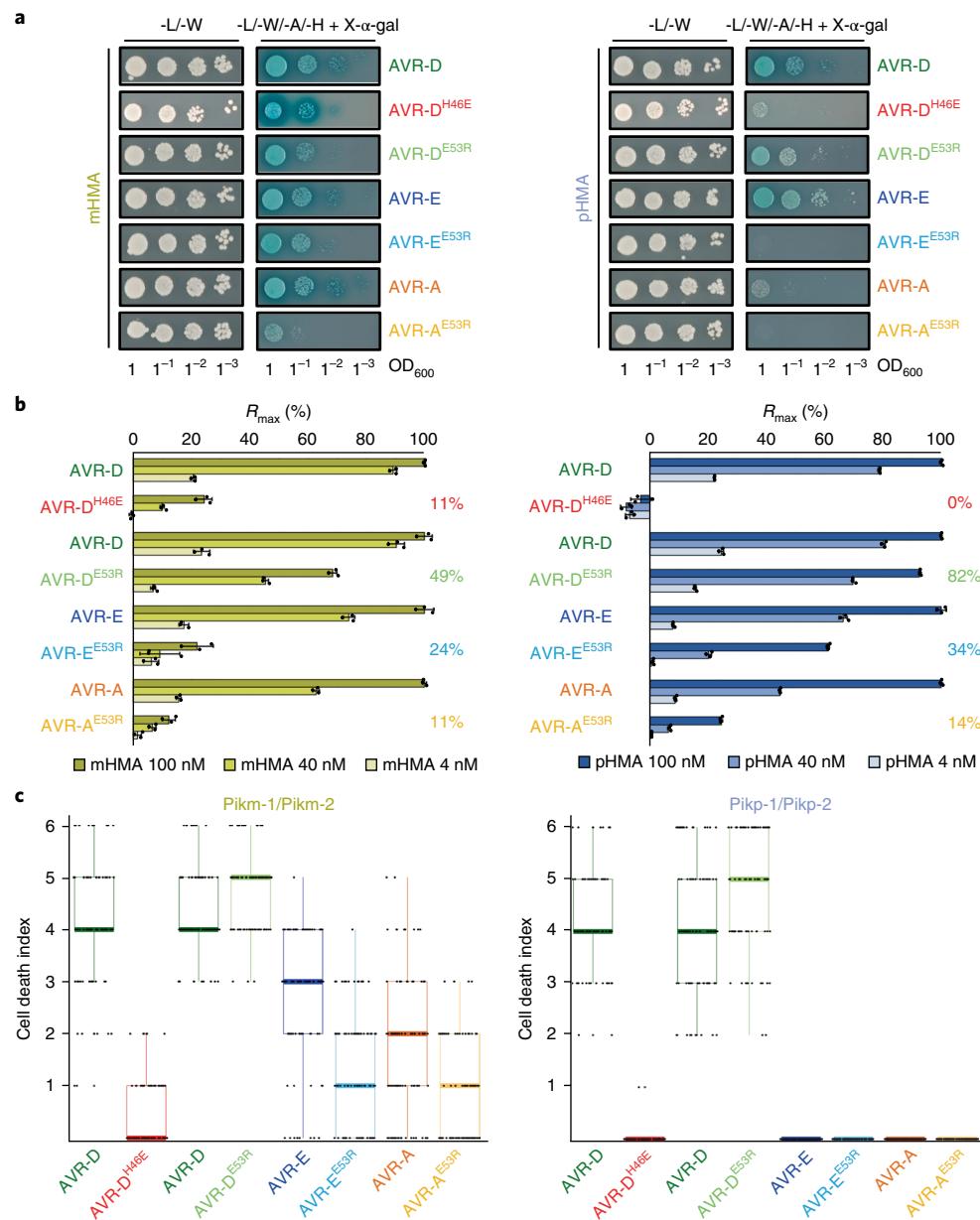


**Fig. 5 | Altered interactions across the interfaces of Pikp-HMA with AVR-PikD and AVR-PikE underpin the differences in recognition and response.**  
**a, b**, Zoomed-in views of the interactions across interface 2 in the Pikp-HMA (pHMA) complexes with AVR-PikD (**a**) and AVR-PikE (**b**). In each panel, the molecular surface of Pikp-HMA is shown. Effector variant residues are coloured as labelled and shown in C- $\alpha$ -worm with side-chain representation.  
**c**, Superposition of panels **a** and **b**, with only selected side chains shown for clarity. The polymorphism at position 46 occupies a very different position, fully flipped out of the His 46 binding pocket in the AVR-PikE structure, which alters the position of residues Asn 44-Pro 50 relative to Pikp-HMA.  
**d–f**, Zoomed-in views of the interactions across interface 3 in the Pikp-HMA complex with AVR-PikD (**d**), the Pikm-HMA (mHMA) complex with AVR-PikD (**e**) and the Pikp-HMA complex with AVR-PikE (**f**). In each panel, the effector is shown as sticks and the molecular surface is also shown and coloured as labelled. Pik-HMA residues are coloured as labelled and shown in the C- $\alpha$ -worm with side-chain representation. The looping out of Asn 261 in Pikp compared to Pikm, when in complex with AVR-PikD, is seen in panels **d** and **e**, and the displacement of residues Gln 259 and Ala 260 in Pikp, between the complexes with AVR-PikD and AVR-PikE, is seen in panels **d** and **f**. **g**, Superposition of panels **d–f**, with only the side chain of Pik-HMA Lys 262 and only the surface of AVR-PikD, shown for clarity.

is coupled with changes at interface 3 (Fig. 5d,f,g). Interestingly, in the Pikp-HMA–AVR-PikE complex, Lys 262 adopts a similar orientation to that found in the Pikm-HMA complexes (Fig. 5e–g). However, to enable this, residues Ser 258–Asn 261 adopt a dramatically different position, by looping out residues Gln 259 and Ala 260 from their positions in the Pikm-HMA complex (Fig. 5e–g), with consequent effects on this interface.

We conclude that interface 2 is key for effector recognition by Pikp and, unlike for Pikm, interfaces 1 and 3 are not able to compensate to enable productive binding.

**Mutations at separate interfaces have differential effects on Pik-HMA–effector interactions and immunity phenotypes.** We subsequently tested whether mutations in the effectors at interfaces 2



**Fig. 6 | Mutations at different interfaces in the Pik-HMA-effector complexes have differential effects on interactions and phenotypes.** **a**, Effector mutations at positions 46 and 53 perturb interactions with Pikm-HMA (mHMA) and Pikp-HMA (pHMA) as assayed by Y2H. Each experiment was repeated a minimum of three times, with similar results. Single-letter amino acid codes are shown. **b**, Changes in in vitro binding for effector mutants with Pikm-HMA and Pikp-HMA, as measured by SPR. The R<sub>max</sub> (%) was calculated as described in the text. To emphasize the altered binding for each effector mutant, the averaged difference R<sub>max</sub> (%) across the three different concentrations measured is shown. Bars represent the average of three measurements, with the error bars representing the standard deviation. **c**, Box plots of Pikm-mediated or Pikp-mediated cell death triggered by the effector mutants. For each sample, the number of repeats was 90. The centre line represents the median, the box limits are the upper and lower quartiles, the whiskers are the 1.5 $\times$  interquartile range and all of the data points are represented as dots.

and 3 have differential effects on Pik-HMA binding and responses by Y2H, SPR and *N. benthamiana*. We used the previously characterized AVR-PikD (His46Glu) mutant at interface 2 and a Glu53Arg mutant at interface 3 in AVR-PikD, AVR-PikE and AVR-PikA. Although AVR-PikD (His 46) occupies a central position at interface 2, AVR-Pik (Glu 53) locates to the Pik-HMA (Lys 262)-binding pocket, at the periphery of interface 3.

As previously observed (although without the C-terminal extension<sup>33</sup>), the AVR-PikD (His46Glu) mutant essentially blocks the Pikp-HMA–effector interaction in Y2H and SPR, and abolishes Pikp-mediated cell death in *N. benthamiana* (Fig. 6a–c and

Supplementary Fig. 5). Interestingly, the AVR-PikD (His46Glu) mutant interacts with Pikm-HMA in Y2H (Fig. 6a). However, when measured by SPR, Pikm-HMA binding to this mutant is reduced to ~11% compared to the wild type (Fig. 6b). This reduction of binding in vitro is reflected in *N. benthamiana*, where we observe weak AVR-PikD (His46Glu)-dependent Pikm cell death (Fig. 6c and Supplementary Fig. 5b–d).

For each of the Glu53Arg effector mutants, we observe little effect on Pikm-HMA interaction in Y2H compared to the wild type, except a reduced interaction of AVR-PikA (Glu53Arg) (Fig. 6a). Interestingly, the Glu53Arg mutant in AVR-PikE abolishes the interaction of this

**Table 1 |** The various interactions and phenotypes between Pik NLR alleles and effector variants in this study

	AVR-D	AVR-E	AVR-A	AVR-C	AVR-D <sup>H46E</sup>	AVR-D <sup>E53R</sup>	AVR-E <sup>E53R</sup>	AVR-A <sup>E53R</sup>
Interaction in Y2H	<b>Pikp</b>	+++	++	+	—	+	++	—
	<b>Pikm</b>	+++	+++	+++	+	+++	++	+
Interaction in SPR	<b>Pikp</b>	+++	++	+	—	—	+++	+
	<b>Pikm</b>	+++	+++	++	—	+	++	—
Recognition in rice plants	<b>Pikp</b>	+++ <sup>a</sup>	+ <sup>a</sup>	— <sup>b</sup>	— <sup>b</sup>	— <sup>a</sup>	ND	ND
	<b>Pikm</b>	+++ <sup>b</sup>	+++ <sup>b</sup>	+++ <sup>b</sup>	— <sup>b</sup>	ND	ND	ND
CD response in <i>N. benthamiana</i>	<b>Pikp</b>	+++	—	—	—	—	+++	—
	<b>Pikm</b>	+++	++	+	—	+	+++	+

SPR and Y2H interactions used the isolated HMA domains, and in planta experiments were performed with full-length proteins. Recognition in rice plant *Pikp* is rice cv. K60. Recognition in rice plant *Pikm* is rice cv. Tsuyuake. CD, cell death; ND, not determined. <sup>a</sup>See ref. <sup>33</sup>. <sup>b</sup>See ref. <sup>34</sup>.

effector with *Pikp*-HMA in Y2H. Using SPR, the AVR-Pik (Glu53Arg) mutants show reduced binding to both *Pik*-HMA domains when compared pairwise to the wild type in each effector background (Fig. 6b). However, in each case, the Glu53Arg mutant has a greater effect in *Pikm*-HMA binding than in *Pikp*-HMA binding. Surprisingly, in the *N. benthamiana* cell death assay, we observe a slight increase in the AVR-PikD (Glu53Arg)-dependent cell death compared to the wild type for both *Pikp* and *Pikm* (Fig. 6c and Supplementary Fig. 5b–d). However, we see a reduction in the intensity of *Pikm*-mediated cell death for the effector variants AVR-PikE (Glu53Arg) and AVR-PikA (Glu53Arg) (Fig. 6c and Supplementary Fig. 5b–d).

We conclude that interactions across interface 2 are critical for effector recognition by *Pikp* and important for *Pikm*, and interface 3 has an important role in the extended response of *Pikm* to AVR-PikE and AVR-PikA.

## Discussion

Despite intensive study, 25 years since the cloning of the first plant NLRs<sup>38–40</sup>, very little is known about the molecular mechanistic basis of how these proteins recognize pathogen effectors and initiate immune signalling. The recent identification of plant NLRs with integrated domains<sup>28–30</sup> has enabled new opportunities to investigate how these receptors directly recognize pathogen effectors at the biochemical and structural level, and how these binding events are linked to disease resistance<sup>33,41–44</sup>. Here, we have generated five structures of different complexes between the integrated domains of an allelic NLR (*Pik*), and the variants of the effector (AVR-Pik) they recognize. When combined with the analysis of biophysical interactions in vitro and cell death responses in the model plant *N. benthamiana*, these structures provide new understanding, and unexpected findings, on how co-evolution has driven the emergence of new plant NLR receptor specificities.

High levels of diversifying selection in allelic plant NLRs and pathogen effectors suggest direct interaction between the proteins. Previous studies where structures of the effectors, but not the interacting NLR domain, were available showed that distributed surface-presented residues on the effectors defined NLR recognition specificity, mediated by polymorphic LRR domains<sup>14,15</sup>. The integrated HMA domains are the most polymorphic regions of the rice *Pik*-1/*Pik*-2-paired NLRs, and *Pik*-HMA amino acids that form the interfaces with effectors are probably under the strongest selective pressure. Thus, during the course of plant–pathogen co-evolution, at least two alternative solutions for recognizing divergent effectors have emerged. One of these involves the integration and diversification of non-canonical domains in the NLR architecture. The second involves the diversification of LRR domains. An important question raised by these studies is what has driven the emergence of these different systems? An advantage of the integrated domain is that (once stably incorporated) it may tolerate the accelerated accumulation of

mutations, followed by selection for function, as mutations may be less likely to disrupt the overall structure and function of the NLRs.

One outcome from this work is the surprising plasticity of the *Pik*-HMA interfaces that supports differential recognition of AVR-Pik variants. Interactions across interface 2 are important for effector binding by *Pikp*-HMA and *Pikm*-HMA. Disruption of interface 2 by amino acid polymorphisms in AVR-PikE and AVR-PikA eliminates *Pikp*-mediated cell death in planta and weakens *Pikm*-mediated cell death. The unique polymorphism that defines AVR-PikC (Ala67Asp) also maps to interface 2 and may result in a steric clash preventing, or severely reducing, *Pik*-HMA binding. Our structural data support a conclusion that more-favourable interactions across interface 3 have evolved in *Pikm*-HMA to, in-part, compensate for the effect of AVR-Pik variation at interface 2 and support cell death signalling. Our biophysical data indicate that quantitative binding differences, visualized as a disruption of interfaces in the structures, underpin differential effector recognition by *Pik*-HMAs and that a threshold of binding is required for the activation of response in planta. These insights will inform future structure–function studies to address whether rational engineering of *Pik*-HMA–effector-binding interfaces can generate NLR receptors with improved recognition profiles. Ultimately, we must understand how the recognition of effectors, through either integrated domains or other mechanisms, results in the triggering of immune responses in the context of the full-length proteins and, potentially, oligomeric states.

## Methods

**Gene cloning.** For details of gene cloning, please see Supplementary Methods.

**Expression and purification of proteins for in vitro binding studies.** pOPINM, which encodes *Pikm*-HMA or *Pikp*-HMA, was transformed into *Escherichia coli* SHuffle cells<sup>35</sup>. Inoculated cell cultures were grown in autoinduction media<sup>46</sup> at 30 °C for 6 h and 18 °C overnight. Cells were harvested and proteins were extracted as previously reported<sup>33</sup>. AVR-Pik effectors with a cleavable N-terminal SUMO (small ubiquitin-like modifier) or MBP (maltose-binding protein) tag and a non-cleavable C-terminal 6xHis tag were produced in and purified from *E. coli* SHuffle cells as previously described<sup>33</sup>, using either autoinduction media<sup>46</sup> or Power Broth (Molecular Dimensions).

The protein concentration of AVR-Pik effectors was determined by absorption at 280 nm using a NanoVue spectrophotometer (GE Life Sciences). Measurements were corrected using the molar extinction coefficient 25,105 M<sup>−1</sup>cm<sup>−1</sup>, as calculated by Expasy (<http://web.expasy.org/protparam>). Owing to the lack of aromatic residues in *Pik*-HMA, protein concentrations were measured using a Direct Detect Infrared Spectrometer (Merck).

## Co-expression and purification of *Pik*-HMA–AVR-Pik effectors for crystallization.

Relevant *Pik*-HMA domains and AVR-Pik effectors were co-expressed in SHuffle cells following co-transformation of pOPINM:*Pik*-HMA and pOPINA:AVR-Pik, as previously described<sup>33</sup>. Cells were grown in autoinduction media (supplemented with both carbenicillin and kanamycin), harvested and processed as described in the Supplementary Methods. Protein concentrations were measured by absorbance at 280 nm using a NanoVue spectrophotometer

and an extinction coefficient of  $25,105\text{ M}^{-1}\text{ cm}^{-1}$  for Pikm-HMA complexes and  $26,720\text{ M}^{-1}\text{ cm}^{-1}$  for Pikp-HMA complexes, as calculated by Expasy (<http://web.expasy.org/protparam>).

**Protein–protein interaction.** *Analytical gel filtration.* Pikm-HMA and the AVR-Pik effectors were mixed in a molar ratio of 2/1 and incubated on ice for 60 min. In each case, a sample volume of 110  $\mu\text{l}$  was separated at  $4^\circ\text{C}$  on a Superdex 75 10/300 size exclusion column (GE Healthcare), pre-equilibrated in buffer B and at a flow rate of  $0.5\text{ ml min}^{-1}$ . Fractions of 0.5 ml were collected for analysis by SDS-PAGE. The Superdex 75 10/300 column has a void volume of 7.4 ml and a total volume of 24 ml.

**SPR.** SPR experiments to analyse protein–protein interactions were performed on a Biacore T200 system (GE Healthcare) using an NTA sensor chip (GE Healthcare). All proteins were prepared in SPR running buffer (20 mM HEPES (pH 7.5), 860 mM NaCl and 0.1% Tween 20). Details of the cycling conditions are given in the Supplementary Methods.

The  $K_D$  for Pikm-HMA binding to AVR-Pik alleles and Pikp-HMA binding to AVR-PikD were determined from multicycle kinetics curves using the Biacore T200 BioEvaluation software (GE Healthcare), with a 1:1 or 2:1 fit model, respectively. For the interaction between Pikp-HMA and AVR-PikE and AVR-PikA, and for both Pik-HMAs and the AVR-Pik mutants, it was not possible to accurately determine the  $K_D$  owing to the insufficient quality of the data. In these cases, the level of binding was expressed as a percentage of the  $R_{\max}$  normalized for the amount of ligand immobilized on the chip. SPR data were exported and plotted using Microsoft Excel. Each experiment was repeated a minimum of three times, with similar results.

**Y2H analyses.** The Matchmaker Gold Yeast Two-Hybrid System (Takara Bio USA) was used to detect protein–protein interactions between Pik-HMAs and AVR-Pik effectors. The DNA encoding the Pik-HMAs in pGBKT7 was co-transformed with either the individual AVR-Pik variants or the mutants in pGADT7 into chemically competent *Saccharomyces cerevisiae* Y2HGold cells (Takara Bio USA). Single colonies grown on selection plates were inoculated in 5 ml SD-Leu-Trp plate and grown overnight at  $30^\circ\text{C}$ . Saturated culture was then used to make serial dilutions of optical density at 600 nm ( $\text{OD}_{600}$ ) 1,  $1^{-1}$ ,  $1^{-2}$  and  $1^{-3}$ , respectively. Of each dilution, 5  $\mu\text{l}$  was then spotted on a SD-Leu-Trp plate as a growth control and also on a SD-Leu-Trp-Ade-His plate containing X- $\alpha$ -gal and auroebasidine, as detailed in the user manual. Plates were imaged after incubation for 60–72 h at  $30^\circ\text{C}$ . Each experiment was repeated a minimum of three times, with similar results.

To confirm protein expression in yeast, the total protein was extracted from transformed colonies by boiling the cells for 10 min in LDS Runblue sample buffer. Samples were centrifuged, and the supernatant was subjected to SDS-PAGE before western blotting. The resulting membranes were probed with anti-GAL4 DNA-BD (Sigma) for the HMA domains in pGBKT7 and anti-GAL4 activation domain (Sigma) antibodies for the AVR-Pik effectors in pGADT7.

***N. benthamiana* cell death assays.** Transient gene expression in planta was performed by delivering T-DNA constructs with *Agrobacterium tumefaciens* GV3101 strain into 4-week-old *N. benthamiana* plants grown at  $22$ – $25^\circ\text{C}$  with high light intensity. Pik-1, Pik-2, AVR-Pik and P19 were mixed at  $\text{OD}_{600}$  0.4, 0.4, 0.6 and 0.1, respectively. Detached leaves were imaged at 5 dpi from the abaxial side. Images are representative of three independent experiments, with internal repeats. The cell death index used for scoring is as presented previously<sup>33</sup> (also included in Supplementary Fig. 1d). The scoring for all replicas is presented as box plots, which were generated using R v3.4.3 (<https://www.r-project.org/>) and the graphic package ggplot2 (ref. <sup>47</sup>). The centre line represents the median, the box limits are the upper and lower quartiles, the whiskers are the  $1.5\times$  interquartile range and all of the data points are represented as dots.

The presence of each protein, as expressed in representative assays, was determined by SDS-PAGE or western blot. For this, the leaf tissue was frozen and ground to fine powder in liquid nitrogen using a pestle and mortar. The leaf powder was mixed with two-times weight/volume ice-cold extraction buffer (10% glycerol, 25 mM Tris (pH 7.5), 1 mM EDTA, 150 mM NaCl, 2% w/v PVP, 10 mM dithiothreitol, 1× protease inhibitor cocktail (Sigma) and 0.1% Tween 20 (Sigma)), centrifuged at 4,200 g at  $4^\circ\text{C}$  for 20–30 min and the supernatant was filtered (0.45  $\mu\text{m}$ ).

**Crystallization, data collection and structure solution.** For crystallization, Pik-HMA–AVR-Pik complexes were concentrated in buffer B (see Supplementary Methods). Sitting drop, vapour diffusion crystallization trials were set up in 96-well plates, using an Oryx nano robot (Douglas Instruments). Plates were incubated at  $20^\circ\text{C}$ , and crystals typically appeared after 24–48 h. For data collection, all crystals were harvested from the Morphew HT-96 screen (Molecular Dimensions) and snap frozen in liquid nitrogen. The details of each crystallization condition are given in the Supplementary Methods.

X-ray data sets were collected at the Diamond Light Source. The data were processed using the xia2 pipeline<sup>48</sup> and AIMLESS<sup>49</sup>, as implemented in CCP4<sup>50</sup>. The structures were solved by molecular replacement using PHASER<sup>51</sup> and the Pikp-HMA–AVR-PikD structure<sup>33</sup>. The final structures were obtained through iterative

cycles of manual rebuilding and refinement using COOT<sup>52</sup> and REFMAC5<sup>53</sup>, as implemented in CCP4<sup>50</sup>. Structures were validated using the tools provided in COOT and MOLPROBITY<sup>54</sup>. More details on data collection and refinement are given in the Supplementary Methods.

**Protein interface analyses.** Protein interface analyses were performed using QtPISA<sup>36</sup>. For each complex, one Pik-HMA–AVR-Pik effector assembly was used as a representative example. QtPISA interaction radars<sup>36</sup> were produced using the reference parameter ‘Total Binding Energy’. The area of the polygon indicates the likelihood of the interface to constitute part of a biological assembly (the greater the area, the more likely the interface constitutes part of a biological assembly). The scales along the beams compare the key interface properties to statistical distributions derived from the PDB. In general, if the radar area is contained within the 50% probability circle, then the interface is considered superficial and its biological relevance is questionable. In cases where the radar area is expanded outside the 50% probability circle, the interface is considered more likely to be significant and biologically relevant<sup>36</sup>.

**Reporting Summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

**Data availability.** The coordinates and structure factors have been deposited in the PDB with accession codes **6FU9** (Pikm-HMA–AVR-PikD), **6FUB** (Pikm-HMA–AVR-PikE), **6FUD** (Pikm-HMA–AVR-PikA), **6G10** (Pikp-HMA–AVR-PikD) and **6G11** (Pikp-HMA–AVR-PikE).

Received: 26 February 2018; Accepted: 5 June 2018;

Published online: 09 July 2018

## References

1. Jones, J. D., Vance, R. E. & Dangl, J. L. Intracellular innate immune surveillance devices in plants and animals. *Science* **354**, aaf6395 (2016).
2. Ronald, P. C. & Beutler, B. Plant and animal sensors of conserved microbial signatures. *Science* **330**, 1061–1064 (2010).
3. Dodds, P. N. & Rathjen, J. P. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* **11**, 539–548 (2010).
4. Win, J. et al. Effector biology of plant-associated organisms: concepts and perspectives. *Cold Spring Harb. Symp. Quant. Biol.* **77**, 235–247 (2012).
5. Bialas, A. et al. Lessons in effector and NLR biology of plant–microbe systems. *Mol. Plant Microbe Interact.* **31**, 34–45 (2017).
6. Ellis, J. G., Lawrence, G. J., Luck, J. E. & Dodds, P. N. Identification of regions in alleles of the flax rust resistance gene *L* that determine differences in gene-for-gene specificity. *Plant Cell* **11**, 495–506 (1999).
7. Allen, R. L. et al. Host–parasite coevolutionary conflict between *Arabidopsis* and downy mildew. *Science* **306**, 1957–1960 (2004).
8. Bhullar, N. K., Zhang, Z., Wicker, T. & Keller, B. Wheat gene bank accessions as a source of new alleles of the powdery mildew resistance gene *Pm3*: a large scale allele mining project. *BMC Plant Biol.* **10**, 88 (2010).
9. Seeholzer, S. et al. Diversity at the *Mla* powdery mildew resistance locus from cultivated barley reveals sites of positive selection. *Mol. Plant Microbe Interact.* **23**, 497–509 (2010).
10. Srichumpa, P., Brunner, S., Keller, B. & Yahiaoui, N. Allelic series of four powdery mildew resistance genes at the *Pm3* locus in hexaploid bread wheat. *Plant Physiol.* **139**, 885–895 (2005).
11. Lu, X. et al. Allelic barley MLA immune receptors recognize sequence-unrelated avirulence effectors of the powdery mildew pathogen. *Proc. Natl Acad. Sci. USA* **113**, E6486–E6495 (2016).
12. Dodds, P. N. et al. Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc. Natl Acad. Sci. USA* **103**, 8888–8893 (2006).
13. Krasileva, K. V., Dahlbeck, D. & Staskawicz, B. J. Activation of an *Arabidopsis* resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. *Plant Cell* **22**, 2444–2458 (2010).
14. Steinbrenner, A. D., Goritschnig, S. & Staskawicz, B. J. Recognition and activation domains contribute to allele-specific responses of an *Arabidopsis* NLR receptor to an oomycete effector protein. *PLoS Pathog.* **11**, e1004665 (2015).
15. Wang, C. I. et al. Crystal structures of flax rust avirulence proteins AvrL567-A and -D reveal details of the structural basis for flax disease resistance specificity. *Plant Cell* **19**, 2898–2912 (2007).
16. Huang, J., Si, W., Deng, Q., Li, P. & Yang, S. Rapid evolution of avirulence genes in rice blast fungus *Magnaporthe oryzae*. *BMC Genet.* **15**, 45 (2014).
17. Raffaele, S. et al. Genome evolution following host jumps in the Irish potato famine pathogen lineage. *Science* **330**, 1540–1543 (2010).
18. Yoshida, K. et al. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *Plant Cell* **21**, 1573–1591 (2009).

19. Dangl, J. L., Horvath, D. M. & Staskawicz, B. J. Pivoting the plant immune system from dissection to deployment. *Science* **341**, 746–751 (2013).
20. Rodriguez-Moreno, L., Song, Y. & Thomma, B. P. Transfer and engineering of immune receptors to improve recognition capacities in crops. *Curr. Opin. Plant Biol.* **38**, 42–49 (2017).
21. Eitas, T. K. & Dangl, J. L. NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Curr. Opin. Plant Biol.* **13**, 472–477 (2010).
22. Wu, C. H., Belhaj, K., Bozkurt, T. O., Birk, M. S. & Kamoun, S. Helper NLR proteins NRC2a/b and NRC3 but not NRC1 are required for Pto-mediated cell death and resistance in *Nicotiana benthamiana*. *New Phytol.* **209**, 1344–1352 (2016).
23. Narusaka, M. et al. *RRS1* and *RPS4* provide a dual *Resistance*-gene system against fungal and bacterial pathogens. *Plant J.* **60**, 218–226 (2009).
24. Sinapidou, E. et al. Two TIR:NBLRR genes are required to specify resistance to *Peronospora parasitica* isolate Cala2 in *Arabidopsis*. *Plant J.* **38**, 898–909 (2004).
25. Ashikawa, I. et al. Two adjacent nucleotide-binding site-leucine-rich repeat class genes are required to confer *Pikm*-specific rice blast resistance. *Genetics* **180**, 2267–2276 (2008).
26. Lee, S. K. et al. Rice *Pi5*-mediated resistance to *Magnaporthe oryzae* requires the presence of two coiled-coil-nucleotide-binding-leucine-rich repeat genes. *Genetics* **181**, 1627–1638 (2009).
27. Wu, C. H. et al. NLR network mediates immunity to diverse plant pathogens. *Proc. Natl Acad. Sci. USA* **114**, 8113–8118 (2017).
28. Cesari, S., Bernoux, M., Moncquet, P., Kroj, T. & Dodds, P. N. A novel conserved mechanism for plant NLR protein pairs: the “integrated decoy” hypothesis. *Front. Plant Sci.* **5**, 606 (2014).
29. Kroj, T., Chanclud, E., Michel-Romiti, C., Grand, X. & Morel, J. B. Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytol.* **210**, 618–626 (2016).
30. Sarris, P. F., Cevik, V., Dagdas, G., Jones, J. D. & Krasileva, K. V. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biol.* **14**, 8 (2016).
31. Wu, C. H., Krasileva, K. V., Banfield, M. J., Terauchi, R. & Kamoun, S. The “sensor domains” of plant NLR proteins: more than decoys? *Front. Plant Sci.* **6**, 134 (2015).
32. Okuyama, Y. et al. A multifaceted genomics approach allows the isolation of the rice *Pia*-blast resistance gene consisting of two adjacent NBS-LRR protein genes. *Plant J.* **66**, 467–479 (2011).
33. Maqbool, A. et al. Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife* **4**, e08709 (2015).
34. Kanzaki, H. et al. Arms race co-evolution of *Magnaporthe oryzae* AVR-Pik and rice *Pik* genes driven by their physical interactions. *Plant J.* **72**, 894–907 (2012).
35. Costanzo, S. & Jia, Y. L. Sequence variation at the rice blast resistance gene *Pi-km* locus: implications for the development of allele specific markers. *Plant Sci.* **178**, 523–530 (2010).
36. Krissinel, E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res.* **43**, W314–W319 (2015).
37. de Guillen, K. et al. Structure analysis uncovers a highly diverse but structurally conserved effector family in phytopathogenic fungi. *PLoS Pathog.* **11**, e1005228 (2015).
38. Bent, A. F. et al. *RPS2* of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science* **265**, 1856–1860 (1994).
39. Mindrinos, M., Katagiri, F., Yu, G. L. & Ausubel, F. M. The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* **78**, 1089–1099 (1994).
40. Whitham, S. et al. The product of the tobacco mosaic virus resistance gene *N*: similarity to toll and the interleukin-1 receptor. *Cell* **78**, 1101–1115 (1994).
41. Le Roux, C. et al. A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell* **161**, 1074–1088 (2015).
42. Sarris, P. F. et al. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell* **161**, 1089–1100 (2015).
43. Zhang, Z. M. et al. Mechanism of host substrate acetylation by a YopJ family effector. *Nat. Plants* **3**, 17115 (2017).
44. Ortiz, D. et al. Recognition of the *Magnaporthe oryzae* effector AVR-Pia by the decoy domain of the rice NLR immune receptor RGA5. *Plant Cell* **29**, 156–168 (2017).
45. Lobstein, J. et al. SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* **11**, 56 (2012).
46. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
47. Wickham, H. *ggplot: Elegant Graphics for Data Analysis* (Springer, New York, NY, 2009).
48. Winter, G. xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
49. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
50. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
51. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
52. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
53. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
54. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

### Acknowledgements

This work was supported by the BBSRC (grants BB/J00453, BB/P012574 and BB/M02198X), the ERC (proposal 743165), the John Innes Foundation, the Gatsby Charitable Foundation and JSPS KAKENHI 15H05779. We thank the Diamond Light Source (beamlines I03 and I04 under proposals MX9475 and MX13467) for access to X-ray data collection facilities. We also thank D. Lawson and C. Stevenson (JIC X-ray Crystallography/Biophysical Analysis Platform) for help with protein structure determination and SPR.

### Author contributions

J.C.D.I.C. and M.F. performed all of the experiments. J.C.D.I.C., M.F. and M.J.B. designed the experiments and analysed the data. A.M. and H.S. assisted with construct design and the initial protein production. R.T. and S.K. analysed the data. J.C.D.I.C., M.F. and M.J.B. wrote the manuscript with input from all authors.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41477-018-0194-x>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to M.J.B.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



# Protein engineering expands the effector recognition profile of a rice NLR immune receptor

Juan Carlos De la Concepcion<sup>1†</sup>, Marina Franceschetti<sup>1†</sup>, Dan MacLean<sup>2</sup>,  
Ryohei Terauchi<sup>3,4</sup>, Sophien Kamoun<sup>2</sup>, Mark J Banfield<sup>1\*</sup>

<sup>1</sup>Department of Biological Chemistry, John Innes Centre, Norwich, United Kingdom;

<sup>2</sup>The Sainsbury Laboratory, University of East Anglia, Norwich, United Kingdom;

<sup>3</sup>Division of Genomics and Breeding, Iwate Biotechnology Research Center, Iwate, Japan; <sup>4</sup>Laboratory of Crop Evolution, Graduate School of Agriculture, Kyoto University, Kyoto, Japan

**Abstract** Plant nucleotide binding, leucine-rich repeat (NLR) receptors detect pathogen effectors and initiate an immune response. Since their discovery, NLRs have been the focus of protein engineering to improve disease resistance. However, this approach has proven challenging, in part due to their narrow response specificity. Previously, we revealed the structural basis of pathogen recognition by the integrated heavy metal associated (HMA) domain of the rice NLR Pikp (Maqbool et al., 2015). Here, we used structure-guided engineering to expand the response profile of Pikp to variants of the rice blast pathogen effector AVR-Pik. A mutation located within an effector-binding interface of the integrated Pikp–HMA domain increased the binding affinity for AVR-Pik variants *in vitro* and *in vivo*. This translates to an expanded cell-death response to AVR-Pik variants previously unrecognized by Pikp in planta. The structures of the engineered Pikp–HMA in complex with AVR-Pik variants revealed the mechanism of expanded recognition. These results provide a proof-of-concept that protein engineering can improve the utility of plant NLR receptors where direct interaction between effectors and NLRs is established, particularly where this interaction occurs via integrated domains.

DOI: <https://doi.org/10.7554/eLife.47713.001>

\*For correspondence:  
Mark.banfield@jic.ac.uk

†These authors contributed equally to this work

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 15

**Received:** 17 April 2019

**Accepted:** 17 September 2019

**Published:** 19 September 2019

**Reviewing editor:** Thorsten Nürnberger, University of Tübingen, Germany

© Copyright De la Concepcion et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

Protein engineering offers opportunities to develop new or improved molecular recognition capabilities that have applications in basic research, health and agricultural settings. Protein resurfacing, where the properties of solvent-exposed regions are changed (often by mutation), has been used extensively in diverse areas from antibody engineering for clinical use to the production of more stable, soluble proteins for biotechnology applications (Chapman and McNaughton, 2016).

Intracellular nucleotide binding, leucine-rich repeat (NLR) receptors are key components of plant innate immunity pathways. They recognize the presence or activity of virulence-associated, host-translocated pathogen effector proteins and initiate an immune response (Kourelis and van der Hoorn, 2018; Cesari, 2018). Because they confer resistance to disease, plant NLRs are widely used in crop breeding programs (Dangl et al., 2013). However, the recognition spectra of plant NLRs tend to be very specific, and pathogens may delete detected effectors from their genome or evolve novel effector variants that are not detected by the NLRs to re-establish disease (Yoshida et al., 2016).

The potential of engineering NLRs to overcome these limitations, or to detect new effector activities, is emerging (Rodríguez-Moreno et al., 2017). Gain-of-function random mutagenesis has

achieved some success in expanding the activation sensitivity or effector recognition profiles of NLRs (Segretin et al., 2014; Giannakopoulou et al., 2015; Harris et al., 2013). An alternative strategy, in which NLRs perceive protease effectors through their activity on engineered host proteins, can lead to expanded recognition profiles (Carter et al., 2018; Helm et al., 2019; Kim et al., 2016). When detailed knowledge of the direct binding interfaces between an effector and an NLR are known, there is potential for protein resurfacing to modify interactions and thereby impact immune signaling.

Plant NLRs are modular proteins that are defined by their nucleotide-binding (NB-ARC) and leucine-rich repeat (LRR) domains, but that also have either an N-terminal coiled-coil (CC) or a Toll/Interleukin-1/Resistance-protein (TIR) signaling domain (Takken and Goverse, 2012). Many NLRs also contain non-canonical integrated domains (Kroj et al., 2016; Sarris et al., 2016; Bailey et al., 2018). Integrated domains are thought to be derived from ancestral virulence-associated effector targets that directly bind pathogen effectors (or host proteins [Fujisaki et al., 2017]) or are modified by them, thereby initiating an immune response (Cesari et al., 2014; Le Roux et al., 2015; Maqbool et al., 2015; Sarris et al., 2015; De la Concepcion et al., 2018). These domains present an exciting target for protein engineering approaches to improve NLR activities. NLRs containing integrated domains (often called the ‘sensor’) typically function in pairs, requiring a second genetically linked NLR (the ‘helper’) for immune signaling (Bialas et al., 2018; Eitas and Dangl, 2010).

Two rice NLR pairs, Pik and Pia, contain an integrated Heavy Metal Associated (HMA) domain in their sensor NLR that directly binds effectors from the rice blast pathogen *Magnaporthe oryzae* (also known as *Pycularia oryzae*) (Maqbool et al., 2015; Cesari et al., 2013; Ortiz et al., 2017; Guo et al., 2018). The integrated HMA domain in the sensor NLR Pik-1 directly binds the effector AVR-Pik. Co-evolutionary dynamics has driven the emergence of polymorphic Pik-1 HMA domains and AVR-Pik effectors in natural populations, and these display differential disease-resistance phenotypes (Costanzo and Jia, 2010; Kanzaki et al., 2012; Yoshida et al., 2009). The Pikp NLR allele only responds to the effector variant AVR-PikD, but the Pikm allele responds to AVR-PikD, AVR-PikE, and AVR-PikA. These phenotypes can be recapitulated in the model plant *Nicotiana benthamiana* using a cell-death assay, and are underpinned by differences in effector–receptor binding interfaces that lead to different affinities in vitro (Maqbool et al., 2015; De la Concepcion et al., 2018).

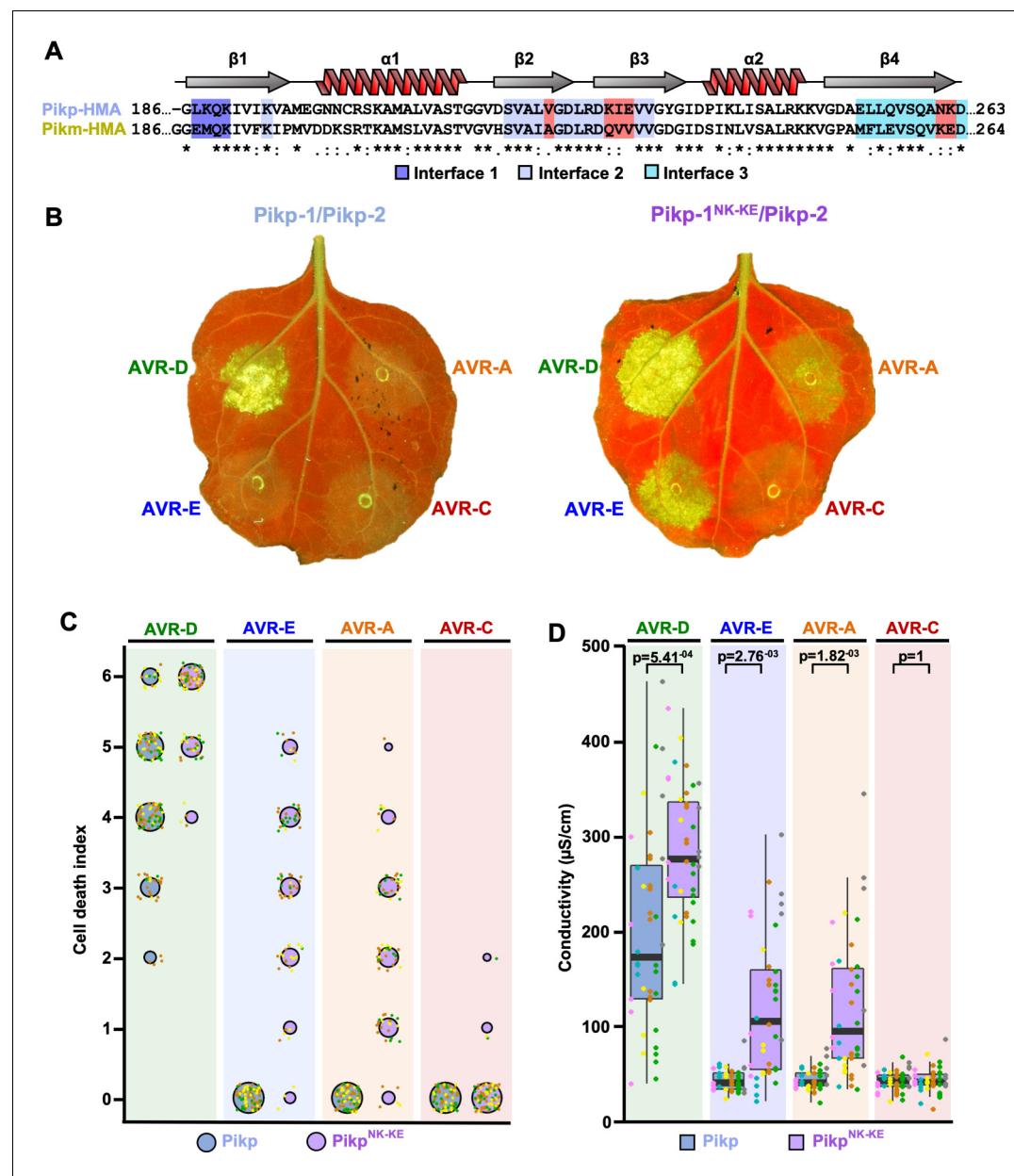
We hypothesized that by combining naturally occurring favorable interactions observed across different interfaces, as defined in different Pik-HMA/AVR-Pik structures (Maqbool et al., 2015; De la Concepcion et al., 2018), we could generate a Pik NLR with improved recognition profiles. Here, we graft an interface from Pikm onto Pikp by mutating two residues in Pikp (Asn261Lys, Lys262Glu), forming Pikp<sup>NK-KE</sup>. This single-site mutation strengthens the cell-death response in *N. benthamiana* to AVR-PikD, and gains a Pikm-like response to AVR-PikE and AVR-PikA. We show that this gain-of-function phenotype correlates with increased binding affinity of the effectors by the Pikp<sup>NK-KE</sup>-HMA domain in vitro and in vivo, and demonstrate that this mutation results in a Pikm-like structure for Pikp<sup>NK-KE</sup> when in complex with AVR-Pik effectors. Finally, we confirm that the newly engineered interface is responsible for the expanded response of Pikp<sup>NK-KE</sup> by mutation of the effectors.

This study serves as a proof-of-concept for the use of protein resurfacing by targeted mutation to develop plant NLR immune receptors with new capabilities. In the future, such approaches have the potential to improve disease resistance in crops.

## Results

### Structure-informed engineering expands Pikp-mediated effector recognition in *N. benthamiana*

By comparing protein interfaces in the structures of Pikp-HMA and Pikm-HMA bound to different AVR-Pik effectors (Maqbool et al., 2015; De la Concepcion et al., 2018), we hypothesized that we could engineer expanded effector recognition capabilities by point mutation of Pikp. We constructed a series of mutations in the previously identified interface 2 and interface 3 regions of Pikp-HMA–AVR-Pik structures (De la Concepcion et al., 2018), swapping residues found in Pikm into Pikp (Figure 1A, Figure 1—figure supplement 1). We then screened these mutations for expanded



**Figure 1.** Structure-informed engineering expands Pikp-mediated effector recognition in *N. benthamiana*. (A) Sequence alignment of the Pikp-1 and Pikm-1 HMA domains. Secondary structure features of the HMA fold are shown above, and the residues that are located at binding interfaces are as colored. Key residues from interface 2 and interface 3 involved in this study are highlighted in red. (B) Representative leaf images showing Pikp- (left) or Pikp-1<sup>NK-KE</sup> (right)-mediated cell death in response to AVR-Pik variants as autofluorescence under UV light. (C) Autofluorescence intensity is scored as previously described (Maqbool *et al.*, 2015; De la Concepcion *et al.*, 2018). Cell death assay scores are represented as dot plots for Pikp and Pikp<sup>NK-KE</sup> (blue and purple, respectively). For each sample, all of the data points are represented as dots with a distinct color for each of the three biological replicates; these dots are plotted around the cell death score for visualization purposes. The size of the centre dot at each cell death value is directly proportional to the number of replicates in the sample with that score. The total number of repeats was 80. Data for Pikp have been previously shown (De la Concepcion *et al.*, 2018), but was acquired at the same time as those for Pikp<sup>NK-KE</sup>. The estimation methods used to visualize differences in the data sets are shown in **Figure 1—figure supplement 3**. (D) Conductivity measurements showing ion leakage as a quantitative measure of cell death. The centre line represents the median, the box limits are the upper and lower quartiles, the whiskers extend to the largest value within Q1 – 1.5x the interquartile range (IQR) and the smallest value within Q3 + 1.5x IQR. All the data points are shown as dots with distinct colors for each biological replicate. For each experiment, six biological replicates with 5 or 10 internal repeats were performed (total data points = 40). 'p' is the p-value obtained from statistical analysis and Tukey's HSD (honestly significant difference) test.

DOI: <https://doi.org/10.7554/eLife.47713.002>

The following source data and figure supplements are available for figure 1:

Figure 1 continued on next page

Figure 1 continued

**Source data 1.** Cell death scoring data used in the preparation of **Figure 1C**.

DOI: <https://doi.org/10.7554/eLife.47713.009>

**Source data 2.** Conductivity measurements used in the preparation of **Figure 1D** and **Figure 1—figure supplement 2B**.

DOI: <https://doi.org/10.7554/eLife.47713.010>

**Figure supplement 1.** Mutations at interface 2 of the *Pikp*-1 HMA domain compromise the response to AVR-Pik effectors.

DOI: <https://doi.org/10.7554/eLife.47713.003>

**Figure supplement 1—source data 1.** Cell death scoring data used in the preparation of **Figure 1—figure supplement 1A**.

DOI: <https://doi.org/10.7554/eLife.47713.004>

**Figure supplement 2.** Response of *Pikp*<sup>NK-KE</sup> to AVR-Pik effectors compared to that of *Pikm*.

DOI: <https://doi.org/10.7554/eLife.47713.005>

**Figure supplement 2—source data 1.** Cell-death scoring data used in the preparation of **Figure 1—figure supplement 2A**.

DOI: <https://doi.org/10.7554/eLife.47713.006>

**Figure supplement 3.** Estimation graphics for cell death, *Pikp* vs *Pikp*<sup>NK-KE</sup>.

DOI: <https://doi.org/10.7554/eLife.47713.007>

**Figure supplement 4.** Estimation graphics for cell death, *Pikm* vs *Pikp*<sup>NK-KE</sup>.

DOI: <https://doi.org/10.7554/eLife.47713.008>

effector recognition by monitoring cell death in a well-established *N. benthamiana* assay (Maqbool et al., 2015; De la Concepcion et al., 2018). We found that one double mutation in two adjacent amino-acid residues contained within interface 3, Asn261Lys and Lys262Glu (henceforth *Pikp*<sup>NK-KE</sup>), induced cell death in response to AVR-PikE and AVR-PikA (Figure 1, Figure 1—figure supplement 1A,B). Mutations that cause loss of response to AVR-PikD are most likely to be compromised in effector binding. Western blots confirmed that all of the proteins were expressed in plants (Figure 1—figure supplement 1C).

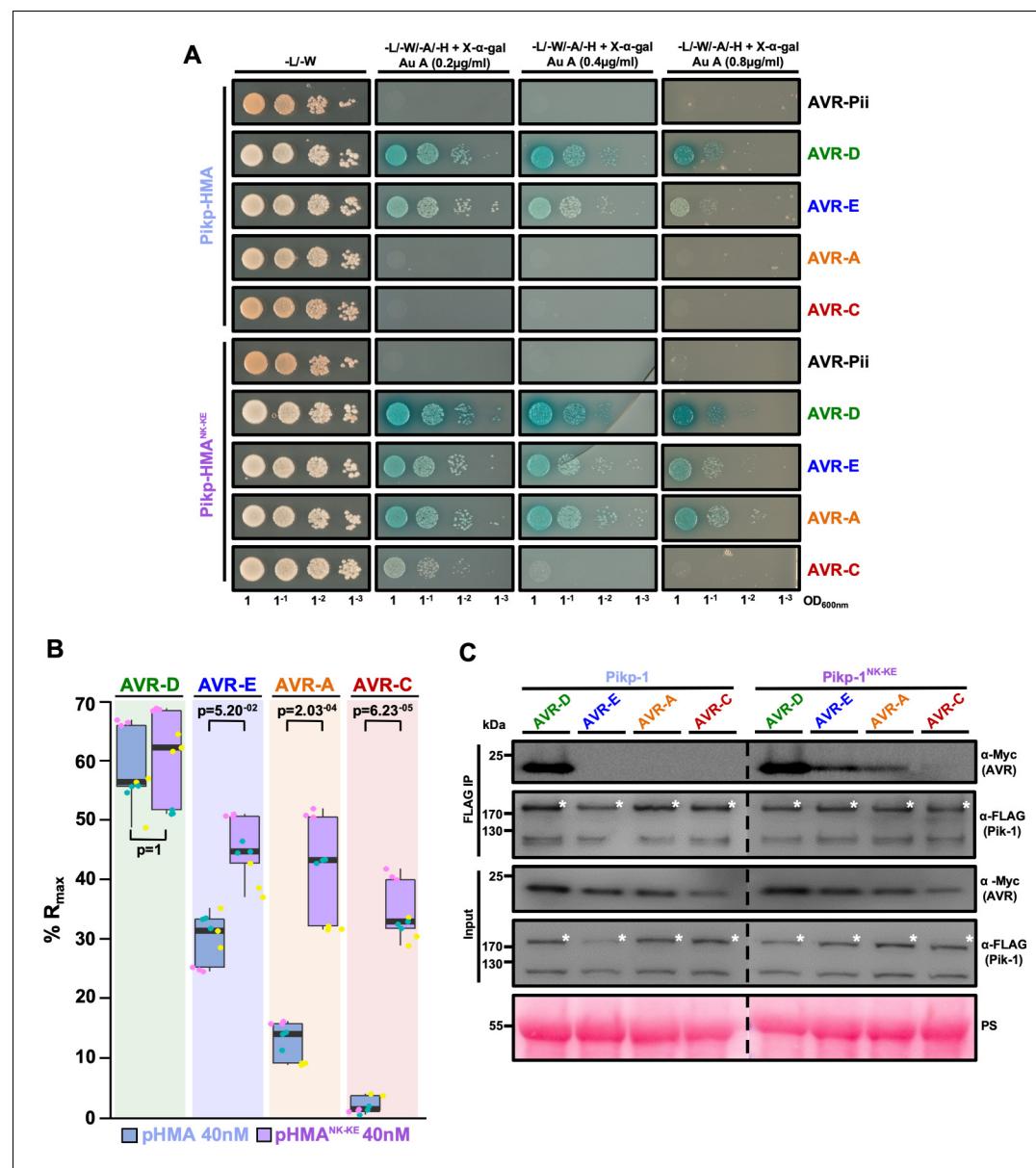
We subsequently focused on this double mutant, and independently repeated the cell-death assay to ensure its robustness (Figure 1B,C, Figure 1—figure supplement 3). When looking at the intensity of cell death mediated by *Pikp*<sup>NK-KE</sup>, as for the *Pikm* allele (De la Concepcion et al., 2018), we observed an effector hierarchy of AVR-PikD > AVR-PikE > AVR-PikA. *Pikp*<sup>NK-KE</sup> shows a comparable, but elevated (not statistically significant in the case of AVR-PikE), response to effector variants when compared to that of *Pikm* (Figure 1—figure supplements 2 and 4). The cell-death response to AVR-PikD mediated by *Pikp*<sup>NK-KE</sup> is similar but elevated when compared to that mediated by *Pikp* (Figure 1B,C). *Pikp*<sup>NK-KE</sup> does not show a response to the stealthy AVR-PikC variant. To obtain a quantitative measure of cell death, we performed ion-leakage assays (Figure 1D, Figure 1—figure supplement 2, Supplementary file 1). The results of these assays correlate well with the cell-death index scores based on autofluorescence described above.

We conclude that the single Asn261Lys/Lys262Glu (*Pikp*<sup>NK-KE</sup>) mutation at interface 3 in the *Pikp* NLR expands this protein's recognition profile to include the effector variants AVR-PikE and AVR-PikA, similar to that observed for *Pikm*.

### The engineered *Pikp*<sup>NK-KE</sup>-HMA mutant shows increased binding to effector variants in vivo and in vitro

We used yeast-2-hybrid (Y2H) and surface plasmon resonance (SPR) to determine whether the expanded *Pikp*<sup>NK-KE</sup> cell-death response in *N. benthamiana* correlates with increased binding affinity of the *Pikp*<sup>NK-KE</sup>-HMA domain for AVR-Pik effectors.

As AVR-PikE and AVR-PikA showed some interaction with *Pikp*-HMA using these approaches (Maqbool et al., 2015; De la Concepcion et al., 2018), we tested interactions with *Pikp*<sup>NK-KE</sup>-HMA side-by-side with interactions with wild-type *Pikp*-HMA. Using Y2H, we observed increases in growth and in blue coloration (both indicative of protein–protein interaction) for *Pikp*-HMA<sup>NK-KE</sup> with effectors AVR-PikE and, particularly, AVR-PikA when compared with *Pikp*-HMA (Figure 2A). This was accentuated with more stringent conditions (imposed by increasing concentration of Aureobasidin A). Unexpectedly, we also observed limited yeast growth for *Pikp*-HMA<sup>NK-KE</sup> with AVR-PikC at the lower stringency (Figure 2A). The unrelated *M. oryzae* effector AVR-Pii was used as a negative control. Expression of all proteins was confirmed in yeast (Figure 2—figure supplement 1).



**Figure 2.** *Pikp*<sup>NK-KE</sup> shows increased binding to effector variants *in vivo* and *in vitro* when compared to wild type *Pikp*. (A) Yeast-two-hybrid assay of *Pikp*-HMA and *Pikp*-HMA<sup>NK-KE</sup> with AVR-Pik alleles. Control plates for yeast growth are on the left, with quadruple dropout media supplemented with X- $\alpha$ -gal and increasing concentrations of Aureobasidin A on the right for each combination of HMA/AVR-Pik. The unrelated *M. oryzae* effector AVR-Pii was used as a negative control. Growth and the development of blue coloration in the selection plate are both indicative of protein–protein interaction. HMA domains were fused to the GAL4 DNA binding domain, and AVR-Pik alleles to the GAL4 activator domain. Each experiment was repeated a minimum of three times, with similar results. (B) Box plots showing %R<sub>max</sub>, as measured by surface plasmon resonance, for *Pikp*-HMA and *Pikp*-HMA<sup>NK-KE</sup> with the AVR-Pik effectors alleles at an HMA concentration of 40 nM. *Pikp*-HMA and *Pikp*-HMA<sup>NK-KE</sup> are represented by blue and purple boxes, respectively. The centre line within each box represents the median, the box limits are the upper and lower quartiles, the whiskers extend to the largest value within Q1 + 1.5  $\times$  the interquartile range (IQR) and the smallest value within Q3 + 1.5  $\times$  IQR. All of the data points are represented as dots with distinct colors for each biological replicate. For each experiment, three biological replicates with three internal repeats were performed. 'p' is the p-value obtained from statistical analysis and Tukey's HSD. For results of experiments with 4 nM and 100 nM HMA protein concentration, see **Figure 2—figure supplement 2**. (C) Co-immunoprecipitation of full length *Pikp*-1 and *Pikp*-1<sup>NK-KE</sup> with AVR-Pik variants. N-terminally 4xMyc tagged AVR-Pik effectors were transiently co-expressed with *Pikp*-1:6xHis3xFLAG (left) or with *Pikp*-1<sup>NK-KE</sup>:6xHis3xFLAG (right) in *N. benthamiana*. Immunoprecipitates (IPs) obtained with anti-FLAG antiserum, and total protein extracts, were probed with appropriate antisera. The dashed line indicates a crop site on the same blot used to compose the figure. Each experiment was repeated at least three times, with similar results. The asterisks mark the *Pik*-1 band. PS = Ponceau Stain.

DOI: <https://doi.org/10.7554/eLife.47713.011>

Figure 2 continued on next page

## Figure 2 continued

The following source data and figure supplements are available for figure 2:

**Source data 1.** Surface plasmon resonance measurements used in the preparation of **Figure 2B**.

DOI: <https://doi.org/10.7554/eLife.47713.018>

**Figure supplement 1.** Western blot confirming the accumulation of proteins in yeast.

DOI: <https://doi.org/10.7554/eLife.47713.012>

**Figure supplement 2.** Binding of the  $\text{Pikp-HMA}^{\text{NK-KE}}$  domain to the AVR-Pik effectors is consistently greater than that of  $\text{Pikp-HMA}$ .

DOI: <https://doi.org/10.7554/eLife.47713.013>

**Figure supplement 2—source data 1.** Surface plasmon resonance measurements used in the preparation of **Figure 2—figure supplement 2**, left panel.

DOI: <https://doi.org/10.7554/eLife.47713.014>

**Figure supplement 2—source data 2.** Surface plasmon resonance measurements used in the preparation of **Figure 2—figure supplement 2**, right panel.

DOI: <https://doi.org/10.7554/eLife.47713.015>

**Figure supplement 3.** Binding of the  $\text{Pikp-HMA}^{\text{NK-KE}}$  domain to the AVR-Pik effectors is consistently greater than that of the  $\text{Pikm-HMA}$  domain.

DOI: <https://doi.org/10.7554/eLife.47713.016>

**Figure supplement 3—source data 1.** Surface plasmon resonance measurements used in the preparation of **Figure 2—figure supplement 3**.

DOI: <https://doi.org/10.7554/eLife.47713.017>

Next, we produced the  $\text{Pikp-HMA}^{\text{NK-KE}}$  domain protein via overexpression in *E. coli* and purified it to homogeneity using well-established procedures for these domains (see Materials and methods; *Maqbool et al., 2015; De la Concepcion et al., 2018*). Using SPR, we measured the binding affinity of the  $\text{Pikp-HMA}^{\text{NK-KE}}$  domain (alongside both wild-type  $\text{Pikp-HMA}$  and  $\text{Pikm-HMA}$ ) to AVR-Pik effectors (**Figure 2B**, **Figure 2—figure supplements 2 and 3**). Response units (RU) were measured following the injection of  $\text{Pik-HMAs}$  at three different concentrations, after capturing AVR-Pik effectors on a Biacore NTA chip. RUs were then normalized to the theoretical maximum response ( $R_{\max}$ ), assuming a 2:1 interaction model for  $\text{Pikp-HMA}$  and  $\text{Pikp-HMA}^{\text{NK-KE}}$ , and a 1:1 interaction model for  $\text{Pikm-HMA}$ , as previously described (*De la Concepcion et al., 2018*). These data showed an increased binding of  $\text{Pikp-HMA}^{\text{NK-KE}}$  to AVR-PikE and AVR-PikA compared to that of wild-type  $\text{Pikp}$  (**Figure 2B**, **Figure 2—figure supplement 2**, *Supplementary file 2*). The binding of  $\text{Pikp-HMA}^{\text{NK-KE}}$  to the AVR-Pik effectors was also greater than that of  $\text{Pikm-HMA}$  (**Figure 2—figure supplement 3**, *Supplementary file 2*), correlating with the results of cell-death assays (**Figure 1—figure supplement 2**). In both cases, statistical analysis has been carried out for the 40 nM analyte ( $\text{Pik-HMA}$ ) data as a representative concentration. Although neither  $\text{Pikp-HMA}$  nor  $\text{Pikm-HMA}$  domains show binding to AVR-PikC by SPR, we observe a gain-of-binding of this effector variant with  $\text{Pikp-HMA}^{\text{NK-KE}}$  (**Figure 2B**, **Figure 2—figure supplements 2 and 3**, *Supplementary file 2*).

These results show that the  $\text{Pikp-HMA}^{\text{NK-KE}}$  mutant has a higher binding affinity for effectors AVR-PikE and AVR-PikA than does the wild-type protein. This suggests that the increased binding affinity to the HMA domain correlates with the expanded cell-death response in planta (**Figure 1B**).

### The engineered $\text{Pikp}^{\text{NK-KE}}$ mutant expands the association of full-length $\text{Pik-1}$ with effector variants in planta

In addition to interaction with the isolated HMA domain, we tested whether the Asn261Lys/Lys262-Glu mutant could expand effector variant binding in the context of the full-length NLR. After generating the mutant in the full-length protein, we co-expressed either  $\text{Pikp-1}$  or  $\text{Pikp-1}^{\text{NK-KE}}$  with the AVR-Pik effector variants in *N. benthamiana*, followed by immunoprecipitation and western blotting to determine effector association.

AVR-PikD shows a robust association with  $\text{Pikp-1}$ . However, although we observe limited binding to the isolated  $\text{Pikp-HMA}$  domain in Y2H and SPR screens, we did not detect association of AVR-PikE and AVR-PikA with the full-length  $\text{Pikp-1}$  in planta (**Figure 2C**). By contrast, we observe clear association of AVR-PikE and AVR-PikA with the  $\text{Pikp-1}^{\text{NK-KE}}$  mutant, albeit with reduced intensity compared to AVR-PikD, correlating with the hierarchical cell-death response observed in planta (**Figure 1C**).

We also detect a very low level of association between full-length  $\text{Pikp-1}^{\text{NK-KE}}$  and AVR-PikC (**Figure 2C**). However, co-expression of  $\text{Pikp-1}^{\text{NK-KE}}$  and AVR-PikC does not result in macroscopic cell death in *N. benthamiana* (**Figure 1C**).

These results show that effector variant association with full-length  $\text{Pikp-1}$  and  $\text{Pikp-1}^{\text{NK-KE}}$  correlates with the in planta cell-death response (**Figure 1C**).

### The effector-binding interface in the $\text{Pikp}^{\text{NK-KE}}$ mutant adopts a $\text{Pikm}$ -like conformation

Having established that the  $\text{Pikp}^{\text{NK-KE}}$  mutant displays an expanded effector recognition profile compared to wild-type  $\text{Pikp}$ , we sort to determine the structural basis of this activity. To this end, we determined the crystal structures of  $\text{Pikp-HMA}^{\text{NK-KE}}$  bound to AVR-PikD, and to AVR-PikE. We obtained samples of  $\text{Pikp-HMA}^{\text{NK-KE}}$ /AVR-PikD and  $\text{Pikp-HMA}^{\text{NK-KE}}$ /AVR-PikE complexes by co-expression in *E. coli* (described in the Materials and methods and by *De la Concepcion et al., 2018*). Each complex was crystallized (see Materials and methods) and X-ray diffraction data were collected at the Diamond Light Source (Oxford, UK) to 1.6 Å and 1.85 Å resolution, respectively. The details of X-ray data collection, structure solution, and completion are given in the Materials and methods and in **Table 1**.

The overall architecture of these complexes is the same as that observed for all  $\text{Pik-HMA/AVR-Pik}$  effector structures. A key interaction at interface 3, one of the previously defined  $\text{Pik-HMA-AVR-Pik}$  interfaces (*De la Concepcion et al., 2018*), involves a lysine residue (Lys262 in  $\text{Pikp}$  and  $\text{Pikm}$ ) that

**Table 1.** Data collection and refinement statistics

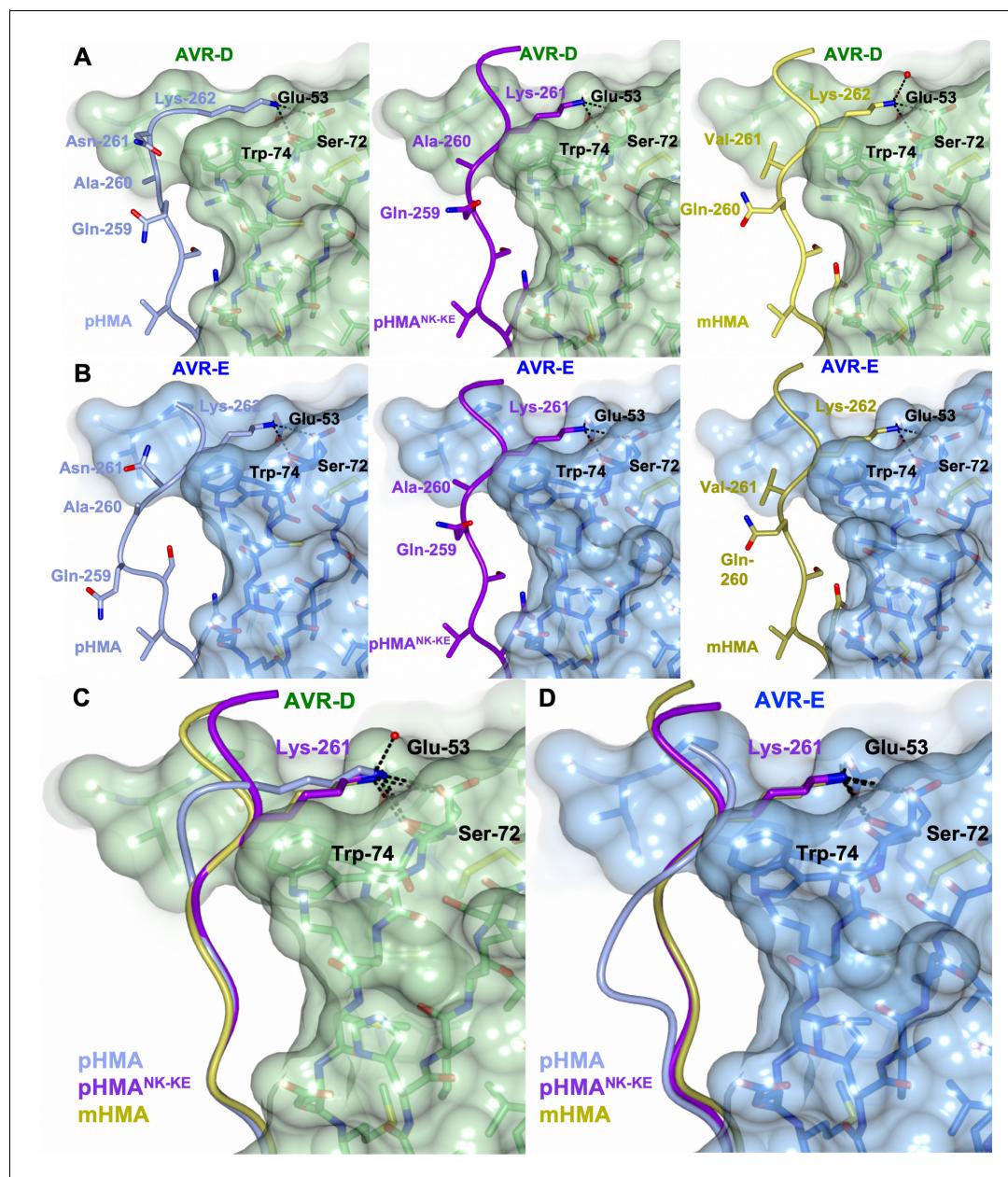
	$\text{Pikp}^{\text{NK-KE}}$ -AVR-PikD	$\text{Pikp}^{\text{NK-KE}}$ -AVR-PikE
<b>Data collection statistics</b>		
Wavelength (Å)	0.9763	0.9763
Space group	$P\ 2_1\ 2_1\ 2_1$	$P\ 2_1\ 2_1\ 2_1$
<b>Cell dimensions</b>		
$a, b, c$ (Å)	29.79, 65.33, 75.86	66.46, 80.70, 105.58
Resolution (Å)*	32.80–1.60 (1.63–1.60)	29.50–1.85 (1.89–1.85)
$R_{\text{merge}}$ (%)#	8.1 (97.1)	5.2 (75.1)
$I/\sigma^{\#}$	16.1 (2.6)	31.0 (4.1)
Completeness (%)#	100 (100)	99.8 (97.8)
Unique reflections#	20,294 (978)	49337 (2963)
Redundancy#	12.8 (13.3)	18.3 (17.8)
$CC^{(1/2)}$ (%)#	99.9 (86.6)	100 (95.2)
<b>Refinement and model statistics</b>		
Resolution (Å)	32.82–1.60 (1.64–1.60)	29.52–1.85 (1.90–1.85)
$R_{\text{work}}/R_{\text{free}}$ (%)	19.7/23.2 (25.5/27.3)	18.6/23.0 (29.1/35.0)
No. atoms (Protein)	1277	3604
B-factors (Protein)	25.6	39.7
<b>R.m.s. deviations</b>		
Bond lengths (Å)	0.009	0.012
Bond angles (°)	1.5	1.4
<b>Ramachandran plot (%)**</b>		
Favored	98.1	97.3
Outliers	0	0.2
MolProbity Score	1.41 (93 <sup>th</sup> percentile)	1.59 (91 <sup>st</sup> percentile)

\*The highest resolution shell is shown in parenthesis.

#As calculated by Aimless, As calculated by Refmac5, \*\*As calculated by MolProbity.

DOI: <https://doi.org/10.7554/eLife.47713.019>

forms intimate contacts within a pocket on the effector surface (**Figure 3**). In order to position this lysine in the effector pocket, Pikp has to loop-out regions adjacent to this residue, compromising the packing at the interface (**De la Concepcion et al., 2018, Figure 3A** [left panel], **B** [left panel], **C**



**Figure 3.** The  $\text{Pikp}^{\text{NK-KE}}$ -HMA mutant adopts a Pikm-like conformation at the effector-binding interface. Schematic view of the different conformations adopted by Pikp-HMA,  $\text{Pikp-HMA}^{\text{NK-KE}}$  and Pikm-HMA at interface 3 when in complex with AVR-PikD or AVR-PikE. In each panel, the effector is shown as sticks with the molecular surface also shown and colored as labeled. Pik-HMA residues are colored as labeled and shown in the  $\text{C}\alpha$ -worm with side-chain representation. (A) Schematic of Pikp-HMA (left),  $\text{Pikp-HMA}^{\text{NK-KE}}$  (middle) and Pikm-HMA (right) bound to AVR-PikD. Important residues in the HMA-effector interaction are labeled as shown. (B) Schematic of HMA residues as for panel (A), but bound to AVR-PikE. (C) Superposition showing Pikp-HMA,  $\text{Pikp-HMA}^{\text{NK-KE}}$  and Pikm-HMA chains (colored in blue, purple and yellow, respectively) bound to AVR-PikD. For clarity, only the Lys-261/262 side chain is shown. (D) Superposition as described before, but bound to AVR-PikE.

DOI: <https://doi.org/10.7554/eLife.47713.020>

The following figure supplement is available for figure 3:

**Figure supplement 1.** Interface 2 is essentially identical in the complexes comprising Pikp-HMA and  $\text{Pikp-HMA}^{\text{NK-KE}}$  bound to AVR-PikD or AVR-PikE.

DOI: <https://doi.org/10.7554/eLife.47713.021>

and D). By contrast, in *Pikm*, where the position of the lysine is shifted one residue to the N-terminus, no looping-out is required to locate the lysine within the pocket (**Figure 3A** [right panel], B [right panel], C and D). In the *Pikp*<sup>NK-KE</sup> mutant, the position of this key lysine is shifted one residue to the N-terminus compared to that in the wild-type, so that it occupies the same position in the *Pikp*<sup>NK-KE</sup> sequence as in the *Pikm* sequence. In the crystal structures of *Pikp-HMA*<sup>NK-KE</sup> in complex with either AVR-PikD or AVR-PikE, we see that this region of the HMA adopts a *Pikm*-like conformation (**Figure 3A** [middle panel], B [middle panel], C and D), with no looping-out of the preceding structure. This confirms that in creating the *Pikp*<sup>NK-KE</sup> mutant, we have resurfaced *Pikp* to have a more robust, *Pikm*-like interface in this region.

We found only limited structural perturbations at either of the other previously defined interfaces (interface 1 or 2 [*De la Concepcion et al., 2018*]) resulting from the binding of the AVR-PikD or AVR-PikE effectors to *Pikp-HMA* or *Pikp-HMA*<sup>NK-KE</sup> (**Figure 3—figure supplement 1**). We therefore conclude that the effects of the *Pikp*<sup>NK-KE</sup> mutant on protein function are mediated by altered interactions at interface 3.

### Mutation in AVR-Pik effectors at the engineered binding interface impacts the in planta response and in vivo binding

To further confirm that the engineered binding interface is responsible for the expanded recognition of AVR-PikE and AVR-PikA by *Pikp*<sup>NK-KE</sup>, we used mutants in the effectors at interface 2 (AVR-PikD<sup>H46E</sup>) and interface 3 (AVR-PikD,E,A<sup>E53R</sup>), which have previously been shown to impact the interactions and in planta responses of wild-type NLR alleles (*De la Concepcion et al., 2018*). We tested whether these mutants affected the cell-death response in *N. benthamiana*, the interactions between effectors and *Pikp-HMA*<sup>NK-KE</sup> (using Y2H), and the interactions between effectors and full-length *Pikp*<sup>NK-KE</sup> (using in planta co-immunoprecipitation [co-IP]).

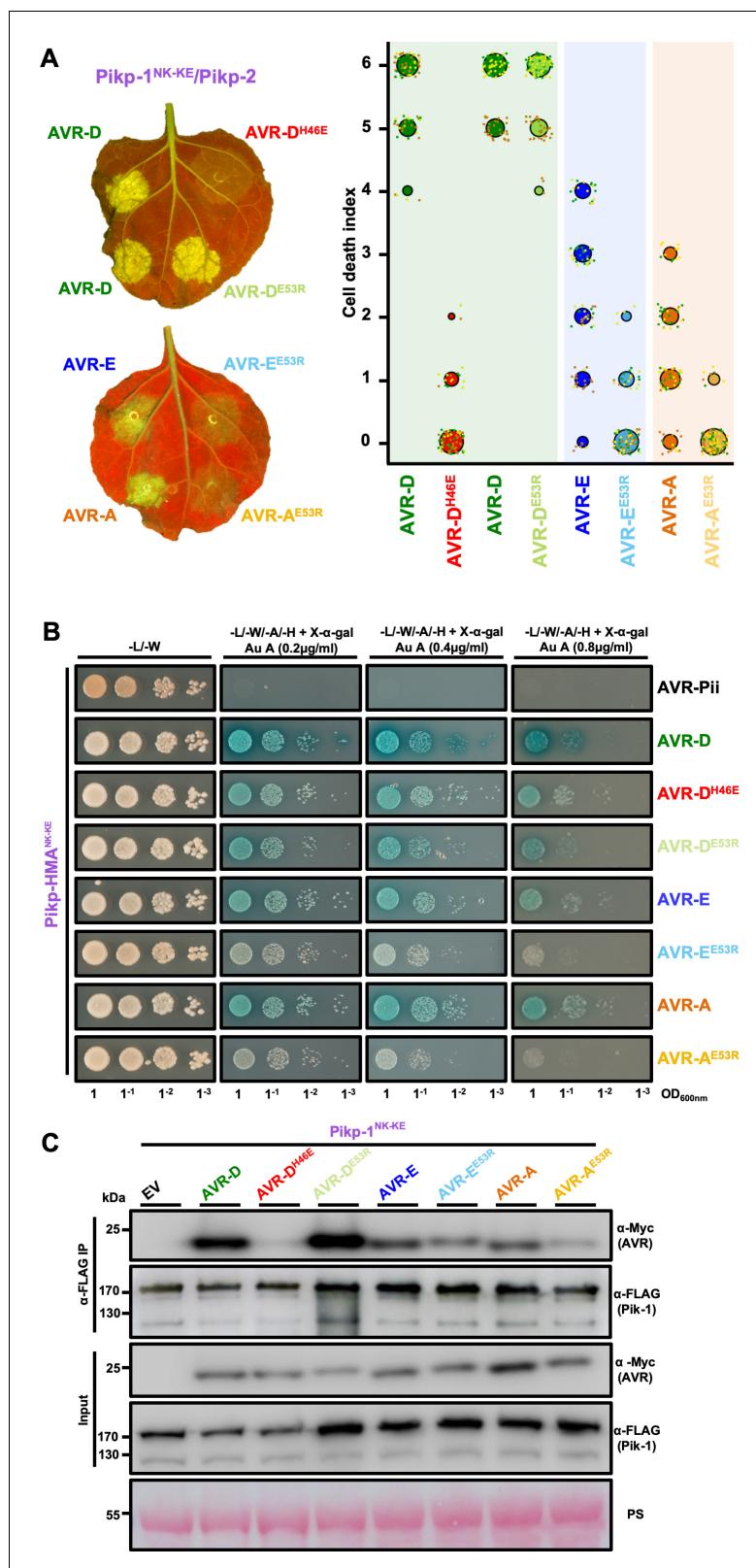
First, we investigated the impact of mutation at interface 2 using the AVR-PikD<sup>H46E</sup> mutant. We found that cell death in *N. benthamiana* is essentially blocked when co-expressing either *Pikp* or *Pikp*<sup>NK-KE</sup> with this mutant (AVR-PikD<sup>H46E</sup>), suggesting that the engineered NLR is still reliant on this interface for response (**Figure 4A**, **Figure 4—figure supplement 1**). Intriguingly, Y2H shows that the AVR-PikD<sup>H46E</sup> mutant displays some interaction with *Pikp-HMA*<sup>NK-KE</sup> (**Figure 4B**), similar to this mutant's interaction with *Pikm-HMA* (*De la Concepcion et al., 2018*), although this interaction is barely observed by co-IP with the full-length NLR (**Figure 4C**).

Second, we investigated the impact of mutations at interface 3 using the Glu53Arg (E53R) mutant in AVR-PikD, AVR-PikE and AVR-PikA. We found that the AVR-PikD<sup>E53R</sup> mutant has essentially no effect on recognition of the effector by *Pikp*<sup>NK-KE</sup> in *N. benthamiana*, and little or no effect on the interaction with *Pikp-HMA*<sup>NK-KE</sup> or full-length *Pikp*<sup>NK-KE</sup> (**Figure 4A,B,C**, **Figure 4—figure supplement 1**). By contrast, the equivalent mutation in AVR-PikE and AVR-PikA restricted the cell-death response in *N. benthamiana*, reduced the binding to *Pikp-HMA*<sup>NK-KE</sup> in Y2H (as shown by the reduced blue coloration) and produced a less intense band for the effector following *Pikp*<sup>NK-KE</sup> co-IP (**Figure 4A,B,C**). The expression of all proteins in yeast was confirmed by western blot (**Figure 4—figure supplement 2**).

These results provide evidence to show that while interactions across interface 2 remain important for the *Pikp*<sup>NK-KE</sup> interaction with AVR-Pik effectors, it is the altered interaction at interface 3, as observed in the structures, that is responsible for the expanded recognition profile of this engineered mutant.

## Discussion

Plants, including food crops, are under continuous threat from pathogens and pests, and new solutions to control disease are required. Although largely elusive to date, engineered plant NLR-type intracellular immune receptors have the potential to improve disease-resistance breeding (*Dangl et al., 2013; Rodriguez-Moreno et al., 2017*). NLR-integrated domains are a particularly attractive target for protein engineering because they interact directly with pathogen effectors (or host effector targets). Further, where tested, their binding affinities in vitro correlate with in planta immunity phenotypes (*Maqbool et al., 2015; De la Concepcion et al., 2018; Guo et al., 2018*), allowing biochemical and structural techniques to inform NLR design directly.



**Figure 4.** Mutation of AVR-Pik effectors at the engineered binding interface compromises binding and response. (A) (Left) A representative leaf image showing Pikp-1<sup>NK-KE</sup>-mediated cell death induced by AVR-Pik variants and mutants as autofluorescence under UV light (the AVR-PikD<sup>H46E</sup> mutant is located at interface 2, whereas the AVR-PikD<sup>E53R</sup>, AVR-PikE<sup>E53R</sup>, and AVR-PikA<sup>E53R</sup> mutants are located at interface 3). Autofluorescence intensity is scored

Figure 4 continued on next page

## Figure 4 continued

as in **Figure 1**. (Right)  $\text{Pikp}^{\text{NK-KE}}$  cell-death assay quantification in the form of dot plots. For each sample, the data points are represented as dots with a distinct color for each of the three biological replicates; these dots are plotted around the cell death score for visualization purposes. The size of the central dot at each cell death value is proportional to the number of replicates of the sample with that score. The number of repeats was 90. (B) Yeast-two-hybrid assay of  $\text{Pikp-HMA}^{\text{NK-KE}}$  with AVR-Pik variants and mutants. Control plate for yeast growth is on the left with quadruple dropout media supplemented with X- $\alpha$ -gal and increasing concentrations of Aureobasidin A on the right for each combination of HMA/AVR-Pik. The unrelated *M. oryzae* effector AVR-Pii was used as a negative control. Growth and the development of blue coloration in the selection plate are both indicative of protein–protein interaction. HMA domains were fused to the GAL4 DNA binding domain, and AVR-Pik alleles to the GAL4 activator domain. Each experiment was repeated a minimum of three times, with similar results. (C) Co-immunoprecipitation of full-length  $\text{Pikp-1}^{\text{NK-KE}}$  with AVR-Pik variants and mutants. N-terminally 4xMyc tagged AVR-Pik effectors were transiently co-expressed with  $\text{Pikp-1}^{\text{NK-KE}}:6x\text{His3xFLAG}$  in *N. benthamiana* leaves. Immunoprecipitates (IPs) obtained with anti-FLAG antiserum, and total protein extracts, were probed with appropriate antisera. Each experiment was repeated at least three times, with similar results. The asterisks mark the  $\text{Pik-1}$  band. PS = Ponceau Stain.

DOI: <https://doi.org/10.7554/eLife.47713.022>

The following source data and figure supplements are available for figure 4:

**Source data 1.** Cell-death scoring data used in the preparation of **Figure 4A**.

DOI: <https://doi.org/10.7554/eLife.47713.025>

**Figure supplement 1.** Estimation graphics for cell death induced by  $\text{Pikp}^{\text{NK-KE}}$  with AVR-Pik variants and mutants.

DOI: <https://doi.org/10.7554/eLife.47713.023>

**Figure supplement 2.** Western blot analysis confirming the accumulation of proteins in yeast.

DOI: <https://doi.org/10.7554/eLife.47713.024>

Here, we show that the recognition profile of the rice NLR  $\text{Pikp}$  can be expanded to include different AVR-Pik variants by engineering the binding interface between these proteins. This strengthens the hypothesis that tighter binding affinity between effectors and integrated HMA domains correlates with increased immune signaling in plants. This was previously shown for both natural alleles of  $\text{Pik}$  (Maqbool et al., 2015; De la Concepcion et al., 2018) and for  $\text{Pia}$  (Guo et al., 2018), but is now also shown for an engineered NLR. We propose this may be a general model for integrated domains that directly bind effectors.

Natural variation in  $\text{Pik}$  NLRs has given rise to different effector recognition profiles, and contributions from different binding interfaces were suggested to underpin this phenotype (De la Concepcion et al., 2018). In particular, a more favorable interaction at one interface (interface 3) in  $\text{Pikm}$ , compared to that in  $\text{Pikp}$ , was concluded to have evolved to compensate for changes in binding at a different site (interface 2). Here, through mutation of residues in  $\text{Pikp}$  (forming  $\text{Pikp}^{\text{NK-KE}}$ ), we have combined favorable interfaces from  $\text{Pikp}$  and  $\text{Pikm}$  into a single protein. This has resulted in an expanded recognition phenotype for effector binding and response in planta. *In vitro*, the binding of  $\text{Pikp}^{\text{NK-KE}}\text{-HMA}$  to effector variants is consistently higher than that of  $\text{Pikm}\text{-HMA}$  (**Figure 2—figure supplement 3**). This is likely to be underpinned by differences in a cluster of key residues within interface 2 (spatially equivalent residues Val222, Lys228, and Glu230 in  $\text{Pikp}$ , and Ala223, Gln229, and Val231 in  $\text{Pikm}$ ) that form more energetically favorable contacts, including water-mediated hydrogen bonds. However, despite a positive trend (which is significant in some cases), the levels of cell death mediated by  $\text{Pikm}$  and  $\text{Pikp}^{\text{NK-KE}}$  in planta are comparable (**Figure 1—figure supplements 2 and 4**).

Although the  $\text{Pikp}^{\text{NK-KE}}$  mutant did not deliver a cell-death response in *N. benthamiana* to the stealthy AVR-PikC effector variant, it did show a gain-of-binding *in vitro*, as well as weak binding in Y2H and in planta co-IP. It is surprising that the gain-of-binding for AVR-PikC *in vitro*, unlike that for AVR-PikA, does not correlate directly with gain of *in vivo* interactions or cell-death response in *N. benthamiana*. We hypothesize that this gain-of-binding for AVR-PikC observed *in vitro* is not sufficiently robust, especially in the context of the full-length NLR, for binding or for triggering immune signaling. Nevertheless, this work sets the scene for future interface engineering experiments that may further improve the response profiles of  $\text{Pik}$  NLRs to currently unrecognized effector variants. It also suggests that future work should test the disease resistance profile of *M. oryzae* strains carrying the different effector variants in rice expressing the engineered receptor.

The integrated HMA domain in the NLR RGA5 (the sensor of the Pia NLR pair in rice) binds to *M. oryzae* effectors AVR1-CO39 and AVR-Pia via a different interface, and it has been suggested that these binding sites are mutually exclusive (Guo et al., 2018). This raises the possibility that an HMA domain could be engineered to bind and respond to multiple effectors (Guo et al., 2018). Recently, the Pikp-HMA domain was shown to interact with AVR-Pia at the same interface as that used by the RGA5-HMA domain, and this probably underpins the partial resistance of rice plants to *M. oryzae* expressing AVR-Pia (Varden et al., 2019). This presents a starting point for the use of Pikp as a chassis for such studies. Although it remains to be seen whether any such resurfaced HMA domain can bind to multiple effectors, these studies suggest that this strategy has potential as a novel approach.

Plant breeding is required to provide new genetic solutions to disease resistance in crops. This is necessary to limit the environmental and social damage caused by pesticides, and to deal with changes in climate and in the globalization of agriculture that result in the spread of pathogens and pests into new environments (Islam et al., 2016; Michelmore et al., 2017; Savary et al., 2019). Classical breeding for disease resistance has been limited by issues such as linkage drag and hybrid incompatibility, as also seen in model plant species (Bomblies et al., 2007). Novel molecular approaches such as engineering 'decoys' (Kim et al., 2016) and protein resurfacing, as described here, combined with modern transformation (Altpeter et al., 2016) and breeding pipelines (Watson et al., 2018), offer the opportunity for more targeted approaches to breeding for disease resistance. These strategies will complement other emerging technologies in NLR identification (Arora et al., 2019) and NLR stacking (Dangl et al., 2013) as methods to develop improved crops for the future.

### Accession codes

Protein structures, and the data used to derive them, have been deposited at the Protein DataBank (PDB) with accession codes 6R8K (Pikp-HMA<sup>NK-KE</sup>/AVR-PikD) and 6R8M (Pikp-HMA<sup>NK-KE</sup>/AVR-PikE).

## Materials and methods

### Gene cloning

For in vitro studies, Pikp-HMA<sup>NK-KE</sup> (encompassing residues 186 to 263) was amplified from wild-type Pikp-HMA by introducing the mutations in the reverse primer, followed by cloning into pOPINM (Berrow et al., 2007). The wild-type Pikp-HMA, Pikm-HMA, and AVR-Pik expression constructs used in this study are as described in De la Concepcion et al. (2018).

For Y2H, we cloned Pikp-HMA<sup>NK-KE</sup> (as above) into pGBT7 using an In-Fusion cloning kit (Takara Bio USA), following the manufacturer's protocol. The wild-type Pikp-HMA domain in pGBT7 and AVR-Pik effector variants in pGADT7 were generated as described in De la Concepcion et al. (2018).

For protein expression in planta, the Pikp-HMA<sup>NK-KE</sup> domain was generated using site-directed mutagenesis by introducing the mutations in the reverse primer. This domain was then assembled into a full-length construct using Golden Gate cloning (Engler et al., 2008) and into the plasmid pICH47742 with a C-terminal 6xHis/3xFLAG tag. Expression was driven by the *Agrobacterium tumefaciens* Mas promoter and terminator. Full-length Pikp-1, Pikp-2, and AVR-Pik variants were generated as described in De la Concepcion et al. (2018).

### All DNA constructs were verified by sequencing

For the expression and purification of proteins for in vitro binding studies, pOPINM constructs encoding Pikp-HMA, Pikm-HMA and Pikp-HMA<sup>NK-KE</sup> were produced in *E. coli* SHuffle cells (Lobstein et al., 2012), using the protocol described in De la Concepcion et al. (2018). Cell cultures were grown in autoinduction media (Studier, 2005) at 30°C for 5–7 hr and then at 16°C overnight. Cells were harvested by centrifugation and re-suspended in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 50 mM glycine, 5% (vol/vol) glycerol, and 20 mM imidazole supplemented with EDTA-free protease inhibitor tablets (Roche). Cells were sonicated and, following centrifugation at 40,000xg for 30 min, the clarified lysate was applied to a Ni<sup>2+</sup>-NTA column connected to an AKTA Xpress purification system (GE Healthcare). Proteins were step-eluted with elution buffer (50 mM Tris-HCl [pH 7.5], 500 mM NaCl, 50 mM glycine, 5% (vol/vol) glycerol, and 500 mM imidazole) and directly injected onto a

Superdex 75 26/60 gel filtration column pre-equilibrated with 20 mM HEPES (pH 7.5) and 150 mM NaCl. Purification tags were removed by incubation with 3C protease (10 µg/mg fusion protein) followed by passing through tandem Ni<sup>2+</sup>-NTA and MBP Trap HP columns (GE Healthcare). The flow-through was concentrated as appropriate and loaded onto a Superdex 75 26/60 gel filtration column for final purification and buffer exchange into 20 mM HEPES (pH 7.5) and 150 mM NaCl.

AVR-Pik effectors, with either a 3C protease-cleavable N-terminal SUMO or a MBP tag and with a non-cleavable C-terminal 6xHis tag, were produced in and purified from *E. coli* SHuffle cells as previously described (Maqbool et al., 2015; De la Concepcion et al., 2018). All protein concentrations were determined using a Direct Detect Infrared Spectrometer (Merck).

### Co-expression and purification of Pik-HMA or AVR-Pik effectors for crystallization

Pikp-HMA<sup>NK-KE</sup> was co-expressed with AVR-PikD or AVR-PikE effectors in *E. coli* SHuffle cells following co-transformation of pOPINM:Pikp-HMA<sup>NK-KE</sup> and pOPINA:AVR-PikD/E (which were prepared as described in De la Concepcion et al., 2018). Cells were grown in autoinduction media (supplemented with both carbenicillin and kanamycin), harvested, and processed as described in De la Concepcion et al. (2018). Protein concentrations were measured using a Direct Detect Infrared Spectrometer (Merck).

### Protein–protein interaction: yeast-two-hybrid analyses

To detect protein–protein interactions between Pik-HMAs and AVR-Pik effectors in a yeast two-hybrid system, we used the Matchmaker Gold System (Takara Bio USA). Plasmid DNA encoding Pikp-HMA<sup>NK-KE</sup> in pGBKT7, generated in this study, was co-transformed into chemically competent Y2HGold cells (Takara Bio, USA) with the individual AVR-Pik variants or mutants in pGADT7, as described previously (De la Concepcion et al., 2018). Single colonies grown on selection plates were inoculated in 5 ml of SD<sup>-Leu-Trp</sup> overnight at 30 °C. Saturated culture was then used to make serial dilutions of OD<sub>600</sub> 1, 1<sup>-1</sup>, 1<sup>-2</sup>, and 1<sup>-3</sup>, respectively. 5 µl of each dilution was then spotted on a SD<sup>-Leu-Trp</sup> plate as a growth control, and on a SD<sup>-Leu-Trp-Ade-His</sup> plate containing X- $\alpha$ -gal and supplemented with Aureobasidin A (Takara Bio, USA). Plates were imaged after incubation for 60–72 hr at 30 °C. Each experiment was repeated a minimum of three times, with similar results.

To confirm protein expression in yeast, total protein extracts from transformed colonies were produced by boiling the cells for 10 min in LDS Runblue sample buffer. Samples were centrifugated and the supernatant was subjected to SDS-PAGE gels prior to western blotting. The membranes were probed with anti-GAL4 DNA-BD (Sigma) for the HMA domains in pGBKT7 and with the anti-GAL4 activation domain (Sigma) antibodies for the AVR-Pik effectors in pGADT7.

### Protein–protein interaction: surface plasmon resonance

Surface plasmon resonance (SPR) experiments were performed on a Biacore T200 system (GE Healthcare) using an NTA sensor chip (GE Healthcare). The system was maintained at 25°C, and a flow rate of 30 µl/min was used. All proteins were prepared in SPR running buffer (20 mM HEPES [pH 7.5], 860 mM NaCl, 0.1% Tween 20). C-terminally 6xHis-tag AVR-Pik variants were immobilized on the chip, giving a response of 200 ± 100. The sensor chip was regenerated between each cycle with an injection of 30 µl of 350 mM EDTA.

For all the assays, the level of binding was expressed as a percentage of the theoretical maximum response (R<sub>max</sub>) normalized for the amount of ligand immobilized on the chip. The cycling conditions were the same as those used in De la Concepcion et al. (2018). For each measurement, in addition to subtracting the response in the reference cell, a further buffer-only subtraction was made to correct for bulk refractive index changes or machine effects (Myszka, 1999). SPR data were exported and plotted using R v3.4.3 (<https://www.r-project.org/>) and the function ggplot2 (Wickham, 2009). Each experiment was repeated a minimum of three times, including internal repeats, with similar results. The proteins used came from three independent preparations for the HMA domains and two independent preparations of the AVR-Pik effectors.

### Protein–protein interaction: in planta co-immunoprecipitation (co-IP)

Transient gene-expression in planta for Co-IP was performed by delivering T-DNA constructs within the *A. tumefaciens* GV3101 strain into 4-week-old *N. benthamiana* plants grown at 22–25°C with high-light intensity. *A. tumefaciens* strains carrying *Pikp-1* or *Pikp-1<sup>NK-KE</sup>* were mixed with strains carrying the AVR-Pik effector, at OD<sub>600</sub> 0.2 each, in agroinfiltration medium (10 mM MgCl<sub>2</sub> and 10 mM 2-(N-morpholine)-ethanesulfonic acid (MES) [pH 5.6]), supplemented with 150 µM acetosyringone. For detection of complexes in planta, leaf tissue was collected 3 days post infiltration (dpi), frozen, and ground to fine powder in liquid nitrogen using a pestle and mortar. Leaf powder was mixed with two times weight/volume ice-cold extraction buffer (10% glycerol, 25 mM Tris [pH 7.5], 1 mM EDTA, 150 mM NaCl, 2% w/v PVPP, 10 mM DTT, 1x protease inhibitor cocktail [Sigma], 0.1% Tween 20 [Sigma]), centrifuged at 4,200 g/4°C for 20–30 min, and the supernatant was passed through a 0.45 µm Minisart syringe filter. The presence of each protein in the input was determined by SDS-PAGE and western blot. *Pik-1* and AVR-Pik effectors were detected by probing the membrane with anti-FLAG M2 antibody (SIGMA) and anti c-Myc monoclonal antibody (Santa Cruz), respectively. For immunoprecipitation, 1.5 ml of filtered plant extract was incubated with 30 µl of M2 anti-FLAG resin (Sigma) in a rotatory mixer at 4°C. After three hours, the resin was pelleted (800 g, 1 min) and the supernatant removed. The pellet was washed and resuspended in 1 ml of IP buffer (10% glycerol, 25 mM Tris [pH 7.5], 1 mM EDTA, 150 mM NaCl, 0.1% Tween 20 [Sigma]) and pelleted again by centrifugation as before. Washing steps were repeated five times. Finally, 30 µl of LDS Runblue sample buffer was added to the agarose and incubated for 10 min at 70°C. The resin was pelleted again, and the supernatant loaded onto SDS-PAGE gels prior to western blotting. Membranes were probed with anti-FLAG M2 (Sigma) and anti c-Myc (Santa Cruz) monoclonal antibodies.

### *N. benthamiana* cell-death assays

*A. tumefaciens* GV3101 vectors carrying *Pikp-1*, *Pikm-1*, or *Pikp-1<sup>NK-KE</sup>* were resuspended in induction media (10 mM MES [pH 5.6], 10 mM MgCl<sub>2</sub> and 150 µM acetosyringone) and mixed with *Pikp-2* (or *Pikm-2* in the *Pikm* cell-death assay), AVR-Pik effectors, and P19 at OD<sub>600</sub> 0.4, 0.4, 0.6 and 0.1, respectively. Four-week-old *N. benthamiana* leaves were infiltrated using a needleless syringe. Leaves were collected at 5 dpi to measure UV autofluorescence (a proxy for cell death) or ion leakage.

### Cell-death scoring: UV autofluorescence

Detached leaves were imaged at 5 dpi from the abaxial side of the leaves for UV fluorescence images. Photos were taken using a Nikon D4 camera with a 60 mm macro lens, ISO set 1600 and exposure ~10 secs at F14. The filter was a Kodak Wratten No.8 and white balance was set to 6250 degrees Kelvin. Blak-Ray longwave (365 nm) B-100AP spot light lamps were moved around the subject during the exposure to give an even illumination. Images shown are representative of three independent experiments, with internal repeats. The cell death index used for scoring was as presented previously (Maqbool et al., 2015). Dotplots were generated using R v3.4.3 (<https://www.r-project.org/>) and the graphic package ggplot2 (Wickham, H., 2009). The size of the centre dot at each cell death value is directly proportional to the number of replicates in the sample with that score. All individual data points are represented as dots.

### Cell-death scoring: ion leakage

For ion-leakage quantification, plants were infiltrated with the relevant constructs on two different leaves. At 5 dpi, leaves were detached and two leaf discs with a diameter of 8 mm (one disc per leaf spot) were collected and floated in 1.5 mL of Milli-Q water. Conductivity (µS/cm) was measured immediately after transferring the leaf disc to water (time zero) using a LAQUAtwin EC-33 conductivity meter (Horiba UK Ltd). Leaf discs were then incubated for 6 h at room temperature with gentle shaking before measuring the final conductivity. The assay was carried out in six biological replicates with a total of 40 technical replicates (2 discs x 40 plants). Conductivity data for each sample were exported and plotted using R v3.4.3 (<https://www.r-project.org/>) and the function ggplot2 (Wickham, H., 2009).

## Crystallization, data collection and structure solution

For crystallization, *Pikp-HMA<sup>NK-KE</sup>* in complex with *AVR-PikD* or *AVR-PikE* was concentrated following gel filtration. Sitting drop vapor diffusion crystallization trials were set up in 96 well plates, using an Oryx nano robot (Douglas Instruments, United Kingdom). Plates were incubated at 20°C, and crystals typically appeared after 24–48 hr. For data collection, all crystals were harvested from the Morpheus HT-96 screen (Molecular Dimensions), and snap-frozen in liquid nitrogen. Crystals used for data collection appeared from the following conditions: (i) *Pikp-HMA<sup>NK-KE</sup>*/*AVR-PikD* (10 mg/ml), Morpheus HT-96 condition D4 (0.12 M alcohols [0.2 M 1,6-hexanediol, 0.2 M 1-butanol, 0.2 M 1,2-propanediol, 0.2 M 2-propanol, 0.2 M 1,4-butanediol, and 0.2 M 1,3-propanediol], 0.1 M Buffer system 1 [1 M imidazole, MES monohydrate (acid) (pH 6.5)], and 50% v/v precipitant mix 4 [25%v/v MPD, 25% v/v PEG 1000, 25% v/v PEG3350]); (ii) *Pikp-HMA<sup>NK-KE</sup>*/*AVR-PikE* (15 mg/ml), Morpheus HT-96 condition A8 (0.06M divalents [0.3 M magnesium chloride hexahydrate, 0.3 M calcium chloride dihydrate]), 0.1M Buffer system 2 (sodium HEPES; MOPS [acid] [pH 7.5]), 37.5% v/v Precipitant mix 4 (25%v/v MPD, 25% v/v PEG 1000, 25% v/v PEG3350).

X-ray data sets were collected at the Diamond Light Source using beamline i03 (Oxford, UK). The data were processed using the xia2 pipeline (Winter, 2010) and CCP4 (Winn et al., 2011). The structures were solved by molecular replacement using PHASER (McCoy et al., 2007) and the coordinates of *AVR-PikD* and a monomer of *Pikp-HMA* from PDB entry 6G10. The final structures were obtained through iterative cycles of manual rebuilding and refinement using COOT (Emsley et al., 2010) and REFMAC5 (Murshudov et al., 2011), as implemented in CCP4 (Winn et al., 2011). Structures were validated using the tools provided in COOT and MOLPROBITY (Chen et al., 2010).

## Statistical analyses

Qualitative cell-death scoring from autofluorescence was analyzed using estimation methods (Ho et al., 2019) and visualized with estimation graphics using the besthr R library (MacLean, 2019). Briefly, in this process all autofluorescence (cell-death) scores in samples under comparison were ranked, irrespective of sample. The mean ranks of the control and test sample were taken and a bootstrap process was begun on ranked test data, in which samples of equal size to the experiment were replaced and the mean rank calculated. After 1000 bootstrap samples, rank means were calculated, a distribution of the mean ranks was drawn and its 2.5 and 97.5 quantiles calculated. If the mean of the control data is outside of these boundaries, the control and test means were considered to be different. Quantitative data from ion-leakage (cell-death) and SPR assays were analyzed by preparing a linear mixed effects model of sample on ion leakage/SPR and post-hoc comparisons performed for sample contrasts using Tukey's HSD method in the R package nlme (Pinheiro and Bates, 2019) and in lsmeans (Lenth, 2016).

## Acknowledgements

This work was supported by the BBSRC (grants BB/J004553, BB/P012574, BB/M02198X), the ERC (proposal 743165), the John Innes Foundation, the Gatsby Charitable Foundation, and JSPS KAKENHI 15H05779. We thank the Diamond Light Source, UK (beamline i03 under proposal MX13467) for access to X-ray data collection facilities. We also thank David Lawson and Clare Stevenson (JIC X-ray Crystallography/Biophysical Analysis Platform) for help with protein-structure determination and SPR, and Andrew Davis and Phil Robinson (JIC Bioimaging facilities) for photography.

## Additional information

### Funding

Funder	Grant reference number	Author
Biotechnology and Biological Sciences Research Council	BB/J004553	Kamoun S Banfield MJ
Biotechnology and Biological Sciences Research Council	BB/P012574	Kamoun S Banfield MJ

Biotechnology and Biological Sciences Research Council	BB/M02198X	Franceschetti M Kamoun S Banfield MJ
H2020 European Research Council	743165	Kamoun S Banfield MJ
John Innes Foundation		De la Concepcion JC Franceschetti M Banfield MJ
Gatsby Charitable Foundation		Sophien Kamoun
Japan Society for the Promotion of Science	15H05779	Ryohei Terauchi

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

#### Author contributions

Juan Carlos De la Concepcion, Formal analysis, Validation, Investigation, Methodology, Writing—original draft, Writing—review and editing; Marina Franceschetti, Formal analysis, Validation, Investigation, Methodology, Writing—review and editing; Dan MacLean, Conceptualization, Resources, Software, Funding acquisition, Validation, Visualization, Project administration, Writing—review and editing; Ryohei Terauchi, Conceptualization, Supervision, Funding acquisition, Project administration, Writing—review and editing; Sophien Kamoun, Conceptualization, Supervision, Funding acquisition, Validation, Writing—original draft, Project administration, Writing—review and editing; Mark J Banfield, Conceptualization, Resources, Software, Supervision, Funding acquisition, Validation, Visualization, Writing—original draft, Project administration, Writing—review and editing

#### Author ORCIDs

Sophien Kamoun  <https://orcid.org/0000-0002-0290-0315>

Mark J Banfield  <https://orcid.org/0000-0001-8921-3835>

#### Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.47713.035>

Author response <https://doi.org/10.7554/eLife.47713.036>

## Additional files

#### Supplementary files

- Supplementary file 1. Table of p-values for all pairwise comparisons of the ion-leakage data in *N. benthamiana*. Underlined values are those presented in the respective figures.

DOI: <https://doi.org/10.7554/eLife.47713.026>

- Supplementary file 2. Table of p-values for all pairwise comparisons of the SPR data including *Pikp* and *Pikp*<sup>NK-KE</sup>. Underlined values are those presented in the respective figures.

DOI: <https://doi.org/10.7554/eLife.47713.027>

- Transparent reporting form

DOI: <https://doi.org/10.7554/eLife.47713.028>

#### Data availability

Protein structures, and the data used to derive these, have been deposited at the Protein DataBank (PDB) with accession codes 6R8K (*Pikp-HMANK-KE/AVR-PikD*) and 6R8M (*Pikp-HMANK-KE/AVR-PikE*).

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Juan Carlos De la Concepcion, Marina	2019	Complex of rice blast ( <i>Magnaporthe oryzae</i> ) effector	<a href="https://www.rcsb.org/structure/6R8K">https://www.rcsb.org/structure/6R8K</a>	Protein Data Bank, 6R8K

Franceschetti, Mark J Banfield	protein AVR-PikD with an engineered HMA domain of Pikp-1 from rice ( <i>Oryza sativa</i> )		
Juan Carlos De la Concepcion, Marina Franceschetti, Mark J Banfield	2019 Complex of rice blast ( <i>Magnaporthe oryzae</i> ) effector protein AVR-PikE with an engineered HMA domain of Pikp-1 from rice ( <i>Oryza sativa</i> )	<a href="https://www.rcsb.org/structure/6R8M">https://www.rcsb.org/structure/6R8M</a>	Protein Data Bank, 6R8M

## References

Altpeter F, Springer NM, Bartley LE, Blechl AE, Brutnell TP, Citovsky V, Conrad LJ, Gelvin SB, Jackson DP, Kausch AP, Lemaux PG, Medford JI, Orozco-Cárdenas ML, Tricoli DM, Van Eck J, Voytas DF, Walbot V, Wang K, Zhang ZJ, Stewart CN. 2016. Advancing crop transformation in the era of genome editing. *The Plant Cell* **28**: 1510–1520. DOI: <https://doi.org/10.1105/tpc.16.00196>, PMID: 27335450

Arora S, Steuernagel B, Gaurav K, Chandramohan S, Long Y, Matny O, Johnson R, Enk J, Periyannan S, Singh N, Asyraf Md Hatta M, Athiayannan N, Cheema J, Yu G, Kangara N, Ghosh S, Szabo LJ, Poland J, Bariana H, Jones JDG, et al. 2019. Resistance gene cloning from a wild crop relative by sequence capture and association genetics. *Nature Biotechnology* **37**:139–143. DOI: <https://doi.org/10.1038/s41587-018-0007-9>, PMID: 30718880

Bailey PC, Schudoma C, Jackson W, Baggs E, Dagdas G, Haerty W, Moscou M, Krasileva KV. 2018. Dominant integration locus drives continuous diversification of plant immune receptors with exogenous domain fusions. *Genome Biology* **19**:23. DOI: <https://doi.org/10.1186/s13059-018-1392-6>, PMID: 29458393

Berrow NS, Alderton D, Sainsbury S, Nettleship J, Assenberg R, Rahman N, Stuart DL, Owens RJ. 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research* **35**:e45. DOI: <https://doi.org/10.1093/nar/gkm047>, PMID: 17317681

Bialas A, Zess EK, De la Concepcion JC, Franceschetti M, Pennington HG, Yoshida K, Upson JL, Chanclud E, Wu CH, Langner T, Maqbool A, Varden FA, Derevina L, Belhaj K, Fujisaki K, Saitoh H, Terauchi R, Banfield MJ, Kamoun S. 2018. Lessons in effector and NLR biology of Plant-Microbe systems. *Molecular Plant-Microbe Interactions* **31**:34–45. DOI: <https://doi.org/10.1094/MPMI-08-17-0196-FI>, PMID: 29144205

Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weigel D. 2007. Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLOS Biology* **5**:e236. DOI: <https://doi.org/10.1371/journal.pbio.0050236>, PMID: 17803357

Carter ME, Helm M, Chapman A, Wan E, Restrepo Sierra AM, Innes R, Bogdanove AJ, Wise RP. 2018. Convergent evolution of effector protease recognition by *Arabidopsis* and barley. *Molecular Plant-Microbe Interactions* : MPMI **32**:550–565. DOI: <https://doi.org/10.1094/MPMI-07-18-0202-FI>

Cesari S, Thilliez G, Ribot C, Chalvon V, Michel C, Jauneau A, Rivas S, Alaux L, Kanzaki H, Okuyama Y, Morel JB, Fournier E, Tharreau D, Terauchi R, Koj T. 2013. The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. *The Plant Cell* **25**:1463–1481. DOI: <https://doi.org/10.1105/tpc.112.107201>, PMID: 23548743

Cesari S, Bernoux M, Moncquet P, Koj T, Dodds PN. 2014. A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis. *Frontiers in Plant Science* **5**:606. DOI: <https://doi.org/10.3389/fpls.2014.00606>, PMID: 25506347

Cesari S. 2018. Multiple strategies for pathogen perception by plant immune receptors. *New Phytologist* **219**: 17–24. DOI: <https://doi.org/10.1111/nph.14877>, PMID: 29131341

Chapman AM, McNaughton BR. 2016. Scratching the surface: resurfacing proteins to endow new properties and function. *Cell Chemical Biology* **23**:543–553. DOI: <https://doi.org/10.1016/j.chembiol.2016.04.010>, PMID: 27203375

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography* **66**:12–21. DOI: <https://doi.org/10.1107/S0907444909042073>, PMID: 20057044

Costanzo S, Jia Y. 2010. Sequence variation at the rice blast resistance gene Pi-km locus: implications for the development of allele specific markers. *Plant Science* **178**:523–530. DOI: <https://doi.org/10.1016/j.plantsci.2010.02.014>

Dangl JL, Horvath DM, Staskawicz BJ. 2013. Pivoting the plant immune system from dissection to deployment. *Science* **341**:746–751. DOI: <https://doi.org/10.1126/science.1236011>, PMID: 23950531

De la Concepcion JC, Franceschetti M, Maqbool A, Saitoh H, Terauchi R, Kamoun S, Banfield MJ. 2018. Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen. *Nature Plants* **4**:576–585. DOI: <https://doi.org/10.1038/s41477-018-0194-x>, PMID: 29988155

Eitas TK, Dangl JL. 2010. NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Current Opinion in Plant Biology* **13**:472–477. DOI: <https://doi.org/10.1016/j.pbi.2010.04.007>, PMID: 20483655

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of coot. *Acta Crystallographica Section D, Biological Crystallography* **66**:486–501. DOI: <https://doi.org/10.1107/S0907444910007493>, PMID: 20383002

**Engler C**, Kandzia R, Marillonnet S. 2008. A one pot, one step, precision cloning method with high throughput capability. *PLOS ONE* **3**:e3647. DOI: <https://doi.org/10.1371/journal.pone.0003647>, PMID: 18985154

**Fujisaki KA**, Kanzaki A, Ito E, Utsushi K, Saitoh H, Bialas H, Banfield A, Kamoun S, Terauchi R. 2017. An unconventional NOI/RIN4 domain of a rice NLR protein binds host EXO70 protein to confer fungal immunity. *bioRxiv*. DOI: <https://doi.org/10.1101/239400>

**Giannakopoulou A**, Steele JF, Segretin ME, Bozkurt TO, Zhou J, Robatzek S, Banfield MJ, Pais M, Kamoun S. 2015. Tomato I2 immune receptor can be engineered to confer partial resistance to the oomycete *Phytophthora infestans* in Addition to the Fungus *Fusarium oxysporum*. *Molecular Plant-Microbe Interactions* **28**:1316–1329. DOI: <https://doi.org/10.1094/MPMI-07-15-0147-R>, PMID: 26367241

**Guo L**, Cesari S, de Guillen K, Chalvon V, Mammri L, Ma M, Meusnier I, Bonnot F, Padilla A, Peng YL, Liu J, Kroj T. 2018. Specific recognition of two MAX effectors by integrated HMA domains in plant immune receptors involves distinct binding surfaces. *PNAS* **115**:11637–11642. DOI: <https://doi.org/10.1073/pnas.1810705115>, PMID: 30355769

**Harris CJ**, Slootweg EJ, Goverse A, Baulcombe DC. 2013. Stepwise artificial evolution of a plant disease resistance gene. *PNAS* **110**:21189–21194. DOI: <https://doi.org/10.1073/pnas.1311134110>, PMID: 24324167

**Helm M**, Qi M, Sarkar S, Yu H, Whitham SA, Innes RW. 2019. Engineering a decoy substrate in soybean to enable recognition of the soybean mosaic virus Nla protease. *Molecular Plant-Microbe Interactions* **32**:760–769. DOI: <https://doi.org/10.1094/MPMI-12-18-0324-R>, PMID: 30676230

**Ho J**, Tumkaya T, Aryal S, Choi H, Claridge-Chang A. 2019. Moving beyond P values: data analysis with estimation graphics. *Nature Methods* **16**:565–566. DOI: <https://doi.org/10.1038/s41592-019-0470-3>, PMID: 31217592

**Islam MT**, Croll D, Gladieux P, Soanes DM, Persoons A, Bhattacharjee P, Hossain MS, Gupta DR, Rahman MM, Mahboob MG, Cook N, Salam MU, Surovy MZ, Sancho VB, Maciel JL, NhaniJúnior A, Castroagudín VL, Reges JT, Ceresini PC, Ravel S, et al. 2016. Emergence of wheat blast in Bangladesh was caused by a south american lineage of *Magnaporthe oryzae*. *BMC Biology* **14**:84. DOI: <https://doi.org/10.1186/s12915-016-0309-7>, PMID: 27716181

**Kanzaki H**, Yoshida K, Saitoh H, Fujisaki K, Hirabuchi A, Alaux L, Fournier E, Tharreau D, Terauchi R. 2012. Arms race co-evolution of *Magnaporthe oryzae* AVR-Pik and rice pik genes driven by their physical interactions. *The Plant Journal : For Cell and Molecular Biology* **72**:894–907. DOI: <https://doi.org/10.1111/j.1365-313X.2012.05110.x>, PMID: 22805093

**Kim SH**, Qi D, Ashfield T, Helm M, Innes RW. 2016. Using decoys to expand the recognition specificity of a plant disease resistance protein. *Science* **351**:684–687. DOI: <https://doi.org/10.1126/science.aad3436>, PMID: 26912853

**Kourelis J**, van der Hoorn RAL. 2018. Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. *The Plant Cell* **30**:285–299. DOI: <https://doi.org/10.1105/tpc.17.00579>, PMID: 29382771

**Kroj T**, Chanclud E, Michel-Romiti C, Grand X, Morel JB. 2016. Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytologist* **210**:618–626. DOI: <https://doi.org/10.1111/nph.13869>, PMID: 26848538

**Le Roux C**, Huet G, Jauneau A, Camborde L, Trémousaygue D, Kraut A, Zhou B, Levaillant M, Adachi H, Yoshioka H, Raffaele S, Berthomé R, Couté Y, Parker JE, Deslandes L. 2015. A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity. *Cell* **161**:1074–1088. DOI: <https://doi.org/10.1016/j.cell.2015.04.025>, PMID: 26000483

**Lenth RV**. 2016. Least-Squares means: the R package lsmeans. *Journal of Statistical Software* **69**:1–33. DOI: <https://doi.org/10.18637/jss.v069.i01>

**Lobstein J**, Emrich CA, Jeans C, Faulkner M, Riggs P, Berkmen M. 2012. SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories* **11**:56. DOI: <https://doi.org/10.1186/1475-2859-11-56>, PMID: 22569138

**MacLean D**. 2019. Besthr. Zenodo. DOI: <https://doi.org/10.5281/zenodo.3374507>

**Maqbool A**, Saitoh H, Franceschetti M, Stevenson CEM, Uemura A, Kanzaki H, Kamoun S, Terauchi R, Banfield MJ. 2015. Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *eLife* **4**:e08709. DOI: <https://doi.org/10.7554/eLife.08709>

**McCoy AJ**, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *Journal of Applied Crystallography* **40**:658–674. DOI: <https://doi.org/10.1107/S0021889807021206>, PMID: 19461840

**Michelmore R**, Coaker G, Bart R, Beattie G, Bent A, Bruce T, Cameron D, Dangl J, Dinesh-Kumar S, Edwards R, Eves-van den Akker S, Gassmann W, Greenberg JT, Hanley-Bowdoin L, Harrison RJ, Harvey J, He P, Huffaker A, Hulbert S, Innes R, et al. 2017. Foundational and translational research opportunities to improve plant health. *Molecular Plant-Microbe Interactions* **30**:515–516. DOI: <https://doi.org/10.1094/MPMI-01-17-0010-CR>, PMID: 28398839

**Murshudov GN**, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D Biological Crystallography* **67**:355–367. DOI: <https://doi.org/10.1107/S0907444911001314>, PMID: 21460454

**Myszka DG**. 1999. Improving biosensor analysis. *Journal of Molecular Recognition* **12**:279–284. DOI: [https://doi.org/10.1002/\(SICI\)1099-1352\(199909/10\)12:5<279::AID-JMR473>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1099-1352(199909/10)12:5<279::AID-JMR473>3.0.CO;2-3), PMID: 10556875

**Ortiz D**, de Guillen K, Cesari S, Chalvon V, Gracy J, Padilla A, Koj 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. DOI: <https://doi.org/10.1105/tpc.16.00435>, PMID: 28087830

**Pinheiro J**, Bates D. 2019. nlme: Linear and Nonlinear Mixed Effects Models. R Package Version. <https://CRAN.R-project.org/package=nlme>

**Rodriguez-Moreno L**, Song Y, Thomma BP. 2017. Transfer and engineering of immune receptors to improve recognition capacities in crops. *Current Opinion in Plant Biology* **38**:42–49. DOI: <https://doi.org/10.1016/j.pbi.2017.04.010>, PMID: 28472757

**Sarris PF**, Duxbury Z, Huh SU, Ma Y, Segonzac C, Sklenar J, Derbyshire P, Cevik V, Rallapalli G, Saucet SB, Wirthmueller L, Menke FLH, Sohn KH, Jones JDG. 2015. A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell* **161**:1089–1100. DOI: <https://doi.org/10.1016/j.cell.2015.04.024>, PMID: 26000484

**Sarris PF**, Cevik V, Dagdas G, Jones JD, Krasileva KV. 2016. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biology* **14**:8. DOI: <https://doi.org/10.1186/s12915-016-0228-7>, PMID: 26891798

**Savary S**, Willocquet L, Pethybridge SJ, Esker P, McRoberts N, Nelson A. 2019. The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution* **3**:430–439. DOI: <https://doi.org/10.1038/s41559-018-0793-y>, PMID: 30718852

**Segretin ME**, Pais M, Franceschetti M, Chaparro-Garcia A, Bos JI, Banfield MJ, Kamoun S. 2014. Single amino acid mutations in the potato immune receptor R3a expand response to Phytophthora effectors. *Molecular Plant-Microbe Interactions* **27**:624–637. DOI: <https://doi.org/10.1094/MPMI-02-14-0040-R>, PMID: 24678835

**Studier FW**. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expression and Purification* **41**:207–234. DOI: <https://doi.org/10.1016/j.pep.2005.01.016>, PMID: 15915565

**Takken FL**, Goverse A. 2012. How to build a pathogen detector: structural basis of NB-LRR function. *Current Opinion in Plant Biology* **15**:375–384. DOI: <https://doi.org/10.1016/j.pbi.2012.05.001>, PMID: 22658703

**Varden FA**, Saitoh H, Yoshino K, Franceschetti M, Kamoun S, Terauchi R, Banfield MJ. 2019. Cross-reactivity of a rice NLR immune receptor to distinct effectors from the rice blast pathogen *Magnaporthe oryzae* provides partial disease resistance. *Journal of Biological Chemistry* **294**:13006–13016. DOI: <https://doi.org/10.1074/jbc.RA119.007730>, PMID: 31296569

**Watson A**, Ghosh S, Williams MJ, Cuddy WS, Simmonds J, Rey MD, Asyraf Md Hatta M, Hinchliffe A, Steed A, Reynolds D, Adamski NM, Breakspear A, Korolev A, Rayner T, Dixon LE, Riaz A, Martin W, Ryan M, Edwards D, Batley J, et al. 2018. Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants* **4**:23–29. DOI: <https://doi.org/10.1038/s41477-017-0083-8>, PMID: 29292376

**Winn MD**, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallographica. Section D, Biological Crystallography* **67**:235–242. DOI: <https://doi.org/10.1107/S0907444910045749>, PMID: 21460441

**Winter G**. 2010. xia2 : an expert system for macromolecular crystallography data reduction . *Journal of Applied Crystallography* **43**:186–190. DOI: <https://doi.org/10.1107/S0021889809045701>

**Yoshida K**, Saitoh H, Fujisawa S, Kanzaki H, Matsumura H, Yoshida K, Tosa Y, Chuma I, Takano Y, Win J, Kamoun S, Terauchi R. 2009. Association genetics reveals three novel avirulence genes from the rice blast fungal pathogen *Magnaporthe oryzae*. *The Plant Cell* **21**:1573–1591. DOI: <https://doi.org/10.1105/tpc.109.066324>, PMID: 19454732

**Yoshida K**, Saunders DG, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, Inoue Y, Chuma I, Tosa Y, Cano LM, Kamoun S, Terauchi R. 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. DOI: <https://doi.org/10.1186/s12864-016-2690-6>, PMID: 27194050



# Preliminary Full wwPDB X-ray Structure Validation Report (i)

Feb 7, 2018 – 12:52 PM EST

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the standalone wwPDB validation server.  
**The structure in question has not been deposited to the wwPDB.**  
**This report should not be submitted to journals.**

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030736

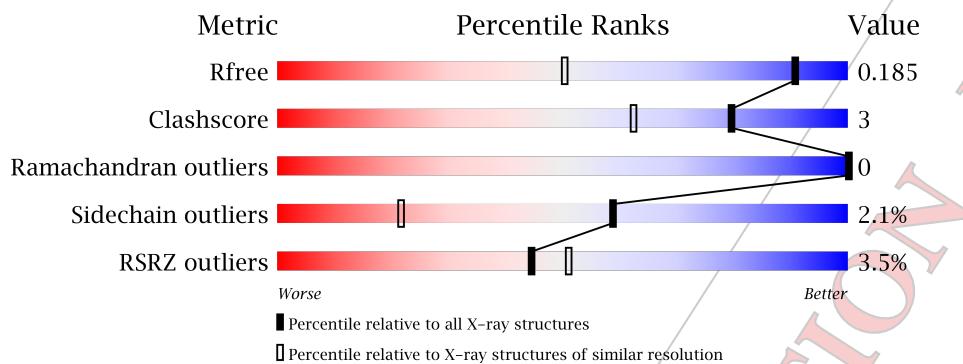
## 1 Overall quality at a glance [\(i\)](#)

The following experimental techniques were used to determine the structure:

### X-RAY DIFFRACTION

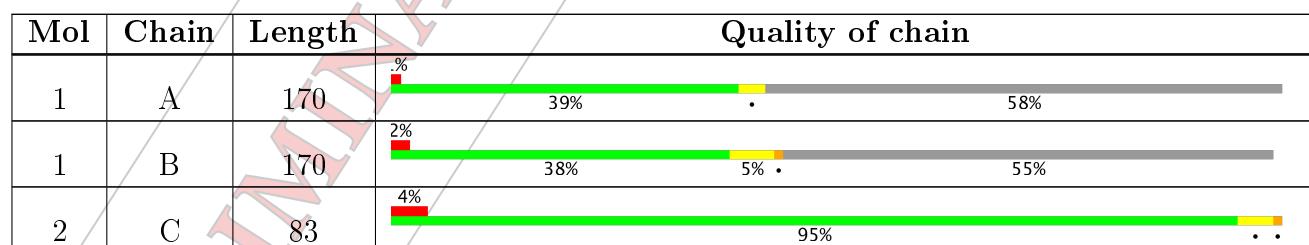
The reported resolution of this entry is 1.35 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	1024 (1.38-1.34)
Clashscore	112137	1063 (1.38-1.34)
Ramachandran outliers	110173	1048 (1.38-1.34)
Sidechain outliers	110143	1048 (1.38-1.34)
RSRZ outliers	101464	1025 (1.38-1.34)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.



The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	CL	D	1	-	-	-	X

PRELIMINARY VALIDATION REPORT

## 2 Entry composition [\(i\)](#)

There are 4 unique types of molecules in this entry. The entry contains 2024 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
1	A	71	Total C N O S 537 339 93 102 3	0	3	0
1	B	76	Total C N O S 592 373 104 112 3	0	5	0

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
2	C	83	Total C N O S 694 447 121 119 7	0	3	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	D	1	Total Cl 1 1	0	0

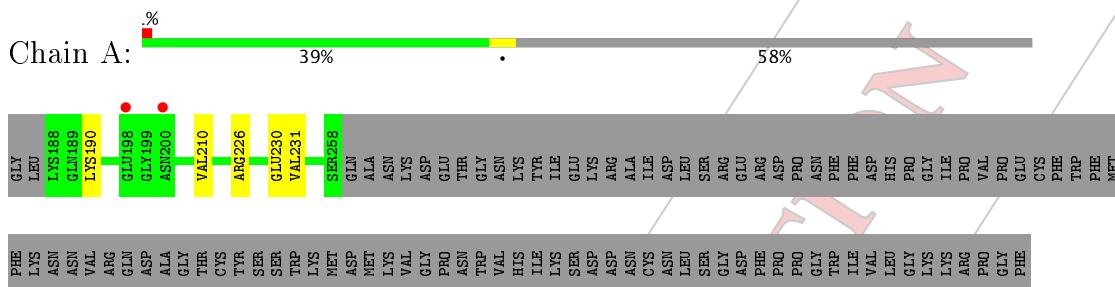
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	E	199	Total O 200 200	0	1

### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

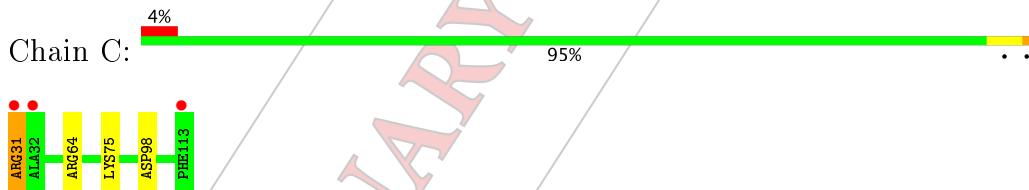
- Molecule 1:



- Molecule 1:



- Molecule 2:



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	40.50 Å   66.26 Å   45.74 Å 90.00°   113.32°   90.00°	Depositor
Resolution (Å)	66.26 – 1.35 42.00 – 1.35	Depositor EDS
% Data completeness (in resolution range)	99.8 (66.26-1.35) 99.8 (42.00-1.35)	Depositor EDS
$R_{\text{merge}}$	(Not available)	Depositor
$R_{\text{sym}}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle^1$	3.53 (at 1.35 Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
$R, R_{\text{free}}$	0.143, 0.184 0.144, 0.185	Depositor DCC
$R_{\text{free}}$ test set	2368 reflections (5.12%)	DCC
Wilson B-factor (Å <sup>2</sup> )	17.1	Xtriage
Anisotropy	0.274	Xtriage
Bulk solvent $k_{\text{sol}}$ (e/Å <sup>3</sup> ), $B_{\text{sol}}$ (Å <sup>2</sup> )	0.40, 56.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50, \langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.025 for h,-k,-h-l	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	2024	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.70% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [\(i\)](#)

### 5.1 Standard geometry [\(i\)](#)

Bond lengths and bond angles in the following residue types are not validated in this section:  
CL

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# $ Z  > 5$	RMSZ	# $ Z  > 5$
1	A	0.80	0/544	0.89	0/729
1	B	0.94	0/599	1.08	2/803 (0.2%)
2	C	0.97	0/728	1.05	4/983 (0.4%)
All	All	0.91	0/1871	1.02	6/2515 (0.2%)

There are no bond length outliers.

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	C	31	ARG	NE-CZ-NH1	7.60	124.10	120.30
1	B	197	MET	CG-SD-CE	-7.12	88.81	100.20
2	C	64	ARG	NE-CZ-NH1	-6.99	116.81	120.30
2	C	64	ARG	NE-CZ-NH2	6.68	123.64	120.30
2	C	98	ASP	CB-CG-OD1	5.56	123.31	118.30
1	B	187	LEU	CA-CB-CG	5.09	127.02	115.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [\(i\)](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	537	0	587	4	0
1	B	592	0	645	8	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	694	0	665	0	0
3	D	1	0	0	0	0
4	E	200	0	0	3	0
All	All	2024	0	1897	12	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 3.

All (12) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:242:ILE:HD11	1:B:254:LEU:HB2	1.19	1.14
1:B:242:ILE:CD1	1:B:254:LEU:HB2	1.99	0.91
1:B:192:VAL:HG22	1:B:230[B]:GLU:HG2	1.67	0.77
1:B:228:LYS:NZ	4:E:242:HOH:O	2.28	0.66
1:A:226:ARG:NH2	4:E:173:HOH:O	2.34	0.60
1:B:242:ILE:HD11	1:B:254:LEU:CB	2.13	0.60
1:B:189[B]:GLN:OE1	1:B:259:GLN:NE2	2.40	0.55
1:B:189[B]:GLN:HE22	1:B:259:GLN:CD	2.12	0.53
1:A:190:LYS:HE2	1:A:230[B]:GLU:HG3	1.95	0.48
1:B:209:LEU:HD11	1:B:248:LYS:HE3	1.96	0.47
1:A:226:ARG:HD3	4:E:62:HOH:O	2.15	0.46
1:A:210:VAL:HG11	1:A:231[B]:VAL:HG22	1.98	0.46

There are no symmetry-related clashes.

### 5.3 Torsion angles [i](#)

#### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	72/170 (42%)	71 (99%)	1 (1%)	0	100 100
1	B	79/170 (46%)	77 (98%)	2 (2%)	0	100 100

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	C	84/83 (101%)	83 (99%)	1 (1%)	0	100 100
All	All	235/423 (56%)	231 (98%)	4 (2%)	0	100 100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains [\(i\)](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	60/144 (42%)	60 (100%)	0	100 100
1	B	66/144 (46%)	63 (96%)	3 (4%)	32 4
2	C	77/74 (104%)	74 (96%)	3 (4%)	37 5
All	All	203/362 (56%)	197 (97%)	6 (3%)	59 11

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	187	LEU
1	B	189[A]	GLN
1	B	189[B]	GLN
2	C	31	ARG
2	C	75[A]	LYS
2	C	75[B]	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	200	ASN
1	A	256	GLN
1	B	200	ASN
1	B	259	GLN

### 5.3.3 RNA [\(i\)](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [\(i\)](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [\(i\)](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [\(i\)](#)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers [\(i\)](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [\(i\)](#)

There are no chain breaks in this entry.

## 6 Fit of model and data (i)

### 6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	71/170 (41%)	0.08	2 (2%) 53 61	14, 23, 37, 43	0
1	B	76/170 (44%)	0.27	3 (3%) 40 46	13, 19, 44, 48	0
2	C	83/83 (100%)	0.06	3 (3%) 43 50	13, 19, 36, 46	0
All	All	230/423 (54%)	0.14	8 (3%) 44 51	13, 20, 38, 48	0

All (8) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	32	ALA	6.4
2	C	31	ARG	4.7
2	C	113	PHE	4.5
1	B	200	ASN	3.3
1	B	261	ASN	3.2
1	A	198	GLU	2.7
1	B	187	LEU	2.5
1	A	200	ASN	2.1

### 6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

### 6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron

density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
3	CL	D	1	1/?	0.99	0.16	2.02	51,51,51,51	0

## 6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.



# Preliminary Full wwPDB X-ray Structure Validation Report (i)

Feb 7, 2018 – 12:35 PM EST

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the standalone wwPDB validation server.  
**The structure in question has not been deposited to the wwPDB.**  
**This report should not be submitted to journals.**

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030736

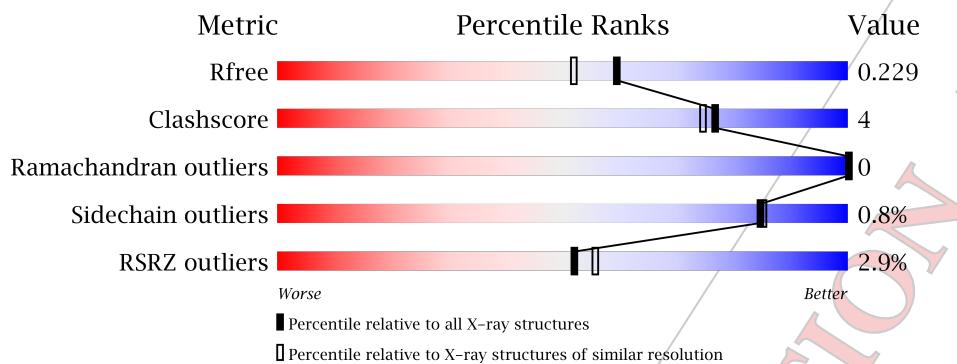
## 1 Overall quality at a glance [\(i\)](#)

The following experimental techniques were used to determine the structure:

### X-RAY DIFFRACTION

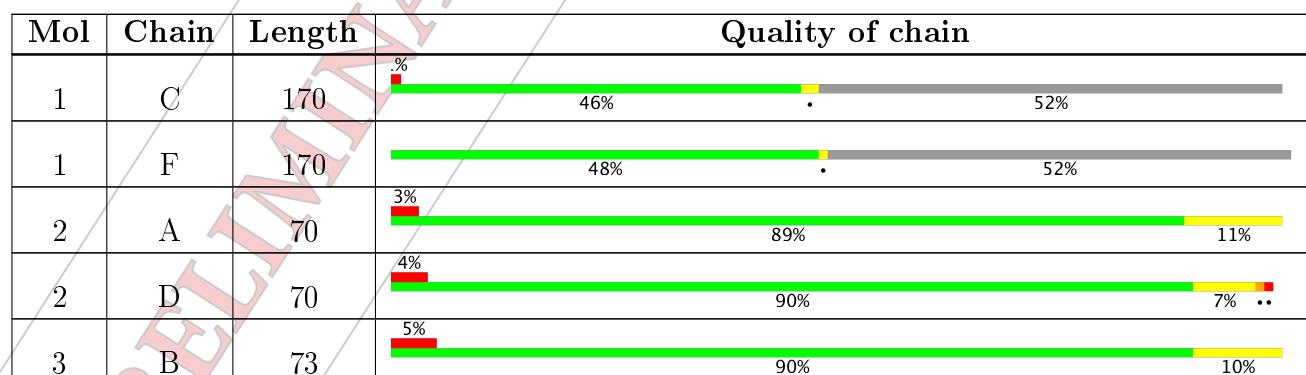
The reported resolution of this entry is 1.90 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	5047 (1.90-1.90)
Clashscore	112137	5731 (1.90-1.90)
Ramachandran outliers	110173	5669 (1.90-1.90)
Sidechain outliers	110143	5670 (1.90-1.90)
RSRZ outliers	101464	5100 (1.90-1.90)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.



Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	E	73	<div style="width: 4%; background-color: red; display: inline-block;">4%</div> <div style="width: 88%; background-color: green; display: inline-block;">88%</div> <div style="width: 12%; background-color: yellow; display: inline-block;">12%</div>

PRELIMINARY VALIDATION REPORT

## 2 Entry composition [\(i\)](#)

There are 4 unique types of molecules in this entry. The entry contains 3712 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
1	C	82	Total C N O S 665 427 114 118 6	0	0	0
1	F	82	Total C N O S 665 427 114 118 6	0	0	0

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
2	A	70	Total C N O S 518 329 91 95 3	0	1	0
2	D	70	Total C N O S 518 329 91 95 3	0	1	0

- Molecule 3 is a protein.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
3	B	73	Total C N O S 561 352 100 106 3	0	7	0
3	E	73	Total C N O S 555 349 98 105 3	0	8	0

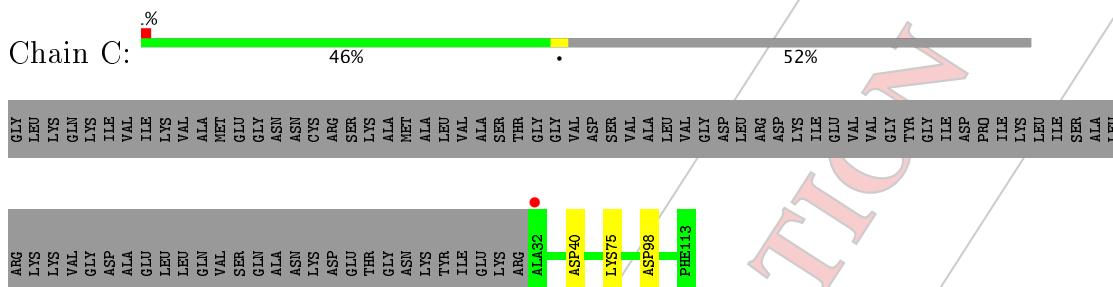
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	H	227	Total O 230 230	0	3

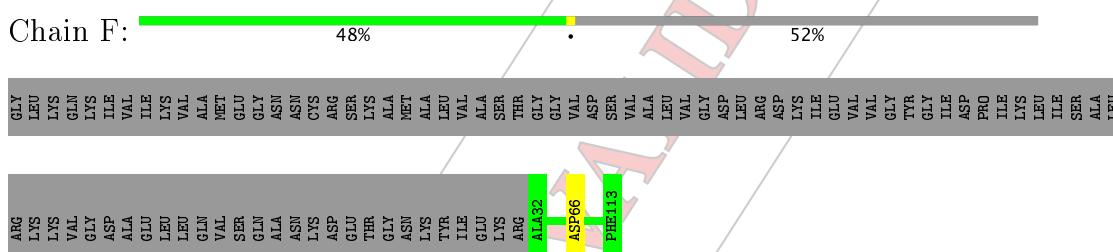
### 3 Residue-property plots [\(i\)](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

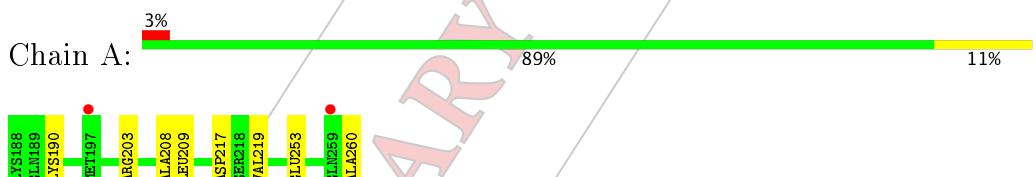
- Molecule 1:



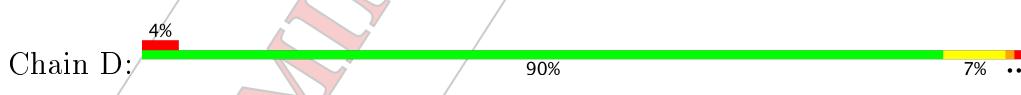
- Molecule 1:



- Molecule 2:



### • Molecule 2:



A horizontal bar composed of four colored segments: red, yellow, yellow, and green. A grey arrow points from the text 'The red segment' to the first red segment.



### • Molecule 3:





- Molecule 3:

Chain E: 4% 88% 12%

Chain E validation bar chart showing 4% in red, 88% in green, and 12% in yellow.



## 4 Data and refinement statistics [\(i\)](#)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	67.20 Å 80.18 Å 105.63 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	105.63 – 1.90 40.09 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.9 (105.63-1.90) 100.0 (40.09-1.90)	Depositor EDS
$R_{\text{merge}}$	(Not available)	Depositor
$R_{\text{sym}}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.73 (at 1.89 Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
$R, R_{\text{free}}$	0.209, 0.224 0.217, 0.229	Depositor DCC
$R_{\text{free}}$ test set	2178 reflections (5.01%)	DCC
Wilson B-factor (Å <sup>2</sup> )	26.2	Xtriage
Anisotropy	0.878	Xtriage
Bulk solvent $k_{\text{sol}}$ (e/Å <sup>3</sup> ), $B_{\text{sol}}$ (Å <sup>2</sup> )	0.33, 49.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49, \langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3712	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 39.14 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 3.3163e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle, \langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [\(i\)](#)

### 5.1 Standard geometry [\(i\)](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	C	0.72	0/689	0.84	1/933 (0.1%)
1	F	0.67	0/689	0.80	0/933
2	A	0.76	0/521	0.90	0/697
2	D	0.72	1/521 (0.2%)	0.87	1/697 (0.1%)
3	B	0.72	0/561	0.86	1/750 (0.1%)
3	E	0.70	0/558	0.87	1/746 (0.1%)
All	All	0.71	1/3539 (0.0%)	0.85	4/4756 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	230	GLU	CD-OE1	5.27	1.31	1.25

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	203	ARG	NE-CZ-NH1	6.05	123.32	120.30
3	B	224	ASP	CB-CG-OD1	5.52	123.27	118.30
1	C	40	ASP	CB-CG-OD1	5.12	122.91	118.30
3	E	224	ASP	CB-CG-OD1	5.11	122.90	118.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [\(i\)](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	665	0	623	1	0
1	F	665	0	623	1	0
2	A	518	0	574	8	0
2	D	518	0	574	8	0
3	B	561	0	609	9	0
3	E	555	0	606	9	0
4	H	230	0	0	2	0
All	All	3712	0	3609	25	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 4.

All (25) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:253:GLU:HG2	4:H:216:HOH:O	2.02	0.58
2:A:190:LYS:HB2	2:A:260:ALA:HB2	1.87	0.56
2:D:219[A]:VAL:CG1	3:E:219:VAL:HG22	2.38	0.54
2:A:219[A]:VAL:CG2	3:B:219:VAL:HG22	2.40	0.52
1:F:66:ASP:OD2	3:E:228:LYS:HE3	2.10	0.52
2:D:230:GLU:OE2	4:H:207:HOH:O	2.19	0.49
2:D:190:LYS:HB2	2:D:260:ALA:HB2	1.95	0.48
3:B:195:VAL:HG23	3:B:197:MET:CE	2.44	0.48
2:D:219[A]:VAL:HG12	3:E:219:VAL:CG2	2.44	0.48
2:D:217:ASP:OD1	3:E:226:ARG:NH2	2.43	0.48
2:A:219[A]:VAL:HG22	3:B:219:VAL:CG2	2.44	0.47
2:A:219[A]:VAL:HG22	3:B:219:VAL:HG22	1.96	0.46
3:B:195:VAL:HG23	3:B:197:MET:SD	2.56	0.45
3:E:195:VAL:HG23	3:E:197:MET:SD	2.56	0.45
2:A:217:ASP:OD1	3:B:226:ARG:NH2	2.47	0.45
3:B:209:LEU:C	3:B:209:LEU:HD23	2.37	0.44
2:A:209:LEU:C	2:A:209:LEU:HD23	2.39	0.43
2:D:209:LEU:C	2:D:209:LEU:HD23	2.39	0.43
3:E:225:LEU:N	3:E:225:LEU:HD12	2.34	0.43
2:D:219[A]:VAL:HG12	3:E:219:VAL:HG22	2.00	0.42
3:E:209:LEU:HD23	3:E:209:LEU:C	2.40	0.42
2:A:208:ALA:HB2	3:B:212:SER:HB3	2.03	0.41
3:E:228:LYS:HE2	3:E:255:LEU:CD1	2.51	0.40
3:B:195:VAL:HG23	3:B:197:MET:HE2	2.03	0.40
1:C:98:ASP:OD2	2:D:203:ARG:HB2	2.21	0.40

There are no symmetry-related clashes.

### 5.3 Torsion angles [\(i\)](#)

#### 5.3.1 Protein backbone [\(i\)](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	C	80/170 (47%)	79 (99%)	1 (1%)	0	100 100
1	F	80/170 (47%)	79 (99%)	1 (1%)	0	100 100
2	A	67/70 (96%)	66 (98%)	1 (2%)	0	100 100
2	D	67/70 (96%)	66 (98%)	1 (2%)	0	100 100
3	B	72/73 (99%)	71 (99%)	1 (1%)	0	100 100
3	E	72/73 (99%)	70 (97%)	2 (3%)	0	100 100
All	All	438/626 (70%)	431 (98%)	7 (2%)	0	100 100

There are no Ramachandran outliers to report.

#### 5.3.2 Protein sidechains [\(i\)](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	C	73/144 (51%)	72 (99%)	1 (1%)	71 69
1	F	73/144 (51%)	73 (100%)	0	100 100
2	A	57/56 (102%)	56 (98%)	1 (2%)	64 60
2	D	57/56 (102%)	56 (98%)	1 (2%)	64 60
3	B	62/59 (105%)	62 (100%)	0	100 100
3	E	62/59 (105%)	62 (100%)	0	100 100
All	All	384/518 (74%)	381 (99%)	3 (1%)	85 85

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	75	LYS
2	A	203	ARG
2	D	203	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA [\(i\)](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [\(i\)](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates [\(i\)](#)

There are no carbohydrates in this entry.

### 5.6 Ligand geometry [\(i\)](#)

There are no ligands in this entry.

### 5.7 Other polymers [\(i\)](#)

There are no such residues in this entry.

### 5.8 Polymer linkage issues [\(i\)](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
3	B	1
2	A	1
2	D	1
3	E	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	197:MET	C	201:ASN	N	7.86
1	E	197:MET	C	201:ASN	N	7.57
1	A	197:MET	C	201:ASN	N	7.56
1	D	197:MET	C	201:ASN	N	7.24

## 6 Fit of model and data (i)

### 6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	C	82/170 (48%)	-0.28	1 (1%) 79 82	21, 28, 44, 56	0
1	F	82/170 (48%)	-0.12	0 100 100	24, 35, 54, 64	0
2	A	70/70 (100%)	-0.02	2 (2%) 52 56	25, 34, 65, 76	0
2	D	70/70 (100%)	0.13	3 (4%) 36 39	24, 37, 64, 70	0
3	B	73/73 (100%)	0.07	4 (5%) 26 29	23, 32, 62, 78	4 (5%)
3	E	73/73 (100%)	0.43	3 (4%) 38 42	24, 30, 56, 87	5 (6%)
All	All	450/626 (71%)	0.03	13 (2%) 52 56	21, 33, 62, 87	9 (2%)

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	B	260[A]	ALA	4.4
2	D	234	TYR	4.2
2	D	260	ALA	4.0
3	B	263[A]	ASP	3.1
2	D	259	GLN	2.9
3	E	197	MET	2.9
2	A	259	GLN	2.8
2	A	197	MET	2.8
3	E	258[A]	SER	2.6
3	E	263[A]	ASP	2.5
3	B	258[A]	SER	2.4
3	B	197	MET	2.3
1	C	32	ALA	2.1

### 6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [\(i\)](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [\(i\)](#)

There are no ligands in this entry.

### 6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.



# Preliminary Full wwPDB X-ray Structure Validation Report (i)

Feb 8, 2018 – 04:58 AM EST

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the standalone wwPDB validation server.  
**The structure in question has not been deposited to the wwPDB.**  
**This report should not be submitted to journals.**

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030736

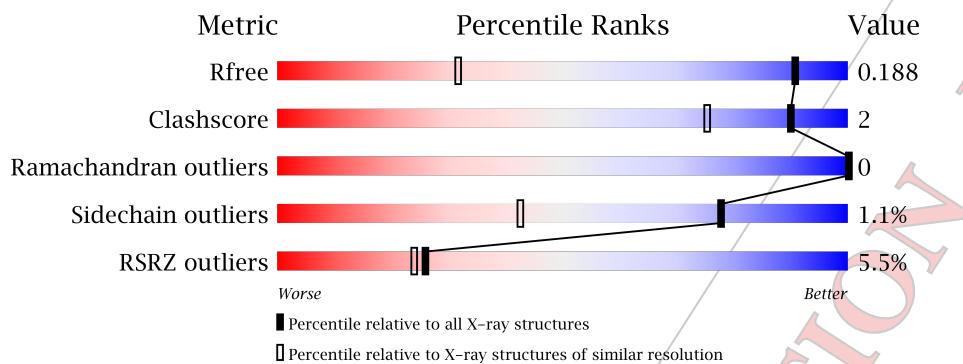
## 1 Overall quality at a glance [\(i\)](#)

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.20 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	1131 (1.24-1.16)
Clashscore	112137	1201 (1.24-1.16)
Ramachandran outliers	110173	1148 (1.24-1.16)
Sidechain outliers	110143	1147 (1.24-1.16)
RSRZ outliers	101464	1132 (1.24-1.16)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ ,  $2$ ,  $1$  and  $0$  types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.



## 2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2919 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
1	A	79	591	371	101	115	4	0	2
1	C	79	609	383	106	116	4	0	5

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	B	84	703	454	122	121	6	0	3
2	D	87	725	468	125	126	6	0	2

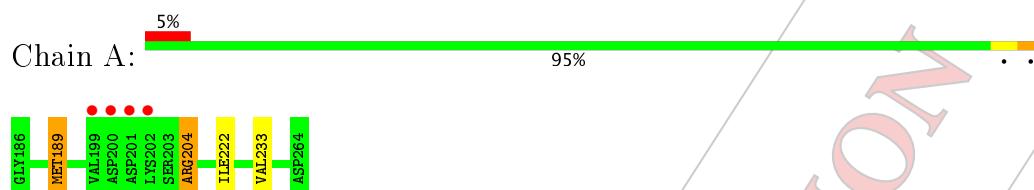
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	E	288	291	291	0	3

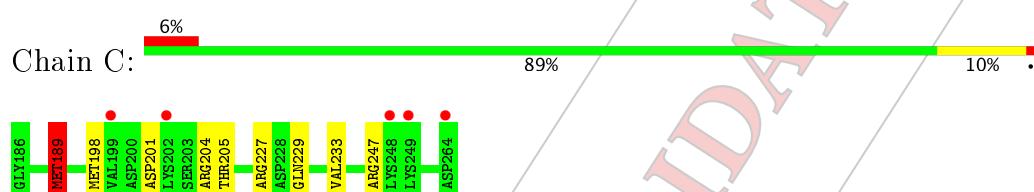
### 3 Residue-property plots [\(i\)](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

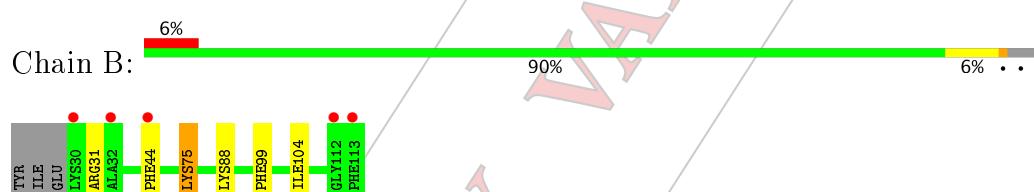
- Molecule 1:



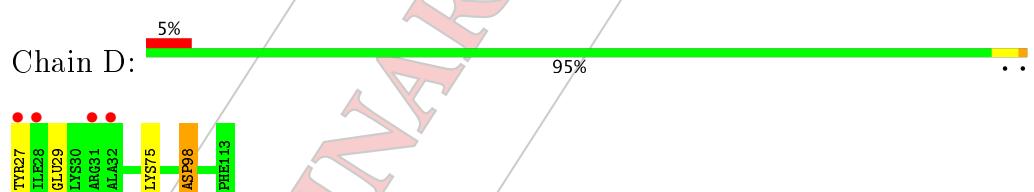
- Molecule 1:



- Molecule 2:



- Molecule 2:



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	37.09 Å 87.13 Å 103.84 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.04 – 1.20 28.52 – 1.20	Depositor EDS
% Data completeness (in resolution range)	98.3 (29.04-1.20) 98.3 (28.52-1.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.88 (at 1.20 Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
$R$ , $R_{free}$	0.148, 0.184 0.158, 0.188	Depositor DCC
$R_{free}$ test set	5033 reflections (5.07%)	DCC
Wilson B-factor (Å <sup>2</sup> )	16.8	Xtriage
Anisotropy	0.180	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34, 39.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	2919	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.10% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [\(i\)](#)

### 5.1 Standard geometry [\(i\)](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	A	0.95	0/601	0.99	2/806 (0.2%)
1	C	0.94	0/629	1.12	7/843 (0.8%)
2	B	0.90	0/737	0.97	4/995 (0.4%)
2	D	0.94	0/757	0.98	1/1022 (0.1%)
All	All	0.93	0/2724	1.01	14/3666 (0.4%)

There are no bond length outliers.

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	247[A]	ARG	NE-CZ-NH2	-8.05	116.28	120.30
1	C	247[B]	ARG	NE-CZ-NH2	-8.05	116.28	120.30
1	C	229	GLN	CA-CB-CG	6.35	127.38	113.40
1	A	189	MET	CG-SD-CE	-6.07	90.49	100.20
1	C	204	ARG	NE-CZ-NH2	-5.89	117.36	120.30
1	A	204	ARG	NE-CZ-NH2	-5.83	117.39	120.30
2	B	31	ARG	NE-CZ-NH1	5.73	123.16	120.30
1	C	189	MET	CG-SD-CE	-5.69	91.09	100.20
2	B	75[A]	LYS	CD-CE-NZ	5.67	124.74	111.70
2	B	75[B]	LYS	CD-CE-NZ	5.67	124.74	111.70
1	C	198	MET	CG-SD-CE	-5.60	91.24	100.20
1	C	227	ARG	NE-CZ-NH1	-5.52	117.54	120.30
2	B	88	LYS	CD-CE-NZ	5.35	124.01	111.70
2	D	98	ASP	CB-CG-OD2	-5.25	113.57	118.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [\(i\)](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen

atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	591	0	626	3	0
1	C	609	0	651	4	0
2	B	703	0	678	3	0
2	D	725	0	694	3	1
3	E	291	0	0	2	1
All	All	2919	0	2649	11	1

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 2.

All (11) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:44:PHE:CD1	1:C:201:ASP:HB3	2.44	0.53
1:C:189:MET:HE3	1:C:233:VAL:CG1	2.38	0.53
1:C:205:THR:HG22	3:E:254:HOH:O	2.11	0.50
2:B:44:PHE:CE1	1:C:201:ASP:HB3	2.47	0.50
2:B:99:PHE:CD2	2:B:104[B]:ILE:HD12	2.49	0.48
2:D:98:ASP:HB3	3:E:231:HOH:O	2.16	0.46
2:D:27:TYR:CZ	2:D:29:GLU:HB2	2.53	0.43
2:D:27:TYR:OH	2:D:29:GLU:HB2	2.20	0.42
1:A:204:ARG:HB3	1:A:222:ILE:HD13	2.01	0.42
1:A:189:MET:HE2	1:A:233:VAL:CG1	2.50	0.42
1:A:204:ARG:HB3	1:A:222:ILE:CD1	2.50	0.41

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:D:27:TYR:N	3:E:245:HOH:O[3_644]	2.12	0.08

### 5.3 Torsion angles [\(i\)](#)

#### 5.3.1 Protein backbone [\(i\)](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	79/79 (100%)	78 (99%)	1 (1%)	0	100	100
1	C	82/79 (104%)	82 (100%)	0	0	100	100
2	B	85/87 (98%)	84 (99%)	1 (1%)	0	100	100
2	D	87/87 (100%)	86 (99%)	1 (1%)	0	100	100
All	All	333/332 (100%)	330 (99%)	3 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains [\(i\)](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	68/66 (103%)	68 (100%)	0	100	100
1	C	71/66 (108%)	70 (99%)	1 (1%)	71	33
2	B	78/78 (100%)	76 (97%)	2 (3%)	51	11
2	D	80/78 (103%)	78 (98%)	2 (2%)	53	12
All	All	297/288 (103%)	292 (98%)	5 (2%)	78	26

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	75[A]	LYS
2	B	75[B]	LYS
1	C	189	MET
2	D	75[A]	LYS
2	D	75[B]	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	241	ASN
2	B	83(C)	ASN
1	C	229	GLN
2	D	62	ASN
2	D	83(C)	ASN

### 5.3.3 RNA [\(i\)](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [\(i\)](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates [\(i\)](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [\(i\)](#)

There are no ligands in this entry.

## 5.7 Other polymers [\(i\)](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [\(i\)](#)

There are no chain breaks in this entry.

## 6 Fit of model and data (i)

### 6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	79/79 (100%)	0.34	4 (5%) 29 27	13, 19, 39, 78	0
1	C	79/79 (100%)	0.50	5 (6%) 21 20	15, 25, 49, 59	0
2	B	84/87 (96%)	0.23	5 (5%) 23 22	13, 19, 32, 75	0
2	D	87/87 (100%)	0.22	4 (4%) 33 32	12, 18, 33, 45	0
All	All	329/332 (99%)	0.32	18 (5%) 26 24	12, 20, 40, 78	0

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	199	VAL	11.9
2	D	27	TYR	9.0
2	B	30	LYS	7.5
2	B	113	PHE	5.5
1	C	264	ASP	4.4
1	C	199	VAL	4.4
1	A	201	ASP	4.0
2	D	31	ARG	3.9
1	C	202	LYS	3.7
1	C	248	LYS	3.0
1	A	200	ASP	2.7
2	D	28	ILE	2.7
2	B	44	PHE	2.5
2	B	32	ALA	2.4
1	A	202	LYS	2.4
1	C	249	LYS	2.3
2	D	32	ALA	2.2
2	B	112	GLY	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains [\(i\)](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates [\(i\)](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [\(i\)](#)

There are no ligands in this entry.

## 6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.



# Preliminary Full wwPDB X-ray Structure Validation Report (i)

Feb 8, 2018 – 05:57 AM EST

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the standalone wwPDB validation server.  
**The structure in question has not been deposited to the wwPDB.**  
**This report should not be submitted to journals.**

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see [references \(1\)](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriaage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030736

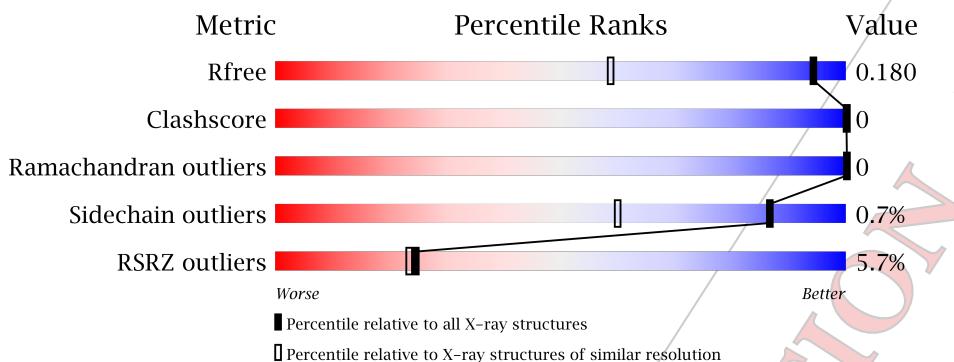
## 1 Overall quality at a glance [\(i\)](#)

The following experimental techniques were used to determine the structure:

### X-RAY DIFFRACTION

The reported resolution of this entry is 1.30 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	1131 (1.32-1.28)
Clashscore	112137	1185 (1.32-1.28)
Ramachandran outliers	110173	1138 (1.32-1.28)
Sidechain outliers	110143	1138 (1.32-1.28)
RSRZ outliers	101464	1133 (1.32-1.28)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ ,  $2$ ,  $1$  and  $0$  types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain	
1	B	93	5%	95% 5%
2	A	81	6%	98% :

## 2 Entry composition [\(i\)](#)

There are 3 unique types of molecules in this entry. The entry contains 1528 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	93	769	491	134	138	6	0	2	0

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	81	615	388	106	116	5	0	4	0

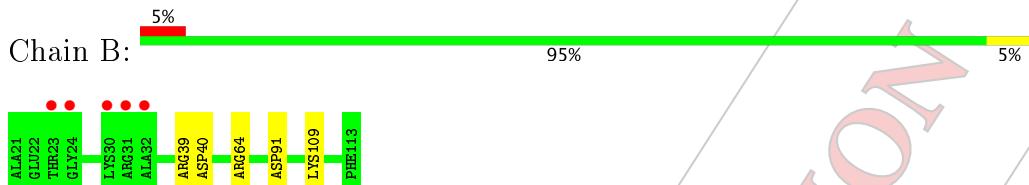
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	D	144	144	144	0	0

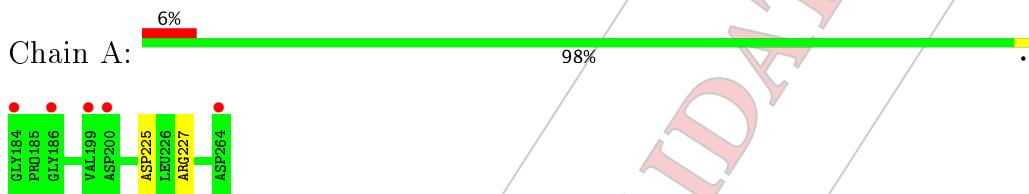
### 3 Residue-property plots [\(i\)](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1:



- Molecule 2:



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	30.02 Å 54.34 Å 48.49 Å 90.00° 90.53° 90.00°	Depositor
Resolution (Å)	54.35 – 1.30 23.70 – 1.30	Depositor EDS
% Data completeness (in resolution range)	99.3 (54.35-1.30) 99.4 (23.70-1.30)	Depositor EDS
$R_{\text{merge}}$	(Not available)	Depositor
$R_{\text{sym}}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle^1$	2.00 (at 1.30 Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
$R, R_{\text{free}}$	0.141, 0.178 0.141, 0.180	Depositor DCC
$R_{\text{free}}$ test set	1959 reflections (5.41%)	DCC
Wilson B-factor (Å <sup>2</sup> )	13.1	Xtriage
Anisotropy	0.807	Xtriage
Bulk solvent $k_{\text{sol}}$ (e/Å <sup>3</sup> ), $B_{\text{sol}}$ (Å <sup>2</sup> )	0.36, 43.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50, \langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.038 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	1528	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.30% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [\(i\)](#)

### 5.1 Standard geometry [\(i\)](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# $ Z  > 5$	RMSZ	# $ Z  > 5$
1	B	1.04	0/800	1.14	8/1080 (0.7%)
2	A	0.72	0/634	1.02	3/850 (0.4%)
All	All	0.91	0/1434	1.09	11/1930 (0.6%)

There are no bond length outliers.

All (11) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	64	ARG	NE-CZ-NH1	-9.41	115.59	120.30
1	B	64	ARG	NH1-CZ-NH2	7.28	127.41	119.40
1	B	91	ASP	CB-CG-OD2	-6.89	112.10	118.30
2	A	227[A]	ARG	NE-CZ-NH1	6.84	123.72	120.30
2	A	227[B]	ARG	NE-CZ-NH1	6.84	123.72	120.30
1	B	40	ASP	CB-CG-OD2	-6.81	112.17	118.30
1	B	64	ARG	NE-CZ-NH2	-6.68	116.96	120.30
1	B	91	ASP	CB-CG-OD1	5.67	123.40	118.30
2	A	225	ASP	CB-CG-OD1	5.44	123.20	118.30
1	B	39[A]	ARG	NE-CZ-NH2	-5.14	117.73	120.30
1	B	39[B]	ARG	NE-CZ-NH2	-5.14	117.73	120.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [\(i\)](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	769	0	732	0	0
2	A	615	0	660	0	0
3	D	144	0	0	0	0
All	All	1528	0	1392	0	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 0.

There are no clashes within the asymmetric unit.

There are no symmetry-related clashes.

### 5.3 Torsion angles [\(i\)](#)

#### 5.3.1 Protein backbone [\(i\)](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	93/93 (100%)	91 (98%)	2 (2%)	0	100 100
2	A	83/81 (102%)	82 (99%)	1 (1%)	0	100 100
All	All	176/174 (101%)	173 (98%)	3 (2%)	0	100 100

There are no Ramachandran outliers to report.

#### 5.3.2 Protein sidechains [\(i\)](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	B	84/82 (102%)	83 (99%)	1 (1%)	75 39
2	A	71/67 (106%)	71 (100%)	0	100 100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	155/149 (104%)	154 (99%)	1 (1%)	87 64

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	109	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	A	260	GLN

### 5.3.3 RNA [\(i\)](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [\(i\)](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates [\(i\)](#)

There are no carbohydrates in this entry.

### 5.6 Ligand geometry [\(i\)](#)

There are no ligands in this entry.

### 5.7 Other polymers [\(i\)](#)

There are no such residues in this entry.

### 5.8 Polymer linkage issues [\(i\)](#)

There are no chain breaks in this entry.

## 6 Fit of model and data (i)

### 6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	B	93/93 (100%)	0.15	5 (5%) 26 26	10, 15, 34, 69	0
2	A	81/81 (100%)	0.10	5 (6%) 21 21	12, 23, 37, 43	0
All	All	174/174 (100%)	0.13	10 (5%) 24 23	10, 19, 37, 69	0

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	31	ARG	6.5
2	A	199	VAL	5.8
1	B	32	ALA	5.8
1	B	23	THR	4.3
2	A	184	GLY	3.3
1	B	24	GLY	3.1
2	A	264	ASP	2.5
2	A	186	GLY	2.2
1	B	30	LYS	2.2
2	A	200	ASP	2.1

### 6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

### 6.4 Ligands (i)

There are no ligands in this entry.

## 6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.

PRELIMINARY VALIDATION REPORT



# Preliminary Full wwPDB X-ray Structure Validation Report (i)

Feb 8, 2018 – 05:28 AM EST

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the standalone wwPDB validation server.  
**The structure in question has not been deposited to the wwPDB.**  
**This report should not be submitted to journals.**

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see [references \(1\)](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriaage (Phenix)	:	1.9-1692
EDS	:	rb-20030736
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030736

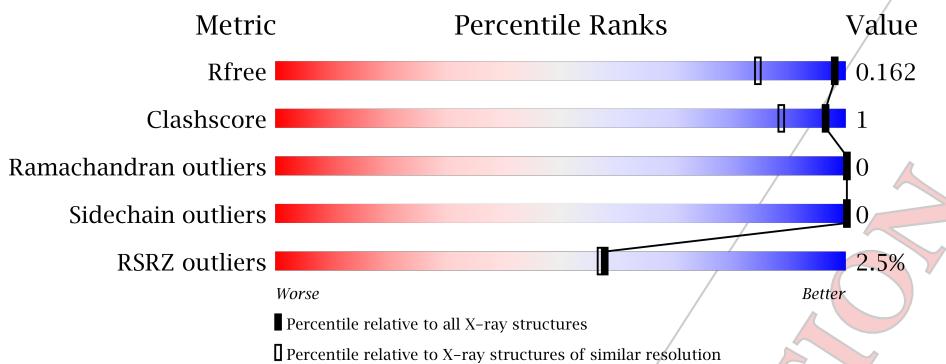
## 1 Overall quality at a glance [\(i\)](#)

The following experimental techniques were used to determine the structure:

### X-RAY DIFFRACTION

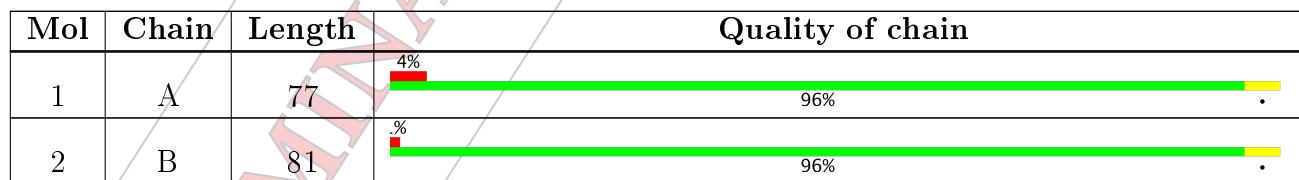
The reported resolution of this entry is 1.30 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	1131 (1.32-1.28)
Clashscore	112137	1185 (1.32-1.28)
Ramachandran outliers	110173	1138 (1.32-1.28)
Sidechain outliers	110143	1138 (1.32-1.28)
RSRZ outliers	101464	1133 (1.32-1.28)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.



## 2 Entry composition [\(i\)](#)

There are 3 unique types of molecules in this entry. The entry contains 1436 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	77	596	377	102	113	4	0	4	0

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	81	696	450	117	123	6	0	6	0

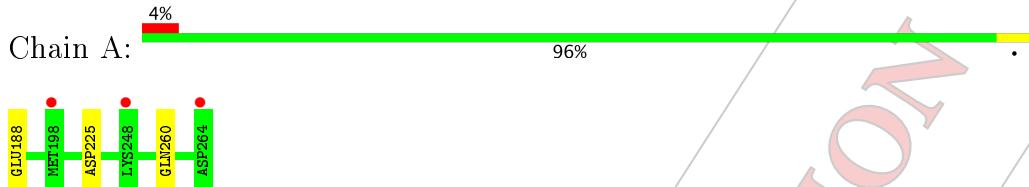
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	C	144	144	144	0	0

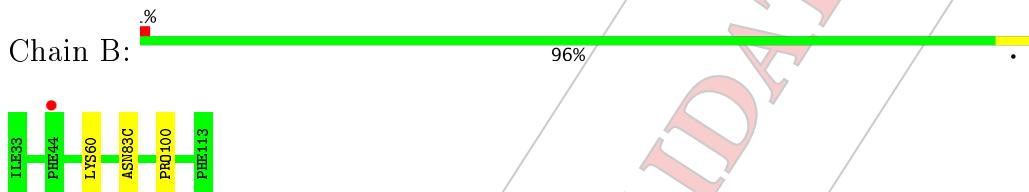
### 3 Residue-property plots [\(i\)](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1:



- Molecule 2:



## 4 Data and refinement statistics <sup>i</sup>

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	30.07 $\text{\AA}$ 54.44 $\text{\AA}$ 46.64 $\text{\AA}$ 90.00° 104.43° 90.00°	Depositor
Resolution ( $\text{\AA}$ )	34.76 – 1.30 34.76 – 1.30	Depositor EDS
% Data completeness (in resolution range)	99.2 (34.76-1.30) 99.2 (34.76-1.30)	Depositor EDS
$R_{\text{merge}}$	(Not available)	Depositor
$R_{\text{sym}}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.71 (at 1.30 $\text{\AA}$ )	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
$R, R_{\text{free}}$	0.125, 0.162 0.126, 0.162	Depositor DCC
$R_{\text{free}}$ test set	1720 reflections (5.08%)	DCC
Wilson B-factor ( $\text{\AA}^2$ )	12.9	Xtriage
Anisotropy	0.662	Xtriage
Bulk solvent $k_{\text{sol}}$ (e/ $\text{\AA}^3$ ), $B_{\text{sol}}$ ( $\text{\AA}^2$ )	0.34, 46.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49, \langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.98	EDS
Total number of atoms	1436	wwPDB-VP
Average B, all atoms ( $\text{\AA}^2$ )	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.07% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [\(i\)](#)

### 5.1 Standard geometry [\(i\)](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# $ Z  > 5$	RMSZ	# $ Z  > 5$
1	A	0.76	0/612	0.94	1/821 (0.1%)
2	B	0.82	0/737	0.83	0/996
All	All	0.79	0/1349	0.88	1/1817 (0.1%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	225	ASP	CB-CG-OD2	-5.42	113.42	118.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [\(i\)](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	596	0	644	1	0
2	B	696	0	672	2	0
3	C	144	0	0	1	1
All	All	1436	0	1316	3	1

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 1.

All (3) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:60[B]:LYS:NZ	3:C:154:HOH:O	1.98	0.96
1:A:188:GLU:C	1:A:260:GLN:HE22	2.17	0.48
2:B:83(C):ASN:HD22	2:B:100:PRO:HB3	1.85	0.41

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:C:93:HOH:O	3:C:100:HOH:O[2_747]	2.02	0.18

### 5.3 Torsion angles [\(i\)](#)

#### 5.3.1 Protein backbone [\(i\)](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	79/77 (103%)	79 (100%)	0	0	100 100
2	B	84/81 (104%)	83 (99%)	1 (1%)	0	100 100
All	All	163/158 (103%)	162 (99%)	1 (1%)	0	100 100

There are no Ramachandran outliers to report.

#### 5.3.2 Protein sidechains [\(i\)](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	70/66 (106%)	70 (100%)	0	100 100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	B	79/73 (108%)	79 (100%)	0	100 100
All	All	149/139 (107%)	149 (100%)	0	100 100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	260	GLN
2	B	83(C)	ASN

### 5.3.3 RNA [\(i\)](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [\(i\)](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates [\(i\)](#)

There are no carbohydrates in this entry.

### 5.6 Ligand geometry [\(i\)](#)

There are no ligands in this entry.

### 5.7 Other polymers [\(i\)](#)

There are no such residues in this entry.

### 5.8 Polymer linkage issues [\(i\)](#)

There are no chain breaks in this entry.

## 6 Fit of model and data [\(i\)](#)

### 6.1 Protein, DNA and RNA chains [\(i\)](#)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	77/77 (100%)	-0.11	3 (3%)	40, 38, 14, 26, 38, 57	0
2	B	81/81 (100%)	-0.34	1 (1%)	79, 79, 11, 15, 31, 39	0
All	All	158/158 (100%)	-0.23	4 (2%)	58, 57, 11, 19, 36, 57	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	44	PHE	3.6
1	A	198	MET	3.0
1	A	264	ASP	2.7
1	A	248	LYS	2.0

### 6.2 Non-standard residues in protein, DNA, RNA chains [\(i\)](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [\(i\)](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [\(i\)](#)

There are no ligands in this entry.

### 6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.



# Preliminary Full wwPDB X-ray Structure Validation Report (i)

Jan 30, 2019 – 05:21 AM EST

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the standalone wwPDB validation server.  
**The structure in question has not been deposited to the wwPDB.**  
**This report should not be submitted to journals.**

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	rb-20031633
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20031633

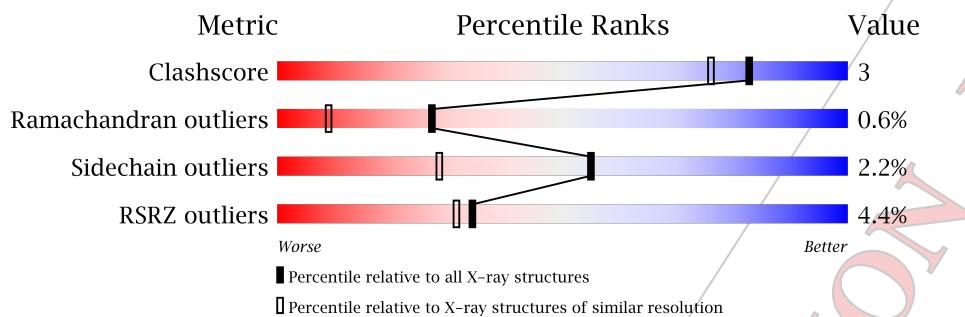
## 1 Overall quality at a glance [\(i\)](#)

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

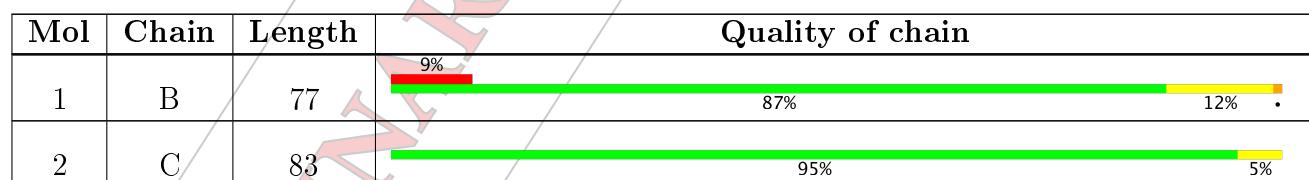
The reported resolution of this entry is 1.60 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	122126	3202 (1.60-1.60)
Ramachandran outliers	120053	3117 (1.60-1.60)
Sidechain outliers	120020	3116 (1.60-1.60)
RSRZ outliers	108989	2883 (1.60-1.60)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.



## 2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1403 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	77	573	359	100	110	4	0	1	0

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	83	704	452	125	121	6	0	5	0

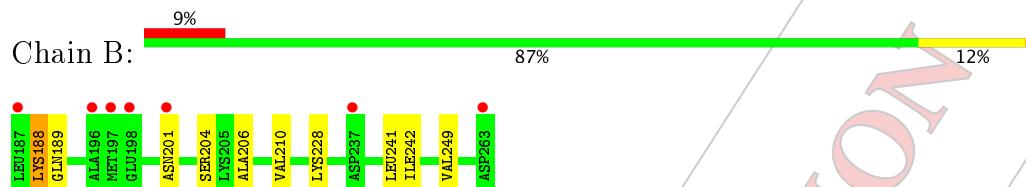
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	D	126	126	126	0	0

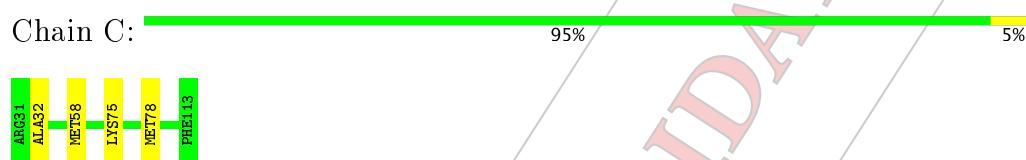
### 3 Residue-property plots [\(i\)](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1:



- Molecule 2:



## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	29.79 Å 65.33 Å 75.86 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.82 – 1.60 32.80 – 1.60	Depositor EDS
% Data completeness (in resolution range)	99.9 (32.82-1.60) 100.0 (32.80-1.60)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.75 (at 1.60 Å)	Xtriage
Refinement program	REFMAC 5.8.0232	Depositor
$R$ , $R_{free}$	0.197, 0.232 0.208, (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	21.3	Xtriage
Anisotropy	0.302	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34, 42.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o$ , $F_c$ correlation	0.96	EDS
Total number of atoms	1403	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.51% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [\(i\)](#)

### 5.1 Standard geometry [\(i\)](#)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	B	0.70	0/577	0.84	0/771
2	C	0.65	0/735	0.79	0/994
All	All	0.67	0/1312	0.81	0/1765

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [\(i\)](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	573	0	620	6	0
2	C	704	0	671	1	0
3	D	126	0	0	3	0
All	All	1403	0	1291	7	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 3.

All (7) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:189:GLN:NE2	3:D:105:HOH:O	2.22	0.72

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:210:VAL:HG22	1:B:241:LEU:HD11	1.88	0.56
1:B:228:LYS:HD3	3:D:141:HOH:O	2.09	0.52
2:C:58:MET:CE	2:C:78:MET:SD	3.00	0.50
1:B:188:LYS:HE3	3:D:35:HOH:O	2.12	0.49
1:B:206:ALA:HB2	1:B:249:VAL:HG11	1.96	0.47
1:B:201:ASN:O	1:B:204:SER:HB2	2.19	0.41

There are no symmetry-related clashes.

### 5.3 Torsion angles [\(i\)](#)

#### 5.3.1 Protein backbone [\(i\)](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	76/77 (99%)	74 (97%)	2 (3%)	0	100 100
2	C	85/83 (102%)	82 (96%)	1 (1%)	2 (2%)	6 1
All	All	161/160 (101%)	156 (97%)	3 (2%)	2 (1%)	27 3

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	32[A]	ALA
2	C	32[B]	ALA

#### 5.3.2 Protein sidechains [\(i\)](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	B	63/62 (102%)	61 (97%)	2 (3%)	42 16
2	C	77/74 (104%)	76 (99%)	1 (1%)	71 53
All	All	140/136 (103%)	137 (98%)	3 (2%)	55 30

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	188	LYS
1	B	242	ILE
2	C	75	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	61	ASN

### 5.3.3 RNA [\(i\)](#)

There are no RNA molecules in this entry.

### 5.4 Non-standard residues in protein, DNA, RNA chains [\(i\)](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates [\(i\)](#)

There are no carbohydrates in this entry.

### 5.6 Ligand geometry [\(i\)](#)

There are no ligands in this entry.

### 5.7 Other polymers [\(i\)](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [\(i\)](#)

There are no chain breaks in this entry.

PRELIMINARY VALIDATION REPORT

## 6 Fit of model and data (i)

### 6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	B	77/77 (100%)	0.46	7 (9%) 9 8	19, 34, 48, 65	0
2	C	83/83 (100%)	-0.29	0 100 100	14, 20, 28, 33	1 (1%)
All	All	160/160 (100%)	0.07	7 (4%) 34 31	14, 23, 45, 65	1 (0%)

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	263	ASP	4.2
1	B	187	LEU	3.5
1	B	198	GLU	3.5
1	B	196	ALA	2.9
1	B	237	ASP	2.8
1	B	201	ASN	2.6
1	B	197	MET	2.6

### 6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

### 6.4 Ligands (i)

There are no ligands in this entry.

## 6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.

PRELIMINARY VALIDATION REPORT



# Preliminary Full wwPDB X-ray Structure Validation Report (i)

Jan 30, 2019 – 05:13 AM EST

This is a Preliminary Full wwPDB X-ray Structure Validation Report.

This report is produced by the standalone wwPDB validation server.  
**The structure in question has not been deposited to the wwPDB.**  
**This report should not be submitted to journals.**

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see [references \(1\)](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix)	:	1.13
EDS	:	rb-20031633
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20031633

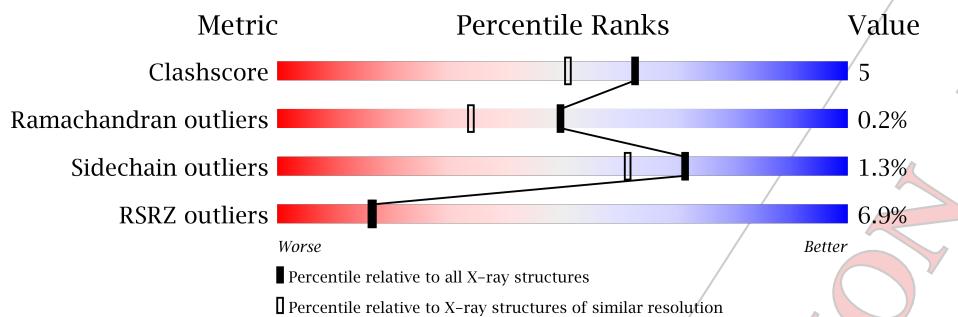
## 1 Overall quality at a glance [\(i\)](#)

The following experimental techniques were used to determine the structure:

### X-RAY DIFFRACTION

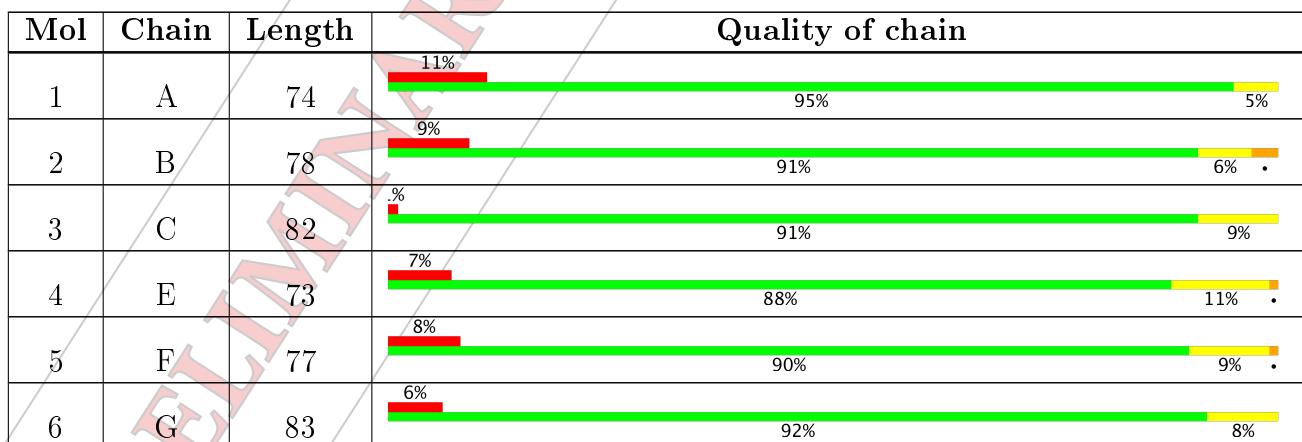
The reported resolution of this entry is 1.85 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive	Similar resolution
	(#Entries)	(#Entries, resolution range(Å))
Clashscore	122126	2258 (1.86-1.86)
Ramachandran outliers	120053	2234 (1.86-1.86)
Sidechain outliers	120020	2234 (1.86-1.86)
RSRZ outliers	108989	2075 (1.86-1.86)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.



The following table lists non-polymeric compounds, carbohydrate monomers and non-standard

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
7	MPD	G	201	X	-	-	-

## 2 Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 3916 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	74	549	348	97	101	3	0	2	0

- Molecule 2 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	78	576	364	99	110	3	0	2	0

- Molecule 3 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	82	676	435	116	119	6	0	2	0

- Molecule 4 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	E	73	549	346	98	102	3	0	2	0

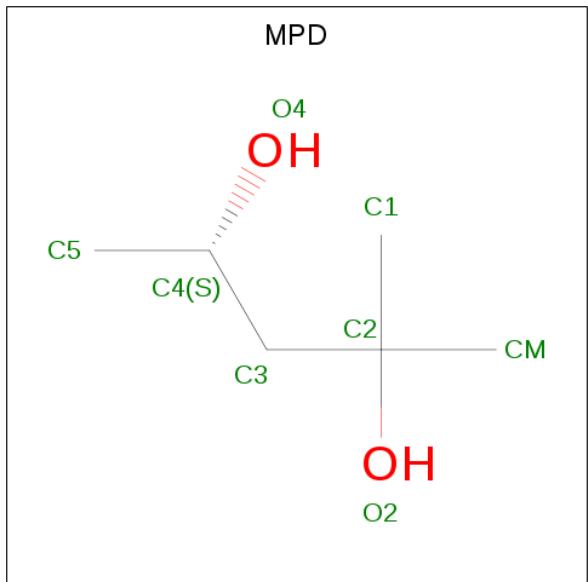
- Molecule 5 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	77	573	364	98	108	3	0	3	0

- Molecule 6 is a protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	G	83	681	436	119	120	6	0	1	0

- Molecule 7 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: C<sub>6</sub>H<sub>14</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	C	1	Total C O 8 6 2	0	0
7	G	1	Total C O 8 6 2	0	0

- Molecule 8 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
8	D	1	Total Cl 1 1	0	0
8	D	1	Total Cl 1 1	0	0

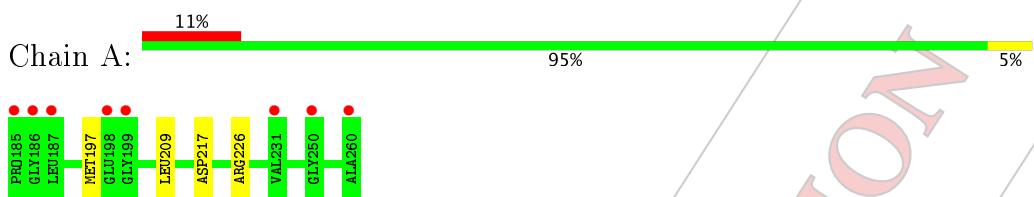
- Molecule 9 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
9	H	294	Total O 294 294	0	0

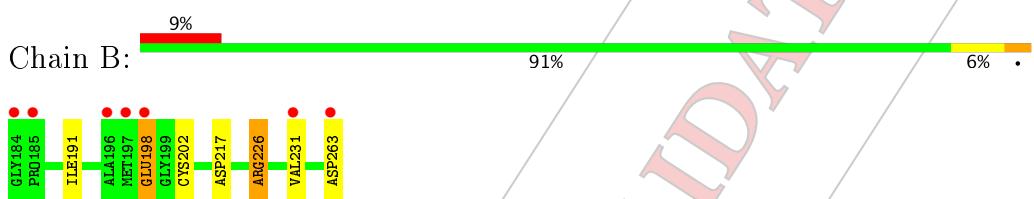
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

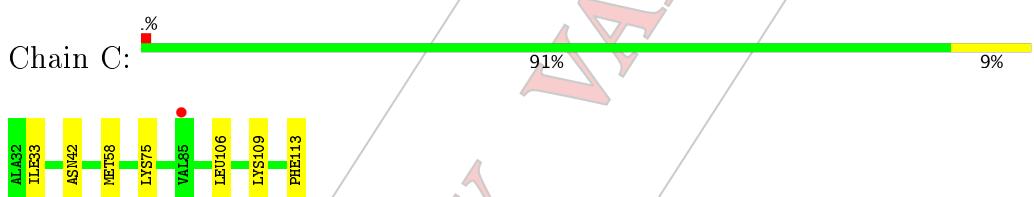
- Molecule 1:



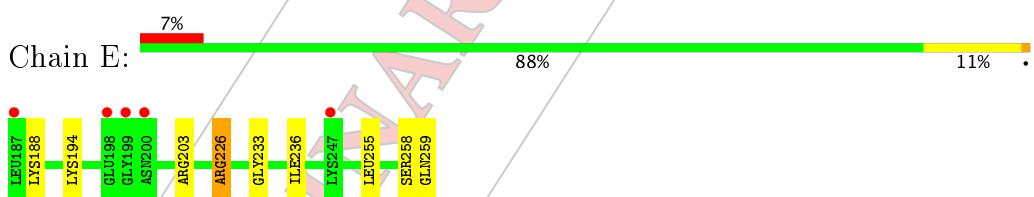
- Molecule 2:



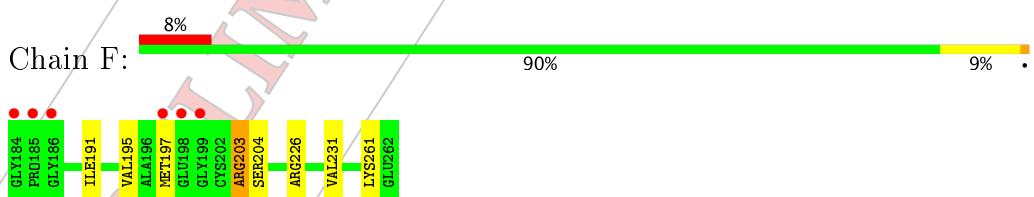
- Molecule 3:



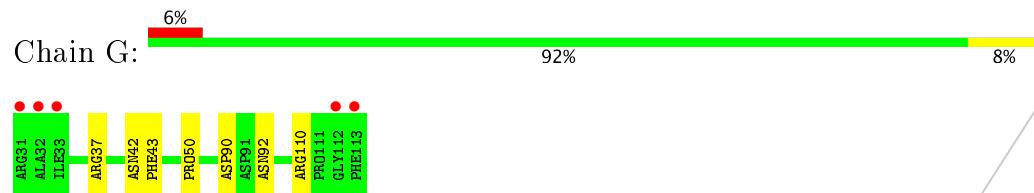
- Molecule 4:



- Molecule 5:



- Molecule 6:



PRELIMINARY VALIDATION REPORT

## 4 Data and refinement statistics [\(i\)](#)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	66.46 Å 80.70 Å 105.58 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.52 – 1.85 29.50 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.8 (29.52-1.85) 99.8 (29.50-1.85)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.15 (at 1.85 Å)	Xtriage
Refinement program	REFMAC 5.8.0232	Depositor
$R$ , $R_{free}$	0.186, 0.230 0.197, (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.4	Xtriage
Anisotropy	0.680	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34, 43.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	3916	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 30.25 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.3492e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [\(i\)](#)

### 5.1 Standard geometry [\(i\)](#)

Bond lengths and bond angles in the following residue types are not validated in this section: MPD, CL

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# $ Z  > 5$	RMSZ	# $ Z  > 5$
1	A	0.63	0/556	0.74	0/742
2	B	0.67	0/583	0.77	0/778
3	C	0.63	0/706	0.76	1/955 (0.1%)
4	E	0.61	0/556	0.80	0/744
5	F	0.71	0/583	0.78	0/779
6	G	0.55	0/708	0.68	0/958
All	All	0.63	0/3692	0.75	1/4956 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
2	B	0	1
4	E	0	2
5	F	0	1
All	All	0	4

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	58	MET	CG-SD-CE	5.10	108.37	100.20

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	226	ARG	Sidechain
4	E	203	ARG	Sidechain
4	E	258	SER	Peptide
5	F	203	ARG	Sidechain

## 5.2 Too-close contacts [\(i\)](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	549	0	608	4	0
2	B	576	0	629	5	0
3	C	676	0	642	3	0
4	E	549	0	605	7	0
5	F	573	0	631	13	0
6	G	681	0	642	9	0
7	C	8	0	14	0	0
7	G	8	0	14	1	0
8	D	2	0	0	0	0
9	H	294	0	0	4	0
All	All	3916	0	3785	35	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 5.

All (35) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:F:195:VAL:CG2	5:F:197:MET:SD	2.65	0.83
5:F:195:VAL:HG23	5:F:197:MET:SD	2.21	0.81
5:F:226:ARG:NH1	6:G:42[B]:ASN:HD21	1.86	0.73
5:F:226:ARG:HD3	9:H:251:HOH:O	1.90	0.70
5:F:197:MET:SD	5:F:203:ARG:HA	2.36	0.66
5:F:226:ARG:CZ	6:G:42[B]:ASN:HD21	2.10	0.64
5:F:195:VAL:HG21	5:F:197:MET:SD	2.38	0.64
5:F:195:VAL:HG23	5:F:197:MET:HG3	1.81	0.63
2:B:198:GLU:HB3	2:B:202:CYS:SG	2.42	0.60
4:E:226[A]:ARG:HG2	4:E:226[A]:ARG:HH21	1.68	0.59

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:F:226:ARG:HG2	6:G:42[B]:ASN:ND2	2.18	0.58
5:F:195:VAL:HG23	5:F:197:MET:CG	2.34	0.58
5:F:191:ILE:HB	5:F:231:VAL:HG13	1.88	0.56
7:G:201:MPD:O2	7:G:201:MPD:O4	2.05	0.56
5:F:261:LYS:O	6:G:50:PRO:HG2	2.05	0.56
6:G:37:ARG:CD	9:H:226:HOH:O	2.54	0.56
4:E:226[B]:ARG:HH21	4:E:226[B]:ARG:HG3	1.73	0.53
4:E:226[A]:ARG:HG2	4:E:226[A]:ARG:NH2	2.25	0.50
1:A:226[B]:ARG:HG3	1:A:226[B]:ARG:HH21	1.77	0.50
2:B:226:ARG:HG2	3:C:42[A]:ASN:ND2	2.28	0.49
4:E:226[B]:ARG:NH2	4:E:226[B]:ARG:HG3	2.30	0.46
5:F:203:ARG:NH2	5:F:226:ARG:O	2.49	0.46
6:G:37:ARG:HD3	9:H:226:HOH:O	2.14	0.46
1:A:226[A]:ARG:NH1	2:B:217:ASP:OD1	2.45	0.45
6:G:37:ARG:HD2	9:H:226:HOH:O	2.14	0.45
3:C:109:LYS:HE2	3:C:113:PHE:OXT	2.16	0.44
1:A:209:LEU:C	1:A:209:LEU:HD23	2.37	0.44
1:A:217:ASP:OD1	2:B:226:ARG:NH2	2.51	0.44
4:E:194:LYS:HB2	4:E:255:LEU:HD11	2.00	0.44
4:E:233:GLY:HA3	4:E:236:ILE:HD11	2.00	0.43
6:G:43:PHE:CE2	6:G:110:ARG:NH2	2.86	0.43
4:E:188:LYS:HG3	4:E:259:GLN:NE2	2.34	0.42
3:C:33:ILE:HG21	3:C:106:LEU:CD1	2.50	0.41
6:G:90:ASP:OD1	6:G:92:ASN:N	2.49	0.41
2:B:191:ILE:HB	2:B:231[A]:VAL:CG1	2.51	0.41

There are no symmetry-related clashes.

### 5.3 Torsion angles

#### 5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	72/74 (97%)	67 (93%)	4 (6%)	1 (1%)	12 3

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
2	B	76/78 (97%)	73 (96%)	3 (4%)	0	100 100
3	C	82/82 (100%)	81 (99%)	1 (1%)	0	100 100
4	E	73/73 (100%)	70 (96%)	3 (4%)	0	100 100
5	F	76/77 (99%)	76 (100%)	0	0	100 100
6	G	82/83 (99%)	81 (99%)	1 (1%)	0	100 100
All	All	461/467 (99%)	448 (97%)	12 (3%)	1 (0%)	49 34

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	197	MET

### 5.3.2 Protein sidechains [\(i\)](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	A	60/58 (103%)	60 (100%)	0	100 100
2	B	63/61 (103%)	61 (97%)	2 (3%)	42 24
3	C	75/73 (103%)	73 (97%)	2 (3%)	48 31
4	E	61/59 (103%)	59 (97%)	2 (3%)	41 22
5	F	63/60 (105%)	61 (97%)	2 (3%)	42 24
6	G	75/74 (101%)	75 (100%)	0	100 100
All	All	397/385 (103%)	389 (98%)	8 (2%)	71 43

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	198	GLU
2	B	263	ASP
3	C	75[A]	LYS
3	C	75[B]	LYS
4	E	226[A]	ARG

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type
4	E	226[B]	ARG
5	F	204[A]	SER
5	F	204[B]	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
4	E	256	GLN
4	E	259	GLN

### 5.3.3 RNA [\(i\)](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [\(i\)](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 5.5 Carbohydrates [\(i\)](#)

There are no carbohydrates in this entry.

### 5.6 Ligand geometry [\(i\)](#)

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
7	MPD	C	201	-	7,7,7	0.34	0	9,10,10	0.79	0
7	MPD	G	201	-	7,7,7	0.49	0	9,10,10	0.56	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral

centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
7	MPD	C	201	-	-	0/5/5/5	0/0/0/0
7	MPD	G	201	-	1/1/2/2	0/5/5/5	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
7	G	201	MPD	C4

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	G	201	MPD	1	0

## 5.7 Other polymers [i](#)

There are no such residues in this entry.

## 5.8 Polymer linkage issues [i](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
2	B	1
1	A	1
5	F	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	F	199:GLY	C	202:CYS	N	4.97
1	B	199:GLY	C	202:CYS	N	3.97
1	A	199:GLY	C	202:CYS	N	3.52

## 6 Fit of model and data (i)

### 6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	74/74 (100%)	0.38	8 (10%) <span style="border: 1px solid red; padding: 2px;">6</span> <span style="border: 1px solid red; padding: 2px;">6</span>	25, 37, 84, 120	0
2	B	78/78 (100%)	0.44	7 (8%) <span style="border: 1px solid red; padding: 2px;">9</span> <span style="border: 1px solid red; padding: 2px;">9</span>	24, 35, 73, 115	0
3	C	82/82 (100%)	-0.19	1 (1%) <span style="border: 1px solid blue; padding: 2px;">79</span> <span style="border: 1px solid blue; padding: 2px;">79</span>	22, 29, 43, 53	0
4	E	73/73 (100%)	0.17	5 (6%) <span style="border: 1px solid red; padding: 2px;">17</span> <span style="border: 1px solid red; padding: 2px;">17</span>	28, 39, 67, 81	0
5	F	77/77 (100%)	0.31	6 (7%) <span style="border: 1px solid red; padding: 2px;">13</span> <span style="border: 1px solid red; padding: 2px;">13</span>	26, 33, 73, 109	0
6	G	83/83 (100%)	0.08	5 (6%) <span style="border: 1px solid red; padding: 2px;">22</span> <span style="border: 1px solid red; padding: 2px;">21</span>	30, 41, 64, 98	0
All	All	467/467 (100%)	0.19	32 (6%) <span style="border: 1px solid red; padding: 2px;">17</span> <span style="border: 1px solid red; padding: 2px;">16</span>	22, 36, 69, 120	0

All (32) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	185	PRO	7.5
2	B	263	ASP	7.1
4	E	200	ASN	5.5
5	F	198	GLU	5.3
2	B	198	GLU	5.0
6	G	31	ARG	5.0
1	A	187	LEU	4.9
5	F	185	PRO	4.8
2	B	184	GLY	4.4
6	G	32	ALA	4.1
2	B	197	MET	4.0
2	B	185	PRO	4.0
5	F	199	GLY	3.9
5	F	197	MET	3.8
1	A	250	GLY	3.4
5	F	184	GLY	3.3
5	F	186	GLY	3.3
1	A	198	GLU	2.9
1	A	186	GLY	2.9

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	260	ALA	2.9
6	G	33	ILE	2.8
2	B	196	ALA	2.5
2	B	231[A]	VAL	2.5
4	E	187	LEU	2.5
4	E	199	GLY	2.4
4	E	198	GLU	2.3
1	A	231	VAL	2.1
3	C	85	VAL	2.0
4	E	247	LYS	2.0
1	A	199	GLY	2.0
6	G	112	GLY	2.0
6	G	113	PHE	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains [\(i\)](#)

There are no non-standard protein/DNA/RNA residues in this entry.

## 6.3 Carbohydrates [\(i\)](#)

There are no carbohydrates in this entry.

## 6.4 Ligands [\(i\)](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
7	MPD	G	201	8/?	0.87	0.16	54,57,61,67	0
7	MPD	C	201	8/?	0.92	0.14	49,52,60,63	0
8	CL	D	1	1/?	0.99	0.09	38,38,38,38	0
8	CL	D	2	1/?	1.00	0.07	28,28,28,28	0

## 6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.