Structural and biochemical investigation of the RRS1/RPS4 immune receptor pair

Hannah Victoria Brown

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

The Sainsbury Laboratory

September 2019

Abstract

Pathogens secrete an arsenal of effectors to facilitate infection, however these effectors are themselves susceptible to molecular detection by plant immune receptors known as NLRs (nucleotide-binding, leucine-rich repeat proteins). Our understanding of how NLRs function mechanistically still remains relatively rudimentary. Mechanistic studies of NLRs have largely been hampered by difficulties in producing sufficient quantities of soluble NLR protein for biochemical and structural investigation. A major goal in the field is to expand our structural understanding of how effectors activate NLRs and to increase the variety and number of protein structures of these receptors.

Using the example of the Arabidopsis paired TIR-NLRs RRS1 and RPS4, this work aimed to evaluate a diverse range of protein expression systems for their suitability for NLR protein production. This study investigated the use of classical heterologous expression systems of *E. coli* and insect cells as well as plant-based systems such as cell-free wheat germ, transgenic lines of *Arabidopsis thaliana* and agroinfiltration of *Nicotiana benthamiana*. Soluble full-length RRS1 protein could be purified from plant-based systems, though protein yield issues have hampered current efforts to gain structural information on this protein. The insights presented in this multi-system screening process should provide a valuable foundation for future studies into NLR purification.

In the second part of this work, I set out to examine the structural basis of the recognition of *Pseudomonas syringae* effector AvrRps4 by RRS1 via RRS1's integrated WRKY transcription factor domain. Previous work has shown that binding of the effector AvrRps4 to RRS1 is required but not sufficient to activate a defence response. I therefore set out to gain a structural insight into this NLR-effector interface to help guide and support our biological understanding of the inter- and intra-molecular interactions involved in the activation of this receptor complex. Through the use of quantitative biochemical techniques including surface plasmon resonance and analytical gel filtration, we gained important insights into the molecular basis of this interaction.

Acknowledgments

I would firstly like to begin by thanking my supervisors Jonathan Jones and Mark Banfield for giving me the opportunity to conduct my PhD and guiding my research over the last four years. I have learnt a lot from you both and thank you for your advice and support throughout the project. My thanks extend to all my supervisory team members past and present for your insightful discussion and inspiration which has shaped my PhD.

Whilst TSL provides an amazing environment for carrying out scientific research, it is the people here that have really made my PhD experience special. The student community in particular has been an enormous source of encouragement, support and at times some much-needed distraction! I am grateful to have had an amazing group of student colleagues in the Jones lab during my PhD who have made every day here at TSL full of laughter, friendship and some strange YouTube videos… I would particularly like to thank Yan Ma who was my mentor throughout the early years of my PhD and is an eternal source of wisdom, scientifically and beyond. I have also been very fortunate to be part of the RRS1/RPS4 team during my time in TSL, members of which past and present have been an invaluable source of scientific inspiration and encouragement and created a great environment for me to grow as a researcher.

I would also like to thank all members of the Banfield lab and Chatt Building for your help and advice throughout my PhD. Particularly to Josie, Freya, Jenny, Hannah, Marina, Nitika, and Sarah whose friendship made my 'thyme' here so memorable, your support throughout all the challenges of the past few years means more than I can say!

A massive thank you has to go to my housemates Josie and Helen without whom my experience these last four years would not have been the same. You have always known how to pick me up and put a smile back on my face whether it's a hug, a laugh or a G&T. I know in you both I have friends for life and I will be always be eternally grateful for your friendship and support and my amazing years in the 'coven' of Cunningham Road!

A big thank you to my partner Alex for your constant support and understanding through my whole PhD journey. Thank you for listening to my numerous and likely incoherent rants

and answering them only with kindness, these last four years would not have been the same without you.

Finally, I would like to thank my family for all their support and encouragement throughout the years and providing a very important sense of perspective! You all continue to inspire me daily and I wouldn't be who I am today without you.

The analogy of a rollercoaster in my experience cannot be more strongly applied than to a PhD. Through highs and lows these people have stuck by and supported me so thank you to you all, it has been a ride to remember!

Table of Contents

Table of Contents

Table of Contents

List of Figures

List of Figures

List of Tables

Abbreviations

Abbreviation	Definition
AIM	Autoinduction media
ADPR	Adenosine diphosphate ribose
ATP	Adenosine triophosphate
BiFC	Biomolecular fluorescence complementation
BIK1	Botrytis Induced Kinase 1
bp	Base pair
CC	Coiled coil
cDNA	Complementary deoxyribonucleic acid
CDPK	Calcium-dependent protein kinase
CL	Crude lysate
CNL	Coiled coil Nucleotide-binding, leucine-rich repeat protein (CC NLR)
CTD	C-terminal domain (RPS4)
DAMP	Damage associated molecular patterns
ddH_2O	Double distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
Dom4	Domain 4
Dom6R/S	Domain 6 of RRS1-R/S allele
dpi	Days past inoculation
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EDS1	Enhanced disease susceptibility 1
EGTA	Egtazic acid
EMSA	Electrophoretic mobility shift assay
ETI	Effector-triggered immunity
FLS2	Flagellin Sensitive 2
gDNA	Genomic DNA
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HF	Hell-Fire tag (6xHis, 3xFLAG epitopes)
Hisx6	6 x Histidine
HMA	Heavy metal associated

Abbreviations

Abbreviation	Definition
HR	Hypersensitive response
HRP	Horseradish peroxidase
HS	HA and Strep (affinity tag)
IB	Instant blue
IMAC	Immobilized Metal Affinity Chromatography
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
JIC	John Innes Centre
K_D	Equilibrium dissociation constant
LB	Lysogeny Broth
LRR	Leucine-rich repeat
MAPK	Mitogen-activated protein kinase
MBP	Maltose binding protein
mins	Minutes
MOI	Multiplicity of infection
MS	Murashige and Skoog
MW	Molecular weight
NAD^+	Nicotinamide adenine dinucleotide
NB-ARC	Nucleotide-binding domain found in Apaf1, R proteins and CED4
NLR	Nucleotide-binding, leucine-rich repeat protein
NRG1	N Requirement Gene 1
OD_{600}	Absorbance at 600 nm
OPPF	Oxford Protein Production Facility
PAGE	Polyacrylamide gel electrophoresis
PAMP	Pathogen associated molecular patterns
PCR	Polymerase chain reaction
PDB ID	Protein Data Bank Identification code
Pkinase	Protein kinase
POI	Protein of interest
PRR	Pattern recognition receptor
PS	Ponceau S stain
PTI	PAMP-triggered immunity

Abbreviations

Abbreviation	Definition
PVPP	Polyvinylpolypyrrolidone
RBOHD	Respiratory Burst Oxidase Homologue D
R protein	Resistance protein
ReDCaT	Re-useable DNA Capture Technique
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rpm	Revolutions per minute
RPS4	Resistance to *Pseudomonas syringae* 4
RRS1	Resistance to *Ralstonia solanacearum* 1
RRS1-R	RRS1 allele from Ws-2 accession (RRS1^{Ws-2})
RRS1-S	RRS1 allele from Col-0 (RRS1^{Col-0})
RU	Response units
SAXS	Small-angle X-ray scattering
SCAF	signalling through cooperative assembly formation
SDS	Sodium dodecyl sulfate
sec	Seconds
SETI	Super Effector-triggered immunity
SF	Soluble fraction
spp	Species pluralis
SPR	Surface plasmon resonance
SUMO	Small ubiquitin-like modifier
TCEP	Tris(2-carboxyethyl) phosphine
TIR	Toll/Interleukin-1 receptor
TNL	Toll/Interleukin-1 receptor Nucleotide-binding, leucine-rich repeat protein (TIR-NLR)
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
tRNA	Transfer ribonucleic acid
TSL	The Sainsbury Laboratory
UV	Ultraviolet
v/v	Volume to volume
w/v	Weight to volume
WT	Wild type
ZAR1	HopZ-Activated Resistance 1

Introduction

Approximately 20-40% of global crop yields are lost each year due to damage by pests and disease. With a global population set to grow by a further 20% by 2050, combined with climatic change, and limited agriculturally viable land, such losses are unsustainable and pose a major threat to our global food security. Only through understanding the complex interplays between plant hosts and their pathogens will we gain insights to help reduce this yield gap and strengthen global food security.

Unlike mammals, plants rely on a multi-layered innate immune system comprised of a repertoire of genetically determined intra- and extra-cellular receptors for pathogen detection and induction of defences in response to pathogen attack. The work in this project focusses on the intracellular receptors involved in plant innate immunity and understanding the mechanism through which these receptors perceive pathogens and activate plant defences. This introductory chapter will discuss a broad overview of plant immunity focussing on the perception of pathogen effectors by intracellular receptors, namely the modular nucleotide-binding, leucine-rich repeat proteins (NLRs).

1.1   Overview of Plant-pathogen interactions

1.1.1   Models for plant innate immunity

Plants are constantly exposed to a wide variety of biotic and abiotic challenges to which they must be able to adapt and respond appropriately in order to reproduce successfully. Challenge by pathogens such as bacteria, fungi, oomycetes, nematodes and viruses has led to the evolution of a multi-layered plant innate immune system which is capable of cell autonomous detection of the presence of pathogens leading to activation of defence responses within the plant tissue. There are a growing number of models to illustrate this multi-layered response with the most current describing two interlinked layers of plant innate immunity referred to as the 'zigzagzig model'[1,2], **Figure 1.1A**. In this model, the primary layer of immunological response is initiated at the plasma membrane, where plants utilise membrane spanning receptors known as PRRs (Pattern recognition receptors) for the

activation of the first layer of the plant immune response, PTI (Pattern-triggered immunity).
These receptors are responsible for the detection of pathogen-derived conserved
molecules, PAMPs (Pathogen associated molecular patterns), as well as plant self-derived
peptides termed DAMPs (Damage associated molecular patterns) which are believed to act
as markers of pathogen-induced cellular damage within the host tissue[3]. To attenuate host
immune responses, pathogens have evolved to secrete an arsenal of effector proteins
which suppress host defence responses and facilitate pathogen infection through
manipulation of host proteins and signalling pathways. However, these effector proteins
are themselves susceptible to molecular detection by the host plant cell through
intracellular immune NLR receptors, often encoded by *Resistance (R)* genes[4]. Activation of
NLRs through direct or indirect effector perception, leads to induction of ETI (effector
triggered immunity) often associated with the programmed cell death hypersensitive
response (HR), **Figure 1.1**.
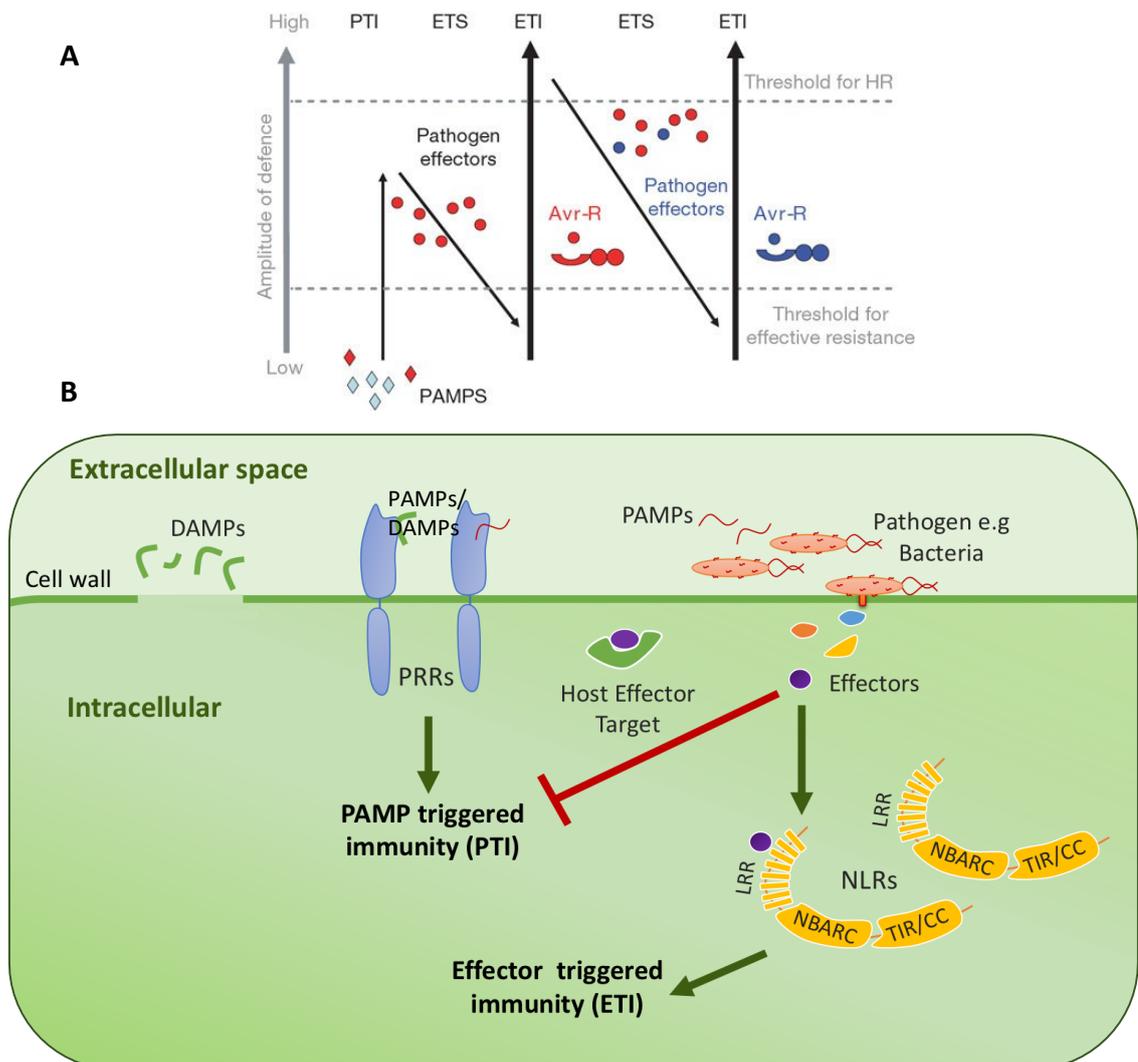
Figure 1.1 The multi-layered model of plant immunity. (A) Zigzagzig model of plant immunity
presenting the quantitative output of the plant immune system in four phases. Phase 1 involves

detection of PAMPs via PRRs resulting in activation of PTI. In phase 2 successful pathogens deliver effectors (Avr-Rs) into the host cell which act to attenuate PTI, effector triggered susceptibility (ETS). Phase 3 involves detection of pathogen effectors by plant NLRs resulting in activation of ETI. In phase 4 pathogen isolates which have lost recognised effectors are selected for leading to ETS, figure taken from Jones and Dangl, 2006[5]. **(B)** Representation of the zigzagzig multi layered plant immune system in plant cellular context. PAMPs (Pathogen associated molecular patterns), DAMPs (Damage associated molecular patterns), PRRs (Pattern recognition receptors), NLRs (Nucleotide-binding, leucine-rich repeats). Plasma membrane associated PRRs can take the form of receptor kinases or receptor like kinases and often function with co receptors (not shown) which bind PAMPs and DAMPs via extracellular LRR domains. NLRs are predominantly intracellular proteins based with a largely canonical modular structure. NLRs can recognise effectors by a number of mechanisms discussed in 1.2.

Since the 'zigzagzig model' was first postulated, other authors have extended or modified this original model as more experimental data has been revealed. For example, the 'invasion model'[6] avoids a strict definition of PAMPs and effectors in favor of a more spectrum-based definition of immunogenic molecules. The 'invasion model' refers to effectors, PAMPs and DAMPs as invasion patterns (IPs) and both PRR and NLR immune receptors as invasion pattern receptors (IPRs) which activate an IP-triggered response. This more generalized model, argues against the strict separation of PTI and ETI signalling pathways in favour of a broad view of a spectrum-based system evolved to detect invasion without the constraints of strict definitions on pathogen-derived molecules.

The strict distinction between widely distributed, conserved PAMPs and race, species or strain-specific effectors is increasingly becoming blurred as our understanding of plant-pathogen interactions evolves[7]. Additionally, some *R*-genes have been found to encode PRRs, for example the *Solanum lycopersicum* PRR Cf-9 which confers recognition of the *Cladosporium fulvum* effector Avr9[8,9]. To accommodate this, a new model referred to as the 'spatial immunity model' has been proposed[10]. The model argues that the definition of a plant immune response should be determined on the basis of where immunogenic molecules are perceived. The model suggests the use of the term extracellular immunogenic pattern (ExIP) or intracellular immunogenic pattern (InIP). The spatial bipartition in the defining immunogenic molecules allows the differentiation of immune signalling events in the plant host to recognition of ExIPs at the cell surface as

extracellularly triggered responses (ExTRs), and intracellular InIPs as intracellularly triggered immunity (InTI)[10].

Whilst our ever-evolving understanding of the complexities of microbe-plant interactions means these models are not likely to represent all examples, such models help to conceptualise the intricacies of the interactions between these organisms. For the purposes of this study I will refer to the players in plant-pathogen interactions using the terms described in the 'zigzagzig model', **Figure 1.1**.

1.1.2 Perception of MAMPS and cell surface immunity

PAMPs/MAMPs are highly conserved throughout the lineages of microbes and through their detection provide plants with a mechanism for perceiving and defending against the vast majority of pathogenic challenges. PAMPs have been identified across the major phylogenetic lineages of plant pathogens. Well described bacterial PAMPs include the flagellin peptide flg22[11], peptidoglycans[12] and elf18 a peptide derived from Elongation factor Tu[13]. PAMPs have also been identified in fungi and oomycetes, for example chitin[14] and heptaglucan[15] respectively.

Plant self-derived molecules termed DAMPs are also capable of activating PRRs. DAMPs are believed to act as a marker for pathogen-induced cellular damage within the host tissue providing an alternative mechanism for extracellular pathogen recognition. Oligogalacturonides (oligomers of alpha-1,4-linked galacturonosyl residues) are an example of a DAMP released via direct damage caused by invading pathogens[16]. During host invasion many fungi species release cell wall degrading enzymes which release oligogalacturonides usually embedded in the matrix of the cell wall. Oligogalacturonides are recognised by PRR WAK1 (wall-associated kinase 1) which transduce this perception into a PTI defence response[17].

PAMPs are released by pathogens proliferating in the apoplast of plant tissue following invasion and are subsequently vulnerable to detection by PRRs. There are two major classes of PRRs: Receptor kinases proteins (RKs) and Receptor-Like proteins (RLPs). It is believed that most RLPs, which lack internal kinase domains, function in conjunction with one or more RKs to allow them to transduce PAMP ligand binding into an internal signalling response[3].

Introduction

Upon binding of PAMPs/DAMPs, PRRs activate a PTI response with some PRRs requiring co-receptors such as BAK1 (Brassinosteroid Insensitive 1-associated receptor kinase 1) or other SERK (Somatic embryogenesis receptor kinases) family members to do so[3,18]. This results in an active signalling complex which activates downstream PTI events including the activation of MAPK (Mitogen-activated protein kinase) and CDPK (Calcium-dependent protein kinase) cascades, reactive oxygen species (ROS) and reactive nitrogen species generation, resulting in large scale reprogramming of the cell transcriptome and metabolome. One such downstream target of FLS2 (Flagellin Sensitive 2) is the NADPH oxidase RBOHD (Respiratory Burst Oxidase Homologue D) which is phosphorylated by BIK1[19,20](Botrytis Induced Kinase 1). This enzyme is responsible for a classic hallmark of PTI, the generation of apoplastic ROS burst. ROS have direct roles in defence as antimicrobial molecules as well as important immune signalling components. The later stages of PTI involve events such as callose deposition, thought to help strengthen plant cell walls against pathogen invasion, and redistribution or restriction of water and nutrient availability with the aim of restricting pathogen growth.

1.1.3   Effectors and effector-triggered immunity

Whilst PTI appears to be remarkably effective at preventing the majority of non-host adapted pathogen colonisations, pathogens have evolved host specific adaptation mechanisms to overcome these defences through the secretion of an arsenal of effector proteins which largely function to attenuate host immunity, **Figure 1.1**. The delivery mechanism of effectors varies amongst pathogens. Many Gram-negative bacteria use a specialised needle-like structure which directly delivers effectors into the host cell, called the Type III secretion system[21]. For example, this mechanism is utilised by bacterial speck and blight causing bacteria *Pseudomonas syringae* and *Ralstonia solanacearum* discussed later in this study. Fungi and oomycetes deliver effectors via haustoria [22], a specialized feeding structure which extends through the plant cell wall and forms an intimate association with the host cell membrane, whilst nematodes utilise a specialized proboscis called a stylet for effector delivery[23]. Alternatively, a subset of pathogens secrete effectors into the apoplast of the host, for example *C. fulvum* [24].

Adapted pathogens deliver an arsenal of effectors into hosts which attenuate host defences and support pathogen proliferation, imposing effector-triggered susceptibility. Effectors target various components of the plant immune signalling pathways as part of their

virulence strategy. Yeast two-hybrid interactome network studies have noted a convergence of effectors from distinct pathogens targeting an overlapping subset of highly interconnected host protein 'hubs'[25]. For example, the PRR receptor complex is a common target of effectors[26–28]. In order to detect the presence of effectors, plants have evolved intracellular R-protein immune receptors encoded by *R* genes. R-protein perception of effectors leads to induction of ETI. Whether ETI is a distinct signalling mechanism or a potentiator of PTI signalling is still debated but the two layers of immune signalling do appear to be interlinked. Activation of ETI, in the presence of PTI, leads to induction of a form of programmed cell death referred to as the hypersensitive response (HR) which is thought to aide immunity by constricting pathogen growth in the plant tissue.

Most *R* genes encode NLRs. NLRs are capable of recognizing effectors directly or indirectly and can function both singly or in a pair, see section 1.2. The canonical structure of NLRs consist of: a TIR (Toll/interleukin 1-receptor) or CC (coiled-coil) class defining N-terminal domain[29]; a nucleotide binding NB-ARC (Nucleotide-binding domain found in Apaf1, R proteins and CED4) domain and a C-terminal LRR (leucine rich repeat) domain[4,30,31]. Plant genomes contain hundreds of NLR proteins likely due to strong selection for novel pathogen recognition capabilities. NLRs are believed to be the most rapidly evolving gene family in the plants with 126 NLRs presently reported in the *Col-0* accession of *Arabidopsis thaliana* alone[32].

Animals also utilize NLRs for activation of innate immunity. Animal NLRs share a similar modular domain architecture with plant NLRs which is thought to have evolved independently but convergently in the two kingdoms[33]. Many animal NLR domain structures comprise; LRRs, a CARD (caspase activation and recruitment domain) or PYD (pyrin domain) N-terminal domain which activates downstream signalling and a NACHT nucleotide binding which enables activation of the NLR complex[33,34]. Upon activation by detection of pathogen-derived ligands, some mammalian NLR proteins form an oligomeric wheel-like structure referred to as an inflammasome. The structure of one inflammasome involving the NLR NLRC4 (NLR family CARD domain-containing protein 4) and NAIP2 (NLR family, apoptosis inhibitory protein 2) has now been solved using cryo-electron microscopy[35]. Given the modular similarities between the two kingdoms NLRs, advances in understanding mammalian NLR structure and function have long been used to make inferences about plant NLRs for which until recently no full-length structures were available[33,36].

1.2 Models for effector perception by NLRs

NLR proteins are involved in the induction of an ETI response upon the detection of effectors. This perception can occur through direct binding of an effector to an NLR or indirectly through detection of an effector's perturbations of the host plant cell, for example phosphorylation or cleavage of host target protein by an effector **Figure 1.2.**
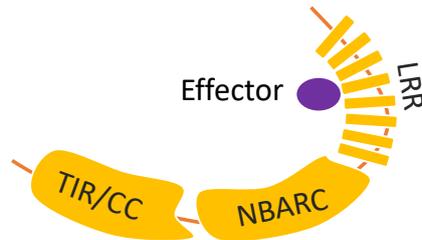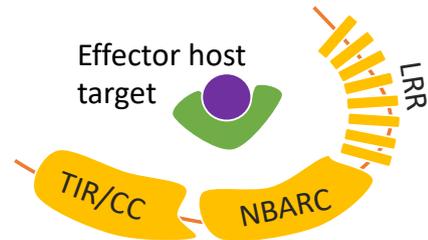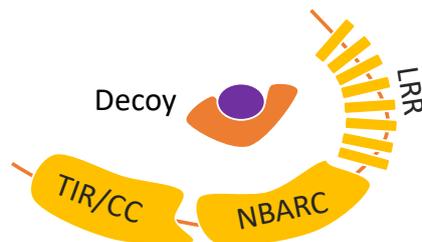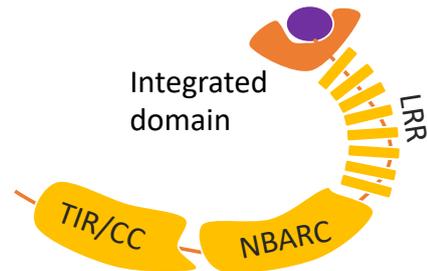
Figure 1.2 Four models of NLR effector perception mechanisms. NLRs can perceive effectors in a variety of ways. Current models include: **(A)** Direct recognition where NLRs bind effectors directly commonly via NLR's LRR domain; **(B)** Guard model in which an NLR monitors host proteins for effector induced perturbations; **(C)** Decoy Model in which an NLRS monitors a decoy mimic of authentic host targets and **(D)** The Integrated Domain model where decoy domains can be integrated into the structure of the NLR to facilitate direct binding of effector.

1.2.1 Direct recognition of effectors

The direct recognition model for NLR effector perception biochemically expands on Flor's gene-for-gene hypothesis[37]. The model describes a receptor-ligand mechanism via which direct binding of an effector to an NLR activates induction of an ETI response[38], **Figure 1.2A**. It is thought this plant-pathogen interaction is largely facilitated through the polymorphic NLR LRR domain. The effector-NLR pair of ATR1 and RPP1 is an example of such a

recognition mechanism by an NLR. *A. thaliana* NLR RPP1 has been shown to directly bind the oomycete *Hyaloperonospora arabidopsidis* (the causal agent of downy mildew) effector ATR1 via RPP1's LRR domain and activate ETI in a ligand-dependent manner[39]. Similarly, recognition of *Melampsora lini* effector AvrL567 by the L5 and L6 flax TIR-NLRs occurs directly via multiple contact points in the NLR's LRR domain with assistance from the N-terminal TIR domains[40,41].

1.2.2 The Guard model of recognition

In addition to binding effectors directly, NLRs can also survey host proteins in the plant cell for perturbations induced by an effector's virulence activity in a scenario described as the 'Guard model'[1] e.g. host protein cleavage by an effector, **Figure 1.2B**. This model also describes how one NLR can monitor for the presence of multiple effectors through a guarding a common virulence target of several effectors. This allows plants to monitor for a range of pathogens via a limited repertoire of NLRs.

A key example of the 'Guard' mechanism of NLR recognition is the *A. thaliana* guardee protein RIN4. RIN4 is a small, unstructured protein seen to associate with PRR complexes and act as a scaffold in both PTI and ETI signalling. It is the virulence target of multiple effectors and monitored by the 'Guard' NLRs RPM1 and RPS2. The effector AvrRpt2 from *P. syringae* cleaves RIN4 as part of its virulence strategy. This cleavage event can be detected by RPS2 and leads to activation of RPS2-mediated ETI defences[42,43]. The *P. syringae* effectors AvrB and AvrRpm1 also target RIN4 by up regulating expression of RIPK which phosphorylates RIN4 residue T166. Phosphorylation of RIN4 at this residue is perceived by RPM1 and leads to activation of ETI[44,45]. In this way both RPM1 and RPS2 can indirectly perceive a pathogen challenge through monitoring a virulence target of effectors for perturbations[44]. Interestingly, the effector HopF2 has evolved to ADP-ribosylate RIN4, hypothesised to inhibit accumulation of S141 phosphorylated RIN4 which promotes PTI without increasing phosphorylation of RIN4 residue T166 monitored by the NLR RPM1 thereby evading NLR detection[46]. This example highlights an on-going evolutionary arms race between pathogen and plant.

1.2.3 The Decoy hypothesis

A variation on the 'Guard model' of NLR activation is the 'Decoy model'[47]. This model is similar to the guard model in that NLRs detect the virulence functions of effectors and the host proteins they target as a proxy for effector detection. However, instead on monitoring the authentic host effector target protein 'Decoy model' NLRs monitor a virulence target mimic for effector-induced perturbations, **Figure 1.2C**.

Through utilisation of this mechanism a single structural decoy or 'effector bait' monitored by one or more NLRs can protect a whole family of effector target proteins. For example, the NLR Prf is kept in an autoinhibited state by the kinase Pto. Pto is cleaved by the effector AvrPto releasing Pto's inhibition of Prf and activating defence[48,49]. Pto itself is not known to have a direct role in plant resistance but it shows great homology to several other kinase domains involved in immunity, namely those of PRR receptors FLS2 and EFR[26]. Similarly, RPS5 guards the decoy PBS1 which negatively regulates the NLR[50,51]. AvrPphB targets PBS1 and cleaves the protein allowing RPS5 immune defence response induction[28]. PBS1 has been shown to play no role in immunity but shows structural similarity to several other virulence targets of AvrPphB which do play important roles in immune signalling including BIK1[26,28].

1.2.4 Integrated domain NLRs

Whilst direct effector perception within the canonical NLR structure is classically associated with effector interactions with the LRR domain[39,40], it is estimated that between 3-10% of plant NLRs utilize non-canonical domains termed 'integrated domains'[52,53] to perceive effectors. Integrated domains are believed to have evolved from the host targets of effectors and facilitate direct binding of effectors to NLRs by acting as bait for pathogen effectors[54,55], **Figure 1.2D**.

Whether these integrated domains are actually true 'decoys' retaining none of their template host protein's biological function themselves is still controversial[56]. The 'integrated decoy' model[54] implies the non-canonical domain integrated into the structure of the NLR has lost its original biochemical function and acts solely as a structural mimic in effector sensing. However, this is not the case for all integrated domain NLRs as the integrated WRKY domain of RRS1 for example still retains the ability to bind W-box DNA

Introduction

like the WRKY transcription factors the domain mimics[57,58]. Subsequent authors proposed that the term 'integrated decoy' should be replaced with 'sensor domain' in order to fully incorporate the possibility that at least some of these integrated domains may retain their ancestral biochemical function, perhaps even for only a short time in their evolutionary history[59]. The authors argue that as the identity or biochemical function of many of the ancestral effector targets or integrated domain NLRs have yet to be identified it is not possible to label these domains simply as biochemically inactive decoys. Furthermore, the evolution of these integrated domain NLRs still remains unclear and it is entirely possible that fully functioning effector target domains may become fused into the structure of NLRs before subsequently losing the ancient biochemical function. This would suggest that more recent NLR atypical domain fusions may not have yet evolved to lose this function and become a true 'decoy' [56]. Some have argued however against the use of the term 'sensor domain' noting the term 'sensor' does not fully convey the principle that these domains are mimicking functional host target template of the integrated domains[60]. It has therefore been suggested that 'integrated domains' may be a more appropriate term as this incorporates the concept of an atypical domain fusion into the NLR structure without assigning a function to the domain [52] and is the term utilised in this study.

Several bioinformatics studies have highlighted the diversity of NLR integrated domains with evidence of such integration events occurring in phylogenetically unrelated monocot and dicot lineages as well as at least one moss species (*Physcomitrella patens*)[52,53]. Bioinformatic studies predict that the most common atypical domains to be structurally integrated into NLRs are Pkinase domains as well as the DNA binding WRKY and Zf-BED domains[52,53]. Pkinase and WRKY domains have classically been associated with plant immunity signalling and increasing evidence is emerging for the role BED domain-containing proteins play as well[53,61–63]. This finding strongly supports the concept postulated by the 'integrated domain' model that domains integrated into NLRs would be linked to playing important roles in plant immunity and hence targeted by effectors. Several NLRs containing these atypical domains have already been cloned. For example: the wheat stem rust resistance (*Puccinia graminis*) NLR RPG5 cloned from Barley (*Hordeum vulgare*) contains an integrated protein kinase domain[64]; Integrated WRKY domains have been identified in the NLR RRS1 which functions in a pair with RPS4 to recognise effector *from P. syringae*, *R. solanacearum* and *Colletotrichum higginsianum*[65,66] and ZfBED domains were found in the cloned *Oryza sativa* resistance NLR Xa1 which recognises an unknown bacterial blight effector[67].

Introduction

Integration of these domains appears to commonly occur in pairs of genetically linked NLRs which function together to elicit an immune response. For example, in the *A. thaliana* NLR pair utilised in this thesis (RRS1 and RPS4) the sensor NLR RRS1 contains an integrated WRKY domain[65,66] and in the well-studied *Oryza sativa* NLR pair Pikp-1 and PikP-2, PikP-1 contains an integrated Heavy-metal associated (HMA) domain[68,69]. As demonstrated in these two integrated NLR pairs, the location of the integrated domain in the structure of an NLR is not consistent, with the HMA domain of PikP1 found between the CC and NB-ARC domain and WRKY of RRS1 at the C-terminus of the protein, **Figure 1.3**. This may reflect different intra- and inter-domain interactions involved in the activation mechanisms for NLRs of this variety. It is likely given the variation in modular domain layout that binding of an effector to the integrated domain of an NLR may have a different effect on the intra and inter molecular interactions with other NLR domains. This suggests that there will not be a single step-by-step activation mechanism of all integrated domain NLRs rather a variety of paths utilised by different NLRs to transduce effector signalling into defence activation. Such differences in activation and inter-domain interactions can even be demonstrated between paralogous RRS1 and RRS1B discussed in 1.5.3.
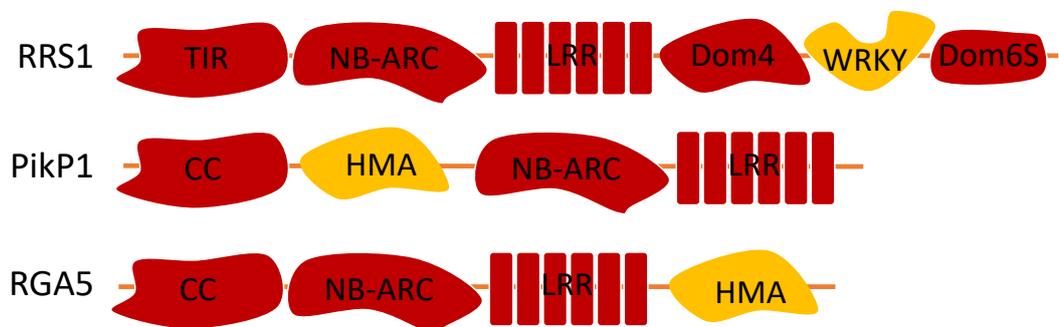
Figure 1.3 Variations in modular domain structure of integrated domain NLRs. Modular structure of NLRs RRS1, PikP1 and RGA5 that contain an integrated WRKY and HMA (Heavy metal associated) integrated domain respectively demonstrating that the site of integrated domains (highlighted in yellow) can vary between NLRs. Size of domains is not to scale.

Understanding the structural basis of NLR integrated domains recognition of effectors is a key step in enabling the future engineering of synthetic NLRs with expanded recognition capabilities. The structures of several NLR integrated domain-effector complexes have now been solved including the *Magnaporthe oryzae* effector Avr-PikD bound to the HMA domain of PikP1[70] and *R. solanacearum* PopP2 bound to the integrated WRKY domain of

RRS1[71], the latter of which interactions are further discussed in 1.5.2 and 5.1. Understanding effector-integrated domain binding events at the structural level allows us to begin to understand the constraints and requirements of effector recognition by integrated domain NLRs which can then be used to guide future engineering of synthetic NLRs. There is also evidence that integrated domain NLRs can interact with pathogen effectors via domains outside of their defined integrated domain. This is demonstrated by the ability of the *M. oryzae* effector AVR-Pia to interact with *O. sativa* NLR RGA5 at sites both within and outside of RGA5's integrated HMA domain[72]. The mechanistic relevance in terms of effector perception and NLR regulation of these non-integrated domain interactions is not yet understood[72] but successful engineering of synthetic NLRs with expanded effector recognition capability will likely rely on a comprehensive understanding of the entire interaction surface between NLRs and effectors.

1.3   The domain structure of NLR proteins

NLRs are modular proteins largely consisting of three classical domains[73]: a N-terminal CC or TIR domain[74]; a central NB-ARC ATPase domain[75] and a C-terminal LRR. Many NLRs show structural variations to this domain pattern including domain duplications and deletions as well as integration of non-canonical domains such as integrated domain NLRs, discussed in section 1.2.4, which contain a variety of non-canonical domains such as WRKY, BED and HMA domains. Truncated NLRs have also been identified such as the TIR only Arabidopsis NLR RBA1[76].

1.3.1   Class defining N-terminal domains

NLRs can be categorized into one of two monophyletic groups based upon the nature of their N-terminal domains: TIR domain NLRs (TNLs) found only in dicots and CC domains NLRs (CNLs) present in both monocots and dicots[74,77]. There is a second small subclass of the CNLs which do not possess the negatively charged 'EDVID' motif found in the rest of the CNLs class. Instead this group shows strong homology for the non NLR R protein RPW8 (Resistance To Powdery Mildew 8) and are therefore categorised as CC_{RPw8} NLRs[78,79]. This CC_{RPw8} class of NLRs also shows difference to the rest of the CNLs in their NB-ARC domains and are thought to form a monophyletic anciently diverged clade within the NLRs[29,80]. The function of these N-terminal domains has been linked to defence signalling with transient

expression of only the N-terminal domains from a variety of NLRs capable of inducing an HR *in planta*[78,81,82].

Through the solving of multiple NLR TIR domain structures two putative dimerization interfaces have been identified in this domain. These two interfaces are clearly highlighted in the crystal structure of *A. thaliana* NLR SNC1's TIRs: the DE interface and AE interface [83,84], see **Figure 1.4** . Mutations in either of these TIR dimerization interfaces were found to disrupt the cell death instigating abilities of the SNC1, L6 and RPS4 TIR domains as well as full-length L6 and RPS4 highlighting the importance of these interfaces in the ability of the NLR to invoke an immune response[84]. It was subsequently postulated that dimerization of the TIR domains may allow the TIR domains to signal for an immune response via signalling through cooperative assembly formation (SCAF)[85].
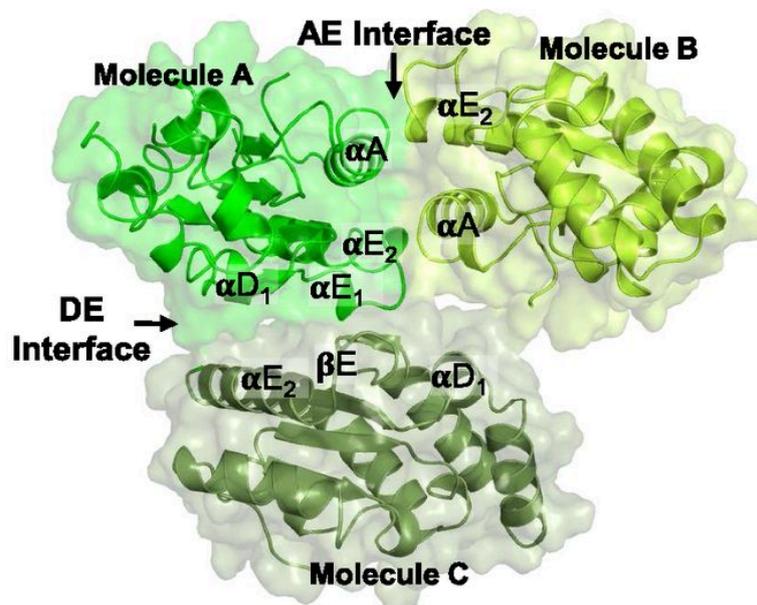
Figure 1.4 Multiple self-association interfaces of the TIR domain. The SNC1 TIR crystal structure reveals two self-association interfaces, the DE interface αD and αE regions of the protein and AE interface involving the αA and αE regions. In the SNC1 TIR structure, Molecule A and B were observed in the asymmetric unit and interact through the AE interface; and molecules A and C interact with a crystallographic symmetry-related molecule through the DE interface.Figure taken from Zhang et al, 2017[84]

It appears these AE and DE interfaces can be widely found throughout TNLs with similar interfaces also being identified in the Arabidopsis NLR RPP1[84]. Identification of these two dimerization interfaces has led to the development of the hypothesis that these interfaces may allow TNLs to oligomerise via their TIR domains forming oligomeric structures similar to those seen in mammalian NLRs[85]. It should be noted however that these observations

are purely based of models of the TIR domains alone and do not include the effects that other NLR domains may have on the formation of this oligomer in terms of domain arrangement or stoichiometry. Though several NLRs have been observed to self-associate upon activation *in planta*[82,86], the downstream signalling importance of dimerization or oligomerisation of plant NLRs upon activation has yet to be fully understood[33].

What has been observed for several NLRs is that transient over-expression of the N-terminal domains of NLRs alone can result in a cell death response in cells. This has been demonstrated for both TNLs (RPS4, N, RPP1, L10, L6)[30,39,87–91] and a subset of CNLs (MLA, NRG1, ADR1)[78] demonstrating this domains' role in at least some NLRs as activators of downstream signalling pathways. This is not observed however in the case of all NLRs. Overexpression of the potato NLR Rx NB-ARC domain alone is sufficient to induce cell death in N. tobacum[90]. Conversely, the expression of the CC or NB-ARC domain of RPS5 was observed to not be capable of inducing HR. However, when expressed together these two domains did induce an HR response in the plant tissue suggesting the CC and NB-ARC[39] work together in RPS5 to activate downstream defences[91]. Whether these findings however are due to variations in method of NLR activation or experimental procedure (e.g epitope tags or expression levels) still needs to be assessed[30].

Recent observations have also suggested that plant TIR domains may possess NADase activity[92,93]. For example, the TIR domains of RUN1 and L6 were observed to cleave NAD+ into Nicatinamide and ADPR. What role this possible NADase activity plays in downstream signalling however is not yet understood[92]. Notably, NADase biochemical activity was not observed for all NLR TIR domains including RPS4, SNC1, RPP1 and ROQ1.

Several structures of CC domains have been published and are hypothesised the capture the different activation state conformations of NLR CC domains. The structure of Sr33's CC domain (four helix bundles) shows close similarities to unrelated CC domain of Rx and is thought to capture the inactive monomeric state of a CC domain. Conversely the dimer of two elongated antiparallel helix-turn-helix monomers of MLA10 CC domain is thought to depict the active dimeric state of an NLR CC domain [94,95,96], **Figure 1.5**.
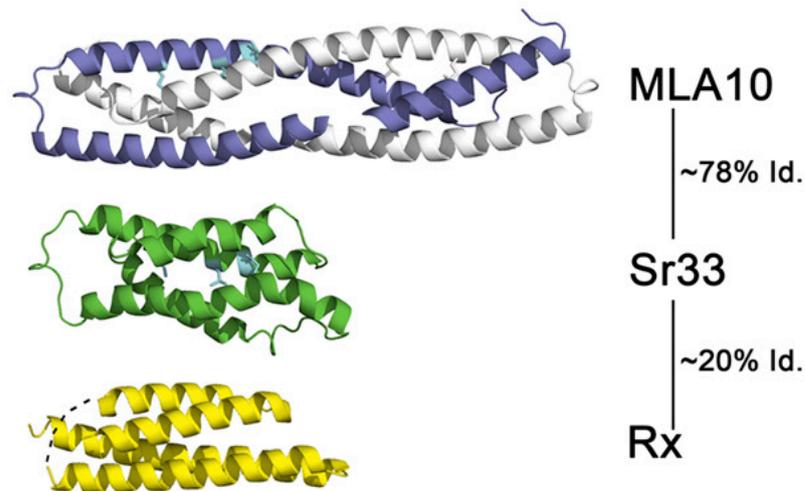
Figure 1.5 The activation structures of plant NLR coiled-coil domains. Elongated dimer of MLA10 CC domain (Purple/white-PDB 3QFL) is thought to represent an active state whilst the compact monomer structure of Sr33 CC (Green-PDB 2NCG) and Rx CC (Yellow-PDB 4M70) are predicted to represent the inactive state. Amino acid sequence identity of protein's CC domains

1.3.2 NB-ARC

Structurally related to the nematode *Caenorhabditis elegans* cell death regulator CED4[97] and *Homo sapien* Apaf-1 domain[98], the plant NLR NB-ARC domain consists of three subdomains: NB (Nucleotide-binding), ARC1 and ARC2. Interestingly plant NLR NB-ARC domains have lost the fourth ARC3 domain found in the *H. sapien* protein Apaf-1. The NB domain is comprised of five-stranded parallel β sheets surrounded by seven α helices. The ARC1 domain consists of four α helix bundles whilst the ARC2 domain adopts a winged helix fold. Plant NLR NB-ARC domains have ATPase activity and contain a number of conserved motifs including: hhGRExE, P-loop (Walker A), Walker B, GxP, RNBS-A to D, and MHD motifs[99,100]. Through structure-informed mutagenesis the roles of a number of these motifs has been investigated. The NB subdomain P-loop motif contains a conserved lysine residue in the consensus sequence GxxxxGKS/T which is key for nucleotide interactions binding α and γ phosphates. In mammalian APAF-1 the MHD motif in the ARC2 subdomain has been found to interact with β phosphate of the bound nucleotide whilst the NB Walker B motif is thought to be responsible for nucleotide hydrolysis. The RNBS-C motif marks the distinction between the NB and ARC1 domain whilst the hhGRExE links the NB to the N-terminal domain and is involved in conformation changes of the protein in APAF-1[100], see **Figure 1.6**.

Introduction
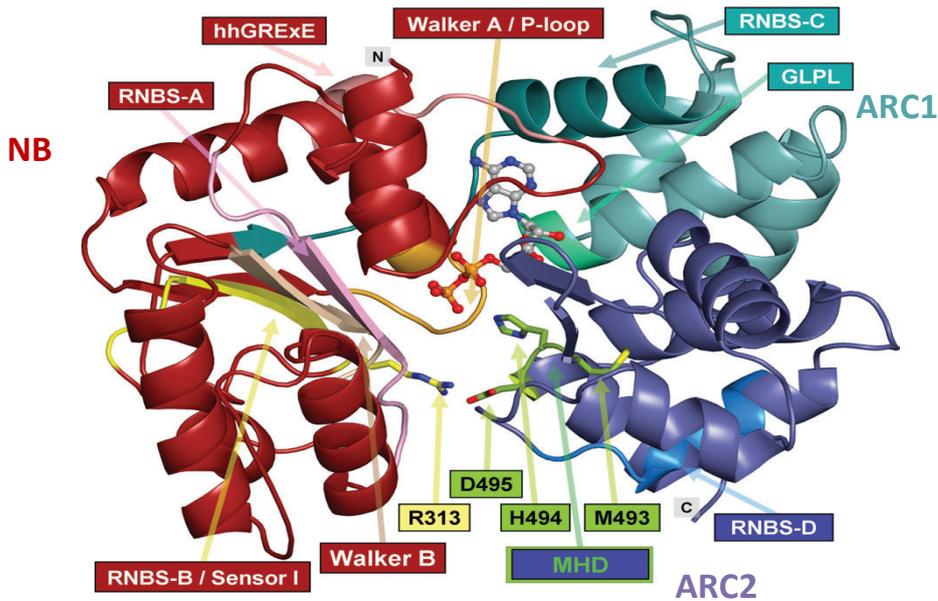
Figure 1.6 Model of plant NB-ARC domain structure and motifs. Structural model for NB-ARC domain of NLR I-2 based on structural template of ADP bound human Apaf-1. Subdomains are coloured red for NB, cyan for ARC1 and purple for ARC2. Figure adapted from Ooijen et al, 2008[264]

1.3.3   LRRs

Plant NLR LRR domains are highly polymorphic and contain varying nu[...] based around the consensus motif of hydrophobic leucines intersper[...] residues (LxxLxLxxN/CxL) as well as non-canonical LRR motifs. The rep[...] form a parallel β-sheet at the concave side of an arc shaped structure [...] residues forming the hydrophobic core, see **Figure 1.7**. Classically asso[...] binding domain of direct recognition ligand-receptor NLRs, the hyper v[...] region and the highly adaptable exposed surface utilised in LRR protei[...] mean this domain is primed for rapid evolution of new recognition spe[...]

Classically associated with being a protein-protein interaction domain, [...] NLR activation appears to vary, largely facilitated by the domain's larg[...] surface. Co-immunoprecipitation experiments suggest that in additio[...] role in the perception of effectors in the case of direct recognition NLR[...] also has a pivotal role in maintaining the NLR in an auto-inhibited state[...] effectors[51]. For example, the LRR domain of RPS5 has been observed t[...] NB-ARC domain of the NLR keeping the NLR in an auto-inhibitive state[...]

Figure 1

effector. This interaction is reliant on the first four LRRs of RPS5. A similar mechanism has been predicted for the mouse NLR NLRC4 based of the crystal structure of the protein. In the case of NLRC4 suppression of the NLR signalling activity is thought to be dependent on the LRRs domain negative suppression of NLRC4's NB-ARC domain[101,102].
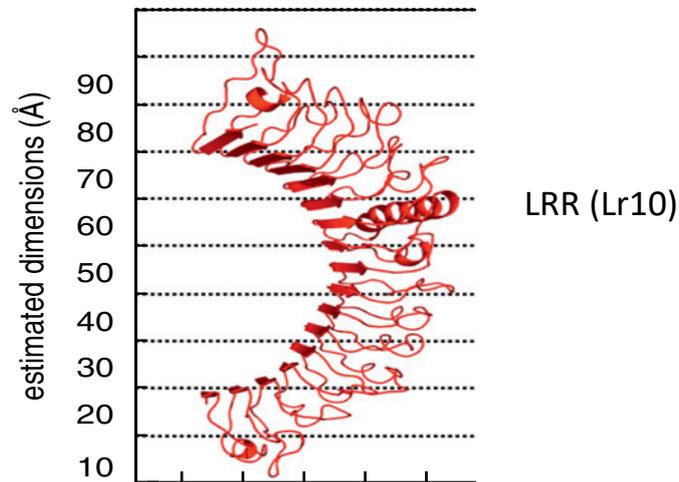
Figure 1.7 Composite Model of Lr10 LRR domain. Composite model of the LRR domain of Lr10 revealed a compact horseshoe-like structure divided into an N-terminal part containing a cluster of positively charged residues and a C-terminal part enriched in aromatic amino acids possibly involved in hydrophobic interactions. Figure adapted from Takken et al, 2012[74].

1.4 NLR activation and signalling

1.4.1 Inter- and intra-molecular interactions of NLR activation

Until the recent publishing of the Arabidopsis NLR ZAR1 structure, models of NLR activation were largely based on biochemical data, single domain structures and inference from mammalian NLRs. Now, utilising the information from the structure of CC-NLR ZAR1 and previous biochemical observations, a model has been put forward for NLR activation. However, it should be noted that whilst the broad principles of this model might apply to most NLRs it is not likely to represent all NLRs universally given the diversity of plant NLRs and effector perception strategies, see section 1.2.

The ATPase function and nucleotide binding status of NB-ARC domains has long been proposed to be the regulator of NLR activation acting as a form of molecular switch [103,104]. In this model, nucleotide binding and exchange is coupled with conformational changes within the NLR and transduction of defence signalling downstream[75]. Key mutations in the

nucleotide binding P-loop of several NLRs including RPS4, N and Rx render the NLR inactive highlighting the critical role of the NB-ARC domain in ETI induction[87,103,105]. The NB-ARC domain is also the focus of the 'equilibrium model' of NLR activation which describes a scenario in which an NLR adopts an 'off' state conformation when the NB-ARC domain binds ADP changing to an 'on' state of defence signalling when ATP is bound[106]. This model hypothesizes that though an NLR is capable of binding ATP or ADP in the absence of an effector, effector binding helps to stabilize the ATP-bound form of the NLR. Stabilization of the ATP bound form of NLR shifts the equilibrium between the two states and allows the ATP bound form to accumulate in the plant cell potentially beyond a threshold upon which defence is activated. However, ADP/ATP exchange has been demonstrated not to be key for the activation of all NLRs. For example the *O. sativa* CC-NLR PBL1 confers broad resistance to blast fungus *M. oryzae* whilst lacking a P-loop motif for ADP/ATP binding[107]. This highlights a key theme to NLR mechanistic work in that it is likely there are multiple mechanism through which different NLRs are activated.

Autoinhibition of NLRs largely involves intramolecular interaction of multiple NLR domains with the LRR domain[74,90]. For example both the N-terminal CC/TIR domains and NB-ARC is predicted to be in close proximity of the LRR in the autoinhibited state with simple models suggesting the LRR and CC/TIR sequester away the NB-ARC preventing nucleotide exchange[74,75,100]. The LRR domain was found to play a key autoinhibited role in the case of the ZAR1 structure, in which the authors found that in the absence of *Xanthomonas campestris* effector AvrAC, ZAR1 precomplexed with pseudokinase RKS1, resides in an autoinhibited conformation with ADP bound in ZAR1's nucleotide binding pocket. Intramolecular interactions involving the LRR domain of ZAR are thought to sequester the NLR in an inactive conformation [108,109], **Figure 1.8.**

The ZAR1 structure has now provided key insights into how activation upon ADP to ATP exchange occurs in this NLR example. The virulence target of AvrAC is BIK1, a key kinase involved in PTI, which the effector uridylylates to prevent activation of PTI. ZAR1-RKS1 guards PBL2, a kinase decoy of BIK1 which is then uridylylated by AvrAC leading to activation of ZAR1[110]. Uridylylated PBL2 interacts exclusively with RKS1 which interacts with ZAR1's LRR domain. Perception of uridylylated PBL2 by RKS1 stabilizes the activation segment of RKS1 which allosterically evicts ADP from ZAR1's nucleotide binding pocket and causes ZAR1's nucleotide binding domain to become more flexible[108]. Upon provision and binding of dATP/ATP in ZAR1's nucleotide binding pocket, ZAR1 is activated and

oligomerises into a multimeric structure observed by cryo-EM to be a pentamer referred to as a resistosome[109]. This pentameric structure is composed of a central α-helical barrel of ZAR1 CC domains[109], **Figure 1.8**. The authors postulate that this α-helical funnel may be inserted into the membrane leading to the formation of pores which directly lead to HR, though more evidence is needed to support the last step of this model. Such hypothesised membrane pore formation is not thought to be the outcome of all activated NLRs given the non-cell membrane associated sub-cellular localisation of some active NLRs and the known reliance on downstream signalling partners such as NRG1, discussed in 1.4.2.

Whilst oligomerisation had long be observed upon activation of mammalian NLRs, exemplified in the structure of 11-mer NLRC4-NAIP2 inflammasome, similar oligomerisation of plant NLRs was not observed until the publishing of the ZAR1 pentameric structure referred to as a resistosome. Interestingly oligomerization of both plant NLR N-terminal TIR and CC domains, discussed in 1.3.1 , and animal N-terminal CARD and PYDs has been demonstrated to be integral for downstream signalling[87,111–114]. In animal NLRs the formation of the inflammasome brings the N-terminal domains of the NLRs into close proximity providing a platform for the recruitment of active signalling components like procaspases[35,102,115]. Conversely the function of multimeric plant NLR structures for facilitating downstream signalling is still widely debated.
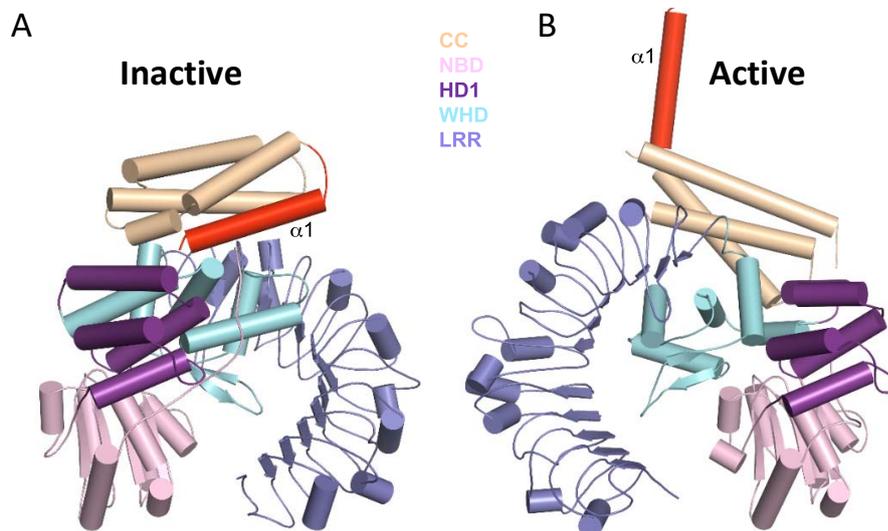
Figure 1.8 Structural remodelling of ZAR1 during activations. The inactive and active states are shown on th... f... ...erminal α-helix ...pted from Wang ...

Introduction

As ZAR1 only represents one type of NLR as a CC-NLR with indirect-decoy effector recognition, it is likely its activation mechanism is not representative of all plant NLRs. The model proposed alongside the structure of ZAR1 still leaves several unanswered questions. For example, the role of nucleotide hydrolysis, as the structure reveals that binding of both the effector-dependent activator and ATP leads to the protein being locked in the ATP-bound state. It is only through attaining further structures of full-length and multidomain NLRs representing a diverse array of effector perception mechanisms and NLR N-terminal domains that we will be able to gain robust insights into how the wider family of NLR receptors are activated.

1.4.2   Downstream signalling of NLRs

The signalling steps which link activation of NLRs to ETI and HR are still not comprehensively understood and remain a black box in our understanding of plant innate immunity. Current thinking suggests that there are two key regulators which work downstream of the different NLR classes: EDS1 (Enhanced disease susceptibility 1) for TNLs and NPR1 (Nonexpresser of PR Genes 1) for CNLs[116].

Nucleo-cytoplasmic EDS1 is a lipase like protein that works in conjunction with PAD4 (Phytoalexin Deficient 4) or structurally related SAG101 (Senescence-Associated Gene 101) as a heterodimer. EDS1 and its interactors PAD4 or SAG101 have been shown to play crucial roles in downstream ETI responses such as generation of ROS, salicylic acid and induction of a HR[117]. For example, induction of paired NLRs RRS1/RPS4 ETI responses have been shown to be dependent on EDS1 and SAG101[118]. Consequently, overexpression of EDS1 and PAD4 proteins leads to a dose dependent enhancement of resistance against bacterial pathogens whilst *eds1* and *pad4* mutants are highly susceptible to virulent *P. syringae* strains whose effector AvrRps4 activates RRS1/RPS4[119]. Whether EDS1 constitutively interacts with TNLs is highly debated. Whilst EDS1 has been reported to directly interact with various TIR-NLRs such as RPS4, RPS6 and SNC[120,121] others have found issues replicating this interaction (unpublished data and correspondence). Additionally, it has been hypothesised that upon effector perception the interaction between EDS1 and TNLs may be disrupted which then allows unbound EDS1 to function downstream[121] though more evidence is required to support this model. The ability of EDS1/PAD4/SAG101 to form both nuclear and cytoplasmic complexes has also been hypothesized to play an important role in transducing NLR signalling across different cellular compartments post effector perception[120].

Introduction

The importance of helper NLRs in NLR functioning is an exciting emerging field of study in plant immunity. The CNL$_{RPW8}$ NRG1(N Requirement Gene 1) [122] for example, has been recently hypothesized to function as a helper NLR for TNLs. This is exemplified in the evolution of NRG1 whose presence correlates with the presence of TNLs in different plant clades[123]. Studies have shown the requirement of NRG1 for the functioning of TNLs WRR4, Roq1, RPP1 and paired RRS1/RPS4 but not two CNLs tested in *N. benthamiana*[123,124]. Interestingly, whilst NRG1 was not required for CNLs RPS5 and MLA function, *nrg1* loss of function mutants showed slightly compromised RPM1 and RPS2 resistance. Interestingly, ADR1 which is thought to work downstream of CNLs is also a RPW8-like NLR and is phylogenetically related to the NRG1 clade[78]. The exact mechanistic role that helper CNL$_{RPW8}$ such as ADR1 and NRG1 play is still unknown but presents an exciting evolving facet of plant innate immunity research. Taken together this could suggest that RPW8-like NLRs are helper NLRs with ADR1 functioning downstream of CNLs and NRG1 downstream of TNLs whilst potentially contributing to a subset of CNLs functions[123]

Additionally the presence of an immune receptor signalling network has recently been revealed in initially *N. benthamiana* and now the wider Solanaceae family[125].The 'NRC' superclade of NLRs in Solanaceae consists of several distinct subclades of helper NLRs, NRC1-4, which function together with a highly diversified and expanded range of sensor NLRs including Rx, R8, Prf and Rpi-blb2 to perceive a broad range of pathogen effectors[126]. In this way, these NLRs form a signalling network with expanded sensor members relative to helpers and high levels of redundancy amongst helpers to form a robust and highly evolvable NLR signalling network.

Several NLRs have also been reported to directly interact with transcription factors or DNA upon pathogen recognition. For example, two TNLs SNC1 and RPS4 have been shown to directly interact with the positive immune regulator transcriptional activator bHLH84[127] whilst NLR N has been observed to interact with transcription factor SPL6[128]. A recent study also identified the ability of potato CNL Rx1 to interact with and melt DNA following effector recognition via the NLR's NB-ARC domain, however specificity of the target DNA region, for example via a binding partner, has not yet been clarified[129]. This has led to the hypothesis that nuclear NLR proteins may interact directly with transcriptional regulators to instigate an immune responses and transcriptional reprogramming upon pathogen perception.

Introduction

The exact downstream outputs of ETI and how these feed into plant immunity are not fully understood. However, a common downstream phenotype of NLR/ETI activation which is often but not always observed in induction of HR (Hypersensitive Response)[130,131]. The exact mechanism which leads to HR induction is not yet understood but this defence output is associated with the shrinkage of the cell cytoplasm, condensing of chromatin, swelling of mitochondria, chloroplast disruption and vacuolization[132]. This leads to a localised area of cell death at the point of pathogen infection thought to help constrict pathogen colonization of the host. New evidence is also now emerging about the requirements of both PTI and ETI in order to establish an ETI-induced HR. This has been demonstrated via unpublished data from the Jones lab showing lack of HR induction when recognised effectors are delivered without activation of PTI in Arabidopsis. Additionally, it has been recently demonstrated that MAPK3 and MAPK6 activation is required for ETI-induced HR[133]. This suggests that there may be a strong interplay between ETI and PTI. The signalling components controlling this output are yet to be defined but these findings present a fascinating challenge to the classical association of HR and ETI.

1.5 Investigating paired NLRs with RRS1/RPS4

The *A. thaliana* TIR-NLRs RRS1 (Resistance to *Ralstonia solanacearum* 1) and RPS4 (Resistance to *Pseudomonas syringae* 4) are a well-studied example of genetically linked paired NLRs. This divergently transcribed nuclear-localised NLR pair provide race specific resistance to a number of bacteria and fungi[134]. The roles of effector perception and activation of downstream signalling are split between these two NLRs. The 'sensor' NLR RRS1 is responsible for effector binding, which is facilitated through an integrated WRKY domain found towards the C-terminus of the NLR, whilst the 'executor' NLR RPS4 is responsible for activation of downstream ETI signalling. This spilt allocation of functional roles between two NLRs is seen in many genetically linked NLR pairs where one NLR contains an integrated domain for effector perception and the other facilitating downstream signalling[68,135,136]. Understanding the functional mechanism of the RRS1/RPS4 immune complex will therefore bring key insights to many paired NLRs.

1.5.1 The genomic and protein domain structure of RRS1 & RPS4

Divergently transcribed RRS1 and RPS4 are found in the *A. thaliana* genome in a head-to-head orientation separated by a small intergenic region of 264 bp thought to function as a

bidirectional promoter[137], **Figure 1.9A**. This type of genomic arrangement is also found with other paired NLRs for example Pikp1/Pikp2 and RGA4/RGA5 and may reflect the need for transcriptional co-regulation of the NLR pair members[68,69,134,138]. Additional RPS4 homologs are situated in the genome paired with other divergently transcribed NLRs carrying integrated domains, suggesting these linked pairs also co-evolved [137].

RRS1 and RPS4 both possess the canonical TIR, NB-ARC and LRR NLR domains. However, each NLR also both possesses additional non-canonical domains which reflect their split roles of 'sensor' and 'executor', **Figure 1.9B**. RRS1 has an integrated C-terminal WRKY domain which allows RRS1 to bind and perceive effectors which target WRKY transcription factors. More than 70% of Arabidopsis WRKY transcription factors are involved in defence, rendering these proteins as likely plant hubs for effector targeting to attenuate host immunity[61,62]. Given the strong involvement of WRKY proteins in defence signalling it is not surprising that bioinformatic studies have predicted that WRKY domains are amongst the most common non-canonical domains to be integrated into an NLR structure[52,53]. The WRKY domain integrated into RRS1 is a group III WRKY and contains the conserved DNA binding motif 'WRKYGQK'[65].

Beyond the WRKY domain, RRS1 possess a C-terminal extension with little homology to any other proteins. This extension, referred to as Domain 6 (Dom6), varies between alleles of RRS1. The RRS1 alleles investigated in this study are RRS1-R, found in *A. thaliana* accession Ws-2 and RRS1-S found in the Col-0 accession, **Figure 1.9B**. These NLRs are near-identical in their amino acid sequence except for their C-terminal Dom6 regions. RRS1-S's C-terminal extension, referred to as Dom6S, contains a premature stop codon relative to RRS1-R leading to a C-terminal extension of only 21 amino acids. Dom6 of the RRS1-R allele (Dom6R) however contains a further 83 amino acids beyond the end of Dom6S, with a total C-terminal extension of 104 amino acids. This extension is vital for supporting extended effector recognition capabilities compared to RRS1-S as discussed in 1.5.2. RRS1 also contains a region of 322 amino acids between the LRR and WRKY domains referred to as Domain 4 (Dom4). This domain shows little homology to other known protein domains and has no assigned biochemical function. Reports have suggested that Dom4 contains a leucine-zipper motif which appears to be conserved in RRS1 alleles from many accessions of *A. thaliana*[139]. Putative leucine zippers have also been found in WRKY 18, WRKY 40 and WRKY 60 and play a role in facilitating physical interactions between WRKY proteins[140], however the role of this motif in RRS1 function is not clear.
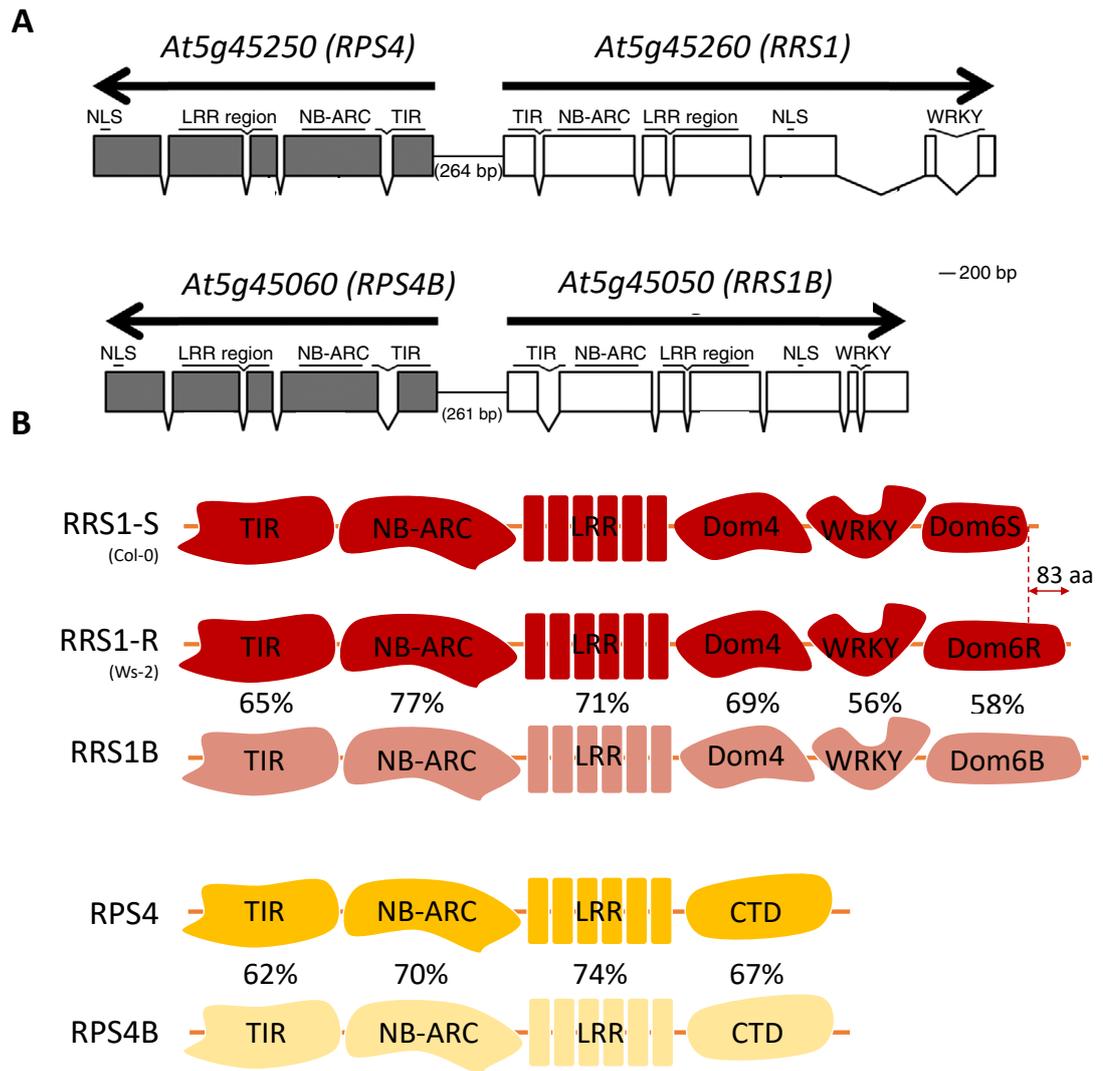
Figure 1.9 Gene pair and protein modular domain structure of RRS1/RPS4 and paralogous RRS1B/RPS4B. (A) Schematic representation of RRS1/RPS4 and RRS1B/RPS4B gene pairs. Exons are depicted as boxes and domains references above relevant regions. Black arrows indicate reading frame direction, NLS (nuclear localisation signal), figure adapted from Saucet et al, 2015[141]. **(B)** Protein domain structure of RRS1 alleles, RPS4 and paralogous RRS1B/RPS4B. Amino acid sequence identity is listed as a percentage between the comparable domains.

In addition to the canonical NLR domains, RPS4 contains a non-canonical 338 amino acid C-terminal domain (CTD) extension which is thought to play an important role in mediating interactions with RRS1. RPS4's CTD shows little homology with any other known protein domains, limiting protein-protein interaction modelling work that can be done with this NLR. Unlike RRS1, the alleles of RPS4 found in Col-0 and Ws-2 which function with RRS1-R and RRS1-S are identical.

1.5.2 Effector recognition capabilities of RRS1 & RPS4

Different allelic variations of RRS1 provide RPS4-dependent recognition of a variety pathogens including *P. syringae*, *R. solanacearum* and *C. higginsianum,* **Figure 1.10***.* Widely distributed Gram-negative *R. solanacearum* is a soil borne bacteria which causes bacterial wilt disease in a variety of Solanaceae crop species including potato, tomato and aubergine as well as soy bean and banana[142]. *P. syringae* is an epiphytic Gram-negative bacterium which can infect a wide range of herbaceous and woody crops such as tomato causing bacterial speck disease. *P. syringae* pathovar DC3000 is also widely used to study plant-pathogen interactions on model species *A. thaliana*[143]. C. *higginsianum* is an ascomycete pathogen which causes anthracnose disease on many economically important cruciferous plants such as Brassica and Raphanus species as well as model species *A. thaliana*[144].
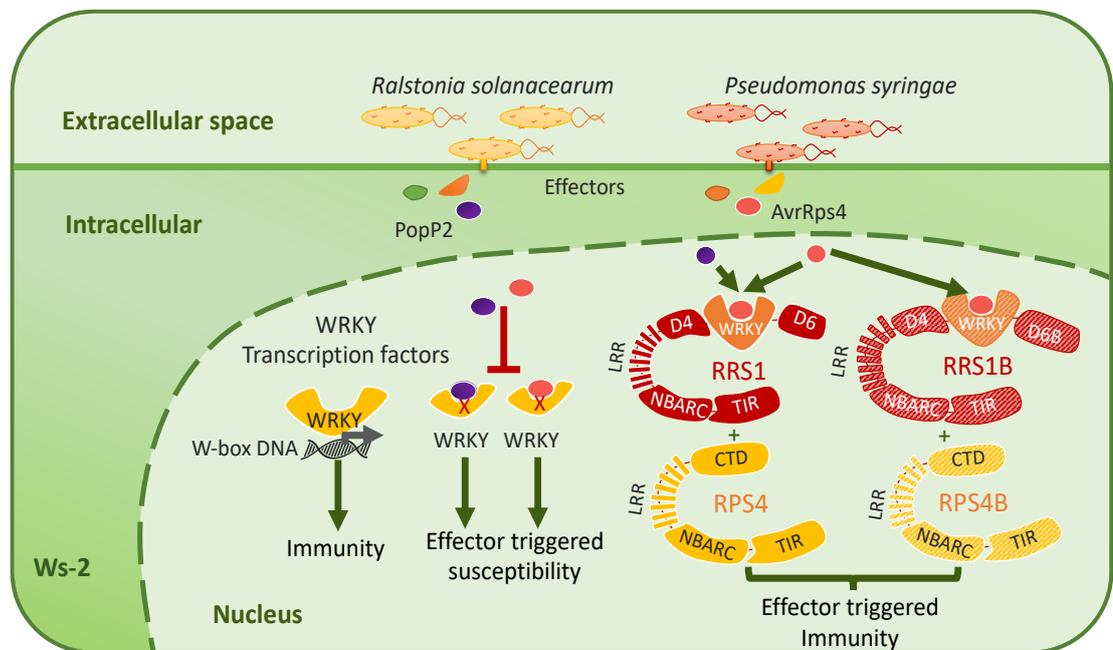
Figure 1.10 The RRS1/RPS4 NLR complex system. Effectors PopP2 (purple circle) from *R. solanacearum* and AvrRps4 (pink circle) from *P. syringae* target WRKY transcription factors involved in plant immunity signalling to trigger effector susceptibility. PopP2 and AvrRps4 are susceptible to molecular detection by the NLR pairs RRS1/RPS4 and RRS1B/RPS4B which bind the effectors via RRS1's or RRS1B's integrated WRKY domain. Activation of these immune complexes leads to induction of effector triggered immunity (ETI). RRS1 alleles can recognise AvrRps4 and PopP2 but RRS1B can only perceive AvrRps4.

Introduction

In the presence of RPS4, RRS1-R is capable of perceiving the effectors AvrRps4 from *P. syringae*[137], PopP2 from *R. solanacearum*[145] and an unknown effector from the fungus *C. higginsianum*[134,146]*,* **Figure 1.10***.* AvrRps4 is a type III secreted effector deployed by *P. syringae f. sp. pisi*. Upon secretion into the plant cell AvrRps4 is cleaved between Gly133 and Gly134 to release a C-terminal anti-parallel coiled coil truncation of residues Gly134-Gly221 referred to as AvrRPS4$_c$[147]. AvrRps4$_c$ is capable of inducing resistance though *in planta* processing is not necessary for perception of AvrRps4 as the un-cleaved AvrRps4 R112L is still recognised[147]. Processing of AvrRps4 may take place in the chloroplast to which AvrRps4 is targeted via a predicted chloroplast transit peptide also utilised by *P. syringae* effector HopK1[148]. This might suggest that whilst AvrRps4$_c$ recognition by RRS1 is thought to occur in the nucleus, AvrRps4 may have an additional virulence function in the chloroplast. Whether the N- and C-terminal regions of AvrRps4 have distinct virulence activities is not known but the N-terminal fragment of AvrRps4 has been shown to interact with EDS1 indicating it may play a role beyond acting as a chloroplast localization transit peptide[149].

The exact virulence mechanism of this effector is not fully understood as the crystal structure of AvrRps4$_c$ revealed no obvious catalytic sites or biochemical activity[150]. It is therefore hypothesised that AvrRps4$_c$ interferes with host immune signalling by direct interactions with WRKY transcription factors. Coimmunoprecipitation data has shown that AvrRps4 can interact with a variety of WRKY proteins including WRKY41/70/60/33 but EMSA studies showed the effector did not disrupt the ability of WRKY proteins to bind their W-box DNA element. As discussed in Chapter 5, the inability to detect WRKY-DNA binding interaction disruption upon AvrRps4 treatment may be due to insensitivity of the EMSA technique. There are two important potential AvrRps4-RRS1 interaction surfaces which have been highlighted in the structure of AvrRps4$_c$. Firstly, residues E175 and E187 are found in an electrostatic negative patch positioned both on the helices on either side of AvrRps4's β-turn and on the β-turn itself and are predicted in protein-protein interaction modelling studies, to directly interact with lysine residues in RRS1's 'WRKYGQK' motif[150], **Figure 5.2**. Secondly, a region at the N-terminus of truncated AvrRps4$_c$ referred to as the KRVY motif (AvrRps4 K135-Y138) is essential for RRS1's recognition of AvrRps4$_c$ but not binding. As the 'KRVY' motif is not in the electron dense region of the structure of AvrRps4$_c$ protein-protein interaction modelling of this region with RRS1's WRKY has not been possible. Interestingly mutating the KRVY motif to alanine residues (KRVY/AAAA) results in the generation of an AvrRps4$_c$ protein which is capable of binding to the WRKY domain of

Introduction

RRS1 but not activating defence. This suggests that whilst binding of AvrRps4$_c$ is required for RRS1 recognition it is not sufficient to activate defence. The reason for this is not known and will likely not be fully understood until we understand the structural basis of RRS1's recognition of AvrRps4. The structure of AvrRps4 and interactions with RRS1's WRKY domain is discussed in further detail in 5.1.

PopP2 is as an acetyltransferase enzyme of the YopJ family[151]. PopP2 acetylates conserved lysine residues in the 'WRKYGQK' motif of defence-related WRKY transcription factors and RRS1 abolishing the DNA binding capabilities of these proteins[65,66]. Similarly to AvrRps4, coimmunoprecipitation assays have demonstrated that PopP2 can bind of WRKY transcription factors including WRKY41/70/60/33. As an acetyltransferase the catalytic core residue of C321 is key for virulence function and recognition by RRS1. PopP2 C321A mutants are both incapable of acetylating WRKY transcription factors and RRS1 WRKY. The recently published crystal structure of the RRS1 WRKY$_{E1195-T1273}$-PopP2 complex revealed that a core lysine residue in RRS1'S WRKY motif, K1221, inserts directly into the active site of PopP2 containing residue C321. The interaction surface between RRS1 WRKY and PopP2 is discussed further in 5.1.

Whilst RRS1-R confers recognition of AvrRps4, PopP2 and an unknown effector from the fungus *C. higginsianum*[134,146], RRS1-S is only capable of perceiving AvrRps4. Experiments truncating Dom6R of RRS1-R back to the Dom6S boundary have shown that RRS1-R's extended recognition capabilities are facilitated by the Dom6R extension. Whilst RRS1-S is still capable of binding and being acetylated by PopP2 this is not transduced into an ETI output as observed in RRS1-R. How Dom6R confers such expanded recognition capabilities is not fully understood. Interestingly acetyl-lysine mimic substitutions of the lysine residues in RRS1-R's WRKY motif to 'WR**Q**YGQ**Q**' activates RPS4-dependent defence activation in the absence of an effector but not RRS1-S. This indicates that Dom6R must be involved in specific intra- and inter-domain interactions which integrate this acetylation event into ETI induction which Dom6S cannot facilitate.

1.5.3   RRS1B/RPS4B: The paralogous B pair

RRS1 and RPS4 (A pair) are linked to second pair of similar divergently transcribed NLRs called RRS1B/RPS4B which are found in both Col-0 and Ws-2 *A. thaliana* accessions[141]. Unlike the A pair, RRS1B/RPS4B can only recognise AvrRps4 and not PopP2 or *C.*

higginsianum, **Figure 1.10**. RRS1B and RPS4B have the same overall modular domain structure as the A pair but only share 64.7% and 65.3% amino acid sequence identity with RRS1-R and RPS4 respectively. This similarity goes down to 56% and 58% sequence identity when comparing the WRKY and Dom6 regions of RRS1 and RRS1B respectively, **Figure 1.9B**. The possible activation mechanism of RRS1/RPS4 will be discussed in 1.5.4 but domain swapping experiments between the B pair and A pair of RRS1/RPS4 suggests these paralogous pairs function in mechanistically distinct ways. For example, unlike RRS1 the Dom6 region of RRS1B, termed Dom6B, is required for RRS1B's perception of AvrRps4. Additionally, truncation of the WRKY Dom6 region results in RPS4-dependent autoactivity in the case of RRS1 but not RRS1B[152] (and unpublished data from Dr Yan Ma). This implies the two NLR pairs are perceiving effectors and activating ETI in a distinct manner. Comparing the activation mechanisms of A and B pair RRS1/RPS4 therefore provides an intriguing tool set for understanding how one effector, AvrRps4, is recognised and activates different NLRs.

1.5.4   Activation mechanism of the RRS1/RPS4 complex

Qualitative biochemical investigations, HR assays and microscopy studies have begun to reveal the activation mechanism of RRS1/RPS4. However, structural insights into the functioning of this complex are now required to further our understanding.

In the absence of RRS1, RPS4 can induce a weak autoimmune response upon transient expression in *N. tabacum*. In order to prevent autoimmunity, RRS1 negatively regulates RPS4 in the absence of effectors. RRS1 itself is kept in the inactive state by RRS1's WRKY domain negative regulation of Dom4. This can be demonstrated by the RPS4-dependent autoactive phenotype observed when RRS1 ΔWRKYDom6 truncations are transiently expressed in *Nicotiana tabacum*[152]. This autoactivity is dependent on RRS1's Dom4 as further deletion of Dom4 (RRS1 ΔDom4-WRKY-Dom6) abolishes this autoactive phenotype. Suppression of Dom4 appears to be specific to RRS1's coevolved integrated WRKY, as substitution of the RRS1 WRKY with similar DNA binding domains such as the bacterial DNA binding domain LexA, AtWRKY41 or the WRKY domain of RRS1B result in generation of an autoactive complex[153]. Bimolecular fluorescence complementation (BiFC) studies have shown that in the absence of effectors the N- and C-termini of RRS1 Dom4-WRKY-Dom6 are in close proximity. Binding of AvrRps4 to RRS1's WRKY domain is observed to disrupt this interaction potentially relieving the WRKY domain's negative regulation of Dom4. However,

activation of the complex in the presence of AvrRps4 cannot be explained by this inter domain interaction disruption alone as mutant AvrRps4$_{KRVY/AAAA}$ is also observed to disrupt this interaction but not activate defence[152]. What these additional interaction changes are is still not understood and likely will require structural information to decipher. What has been observed however is the ability of Dom4 to associate with RPS4's CTD. It is therefore predicted that derepressed RRS1 Dom4 activates RPS4 via its CTD. The changes in intramolecular interactions within RPS4 which occur upon activation by RRS1 are not fully understood. What has been demonstrated however, is the ability of the RPS4 TIR domain to activate a cell death response when transiently overexpressed in isolation. This has led to the proposed model that upon effector binding by RRS1, suppression of RPS4 TIR is released allowing the formation of a signalling-competent RPS4 TIR domain homodimer[114].

The mechanism by which AvrRps4 activates the RRS1/RPS4 complex appears to differ from PopP2. For example, PopP2 was not observed to disrupt interactions between RRS1 Dom4 and WRKY-Dom6R in BiFC or coimmunoprecipitation assays. It is likely therefore that distinct changes are occurring upon binding of AvrRps4 and PopP2. These distinctions likely involve Dom6R which has been shown to be specifically required for PopP2 perception. Further evidence for distinct AvrRps4 and PopP2 RRS1 activation mechanisms comes from the identification of mutants S983F and E1070K in RRS1-R Dom4, and C887Y in RPS4 CTD which impair PopP2 triggered HR but not AvrRps4[152]. This indicates the requirement of distinct residues within RRS1 and RPS4 for PopP2 recognition implying the mechanism of intra and inter-molecular changes occurring within the RRS1/RPS4 complex differs between effectors. The role these residues play in the activation of RRS1/RPS4 is still not understood and will likely require gaining structural information on these domains to assign precise functional roles.

The AvrRps4 activation mechanism of the A pair and B pair of RRS1/RPS4 also appears to differ. For example, whilst RRS1B required Dom6B for AvrRps4 induced activation truncated RRS1 ΔDom6 is still capable of inducing AvrRps4-dependent HR in the presence of RPS4 (unpublished data Dr Yan Ma). Additionally, whilst Dom4 and Dom6 of RRS1 were observed to interact in BiFC studies and coimmunoprecipitation assay no such interaction was observed between RRS1B Dom4 and Dom6B[152]. This suggests that RRS1B Dom4-WRKY-Dom6B has a different conformation from this region in RRS1 suggesting different inter- and intra-domain interactions are involved.

Introduction

What intra- and inter-domain interactions changes occur within RPS4 once RRS1 suppression of RPS4 is released is not understood. RPS4's NB-ARC domain is thought to play an important role in induction of ETI as mutations in RPS4's nucleotide binding P-loop render the RRS1/RPS4 complex non-functional[114]. Conversely a functioning RRS1 P-loop is not required for complex activation highlighting RPS4's role as the 'executor' NLR in the pair[114]. In addition to the functioning RPS4 NB-ARC domain, heterodimerization of the RRS1/RPS4 TIR domains has been shown to be crucial in the maintenance of a functioning effector recognition complex. The TIR domain of RPS4 can activate defence in an effector independent manner whilst RRS1 TIR does not, and can inhibit RPS4 constitutive action of defence via its heterodimerization interface[114]. This interface involves the residues within the αA and αE helices and EE loops of RPS4 and RRS1 TIR and the DD loops of RRS1 TIR, **Figure 11**. At the core of this interface resides a stabilising stacking interface of Histidine residues (RPS4 H34 and RRS1 H26) as well as interactions between the serines in each TIR which precede this histidine residue (SH motif) which together stabilise the heterodimer, **Figure 11.B**. Mutation of this SH motif also prevented homodimerization of RPS4 TIR and induction of RPS4-dependent cell death highlighting the importance dimerization of TIRs plays in induction of RRS1/RPS4 dependant ETI[114].
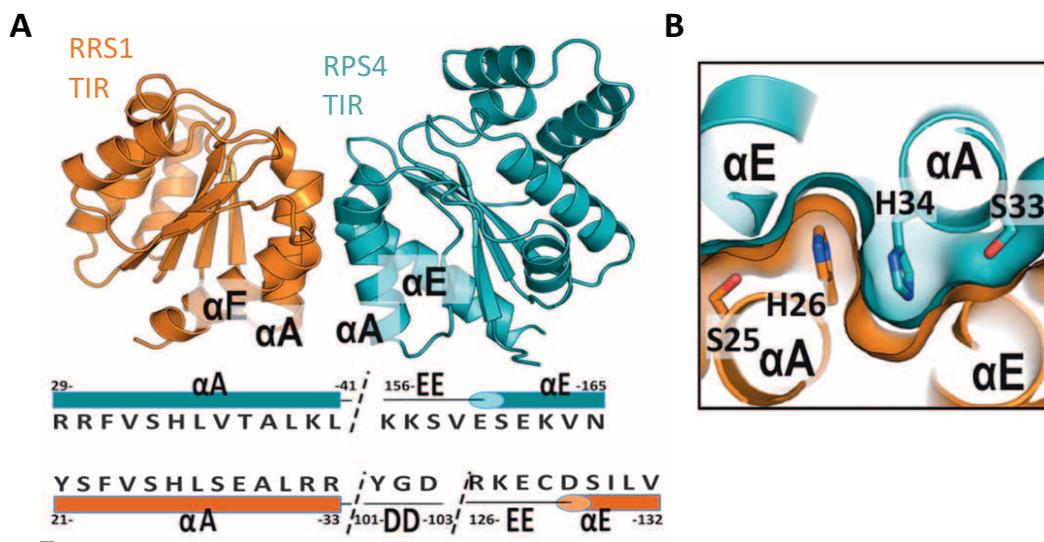
Figure 1.11 Heterodimerization interface of RRS1 and RPS4 TIR domains. (A) Crystal structure of RRS1 TIR (orange) and RPS4 TIR (teal) heterodimer interface shown in cartoon representation. Residues contributing to interface are highlighted in the amino acid and secondary structure below. **(B)** The position of serine and histidine residues within the SH motif heterodimerization interface. Figure taken from Williams et al, 2014[114]

It is predicted that upon effector perception, RRS1 TIR's inhibition of RPS4 is released enabling the formation of signalling competent RPS4 TIR homodimer. The exact role the TIR

domains play in instigating ETI upon activation by effector perception is not fully understood. It has recently been hypothesised that TIR domains may possess NADase activity. These investigations came following the identification that mammalian SARM1(sterile alpha and Toll/interleukin-1 receptor motif-containing 1) TIR domain shows intrinsic NADase activity cleaving NAD^+ (nicotinamide adenine dinucleotide) into ADPR (ADP ribose), cyclic ADPR and nicotinamide[154]. Whilst an ability to cleave NAD^+ into nicotinamide and ADPR has now been demonstrated for the TIR domains of NLRs RUN1 and L6, albeit at distinctly lower enzymatic levels than observed with SARM TIR domain, RPS4 TIR domains were not found to have such biochemical activity[92,93]. Conceivably the correct substrate for RPS4 TIR domain cleavage may have not yet been identified or RPS4 TIR may function in a disparate manner that L6 and RUN1 potentially via signalling through cooperative assembly formation (SCAF)[85]acting a structural platform.

1.6 Project aims and objectives

If we are to be in a position to engineer synthetic NLRs with the ability to perceive currently unrecognised effectors we need to gain a comprehensive understanding of how NLR receptors function mechanistically. Our mechanistic understanding of plant NLRs is still hindered by a lack of structural information of these receptors. The structures we do have of plant NLRs are largely of single domains in isolation and predominantly limited to the N-terminal CC or TIR domains. Though the publishing of the ZAR1 structure provided a significant leap forward in our understanding of plant NLR activation[108,109], this structure only represents one type of NLR as a CC-NLR with indirect-decoy effector recognition. Given the great diversity of known plant NLRs and effector perception strategies it is highly unlikely ZAR1 will provide a universal mechanism through which all NLRs function. This is important to recognise as only in attaining a comprehensive understanding of NLR receptor mechanisms will the field have the foundation to enable effective engineering or NLRs with expanded effector recognition capability. It is therefore vital for the progress of the plant NLR field that we gain structural and quantitative biochemical insights of other NLRs in order to understand the structural basis through which NLRs function and validate and challenge our current models of thinking.

The lack of multidomain structures of NLRs beyond ZAR1 is likely not the result of a lack of effort in the field. Instead this reflects the notorious difficulties of expressing these proteins heterologously. Both the animal and plant NLR fields have struggled for decades to produce

Introduction

NLRs in various expression systems with breakthroughs of full-length NLR expression for structural studies only achieved in 2013 and 2019 respectively[101,108,109].

The primary aim of this study is to gain biochemical and structural insights into the mechanistic functioning of the RRS1/RPS4 NLR pair.

Within this aim, the first objective of this study is to evaluate a diverse range of heterologous protein expression systems for the ability to produce RRS1 and RPS4 protein, both full-length and domain truncations. By completing a comprehensive study of prokaryotic and eukaryotic systems including plant-based expression systems, this study aims to evaluate the expression requirements of RRS1 and RPS4. In addition to providing a pipeline for RRS1/RPS4 protein production, the insights this work will bring will also provide a valuable foundation for other studies of a similar nature looking to express plant NLRs.Through completing this first objective, the study would establish a pipeline for expressing soluble RRS1 and RPS4 protein to utilise in downstream structural and quantitative biochemical studies. As such, the second objective of this work is to begin to utilise these identified pipelines to produce RRS1 and RPS4 protein to begin to dissect the mechanism through which these NLRs recognise effectors and activate defence responses utilising techniques such as analytical gel filtration and Blue Native PAGE.

The third objective of this study focusses on understanding the structural basis of RRS1 WRKY domain's recognition of effectors. Given the diversity of RRS1 variations we have in both the A and B pair of the protein, this system provides a unique opportunity to understand the inherent flexibilities in effector recognition and activation of integrated domain NLRs. This study looks to gain a further understanding of this plant/NLR interface than biological work has previously provided through utilising structural biology and quantitative and qualitative biochemical techniques. These techniques include SPR (Surface plasmon resonance) and analytical gel filtration.

Drawing information together from work looking at evaluating the biochemistry of multi domain RRS1 and RPS4 protein with a deeper understanding of how effectors are perceived by RRS1 will bring fresh insights into the mechanistic behavior of this NLR pair. In this way this study will provide an expanded understanding of plant-pathogen interactions at the molecular level and contribute towards a growing body of the work in the field which can be utilised to strengthen global food security.

Materials & Methods

2.1 Materials

2.1.1 Chemicals and Reagents

Chemicals used in this study were purchased from Sigma-Aldrich, Melford Laboratories, Honeywell Fluka, Avantor or Thermo Fisher Scientific unless otherwise stated.

2.1.2 Antibiotics

Stock and selective media working concentrations for *Escherichia coli* as well as solutes are as listed in **Table 2.1**. Stock solutions of Chloramphenicol and Gentamycin were stored at -20^{0}C whilst all other antibiotic stocks were stored at 4^{0}C. Stock solutions were filter sterilised using a 0.3 µm Ministart® filters prior to storage.

Table 2.1 Antibiotics used in this study

Antibiotic	Stock concentration	Working Concentration
Carbenicillin	100 mg/ml in H$_2$0	100 µg/mL
Chloramphenicol	34 mg/ml in ethanol	34 µg/mL
Gentamycin	10 mg/ml in H$_2$0	20 µg/mL
Kanamycin	30 mg/ml in H$_2$0	30 µg/mL
Rifampicin	10 mg/ml in methanol	50 µg/mL

2.1.3 Bacterial Media: Lysogeny Broth (LB)

Lysogeny broth media (LB) (LB broth Miller, Formedium) consisting of 1.0% (w/v) tryptone, 0.5% (w/v) yeast extract and 1.0% (w/v) sodium chloride at pH 7.0, was made up in de-ionised water prior to autoclaving. Solid LB media plates were prepared with 1% (w/v) microbiology grade agar before autoclaving. During transformation electro-transformed or chemically transformed cells were rescued with L media, which has a lower salt content of 0.5% (w/v) compared to 1.0% (w/v) of LB broth.

2.1.4 Bacterial Media: Autoinduction media (AIM)

AIM (Auto-induction media) was prepared with 1% (w/v) Yeast extract, 0.5% w/v Tryptone, 0.035% (w/v) auto induction medium micro element mix (Formedium), 0.33% (w/v)

ammonium sulphate, 0.68% potassium dihydrogen phosphate, 0.71% (w/v) Disodium hydrogen phosphate, 0.05% (w/v) glucose, 0.2% (w/v) α-lactose, 0.015% (w/v) magnesium, trace elements. Media was prepared in de-ionised water and autoclaved prior to use.

2.1.5   Bacterial Media: Power Broth

PowerBroth™ media powder was purchased from Molecular Dimensions and prepared as 5.2% (w/v) PowerBroth™ powder with 0.4% (v/v) glycerol.

2.1.6   Sf9 Baculovirus media

Spodoptera frugiperda clonal isolate Sf9 insect cells were grown in Sf-900™ II medium purchased from Invitrogen. Sf-900™ II medium is a serum-free and protein-free insect cell culture medium which has been optimised for the growth and maintenance of Sf9 and Sf21 *S. frugiperda* cells as per OPPF's standard operating procedure for 'Insect Expression' produced by Joanne Nettleship (https://www.oppf.rc-harwell.ac.uk/OPPF/protocols/).

2.1.7   Bacterial strains: *Escherichia coli*

For plasmid preparation and storage cloning purposes bacterial strains DH10B, DH5α and ElectroMAX™Stbl4™ were used. High-throughput protein expression screening at OPPF (Oxford Protein Production Facility) was conducted in Lemo21(DE3) and Rosetta™(DE3) pLyS *E. coli* cells. Rosetta™(DE3) pLyS and Rosetta™ 2(DE3) strains were grown in media supplemented with chloramphenicol in order to maintain the pRARE2LysS and pRARE2 plasmids respectively. Sequential protein expression experiments extended to included SHuffle®T7, BL21(DE3) and Rosetta™ 2(DE3) *E. coli* strains. The genotypes and suppliers of all *E. coli* cell strains used in the study can be found in **Table 2.2.**

Table 2.2 *E. coli* strains and genotypes used in this study

E. coli strain	Genotype
BL21(DE3) (NEB)	*fhuA2 [lon] ompT gal [dcm] ΔhsdS*
DH10B (ThermoFisher)	F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ– *rps*L(StrR) *nup*G

DH5α (ThermoFisher)	F– φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17(rK–, mK+) *pho*A *sup*E44 λ– *thi*-1 *gyr*A96 *rel*A1
ElectroMAX™Stbl4™ (Invitrogen)	*mcr*A Δ(*mcr*BC-*hsd*RMS-*mrr*) *rec*A1 *end*A1 *gyr*A96 *gal*⁻*thi*-1 *sup*E44 λ-*rel*A1 Δ(*lac-pro*AB)/F' *pro*AB⁺ lacI�q ZΔM15 *Tn*10 (TetR)
Lemo21(DE3) (NEB)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo*(CamR) *λ DE3 = λ sBamHlo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5* pLemo = pACYC184-*PrhaBAD-lysY*
Rosetta™ 2 (DE3) (Merck Millipore)	F- *ompT hsdS*B(rB- mB-) *gal dcm* (DE3) pRARE2 (CamR)
Rosetta™(DE3) pLyS (Novagen)	F⁻ *ompT hsdS_B* (r_B⁻ m_B⁻) *gal dcm* (DE3) pLysSRARE (CamR)
SHuffle® T7 (NEB)	F´ *lac, pro, lacIq / Δ(ara-leu)7697 araD139 fhuA2 lacZ::T7 gene1 Δ(phoA)PvuII phoR ahpC* galE (or U) galK λatt*::pNEB3-r1-*cDsbC*(SpecR, *lacIq*) *ΔtrxB rpsL150*(StrR) *Δgor Δ(malF)3*

2.1.8 Bacterial Strains: *Agrobacterium tumefaciens*

Two strains, GV3101 and AgL1, of *Agrobacterium tumefaciens* strains were used for transient transformation of both *Nicotiana benthami*ana and *Nicotiana tabacum*. Both strains carried rifampicin resistance with GV3101 carrying additional gentamycin resistance for maintenance of pMP90 helper plasmid.

2.1.9 DNA Oligonucleotides

Oligonucleotides were synthesised by Merck Millipore on 0.025 mg scale unless otherwise stated using a desalting purification method to remove small-molecule impurities. Lyophilised primers were resuspended in ddH$_2$O to a final concentration of 100 µM and stored at -20^{0}C. 100 µM stocks were diluted to a working concentration of 20 µM and also stored at -20^{0}C. A full list of oligonucleotide primers used in this study can be found in Appendix 1.

2.1.10 DNA fragment synthesis

DNA fragments used in this project were synthesised from Twist Biosciences.

2.1.11 Plant material: *Nicotiana spp*

Nicotiana benthamiana and *Nicotiana tabacum* cultivar 'Petit Gerard' were grown in long days (16 hours light, 8 hours dark) under controlled conditions at 24^0C prior to *A. tumefaciens*-mediated transient transformation of leaf tissue.

2.1.12 Plant material: *Arabidopsis thaliana*

Approximately 50mg of transgenic *Arabidopsis thaliana* seed was sterilised in 1ml of sterilisation liquid (1% NaClO, 0.1% Tween20) for 5 mins with periodic agitation. Seeds were spun down at 1000 rpm for 30 seconds and the sterilisation liquid supernatant discarded. The sterilised seeds were then washed twice with ELGA water spinning down seeds in between each wash at 1000 rpm for 30 seconds. In a flow hood seeds were then sowed in 100 ml of 1% Murashige and Skoog (MS) media[155] in 250 ml conical flasks. Seeds was grown in shaking liquid culture at 100 rpm for ~2 weeks in a controlled growth room at 22^0C under long day conditions (16 hours light, 8 hours dark) before tissue was harvested and snap frozen in liquid nitrogen and stored at -80^0C or used fresh.

2.2   Molecular Biology

2.2.1   Expression Vectors

Vectors used for expression in *E. coli*, Sf9 Insect cell, *N. benthamiana*, *N. tabacum* and wheat germ cell free heterologous systems in this study can be found in **Table 2.3.** Maps of all protein expression vectors used in this study can be found in Appendix 2. A list of all cloned plasmids used in this study can be found in Appendix 4.

Table 2.3 Expression vectors used in this study

Vector	Expression System	Feature	Cloning Method	Antibiotic Resistance	Reference
pOPINF	Bacterial/ Insect	3C protease cleavable N-terminal histidine tag	Infusion®	Carbenicillin	Berrow et al,2007[156]

POPINS3C	Bacterial/ Insect	3C protease cleavable N-terminal histidine SUMO tag	Infusion®	Carbenicillin	Bird, 2011[157]
pOPINM	Bacterial/ Insect	3C protease cleavable N-terminal histidine MBP tag	Infusion®	Carbenicillin	Berrow et al, 2007[156]
pOPINA	Bacterial/ Insect	No affinity tag	Infusion®	Kanamycin	OPPF
pOPINE	Bacterial/ Insect	Non-cleavable C-terminal histidine tag	Infusion®	Carbenicillin	Berrow et al, 2007[156]
pEU-GG	Wheat Germ	No affinity tag, modified from pEU-E01-MCS	Golden Gate	Carbenicillin	Cell Free Sciences (Modified by
pICSL86977	*Nicotiana spp (Agrobacterium)*	CaMV35s promoter Level 1 acceptor	Golden Gate	Kanamycin	Mark Youles TSL Synbio)
pMCSG7	Bacterial	N-terminal TEV protease cleavable His-tag	Ligation independent cloning	Carbenicillin	Stols et al, 2002[158]

2.2.2 Polymerase Chain Reaction (PCR)

Polymerase chain reactions (PCRs) were carried out to amplify regions of DNA for cloning purposes and to confirm the presence of target plasmids in transformed bacterial colonies referred to as colony PCRs. For colony PCRs, TAQ DNA polymerase (NEB) was used whilst cloning PCR products were generated using high fidelity Phusion DNA polymerase (NEB) or KAPA HiFi DNA Polymerase (Roche). Reaction mixes were setup as per manufactures instructions and reactions performed in a C1000 Touch thermocycler (Bio-Rad). Plasmid DNA or cDNA was used for PCR template for cloning PCR product purposes whilst for colony PCRs single transformed colonies were diluted in 30 µl of water and 1 µl of this solution added to each 15 µl PCR reaction. Annealing temperatures and elongation times were optimised based on the properties of the primers and the amplification fragment. PCR program parameters for colony PCRs and Phusion PCRs can be found in **Table 2.4**.

In the case of full-length cDNA of RRS1-R it was found that a significantly longer than predicted elongation time of 6 minutes was optimal for amplification of fragment.

Table 2.4 PCR parameters for Colony PCR (Left) and Phusion PCR (Right)

PCR step	Temp (°C)	Time (sec)	Cycle Number	PCR step	Temp (°C)	Time (sec)	Cycle Number
Priming	95	120	1	Priming	98	300	1
Denaturing	95	30		Denaturing	98	10	
Annealing	55	15	29	Annealing	57-60	30	30
Extension	72	30s/1KB		Extension	72	15s/1KB	
Final extension	68	300	1	Final extension	72	600	1
Hold	16	∞	1	Hold	16	∞	1

2.2.3   Gel Electrophoresis

1% (w/v) agarose gels in TAE (40 mM Tris pH 7.6, 20 mM acetic acid, 1 mM EDTA, 0.5 µg/ml ethidium bromide) were used to separate DNA fragments according to size following PCR or restriction enzyme digest. DNA samples were mixed with 10x loading dye (40% v/v glycerol, 0.5% (w/v) SDS, 10 mM EDTA, Orange G) prior to gel loading with an accompanying well loaded with DNA ladder. Ladders used in this study were Quick-Load Purple 100bp DNA ladder (NEB) or Quick-Load Purple 2-Log DNA ladder (NEB). Samples were run using horizontal electrophoresis at 90-130V for 15-20 min before imaging using a UV transilluminator (Bio-Rad). DNA bands for cloning purposes were then excised using a razor and purified from the agarose gel using QIAquick DNA Gel extraction kits (QIAGEN) according to the kit protocol.

2.2.4   Golden Gate Cloning

Golden Gate cloning utilises Type IIS endonucleases and T4 DNA ligase for the simplified assembly of multiple DNA fragments and standard parts into a destination vector in a few reactions[159,160]. Golden Gate cloning was used in this study to generate constructs for *in planta* expression and wheat germ expression. This study used the Type IIS endonuclease *BsaI* which cleaves outside of its 5'GGTCTC3' recognition site to leave a 4 bp overhangs after digestion. Constructs were assembled using the TSL Synbio long protocol using a

reaction mix consisting of; 100 ng of acceptor vector, 1x T4 Ligase buffer (NEB), 0.1 mg/ml

Bovine Serum Albumin, 20 U/µl T4 DNA ligase (NEB), 1 U/µl BsaI-HF (NEB) and insert

fragments in 2:1 insert:acceptor vector go through the digestion-ligation reaction protocol

shown in **Table 2.5** Following the digestion-ligation protocol, the reactions were desalted

using a sepharose column before transformation into *E. coli*.

Table 2.5 Long Protocol Digestion-Ligation reaction protocol

Time	Temperature	Number of Cycles
20 sec	37^0C	1
3 min	37^0C	
4 min	16^0C	26
5 min	50^0C	1
5 min	80^0C	1
∞	16^0C	

2.2.5   Infusion Cloning

Infusion (Clontech) cloning was used for the generation of the *E. coli* and insect cell

expression pOPIN vector constructs developed by Oxford Protein Production Facility

(OPPF), vector maps of heterologous protein expression vectors found in Appendix 2.

Infusion cloning is a ligation-independent method based on the annealing on

complementary ends of a PCR insert fragment and linearized cloning acceptor vector[70].

Firstly, PCR products with specific primer extensions are generated to allow for the direct

cloning of these PCR fragments into pOPIN vectors. Prior to cloning, pOPIN acceptor vectors

must be linearized with restriction enzymes to allow for insertion of PCR fragment.

Acceptor pOPINF/S3C/M vectors were linearized by digestion with *KpnI* and *HindIII*, pOPINE

with *NcoI* and *PmeI* and pOPINA with *NcoI* and *DraI* for 2 hours at 37^0C before gel

purification (see section 2.2.3). Infusion reactions were setup with 100 ng of linearized

destination pOPIN vector, 50-100 ng purified PCR product (Section 2.2.3), 1 µl Infusion

enzyme pre-mix (Clontech) and made up to 5 µl with ddH$_2$O following the manufacturer's

instructions. Reactions were incubated for 30 mins at 42^0C followed by 15 mins at 50^0C

before cooling on ice prior transformation of reactions into DH10B *E. coli* and plating onto

IPTG and X-gal supplemented plates for blue-white selection. All RRS1 pOPIN constructs

were cloned from a RRS1-R cDNA construct generated from *N. benthamiana* cDNA

transiently overexpressing 35s::RRS1-R:HF generated by myself. RPS4 constructs were cloned from various cDNA fragments cloned by myself and Dr Zane Duxbury from reassembling the exons of a RPS4 gDNA construct.

2.2.6   Plasmid purification & confirmation

Once colony PCRs (section 2.2.2) were carried out to identify colonies carrying correct predicted construct each colony was inoculated into 5 ml of L medium and grown overnight 37^0C shaking at 200 rpm. Cells were harvested and plasmid purified using QIAquick Spin Miniprep kit (QIAGEN) as per manufacturer's instruction. Each plasmid prep was eluted in 35 μl of warmed ddH$_2$O. Non-pOPIN vector construct plasmids were then checked for correct assembly using restriction enzyme digest analysis. All plasmids sequences were then verified by sanger sequencing using Eurofins Genomics LightRun sequencing services (previously GATC Biotech). Verified plasmids were retransformed into DH10B cells and cultured to make glycerols by mixing cultures with 20% glycerol prior to freezing in liquid nitrogen and storage at -80^{0}C.

2.2.7   Transformation of competent bacterial cells

Transformation of plasmids and ligation reactions into electrocompetent *E. coli* DH10B and ElectroMAX Stbl4 or *A. tumefaciens* GV3101 were conducted by electroporation. 50 μl aliquots of electrocompetent were thawed on ice before adding 100-400 ng of plasmid or a 5 μl desalted ligation reaction. The cell-DNA mix was then placed in a 0.1 cm cooled electroporation cuvette and electroporation conducted using a MicroPulser (Bio-Rad). MicroPulser conditions for *E. coli* were: 1800 V, 25 μF capacitance, 200 Ω resistance and for *A. tumefaciens* were 2400 V, 25 μF capacitance, 200 Ω resistance. All other cells, unless otherwise stated, used in this study were transformed using heat shock treatment of chemically competent cells. For heat shock treatment, aliquots of chemically transformed cells were thawed on ice and mixed with 100-400 ng of plasmid before being incubated on ice for 20 minutes. Cells were then heat shock treated for 42 second at 42^{0}C followed by a 5 minute incubation on ice.

Following electroporation or heat shock treatment, cells were diluted with 250 μl of L media and left shaking at 200 rpm for 1 hour at 37^{0}C or 28^{0}C for *E. coli* and *A. tumefaciens* respectively to allow cells to recover before plating. 50-250 μl of cells were then plated on appropriate antibiotic selection LB plates and left to grow overnight at 37^{0}C or 28^{0}C for *E.*

coli and *A. tumefaciens* respectively. Blue/white selection plates were generated by spreading 10 µl of 1 M IPTG and 40 µl of 40 µg/ml on plates prior to plating of cells.

2.2.8 Isolation of total RNA & cDNA from *N. benthamiana*

N. benthamiana tissue harvested 2 dpi was ground to a fine powder using precooled pestle and mortar. ~150 mg of ground tissue was then transferred to a precooled 2ml Eppendorf and the tissue vortexed with 1 ml TRI reagent (Sigma) and incubated for 5 minutes on ice. The sample was then centrifuged at 4^0C 12,000 x g for 5 minutes. 150 µl of BCP (1-bromo 3-chloropropane, Sigma) was then added and the sample vortexed before incubation at room temperature for 10 minutes. Samples were then centrifuged for 10 minutes at 12,000 x g 4^0C. The aqueous upper phase, ~800 µl was then transferred to a fresh 2 ml Eppendorf and mixed with 400 µl Isopropanol and 400 µl high salt precipitation solution (0.8 M sodium citrate, 1.2M NaCl) to precipitate the RNA. Samples were mixed well by turning before incubating for 5 minutes at room temperature. Samples were then centrifuged for 15 minutes at 12,000 x g at 4^0C and the supernatant discarded. The RNA pellet was then washed with 1.5 ml 75% ethanol then centrifuged for 5 minutes at 7,500 x g at 4^0C. The resulting pellet was air dried and then resuspended in 30 µl DEPC treated H_2O. To remove any contaminating DNA, the RNA samples were treated with a DNAse treatment using RNAse free DNAseI (Roche) as per manufacturer's instructions. In order to deactivate the DNAse, samples were then treated with proteinase K for 15 minutes at 42^0C. RNA was then purified from the sample using QIAGEN RNeasy clean-up kits according to the kit protocol. RNA was eluted from RNeasy columns in 50 µl RNAse-free water. The concentration of RNA was quantified using a nanodrop and integrity of the RNA observed by gel electrophoresis running 1 µl of sample.

cDNA was generated through reverse transcription of 2-5 µg RNA using SuperScript II Reverse Transcriptase (Invitrogen). A reaction mixture of purified RNA, Oligo(dT)$_{12-18}$ and dNTPs was incubated at 70^0C for 5 mins before placing back on ice to disrupt the secondary structures of RNA and primers. DTT, SuperScript II buffer, RNAseOUT and SuperScript II Reverse transcriptase were then added to the mix and incubated at 42^0C for 75 mins followed by 70^0C for 15 mins to deactivate the enzymes before being transferred back to ice. The cDNA volume was then adjusted to 60-100 µl and stored at -80^0C.

2.3 Transient expression in *Nicotiana spp* (Agroinfiltration)

Agrobacterium strains were streaked on to solid LB media plates supplemented with the appropriate antibiotics from glycerols stored at -80^{0}C 3 days prior to the infiltration date and grown at 28^{0}C for ~2 days. 10 ml overnight cultures of agrobacterium strains were then set up from these plates in LB media supplemented with the appropriate antibiotics and grown shaking overnight at 28^{0}C. Cells were then pelleted by centrifugation at 3500 x g for 10 mins. Pelleted cultures were resuspended in 2 ml of infiltration buffer (10 mM MgCl$_2$, 10 mM MES, pH 5.6). Cultures were then diluted in infiltration buffer to OD$_{600}$= 0.5 to make the final mix for infiltration. For co-expression of multiple agrobacterium strains, each strain was adjusted to be OD$_{600}$= 0.5 in the final mixture to ensure the same quantity of each expression in the infiltration mix. The agrobacterium mixtures were then incubated at room temperature for ~1 hour prior to infiltration. The abaxial leaf surface of *N. benthamiana* or *N. tabacum* leaves were infiltrated ~5 weeks after sowing with a blunt ended 1 ml syringe. *N. tabacum* leaves were pricked with a sterile needle prior to infiltration to aid the delivery of agrobacterium to the leaf apoplast. For protein expression, *N. benthamiana* leaves were harvested and the mid-vein removed before being frozen in liquid nitrogen and stored at -80^{0}C 2 dpi (days post inoculation). For hypersensitive response (HR) assays, *N. tabacum* leaf images were taken 3-5 dpi. This method is based on that described in Sarris et al, 2015[58].

2.4   Biochemical Techniques

2.4.1   Total protein extraction from plant tissue

Harvested frozen plant tissue was ground to a fine powder in liquid nitrogen with a precooled pestle and mortar. ~0.5 ml of powder (~0.15 g of powdered tissue) was then transferred to a pre-cooled 2 ml Eppendorf and mixed thoroughly with 1 ml of extraction GTEN buffer (150 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 10 % v/v glycerol) supplemented with 10 mM DTT, 0.2%Nodinet-40 (NP40), anti-protease tablet (cOmplete™ EDTA-free, Sigma Aldrich). For *N. benthamiana* tissue, the extraction buffer also contained 2% (w/v) PVPP (Polyvinylpolypyrrolidone). Samples were then incubated in the extraction buffer at 4^{0}C on a rotor. Incubated samples were then centrifuged at 5000 x g for 20 minutes at 4^{0}C and the supernatant collected. To check for protein expression in the total

extract, 100 µl of supernatant was then heated with SDS-PAGE loading buffer to 95^0C for 10 mins. This method was adapted from Sohn et al, 2014 [161].

2.4.2   SDS-PAGE

This project used two different SDS-PAGE (Sodium dodecyl sulfate polyacrylamide gel electrophoresis) systems.

Method 1: SDS-PAGE gels were prepared in the lab with a resolving gel layer of 12% or 17% w/v polyacrylamide diluted in 375 mM Tris-HCl pH 8.8 and 0.1% w/v SDS. Immediately prior to pouring into Mini-PROTEAN 1 mm casting glass plates (Bio-Rad) 0.1 % (w/v) ammonium persulfate and 0.04 % (v/v) N,N,N',N'-tetraacetylethylenediamine was added to the polyacrylamide SDS solution. Water saturated butanol was then applied to the top of the resolving gel layer whilst setting to ensure a level boundary between the stacking and resolving gel layers. This butanol solution was removed once the resolving gel layer had set and stacking gel mixture poured on top (5 % w/v polyacrylamide, 63 mM Tris-HCl, pH 6.8 with 0.1 % (w/v) SDS, 0.1 % (w/v) ammonium persulfate and 0.1 % (v/v) N,N,N',N'-tetramethylethylenediamine). Plastic gel casting combs were then added and the gels left to set. Gels were subsequently wrapped in clingfilm and stored at 4^0C prior to use. Protein samples were heated at 95^0C for 10 mins in 4x SDS-loading buffer (final concentrations: 50 mM Tris-HCl, pH 6.8, 100 mM dithiothreitol (DTT), 2.0 % (w/v) SDS, 0.1 % bromophenol blue, 10 % glycerol). Gel combs were removed prior to loading denatured protein samples and gels run in a Bio-Rad miniPOTEAN tetra cell system in SDS-running buffer (25 mM Tris-HCl, 250 mM NaCl, 0.1 % w/v SDS). Protein samples were loaded along with a pre-stained protein standard (RunBlue™ Prestained Molecular Weight Marker, Expedeon or Plus2 Pre-stained Protein Standard, Invitrogen) and run at 140-180V for 60-75 minutes.

Method 2: Pre-cast 16%, 12% or gradient 4-20% (w/v) polyacrylamide Teo-Tricine SDS gels were purchased from Expedeon. Gels were run in RunBlue™ tanks with 1x RunBlue™ Teo-Tricine SDS running buffer. Prior to loading, protein samples were heated at 95^0C for 10 mins in 4x LDS Sample Buffer (Expedeon) supplemented with 100 mM DTT. Gel wells were flushed with buffer prior to sample loading and protein samples loaded alongside a pre-stained protein standard (RunBlue™ Prestained Molecular Weight Marker (Expedeon) or Precision Protein™ Dual Xtra Prestained Protein Standard (Bio-Rad). Gels were run at 120-180V until the dye front reached the bottom of the gel.

For visualising proteins, gels prepared by either method were stained overnight shaking at room temperature in InstantBlue™ Coomassie Protein Stain (Expedeon) before rising in deionised water and imaging using a G:BOX gel doc system (Syngene) and GeneSys software.

2.4.3   Immunoblotting (Western blotting)

SDS-PAGE gels were run as described in 2.4.2. Proteins in the gel were then transferred to a PVDF (polyvinylidene difluoride, Merck Millipore) membrane using semi-dry transfer apparatus Trans-Blot® TurboTM Transfer System (Bio-Rad) using the 'High-Molecular Weight' programme according to the manufactures instructions. Following protein transfer, gels were blocked for 1 hour in TBST (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1 % Tween®-20) supplemented with 5% (w/v) non-fat dried milk powder with gentle agitation. Following blocking, membranes were incubated with the appropriate antibody in a solution of TBST with 5% (w/v) non-fat dried milk powder overnight at 4^0C with gentle agitation. Antibody concentrations can be found in **Table 2.6**. If a secondary HRP-conjugated antibody was required membranes were then washed 3 times with TBST for 10 minutes. Following antibody incubation, membranes were washed 3 times in TBST for 5 minutes then in TBS (50 mM Tris-HCl, pH 8.0, 150 mM NaCl) for 10 minutes. Chemiluminescence visualisation of blots was then carried out by thinly covering membranes in 500 µl of HRP substrate developing reagent (SuperSignal West Pico Plus or Femto solution, Thermo Scientific). Blots were imaged using an ImageQuant LAS 4000 (Life Sciences) or X-ray film exposure (Fujifilm). This method is based on that described in Sarris et al, 2015[58].

Table 2.6 Antibodies and dilutions used in this study

Antibody	Dilution concentration	Supplier
α-FLAG HRP M2	1: 10000	Sigma Aldrich
α-HA HRP	1: 3000	Sigma Aldrich
α-mCherry	1:3000	Abcam
α-mNeon	1:2000	Chromotek
α-rabbit IgG HRP	1:10000	Sigma Aldrich

2.4.4 Native-PAGE

Following protein extraction, protein samples were mixed with 10x Native loading buffer and loaded onto a precast 3-12% Native PAGE Bis-Tris Gel (Life Technologies) alongside NativeMark Unstained Protein Standard (Thermo Scientific) ladder. The NativePAGE gel was run at 150 V for 60 minutes followed by 250 V for 30-60 minutes at 4^0C until the dye front had reached the base of the gel as per the manufacturer's protocol. Following gel electrophoresis, proteins were transferred to a PVDF membrane in 1x Novex transfer buffer. Proteins were transferred as per manufactures instructions using the 'High-Molecular Weight' programme on the semi-dry transfer Trans-Blot® TurboTM Transfer System apparatus (Bio-Rad). Following transfer membranes were immediately fixed by washing in a solution of 25% methanol and 10% acetic acid for 10 minutes. Membranes were then twice washed with distilled water and left to air dry overnight. Once dried membranes were reactivated in methanol for 30 seconds and then blocked, probed and imaged as described in section 2.4.3. This method is based on that described in Sarris et al, 2015[58].

2.5   High-throughput protein expression screening

Constructs for high-throughput protein expression screening in *E. coli* and Sf9 Baculovirus transfected insect cells were generated using Infusion cloning in pOPIN vectors according to section 2.2.5. A list of constructs generated for this expression screen can be found in **Table 3.2**.

2.5.1 Rational Design of Protein constructs

RRS1 and RPS4 domain boundaries on which protein expression boundaries were based were defined on the basis of homology to other NLRs. Protein disorder prediction analysis was conducted using RONN protein disorder prediction analysis[162] (https://www.strubi.ox.ac.uk/RONN) or IUPred2A[163] (https://iupred2a.elte.hu/). Secondary structure prediction analysis was conducted using PHYRE2[164] (www.sbg.bio.ic.ac.uk/phyre2/).

2.5.2 *E. coli* expression screening

High-throughput screening of RRS1 and RPS4 constructs in *E. coli* was conducted at the OPPF. All constructs except for constructs 39-43 listed in **Table 3.2** were screened by myself

with the supervision of Heather Rada at OPPF. Each pOPIN construct was screened in two *E. coli* strains, Lemo21(DE3) and Rosetta (DE3) pLysS, grown in two different medias, Power Broth and AIM (Overnight Express™ Instant TB medium) at 20^0C and 25^0C respectively. This meant that each construct was tested in four different expression conditions. Alongside the RRS1 and RPS4 pOPIN constructs an eGFP construct was also transformed as a control. The protocol described below follows OPPF's standard operating procedure for 'small-scale cloning and expression' for *E. coli* produced by Joanne Nettleship (https://www.oppf.rc-harwell.ac.uk/OPPF/protocols/).

Expression screening cell growth:

Sequence verified pOPIN vector constructs were transformed into Lemo21(DE3) and Rosetta (DE3) LysS but conducting a heat shock treatment. ~300 ng of pOPIN plasmid was added to aliquots of chemically competent *E. coli* cells and incubated on ice for 30 mins before heating at 42^0C for 30 seconds in a water bath. Cells were then returned to ice for 2 minutes before 300 µl of Power broth with no antibiotic selection was added to each tube and tubes incubated at 37^0C for 1 hour. Following incubation cells were plated on solid LB media plates supplemented with the correct antibiotic. Plates were incubated overnight at 37^0C. 24 Deep-well plates were then setup with 0.7ml of Power Broth supplemented with the appropriate antibiotic. Colonies were then picked from the solid LB media plates and added to each well. Deep-well blocks were then sealed with gas permeable adhesive seals (ABgene AB-0718) and grown overnight at 37^0C shaking at 200-225 rpm.

For expression screening, 24 deep-well plates were setup with 3ml of media in each well supplemented with the appropriate antibiotic. Cells were screened in both Power Broth media with IPTG induction and autoinduction media (Overnight Express™ Instant TB medium). 150 µl of Lemo21(DE3) or 250 µl of Rosetta cells were added to wells of media. The plates were resealed with gas permeable adhesive seals and grown shaking at 240 rpm at 37^0C for 3-5 hours until the cells reached ~OD_{600}=0.5. Cells growing in Power Broth were then cooled by shaking at 240 rpm at 20^0C for 20 minutes. Power Broth Cells were then induced by the addition of IPTG to a final concentration of 1 mM per well and cells grown overnight (~18 hours) at 240 rpm at 20^0C. Cells growing in autoinduction media were grown at cells reached ~OD_{600}=0.5 at 37^0C before growing the cells overnight (~20-24) hours at 25^0C shaking at 240 rpm.

Expression screen cell harvesting:

The following overnight induction, 1 ml of culture from each well was transferred into a 96-well deep-well block (BC Falcon 353966) maintaining the plate layout. Cells were harvested by centrifugation (6000 x g for 10 minutes) and the supernatant carefully discarded. Plates were then sealed with corning foils and frozen at -80^0C for a minimum of 20 minutes to aid later cell lysis.

Robotic miniature Ni^{2+}-NTA purification

Expressed protein of interest (POI) were then purified using an AVISO Theonyx robotic platform using a method adapted from the QIAGEN BioRobot 8000 protocol.

Harvested frozen cell pellets were resuspended in 210 μl of lysis buffer (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 10 mM Imidazole, 1% v/v Tween20, pH 8.0) supplemented with 1mg/ml lysozyme (Sigma Aldrich) and 400 U/ml DNAse I (Sigma Aldrich). Cells were incubated in lysis buffer for 30 minutes before centrifugation at 5000 x g for 30 minutes at 4^0C. The supernatant of the lysed cell mixture was then transferred to a 96-well flat bottomed magnet compatible microtiter plate (MTP)(Greiner 655101) which contained 20 μl of Ni-NTA magnetic bead suspension (GE Healthcare) in each well. The insoluble pellet of each lysis reaction was then stored at -80^0C. Each supernatant and Ni-NTA mix was then incubated for 30 minutes at room temperature on an MTP shaker. The 96-well MTPs were then placed on a 96-well magnet for 1 min before the supernatant was removed. Ni-NTA beads were then washed with 200 μl of wash buffer (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 20 mM Imidazole, 0.05% v/v Tween20, pH 8.0) for 5 minutes with agitation before placing back on the 96-well magnet and the supernatant was removed. This wash step was repeated twice. POIs were eluted from the Ni-NTA beads by mixing with 50 μl of elution buffer (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 250 mM Imidazole, 0.05% v/v Tween20, pH 8.0) for 1 minute before collecting the eluate supernatant. The eluate was then analysed by running samples on 10% (w/v) polyacrylamide SDS-PAGE gels and staining with InstantBlue™ Coomassie Protein Stain, see section 2.4.2.

2.5.3   Baculovirus transfected Sf9 insect cells

High-throughput screening of RRS1 and RPS4 constructs in Baculovirus transfected Sf9 insect cells, a clonal isolate derived from the *Sfrugiperda* IPLB-Sf21-AE parental cell line*,* was conducted at the OPPF. Each pOPIN construct was screened in Sf9 cells transfected with two different titres of P1 virus. The protocol described below follows OPPF's standard

operating procedure for 'Insect expression' produced by Joanne Nettleship (https://www.oppf.rc-harwell.ac.uk/OPPF/protocols/). All Sf9 cell work prior to protein Ni-NTA purification was conducted under sterile conditions. Alongside the RRS1 and RPS4 pOPIN constructs an eGFP construct was also transfected as a control. Cell density was monitored by mixing 10 μl of Sf9 cells with 10 μl of trypan blue and pipetting into the reservoir of a countess slide (Invitrogen). Cells were then counted and checked for dead cell to attain viability count and the ratio of cell stock to Sf900II medium needed for a specific dilution calculated.

Sf9 cell transfection for P0 virus stock generation:

Although expression screening can be conducted with P0 virus (Passage 0), P1 (Passage 1) virus stock is more reliably used for small to medium scale screening. To generate the P0 virus stock 500 μl of Sf9 cells (density: 5×10^5 cells/ml) in Sf-900™ medium (Invitrogen) were added to each well of a 24-well culture plate and the cells left to attach to the plate at room temperature for 1 hour. Separately, the following transfection mix reagents were mixed: 100-500 ng pOPIN vector, 250 ng linearized bacmid, 50 μl Sf-900™II medium, 1.5 μl FuGene HD Transfection Reagent (Promega). The transfection reaction mix was then gently mixed and incubated at room temperature for 30 minutes. This transfection mix was then slowly added to the Sf9 cells in the 24-well culture plate to avoid disrupting the cell monolayer and gently swirled to distribute transfection mix across the well. The plates were then incubated for 6-7 days at 27^0C. The viral supernatant was then collected and stored in the dark at 4^0C in a sealed 96-well storage block. This supernatant is the P0 virus stock.

Viral amplification of P1 virus stock:

To generate the P1 virus stock 500 μl of Sf9 cells (density: 1×10^6 cells/ml) in Sf-900™ II medium (Invitrogen)were added to each well of a 24-well culture plate and left to allows cells to attach to plate at room temperature for 30-60 minutes. 5μl of P0 virus was then added to the Sf9 cells and left to incubate for 6-7 days at 27^0C. The eGFP control virus was monitored to ensure the amplification was working efficiently. Following incubation, the virus supernatant was harvested and stored in the dark at 4^0C in a sealed 96-well storage block.

Expression testing:

A 24-well deep block plate was prepared with 3ml of Sf9 cells (density: 1×10^6 cells/ml) in each well. Sf9 cells well infected with two different titres of P1 virus stock, 3 μl or 30 μl.

Cells were incubated for 3 days at 27^{0}C shaking at 250 rpm. To asses for protein expression levels, 1 ml of each well culture was transferred to a 96-well and centrifuged for 15 min at 6000 x g. The supernatant was then removed and cells frozen at -80^{0}C for a minimum of 30 minutes. Ni-NTA purification was then performed as per 'Robotic miniature Ni^{2+}-NTA purification' protocol described in section 2.5.1.

2.6   *E. coli* Protein Purification

Protein purification methods were based of the protocol developed in the Banfield lab[70].

2.6.1   Small-scale expression screening

In addition to the high-throughput expression screens conducted at OPPF (Section 2.5) screening of constructs for expression in *E. coli* cells was also conducted by myself in the Banfield laboratory at JIC. For these screens, constructs were transformed into *E. coli* strains as described in section 2.2.7. As a starting point for expression screening, constructs were first transformed into Rosetta™ 2 (DE3) and BL21(DE3) cells as previous experience in the lab highlighted these two strains as having good protein expression rates. Some screens were subsequently expanded out to include SHuffle® T7 and Lemo21 (DE3) cells as well. 10 ml universals of LB media supplemented with the correct antibiotics were inoculated with colonies from the transformant plates or a scraping from a glycerol stock and grown shaking at 200 rpm overnight at 37^{0}C. A 24 deep-well block with 5ml of media per well or 10 ml universals were then supplemented with the appropriate antibiotics. Each construct was initially tested in two types of media, LB and autoinduction media. Fresh 10 ml universals or 5ml deep-well blocks were then inoculated with 200-400 µl of overnight bacterial cultures and grown shaking at 200 rpm at 37^{0}C until cells reached a density of OD$_{600}$=0.4-0.8. Cells grown in LB media were then cooled and induced with a final concentration of 1 mM IPTG unless stated otherwise, cells grown in LB media were grown in duplicate cultures where one culture was not induced with IPTG to help with identifying POI bands by SDS-PAGE analysis. Cell cultures were then grown overnight at 18 or 20^{0}C. Cells were harvested by centrifuging 1ml of each culture at 3800 x g for 10 minutes and the supernatant discarded. 1ml cell pellets were then lysed in 300 µl 1x BugBuster® Protein Extraction Reagent (Merck) supplemented with 0.5 µl Benzonase (Merck). Cells were incubated in lysis buffer for 5 minutes at room temperature with gentle agitation. A sample of the total lysate was then collected for SDS-PAGE analysis. The lysate was then centrifuged at 13,300 x g for 5 minutes at 4^{0}C to pellet the insoluble fraction of the mixture.

A sample of the supernatant was then taken for SDS-PAGE analysis of the soluble protein fraction. SDS-PAGE gels were run as per section 2.4.2. Should these small-scale screens identify conditions for soluble protein production the conditions were scaled up as per 2.6.2.

2.6.2 Large-scale culture growth

50ml of LB media in 250 ml conical flasks supplemented with the appropriate antibiotics were inoculated with a scraping taken from a glycerol stock and grown overnight shaking at 200 rpm at 37^0C. 1 L of fresh expression media (LB or AIM) was then inoculated with the overnight culture to OD_{600}=0.07 (usually 25-35ml of overnight) and grown shaking at 200 rpm at 37^0C until culture reached OD_{600}=0.4-0.8. A total of 6-8 L of culture were grown for a large-scale purification. For LB grown cultures, cells were then induced with IPTG (final concentration 1 mM unless otherwise stated) and cultures grown overnight (~16 hours) at the appropriate temperature. For cells grown in AIM, no additional inducer needs to be added. As cells grow in AIM they initially metabolise glucose preferentially over lactose. Once the glucose in the AIM is depleted lactose is taken up and converted by β-galactosidase to inducer allolactose. Allolactose then release the lac repressor thereby inducing the expression of T7 RNA polymerase which allows expression of POI[165]. AIM grown cultures were grown at 200 rpm at 37^0C until culture reached OD_{600}=0.4-0.8 then grown overnight at 18-25^0C. Cells were harvested by centrifugation at 5000 x g for 10 minutes at 4^0C. Pellets were then completely resuspended in ~50 ml of lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM Imidazole, 50 mM glycine, 5% v/v glycerol with 1 protease inhibitor tablet per 50 ml buffer) for every 2 L of culture unless otherwise stated. Pellets were then frozen for a minimum of 16 hours at -80^0C.

2.6.3 Large-scale *E. coli* purification

Frozen cell pellets were defrosted at room temperature before cells were sonicated on ice using a single 10-12 mm probe Vibra-Cell™ sonicator (Sonics). Cells were sonicated at 40% amplitude for 1 second on and 3 seconds off for a total of ~10 minutes on. A total protein sample was then taken for later SDS-PAGE analysis. The cell lysate was then sonicated for 30 minutes at 18,500 x g at 4^0C to pellet the insoluble fraction. The soluble protein fraction supernatant was then collected and a sample taken for later SDS-PAGE analysis. The POI was then purified from the cell lysate using an automated protein purification system on an ÄKTAxpress system (GE healthcare). The soluble cell lysate was first loaded onto a nickel-

charged 5ml HisTrap™ HP IMAC column (GE healthcare) and unbound protein washed out of column with binding buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 20 mM Imidazole, 50 mM glycine and 5% v/v glycerol). IMAC column bound protein were eluted by washing with elution buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 500 mM Imidazole, 50 mM glycine and 5% v/v glycerol). Eluted protein was immediately loaded onto a Superdex™ 75 HiLoad™ 26/600 gel filtration column (GE Healthcare) pre-equilibrated in gel filtration running buffer (standard buffer: 20 mM HEPES pH 7.5, 150 mM NaCl). See **Table 2.7** for a full list of construct extraction and running buffers. Gel filtration column elution was collected in 8 ml fractions in 24 deep-well blocks and samples from wells corresponding to UV absorbance 280 nm trace peak analysed by SDS-PAGE for presence of POI. Gel filtration fractions containing POI were pooled and concentrated ~20ml. Proteins were concentrated by ultrafiltration using Vivaspin® concentrators (Satorius)of various molecular weight (MW) cut-off sizes depending on the size of the POI. A concentrator MW cut-off of less than half the POI's MW was used and concentrators centrifuged as per manufactures instructions. Once concentrated, POI containing a 3C protease cleavable His/solubility tag were incubated with 12µg of His-tagged 3C protease (produced by Richard Hughes) per mg of POI overnight at 4^0C to remove His and solubility tags from the POI. To remove the uncleaved POI, His-tagged 3C protease and cleaved solubility tags the cleaved protein mixture was manually loaded onto a 5 ml HisTrap™ HP IMAC column (GE healthcare) pre-equilibrated in binding buffer. The unbound fraction containing the cleaved POI was washed out and collected with 15ml binding buffer. Uncleaved POI, 3C protease and cleaved solubility tag was eluted in elution buffer and disposed of. The cleaved POI was then loaded onto Superdex™ 75 HiLoad™ 26/600 gel filtration column (unless otherwise stated) pre-equilibrated in gel filtration running buffer to undergo a second gel filtration purification. Gel filtration fractions from the UV_{A280} trace peak were analysed by SDS-PAGE for presence of POI. Fractions containing pure POI were concentrated in an appropriately sized Vivaspin® concentrator and purified protein used immediately or flash frozen in aliquots of 50-75 µl in liquid nitrogen and stored at -80^0C.

Protein complexes were produced my three methods. **Method 1** involved co-expressing the two proteins of interest in the same *E. coli* cell line where one protein contained a His-tag and one protein did not (pOPINA vector). Once pelleted, these cells were processed as described above. For complexes where the two-proteins required different growth conditions **Method 2** was used. In this strategy two cultures are grown separately one expressing protein with a 6xHis-tag and one with untagged protein (pOPINA vector).

Following culture growth, the cells pellets were resuspended in lysis buffer and frozen separately but the cell lysates were mixed after thawing. The rest of the process was then conducted with mixed lysates. **Method 3** was used for the production of RRS1B WRKY-AvrRps4 complex only. In this method cells were processed separately as described above until the manual Ni-IMAC column to remove His-tagged proteins where both cleaved POI were eluted into the same vessel and incubated together on ice for 1 hour prior to the second gel filtration. Cell lines, buffer amendments (relative to those listed above) and *E. coli* growth conditions used for each large-scale purified construct in this study can be found in **Table 2.7.**

Table 2.7 Growth and purification conditions for large-scale *E. coli* preps in this study

Construct	Vector	*E. coli* cell strain	Growth media	Overnight growth temperature (^{0}C)	Lysis buffer amendments	Gel filtration buffer amendments
RRS1 WRKY Dom6S$_{S1184-C1290}$	pOPINF	SHuffle® T7	AIM	18		
RRS1 WRKY Dom6S$_{E1195-C1290}$	pOPINS3C	Rosetta™ 2 (DE3)	LB	20	+1 mM TCEP	+1 mM TCEP and 500 mM NaCl
RRS1 WRKY$_{E1195-T1273}$[71]	pOPINS3C	Rosetta™ 2 (DE3)	LB	20	+1 mM TCEP	+1 mM TCEP and 500 mM NaCl
RRS1B WRKY$_{N1163-H1237}$	pOPINS3C	Rosetta™ 2 (DE3)	LB	20	+1 mM TCEP	+1 mM TCEP and 500 mM NaCl
AvrRps4$_{G134-Q221}$[166]	pOPINF/A	BL21(DE3)	LB	18		
RRS1 TIR$_{K6-G153}$[114]	pMCGS7	BL21(DE3)	AIM	20		+1 mM TCEP

2.6.4 Measuring protein concentration

Protein concentration was measured by two methods depending of the number of aromatic residues contained within the POI and thereby the protein's absorbance at 280 nm. The concentration of proteins containing aromatic residues was measured using a NanoDrop™ One spectrophotometer (Thermo Fisher) by absorbance at 280 nm.

Absorbance readings were corrected using extinction coefficient of the POI which adjust the value based on the aromatic residue composition of the protein. The extinction coefficient for a POI was calculated using the ExPASy ProtPram online tool[167]. For proteins which contained little to no aromatic residues (e.g. AvrRps4) protein concentration was determined using a Direct Detect® Infrared Spectrometer (Merck). This apparatus determines protein concentration based on the amide bonds in the protein chain and therefore is independent of the POI's amino acid composition.

2.6.5 Mass Spectrometry

Protein samples were submitted to JIC Proteomic platform to determine a POI's intact mass using a Synapt G2-Si mass spectrometer coupled to an Acquity UPLC system (Waters). Samples were run and analysed by Dr Gerhard Saalbach. In order to confirm the peptide coverage of a POI SDS-PAGE gel bands thought to contain the POI were excised using a sterile blade. Excised gel samples were prepared for mass spec analysis by in-gel trypsin digestion performed by Dr Paul Derbyshire as described by Bender et al, 2018[168] . Samples were run on a Thermo Scientific™ Orbitrap Fusion™ Tribid™ Mass Spectrometer.

2.7 Cell-free wheat germ expression system

Proteins were expressed in two different cell-free wheat germ sources; one system was using the Protein Research Kit S from Cell Free Sciences and the other using a wheat germ extract supplied by Professor Yasuomi Tada from Nagoya University. All proteins were expressed in a pEU-E01-MCS vector which had been modified to enable golden gate cloning by Dr Cheng Chang. This modified vector allowed for the golden gate constructs to be generated with HellFire (6xHis with 3xFLAG) tagged POI under a SP6 promoter. The plasmid preparation and transcription protocols were the same for both wheat germ systems and followed the protocol outlined in Cell Free Sciences 'ENDEXT® Technology Protein Research Kit (S,H,G) for 'Protocol for plasmid DNA-based protein synthesis'.

Plasmid preparation and transcription:
Sequence verified pEU expression plasmids containing the POI were transformed as per section 2.2.7 in to Dh10B *E. coli* cells and 25 ml overnight cultures grown from the subsequent transformed colonies. The pEU plasmid was then prepped from the 25 ml overnight cultures and purified using QIAGEN Plasmid Plus Midi Kit as per the

manufacturer's instructions. Plasmid was eluted from spin columns in 400 µl of TE buffer and plasmid concentration adjusted to 1 µg/ul. Plasmid purity was checked by measuring absorbance using a spectrometer ensuring the 260/280 nm absorbance ratio was between 1.7-1.85. 2000 ng of high purity pEU plasmid was then added to each thawed Transcription premix LM reactions supplied with the Cell Free Sciences Protein Research Kit S and the reaction gently mixed by pipetting. The transcription reaction was then incubated for 1 hour at 37^0C in a C1000 Touch thermocycler (Bio-Rad). After incubation, the quality of the transcribed mRNA was assessed by gel electrophoresis of 1 µl of the reaction on a 1% agarose TAE gel. Clear bands should be visible when imaging the gel indicative of a high-quality mRNA prep. A laddering pattern or smearing below 500 bp indicates possible degradation of mRNA by RNase and the transcription reaction should be repeated.

Method 1: Cell Free Sciences Protein Research Kit S

The Cell Free Sciences kit used WEPRO® 9240 extract (wheat germ extract mixed with creatine kinase) for expression of proteins. WEPRO® 9240 extract and SUB-AMIX®SGC (translation mix including amino acid mix) were removed from -80^0C storage and thawed on ice. Once thawed the two reagents were spun down and the SUB-AMIC®SGC gently resuspended by pipetting. 10 µl of the 20 µl mRNA transcription reaction was then mixed with the WEPRO® 9240 extract my gentle pipetting. A bilayer translation reaction was then setup. The WEPRO® 9240 extract and mRNA mix were gently pipetted to the bottom of the tube containing SUB-AMIX®SGC carefully dispensing the extract-mRNA mix to avoid mixing with the SUB-AMIX®SGC such that a bilayer is created, see **Figure 2.1**. This bilayer setup allows for diffusion-controlled translation which can be sustained for longer period of time than the standard mixed batch setup which typically can only be sustained for a few hours dependant on the wheat germ extract concentration. This bilayer setup therefore allows for a higher protein yield to be attained[169,170]. The bilayer reaction tubes were sealed with aluminium foils and the whole reaction incubated at 15^0C for 20 hours to allow for translation. Following the translation incubation, the entire contents of the reaction were mixed and a 3 µl sample analysed for protein expression by SDS-PAGE.
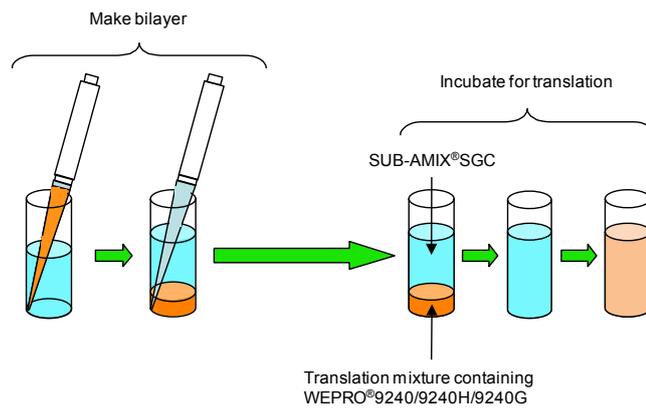
Figure 2.2.1 Bilayer setup for wheat germ protein translation. Figure taken from Cell Free Sciences 'ENDEXT® Technology Protein Research Kit (S,H,G)' protocol

Method 2: Wheat germ extract from Professor Yasuomi Tada

Plasmid preparation and mRNA transcription was conducted as described above. A translation bilayer reaction was setup as described for Cell Free Sciences Protein Research Kit S but instead of using WEPRO® 9240 extract wheat germ purified in Professor Yasuomi Tada's lab was used to enable testing of protein production efficiencies of different wheat germ sources.

His-tagged protein purification from wheat germ

To purify the His-tagged POI from wheat germ translation reactions a protein pull down with Ni-Sepharose resin was performed as per the method described in Cell Free Sciences 'ENDEXT® Technology Protein Research Kit (S,H,G)'. 200 µl of the completed translation reaction was transferred to a fresh 1.5 ml Eppendorf and mixed with 6.7 µl of 600 mM Imidazole (pH 8.0). 15 µl of Ni-Sepharose High performance resin (GE Healthcare) was then added to the tubes and the mixture incubated at 4^0C for 1 hour with gentle agitation. Tubes were then centrifuged to collect the Ni-Sepharose resin at 500 x g for 5 minutes at 4^0C. Before being discarded a sample of the supernatant was collected for SDS-PAGE analysis for the unbound fraction. The pelleted resin was then washed in IP-wash buffer (20 mM Na_3PO_4 pH 7.5, 0.3 M NaCl, 20 mM Imidazole) and the tubes 5 minutes at 4^0C with gentle agitation before the resin was pelleted again by 500 x g for 5 minutes at 4^0C. This wash step was repeated 4 times in total. POI was eluted from the resin by incubating in 30 µl of elution buffer (20 mM Na_3PO_4 pH 7.5, 0.3 M NaCl, 500 mM Imidazole) for 10 minutes at 4^0C. The resin was pelleted by centrifugation at 500 x g for 5 minutes at 4^0C and the supernatant collected as the elution fraction. This elution step was repeated 3 times in total generating

90 μl of elution fraction. POI yield was analysed by SDS-PAGE and western blot analysis described in section 2.4.2 and 2.4.3.

2.8 Large-scale protein purification from plant tissue

For *A. thaliana* plant tissue protein purification purposes, seeds were grown shaking in conical flasks as per 2.1.11. For harvesting from LexA:: inducible transgenic *A. thaliana* inducible lines the growth 1% MS media was poured out from each flask and replaced with fresh 1% Murashige and Skoog (MS) media[155] containing estradiol, control non-induced lines had media replaced with fresh 1% MS media only. *N. benthamiana* tissue for purification purposes was grown as per 2.1.10 and infiltrated with agrobacterium carrying expression constructs as per 2.3 before harvesting 2 dpi. The purification was carried out in a cold room at 5^0c.

2.8.1 Protein extraction of from nuclear preparation

The method used for nuclear extraction from *A. thaliana* tissue is adapted from a protocol developed by Dr Pingtao Ding. 300 ml of ice-cold grinding buffer (40 mM MES, 4 mM EGTA, 80 mM KCl, 250 mM Sucrose, 0.5 mM Spermidine, 0.1 mM Spermine, 5 mM 2-ME, 0.5% v/v TritonX-100 and anti-protease tablet (cOmplete™ EDTA-free, Sigma Aldrich) was added to 100 x g of *A. thaliana* tissue and blended in short bursts (~7x 10 second) in a whirring blender until homogenized. Using a Pasteur pipette the blended mix was filtered through two layers of miricloth to remove large pieces of cell debris. The mixture was left to flow through the miricloth by gravity and collected in a clean tube. The collected flow through was then centrifuged at 2000 x g for 20 minutes at 4^0C to pellet the nuclei. A sample of the supernatant for later analysis was collected before the supernatant was discarded. The nuclei pellet was then resuspended in nuclear resuspension buffer (10% v/v glycerol, 150 mM Tris-HCl pH 7.5, 1 mM EDTA, 150 mM NaCl, 0.1% v/v Tween20 and anti-protease tablet). Following initial trials to evaluate the effectiveness of different sonication treatments for nuclear disruption, a treatment of 5 repeats of 10 seconds on, 120 seconds off at 40% amplitude (Vibra-Cell™ sonicator, Sonics) was decided on. The resuspension was sonicated in two batches. The lysed nuclei were then centrifuged at 5000 x g for 20 minutes at 4^0C. If the suspension was not clear after this centrifugation the lysate was subsequently re-centrifuged at 5000 x g for 20 minutes at 4^0C. The supernatant was then collected and split into 50 ml falcon tubes to be taken forward for immunoprecipitation purification.

2.8.2 Protein extraction from total protein preparation

This method of protein purification from a total plant protein extract was used for both *A. thaliana* and *N. benthamiana* tissue. The method is adapted from a protocol developed by Dr Benjamin Petre and Juan-Carols De la Concepcion. 100 x g of liquid nitrogen frozen leaf tissue was blended to a fine powder using a pre-cooled pestle and mortar. The ground leaf powder was then vortexed in 300% (v/w) ice-cold protein extraction buffer (150 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 10 % v/v glycerol) supplemented with 10mM DTT, 0.2%Nodinet-40 (NP40), anti-protease tablet (cOmplete™, EDTA-free, Sigma Aldrich) until the powder was completely thawed. For *N. benthamiana* tissue the extraction buffer also contained 2% (w/v) PVPP (Polyvinylpolypyrrolidone). The suspension was then filtered through 2 layers of miricloth to remove large pieces of cell debris and the filtrate collected in a beaker on ice. The filtrate was then centrifuged for 30 minutes at 4000 x g at 4^0C. The supernatant was then transferred to fresh tubes and centrifuged again at 50,000 x g for 90 minutes at 4^0C. The supernatant was then collected and split into 50 ml falcon tubes to be taken forward for immunoprecipitation purification.

2.8.3 Immunoprecipitation purification

POI was purified from protein extraction solution of either nuclear or total extraction methods described above using immunoprecipitation with anti-FLAG or anti-HA resins depending on the tag on the POI. Anti-FLAG/HA affinity bead resin were prepared as per manufactures instructions and equilibrated in IP buffer (150 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 10 % v/v glycerol, 0.2%Nodinet-40, anti-protease tablets). 50 µl of resin was added to each 50ml aliquot of protein suspension. The protein solution was then incubated for 2 hours at 4^0C with gentle rotation. The anti-FLAG/HA affinity resin was then collected by centrifugation at 800 x g for 5 minutes at 4^0C and the supernatant removed. The pelleted resin was then pooled resuspended in 2x 45 ml of wash buffer (150 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 10 % v/v glycerol, 0.2%Nodinet-40). The resin was then pelleted again by centrifugation at 800 x g for 5 minutes at 4^0C and the supernatant removed. The resin from both tubes was then resuspended and pooled in 1x 45 ml of wash buffer and the resin pelleted as before. This step was then repeated twice. The resin was then resuspended and pooled together in 2 ml of wash buffer and then the resin was pelleted as before. This step was then repeated. Proteins were then eluted from the resin by incubating 2 ml of wash buffer supplemented with 150 ng/µl 3xFLAG peptide for 1 hour.

The resin was then pelleted as before and supernatant collected. This step was then repeated with 2 ml of wash buffer supplemented with 150ng/μl 3xFLAG peptide. The eluted supernatant was pooled and centrifuged at 15,000 x g for 30 seconds at 4^0C to pellet any residual resin or aggregated protein. For analytical gel filtration analysis, the eluate was concentrated by ultrafiltration to ~120 μl using Vivaspin® concentrators (Satorius) with a 30 kDa molecular weight cut-off.

2.9   Analytical Gel Filtration

Analytical gel filtration was conducted at 4^0C using ~100 μl of protein per run. For proteins and complexes less then 75 kDa, proteins were loaded on Superdex™ 75 10/300 GL column (GE Healthcare) pre-equilibrated in gel filtration buffer, see **Table 2.7** for buffer amendments. For proteins larger than 75 kDa proteins were loaded onto Superose™ 6 Increase column (GE Healthcare) pre-equilibrated in running buffer (150 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA). Protein samples were eluted from the gel filtration column in a total of 1.5 column volumes at a flow rate of 0.5ml/min. Column elution was collected in 0.5ml fractions and analysed by SDS-PAGE. To analyse protein complex formation, proteins of interest were mixed in a 1:1 molar ratio and incubated on ice for 2 hours or overnight at 4^0C. Prior to loading proteins were centrifuged at 17,000 x g for 10 minutes at 4^0C to pellet any contaminants or protein aggregates. Molecular weights of eluted proteins were estimated using molecular weight calibration curves produced for each analytical gel filtration column generated by Dr John Steele and Dr Abbas Maqbool, see Appendix 3. These methods were based of the protocol developed in the Banfield lab[70].

2.10 Surface Plasmon Resonance (SPR)

SPR is a method which allows the analysis of biomolecular interactions e.g. protein-protein and protein-DNA. The system measure changes in the refractive index on the surface of a chip to which one bio-molecule is immobilized as another biomolecule is injected over the chip surface in the analyte. The system provides real-time information of interactions between bio-molecules. In the correct experimental conditions, kinetic information of this interactions can also be obtained. All SPR experiments were conducted using a Biacore T200 system (GE Healthcare) and based of a method developed in the Banfield lab[70].

2.10.1 Protein-Protein interactions

Protein-protein interactions were quantitatively measured using an SPR Sensor NTA chip (GE Healthcare). This allowed for the immobilization of a His-tagged protein, cloned in pOPINE with a non-cleavable 6xHis-tag, onto the NTA sensor chip surface. Initial trials using this system showed that RRS1 WRKY$_{E1195-T1273}$ and WRKY Dom6S$_{E1195-C1290}$ were prone to sticking to the surface of the NTA sensor chip non-specifically. Optimisation of the NaCl levels in the SPR buffer was found to help reduce these non-specific interactions. The SPR buffer used for the final R$_{max}$ experiments was 20 mM HEPES pH 7.5, 0.1% Tween20 with 500 mM NaCl for RRS1 WRKY$_{E1195-T1273}$ and 860 mM NaCl for RRS1 WRKY Dom6S$_{E1195-C1290}$ analysis. Following additional optimisation experiments, R$_{max}$ data was collected at 8^0C. All proteins were diluted fresh in SPR running buffer. For R$_{max}$ experiments the NTA sensor chip was activated with 30 µl of 0.5 mM NiCl$_2$ then 30 µl of His-tagged AvrRps4 effector protein was immobilized on the surface of the NTA sensor chip. The concentration of wild-type and mutant effectors used was adjusted to give a reliable capture rate of 25-35 response units (RU). Following effector immobilization, either 500 nM or 3 µM RRS1 WRKY/WRKY Dom6S protein was then injected over the chip surface at a rate of 30 µl/min for 120 seconds contact time followed by a 60 second dissociation time. Following RRS1 WRKY/WRKY Dom6S binding the chip was regenerated in 30 µl of 350 mM EDTA pH 8.0. Before the next experiment cycle started the chip was then further washed for 30 seconds in 15 µl of SPR buffer to remove any residual EDTA prior to nickel loading. For R$_{max}$ calculations both the reference cell response unit (flow cell 2-1) and a buffer blank (average of three buffer blank cycles) was subtracted from each data point to reduce the effect of non-specific binding in calculations. R$_{max}$ calculations, were generated using binding stability or level values. The R$_{max}$ of each experiment was calculated using the equation below which is dependent on the molecular weight (MW) of the analyte and ligand, the amount of ligand immobilized on the surface of the sensor chip (R$_{Ligand}$) and stoichiometry of the interaction between ligand and analyte. %R$_{max}$ plots were generated using a ggplot2 package[171] in R using code written by Josephine Maidment.

$$R_{max} = \frac{MW \text{ of analyte (RRS1 WRKY)}}{MW \text{ of Ligand (AvrRps4}_c)} \times R_{Ligand} \times Stoichometry$$

Equilibrium dissociation constant values (K$_D$) were attempted to be calculated using multicycle kinetics curves generated by the Biacore machine software.

2.10.2 Protein-DNA interactions

To analyse protein-DNA interactions between the WRKY domain of RRS1 and WRKY motif recognised W-box DNA, the SPR ReDCaT (Re-usable DNA Capture Technique) chip method developed by Dr Clare Stevenson at JIC was used[172] . This method of SPR allows the reversible application of different DNA sequences to the surface of a streptavidin SPR chip over which varying proteins can be flown as analyte to investigate protein-DNA interactions, see **Figure 2.2**.

These experiments were conducted using a single sensor SA chip (GE Healthcare) which contains four cells with pre-immobilized streptavidin on a carboxymethylated dextran matrix as described by Stevenson et al, 2013[172]. The immobilized streptavidin allows for the convenient binding of biotinylated molecules on the chip surface. For the purpose of this experiment, biotinylated single stranded DNA fragments referred to as 'ReDCaT linkers' were then flown over and attached to the SA chip, this stage was conducted by Dr Clare Stevenson. This then allows the attachment of DNA fragments with a double stranded DNA sequences for DNA region of interest (i.e. the test region presented to protein for binding) with a single stranded DNA region complementary to the attached ReDCaT linked DNA. These DNA fragments were synthesised as oligonucleotides (Merck Millipore) with the reverse primer containing the ReDCaT linker complementary extension at the 3'end plus the region of interest DNA sequence and the forward primer containing on the region of interest DNA sequence. A list of primers used in this experiment can be found in **Table 2.8.**

The experiment used both one and three time repeats of the test DNA sequences. In addition to wild-type w-box DNA sequences[173], protein binding was also tested against 3 mutants. Mutant 1 was taken from previous EMSA work in the Jones lab[65], Mutant 2 was a mutant sequence often used in the literature[174,175] and Mutant 3 was produced by scrambling the wild-type w-box sequence using an online tool from GenScript (https://www.genscript.com/tools/create-scrambled-sequence). The forward and reverse primers were diluted to 100 µM in ReDCaT running buffer (10 mM HEPES, 300 mM NaCl (varied as stated), 3 mM EDTA, 0.05% v/v Tween20). To anneal the two primers together, 45 µl of Reverse strand primer was then mixed with 55µl of forward strand primer and heated to 95^{0}C for 10 minutes before being allowed to cool gradually. This generated a

stock of 45 µM double stranded DNA which was then diluted to 1 µM in ReDCaT running

buffer prior to the ReDCaT experiment.
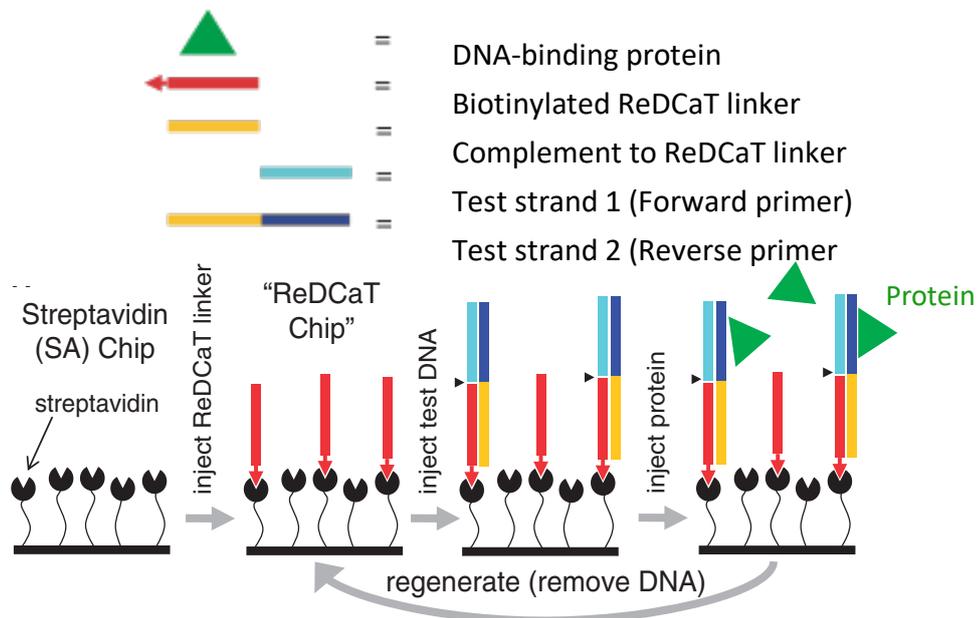
Figure 2.2 ReDCaT SPR experimental setup. Procedure for creating and regenerating ReDCaT chip

for DNA-protein interaction analysis, adapted from Stevenson et al, 2013[172]

Table 2.8 ReDCaT SPR experiment primers

Primer name	Sequence	Reference
ReDCaT Linker	GGATGGGATGCAGGAGGACG	Stevenson et al, 2013[172]
WT w-box 1x repeat	CGTTGACCG (Fwd)	Maeo et al, 2001[173]
WT w-box 1x repeat	CGGTCAACGCCTACCCTACGTCCTCCTGC (Rev)	
WT w-box 3x repeat	CGTTGACCGTTGACCGAGTTGACTTTTTA	Maeo et al, 2001[173]
WT w-box 3x repeat	TAAAAAGTCAACTCGGTCAACGGTCAACGCCTACCCTACGTCCTCCTGC	
Mutant 1 w-box 1x repeat	CGTAGACGG	Sarris et al, 2015[65]
Mutant 1 w-box 1x repeat	CCGTCTACGCCTACCCTACGTCCTCCTGC	
Mutant 1 w-box 3x repeat	CGTAGACGGTAGACGGAGTAGACGTTTTA	Sarris et al, 2015[65]
Mutant 1 w-box 3x repeat	TAAAACGTCTACTCCGTCTACCGTCTACGCCTACCCTACGTCCTCCTGC	

Mutant 2 w-box 1x repeat	CGTTGCACGG	Zhou et al, 2008[175]
Mutant 2 w-box 1x repeat	CCGTGCAACGCCTACCCTACGTCCTCCTGC	Wang et al, 2009[174]
Mutant 2 w-box 3x repeat	CGTTGCACGGTTGCACGGTTGCACGG	Zhou et al, 2008[175]
Mutant 2 w-box 3x repeat	CCGTGCAACCGTGCAACCGTGCAACGCCTACCCTACGTCCTCCTGC	Wang et al, 2009[174]
Mutant 3 w-box 1x repeat	TGATCGC	
Mutant 3 w-box 1x repeat	GCGATCACCTACCCTACGTCCTCCTGC	Scrambled (GenScript)
Mutant 3 w-box 3x repeat	ATGCTTAACGTTCTACGTGTGCGTTACGT	
Mutant 3 w-box 3x repeat	ACGTAACGCACACGTAGAACGTTAAGCATCCTACCCTACGTCCTCCTGC	Scrambled (GenScript)

For ReDCaT experiments 10 µl of 1 µM test primers were injected over the ReDCaT chip (pre-prepared with annealed linker DNA fragments). RRS1 WRKY Dom6S$_{S1184-C1290}$ protein was then flowed over the chip at a rate of 30 µl/min with a contact time of 60 seconds and 60 seconds dissociation time. The chip was then regenerated in 10 µl of 1 M NaCl 50 mM NaOH. Prior to the start of the next experimental cycle the chip was further washed for 60 seconds in ReDCaT running buffer to remove any residual 1 M NaCl 50 mM NaOH. For R$_{max}$ calculations, the reference cell response unit (flow cell 2-1) was subtracted from each data point to reduce the effect of non-specific binding in calculations.

2.11 Crystallography

For protein crystallisation screening, protein was freshly prepped or previously flash frozen aliquots of protein thawed on ice. Prior to screen setup, proteins were centrifuged at 17,000 x g for 10 minutes at 4^{0}C to pellet any contaminants or protein aggregates. Crystallisation screens were setup as sitting drop vapour-diffusion experiments. This method of crystallisation screen involves a setup with a sealed chamber containing a reservoir of solution containing buffers, additives and precipitants, see **Figure 2.3**. A drop of protein mixed with the reservoir solution is placed on a raised well above the reservoir solution and the chamber sealed with a foil seal to prevent evaporation. As the

concentration of solute is higher in the reservoir than the protein drop, water diffuses from the protein drop to the reservoir. As equilibrium is reached, the concentration of protein, reservoir solution solutes, and precipitants increases in the protein drop to the point where the protein precipitates. If this precipitation occurs in a slow and controlled manner this precipitation can lead to protein crystal formation. The conditions which support this form of gradual precipitation are often narrow and therefore the screening of many different reservoir solutions is often required to obtain suitable crystallisation conditions.

A range of commercial crystallisation screens were used in this study which covered a large range of different buffer compositions, salts and additives. For this study these screens were: Morpheus® (Molecular Dimensions), JCSG-plus™ (Molecular Dimensions), PropPlex™ (Molecular Dimensions), PACT Premier™ (Molecular Dimensions), MIDAS™(Molecular Dimensions), PEGs suite (Qiagen), Structure(Molecular Dimensions) and KISS (a custom screened designed by Dr Clare Stevenson and Dr Dave Lawson in JIC's crystallography department). These screens cover a wide range of the crystallisation space including sparse matrix screens (e.g. JCSG-plus™), screens with low molecular weight ligands observed to promote crystal formation (e.g. Morpheus®) and screens targeted towards gaining protein complex structures (e.g. PropPlex™- Molecular Dimensions).

Crystallisation screens were setup in 96-well sitting drop vapour-diffusion plates which contained 2 wells (A and B well) per solution reservoir. Each crystallisation screens were set up at two different protein concentrations with well B setup with half the concentration of protein compared to well A. This meant the protein precipitation rates could be better analysed in each reservoir solution condition. Sitting drops containing 0.3 µl of protein and 0.3 µl of reservoir solution were dispensed in each well using an Oryx Nano crystallisation robot (Douglas Instruments). Following screen setup, plates were immediately sealed to prevent evaporation and placed in an incubator at 20^0C or 4^0C. Plates incubated at 20^0C were imaged at regular intervals with a Minstrel Crystallisation Imager (Rigaku) which imaged plates in both visible and UV light to check for crystal formation. Plates incubated at 4^0C were checked for crystal formation manually by eye using a Nikon SMZ800 microscope.
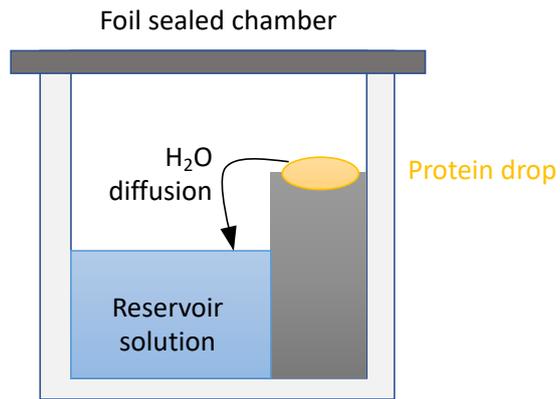
Figure 2.3 Sitting drop vapour diffusion crystal screen experimental setup. A protein drop of purified protein and reservoir solution is placed in a sealed chamber with a partially filled reservoir of solution. As water gradually diffuses from the protein drop to the reservoir, the concentration of protein, solutes and precipitants increases in the protein drop causing the protein to precipitate.

Investigating heterologous expression systems for plant NLR expression

3

3.1 Introduction & chapter aim

Since the cloning of the first NLR over 25 years ago our understanding of NLRs structure and biochemical behaviour has significantly lagged behind our genetic understanding of these receptors. This disparity has largely been due to the difficulties with expressing soluble NLR protein at yields required for biochemical and structural characterisation. Until the recent publication of the Arabidopsis NLR ZAR1 structure[108,109], structures of plant NLRs were largely restricted to the class defining N-terminal CC and TIR regions or integrated domains of NLR proteins[176]. This restriction in available NLR structural data has been a major obstacle to understanding the structural basis of NLR's activation mechanisms and the rational design of NLRs with extended pathogen recognition capabilities. There is therefore a high incentive in the field to expand our structural understanding of NLRs and increase the variety and number of available protein structures of NLRs.

Heterologous systems used for expression of both animal and plant NLRs have largely focussed on the use of the well-studied systems of *S. frugiperda* insect cells and *E. coli*. A list of published structures of plant NLR proteins and the heterologous expression systems used in the studies can be found in **Table 3.1**. In addition to the examples listed in **Table 3.1**, other plant NLRs and truncations of have been heterologously expressed but not resulted in the determination of a protein structure with studies often relying on the use of refolded protein[104] or very low protein expression yields. For example, full-length soluble expression of *Linum usitatissimum* NLR M[177] and *Hordeum vulgare* MLA27[178], expressed in *Pichia pastoris* and *S. frugiperda* Sf21 cells respectively, has been observed but to very low yield levels.

There are several factors to consider when choosing a heterologous expression system which focus around the core facets of the: protein folding environment provided, ability to conduct high-throughput screening, feasibility of physical scalability and the economics of scalability. The *S. frugiperda* insect cell and *E. coli* systems excel in many of these areas and

Investigating heterologous expression systems for plant NLR expression

have consequently historically been widely utilised for heterologous protein expression studies.

Table 3.1 Summary of published plant NLR structures listing the heterologous expression system utilised. Adapted from Burdett et al, 2019[176]

NLR	Domain	Species of origin	Heterologous expression species	Reference
L6	TIR	*Linum usitatissimum*	*E. coli*	Bernoux et al, 2011[87]
SNC1	TIR	*Arabidopsis thaliana*	*E. coli*	Hyun et al, 2016[83] & Zhang et al, 2017[179]
RPS4	TIR	*Arabidopsis thaliana*	*E. coli*	Williams et al, 2014[114]
RPV1	TIR	*Vitis rotundifolia*	*E. coli*	Zhang et al, 2017[179]
RRS1	TIR	*Arabidopsis thaliana*	*E. coli*	Williams et al, 2014[114]
	WRKY	*Arabidopsis thaliana*	*E. coli*	Zhang et al, 2017[71]
MLA10	CC	*Hordeum vulgare*	*E. coli*	Casey et al, 2016[95] & Maekawa et al, 2011[178]
Sr33	CC	*Aegilops tauschii*	*E. coli*	Casey et al, 2016[95]
Rx	CC	*Solanum tuberosum*	*E. coli*	Hao et al, 2013[94]
NRC1	NB-ARC	*Solanum lycopersicum*	*E. coli* & *S. frugiperda* (Sf9)	Steele et al, 2019[180]
PikP-1	HMA	*Oryza sativa ssp. japonica*	*E. coli* & *S. frugiperda* (Sf9)	Maqbool et al, 2015[70] & De la Concepcion et al, 2018[181]
Pikm-1	HMA	*Oryza sativa ssp. japonica*	*E. coli*	De la Concepcion et al, 2018[181]
RGA5	HMA	*Oryza sativa ssp. japonica*	*E. coli*	Guo et al, 2018[182]
ZAR1	Full-length	*Arabidopsis thaliana*	*S. frugiperda* (Sf21)	Wang et al, 2019[183]

E. coli excels in its ease of use in terms of high-throughput screening and scalability. The system is quick to setup and scale and has a relatively low running cost compared to other expression systems as well as a wide range of available compatible expression vectors. However, this system does come with several major disadvantages. Firstly, codon usage is often a problem when expressing eukaryotic proteins in *E. coli* using the POI (protein of interest) origin organism's cDNA sequence. *E. coli's* bias towards preferred codons often

significantly differs from other organisms leading to problems with gene translation and protein yield[184]. This issue can be circumvented however by codon optimising the cDNA sequence of the POI for *E. coli* expression[185] either by point mutagenesis or gene synthesis. Secondly, as a prokaryotic system a major disadvantage associated with this system is *E. coli's* poor folding environment for eukaryotic protein production and the lack of eukaryotic post translational modifications. The poor folding environment is caused by several factors including the inability of the majority of *E. coli* strains to catalyse disulphide bridges and carry out correct eukaryotic post-translational modifications such as glycosylation and fatty acid acylation[186]. The chaperone repertoire of *E. coli* may also be insufficient to aid the correct folding of the POI and may lack key complex chaperones the POI has co-evolved with in the original host species to enable correct folding. Incorrect folding may lead to misfolded insoluble protein aggregates which often accumulate in the bacterial cytoplasm in inclusion bodies[186]. Whilst proteins can be recovered, solubilized and refolded from inclusion bodies, as demonstrated with N-terminal domains of tomato NLR I-2[104], the recovery yield of recovered bioactive protein is often low, at around 15-25% of total protein[187].

Whilst the majority of published plant NLR structures have utilised an *E. coli* expression system, these have thus far been restricted to single domains of plant NLRs such as the N-terminal regions and integrated domains, see **Table 3.1**. For example, *E. coli* systems have already proved successful for expressing both the TIR domains of RRS1 and RPS4[114] and RRS1's WRKY domain[71]. The ease of high-throughput screening and proven success with expressing singular domains of RRS1 and RPS4 therefore make *E. coli* an attractive system to trial for the purposes of this study.

As a eukaryotic expression system, the *S. frugiperda* insect cells provide a protein folding environment more akin to the native environment of plant NLRs. Widely utilised in the pharmacological and vaccine fields, insect cells have also proven to be a suitable expression system for both the full-length plant NLRs such as ZAR1[108,109] and MLA27[178] and mammalian NLRs such as NLRC4[101]. The major benefit offered by insect cell expression systems compared to bacterial based systems is in insect cells' eukaryotic protein processing capabilities. Insect cells are capable of processing proteins with post translation modifications such as glycosylation and the formation of disulphide bonds both of which are lacking in bacterial systems. It should be noted however that the glycosylation patterns

Investigating heterologous expression systems for plant NLR expression

of insects differ from other organisms which can lead to issues with recombinant protein bioactivity[188].

Though the scaling up of insect cell production once a system is in place is relatively straightforward, high-throughput screening in this system is not as quick or simple as bacterial systems due to the need to amplify Baculovirus viral stocks for expression transfection which can take several weeks. Insect cells also require more complex and costly culturing conditions than bacterial cells in terms of scale up and maintenance of cells. Using dedicated high-throughput screening facilities such as those provided at OPPF however make screening in Baculovirus-insect cell systems a viable option for this study. This combined with the proven success of multi NLR-domain soluble protein expression in insect cells highlighted this system as a strong candidate for trialling expression of domains of RRS1 and RPS4.

This chapter will investigate the use of *E. coli* and *S. frugiperda* Sf9 insect cell expression systems for the expression of RRS1 and RPS4 protein. The aim of the work in this chapter was to conduct a large-scale screen of single and multi-domain constructs of RRS1 and RPS4 to identify potential constructs to take forward for scale up purification for biochemical and structural studies. The screen focussed on the well-studied NLR pair RRS1 and RPS4 such that any structural or biochemical data produced could be used to support and further probe our biological understanding of the activation mechanism of these NLRs.
This chapter will discuss the results of these screens of which full methodologies can be found in chapter 2. A list of the cloned vectors and maps can be found in Appendix 2 and 4.

3.2   Producing RRS1 and RPS4 cDNA for heterologous expression

In order to generate truncation variants of RRS1 and RPS4 for expression trialling in heterologous systems, cDNA constructs of RRS1 and RPS4 was generated, methods for which are described in 2.2.8. The first strategy for generation of these cDNA constructs was to purify RRS1 and RPS4 cDNA in *N. benthamiana* to be used as a template for PCR amplification in subsequent cloning. As such, *N. benthamiana* leaves were infiltrated with agrobacterium carrying vectors for 35S::RRS1^{Ws-2}-HF(gDNA) or 35S::RPS4$_{col-0}$-HF (gDNA) (plasmids were generated by Dr Yan Ma). Tissue was subsequently harvested 3 dpi and total RNA purified. RNA quality checked by gel electrophoresis before subsequent cloning, **Fig 3.1A**. cDNA was then generated from the RNA library by reverse transcription. Initial

attempts to produce full-length cDNA constructs of RRS1 and RPS4 focussed on amplifying full-length coding sequences of each gene from the purified *N. benthamiana* cDNA library which would then be subsequently cloned into a golden gate level 1 acceptor vector with a C-terminal HF-tag. Following optimisation of PCR conditions, which looked at trialling annealing temperature and extension time, a full-length PCR product of RRS1^{Ws-2} was generated (annealing temperature 58^{0}C extension time 6 minutes), **Figure 3.1B**, and cloned into a golden gate level 1 acceptor vector. However, attempts to amplify a full-length RPS4^{col-0} coding sequence PCR product resulted incorrectly assembled transformants despite multiple attempts and therefore could not be achieved.
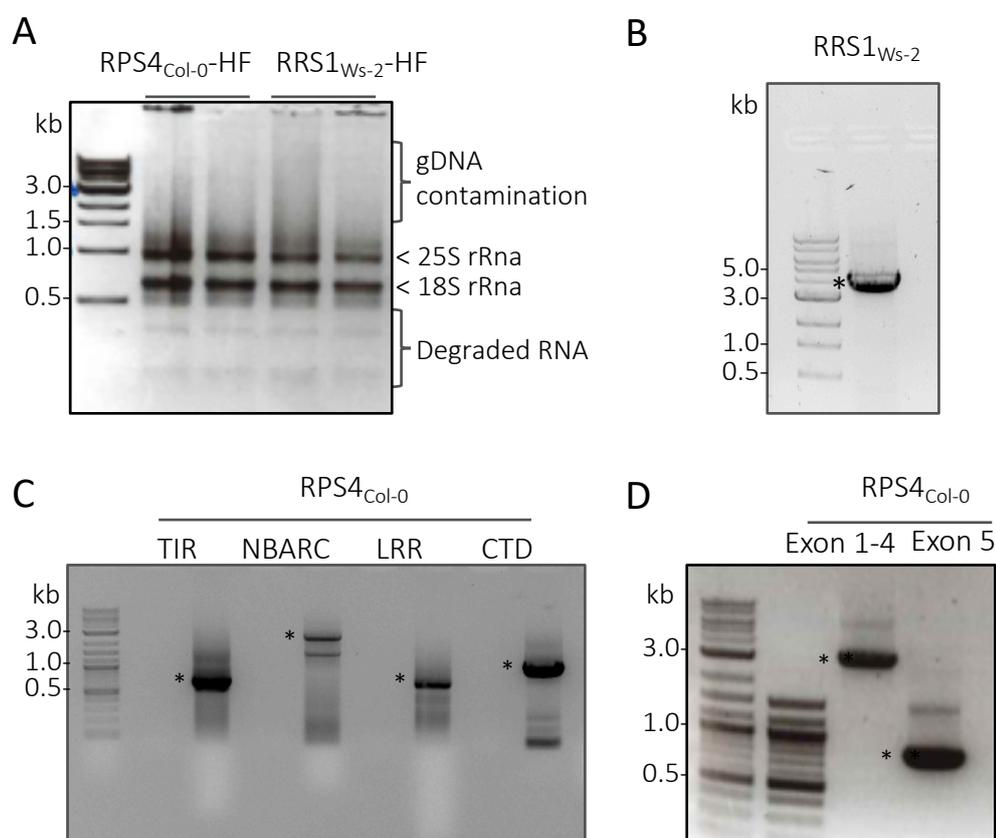
Figure 3.1 Cloning of RRS1^{Ws-2} and RPS4 cDNA. **(A)** Checking for *N. benthamiana* transiently expressing 35s::RPS4-HF or 35s::RRS1-HF purified RNA integrity by gel electrophoresis. **(B)** RRS1^{Ws-2} cDNA full-length PCR product cloned from cDNA library. **(C)** PCR amplification of individual domains of RPS4 cDNA from *N. benthamiana* cDNA. **(D)** Amplification of ligated exons 1-4 and exon 5 of RPS4 from RPS4 gDNA plasmid. Asterisks indicate band of expected size, DNA and RNA visualised on 1% TAE agarose gel with 1Kb DNA ladders.

As there were problems in cloning a single full-length PCR product of RSP4, the next strategy involved PCR amplifying each individual domain of RPS4 (TIR, NB-ARC, LRR and CTD) from the *N. benthamiana* cDNA library, **Figure 3.1C**, which were then assembled into a

single construct using digestion-ligation reaction. Although this method resulted in the generation of a few transformants, all contained mutations or base pair dropouts in the cDNA sequence. Due to issues in attaining reliable quantities of RPS4 amplified coding sequence from the cDNA library, I also tried amplifying each of RPS4's five exons from a RPS4$_{col-0}$ gDNA plasmid generated by Dr Pingtao Ding. These exon PCR products were then used to produce a full-length cDNA through an exon reassembly digestion-ligation reaction, method found in 2.2.4 . The aim of this was to increase the efficiency of the digestion-ligation reaction, resulting in the generation of more transformants which could then be screened for correct assembly of RPS4 cDNA sequence. This however also failed to generate a full-length cDNA construct with the longest correct construct reassembled including only RPS4 exons 1-4 (1-2856 bp cDNA). Attempts to ligate PCR products of exons 1-4 and exon 5, **Figure 3.1D**, resulted in either no transformants or recovered transformants displaying a consistent 63 bp deletion in the CTD of RPS4 (3442-3504 bp), **Figure 3.2**.

Figure 3.2 RPS4 69 bp deletion found in cDNA clones. Emboldened text indicates region of gene consistently found deleted in recovered RPS4 cDNA *E. coli* clones representing RPS4 G1148-N1168.

This led to the hypothesis that RPS4 cDNA may be lethal in *E. coli*. This would explain our ability to only recover transformants which had appeared to undergo a deletion in the CTD

of RPS4 which may render the RPS4 protein non-toxic. As such deletions may be result of homologous recombination events, I investigated cloning with RPS4 exons fragments in ElectroMAX™Stbl4™ strain of *E. coli.* This strain is recommended for use in cloning with unstable inserts and are optimised to reduce frequency of homologous recombination. However, the same unsuccessful transformant pattern was also observed using this strain. Dr Zane Duxbury also tried using the Gibson Assembly method of cloning to reassemble RPS4 exons but again was unable to recover full-length transformants. I subsequently submitted the full-length $RPS4_{Col-0}$ cDNA sequence for commercial synthesis (Twist Bioscience) but the company was also unable to recover a verified full-length sequence clone.

This led us to believe that full-length $RPS4_{Col-0}$ cDNA is likely toxic in *E. coli.* Using the various exon reassembly constructs that were generated during attempts to create a full-length cDNA construct, I was able to clone a variety of domain truncations of RPS4 for expression testing. Interestingly, none of the truncations appeared toxic to *E. coli* suggesting that all domains are required in situ in order for RPS4 to be toxic to *E. coli*.

3.3 RRS1-R cDNA is functional *in planta*

To confirm the biological functionality of the RRS1 cDNA construct generated in 3.2, RRS1-R_{cDNA}-HF was transiently expressed in the plant model organisms *N. benthamiana* and *N. tabacum* by agroinfiltration. Agroinfiltration involves infiltrating leaf tissue with *A. tumefaciens* transformed with a gene of interest. Once in the intracellular space of the leaf, the agrobacterium transfers tDNA of the gene of interest to the plant cells which leads to transient expression of the encoded genes *in planta*[189].

To confirm functionality, RRS1-R_{cDNA}-HF was transiently expressed in *N. tabacum* along with RRS1's partner RPS4-HA and recognised effector AvrRps4-mcherry. Agroinfiltration of RRS1-R_{gDNA}-HF (construct generated by Dr Yan Ma) was used as a positive control. 3 days post infiltration (3 dpi), a cell death HR was visible in tissue infiltrated with RPS4-HA and AvrRps4-mcherry with either RRS1-R_{cDNA}-HF or RRS1-R_{gDNA}-HF, **Figure 3.3A**. This indicated that the RRS1 cDNA construct translated to a functional protein *in planta* that was capable of activating RPS4 in an effector dependent manner. *N. tabacum* shows induction of a rapid HR upon activation of the RRS1/RPS4 complex. HR degrades the leaf tissue such that sampling infiltrated leaves to confirm the presence of expressed protein is not possible in

this species. Conversely, *N. benthamiana* shows a delayed and leaf-position dependant cell-death HR following AvrRps4-RRS1-RPS4 induced activation. This means it is possible to collect leaf tissue for protein expression confirmation before cell death degrades the leaf tissue. Therefore, to confirm expression of proteins *in planta* the same constructs used for *N. tabacum* HR assays were infiltrated in to *N. benthamiana* tissue for protein expression confirmation. Accordingly, RPS4-HA, AvrRps4-mcherry, mcherry, RRS1-R$_{cDNA}$-HF and RRS1-R$_{gDNA}$-HF constructs were transiently expressed in *N. benthamiana* leaves before tissue was harvested 3dpi.

Total protein extracts were purified from each leaf sample and protein expression confirmed by immunoblot analysis, **Figure 3.3B,C**. RRS1-R$_{cDNA}$-HF and RRS1-R$_{gDNA}$-HF appeared to express to proteins of the same size and expression level confirming the appropriate expression and functionality of RRS1-R$_{cDNA}$-HF. HR assays were repeated across three technical replicates and protein expression in two biological replicates with the same result observed across all repeats.

Similar analysis was conducted with RRS1B cDNA which was also used for heterologous expression in this study. A RRS1B$_{cDNA}$-HF construct was generated by Dr Maximiliano Jiménez Dalmaroni which I subsequently tested for protein expression *in planta* in *N. benthamiana* alongside a RRS1B$_{gDNA}$-HF construct generated by Dr Yan Ma. Western blot analysis confirmed appropriate expression of both RRS1B$_{cDNA}$-HF and RRS1B$_{gDNA}$-HF in *N. benthamiana*, **Figure 3C**. This protein expression experiment was conducted across 2 biological replicates with the same result observed across all repeats.
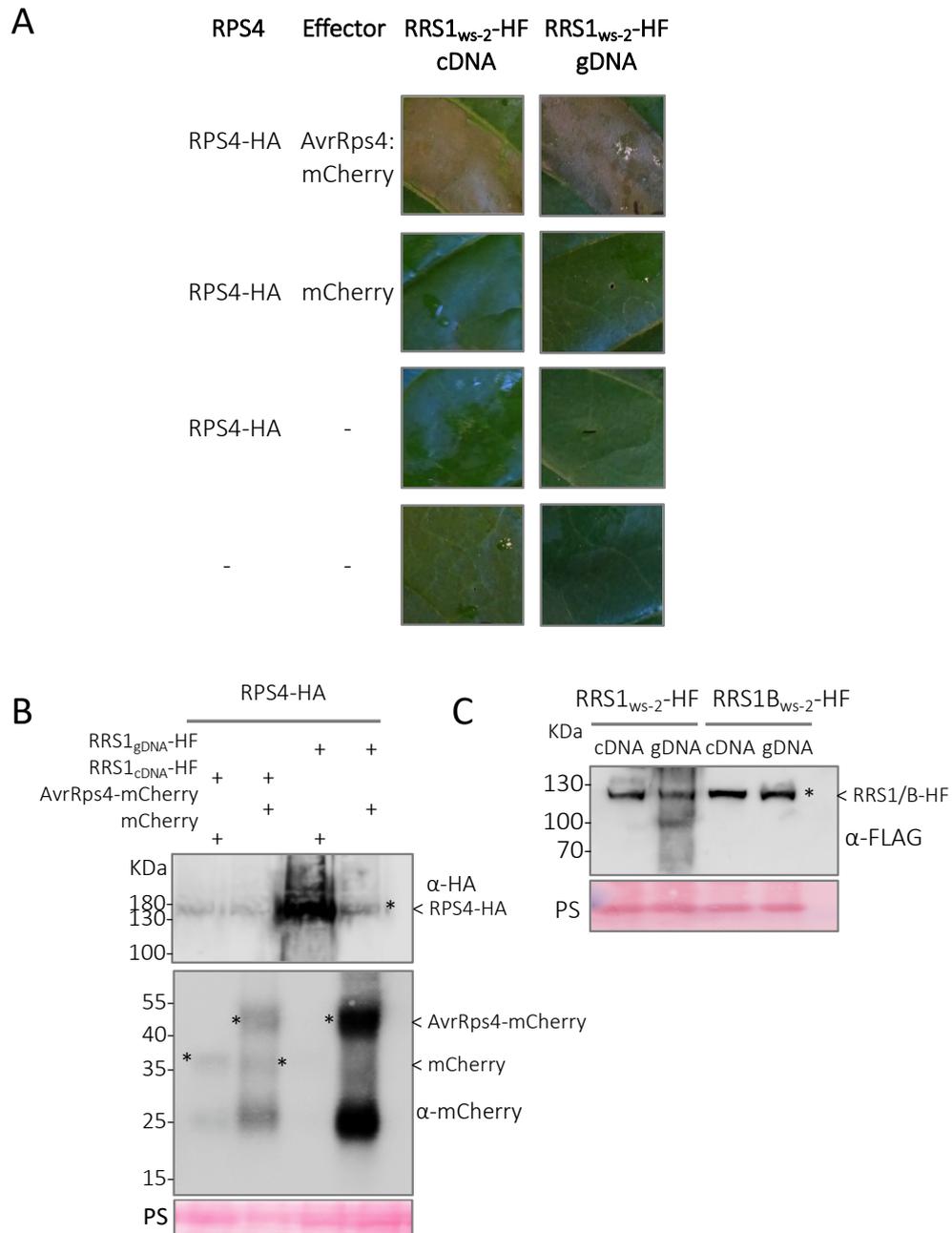
Figure 3.3 RRS1-HF cDNA is expressed and functional in planta. (A) Transient assay in *N. tabacum* leaves using agrobacterium infiltration shows functionality of RRS1-HF cDNA. Each leaf section was co-infiltrated to express different combinations of a RRS1-HF cDNA/gDNA, RPS4-HA and mcherry/AvrRps4-mcherry. Pictures were taken 5 days post infiltration, all pictures to same scale. **(B)** Immunoblot confirmation of protein expression RPS4, mcherry and AvrRps4-mcherry constructs in *N. benthamiana* total protein extract 3 dpi. **(C)** Immunoblot detection of RRS1-HF cDNA/gDNA and RRS1B-HF cDNA/gDNA expression in *N. benthamiana* total protein extract. Bottom panel shows ponceau staining of membrane as loading control. Protein were visualised using an HRP-conjugated anti-FLAG or anti-HA antibody or primary anti-mcherry followed by HRP-conjugated α-rabbit IgG. Asterisks indicates presence of protein of predicted molecular weight. Experiment A and C were conducted twice with similar results found across repeats, experiment B was conducted once.

3.4 High-throughput heterologous system expression screening of RRS1 and RPS4

Note: All RRS1 and RPS4 constructs described in 3.4 were cloned at TSL by myself with the exception of construct 39-43 which were cloned by Dr Lennart Wirthmüller. I then conducted screening of constructs in E. coli and Sf9 cells at OPPF under the supervision of Heather Rada (OPPF). All E. coli screening was conducted by myself with the exception of samples with gel codes A7-B8 which were cloned by myself but screened by Dr Maximiliano Jiménez Dalmaroni (JIC) at OPPF. I generated the P0 virus stock for Sf9 insect screening under the supervision of Heather Rada who was then responsible for subsequent expression transfection and harvesting of Sf9 cells.

Success of downstream structural and biochemical studies strongly hinges on the rational design of protein expression constructs. Often the native termini of eukaryotic proteins are not suitable for expression in heterologous systems due to the presence of disordered protein in these regions[190]. Therefore, expression of truncated protein or individual domains of multi-domain proteins of interest, often increases the chances of soluble protein expression and subsequent crystallization. Consequently, I decided to conduct a large-scale expression screen of various domain truncations of RRS1 and RPS4, to identify constructs for use in downstream studies. I decided to conduct trials in both *E. coli* and Sf9 insect cells, as I predicted the more complex eukaryotic cell environment of Sf9 cells may better facilitate soluble expression of multi-domain constructs NLR proteins, whilst *E. coli* has proven a successful system for singular domain NLR expression, see **Table 3.1**. These screens were conducted at OPPF (Oxford Protein Production Facility) using their high-throughput protein heterologous expression screenings systems designed to increase the likelihood of identifying constructs with soluble expression.

Truncation variants for expression trialling of RRS1 and RPS4 were designed by myself and Dr Lennart Wirthmüller (JIC). Constructs were designed to span the entire lengths of RRS1 and RPS4 and were designed by two methods. Firstly, using the domain boundaries of RRS1 and RPS4 previously predicted by members of the Jones lab. The TIR, NB-ARC and LRR domains of RRS1 and RPS4 had been previously defined on the basis of sequence homology with boundaries in other characterised plant TNLs, and the WRKY domain as the 22-25 amino acids preceding the 'WRKYGQK' amino acid motif till the end of the zinc-finger ('HNH' amino acids sequence). The second method involved using RONN protein disorder

Investigating heterologous expression systems for plant NLR expression

prediction analysis[162] (https://www.strubi.ox.ac.uk/RONN) or IUPred2A[163] (https://iupred2a.elte.hu/), see **Figure 3.4**, to design constructs which excluded regions of protein disorder at their N- and C-terminuses of constructs, inclusion of which can often hinder soluble expression of a protein. This information was supplemented with secondary protein structure prediction analysis using PHYRE2[164] (www.sbg.bio.ic.ac.uk/phyre2/) which was used to ensure the N- and C-terminal regions of protein expression constructs did not terminate mid-secondary structure which can lead to the destabilisation of a protein.

Previous bimolecular fluorescence complementation (BiFC) studies using a split cCFP and nVenus in the Jones laboratory had shown that Dom6S and Dom4 interacted using the construct cCFP-RRS1 Dom4-Dom6S-nVenus[152]. I hypothesised that the interaction of the two fluorescent molecules may help to lock down the conformation of these RRS1 domains and stabilise the protein, aiding solubility. To test this, two constructs containing split cCFP and nVenus on either end of RRS1 Dom4-WRKY-Dom6S were also considered for expression testing (constructs 23 and 24).

In addition to RRS1 and RPS4, three WRKY transcription factors (WRKY18/40/41) were also included in expression trials due to these proteins varying interactions with the RRS1 recognised effectors PopP2 and AvrRps4. *At*WRKY41 belongs to the same group of WRKYs as RRS1-R, group III, and also interacts with both the effectors AvrRps4 and PopP2, with the latter acetylating the WRKY domain[58]. *At*WRKY18 and *At*WRKY40 belong to group IIa WRKY proteins but interestingly only *At*WRKY18 is acetylated my PopP2[66].Therefore, *At*WRKY41 represents an interesting positive control as a possible host target of AvrRps4 and PopP2 and *At*WRKY18 represents a negative control for biochemical interaction studies as a WRKY proteins which does not interact with these effectors. The three constructs included expressed *At*WRKY18$_{A157-E240,}$ *At*WRKY40$_{V124-N213}$ and *At*WRKY41$_{L121-E208}$ in the expression trial, and were cloned by Dr Lennart Wirthmüller

Though no single tag is considered a 'silver-bullet' for soluble expression of all proteins universally, fusion of solubility tags to POIs can often improve the stability and yield of proteins[191]. For this study, constructs were cloned into pOPIN vectors under T7 promoters with a variety of solubility tags which have been observed in the literature to improve protein solubility and yield along with a 6xHis affinity tag required for Nickel-NTA protein purification. Tags trialled in this screen included SUMO tag[192] (pOPINS3C) and MBP[193] (pOPINM) as well 6xHis tag only (pOPINF). This resulted in the generation of 43 constructs

listed in **Table 3.2** which I then took to OPPF for high-throughput expression screening. The pipeline for screening in *E. coli* and Sf9 cells can be seen in **Figure 3.5**.

RRS1

MTNCEKDEEFVCISCVEEVRYSFVSHLSEALRRKGINNVVVGV**DSDDLLFKESQA**KIEKAGVSVMVLPGNCDPSDVWLDKFAK
VLECQRNNKDQAVVPVLYGDSLLRDQWLSELD**FKGLSRIHQSRKECS**DSILVEEIVRDVYETHFYVGRIGIYSKLLEIENMVN

TIR: 1M-G153

KQPIGIRCVGIWGMPGIGKTTLAKAVFDQMSSAFDASCFIEDYDKSIHEKGLYCLLEEQLLPGNDATIM**KLSSL**RDRLNSKRV
LVVLDDVRNALVGESFLEGFDWLGPGSLIIITSRDKQVFCLCGINQIYEVQGLNEKEARQLFLLSASIKEDMGEQNLQELSVR
VINYANGNPLAINVYGRELKGKKK**LSEMETA**FLKLKRRPPFKIVDAFKSTYDTLSDNEKNIFLDIACFFQGENVNYVIQLLEG
CGFFPHVEIDVLVDKCLVTISENRVWLHKLTQDIGREIINGETVQIERRRRLWEPWSIKYLLEYNEHK**ANGEPKTTFKRAQGS**
EEIEGLFLDTSNLRFDLQPSAFKNMLNLRLLKIYCSNPEVHPVINFPTGSLHSLPNELRLLHWENYPLKSLPQNFDPRHLVEI
NMPYSQLQKLWGGT**KNLEMLRTIRLCHSHHLVDIDDLLKAENLEVIDLQGCTRLQNFPAAGRLLRLRDVNLSGCIKIKSVLEI**

NB-ARC: I154-T595

PPNIEKLHLQGTGILALPVSTVKPNHREL**VNFLTEIPGLSEASKLERLTSLLESNS**SCQDLGKLICLELKDCSCLQSLPNMAN
LDLNVLDLSGCSSLNSIQGFPRFLKQLYLGGTAIREVPQLPQSLEILNAHGSCLRSLPNMANLEFLKVLDLSGCSELETIQGF
PRNLKELYFAGTTLREVPQLPLSLEVLNAHGSDSEKL

LRR: K596-L867

PMHYKFNNFFDLSQQVVNDFFLKTLTYVKHIPRGYTQELI**NKAPTF**
SFSAPSHTNQNATFDLQPGSSVMTRLNHSWRNTLVGFGMLVEVAFPEDYCDATDFGISCVCRWSNKEGRSCRIERNFHCWAPG
KVVPKVRKDHTFV**FSDVNMRPSTGEG**NDPDIWAGLVVFEFFPINQQTKCLNDRFTVTRCGVRVINVATGNTSLENISLVLSLD
PVEVSGYEVLRVSYDDLQEMDKVLFLYIASLFNDEDVDFVAPLIAGIDLDVSSGLKVLADVSLISVSSNGEIVMHSLQRQMGK

Dom4: P868-K1189

EI**LHGQSMLLSDCESSMTENLSDVPKK**EKKHRESKVKKVVSIPAIDEGDLWTWRKYGQKDILGSRFPRGYYRCAYKFTHGCKA

WRKY: E1190-H1269

TKQV**QRSETDSNMLAITYLSEHNH**PRPTKRKALADSTRSTSSSIC**SAITTSASSRVFQNKDEPNKPHLPSSSTPPGNAAV**LFK

Dom6S | Dom6R

MTDMEEFQDNMEVDNDVVDTRTLALFPEFQHQPEEEYPWSTFFDY

Dom6: P1270-Y1373

RPS4

METSSISTVEDKPPQHQVFINFRGADLRRRFVSHLVTALKLNNINVFIDDYEDRGQPLDVLLKRIEESKIVLAIF
SGNYTESVWCVRELEKIKDCTDEGTLVAIPIFYKLEPSTVRDLKGKFGDRFRSMAKGDERKKKWKEAFNLIPNIM
GIIIDKKSVESEKVNEIVKAVKTALTGIPPEGSHN**AVVGALGNSNAGTSSGDKKHETFGNEQRLK**DLEEKLDRDK

TIR: 1M-G182

YKGTRIIGVVGMPGIGKTTLLKELYKTWQGKFSRHALIDQIRVKSKHLELDRLPQMLLGELSKLNHPHVDN**LKD**P
YSQLHERKVLVVLDDVSKREQIDALREILDWIKEGKEGCSIVIPTSDMSLTNGLVDDTYMVQNLNHRDSLQLFHY
HAFIDDQANPQKKDFMKLSEGFVHYARGHPLALKVLGGELNKKSMDHWNS**KM**KKLAQSPSPNIVSVFQVSYDELT
TAQKDAFLDIACFRSQDKDYVESLLASSDLGSAEAMSAVKSLTDK**F**LINTCDGRVEMHDLLYKFSREVDLKASNQ
DGSRQRRLWLHQHIIKGGIINVLQNKMKAANVRGIFLDLSEVEDETSLDRDHFINMGNLRYLKFYNSHCPQECKT
NNKINIPDKLKLPLKEVRCLHWLKFPLETLPNDFNPINLVDLKLP**Y**SEMEQLWEGDKDTPCLRWVDLNHSSKLCS

NB-ARC: S183-T659

LSGLSKAEKLQRLNLEGCTTLKAFPHDMKKMKMLAFLNLKGCTSLESLPEMNLISLKTLTLSGCSTFKEFPLISD
NIETLYLDGTAISQLPMNMEKLQRLVVLNMKDCKMLEEIPGRVGELKALQELILSDCLNLKIFPEIDISFLNILL
LDGTAIEVMPQLPSVQYLCLSRNAKISCLPVGISQLSQLKWLDLKYCTSLTSVPEFPPNLQCLDAHGCSSLKTVS

LRR: P670-P879

KPLARIMPTEQNHSTFIFTNCENLEQAAKEEITSYAQRKVT**Y**SYARKRYNGGLVSESLFSTCFPGCEVPSWFCH
ETVGSELEVKLLPHWHDKKLAGIALCAVVSCLDPQDQVSRLSVTCTFKVKDEDKSWVAYTCPVGSWTRHGGGKDK
IELDHVFIGYTSCPHTIKCHEEG**N**SDECNPTEASLKFTVTGGTSENGKYKVLKCGLSLVYAKDKDKNSALETKYD
MLIGKSFQETSE**GVD**GR**VKKTK**GK**YVMPVEKNFQETTEGVDGRVNKKKKTRMDNGRPKKK**QRSGRDDNQTRMQVE
LQEGNINSVIM**H**TVKNF

CTD: E880-F1217

Figure 3.4 Predicted disordered region of RRS1 and RPS4. Emboldened amino acids highlight predicted disordered regions. Disorder precited using RONN[162] and IUPred2A[163] analysis. Predicted domains of RRS1 and RPS4 are highlighted in various colours with N- and C-termini domain boundary amino acids listed to the right.

Table 3.2 RRS1 & RPS4 expression construct trialled at OPPF. Gel code refers to lane reference in SDS-PAGE gels, MW indicates predicted molecular weight of expressed protein with solubility tag

Construct Number	Gene	Domain	Amino acids	Vector	Gel code	MW (kDa)
1	RRS1	TIR-Dom6R	M1-Y1373	pOPINF	A1	157.6
2	RRS1	NB-ARC	I154-T595	pOPINF	E1	52.9
3	RRS1	NB-ARC-Dom6S	I154-C1290	pOPINF	C7	130.7
4	RRS1	NB-ARC-Dom6S	I154-C1290	pOPINS3C	D7	141.6
5	RRS1	NB-ARC-Dom6R	I154-Y1373	pOPINF	E7	140.0
6	RRS1	NB-ARC-Dom6R	I154-Y1373	pOPIS3C	F7	151.0
7	RRS1	NB-ARC-LRR	I154-T595	pOPINS3C	H7	93.8
8	RRS1	LRR	K596-L867	pOPINS3C	G7	43.4
9	RRS1	LRR-Dom4	K596-K1189	pOPINF	F1	68.4
10	RRS1	LRR-Dom6R	K596-Y1373	pOPINF	G1	89.5
11	RRS1	LRR-Dom6R	K596-Y1373	pOPINS3C	A8	100.4
12	RRS1	Leucine Zipper motif-Dom6S	L1089-C1290	pOPINF	C1	25.0
13	RRS1	Leucine Zipper motif-Dom6R	L1089-Y1373	pOPINF	B1	34.4
14	RRS1	Leucine Zipper motif-Dom6R	L1089-Y1373	pOPINS3C	D1	45.3
15	RRS1	Dom4	P868-K1189	pOPINF	H1	38.4
16	RRS1	Dom4	P868-K1189	pOPINS3C	A2	49.3
17	RRS1	Dom4-WRKY	L929-T1273	pOPINF	F4	41.1
18	RRS1	Dom4-WRKY	A1063-T1273	pOPINF	G4	25.8
19	RRS1	Dom4-Dom6S	P868-C1290	pOPINF	D2	50.0
20	RRS1	Dom4-Dom6S	P868-C1290	pOPINS3C	E2	60.9
21	RRS1	Dom4-Dom6R	P868-Y1373	pOPINF	B8	59.5
22	RRS1	Dom4-Dom6R	P868-Y1373	pOPINS3C	C2	70.4
23	RRS1	cCFP-Dom4-Dom6S-nVenus	P868-C1290	pOPINS3C	A7	93.6
24	RRS1	cCFP-Dom4-Dom6S-nVenus	P868-C1290	pOPINF	B7	82.7
25	RRS1	WRKY-Dom6R	S1184-Y1373	pOPINF	F2	24.0
26	RRS1	WRKY-Dom6R	E1209-Y1373	pOPINM	G2	59.0
27	RRS1	WRKY-Dom6R	R1194-Y1373	pOPINS3C	H2	33.7

Construct Number	Gene	Domain	Amino acids	Vector	Gel code	MW (kDa)
28	RRS1 *K1219Q*	WRKY-Dom6S	E1209-C1290	pOPINF	A3	11.7
29	RRS1 *K1219Q*	WRKY-Dom6S	E1209-C1290	pOPINS3C	B3	22.6
30	RRS1 *K1219R*	WRKY-Dom6S	E1209-C1290	pOPINF	C3	11.7
31	RRS1 *K1219R*	WRKY-Dom6S	E1209-C1290	pOPINS3C	D3	22.6
32	RRS1 *slh1*	WRKY-Dom6S	E1209-C1290	pOPINS3C	E3	22.6
33	RRS1 *slh1*	WRKY-Dom6S	E1209-C1290	pOPINF	F3	11.7
34	RPS4	TIR-LRR	M1-P879	pOPINF	G3	102.0
35	RPS4	TIR-LRR	M1-P879	pOPINS3C	H3	112.9
36	RPS4	NB-ARC	S477-T659	pOPINF	C4	56.7
37	RPS4	NB-ARC	S477-T659	pOPINS3C	D4	67.7
38	RPS4	LRR-CTD	P660-F1217	pOPINS3C	A4	75.5
39	RPS4	CTD	E880-F1217	pOPINF	B4	40.0
40	RPS4	CTD	L954-L1120	pOPINF	E4	20.5
41	*At*WRKY 41	WRKY	L121-E208	pOPINF	H4	12.8
42	*At*WRKY 18	WRKY	A157-E240	pOPINF	A5	11.5
43	*At*WRKY 40	WRKY	V124-N213	pOPINF	B5	12.5
44	RRS1/RPS4	RRS1 LRR-Dom4 & RPS4 LRR-CTD	K596-K1189 & P660-F1217	pOPINF (Both)	C8	68.3 +75.5
45	RRS1/RPS4	RRS1 LRR-Dom6R & RPS4 LRR-CTD	K596-Y1373 & P660-F1217	pOPINF (Both)	D8	115.5 +75.5

3.4.1 *E. coli* high-throughput screen design

Expression trialling was conducted in *E. coli* with a number of variables outlined in **Figure 3.5**. Firstly, two different expression strains of *E. coli* were used Rosetta (DE3) pLys and Lemo21 (DE3). Rosetta (DE3) pLys strain has been developed to enhance expression of eukaryotic proteins often limited by codon usage in *E. coli* by supplying rare tRNAs scarcely used by *E. coli*. Rosetta (DE3) carries rare codons for amino acids Arginine (AGG, AGA and CGG), Isoleucine (AUA), Leucine (CUA), proline (CCC) and glycine (GGA). In this way the strain provides a more 'universal' translation environment which was of particular use to

this study as I was using RRS1 and RPS4 cDNA sequence which had not been codon optimised for *E. coli* use. The second strain trialled was Lemo21 (DE3). This strain allows for tuneable protein expression levels through the addition of L-rhamnose which varies the level of lysozyme, the natural inhibitor of T7 RNA polymerase, thereby modulating expression of the POI. This tuneable expression is of particular use for toxic proteins and those prone to insoluble expression as well as membrane proteins. The workflow used by OPPF initially expresses proteins in Lemo21 (DE3) without the addition of L-rhamnose with the use of L-rhamnose tuning used as a secondary optimisation of constructs which have low expression in the absence of L-rhamnose.

There is no universal optimal media for all *E. coli* protein expression[194]C. Consequently, trialling of multiple media conditions in parallel is required to identify a POI's optimal expression conditions. The screen at OPPF was conducted by growing cells in either Power Broth or Autoinduction media. These two medias offer different methods of protein expression induction which can have a large effect on soluble protein expression[195].

 Cells grown in Power Broth were induced with 1mM IPTG once cultures reached a set optical density (OD_{600}=0.5) whilst induction of AIM cultures is self-regulated by the culture's growth. AIM induction utilises the regulatory elements of the lac operon to deliver cell culture density dependant protein induction. The system relies on the switching from glucose to lactose metabolism once the preferentially metabolised glucose levels have been depleted by the growing *E. coli* cell culture[196]. Lactose metabolism then produces β-galactosidase which relieves the repression of the lac operon controlling T7 RNA polymerase expression. The use of AIM thereby places the transition from un-induced to induced under the metabolic control of growing culture. Cultures grown in AIM often reach higher cell densities than those grown in IPTG induced cultures. This can increase soluble yields of POI but can also lead to adverse issues with aeration.

Combining conditions tested between different *E. coli* strains and growth medias meant that each construct was tested in four varying conditions totalling 180 individual trials in total. The pipeline for screening in *E. coli* conducted at OPPF can be seen in **Figure 3.5**. All screens included expression of a GFP positive control vector to confirm to asses for correct transformation and growth of *E. coli* and insect cell cultures.
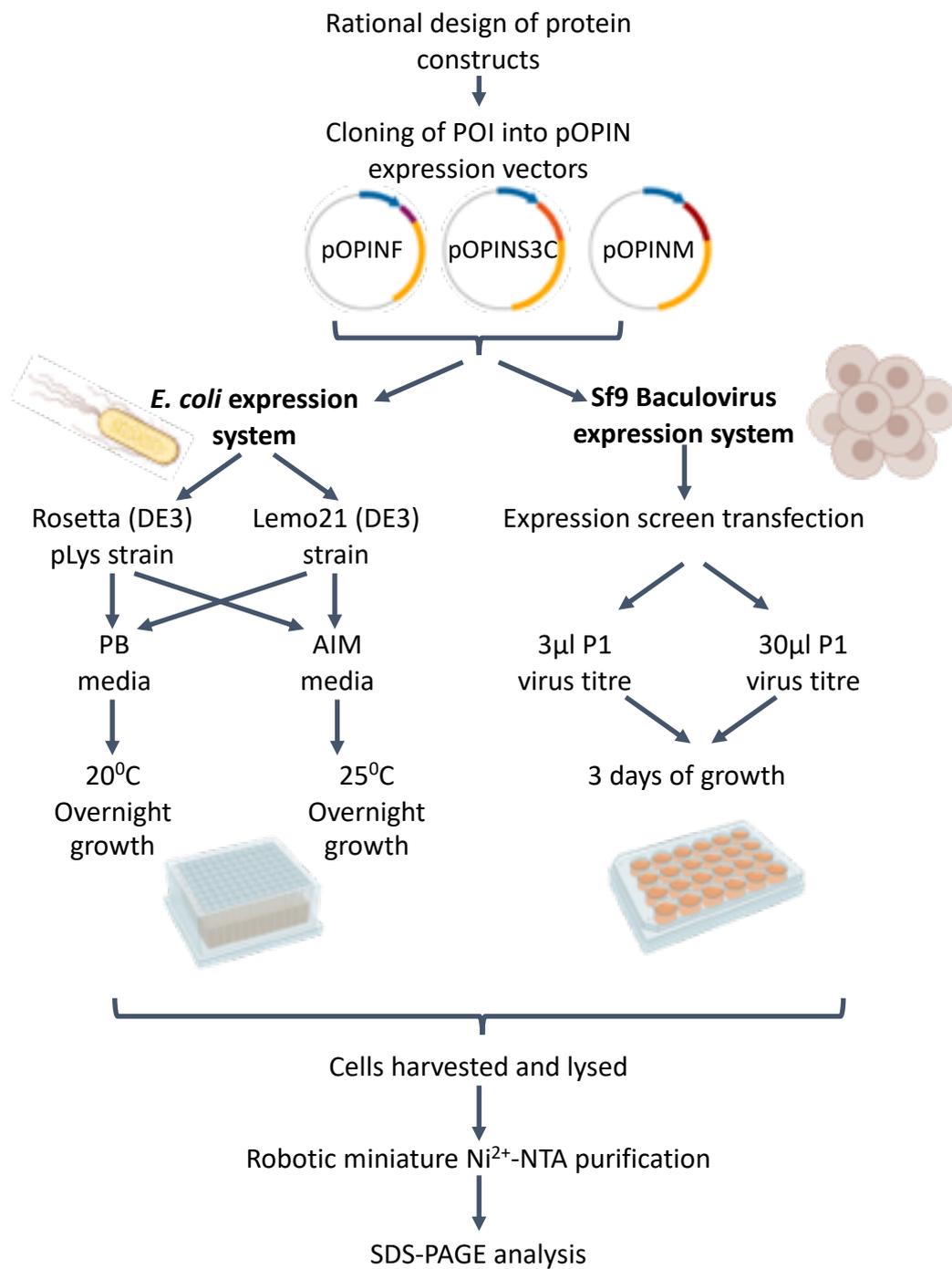
Figure 3.5 *E. coli* and Sf9 insect cell expression screening pipeline conducted at OPPF. Figure created using BioRender

3.4.2 *E. coli* high-throughput expression screen results

The *E. coli* expression screen highlighted five construct/condition combinations of RRS1 which showed soluble protein expression, although all except one combination yielded very low protein expression levels details of which can be found in **Table 3.3**. The one construct which yielded medium expression levels was construct 15-RRS1 Dom4$_{P868-K1189}$ pOPINF (gel code H1) expressed in Rosetta (DE3) pLys cells in AIM, **Figure 3.7D**. However, the expressed protein appears to be a truncation of the protein which this construct should express, as the SDS-PAGE gel band runs at around ~29 kDa rather than the expected size of 38.35 kDa. This band also correlates with a band seen in for this construct in Sf9 screen, section 3.4.4, which mass spectrometry analysis suggests represents P868-K1101 of RRS1 Dom4. This band was also visible in Rosetta (DE3) pLys cells grown in Power broth media but at lower expression levels than cells grown in AIM and was not observed to be expressed in Lemo21 cells. Subsequent scale up expression of this construct on a 6 litre growth media scale however resulted in very low protein yields and hence this construct was not taken forward for further study, data not shown.

Similarly, a truncated variant of ~36 kDa compared to 93.8 kDa predicted translated protein size of construct RRS1 NB-ARC LRR$_{I154-T595}$ pOPINS3C (gel code H7) was seen to be expressed in Rosetta (DE3) pLys cells grown in AIM, **Figure 3.7F**. Expression however was very low and therefore was not followed up with a large-scale purification. Constructs RRS1 NB-ARC$_{I154-T595}$ pOPINF (gel code E1) and RRS1 Leucine-zipper motif-Dom6R$_{L1089-Y1373}$ pOPINF (gel code B1) both appeared to be expressed but at very low levels in Rosetta (DE3) pLys cells grown in AIM, **Figure 3.7D**. The SDS-PAGE band size for both these constructs matches the predicted protein size for fully translated constructs suggesting the proteins were not being truncated in vitro. However, the expression levels observed in this is screen for these constructs are too low to carry forward with larger scale-up expression and purification.

One construct RRS1 WRKY-Dom6R$_{E1209-Y1373}$ appeared to be expressed in the soluble fraction of Lemo21 cells grown in AIM though expression was observed to be very low again and not suitable for scale up purifications, **Figure 3.6D**.

Crude lysate samples of Rosetta (DE3) pLys and Lemo21 cells grown in AIM were analysed by SDS-PAGE to assess for protein expression in the total fraction. However, bands were not clear enough to confirm presence of POI in these samples.
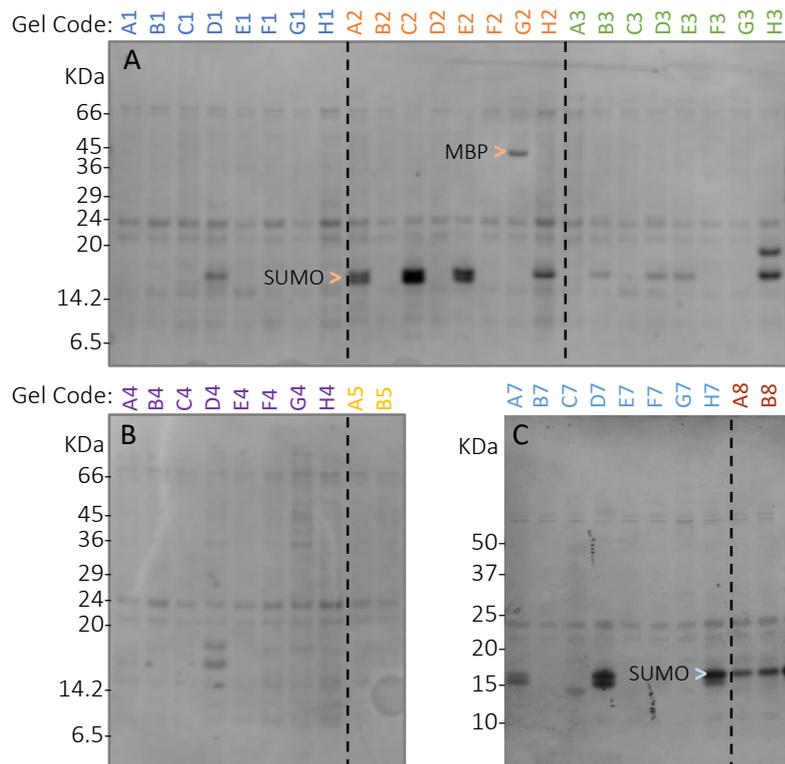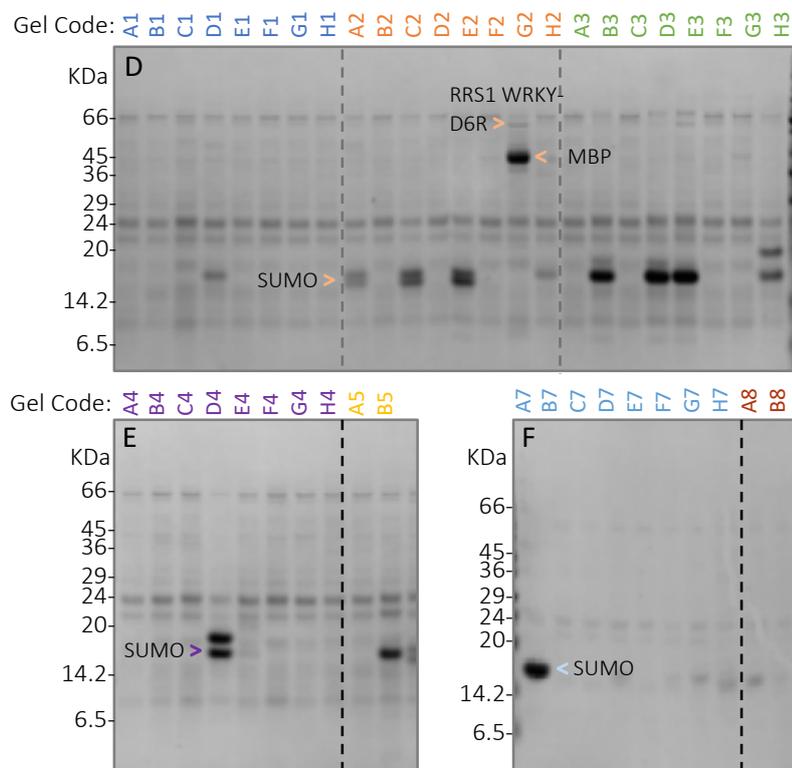
Figure 3.6 Coomassie stained SDS-PAGE of Lemo21(DE3) *E. coli* expression of RRS1 and RPS4 domains. SDS-PAGE soluble extract from Lemo21(DE3) *E. coli* cells following Nickel IMAC. Soluble protein expression is indicted by arrow. **(A-C)** cells were grown in AIM with overnight expression at 25°C, **(D-F)** cells were grown in PB media with overnight expression at 20°C. SDS-PAGE stained with Coomassie dye. Experiment was conducted once.
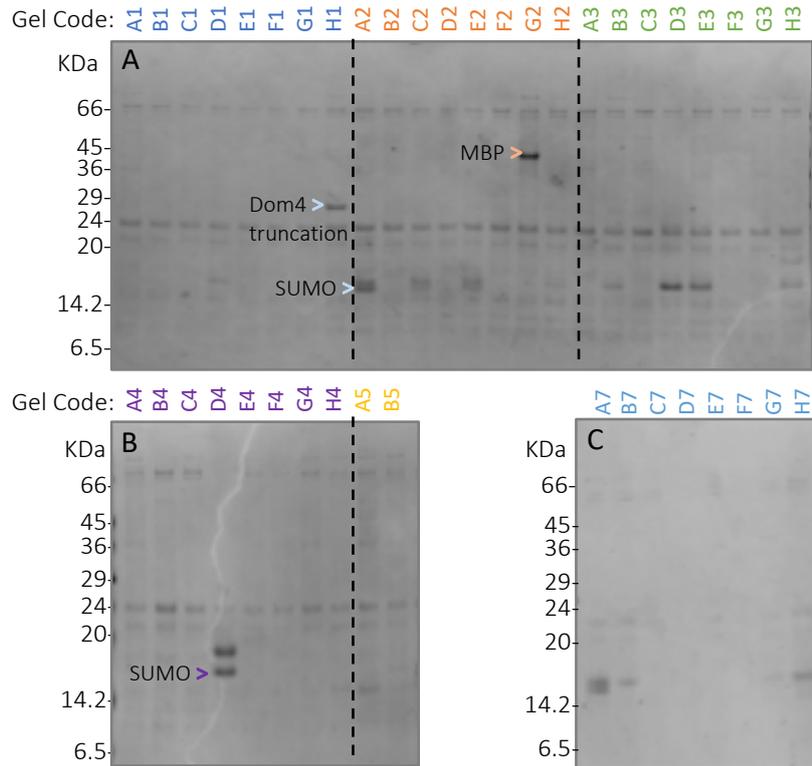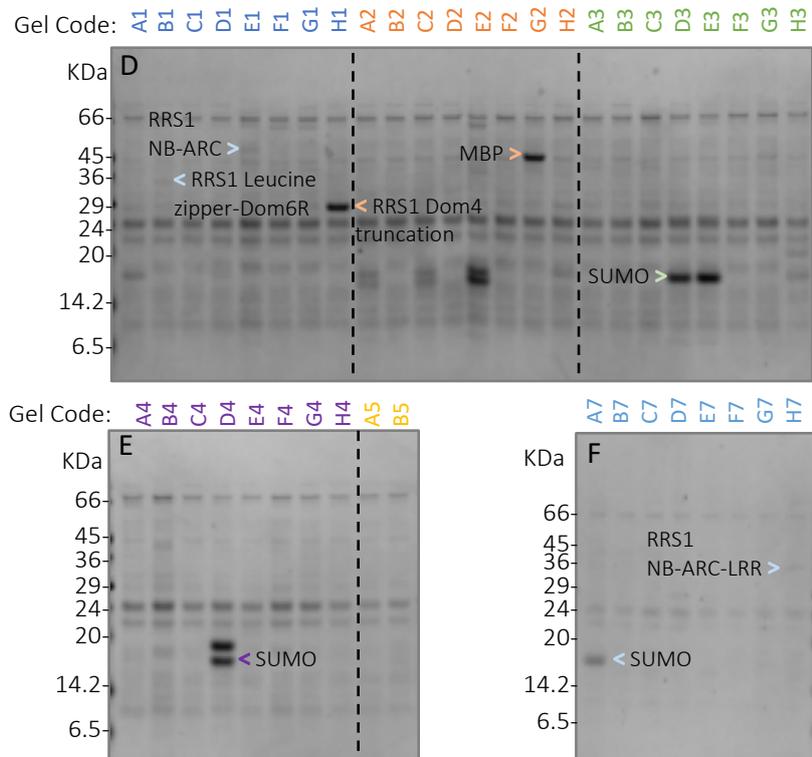
Figure 3.7 Coomassie stained SDS-PAGE of Rosetta (DE3) pLys *E. coli* expression of RRS1 and RPS4 domains. SDS-PAGE soluble extract from Rosetta (DE3) pLys *E. coli* cells Following nickel IMAC. Soluble protein expression is indicted by arrow. **(A-C)** cells were grown in AIM with overnight expression at 25°C, **(D-F)** cells were grown in PB media with overnight expression at 20°C. Experiment was conducted once.

Table.3.3 RRS1 and RPS4 soluble expression results from *E. coli* trials. Construct number is in reference to Table 3.2.

E. coli expression results

Construct Number	Gel code	Gene	Domain	Amino acids	Vector	Media	Expression line	Expression Level	Comments
2	E1	RRS1	NB-ARC	I154-T595	pOPINF	AIM	Rosetta	Very low	
7	H7	RRS1	NB-ARC-LRR	I154-T595	pOPINS3C	AIM	Rosetta	Very Low	Only ~36KDa truncation expressed
13	B1	RRS1	Leucine zipper motif-Dom6R	L1089-Y1373	pOPINF	AIM	Rosetta	Very low	
15	H1	RRS1	Dom4	P868-K1189	pOPINF	AIM	Rosetta	High	
26	G2	RRS1	WRKY-Dom6R	E1209-Y1373	pOPINM	AIM	Lemo	Very low	Only ~29KDa truncation expressed

Investigating heterologous expression systems for plant NLR expression

In summary it appears that *E. coli* is not a suitable heterologous expression to be used for the expression of full-length or truncated RRS1 and RPS4. This correlates with previous struggles in the field to express soluble plant NLR proteins in this host system. The expression of truncated constructs, albeit to very low levels, appeared to be significantly more successful in Rosetta (DE3) pLys cells than Lemo21 cells with 4/5 of occurrence of soluble expression in the screen seen in Rosetta (DE3) pLys. I hypothesise this is due to the fact that Rosetta (DE3) pLys cells carry several tRNAs which are rarely used in *E. coli* but are found in the cDNA sequence of RRS1^{Ws-2} and RPS4$_{Col-0,}$ see **Table 3.4.** Unfortunately, the bands in the total fraction samples from this screen which were also analysed by SDS-PAGE were not clear enough to confirm expression of constructs in the total fraction so this hypothesis cannot be confirmed (gels not shown as bands are unreadable).

Table 3.4 Codon occurrence in RRS1 and RPS4 of tRNAs supplied in Rosetta (DE3) cells (GenScript)

Amino acid	Codon	Codon occurrence in cDNA		Fraction for each codon within synonymous family		
		RRS1^{Ws-2}	RPS4^{Col-0}	*A. thaliana*	*E. coli*	*S. frugiperda*
Arg	AGG	10	13	0.2	0.04	0.21
	AGA	24	17	0.35	0.07	0.16
Ile	AUA	25	13	0.24	0.11	0.12
Leu	CUA	21	14	0.11	0.04	0.08
Pro	CCC	8	5	0.11	0.13	0.31
Gly	GGA	25	29	0.37	0.13	0.27

3.4.3   Baculovirus transfected Sf9 insect cells high-throughput screen design

High-throughput screening of Baculovirus transfected Sf9 insect cells was conducted using a P1 virus stock as though screening can be conducted with P0 virus, screening using P1 virus stocks gives a more reliable expression result. Infectious titre of Baculovirus viral stock has been shown in the literature to have potentially significant effects on the yield of soluble recombinant protein production[197]. To investigate this, screening at OPPF was setup using two different titres or MOI (multiplicity of infection) of P1 virus stock, 3 μl and 30 μl, **Figure 3.5**. As a general principle, if Sf9 cells are transfected with too little virus this can lead to poor protein yields as a synchronous infection across all Sf9 cells will not be established. If a subset of Sf9 cells are not infected after the addition of the virus stock then uninfected cells

may continue to multiply and increase the cell density to the stationary phase of culture growth where virus replication is inhibited. This results in only a small fraction of the Sf9 cell culture expressing the POI and reduces yield of the recombinant protein. Conversely, the addition of too higher titre of virus stock can be detrimental to the growth of the Sf9 cell culture and reduce yields of recombinant protein. Previous studies have shown the effect of a low and high MOI on recombinant protein expression appears to vary between proteins[198], with no universal correlation pattern and therefore different viral titres were investigated in this study.

3.4.4 Baculovirus transfected Sf9 insect cells high-throughput expression screen results

The Sf9 cell expression screen highlighted four construct/condition combinations of RRS1 and one of RPS4 which showed soluble protein expression. Whilst most constructs which exhibited soluble expression in the Sf9 trials were at expression levels too low to carry forward for scale up purification, RRS1 WRKY-Dom6R$_{E1209\text{-}Y1373}$ pOPINM (gel code G2) and RPS4 CTD$_{L954\text{-}L1120}$ pOPINF (gel code E4) showed expression levels which could yield good quantities of protein upon scale up, **Figure 3.8A,B,D**. Work continues at JIC by Dr Nitika Mukhi and Dr Richard Hughes to take forward these two constructs for scale up expression and purification to produce protein for crystallography and biochemical analysis.

The screen also showed that the *At*WRKY40$_{V124\text{-}N213}$ and *At*WRKY18$_{A157\text{-}E240}$ constructs to express well in Sf9 cells transfected with 3 µl of P1 Baculovirus stock. As expression of the potential host target *At*WRKY41 was not soluble however, scale up purifications of the negative effector interaction controls *At*WRKY40 and *At*WRKY18 have not been currently taken forward, **Figure 3.8B.**

Whilst RRS1 Dom4$_{P868\text{-}K1189}$ showed low levels of soluble expression at the predicted molecular weight when fused with a SUMO tag in the pOPINS3C vector (gel code H1), when expressed in the pOPINF vector (gel code A2) only a truncation of this construct appeared to show soluble expression, **Figure 3.8A,D**. In gel digestion mass spectrometry suggested this truncated soluble construct of ~ 28 kDa represented P868-K1101 of RRS1 Dom4 by peptide coverage. This might suggest the SUMO tag is helping to stabilize this protein in vitro. pOPINS3C however did not universally help against the expression of truncations of

Investigating heterologous expression systems for plant NLR expression

the cloned protein construct as only a ~27 kDa variant of the predicated 93.8 kDa RRS1 NB-ARC-LRR$_{I154-T595}$ in pOPINS3C showed low levels of soluble expression.

Sf9 insect cells: 3µl P1 virus
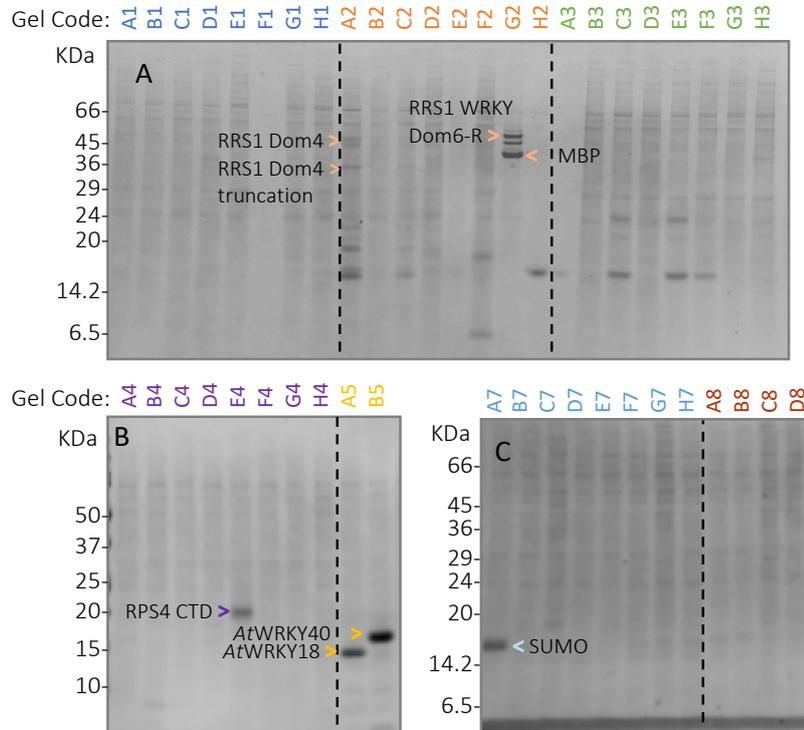
Sf9 insect cells: 30µl P1 virus
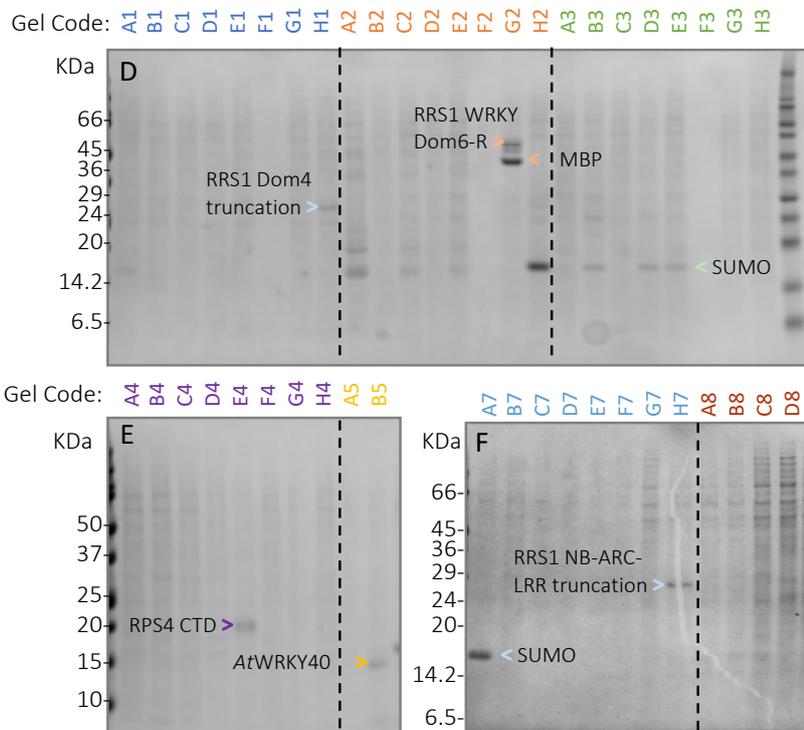
Figure 3.8 Coomassie stained SDS-PAGE of Sf9 insect cell expression of RRS1 and RPS4 domains.

SDS-PAGE soluble extract from Baculovirus transfected Sf9 insect cells following nickel IMAC. Soluble

protein expression is indicted by arrow. **(A-C)** Sf9 cells were transfected with 3μl of P1 Baculovirus,

(D-F) Sf9 cells were transfected with 30μl of P1 Baculovirus. Experiments were only conducted once.

Table 3.5 RRS1 and RPS4 soluble expression results from Sf9 insect cell trials. Construct number is in reference to Table 3.2.

Sf9 transfected Baculovirus expression results

Construct Number	Gel code	Gene	Domain	Amino acids	Vector	P1 virus titre	Expression Level	Comments
7	H7	RRS1	NB-ARC-LRR	I154-T595	pOPINS3C	3uL	Low	
15	H1	RRS1	Dom4	P868-K1189	pOPINF	30uL	Low	~28KDa truncation, mass spec predicts P868-K1101
16	A2	RRS1	Dom4	P868-K1189	pOPINS3C	3uL	Low	Mass spectrometry full peptide coverage
26	G2	RRS1	WRKY-D6R	E1209-Y1373	pOPINM	3 & 30uL	Medium	
40	E4	RPS4	CTD	L954-L1120	pOPINF	3uL	Medium	Near full mass spec coverage by mass spec
42	A5	AtWRKY 18	WRKY	A157-E240	pOPINF	3uL	High	Mass spectrometry full peptide coverage
43	B5	AtWRKY 40	WRKY	V124-N213	pOPINF	3uL	High	Mass spectrometry full peptide coverage

I also tried co-expressing RRS1 LRR-Dom4 pOPINF with RPS4 LRR-CTD pOPINF and RRS1 LRR-Dom6R pOPINF with RPS4 LRR-CTD pOPINF under the hypothesis that co-expression may help to stabilise these two proteins which are predicted to interact. However soluble expression of either RRS1 or RPS4 domain in insect cells was not observed, **Figure 3.8C,F**.

3.5   Discussion

The aim of this project was to trial the expression of full-length and truncated RRS1 and RPS4 constructs in the heterologous expression systems of *E. coli* and insect cells to produce soluble protein for downstream experiments. Previous work in the NLR field to produce soluble NLR protein has faced many difficulties with obtaining good yields of soluble multi-domain NLR protein. The lack of published structural data of multi-domain NLR constructs does likely not reflect a lack of effort to attain such data in the field and as such I wanted to trial a range of experimental conditions with these systems to increase the likelihood of achieving successful expression. We chose to focus our initial round of NLR expression trials on the heterologous systems of *E. coli* and Sf9 insect cells. A summary figure of the results of this screen can be found in **Figure 3.9** and list of all soluble expression results in study found in **Appendix 5.**

The majority of published crystal structures of domains of plant NLR proteins have utilised *E. coli* for protein expression, see **Table 3.1**. *E. coli*'s rapid growth in simple culturing conditions requiring little specialist equipment make the system a straightforward and economic system for a high-throughput expression screen such as this project looked to conduct. The system however is not without significant drawbacks outlined in detail in 3.1 including protein folding capabilities, lack of post translational modifications and codon *E. coli* usage bias.

Figure 3.3.9 Summary of RRS1 and RPS4 expression trials in *E. coli* and Sf9 insect cells. Each horizontal bar indicates the summarised results for all pOPIN vectors expressing that protein construct in both *E. coli* and insect cell. Bar colours indicate expression result with blue bars indicating low levels of soluble protein expression, green bars medium-high levels of soluble expression and grey bars no soluble protein expression as determined by SDS-PAGE analysis. A dashed line indicated a truncation of the cloned protein predicted molecular weight was observed. If expression was observed, the heterologous species in which expression was seen is listed on the left of the bar with 'E.c' indicating *E. coli* and 'Sf9' indicating Sf9 insect cells. Total number of conditions each protein construct was expressed in is listed on the right of each bar for *E. coli* and Sf9 insect cells.

Investigating heterologous expression systems for plant NLR expression

Given the failure of most RRS1 and RPS4 constructs tested in this project to show soluble expression in *E. coli,* it is likely these issues hinder the soluble expression of these plant NLRs. What was observed was a clear trend of higher soluble expression rates of Rosetta (DE3) pLys over Lemo21(DE3). The main difference in terms of expression capabilities of these two BL21 derivative lines is that Rosetta (DE3) pLys cells are supplied with extra copies of rare tRNAs uncommonly found in *E. coli*. As these rare codons are found in the cDNA sequence of RRS1 and RPS4 used in this study, see **Table 3.4**, and cDNA sequences were not codon optimised for *E. coli*, it could be hypothesised that supply of these rare tRNAs allowed for more efficient translation of RRS1 and RPS4 protein in Rosetta (DE3) pLys than Lemo21 (DE3) cells[184]. This potential bottleneck in NLR protein expression could be removed in future studies by codon optimising the cDNA sequence of the POI for *E. coli* to ensure for efficient protein translation. As SDS-PAGE protein bands of the total fraction from each trial were not clear enough to confirm the presence of the POI, we are unable to say what effect inefficient protein translation had on the screen's success and how much of a bottleneck translation efficiency was on soluble protein production.

Another potential avenue for improving the folding of NLRs in *E. coli* is through co-expression of NLRs with eukaryotic chaperone proteins. The dependency of certain plant NLRs on co-evolved chaperones for appropriate protein folding has been documented in the literature. For example, the dependency of tomato NLR I-2 on the small heat shock chaperone RSI2[199] and RPS4's dependency on eukaryotic conserved chaperones complex of SGT1-Hsp90[88,200] has been demonstrated. Going forward, this knowledge of NLR required chaperones could be expanded by conducting ribosome profiling studies[201] of RRS1 and RPS4 to provide a list of chaperone proteins utilised in the folding of these proteins in their native *A. thaliana* environment. A line of *E. coli* could then be generated which heterologously expressed these chaperones in which RRS1 and RPS4 protein could be additionally expressed. This would create a line of *E. coli* with a chaperone repertoire more akin to Arabidopsis which may aid soluble expression of NLRs in this heterologous expression system. A similar pipeline was recently applied for the expression of the Arabidopsis ~540 kDa RuBisCo complex[202]. In this study researchers achieved for the first-time soluble expression of the functional RuBisCo complex in *E. coli* by co-expressing the RuBisCo subunits with five additional chloroplast chaperones thereby enabling correct protein folding in the prokaryotic system. Generation of an 'NLR chaperone' *E. coli* expression line which co-expresses a range of chaperones known to be involved in NLR

folding but are not present in prokaryotic systems, could therefore provide a viable avenue for increasing soluble expression of NLRs in this system.

A variable I tested as part of the *E. coli* screen at OPPF was the cell growth media. Two different medias were tested, AIM and PB, which both employ different POI induction methods previously discussed in 3.4.1. The screen results show that AIM appeared to better support soluble protein expression of RRS1 and RPS4. This may be due to the fact cultures grown in AIM often grow to a higher cell density than IPTG-induced media's leading to higher yields of POI. It should be noted that is not the case for NLR expression universally as for example the RRS1 WRKY constructs used in Chapter 5 were found to express best in LB media induced with a low level of IPTG. This result does emphasise the need however to screen for multiple media conditions as in this project media was found to play an important role in soluble protein production levels. If this screen was to be expanded a range of IPTG induction concentrations used for PB grown cells could be tested as can have significant effects POI yield[203] potentially due to the metabolic burden on *E. coli* cell of the POI[204].

Other variables which could be tested could include the temperature at which cells are grown, especially during the overnight protein induction phase of growth. For example, lowering the cultivation temperature often increases levels of correct protein folding and helps prevent the formation of aggregated inclusion bodies[205,206]. Additionally further solubility tags could be tested for their ability to enhance the POI solubility[187,191]. For example, possible fusion tags to explore include: Glutathione S-transferase (GST) which protects against intracellular proteolysis and aids stability[207], thioredoxin which helps refold proteins in reducing environments[208] or N-utilization substance (NusA) which works to slow protein translation allowing a long time frame for protein folding[209]. Other strains of *E. coli* could also be tested as protein folding environments can greatly differ between strains. For example, to facilitate proper disulphide bond formation, strains containing mutated in glutathione reducatase (gor) and thioredoxin reductase (trxB), the enzymes responsible for reducing thioredoxins, could be used such as Origami and SHuffle[210].

The second heterologous system I explored for RRS1 and RPS4 protein production was Baculovirus transfected Sf9 insect cells. Though insect cells are significantly more time consuming and costly to grow and culture than *E. coli*, as a eukaryotic system they provide a more complex protein folding environment that we hypothesised would better support

for soluble expression of NLRs. However, similarly to *E. coli*, expression of RRS1 and RPS4 in Sf9 insect cells showed very limited success particularly at the multidomain level. With the exception of RRS1 WRKY-D6R$_{E1209-Y1373}$ pOPINM and RPS4 CTD$_{L954-L1120}$ pOPINF all soluble expression observed was of a yield too low to take forward for subsequent scale up experiments. Work to scale up production of RRS1 WRKY-D6R$_{E1209-Y1373}$ and RPS4 CTD$_{L954-L1120}$ is currently being conducted in the lab by Dr Nitika Mukhi and Dr Richard Hughes. Should scale up of these constructs prove successful, structural and biochemical interaction information of both these constructs could provide interesting biological insights into the functioning of RRS1 and RPS4.

Production of RRS1 WRKY-D6R$_{E1209-Y1373}$ would enable the exploration of the structural basis of RRS1-R's ability to bind structurally and enzymatically distinct effectors by comparing binding of AvrRps4 and PopP2 to this domain and provide intriguing insights into NLR effector recognition strategies. Genetic and biochemical evidence suggests that the post-LRR CTD region of RPS4 plays an important role in the functioning of RRS1-RPS4 complex[152]. RPS4 CTD has been observed in co-immunoprecipitation studies to interact with RRS1 Dom4-D6R but the activation of RRS1-RPS4 complex cannot be explained by the presence or absence of this interaction alone. This implies there are more subtle changes in domain interactions occurring during complex activation which only structural biology can provide insights on. As our ability to model RPS4's CTD is very low due to low homology with other protein domains, gaining structural insights of this domain would provide vital biological insights into the functioning of this domain. As a truncation of RPS4's CTD the RPS4 CTD$_{L954-L1120}$ pOPINF construct covers the central ~50% of this domain. Whilst this does include residues previously identified to be important in RRS1-RPS4 immune signalling such as RPS4 G997 it does not include other RPS4 CTD residues also known to impact the complex's signalling abilities such as S914 and G952[161]. Future work could therefore investigate identifying longer constructs of RPS4's CTD with soluble expression which covers all these key residues.

Despite having eukaryotic protein folding capabilities, Sf9 insect cells did not show a greatly increased range of soluble RRS1 and RPS4 protein expression. It could therefore by hypothesised that lack of appropriate chaperones may also be limiting the production of soluble RRS1 and RPS4 protein in this system similar to *E. coli*. Future studies could employ a similar strategy to that described above for *E. coli* in co-expressing NLRs with chaperones identified in the literature or ribosome profiling to be required for NLR functioning *in*

planta. Similar to *E. coli*, the SDS-PAGE bands of POI from insect cell trials total fraction were unclear to see we do not know the extent protein translation efficiency had on soluble protein expression and future studies could benefit from codon optimising NLR cDNA for insect cell expression.

Looking across both *E. coli* and insect cell systems there are several options to consider should future studies wish to further explore the use of these heterologous expression systems for RRS1 and RPS4 or alike NLR protein production which could help improve the protein folding environment of these hosts. In addition to previously discussed use of co-expression with chaperones future studies could explore the use of chemical chaperones to aid soluble protein production. Such chemical chaperones include: ethanol which is thought to induce expression of heat shock proteins[211], glycerol which helps with protein stability[212,213] or DMSO which is hypothesised to protect proteins from thermal denaturation and aggregation[213,214]. Beyond changing expression conditions, future studies of this kind could employ new screening technologies which expand on traditional PCR cloning methods such as ESPRIT screening technology developed at EMBL by Dr Darren Hart. This method involves the creation of unidirectional truncations of a target gene using exonuclease degradation to create a diverse library of potential expression constructs. Constructs are then "printed" onto nitrocellulose membranes and soluble POI expression screened for by hybridisation of fluorescent antibodies[215,216]. This enables the screening of up to 30,000 individual clones for yield and soluble POI expression in a single experiment and has already been utilised for similar studies with the *Symphytum tuberosum* NLR R3a[217] conducted in the Banfield laboratory.

What the results in the chapter highlight is that both *E. coli* and Sf9 insect cells appear to be an inappropriate expression system for the production of multi-domain soluble RRS1 and RPS4 protein, likely due to insufficient protein folding environment. The expression trials described in this chapter and chapter 5 suggested that whilst singular domains or truncations of RRS1's WRKY and Dom6 could be expressed in *E. coli* and Sf9 insect cells, multi domain expression was not supported. Work taking forward RRS1 WRKY Dom6 expression is described in chapter 5. The expression screen results in this chapter suggested that going forward with trials to produce soluble full-length RRS1 and RPS4 protein, I would need to utilise systems which provided the protein folding machinery and conditions more akin to that of the RRS1/RPS4's native *A. thaliana* environment such that would support the correct folding of these NLR proteins.

Investigating plant-based expression systems for plant NLR purification

4

Notes: Figures and results presented in section 4.3 of this chapter include my contribution to Ngou et al, 2019[218]. Contributions of other authors from this paper are acknowledged throughout. Work presented in 4.4 was conducted with Freya Hartshorn, an undergraduate summer student under my supervision. A list of the cloned vectors and maps used in this chapter can be found in **Appendix 2** and **Appendix 4.** Summary of all soluble expression results from study can be found in **Appendix 5.**

4.1   Introduction & chapter aims

High-throughput screens of RRS1 and RPS4 in *E. coli* and Sf9 insect cells described in chapter 3 highlighted that these heterologous systems are inappropriate for the production of soluble multi-domain RRS1 and RPS4 protein. I hypothesised that the inability of these systems to produce multi-domain soluble protein of RRS1 and RPS4 was likely due to failure to support protein folding of these NLRs. Consequently, I next wanted to test expression systems with protein folding machinery and conditions more akin to RRS1 and RPS4's native *A. thaliana*. For this I decided to expand trials to plant-based protein expression systems with the rationale that these systems would provide a protein folding environment better suited for soluble production of RRS1 and RPS4.

Plant-based expression systems are increasingly being utilised as a general platform for production of recombinant proteins, particularly in the pharmaceutical field[219,220]. A variety of different plant-based expression systems have been developed utilising a wide range of production and transformation strategies including: stable transgenic plants[221,222], agrobacterium mediated transient expression[223,224], virus infected plants[225], cell cultures (e.g. tobacco BY2 cells)[226] and wheat germ cell-free expression[170]. These systems are generally valued for their eukaryotic folding capabilities, economic scalability and production timescales all of which are highly relevant to this project.

Recombinant protein production relies on three expression strategies: chemical synthesis, in vivo expression and cell-free protein synthesis. As previously discussed, NLR protein

production is recognised to be notoriously difficult, so to increase any chances of success I set out to explore a diverse range of protein synthesis systems for production of RRS1 and RPS4 protein. Whilst chemical synthesis is not suitable for proteins of the length of RRS1 and RPS4, I wanted to investigate the use of cell-free expression systems in addition to the various in vivo strategies. The three best established cell-free systems used for protein production are rabbit reticulocyte, *E. coli* cell and wheat germ extract[227]. Of these systems wheat germ extract seemed the most appropriate for this study given the limited success of in vivo *E. coli* trials described in chapter 3 and observations in the literature that protein yields in wheat germ tend to be higher than rabbit reticulocyte extracts[228]. Due to its role in wheat kernel germination, wheat germ contains all the components required for wheat protein translation, with the exception of mRNA, and as such we hypothesised that this eukaryotic plant system would be a suitable environment for soluble expression of NLRs. Whilst examples are limited, use of wheat germ extract for NLR expression has been demonstrated, for example in the expression of *Oryza sativa* coiled-coil NLR Pb1[107], though protein production was not to levels required for structural or quantitative biochemical work.

For whole-plant based expression systems, I tested both transgenic *A. thaliana* plants and agroinfiltrated *N. benthamiana*. Though initially time-consuming and labour-intensive, once generated, stable transgenic plants provide an economical and relatively straightforward scalable production system, especially for the protein quantities I set out to produce RRS1 and RPS4. The physical ease of scaling transgenic *A. thaliana* for this project was enhanced by our ability to grow sterile *A. thaliana* seedlings shaking in liquid media [229]. Alternatively, transgenic lines can be grown on soil for several weeks before leaf tissue is harvested. Whilst this would be a productive strategy if a protein of interest (POI) was restricted to leaf tissue expression, the promoters used for this work (35S, pAt2 and pAt3) are expressed across *A. thaliana* cell types. As RRS1 and RPS4 are native to *A. thaliana*, transgenic *A. thaliana* lines have the complete protein folding capabilities for producing functional correctly folded and post translationally modified RRS1 and RPS4 protein.

The final expression strategy I set out to evaluate is agroinfiltrated *N. benthamiana.* Whilst increasingly used for the purposes of protein production in the pharmaceutical field[220], use of this system for production of NLRs for structural biology and quantitative biochemical work has only recently been investigated. Prior to this development, the system was widely used for small-scale (single leaf) NLR production for the purpose of co-immunoprecipitation

studies, microscopy and HR assays including RRS1 and RPS4 studies[65,114]. The benefits the system offers over transgenic plants is the speed with which proteins can be expressed and harvested, enabling faster evaluation of the performance of different expression constructs. This method utilises *A. tumefaciens* to deliver genes to leaf cells of *N. benthamiana* which can then be harvested for protein purification ~3 dpi. Protein yields from this system vary but yields of 1 gram of protein product per kilogram of leaves have been recorded[220]. Previous work with small-scale (single leaf) experiments have shown this system is capable of producing full-length RRS1 and RPS4 protein so it was well established that this system provides a suitable folding environment for these NLR proteins. Compared to transgenic plants, agroinfiltration offers more flexibility in trialling a range of expression conditions for protein production. As constructs can be transformed into agrobacterium and protein harvested from leaves days later, this system can be used to investigate a range of variables in protein production including trialling different expression vectors or affinity tags for the purification of POI.

Following on from *E. coli* and Sf9 insect cell expression screens in the previous chapter, this chapter will investigate the use of a range of plant-based expression systems for the production of RRS1 and RPS4 protein. Strategies evaluated for protein production are: cell-free wheat germ system, transgenic *A. thaliana* lines expressing full-length and truncated RRS1 and RPS4 and agrobacterium-mediated transformation of *N. benthamiana* to express full-length and truncated RRS1 and RPS4 protein. This chapter will discuss the results of these screens of which full methodologies can be found in chapter 2.

4.2 Expression of full-length and truncated RRS1 in a cell-free wheat germ system

4.2.1 Full-length RRS1^{Ws-2} and RRS1 WRKY Dom6S show soluble expression in a wheat germ cell-free system-CellFree Sciences WEPRO® extract

Full-length coding sequence of RRS1^{Ws-2} and RRS1 WRKY Dom6S with C-terminal HF tag were cloned in to pEU wheat germ expression vector (CellFree Sciences) using golden gate cloning. This pEU vector was made golden gate compatible by Dr Cheng Chang using PCR mutagenesis. I chose to trial expression of full-length RRS1^{Ws-2} to evaluate the wheat germ cell-free expression system for its suitability to express full-length NLRs which had previously proved insoluble in insect cell and *E. coli* expression systems. Additionally, I

chose to express RRS1 WRKY D6S alongside full-length RRS1 as a form of positive control given we knew this construct expressed well in *E. coli* and therefore was predicted to express in wheat germ's folding environment. I initially trialled a wheat germ expression system using a CellFree Sciences ENDEXT® Technology Protein Research Kit S (https://www.cfsciences.com/eg/products/kit/138-kit/products-for-research-premixed-kit/465-protein-reseach-kit-s) using the workflow described in **Figure 4.1.**

Cloned RRS1[Ws-2]-HF pEU vectors were transformed into *E. coli* DH10B cells and plasmid midiprepped from a 25 ml culture. High purity mRNA was then generated for each construct (from 2000 ng of purified pEU plasmid) using a transcription mix supplied by CellFree Sciences. Following transcription, mRNA integrity was assessed by gel electrophoresis.

There are several methods for setting up a wheat germ translation system which vary in their potential protein yields and simplicity of setup[170]. The traditional method is a batch reaction format where all reagents are mixed and incubated together for a limited time period. This is the simplest method to setup and typically has the lowest protein yields. A variation on this is the repeat-batch method in which the reaction is set up in the batch format, but in a reaction vial with a membrane which allows through small molecular weight translation by-product inhibitors from the reaction mix. These by-products can then be periodically removed from the reaction mix by centrifugation and the translation reaction topped up with fresh substrate buffer. Proteins can also be expressed in a dialysis protein expression format where the translation reaction is setup in a dialysis cup which is in contact with a feeding substrate buffer that supplies the reaction with fresh substrate whilst removing small molecular weight inhibitors like the repeat batch format. These two methods have been shown to offer higher yields of POI for a given volume of wheat germ extract but with a more complex setup which works better with a larger volume of wheat germ extract[170]. Alternatively, the bilayer reaction method is setup with the substrate buffer (amino acids, ATP, GTP etc) added on top of a translation mix which contains a mixture of mRNA and wheat germ extract to form two separate layers[169]. This reaction setup allows for a diffusion-controlled translation process which can be run for longer than a traditional mixed batch reaction format which will end after a few hours. Conversely, bilayer reactions can by maintained for up to 24 hours yielding substantially higher protein yields[170].
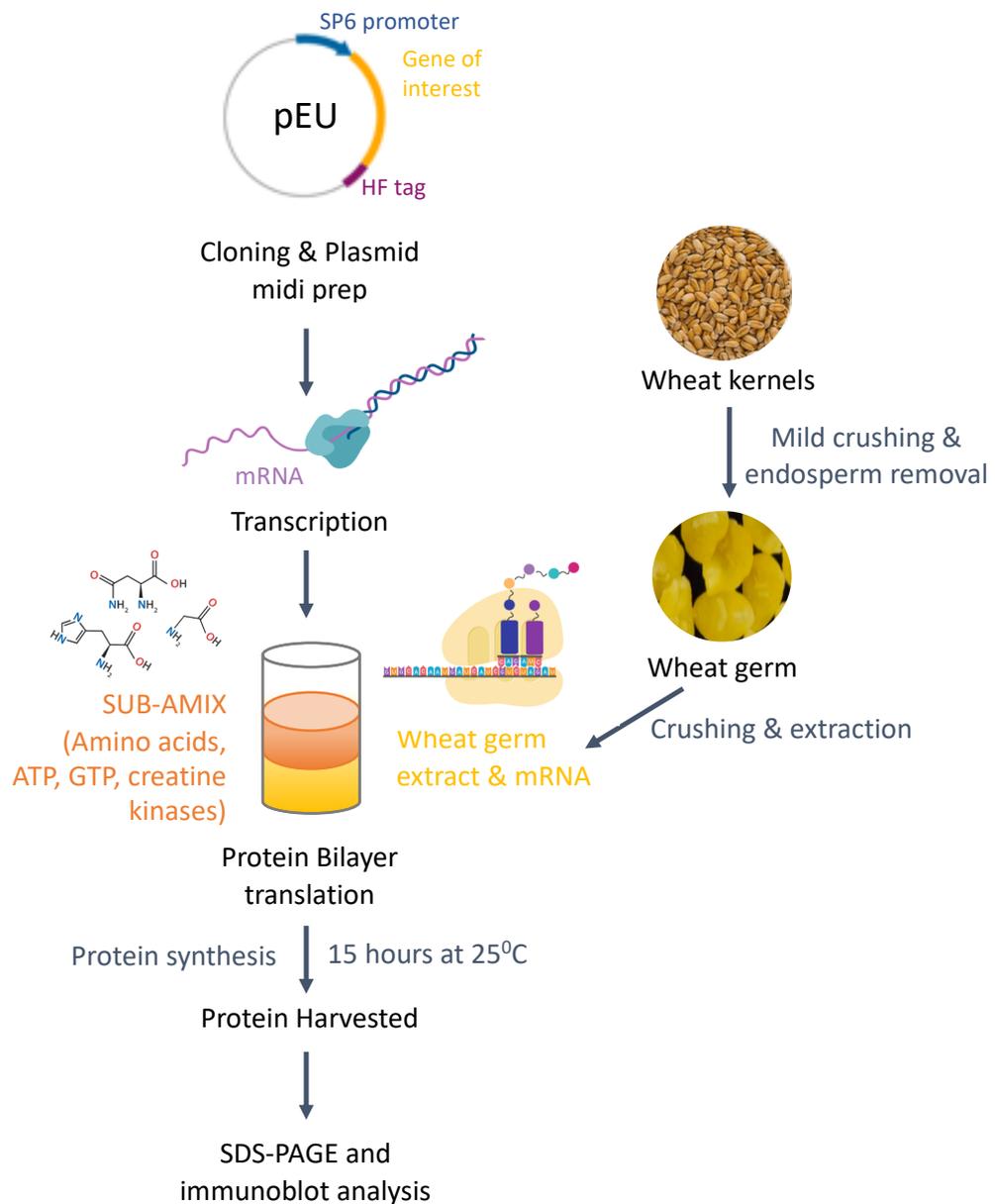
Figure 4.1 Wheat germ cell expression system workflow. Following cloning of gene of interest into a pEU expression vector with a Hell-Fire (HF) tag, plasmid was midiprepped from *E. coli*. mRNA of gene of interest was then generated from midiprepped DNA template in a transcription reaction. mRNA was then mixed with wheat germ extract and setup in a bilayer translation reaction with a top layer of sub-AMIX mixture which contained amino acids, ATP and GTP. Protein translation reaction continued at 25^{0}C for 15 hours before the two translation layers were mixed and POI purified using anti-FLAG affinity beads followed by SDS-PAGE and immunoblot analysis. Figure created using BioRender.
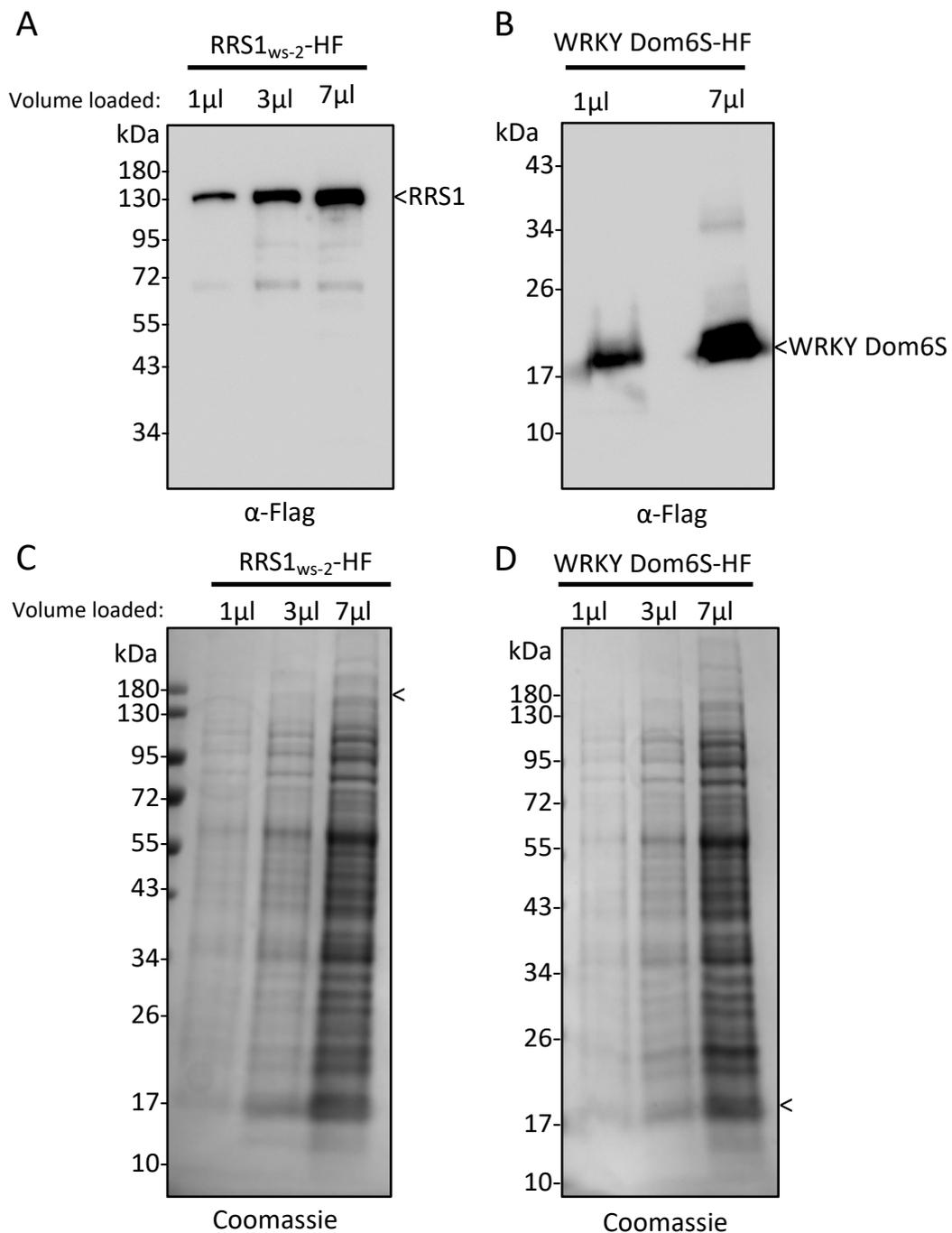
Figure 4.2 Expression of full-length RRS1^{Ws-2}-HF and truncated WRKY Dom6S-HF in CellFree Sciences wheat germ cell-free system. (A&B) Immunoblot analysis of RRS1^{Ws-2}-HF and RRS1 WRKY Dom6S-HF in total extract of CellFree Sciences Protein Research Kit (S) wheat germ cell-free extract, bands were visualized using Pico PLUS chemiluminescent substrate with 30 second exposure, arrows indicate the expected protein bands. Protein were visualised using an HRP-conjugated anti-FLAG antibody. **(C&D)** Coomassie stained SDS-PAGE of total protein of wheat germ cell-free extract arrows indicate the expected band size of POI. Volume loaded of total translation reaction mix is indicated above gel lane. Experiment conducted jointly with Dr Yan Ma. Experiments were repeated twice with similar results found across repeats.

For this investigation, protein translation was set up in a bilayer system as it offers better yields than batch reactions and can be easily set up with the small reaction volumes (226 µl) I was working with. Following translation, the substrate SUB-AMIX buffer and translation mix were mixed and protein yield assessed by SDS-PAGE and immunoblot analysis. Following a 15-hour translation reaction, immunoblot analysis showed that the wheat germ cell-free system was capable of expressing both full-length RRS1^{Ws-2}-HF and RRS1 WRKY Dom6S-HF, **Figure 4.2A,B**. However, SDS-PAGE analysis showed the yields of these protein were not high enough to visualise via Coomassie staining, **Figure 4.2C,D.**

4.2.2 Full-length RRS1^{Ws-2} and RRS1 WRKY Dom6S show soluble expression in a wheat germ cell-free system- Wheat germ extract from Professor Yasuomi Tada

In addition to the wheat germ WEPRO® extract supplied in the CellFree Sciences ENDEXT® Technology Protein Research Kit S, I also tried expressing full-length RRS1^{Ws-2}-HF and truncated RRS1 WRKY Dom6S-HF using an alternative wheat germ extract sourced from Professor Yasuomi Tada at Nagoya University to investigate how this would affect the yield of the POI. Translation reactions were setup in a bilayer format as mentioned before with the reaction run for 15 hours at 25^{0}C.

Similar to the CellFree Sciences system, **Figure 4.2,** whilst full-length expression of RRS1^{Ws-2} and RRS1 WRKY Dom6S was observed by immunoblot analysis using this system, **Figure 4.3A,B** , the proteins of interest were still of an insufficient quantity to be observed by Coomassie staining of an SDS-PAGE gel, **Figure 4.3C,D**. Inferring from the immunoblot conditions required to visualise the POI in **Figure 4.2A & 4.3A**, which were 30 seconds with pico chemiluminescent substrate and 120 seconds with 1:1 pico:femto chemiluminescent substrate respectively, it appeared full-length RRS1^{Ws-2}-HF was expressed to higher levels in the CellFree Sciences WEPRO® extract. Therefore, this source of wheat germ extract was taken forward for subsequent studies.
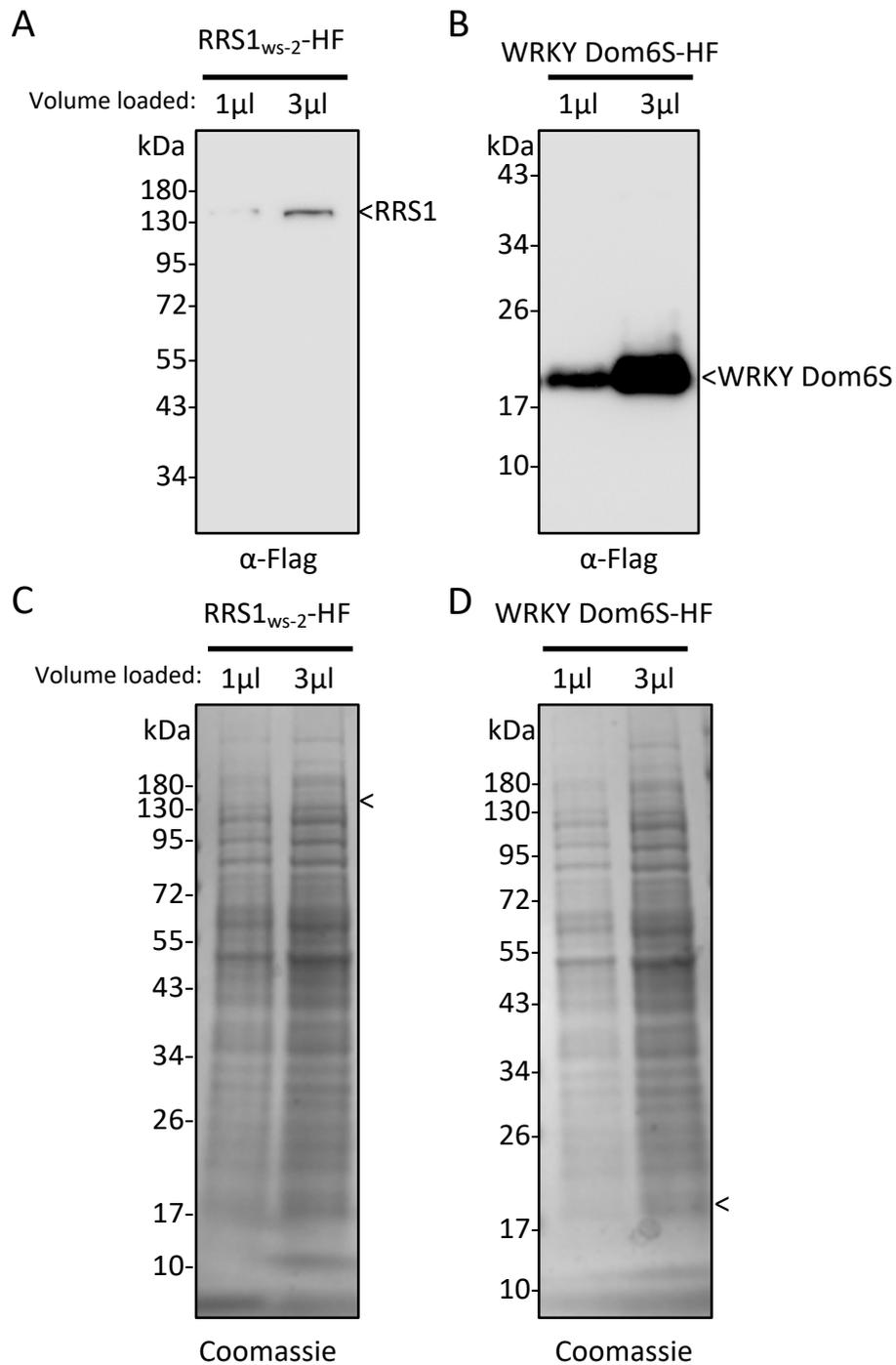
Figure 4.3 Expression of full-length RRS1[Ws-2]-HF and truncated WRKY Dom6S-HF in wheat germ extract. (A&B) Immunoblot analysis of RRS1 [Ws-2]-HF and RRS1 WRKY Dom6S-HF in wheat germ cell-free extract from Professor Yasuomi Tada's laboratory, bands were visualized using Pico PLUS and Femto chemiluminescent substrate in a 1:1 ratio with 120 second exposure, arrows indicate the expected protein bands. Protein were visualised using an HRP-conjugated anti-FLAG antibody. **(C&D)** Coomassie stained SDS-PAGE of total protein of wheat germ cell-free extract arrows indicate the expected band size of POI. Volume loaded of total translation reaction mix is indicated above gel lane. Experiment conducted jointly with Dr Yan Ma.

4.2.3 HF-tagged proteins can be purified from wheat germ cell-free expression system using anti-FLAG affinity beads

In order to establish the best method for purifying POI expressed in CellFree Sciences wheat germ WEPRO® extract, both full-length RRS1[Ws-2]-HF and RRS1 Dom4 WRKY Dom6R-HF was expressed in vitro from mRNA generated from pEU vectors. As each of these proteins were fused to an HF-tag (6xHis, 3xFLAG epitopes) at the C-terminus, I first tried purifying the POI by its 6xHis tag using the well utilised system of Ni-NTA resin[230] as used in previous *E. coli* and insect cell work in this study. However, this method failed to purify proteins from the wheat germ extract. Secondly, I tried to purify the HF-tagged proteins using anti-FLAG affinity beads. This purification method proved successful in immunoprecipitating the POI, **Figure 4. 4**. However, release of POI from the anti-FLAG beads by competitive elution with 3 x FLAG peptide was inefficient, as boiling of the anti-FLAG affinity beads post FLAG peptide elution showed significant amount of protein still remained bound to the beads.
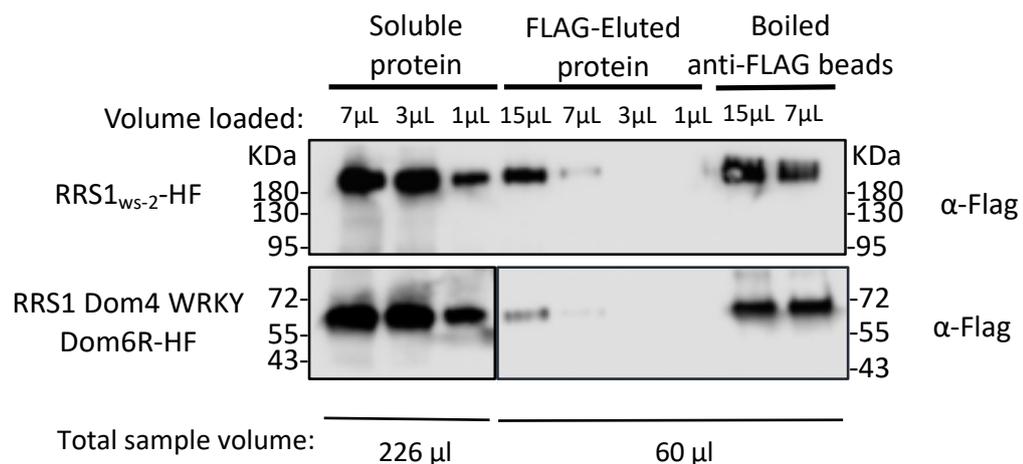
Figure 4.4 Anti-FLAG affinity bead purification of full-length RRS1[Ws-2]-HF and Dom4 WRKY Dom6R truncated protein from wheat germ extract. For immunoblotting, samples were taken throughout purification process from: total soluble input, competitive elution with FLAG peptide and boiled anti-FLAG affinity beads post FLAG peptide elution. Volume loaded for each of the samples is indicated above gel lane. Total sample volume indicates the total volume of sample at each stage of purification of which 20 µl was loaded on an SDS-PAGE gel. Protein were visualised using an HRP-conjugated anti-FLAG antibody.

4.2.4 RRS1[Ws-2]-HF appears to form a multimeric complex when expressed in wheat germ cell-free extract

There has been much discussion in the NLR field as to the ability of NLRs to form multimeric complexes[231,232]. To investigate whether full-length RRS1[Ws-2]-HF could form multimeric complexes when expressed in wheat germ cell-free extract, post translation samples were analysed using Blue Native PAGE. Blue Native PAGE showed a smearing of FLAG-tagged protein between 480-1048 kDa suggesting that the RRS1[Ws-2]-HF protein was not in a monomeric state (161 kDa), **Figure 4.5A.**

Figure 4.

Native PA

indicate

electropl

conjugat

Investigating plant-based expression systems for plant NLR purification

(B) Presence of RRS1-HF protein in Blue Native PAGE sample was shown by SDS-PAGE analysis of same protein sample followed by immunoblotting with an HRP-conjugated anti-FLAG antibody. Volume loaded of total sample is indicated above gel lane. Experiment was only conducted once.

Whilst we cannot infer the exact size of complex formed by RRS1^{Ws-2}-HF in wheat germ due to inaccuracies with the size prediction of the Blue Native PAGE method, this result gives us a promising lead that RRS1^{Ws-2} may form a multimeric complex in vivo. This result is also in accordance with previous findings in which RRS1 was shown to self-associate in the absence of RPS4[233]. Furthermore, the retained ability of the wheat germ extract-produced RRS1 to self-associate is consistent with the protein produced in this system being correctly folded. Correct folding can be inferred from this observation as incorrectly folded protein would likely aggregate and now elute from the column in a single peak. Further analysis such as circular dichroism could be used in the future to verify this.

In addition to establishing the size of complex this RRS1 material was forming in wheat germ, another interesting aspect to be investigated would be whether any non RRS1 proteins were contained in this complex. If so, this would highlight potential downstream interactors of RRS1. However, as the yield of POI in this system after immunoprecipitation with anti-FLAG affinity beads was of a quantity which cannot be visualised by Coomassie staining of SDS-PAGE gels, we are unable to say if this complex of RRS1^{Ws-2}-HF contains other proteins as well.

Ultimately, we would want to investigate the NLR complex of RRS1 and RPS4 together as this is how they work in their native environment and therefore presents a more valuable model to investigate. However, given our current inability to produce RPS4 cDNA I was unable to co-express both RRS1 and RPS4 in this system, to test whether we could recapitulate formation of an RRS1-RPS4 complex. Should RPS4 cDNA become available in the future this would be an interesting goal for future studies.

Though RRS1^{Ws-2}-HF protein appears to show soluble expression and correct folding in the wheat germ cell-free system, the inability to visualise expressed protein on a Coomassie stained SDS-PAGE indicates the yields of protein we are producing in this system is insufficient for follow up quantitative biochemical and structural work. Whilst theoretically protein yield could be increased by increasing the volume of wheat germ extract used for translation, given the quantities of POI yields observed in this study, the volume of wheat

germ extract which would be required to scale up expression was beyond the scope of this project.

4.3 Purifying RRS1 and RPS4 from transgenic over-expression lines of *Arabidopsis thaliana*

Successful expression of RRS1 in cell-free wheat germ extract but not insect cells indicated that a plant-based protein folding environment is better suited for soluble expression of RRS1. However, the major bottleneck with wheat germ cell-free system was our inability to economically scale-up the production of RRS1 protein to quantities sufficient for biochemical and structural studies. I therefore needed a system which could provide a plant-based folding environment for the production of soluble RRS1 and RPS4, and could also be scaled with suitable physical and economical ease to provide a viable source of RRS1 and RPS4 protein for further study. Our inability to produce RPS4 cDNA also meant that switching to a system which utilised gDNA would enable us to investigate purifying both these NLRs together, preferably in complex. At this time, Dr Pingtao Ding had generated a transgenic line of *A. thaliana* by crossing two lines each carrying 35S::RPS4-HS (HA and Strep affinity tags) and 35S::RRS1^{Ws-2}-HF both in a Col-0 background[233], **Figure 4.6A**. Dr Pingtao Ding had shown that this line could produce biologically functional RRS1^{Ws-2}-HF and RPS4-HS protein as demonstrated by the line's extended ability to perceive PopP2 effector delivered by Pf0-1, **Figure 4.6B.**

This transgenic *A. thaliana* line therefore presented itself as a valuable tool for production of soluble full-length RRS1 and RPS4, as protein purified from this system would be folded in the protein's native Arabidopsis environment, albeit under higher expression levels. Furthermore, transgenic *A. thaliana* plants can easily and scalably be grown by shaking sterilized seed in liquid 1% MS media cultures. For this, sterilised seed grown in liquid media and cultivated under long day conditions for ~2 weeks prior to harvesting. Tissue was then snap frozen in liquid nitrogen and stored at -80^{0}C for later use or used fresh as described in **Figure 4.7.**

A

35S::RRS1^{Ws-2}-HF   *35S::RPS4-HS / 35S::RRS1^{Ws-2}-HF*

B

Pf0-1:

AvrRps4_{WT}   AvrRps4_{KRVY-AAAA}   PopP2_{WT}   Mock

35S::RPS4-HS / 35S::RRS1^{Ws-2}-HF

C   RRS1-HF

30μl   20μl   10μl   5μl

kDa
180-
130-   < RRS1^{Ws-2}-HF
95-   α FLAG

PS

D   RPS4-HS

30μl   20μl   10μl   5μl

kDa
180-   < RPS4-HS
130-
                                        α HA
95-

PS
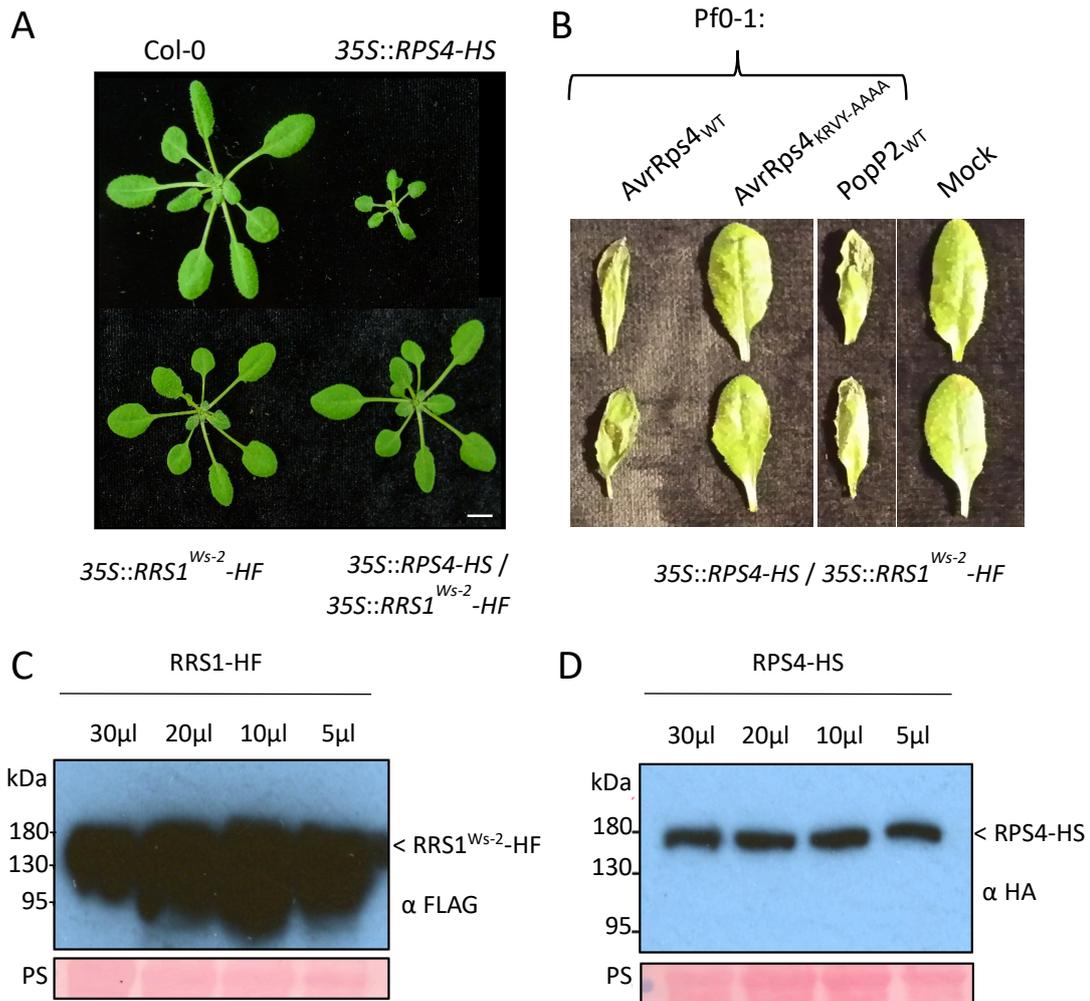
Figure 4.6 Overexpression of RRS1^{Ws-2}-HF rescues 35S::RPS4-HS phenotype and confers recognition of PopP2 to Col-0 plants (A) Stunting and dwarf phenotype of Arabidopsis transgenic line stably overexpressing RPS4 is attenuated by crossing with RRS1^{Ws-2} transgenic Arabidopsis line to generate a 35S::RPS4-HS / 35S::RRS1^{Ws-2}-HF transgenic line. Images were taken with 4-week-old short-day conditions at 22°C, Scale bar = 1.0 cm, figure adapted from Huh et al, 2017. **(B)** Over-expression of RPS4 and RRS1^{Ws-2} confers recognition of PopP2 in Col-0. Col-0 is known as the accession carrying RRS1-S that can only recognize WT AvrRps4 and activate hypersensitive cell death response (HR), but not able to recognize PopP2 unless RRS1^{Ws-2} (RRS1-R) is artificially supplied. All plant leaves were infiltrated with *Pseudomonas fluorescence* (Pf) 0-1 strains carrying wither wild-type AvrRps4_{WT,} unrecognized mutant AvrRps4_{KRVY135-138AAAA} or wild typePopP2. figure generated by Dr Pingtao Ding. **(C)** Immunoblot analysis shows expression of RRS1^{Ws-2} using an HRP-conjugated anti-FLAG antibody or **(D)** RPS4-HA using an HRP-conjugated anti-HA antibody in transgenic 35S::RPS4-HS / 35S::RRS1^{Ws-2}-HF lines. Total extract of A. *thaliana* tissue was loaded onto 4-20% gradient SDS-PAGE gel and Ponceau S Staining (PS) of Rubisco large subunit was used as a loading control. Volume of total extract loaded is indicated above gel lane with the experiment repeated twice with similar results observed.
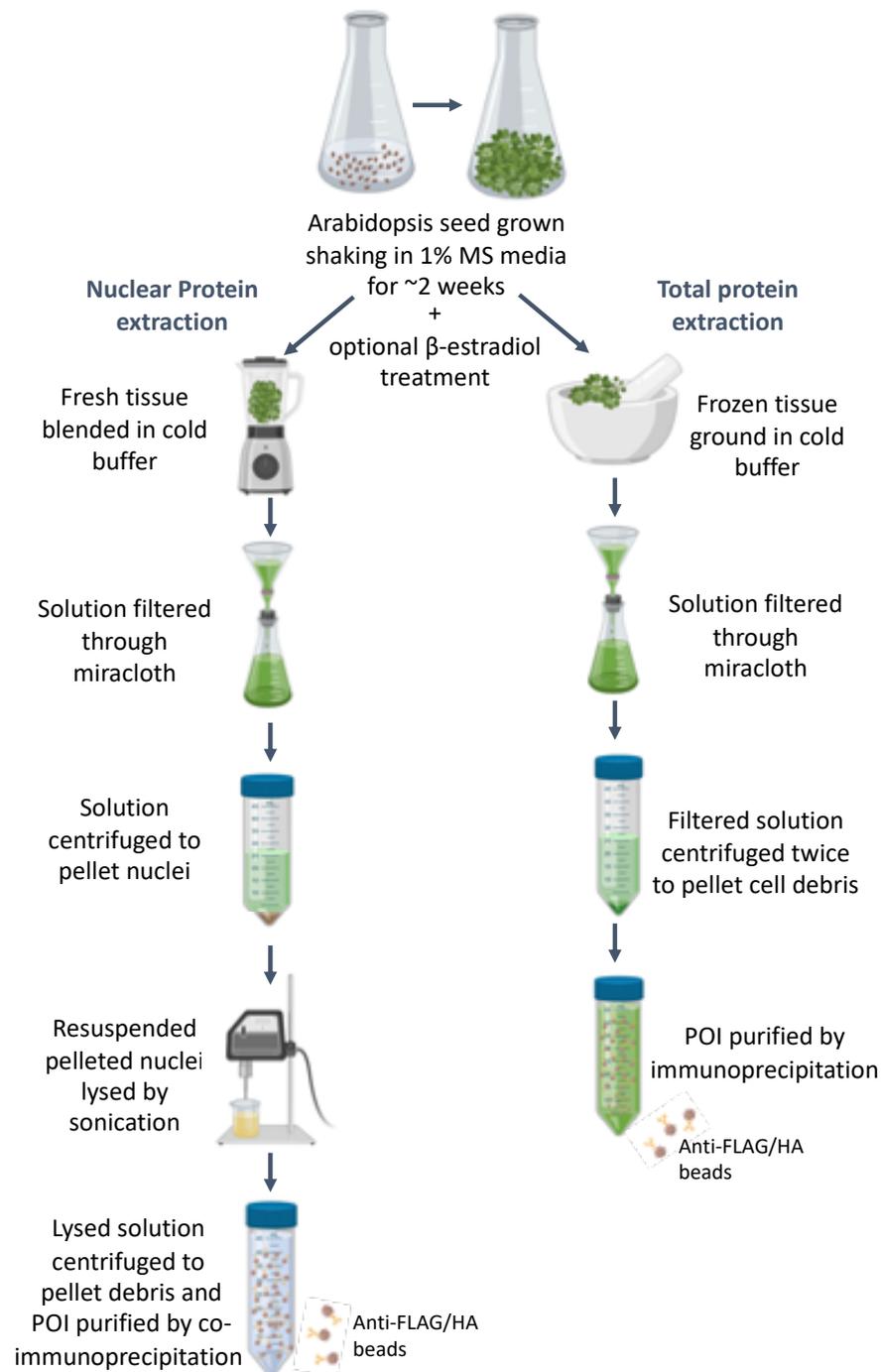
Figure 4.7 Purification of full-length RRS1 and RPS4 from transgenic Arabidopsis workflow.

Arabidopsis seed was sterilized and sown in 1% MS liquid media and grown shaking for ~2 weeks. Proteins were then purified by two different methods. In the nuclear extraction method fresh tissue was blended in nuclear extraction buffer, filtered and nuclei harvested by centrifugation, nuclei were then lysed by sonication and the lysate recentrifuged to pellet debris. In the total protein extraction method, tissue was harvested and frozen in liquid nitrogen, tissue was then ground in extraction buffer, filtered and centrifuged twice to pellet cell debris. Following both methods POI was then purified from lysate by immunoprecipitation incubation with anti-FLAG or anti-HA affinity beads. Figure created using BioRender.

4.3.1 Purifying pre-activation RRS1 and RPS4 from transgenic overexpression *A. thaliana*

Full-length inactive RRS1 and RPS4 was purified using the transgenic *A. thaliana* line expressing 35S::RPS4-HS/ 35S::RRS1 [Ws-2]-HF generated by Dr Pingtao Ding. Firstly, I confirmed the expression of RPS4-HS and RRS1 [Ws-2]-HF proteins in total extract of these plants grown in shaking liquid media culture by immunoblotting, **Figure 4.6C,D.** As RRS1 and RPS4 have been shown previously to localise into the nucleus of the plant cell[65,233], I hypothesised that purifying RPS4-HS and RRS1 [Ws-2]-HF proteins from a nuclear enriched sample would improve the purity and yield of these NLR proteins. To do this, I used a method adapted from Dr Pingtao Ding to purify the nuclei from ~2 week old 35S::RPS4-HS/ 35S::RRS1 [Ws-2]-HF material from which RPS4-HS and RRS1 [Ws-2]-HF protein was purified by co-immunoprecipitation following lysis of nuclei via sonication, see **Figure 4.7**. I trialled three different sonication treatments for nuclei lysis based on recommendations in the literature: 1:3 seconds on:off at 20% amplitude for a total of (i) 30 seconds on and (ii) 15 seconds on and (iii) 10:120 seconds on:off at 40% amplitude for a total of 50 seconds on. These trials suggested treatment iii (10:120 seconds on:off at 40% amplitude for a total of 50 seconds on) was most effective for rupturing nuclei of *A. thaliana* tissue and was therefore taken forward for subsequent studies, **Figure 4.10.**

To purify RRS1 [Ws-2]-HF and associated proteins, I used anti-FLAG affinity beads to pull down RRS1 [Ws-2]-HF from lysed nuclear extract. For immunoblot analysis, samples were taken throughout the purification process to track the progression of RRS1 [Ws-2]-HF and RPS4-HS protein throughout purification, **Figure 4.8.** Interestingly RPS4 protein appeared to be nearly exclusively present in the nuclear fraction of the *A. thaliana* extract whilst RRS1 protein appeared in both the nuclear and non-nuclear fraction. This suggests that RPS4 might predominantly reside in the nucleus of the *A. thaliana* cell whilst RRS1 may be enriched in the nucleus (as suggested by previous microscopy data in the Jones lab), but may also be present in the non-nuclear fractions of the cell as well. This observation was an important lead and could be used for future our work to purify RPS4 from plant material as purification strategies will need to ensure nuclei are adequately lysed during the purification process to release RPS4 protein through the use of sonication or detergents. It should be noted that this observation was not seen in all purifications of RRS1 [Ws-2]-HF and RPS4-HS protein from this line, but we hypothesise this may be due to variations between preps in effectiveness of pelleting of nuclei in the *A. thaliana* extracts.
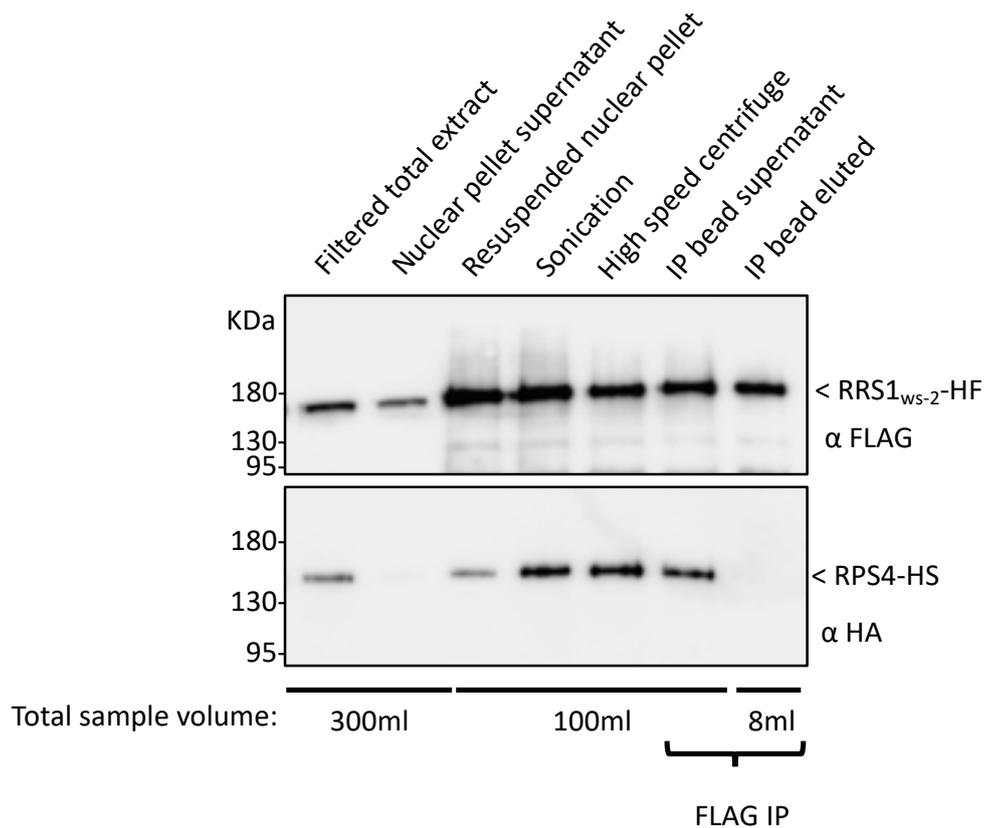
Figure 4.8 Anti-FLAG affinity bead purification of RRS1 and RPS4 from 35S::RPS4-HS/ 35S::RRS1 Ws-2-HF *A. thaliana* by nuclear extraction. For immunoblot analysis, samples were taken throughout the purification process. Proteins were eluted from anti-FLAG beads by competitive binding with 3xFLAG peptide. Samples were loaded on 4-20% gradient SDS-PAGE gel and proteins detected by immunoblot analysis with an HRP-conjugated anti-FLAG or anti-HA antibody. Total sample volume indicates the total volume of sample at each stage of purification of which 20 µl was loaded on SDS-PAGE gel. Experiments were repeated twice with similar results found across repeats.

Immunoblot analysis of samples taken throughout the RRS1 Ws-2-HF purification process revealed that major protein losses occurred during immunoprecipitation of the RRS1ws-2-HF protein. This is likely due to both the inefficiency in the release of RRS1 Ws-2-HF from the anti-FLAG affinity beads by competitive binding with 3xFLAG peptide, and the quantity of beads used to immunoprecipitate the RRS1 Ws-2-HF protein which come with cost limitations. Whilst previous work has observed RRS1 Ws-2 and RPS4 to form a complex pre-effector activation[233], attempts to coimmunoprecipitate RPS4-HS with RRS1 Ws-2-HF and RRS1 Ws-2-HF with RPS4-HS in this study failed, **Figure 4. 9 & 4.10**. The reason for this is not established but could reflect non-optimized conditions used to extract the NLRs, such as buffer composition. However, it should be noted that previous studies which observed this interaction utilised agrobacterium mediated transient expression of RRS1 Ws-2-HF and RPS4-

Investigating plant-based expression systems for plant NLR purification

HA in *N. benthamiana* leaves. We therefore cannot rule out the possibility this interaction may not occur, or not with enough strength, in the native Arabidopsis environment and in the absence of PTI activation as agrobacterium infiltration triggers a PTI response in the *N. benthamiana* tissue.

Additionally, I set out to purify RPS4 protein using anti-HA affinity beads using the same flow path as RRS1 [Ws-2]-HF immunoprecipitation purification with anti-FLAG affinity beads. Similar problems with release of protein from anti-FLAG affinity beads were seen too with release of RPS4-HS protein from the anti-HA affinity beads with competitive binding with HA peptide, **Figure 4.9**. Mirroring results in **Figure 4.8,** co-immunoprecipitation of RPS4-HS failed to pull down RRS1 [Ws-2]-HF which remains unbound in the supernatant of the anti-HA affinity beads. Similarly to RRS1 co-immunoprecipitation experiments, we cannot conclusively say if inability to detect an interaction between RRS1[Ws-2]-HF and RPS4-HS reflects biological state of these proteins, or is due to our experimental conditions e.g. buffer composition.
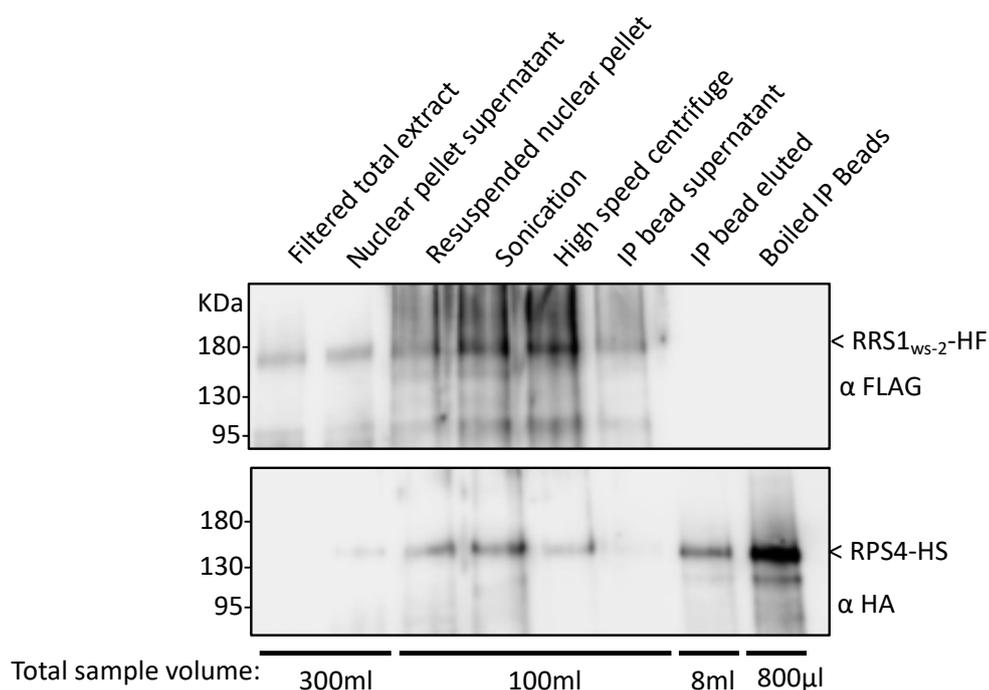
Figure 4.9 Anti-HA affinity bead purification of RRS1 and RPS4 from 35S::RPS4-HS/ 35S::RRS1[Ws-2]-HF *A. thaliana* by nuclear extraction. Immunoblot analysis samples taken throughout purification process. Proteins were eluted from anti-HA beads by competitive binding with HA peptide. Samples were loaded on a 4-20% gradient SDS-PAGE gel and proteins detected by immunoblot analysis with an HRP-conjugated anti-FLAG or anti-HA antibody. Total sample volume indicates the total volume of sample at each stage of purification of which 20 µl was loaded on the SDS-PAGE gel.

4.3.2 RRS1 [Ws-2]-HF appears to form a multimeric complex when overexpressed in *A. thaliana*

Having established I could purify quantities of RRS1 [Ws-2]-HF suitable for initial biochemical characterisation, I first wanted to determine the size of the RRS1 [Ws-2]-HF complex. Previous published work[233] as well as my own observations using a wheat germ cell-free expression system suggested RRS1 [Ws-2]-HF may form a multimeric complex. I therefore set out to measure the size of RRS1 [Ws-2]-HF purified from transgenic 35S::RPS4-HS/ 35S::RRS1 [Ws-2]-HF *A. thaliana* using analytical gel filtration chromatography. Analytical gel exclusion chromatography separates proteins by size. Sample protein is run through a porous resin packed column containing pores of various sizes. Larger protein/complexes cannot enter smaller pores and elute from the column earlier than smaller proteins which are retained in the porous resin, slowing their passage though the column. The elution volume of proteins is detected through absorption at 280 nm. Analytical gel filtration can thus be used to estimate the size of a protein complex depending on its retention volume. However, whilst estimates can be drawn, an exact size of complex cannot always be concluded from data of this kind, especially at high molecular weights. For example, proteins of the same molecular weight can have different retention volumes depending on the shape the protein/complex resides in, e.g. globular proteins will have a lower retention volume than elongated proteins.

RRS1 [Ws-2]-HF was purified from 100 g of fresh 35S::RPS4-HS/ 35S::RRS1 [Ws-2]-HF *A. thaliana* seedlings and concentrated to ~100 µl in volume. This sample was loaded onto a Superose™ 6 Increase column (GE Healthcare) pre-equilibrated in running buffer (150 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA). RRS1 [Ws-2]-HF protein eluted at 14.39 ml, **Figure 4.10A,** with presence of RRS1 [Ws-2]-HF in the fractions of this peak confirmed by immunoblot analysis, **Figure 4.10**. A monomer of RRS1 [Ws-2]-HF (161 kDa) would be predicted to run at 16.86 ml (calibration curve generated by Dr Abbas Maqbool) whilst an elution volume of 14.3 ml equates to a ~640 kDa complex using the calibration curve. Whilst we cannot accurately state that RRS1[Ws-2]-HF is running in a complex of this size, this earlier elution volume did suggest that the protein is forming a multimeric complex. Samples from the RRS1[Ws-2]-HF analytical gel filtration peak were also tested for the presence of RPS4-HS

protein, however, no RPS4-HS protein was detected. It should be noted however that no positive control for HA antibody activity was included in this experiment.

Samples of RRS1 $^{Ws-2}$-HF purified from fresh 35S::RPS4-HS/ 35S::RRS1 $^{Ws-2}$-HF were also analysed for multimeric complex formation by Blue Native-PAGE. Samples were prepared from a singular nuclear extraction preparation of fresh 35S::RPS4-HS/ 35S::RRS1 $^{Ws-2}$-HF tissue. I then split this extract into three samples each of which was lysed by a different sonication treatment as per 4.3.1 before being centrifuged to remove cell debris. This enabled me to evaluate if varying sonication treatment affected the size of the RRS1 $^{Ws-2}$-HF protein complex. I found that apart from effectiveness in ability to lyse nuclei, the various sonication treatments did not affect the apparent complex size of RRS1 $^{Ws-2}$-HF. Immunoblot analysis of Blue Native-PAGE gel showed that RRS1 $^{Ws-2}$-HF was running in a multimeric complex of a similar size to that observed in wheat germ cell-free extract, **Figure 4.5**, and via analytical gel filtration, **Figure 4.10B.** RPS4-HS protein was not able to be purified in sufficient quantities to conduct a similar analysis with.

A

B
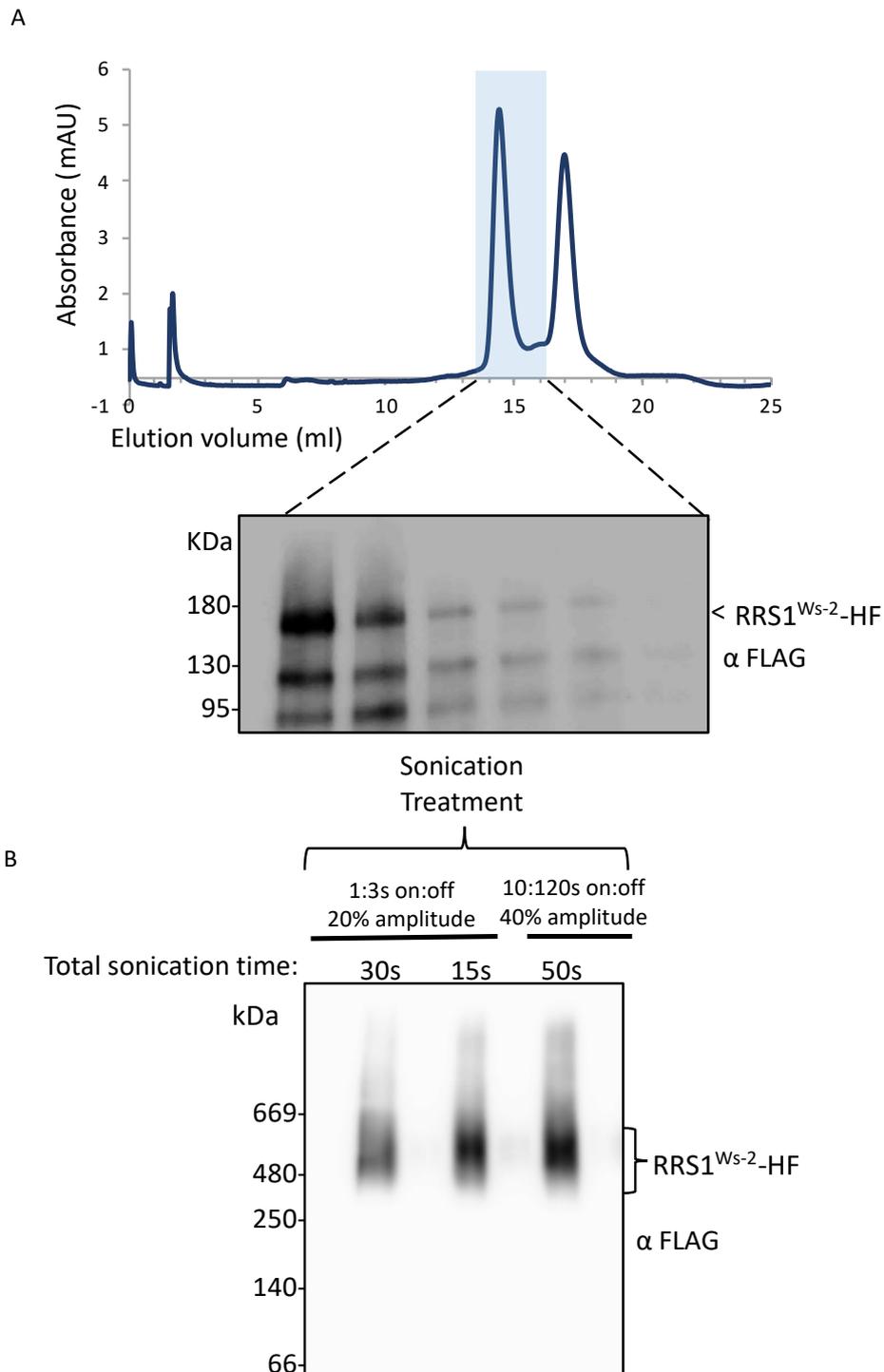
Figure 4.10 RRS1 [Ws-2]-HF purified from 35S::RPS4-HS/ 35S::RRS1 [Ws-2]-HF *A. thaliana* lines forms a multimeric complex. (A) Gel filtration analysis of anti-FLAG bead 3xFLAG peptide eluted RRS1 [Ws-2]-HF protein, presence of protein in peak highlighted in blue was confirmed by immunoblot analysis of fractions from this peak using anti-FLAG antibody. These fractions were also immunoblot probed for RPS4-HS protein but no protein was detected, experiment repeated twice with similar results observed. **(B)** Blue-Native PAGE analysis of samples of 35S::RPS4-HS/ 35S::RRS1 [Ws-2]-HF *A. thaliana* nuclear extract sonicated with different treatments shows a multimeric speared band of RRS1 [Ws-2]-HF protein as visualised by immunoblot analysis using an HRP-conjugated anti-FLAG antibody.

4.3.3 Purifying activated RRS1 and RPS4 from transgenic *A. thaliana* with inducible AvrRps4 expression

In order to gain mechanistic insights into the functioning of the RRS1-RPS4 complex, we need to compare biochemical and structural properties of both pre- and post-effector activated RRS1 and RPS4. This requires a source of effector-activated RRS1-RPS4 protein. Dr Pingtao Ding and Bruno Ngou had developed a resource for this material in the form of a transgenic line of *A. thaliana* expressing pAt3::RPS4-HA/ pAT2::RRS1[Ws-2]-HF with an β-estradiol inducible promoter LexA::AvrRps4-mNeon named the Super-ETI line (SETI line), **Figure 4.11.** This line expressed RRS1 and RPS4 under the moderately expressed promoters of pAt2 and pAt3 respectively, but crucially also used an β-estradiol-inducible system[234] for expression of the *P. syringae* effector AvrRps4. Generation of this line is described in Ngou et al, 2019[218]. This line therefore represents a system in which RRS1 and RPS4 can be activated by inducing the expression of AvrRps4-mNeon using β-estradiol without activating PTI in the plant cell. Expression of AvrRps4-mNeon through β-estradiol induction results in a biologically relevant activation of RRS1 and RPS4 as demonstrated by SETI line seedlings grown on 50 µM β-estradiol containing growth media which display severe growth arrest indicative of an activated immune response, **Figure 4.11B.**

As well as a line expressing WT AvrRps4-mNeon, a line expressing mutant AvrRps4$_{KRVY135-138AAAA}$-mNeon was also generated as a control. In the AvrRps4$_{KRVY-AAAA}$ mutant, residues 135-138 (KRVY amino acids) are mutated to alanine residues resulting in a AvrRps4 protein which does not activate RRS1-RPS4 dependent immunity[147], but retains the ability to bind to the WRKY domain of RRS1[58], and therefore serves as a negative control for WT AvrRps4 activation.

For protein purification, pAt3::RPS4-HA/ pAT2::RRS1 [Ws-2]-HF/LexA::AvrRps4-mNeon *A. thaliana* seeds were sown in liquid 1% MS media as per 4.3.1 and grown for ~ 2 weeks. As β-estradiol can permeate cell membranes, this method of induction represents a method for inducing AvrRps4 expression in *A. thaliana* plants on a large scale by placing plants in liquid media containing β-estradiol. After two weeks of growth in 1% MS media, seedlings were treated with β-estradiol by replacing liquid growth media with fresh 1% MS media containing 50 µM β-estradiol. For mock treatment, plant growth medium was replaced with fresh 1% MS medium.
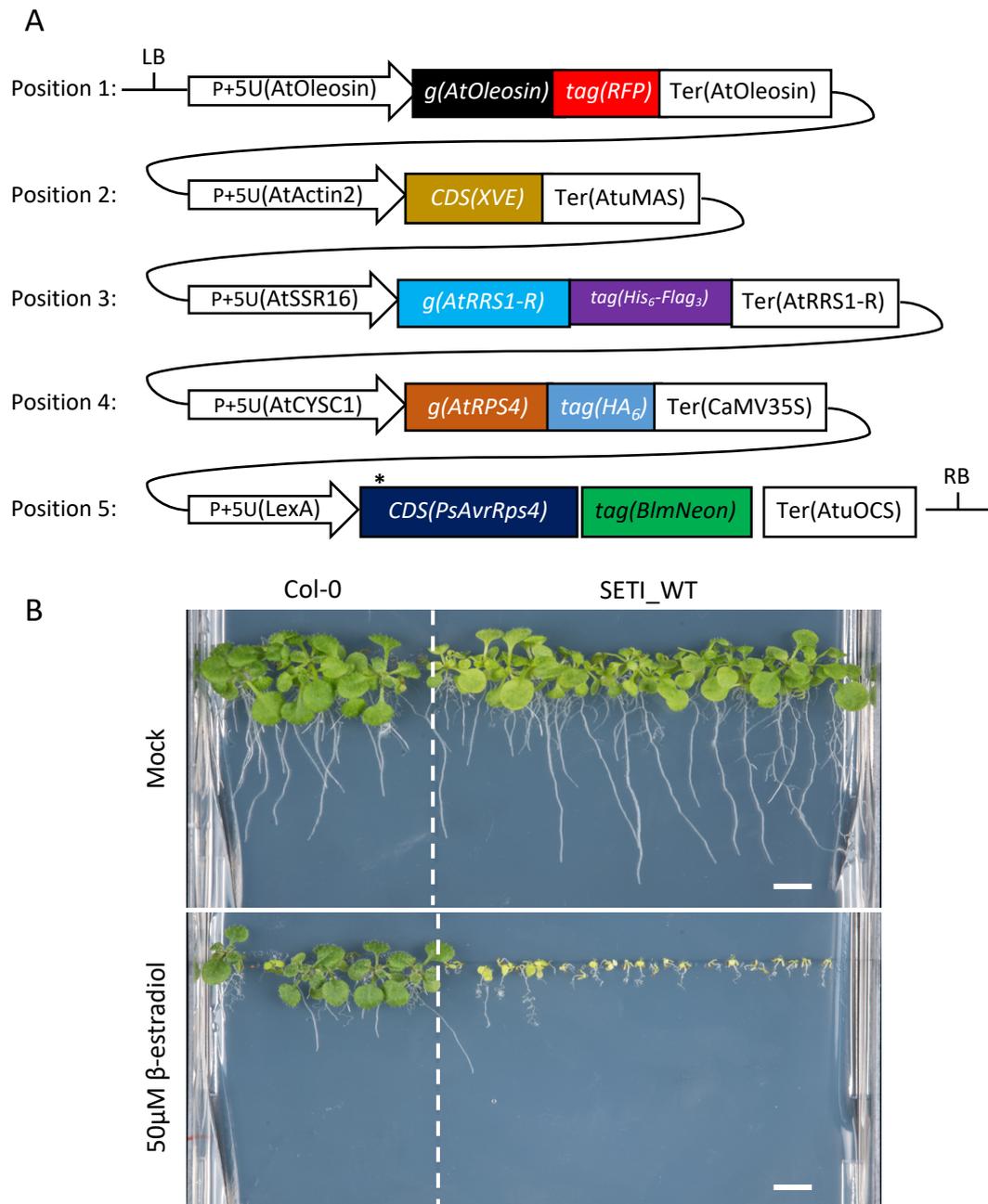
Figure 4.11 Single T-DNA expresses RRS1 [Ws-2]**-HF, RPS4-HA and inducible wild-type AvrRps4-mNeon or AvrRps4 mutant variants (A)** Illustrative layout of the SETI construct. There are five individual expression units listed, which are indicated position 1 to position 5. Position 1; expression unit of the FastRed selection marker (Shimada et al. 2010). Position 2 & 5; chimeric transactivator XVE (LexA-VP16-ER) and the corresponding LexA inducible system to express AvrRps4 or its mutant variants under the control of β-estradiol treatment. Position 3 & 4; full-length RRS1-R and RPS4 proteins with HF and HA epitope tags, respectively. LB: Left Border, RB: Right Border, CDS: Coding sequence, Ter: Terminator **(B)** Seedling phenotype of SETI Arabidopsis transgenic line at 14 days after germination in growth media containing Mock (0.1% DMSO) or 50μM β-estradiol. Col-0 was sown as control for

the effect of β-estradiol on seedling growth. Scale bar = 0.5cm. Figure taken from Ngou et al, 2019[218].

To establish the best time point to harvest RRS1^Ws-2-HF and RPS4-HA protein from transgenic seedlings post β-estradiol induction, I sampled seedlings for immunoblot analysis of AvrRps4-mNeon protein expression levels from 0-960 minutes post β-estradiol induction. This showed that though AvrRps4 expression could be observed shortly after β-estradiol induction, accumulation of AvrRps4$_{WT}$-mNeon protein appeared to peak at ~360 minutes post induction, before decreasing by 960 minutes, **Figure 4.12**. The immunoblots also suggested that the LexA::AvrRps4-mNeon may show very low levels of leaky expression of AvrRps4-mNeon prior to β-estradiol induction. However as demonstrated in **Figure 4.11B** seedlings grown on solid growth media containing no β-estradiol do not show the immune-activated stunting of seedlings grown on β-estradiol-containing media suggesting that leaky expression of LexA::AvrRps4-mNeon is insufficient to activate a defence response.
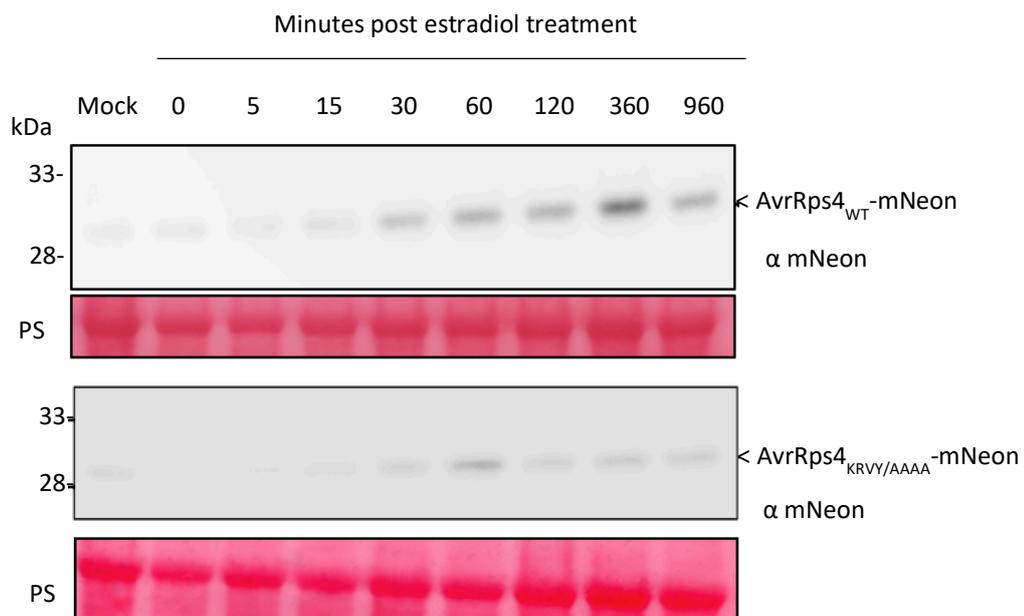
Figure 4.12 Accumulation of AvrRps4 protein in SETI line post β-estradiol induction. Immunoblot analysis of accumulation of AvrRps4$_{WT}$ and AvrRps4$_{KRVY/AAAA}$ protein from 0-960 minutes post β-estradiol treatment. Total protein extract was loaded onto a 4-20% SDS-PAGE gradient gel and proteins detected with immunoblot analysis using anti-mNeon primary antibody and anti-rabbit secondary antibody. Ponceau S Staining (PS) of the Rubisco large subunit was used as a loading control.

Having established AvrRps4-mNeon protein accumulation appeared to peak ~360 minutes post β-estradiol treatment, I next investigated fluctuations in RRS1-HF and RPS4-HA protein levels in this time course. There has been discussion in the field as to whether effector activation of NLRs can lead to stabilization of NLR proteins as part of the ETI response, particularly in regards to nucleocytoplasmic NLRs such as RPS4[235]. In this model, pathogen perception leads to an accumulation of NLRs in the nucleus where receptors then activate immune signalling responses via transcriptional reprogramming[236]. I hypothesised that we might observe changes in the accumulation of RRS1 or RPS4 protein in the SETI lines following AvrRps4 expression by β-estradiol induction. To investigate this, I sampled β-estradiol SETI line seedlings for immunoblot analysis of RRS1-HF and RPS4-HA protein levels from 0-480 minutes post β-estradiol induction. Immunoblot analysis showed that whilst RPS4-HA protein levels appeared to strongly accumulate over this time span, RRS1 protein levels appeared to decrease, **Figure 4.13**.

Minutes post estradiol treatment
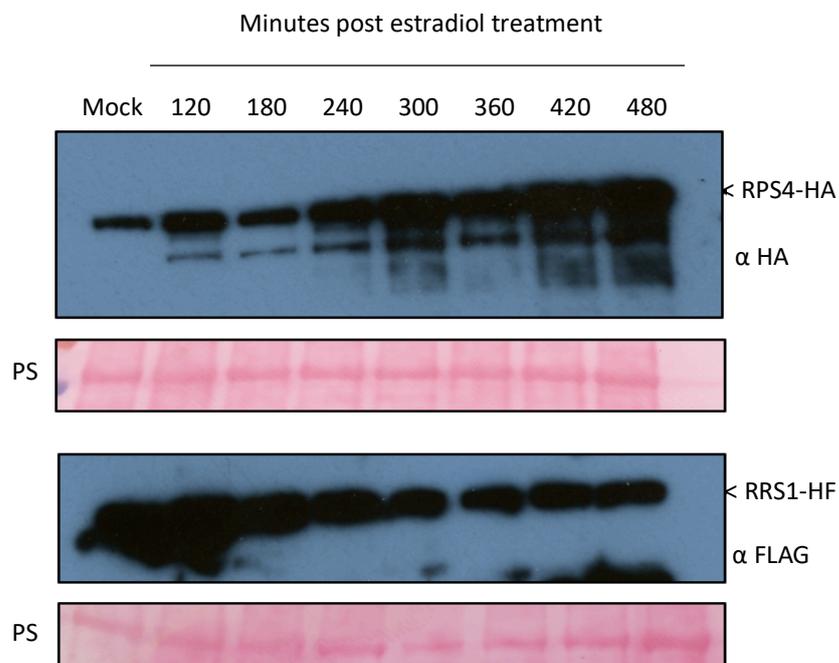
Figure 4.13 Changes in RRS1 and RPS4 protein levels post β-estradiol induction of AvrRps4 in SETI lines. Immunoblot analysis of changes in protein level of RRS1-HF and RPS4-HA 0-480 minutes post β-estradiol induction of AvrRps4. Total protein extract was loaded onto a 4-20% gradient SDS-PAGE gel and proteins visualised using an HRP-conjugated anti-FLAG or anti-HA antibody. Ponceau S Staining (PS) of Rubisco large subunit was used as a loading control. Experiment was repeated twice with similar results observed each time.

This observation may provide interesting insights into the mechanism of this immune receptor complex. Conceivably, RRS1 perception of AvrRps4 and subsequent activation of the receptor may help to stabilise RPS4 protein allowing RPS4 to accumulate to the threshold levels required for the activation of immune signalling. Further investigation is needed to validate this hypothesis. This model however does fit with observations in the literature which have shown for example that the Barley NLR MLA10 accumulates in the nucleus upon perception of the barley mildew effector AvrA10[237].

For purifying activated RRS1-RPS4 complex from SETI *A. thaliana* tissue, seeds were grown as described above and AvrRps4-mNeon protein induced by replacing the liquid growth medium with 1% MS supplemented with 50 µM β-estradiol. Tissue was then harvested and frozen in liquid nitrogen 360 minutes post AvrRps4-mNeon induction. Protein was purified using the total extract method described in **Figure 4.7,** and protein immunoprecipitated using anti-FLAG affinity beads to pull down RRS1-HF and its associated proteins. Samples were taken for immunoblot analysis of RRS1-HF and RPS4-HA protein throughout the purification process, **Figure 4.14**. However, a significant amount of RPS4-HA protein was found in the supernatant fraction of the anti-FLAG affinity beads.

Purified RRS1-HF was concentrated and loaded onto a Superose™ 6 Increase column (GE Healthcare) pre-equilibrated in running buffer (150 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA) but unfortunately due to machine malfunctions I was unable to analyse the complex size of this protein by analytical gel filtration. This work is now being followed up by Dr Hee-Kung Ahn who is investigating purifying RRS1 and RPS4 protein from this line for biochemical characterisation.
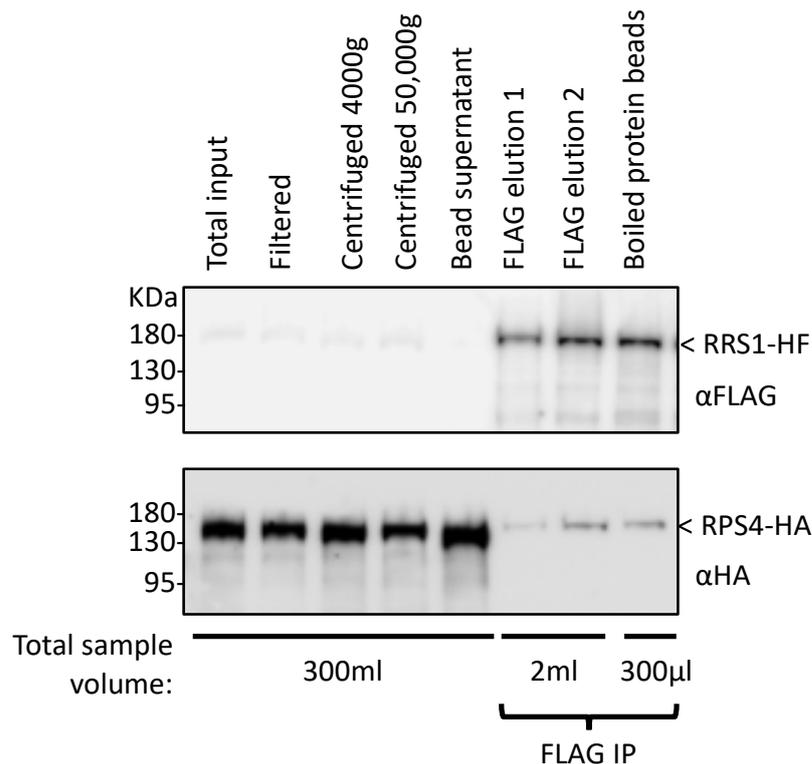
Figure 4.14 Purification of RRS1^{Ws-2}-HF and RPS4-HA via anti-FLAG affinity beads post AvrRps4 β-estradiol induction. Immunoblot analysis samples taken throughout purification process. Following immunoprecipitation, proteins were eluted from anti-FLAG beads by competitive binding with 3xFLAG peptide. Samples were loaded on 4-20% gradient SDS-PAGE gel and proteins detected by immunoblot analysis. Proteins were visualised using an HRP-conjugated anti-FLAG or anti-HA antibody. Total sample volume indicates the total volume of sample at each stage of purification of which 20 µl was loaded on SDS-PAGE gel. Experiments were repeated twice with similar results found across repeats.

4.3.4   Conclusions of expressing RRS1 and RPS4 in transgenic *A. thaliana*

Whilst I was able to show that I could purify both RRS1 and RPS4 from 35S::RPS4-HS/ 35S::RRS1^{Ws-2}-HF *A. thaliana* and both NLRs together in complex from pAt3::RPS4-HA/ pAT2::RRS1 $^{Ws-2}$-HF/LexA::AvrRps4-mNeon following β-estradiol induction, it was clear from immunoblot analysis of the purification processes that the limiting factor for producing large quantities of NLR protein was in the economic constraints of using anti-FLAG/HA affinity beads. Whilst *A. thaliana* seedlings provided a good source of economically and physically scalable plant material, the use of this pipeline was limited in this study due to the purification method utilised of anti-FLAG/HA affinity beads to purify POI. As these

affinity beads are not rechargeable, the cost of the quantity of beads required to purify POI to sufficient quantities for biochemical analysis was not viable for this project.

4.4 Transient expression of RRS1 and RPS4 in *Nicotiana benthamiana*

Based on the success of producing and purifying soluble RRS1 in *A. thaliana,* I wanted to trial an alternative *in planta* expression system for RRS1 and RPS4 protein. As the limiting factor for purification of RRS1 and RPS4 using transgenic *A. thaliana* lines was the use of anti-FLAG/HA affinity beads, I next wanted to test a system which would enable us to trial a variety of different affinity tags for purification whilst keeping the scalability and folding environment that was successful in the *A. thaliana* trials. For this we decided to utilise agrobacterium-mediated transient expression in *N. benthamiana*. We hypothesised this system would offer a similar protein folding environment akin to *A. thaliana* and small-scale co-immunoprecipitation work in the Jones laboratory had previously showed that RRS1 and RPS4 could be expressed in this system[58]. Additionally, unlike the use of stable transgenics, a variety of different purification tags and expression vectors can be trialled in this system on a rapid timeline.

To demonstrate that agrobacterium mediated transient transformation of *N. benthamiana* is a viable source of the production of RRS1 and RPS4 protein, I set out to repeat the expression and purification pipelines used for the purification of these proteins from the total extract of transgenic *A. thaliana* lines in *N. benthamiana*, **Figure 4.15**. Following this method, I showed that I could immunoprecipitate full-length RRS1^{Ws-2}-HF and RRS1B-HF from cDNA and gDNA expressing constructs, **Figure 4.16**. However similar to *A. thaliana* trials, significant amounts of RRS1^{Ws-2}-HF and RRS1B-HF protein could be seen in the unbound fraction post treatment with anti-FLAG affinity beads. This suggests that the anti-FLAF affinity beads were saturated such that significant quantities of protein were not coimmunoprecipitated and implies therefore the limiting factor for yield of POI is quantity of affinity beads, **Figure 4.16**.

In addition to full-length RRS1^{Ws-2}, I wanted to explore whether shorter domain truncations of RRS1 and RPS4 could be also be expressed and purified using this method. I focussed these experiments on regions of high interest of RRS1 and RPS4, namely RRS1's Dom4 WRKY Dom6R/S and RPS4's CTD. Following the same pipeline seen in **Figure 4.15**, I was able to immunoprecipitate RRS1 Dom4 WRKY Dom6R/S and RPS4 CTD using anti-FLAG affinity

beads, **Figure 4.17A,B**. However, Coomassie staining of boiled affinity beads post immunoprecipitation showed no bands of RRS1 Dom4 WRKY D6R/S-HF or RPS4 CTD implying that protein yields were likely not high enough for the downstream biochemical and structural experiments we wished to conduct such as crystallography and ITC, **Figure 4.17C**.
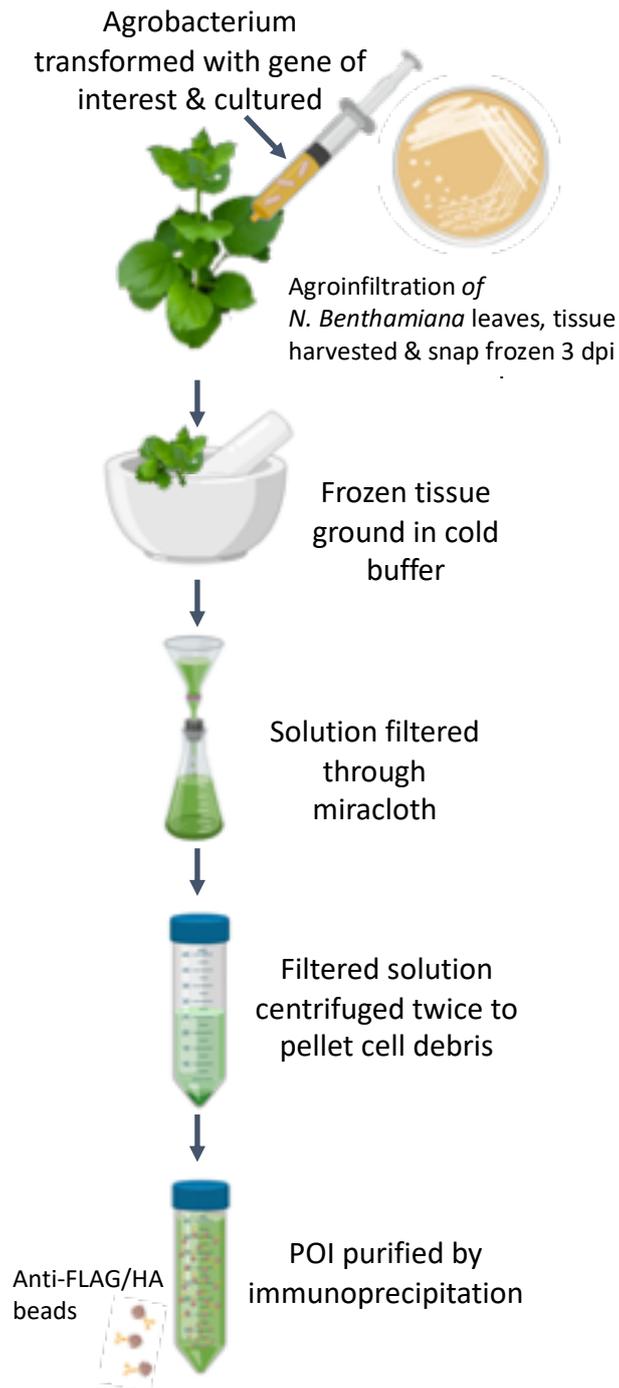
Figure 4.15 Purification of NLRs from agroinfiltrated *N. benthamiana* leaves workflow. The gene of interest was transformed in a level 1 golden gate vector into agrobacterium which was then grown and cultured. *N. benthamiana* leaves were then agroinfiltrated with this agrobacterium and leaves harvested and frozen in liquid nitrogen 3 dpi. Tissue was then ground in extraction buffer, filtered and centrifuged twice, low-speed and high-speed spin, to pellet cell debris. POI was then purified from lysate by immunoprecipitation using anti-FLAG beads. Figure generated using BioRender.
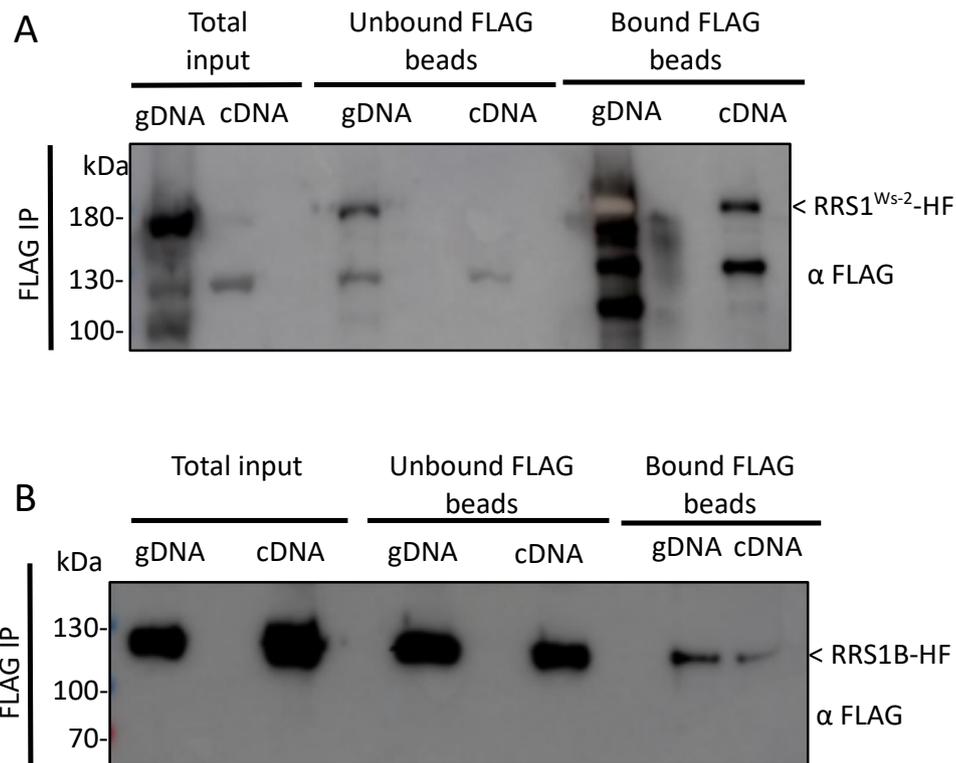
Figure 4.16 Anti-FLAG affinity bead purification of RRS1^{Ws-2}-HF and RRS1B-HF from agroinfiltrated *N. benthamiana* **leaves expressing cDNA and gDNA constructs**. **(A)** immunoblot samples of total input, unbound anti-FLAG bead fraction and immunoprecipitated boiled anti-FLAG affinity bead fraction from the total extract of *N. benthamiana* leaves expressing cDNA and gDNA constructs of RRS1^{Ws-2}-HF or **(B)** RRS1B-HF. Leaves were infiltrated with transformed agrobacterium and leaves harvested 3 dpi. Protein samples was run on 4-20% SDS-PAGE gel and proteins visualized using immunoblot analysis with HRP-conjugated anti-FLAG antibody. POI was released from anti-FLAG beads by boiling in SDS buffer. Experiment was repeated twice with similar results observed each time. Experiment conducted with Freya Harthshorn.

The major advantage of using agroinfiltration of *N. benthamiana* as a source of the production of RRS1 and RPS4 protein compared to transgenic *A. thaliana* is the speed of testing different purification variables, especially protein construct and affinity tag. Trials, particularly with full-length RRS1^{Ws-2}-HF and RRS1B-HF, suggested that whilst these proteins were well expressed in the system, the bottleneck in protein yields was in the cost of anti-FLAG affinity beads. Future studies will therefore be able to utilise this system to quickly and efficiently trial different affinity tags that would allow for more economical purification techniques. For example, using a Streptavidin based system where recombinant protein is expressed with a Strep® epitope tag, purified using a Strep-Tactin® resin and eluted with competitive binding with biotin, after which the resin can be stripped and recharged. Work

Investigating plant-based expression systems for plant NLR purification

continues with Dr Nitika Mukhi to trial other affinity tags in this system including use of Streptavidin tag and an Im9-E9 tag system.
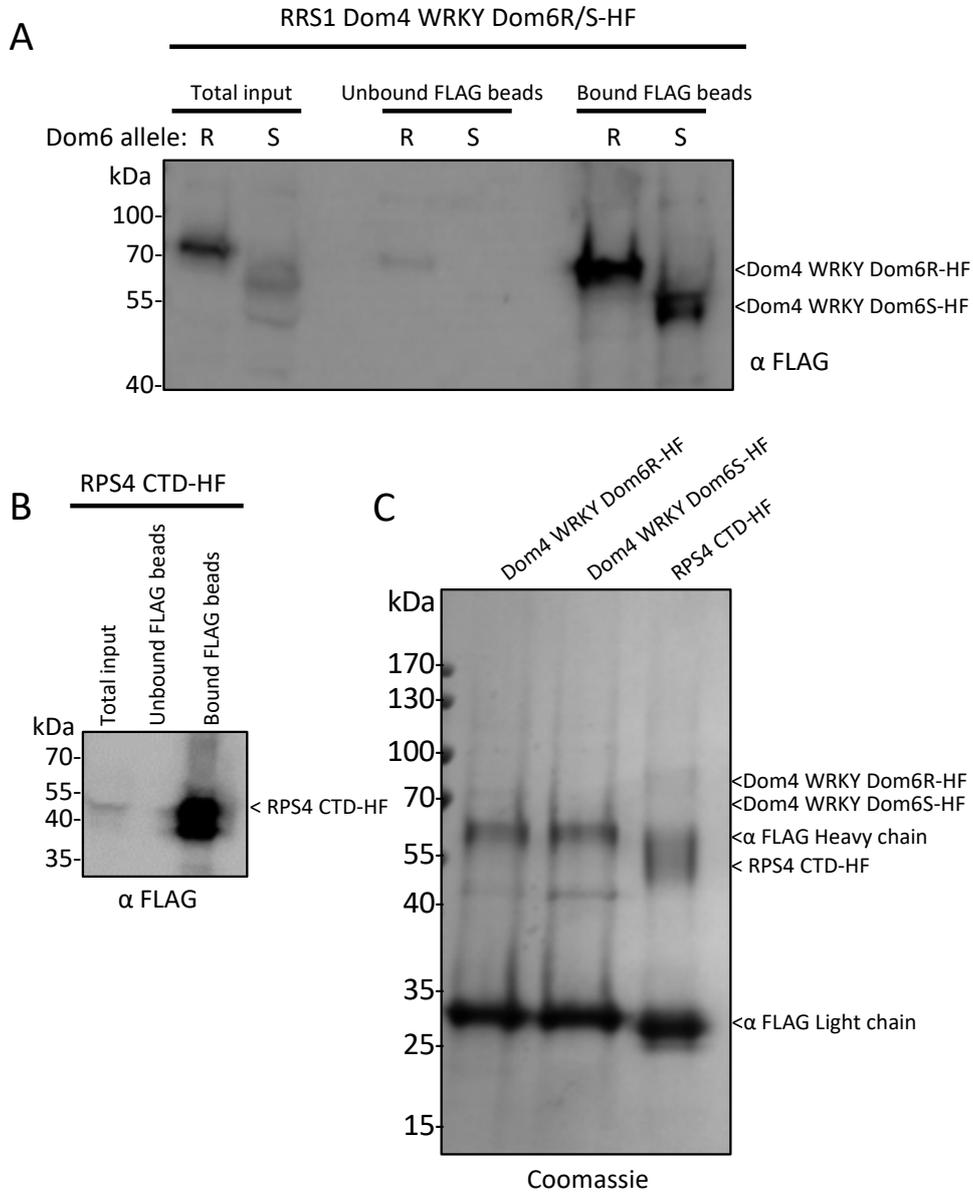
Figure 4.17 Anti-FLAG affinity bead purification of RRS1 and RPS4 domain truncations from agroinfiltrated *N. benthamiana* leaves. (A) Immunoblot samples of total input, unbound anti-FLAG bead fraction and immunoprecipitated boiled anti-FLAG affinity bead fraction from the total extract of *N. benthamiana* leaves expressing RRS1 Dom4 WRKY Dom6R/S-HF or **(B)** RPS4 CTD-HF. Leaves were infiltrated with transformed agrobacterium and leaves harvested 3 dpi. **(C)** Samples of boiled anti-FLAG affinity beads were run analyzed for POI yields by Coomassie staining, arrows indicate the predicted protein size for each protein and antibody heavy and light chain. Protein samples was run on 4-20% SDS-PAGE gel and proteins visualized using immunoblot analysis with HRP-conjugated anti-FLAG antibody. Experiment conducted with Freya Harthshorn.

4.5   Discussion

4.5.1   Comparing plant-based expression systems for RRS1 and RPS4 protein production

Following failures to produce soluble multidomain RRS1 and RPS4 protein in the heterologous systems of *E. coli* and Sf9 insect cells, I hypothesised that RRS1 and RPS4 proteins may require a folding environment more akin to their native *A. thaliana* to support soluble expression. I therefore investigated the use of a variety of plant-based expression systems for their ability to produce soluble RRS1 and RPS4 protein in an economical and physically scalable manner. As well as in vivo systems, I evaluated use of a cell-free expression system to expand the conditions investigated. The systems subsequently evaluated were a cell-free wheat germ system, transgenic *A. thaliana* and agroinfiltrated *N. benthamiana,* a list of successful soluble protein expression trials can be found in ***Appendix 5.***

The first system trialled was a cell-free wheat germ extract. I showed that soluble expression of both truncated WRKY Dom6S and full-length RRS1^{Ws-2} was supported in this system. This suggested that as hypothesised, using an expression system with a folding environment and capabilities more analogous to RRS1's native *A. thaliana,* facilitated soluble expression of this NLR protein. Expression was observed in wheat germ extract from two different sources, commercially produced CellFree Sciences and from extract generated by Professor Yasuomi Tada at Nagoya University. Evaluating on the basis of immunoblot signal development conditions it appeared the extract from CellFree Sciences expressed the POI to a higher level though this is not an accurate enough method of protein level comparisons to draw a conclusive statement. However, neither system was capable of producing RRS1 protein, truncated or full-length, to quantities such that could be observed by Coomassie staining of an SDS-PAGE gel.

This study used a bilayer method for protein translation in wheat germ extract as previous reports have shown this method to produce higher protein yields compared to traditional batch method style. For example, one study found that 1 ml of wheat germ extract was capable of producing 3.2 mg of GFP by bilayer translation compared to 1.6 mg for batch method setup[170]. There are alternative higher yielding methods for protein translation with wheat germ extract which were not trialled in this study as they were not practical for

quantities of wheat germ I was using. Methods such as dialysis for example was shown in the same study described above to produce 20 mg of GFP protein per 1 ml of wheat germ extract[170]. Future studies using this system to produce NLRs could therefore investigate use of other translation method systems in efforts to boost yields. However, given the quantities of RRS1 protein I detected from this system in this study it is clear, at least in the case of RRS1, that higher quantities of wheat germ extract would be needed to produce the quantity of protein required for structural studies such that the system is too costly for our purpose. It should be noted that whilst this system may be inappropriate for producing RRS1 protein quantities for structural work, it could potentially be utilised for quantitative biochemical techniques such as SPR (surface plasmon resonance). SPR requires significantly lower protein quantities that other interaction measuring techniques such as ITC (isothermal titration calorimetry). Furthermore, in SPR analysis the POI can be purified directly from extracts by using various SPR immobilization matrix chip coatings such as biotin or nickel-NTA. In this manner, SPR could be well suited for downstream studies of wheat germ extract-produced protein.

Work with expressing RRS1 and RPS4 protein in cell-free wheat germ extract demonstrated the importance of a plant protein folding environment for the soluble expression of these proteins. However, the required scale up of this to the yields of protein needed for downstream biochemical and structural studies were too costly. I therefore tested a system which provided a plant-based protein folding environment but which could be scaled in a more economical manner. For this I investigated a transgenic line of *A. thaliana* overexpressing RRS1 and RPS4 protein (35S::RPS4-HS / 35S::RRS1 $^{Ws-2}$-HF) in a Col-0 background. By growing these lines shaking in liquid 1% MS media I was able to produce ~100 g of fresh weight tissue from ~10 x 250 ml flasks making this system highly scalable.

Using this system, I was able to produce soluble RRS1 protein to quantities such that the protein could be detected during analytical gel filtration analysis. This result is a very promising lead for future biochemical and structural work with RRS1 protein as it suggests this system could be used to produce significant quantities of protein for downstream study. Whilst yields required for crystal trials may still be unachievable, further optimisation and scale up could feasibly produce quantities of RRS1 protein appropriate for SAXS (small-angle X-ray scattering) analysis. Should a SAXS envelope of RRS1 protein be attained, the published structures of RRS1's WRKY domain[71] and TIR domain[114] along with modelling, based on published structures such as ZAR1[109], could then be used to gain a full-length

structure of RRS1. In doing so this would be the first full-length structure of a plant paired NLR protein. RPS4 protein however could not be purified to adequate quantities from the same fresh weight of 35S::RPS4-HS / 35S::RRS1[Ws-2]-HF tissue, though RPS4 protein was detectable by immunoblot analysis. The reason for this is not understood and may reflect a the potentially lower stability of RPS4 protein compared to RRS1. It could be hypothesized that post-translation regulation of RPS4 protein may be preventing accumulation of RPS4 to levels comparable to RRS1, possibly as a mechanism to control against autoimmunity.

Using an *A. thaliana* transgenic line expressing pAt3::RPS4-HA/ pAT2::RRS1[Ws-2]-HF/LexA::AvrRps4-mNeon (SETI line) I was able to purify RRS1-RPS4 complex post β-estradiol induction of AvrRps4. This result demonstrates an exciting avenue for future work as this line could provide a feasible source of activated RRS1-RPS4 complex for downstream structural studies, particularly SAXS analysis. This transgenic line also provides an exciting resource for investigating the biochemical functions of the RRS1-RPS4 complex and particularly the interplay of PTI and ETI on activation of this complex as discussed in 4.5.2.

Together these results highlight the valuable resource transgenic plants can provide for producing plant NLR protein. Transgenic plants offer a system with sufficient protein folding capabilities and both physical and economical scalability, excelling where previous work in this study with *E. coli*, insect cells and wheat germ cell-free extract failed. The bottleneck I found in work using this system however came during the purifying of the POI from the transgenic tissue extract. The workflow in this study relied on the use of antibody affinity beads which are expensive to use and therefore provide a limitation on the quantity of protein which can be purified using them. Additionally, RRS1 and RPS4 protein were inefficiently released from the affinity beads using competitive binding peptide elution. This suggests that future use of transgenic plants for NLR purification may wish to investigate the use of other epitope and affinity tags with different purification methods which utilise other rechargeable affinity systems. Using rechargeable affinity beads would allow for increased quantities of affinity beads to be used per prep due to a reduction of cost constraints ensuring more of the POI would be captured and purified from a given volume of plant material. For example, use of the Strep® epitope tag and Strep-Tactin® columns which are rechargeable and therefore more economical in their use[238].

One system which could be used to trial the efficiencies of various affinity tags is agroinfiltration of *N. benthamiana*. The major benefit transient expression using

agroinfiltration in *N. benthamiana* offers over transgenic lines is in the ability to trial multiple purification and expression conditions for protein production in a short time frame. This system could be therefore be of great use to trial a variety of purification variables before a transgenic line was then generated as this would combine the speed of trialling conditions of agroinfiltration with the economic and physical scalability of transgenic plants. Preliminary work in this study showed the viability of this system to produce soluble full-length RRS1 [Ws-2] and RRS1B protein as well as truncations of RRS1 and RPS4 which could be immunoprecipitated using anti-FLAG beads. Whilst POI purified from agroinfiltration *N. benthamiana* tissue were detectable by immunoblot analysis, proteins were not detectable by Coomassie staining of SDS-PAGE gels. This suggests that to produce yields of protein for downstream structural studies, production in this system would require significant scaling up which would be costly with the use of anti-FLAG beads. This again supports the concept of expanding purification trials to include the use of affinity tags which utilise rechargeable affinity resins and beads for purification.

Whilst this study investigated the use of a diverse range of plant-based expression systems it was not exhaustive and a number of alternative plant-based systems and methods could be explored for future studies of this kind. For example, the use of plant cell cultures such as *N. tabacum* BY-2 or NT-1 cells[226]. The plant cell culture system offers simple transformation and propagation pipelines as well as precise control over growth conditions and batch-to-batch consistency with recombinant protein production. These factors combined with the benefits of a plant eukaryotic protein folding environment, offer future studies and interesting alternative to whole plant systems for investigating NLR protein production.

4.5.2 Investigating the interactions of RRS1 and RPS4 protein pre- and post-AvrRps4 activation

Whilst investigating expression and purification of RRS1 and RPS4 from various plant-based expression systems a number of interesting observations were made on the interactions of these NLRs pre- and post-AvrRps4 activation.

Changes were observed in the total quantity of RRS1 and RPS4 protein post-AvrRps4 induction. RPS4 protein was seen to accumulate post induction of AvrRps4 whilst RRS1

protein appeared to show a small decrease. Previous studies have shown that RPS4 TIR domain has a self-association surface and that RPS4 TIR homodimerization is required to activate RPS4-mediated HR in *N. tobaccum*[114] . Following on from this, subsequent work suggested that a threshold of RPS4 protein accumulation may need to be reached for RPS4 protein to self-associate and also found that self-association of RPS4 protein requires the presence of RRS1[233]. Taken together, it could therefore be hypothesised that upon perception of AvrRps4 by RRS1, RPS4 protein is stabilised such that RPS4 protein accumulates enabling dimerization of RPS4's TIR domains and induction of a defence response and HR.

The concept of NLR oligomerisation or formation of multimeric complexes has been widely discussed in the field. The recently published structure of the ZAR1 pentameric 'resistosome' shows that oligomerization of NLRs may have a direct role in cell death by the insertion of NLR multimers directly into the cell membrane leading to the formation of membrane pores[109]. Whilst this is likely not the case with all NLRs (e.g. RRS1 and RPS4 are found in the nuclei rather than the cell membrane[58]), it raises an important discussion around the role of oligomerisation in NLR activation. I observed that RRS1 [Ws-2]-HF protein produced in both the cell-free wheat germ extract and transgenic *A. thaliana* plant appeared to form a multimeric complex in a consistent molecular weight range. This was observed in both Blue Native PAGE and analytical gel filtration data. This result supports previous observations in the literature on the ability of RRS1 to self associate[233]. Whilst inabilities to produce RPS4 cDNA meant I was unable to investigate RRS1-RPS4 interactions in the wheat germ cell-free extract, the high molecular weight complexes I observed formed by RRS1 purified from 35S::RPS4-HS/ 35S::RRS1 [Ws-2]-HF did not appear to contain RPS4. Whether this is due to purification conditions or biologically meaningful is not yet understood as interactions between these proteins pre-AvrRps4 induction have been observed in other studies[218,233].

RRS1 [Ws-2]-HF purified from 35S::RPS4-HS/ 35S::RRS1 [Ws-2]-HF migrated as a diffuse Blue Native PAGE band in the range of 480 to approximately 900 kDa indicating the formation of heterogeneous high molecular weight complexes. Interestingly work conducted by Dr Hee-Kyung Ahn in Ngou et al 2019[218] found similar heterogeneous forms of complexes in a similar molecular weight range were formed by RRS1 and RPS4 protein in pAt3::RPS4-HA/ pAT2::RRS1[Ws-2]-HF/LexA::AvrRps4-mNeon (SETI) *A. thaliana* lines pre- and post-AvrRps4 induction. This suggests that there might not be a change in the overall size of the

complexes of RRS1 and RPS4 protein upon effector activation. Given the observed changes in quantity of RRS1 and RPS4 protein post AvrRps4 induction, one could hypothesise that the composition or conformation of the heterogeneous high molecular weight complexes may change post effector activation. Following from the observation of elevated accumulation of RPS4 protein post-AvrRps4 induction, it could be hypothesised that whilst the overall size of complex formed by RRS1 and RPS4 protein does not differ post AvrRps4 activation, the ratio of RRS1:RPS4 protein may, with an increasing proportion of RPS4 protein found in the complex. Alternatively, other proteins may be recruited into the RRS1-RPS4 complex post activation replacing pre-activation complex proteins such that little change in the overall size of RRS1-RPS4 complex is seen by Blue Native PAGE immunoblot analysis. This could provide an interesting avenue of work for future mechanistic studies.

An exciting area of study in the field of NLR mechanisms is in the interplay of PTI and ETI. Up until the development of effector inducible lines such as the SETI line, previous work on NLR mechanisms was largely done using agroinfiltration meaning observations were made in the presence of both ETI and PTI. Using effector-inducible lines such as the SETI line, we can now investigate the behaviour of NLRs upon ETI activation without the complications of co-activating PTI. It will be exciting to see what future insights these lines can bring in investigating conformational and oligomeric changes involved in NLR activation and teasing apart the roles that ETI and PTI play in defence induction.

Gaining biochemical & structural insights into AvrRps4 recognition by RRS1

5

5.1 Introduction & chapter aim

Understanding the structural basis of pathogen effector recognition is a primary objective in the field of NLR biology. Such knowledge is key for achieving future research goals in engineering synthetic NLRs with extended pathogen recognition capabilities. In particular, NLRs containing integrated domains offer an excellent foundation for the engineering of such synthetic receptors. If we can decipher the requirements of integrated domains to bind effectors, this information could then be used to engineer NLRs to target previously unrecognized effectors. In order to achieve such outputs, we need to first expand our understanding of how NLRs bind and perceive effectors at the structural level, and how these binding events leads to NLR activation and induction of ETI. To date, attempts to engineer NLRs with expanded recognition capabilities via wholescale integrated domain switching has largely resulted in the generation of autoactive phenotypes. For example, when the WRKY domain of RRS1 is switched for the WRKY domain of the PopP2 targeted transcription factor WRKY 41, the resulting NLR is autoactive in an RPS4 dependent manner[152]. It is therefore likely that only when we understand the intricacies of NLR-effector binding events at the structural level will we have the tools to begin engineering functional synthetic NLRs.

The primary goal of the work in this chapter is to understand the structural basis of effector recognition by the WRKY domain of RRS1. Given both the allelic variation of RRS1 A pair proteins, primarily RRS1-R and RRS1-S, and the identification of a paralogous B pair, RRS1B, this system provides a unique opportunity to understand the inherent flexibilities in effector recognition and activation of integrated domain NLRs. Given the limited sequence identity of only 56% between the WRKY domains of RRS1 and RRS1B, should structures of AvrRps4 in complex with RRS1 WRKY and RRS1B WRKY be obtained we could then gain insights into how a single effector protein binds and activates different NLRs, **Figure 5.1**. Additionally, the RRS1-R allele from the *A. thaliana* Ws-2 accession provides an intriguing resource for understanding the activation of an NLR containing an integrated domain, given its ability to recognise both structurally and functionally distinct effectors in AvrRps4 and

PopP2, **Figure 5.1**. Following publication of the structure of the RRS1 WRKY domain in complex with PopP2[71], we are now in an exciting position to investigate how one integrated domain can perceive and bind two structurally distinct effectors. However, this is dependent on attaining a structure of the RRS1 WRKY domain in complex with AvrRps4. Additionally, the identification of specific mutations in both RRS1-R and RPS4 that abolish recognition of either AvrRps4 or PopP2, but not both, suggest that this NLR pair is capable instigating different activation mechanisms through the same NLR complex[153]. Understanding how these different activation mechanisms are coordinated and maintained would thereby provide invaluable insights as to how we could synthetically design NLRs with multiple effector recognition specificities in the future.
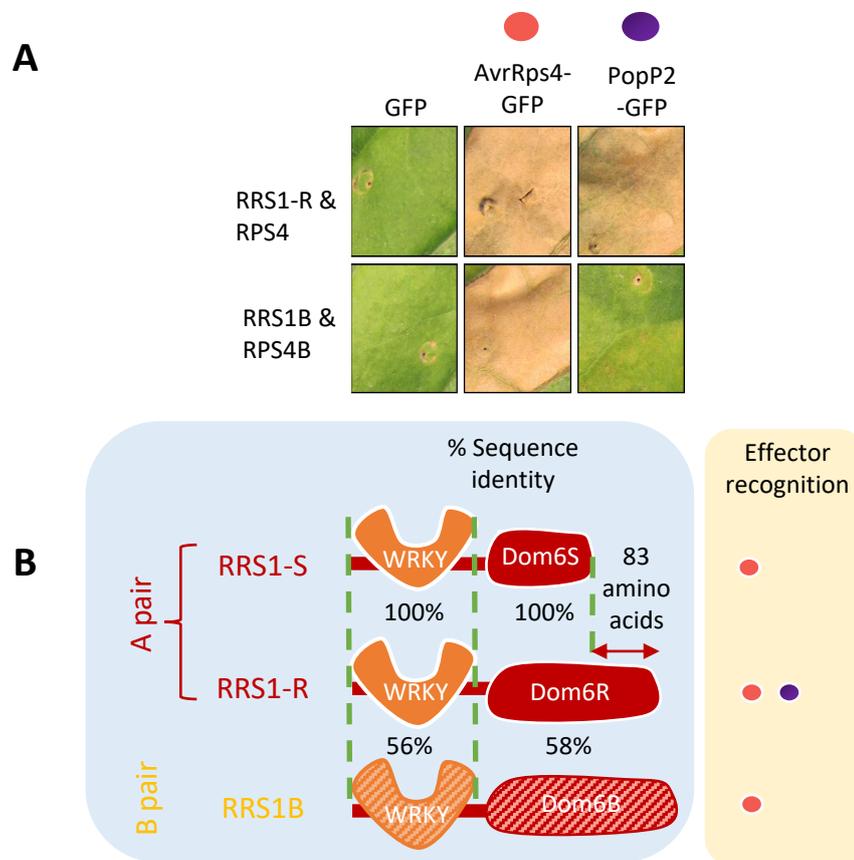
Figure 5.1 Effector recognition capability and sequence similarity of RRS1 alleles and paralogous B pair (A) RRS1-R/RPS4 pair together conger recognition of AvrRps4 and PopP2 whilst RRS1B/RPS4B confer recognition to only AvrRps4. Figure taken from Saucet et al 2015[153] showing the recognition of transiently expressed AvrRps4 and PopP2 by RRS1-R/RPS4 and RRS1B/RPS4B in N. tabacum. **(B)** Sequence identity of the WRKY domain and C-terminal regions of RRS1-S, RRS1-R and RRS1B. Effector recognition capabilities are shown in the panel on right with AvrRps4 recognition shown by pink circles, PopP2 by purple.

Previous work has provided key insights into the role of the WRKY domain in the activation of RRS1/RPS4 and RRS1B/RPS4B. Studies of allelic variation in the RRS1 protein has shown that the C-terminal Dom6R 83 amino acid extension beyond the end of Dom6S, in conjunction with the WRKY domain, is responsible for facilitating the extended recognition capability of RRS1-R to PopP2. The exact mechanism through which this extension facilitates expanded recognition is not yet thoroughly understood. As RRS1-S can still bind and be acetylated by PopP2, it is hypothesized that the RRS1 Dom6R extension may assist in specific inter-domain reconfigurations both within RRS1 or with RPS4, which allow RRS1-R specifically to translate PopP2 binding into activation of the RRS1-R/RPS4 complex[65,66,152]. In contrast to PopP2 perception, recognition of AvrRps4 by RRS1 does not require Dom6, and the truncation RRS1 ΔDom6S is still capable of activating an immune response in the presence of AvrRps4 in an RPS4 dependent manner[152].

It has been established that recognition of both PopP2 and AvrRps4 requires an intact 'WRKYGQK' motif in RRS1's WRKY domain. Specifically, mass spectrometry experiments have led to the identification of two lysine residues, K_1 (RRS1-R K1217, RRS1-S K1215) and K_2 (RRS1-R K1221, RRS1-S K1219), in the WRK_1YGQK_2 motif which are acetylated by PopP2 and play a key role in the recognition of both effectors[65,66]. Work expressing mutants in *N. tabacum* leaves showed that RRS1-R WRKY mutant WRKYGQ**R** (K1221R), but not WR**R**YGQK (K1217R), loses recognition of both AvrRps4 and PopP2. AvrRps4 recognition was however disrupted by acetyl-lysine mimic mutants WR**Q**YGQK (K1217Q) or WRKYGQ**Q** (K1221Q). This observation suggests that ability of RRS1 to perceive AvrRps4 is sensitive to subtle changes in the K_2 residue of the WRKY motif.

The importance of the conserved 'WRKYGQK' motif in RRS1's recognition of PopP2 was highlighted in the published crystal structure of the RRS1 WRKY domain and PopP2 complex[71]. This site formed the majority of interactions with PopP2, facilitated by a number of hydrogen bonds and van der Waals contacts. The structure also highlighted the K_1 and K_2 residues in the WRKY motif as facilitating key interactions with PopP2. The backbone and side chain of K1217(K_1) formed hydrogen bonds with PopP2 residues D292 and N296, and E284, respectively, whilst K1221(K_2) was inserted directly into the active site of PopP2[71]. It should be noted that this structure did not include any of RRS1's C-terminal Dom6R, which is known to be required for PopP2 recognition by RRS1-R[153].

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

Regarding interaction of the RRS1 WRKY domain with AvrRps4, the crystal structure of AvrRps4$_c$ has provided some key insights into the interactions that may underpin this binding event. The electrostatic surface of AvrRps4$_c$ revealed a prominent negative patch, highlighted in **Figure 5.2,** which contained two glutamic acid residues, E175 and E187. Subsequent mutation analysis showed that bacteria expressing AvrRps4 E175A or E187A were not capable of triggering RRS1/RPS4 dependent HR. Interestingly protein-protein interaction modelling of RRS1 or RRS1B WRKY with AvrRps4$_c$, **Figure 5.7**, predicts that E175 and E187 interact with the K$_2$ residue of the WRKY motif. This suggests that these residues may form a key interaction surface in the RRS1/RRS1B WRKY-AvrRps4 binding event. This hypothesis is further supported by co-immunoprecipitation data which shows that mutant AvrRps4 E187A shows reduced binding to RRS1 Dom6R and complete loss of binding for double mutant AvrRps4 E187A/E175A[153].
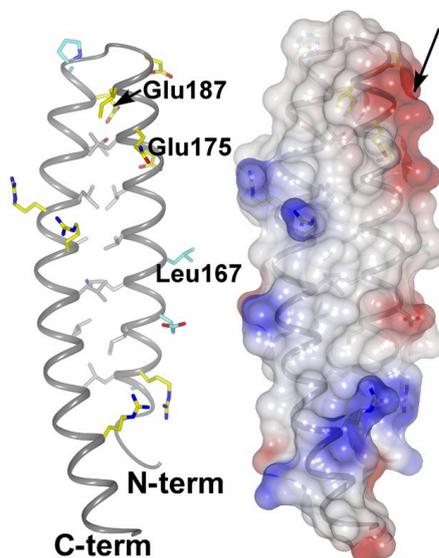
Figure 5.2 Antiparallel α-helical coiled coil structure of AvrRps4$_c$. (Left) Cartoon representation of the α-helices with residues in sticks on left of figure highlighting location of Glu187 and Glu175. (Right) Electrostatic surface representation of AvrRps4$_c$ with arrow highlighting prominent negative patch containing Glu187 and Glu175, Figure taken from Sohn et al, 2012[166]

In addition to E187 and E175, other residues of interest have also been highlighted in AvrRps4. Mutations of AvrRps4 residues K135-Y138, referred as the 'KRVY' motif, to alanine also abolishes RRS1/RPS4 dependent AvrRps4-triggered HR in *N. tabacum*. Interestingly, co-immunoprecipitation experiments show that unlike AvrRps4 E187A/E175A, KRVY/AAAA mutants retain the ability to bind to the RRS1 WRKY domain[147,166]. This suggests that AvrRps4 binding to the RRS1 WRKY domain is required, but not sufficient, to induce RRS1/RPS4 dependent HR. Such observations highlight the need for structural information

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

on this interaction as only in gaining an atomic structure of this NLR-effector interface will we be able to understand the requirements of this binding event.

The importance of DNA binding in RRS1 activation is still not fully understood. It was initially proposed that loss of DNA binding might provide a mechanism for RRS1/RPS4 complex activation by PopP2. However, experiments showing that the RRS1-S WRKY domain mutant K1219R abolishes DNA binding of this domain without activating plant defence signalling threw this hypothesis into contention[65]. Alternatively, loss of DNA binding by acetylation of this domain by PopP2 may be an artefact, in that PopP2 acetylates the integrated WRKY domain as it would its virulence target. Whilst evidence suggests that PopP2 disrupts WRKY domain DNA binding through acetylation of lysine residues in the WRKY motif, the mechanism for the non-enzymatic AvrRps4 is not understood. One finding which remains in contention with RRS1 DNA binding investigations is the observation in EMSA studies that AvrRps4 appeared to not disrupt the ability of RRS1-R to bind W-box DNA. This observation seems contradictory given the fact that the K_1 and K_2 lysine residues in the WRKY motif have been shown to directly interact with DNA[173,239] and are predicted to directly interact with AvrRps4 by protein-protein interaction modelling such that mutation of these residues abolishes AvrRps4 binding. This data would therefore confusingly suggest that the K_1 and K_2 residues can simultaneously bind AvrRps4 and W-box DNA. One explanation for this could be that the previously conducted EMSA study is not sensitive enough to detect changes in DNA binding abilities of RRS1 in the presence of AvrRps4. It is therefore one of the goals of this chapter to investigate this interaction in a more quantitative manner.

The aim of this chapter was to investigate the structural basis of AvrRps4 recognition by the RRS1 WRKY domain. Given the observation that binding of AvrRps4 is required but not sufficient to activate RRS1/RPS4 dependent HR[147,150], we require structural and more quantitative biochemical information on this interaction to understand this binding event. This chapter looked to gain structural, quantitative and qualitative information on AvrRps4 binding to RRS1 using the diversity of RRS1 alleles and paralogous proteins which make this NLR an intriguing system to understand the flexibilities of NLR effector recognition. The ultimate goal is to gain an atomic structure of RRS1 or RRS1B WRKY bound to AvrRps4 to allow us to investigate two key questions. Firstly, how a single NLR, RRS1-R, binds functionally and structurally distinct effectors by comparing the binding interface of RRS1 WRKY-AvrRps4 with the structure of RRS1 WRKY-PopP2. Secondly, how a single effector,

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

AvrRps4, binds different NLRs by comparing the interaction of AvrRps4 with both RRS1 and RRS1B WRKY domains which only share 56% sequence identity.

Through the quantitative and qualitative biochemical techniques of analytical gel filtration and surface plasmon resonance (SPR) this chapter looked to validate previous hypothesises about RRS1 and AvrRps4 residues involved in this binding event, and to provide mechanistic insights into the function of the RRS1/RPS4 complex. In order to gain structural and quantitative biochemical information on the RRS1 WRKY-AvrRps4 interface, it was necessary to produce soluble protein for each component. This chapter looked to identify purification pipelines for RRS1 and RRS1B WRKY domain protein for further studies. Following protein purification, this chapter looked to gain qualitative information on RRS1 WRKY-AvrRps4 binding by analytical gel filtration and quantitative data with SPR (surface plasmon resonance), and ultimately gain an atomic structure of this complex. Full methodologies used in this chapter can be found in chapter 2.

Regarding nomenclature used in this chapter, definitions for RRS1-R WRKY (E1190-H1269), Dom6S (P1270-C1290) and Dom6R (P1270-Y1273) are described in 1.5.1. The residues of these domains expressed for protein purification work are denoted as subscript or as brackets. Amino acid numbers are given in reference to RRS1-R unless otherwise stated.

5.2   RRS1 WRKY Dom6S interacts with AvrRps4$_c$ in vitro

5.2.1   Purifying RRS1 WRKY Dom6S and AvrRps4$_C$

Initial work with RRS1 WRKY Dom6S focussed on the construct described in 5.4, which expressed RRS1-R residues S1184-C1290 in the pOPINF vector. Following publication of the RRS1 WRKY structure by Zhang et al, 2017[71] (PDB ID 5W3X), I moved to working with the RRS1 WRKY construct described in the study, to allow for future comparative work. The construct used in the Zhang et al, 2017[71] study covers all five of RRS1's WRKY β-strands expressing residues E1195-T1273, **Figure 5.3A**. Defining the N-terminus as RRS1-R residue E1195, I trialled the expression of constructs RRS1 WRKY (E1195-T1273), WRKY Dom6S (E1195-C1290) and WRKY Dom6R (E1195-Y1373), **Figure 5.3A**. These constructs were all expressed from pOPINS3C vectors which include an N-terminal cleavable SUMO solubility tag, as was used by Zhang et al, 2017[71]. Further, Zhang et al, 2017[71] used BL21(DE3) RIL cells to express RRS1 WRKY$_{E1195-T1273}$, which are similar to Rosetta™ 2 (DE3) cells in that they carry extra copies of rare *E. coli* tRNAs that can limit translation of heterologous proteins in

E. coli. As such, I trialled the expression of RRS1 WRKY Dom6 construct variants in Rosetta™ 2 (DE3), BL21(DE3) and SHuffle® T7 cell lines. All cell lines were grown in LB media and protein expression induced with 0.4 mM IPTG. Soluble expression of RRS1 $WRKY_{E1195-T1273}$ and RRS1 WRKY $Dom6S_{E1195-C1290}$ was found to be highest in Rosetta™ 2 (DE3) in small-scale expression trials, **Figure 5.3B** (data only shown for Rosetta™ 2 (DE3) expression). However, RRS1 WRKY $Dom6R_{E1195-Y1373}$ did not show soluble expression in any condition tested. Expression of RRS1 WRKY Dom6R in the crude lysate (CL) extract can be seen in **Figure 5.3B,** demonstrating that expression problems with this construct were due to production of soluble folded protein, not translation efficiency. The reasons for this are, as yet, undetermined but we hypothesise that the predicted highly disordered nature of the Dom6R extension beyond the Dom6S boundary may impeded solubility of this construct (secondary structure and disorder prediction conducted with Phyre2 modelling[164] and RONN[162].)
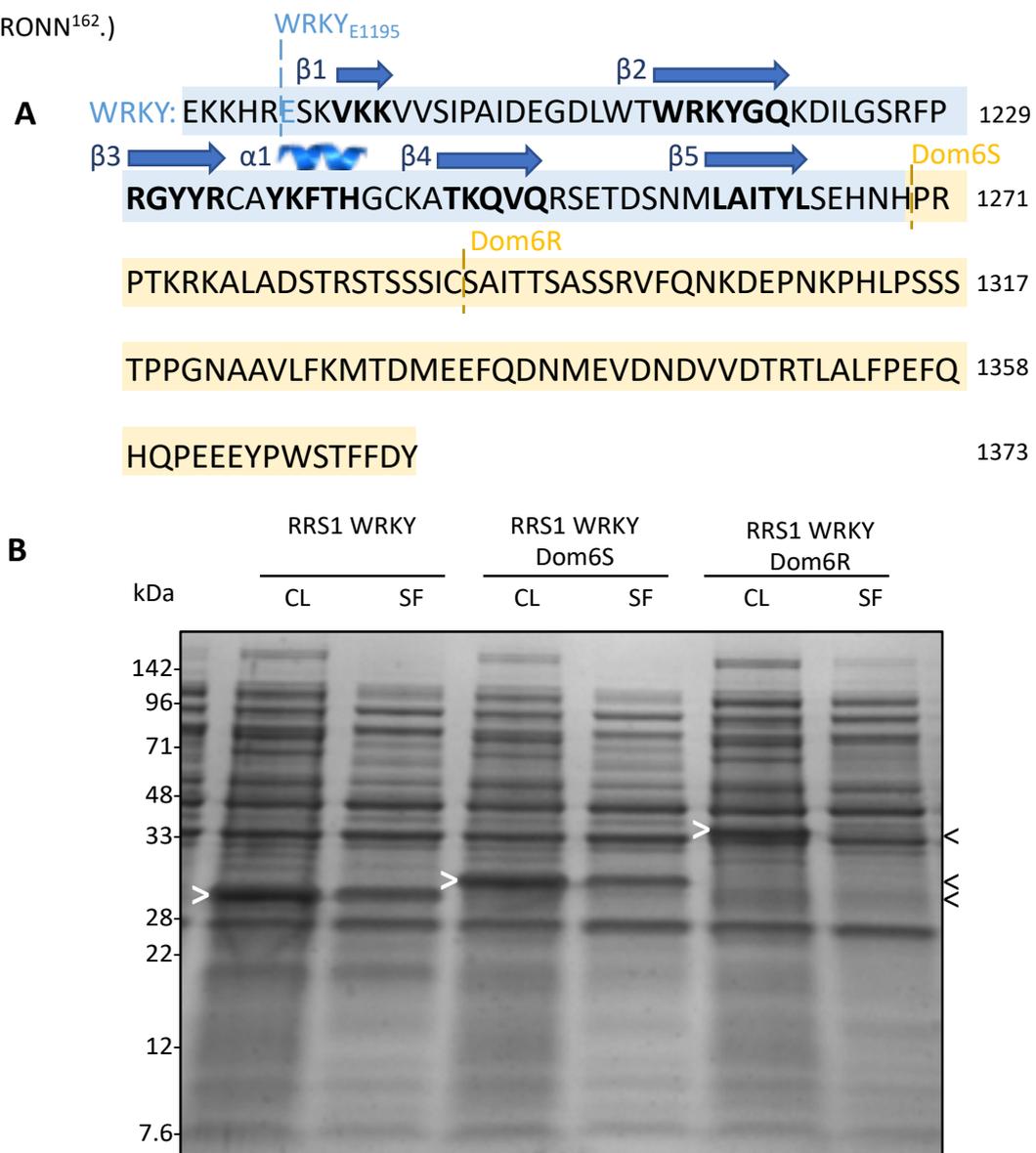
Figure 5.3 Small-scale expression testing RRS1 WRKY Dom6 constructs. (A) Protein amino acid

(cont) sequence for RRS1 WRKY Dom6, WRKY domain is highlighted in blue whilst Dom6 is shown in yellow highlighting the domain boundaries for Dom6S and Dom6R extension. The secondary structure features of RRS1 WRKY are depicted above the sequence numbered β strand1-5 and α helix 1 as defined in crystal structure of RRS1 WRKY (PDB ID 5W3X)[71]. N-terminal expression boundary E1195 is highlighted prior to the start of β strand 1. (B) Small-scale expression trial results for expressing RRS1 WRKY$_{E1195-T1273}$, WRKY Dom6S$_{E1195-C1290}$ and WRKY Dom6R$_{E1195-Y1373}$ from pOPINS3C vector in Rosetta 2 (DE3) cells. Protein samples were run on 16% SDS-PAGE gel and stained using Coomassie dye. White arrows indicate protein bands of correct predicted MW size. CL: crude lysate, SF: Soluble fraction.

Having identified conditions for soluble expression of RRS1 WRKY Dom6S$_{E1195-C1290}$ (expression in pOPINS3C expressed in Rosetta™ 2 (DE3) grown in LB media at 37^{0}C prior to induction with 0.4 mM IPTG and overnight growth at 20^{0}C), purification was scaled up to 6-8L cultures as per **Table 2.7**. Following cleavage of the N-terminal 6xHis-SUMO tag, the protein underwent a second round of gel filtration using a Superdex™ 75 26/600 column from which fractions containing pure RRS1 WRKY Dom6S$_{E1195-C1290}$, as judged by SDS-PAGE, were pooled and concentrated for downstream experiments, **Figure 5.4.**
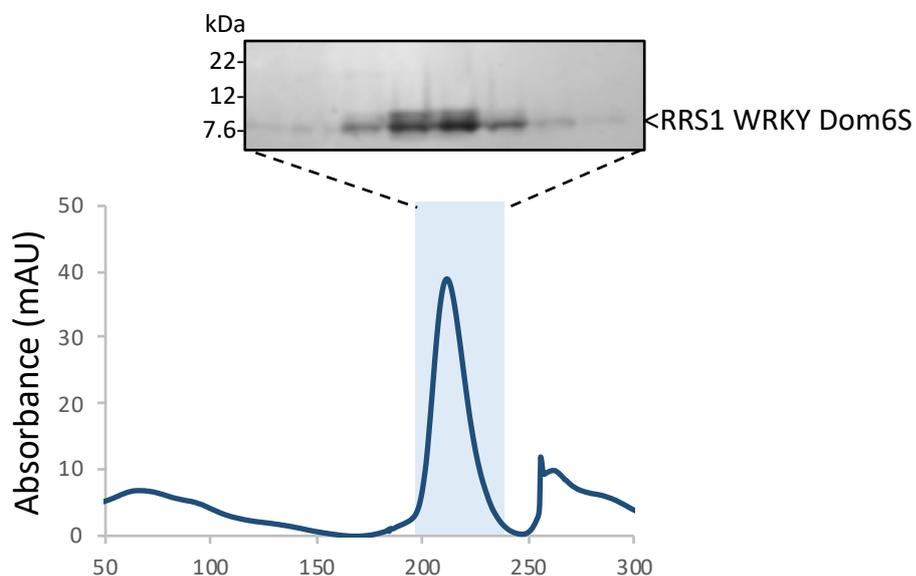
Figure 5.4 Purification of RRS1 WRKY Dom6S (E1195-C1290) in pOPINS3C. Gel filtration trace of RRS1 WRKY Dom6S eluting from a Superdex 75 26/600 gel filtration column peaking at 213 ml following 3C protease cleavage of SUMO tag, accompanying SDS-PAGE gel above trace shows fractions from peak highlighted in blue. Experiments were repeated twice with similar results found across repeats.

To conduct experiments examining the interactions between RRS1 WRKY Dom6S and AvrRps4, AvrRps4$_c$ protein was produced as per **Table 2.7**. This construct of AvrRps4, G134-

Q221, was previously used to obtain the crystal structure of AvrRps4$_c$ (PDB ID 4B6X)[114]. AvrRps4$_c$ (G134-Q221) was expressed from pOPINF in SHuffle® T7 cells grown at 37^{0}C prior to induction with 1 mM IPTG and overnight growth at 18^{0}C. Following cleavage of the N-terminal 6xHis affinity tag, the protein underwent a second round of gel filtration using a Superdex™ 75 26/600 column from which fractions containing pure AvrRps4$_c$, as judged by SDS-PAGE, were pooled and concentrated for downstream experiments, **Figure 5.5A**. Following purification of AvrRps4$_c$, the presence and integrity of the protein was confirmed by intact mass spectrometry which indicated no protein degradation had occurred, **Figure 5.5B**. The expected molecular mass for this protein (10083.19 Da) matched the peak from mass spectrometry of 10083.5215 Da.
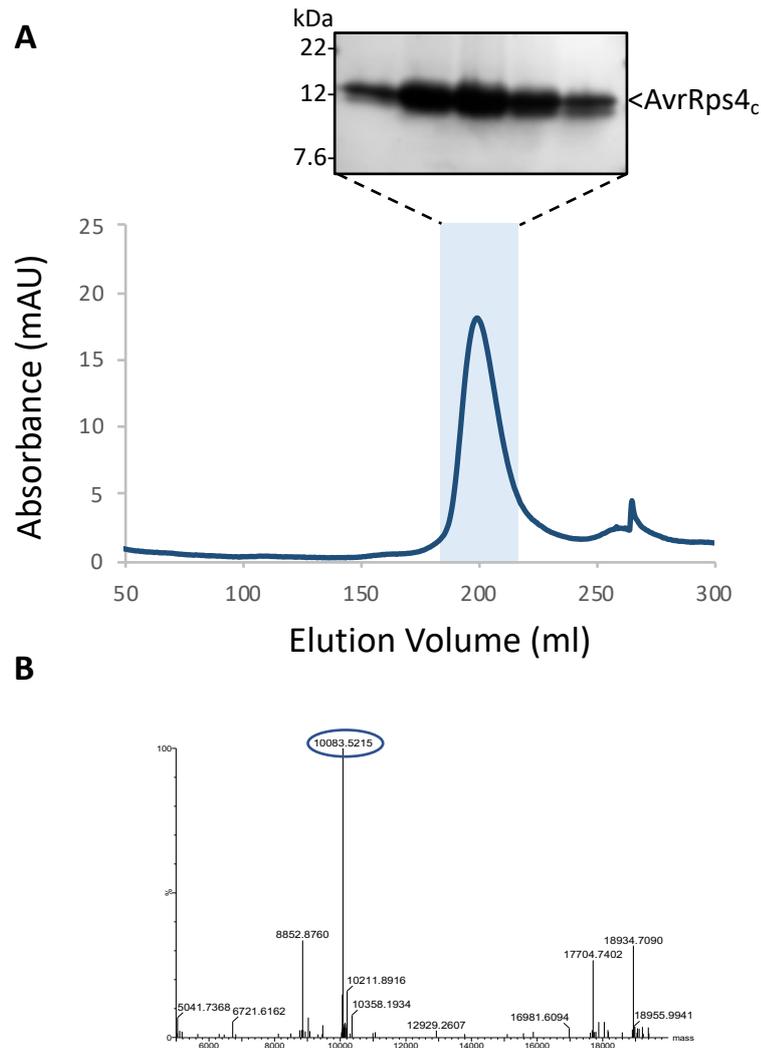
Figure 5.5 Purification of AvrRps4$_c$ (G134-Q221) in pOPINF. (A) Gel filtration trace of AvrRps4$_c$ eluting from a Superdex 75 26/600 gel filtration column peaking at 201 ml following 3C protease cleavage of SUMO tag, accompanying SDS-PAGE gel above trace shows fractions from peak highlighted in blue. **(B)** Intact mass spectrometry analysis of purified AvrRps4$_c$. Spectra show different species identified by peaks in intact mass spectrometry analysis, peak of interest for AvrRps4$_c$ is circled (10083.52150 Da), exact mass (in Daltons) is labelled above each peak. Experiments, except for mass spectrometry, were repeated twice with similar results found across repeats.

5.2.2   RRS1 WRKY Dom6S and AvrRps4$_c$ form a complex in vitro

Having purified WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ protein, I wanted to determine whether these proteins could form a complex in vitro. This demonstrated that the WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ constructs I was working with in vitro covered the minimum viable region required to facilitate the interaction between these proteins observed in vivo.

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

For this, I first used the qualitative method of analytical gel filtration. Analytical gel filtration is an approach used to monitor protein elution from a size exclusion chromatography column by absorbance at 280 nm. Elution volume is dependent on the size of the protein or complex, with those of a larger size eluting earlier from the column. Therefore, if two proteins form a complex when mixed they will elute earlier from the column, observed as a peak shift to the left of the trace, then if the proteins are run individually. For this study, RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ were first run individually on a Superdex™ 75 10/300 GL column. To test for complex formation, RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ were mixed in a 1:1 molar ratio and incubated on ice for 1 hour. The resulting trace from analytical gel filtration of the mixed proteins shows a clear shift to a single earlier elution peak at 11.8ml compared to 12.9 ml and 14.2 ml for AvrRps4$_c$ and RRS1 WRKY Dom6S$_{E1195-C1290}$ respectively, **Figure 5.6**. Presence of proteins of interest were confirmed by SDS-PAGE analysis, **Figure 5.6**.

Protein-protein interaction modelling conducted by Dr Lennart Wirthmüller and Dr Yan Ma of the interaction between RRS1 WRKY and AvrRps4$_c$ highlighted two AvrRps4 glutamic acid residues, E175 and E187, as having a key role in facilitating AvrRps4's interaction with RRS1 WRKY, **Figure 5.7**. To help validate this interaction model, I generated single and double mutant AvrRps4$_c$ E187A and E187A/E175A proteins that were expressed and purified using the same conditions as wild type AvrRps4$_c$. Interestingly, a peak shift is not observed on analytical gel filtration when RRS1 WRKY Dom6S$_{E1195-C1290}$ is mixed with double mutant AvrRps4$_c$ E187A/E175A, or single mutant AvrRps4$_c$ E187A, indicating the importance of these residues in the RRS1 WRKY-AvrRps4 interface, **Figure 5.6**. When RRS1 WRKY Dom6S$_{E1195-C1290}$ is mixed with either E187A or double mutant E187A/E175A AvrRps4$_c$, the two proteins elute in separate peak indicating a complex is not being formed, note that AvrRps4$_c$ has a very low extinction coefficient and therefore absorbs poorly at 280 nM. This result supports both the protein-protein interaction model seen in **Figure 5.7** and *in planta* work which demonstrates the inability of AvrRps4$_c$ E187A or E175A mutants to induce RRS1/RPS4 dependent HR and reduced binding to RRS1 WRKY Dom6R in coimmunoprecipitation assays[153].
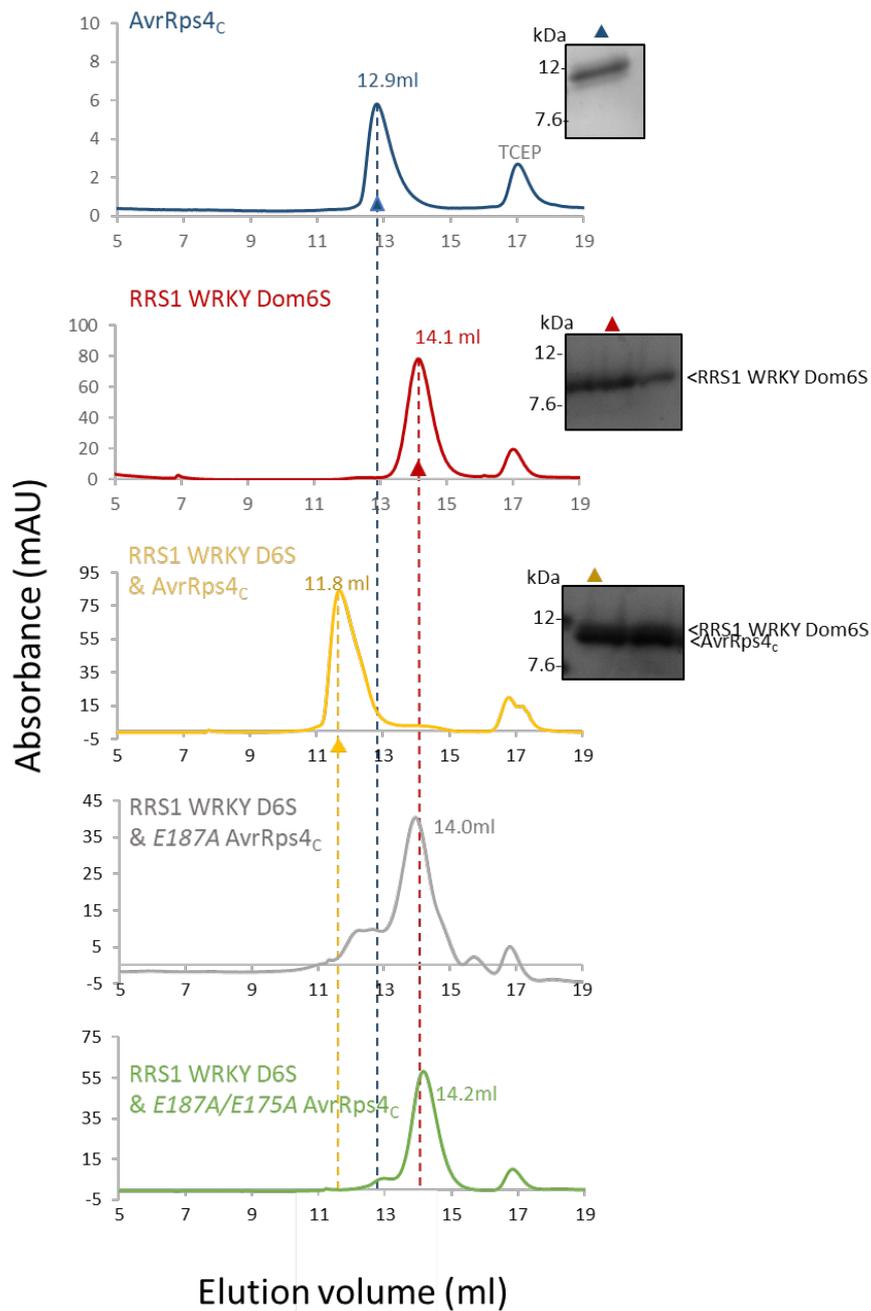
Figure 5.6 RRS1 WRKY Dom6S and AvrRps4c form a complex in vitro (A) Analytical gel filtration traces showing elution volume of AvrRps4c (blue), RRS1 WRKY Dom6S$_{E1195-C1290}$ (red), RRS1 WRKYDom6S$_{E1195-C1290}$-AvrRps4c complex (yellow), mixing of RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4c E187A (grey) and mixing of RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4c E187A/E175A (green). SDS-PAGE analysis shows the fractions taken from volumes on trace indicated by coloured triangles. Peak volumes shown indicate elution volume. Experiments were repeated twice with similar results found across repeats.
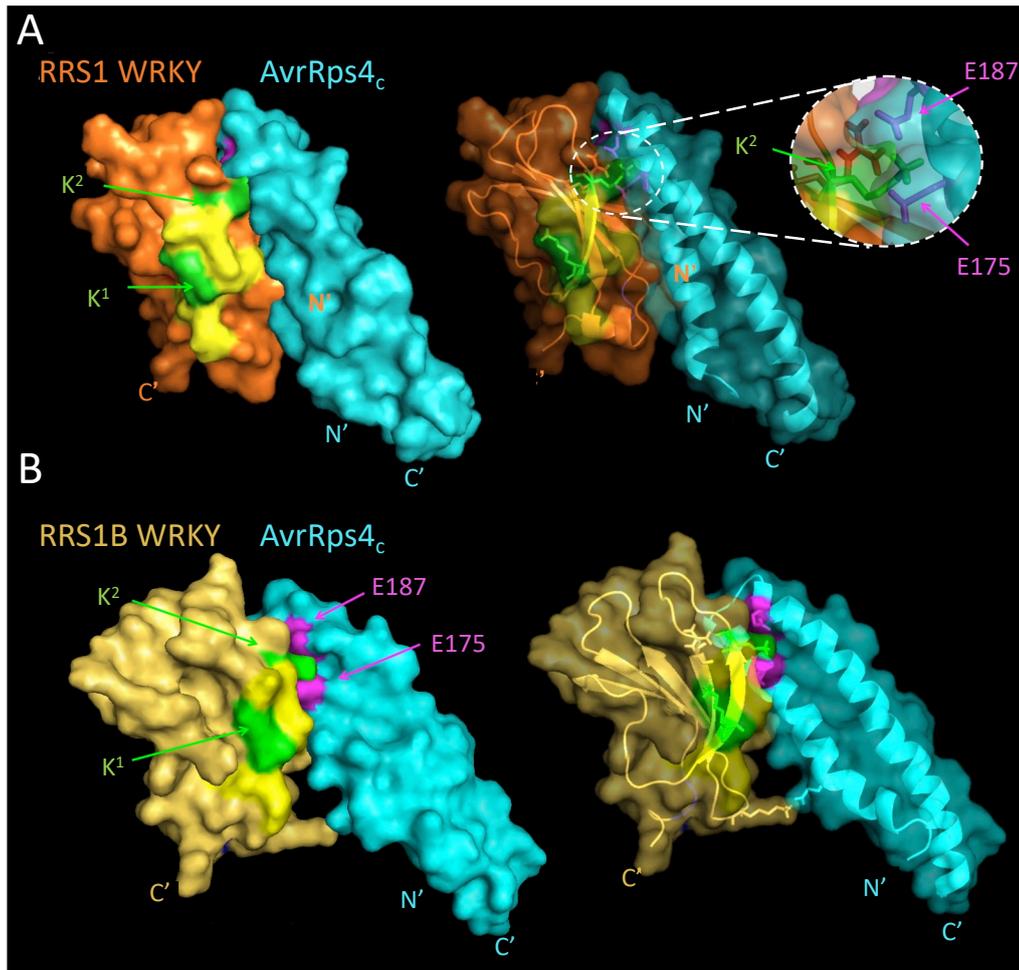
Figure 5.7 Structural modelling of the possible interaction surface between AvrRps4c (c-terminus) and the WRKY domains of RRS1-R and RRS1B. **(A)** Precited protein-protein interaction between RRS1 WRKY$_{E1190-H1269}$ and AvrRps4$_c$ generated using ClusPro 2.0[265] with RRS1 WRKY$_{E1190-H1269}$ shown in orange and AvrRps4$_c$ in cyan. **(B)** Predicted interaction of RRS1B WRKY$_{G1166-T1241}$ shown in yellow and AvrRps4$_c$. The structures of RRS1-R and RRS1B WRKY domain were predicted with 100% confidence using Phyre2[164] using the template of the crustal structure of WRKY1 (PDB ID:c2aydA[71]). AvrRps4c (PDB ID:4B6X[166]) is shown in cyan. The sequence identity between RRS1 WRKY and RRS1B WRKY and the template WRKY1 is 41% and 48% respectively. The solid molecular surface of each complex is presented on the left whilst the secondary structure ribbons are shown on the right. Amino acids predicted at the interface are shown as sticks, with lysine (K_1 and K_2) residues of RRS1-R or RRS1B WRKY motif highlighted in green, and glutamic acid residues (E187 and E175) of AvrRps4 highlighted in purple. Modelling was conducted by Dr Lennart Wirthmueller and Dr Yan Ma, Figure adapted from Ma et al, 2016[266].

5.2.3 Quantitative binding analysis of RRS1 WRKY Dom6S and AvrRps4$_c$ in vitro

To further investigate the interaction between RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$, and to gain quantitative insights on this binding event, I used surface plasmon resonance (SPR). This technique allows for quantitative analysis of biomolecular interactions, e.g. protein-protein[240]. SPR measures changes in the refractive index on the surface of a sensor chip upon which one binding partner has been immobilized, referred to as the ligand, when the second binding partner flows over the chip surface in the analyte. Should the two binding partners interact, the mass bound to the surface of the chip increases and changes the refractive index of a polarised light beam directed towards the sensor chip surface. In this way you can gain quantitative real-time information on a protein-protein binding event[240].

For the purposes of this study, AvrRps4$_c$ was used as the ligand immobilised on the surface of the sensor chip. This was enabled by expressing AvrRps4$_c$ with a non-cleavable C-terminal Hisx6 tag (from the pOPINE vector) allowing the protein to bind to a Ni-NTA sensor chip surface. Alongside wild-type AvrRps4$_c$, single mutant AvrRps4$_c$ E187A and double mutant E187A/E175A were also used as negative binding controls. After immobilizing AvrRps4$_c$ to the surface of the sensor chip, RRS1 WRKY Dom6S$_{E1195-C1290}$ protein was injected over the chip surface as the analyte. In SPR, refractive differences caused by binding between the analyte and ligand is shown in a change of response units (RU). The maximum binding capacity of analyte to the ligand chip, assuming all ligand is active and binding sites are available, is referred to as the R$_{max}$. The theoretical R$_{max}$ can be calculated using the formula below and reflects the theoretical maximum amount of analyte bound to a given amount of immobilized ligand on a chip surface. The R$_{max}$ is dependent on the molecular weight (MW) of the analyte and ligand, the amount of ligand immobilized on the surface of the sensor chip (R$_{Ligand}$) and stoichiometry of the interaction between ligand and analyte.

$$R_{max} = \frac{\text{MW of analyte (RRS1 WRKY)}}{\text{MW of Ligand (AvrRps4}_c)} \text{ X } R_{Ligand} \text{ X Stoichiometry}$$

Once the theoretical R$_{max}$ has been calculated for a given interaction it can be compared with the experimental data RU for binding between the ligand and analyte which is then expressed as a percentage of the theoretical R$_{max}$ value (% R$_{max}$). This value provides quantitative data on the binding affinity of the ligand and analyte.

Initial SPR experiments with RRS1 WRKY Dom6S$_{E1195-C1290}$ showed issues with non-specific binding of RRS1 WRKY Dom6S$_{E1195-C1290}$ to the surface of the chip. Extensive optimisation work required varying both the temperature at which the SPR analysis was conducted from 4-25^{0}C, the concentration of NaCl in the SPR running buffer conditions, as well as the flow rate of analyte before conditions were identified which were suitable for conducting the binding experiments. It was observed that conducting the experiment at lower temperatures (4^{0}c) slowed the dissociation rate of the RRS1 WRKY Dom6S$_{E1195-C1290}$ − AvrRps4$_c$ complex improving the quality of data collected, **Figure 5.8.** The final conditions included conducting the analysis at 8^{0}C in running buffer containing 860 mM NaCl and an analyte flow rate of 30 μl/min for 120 seconds contact time and 60 seconds dissociation time. These conditions were used for all subsequent SPR work with these proteins.
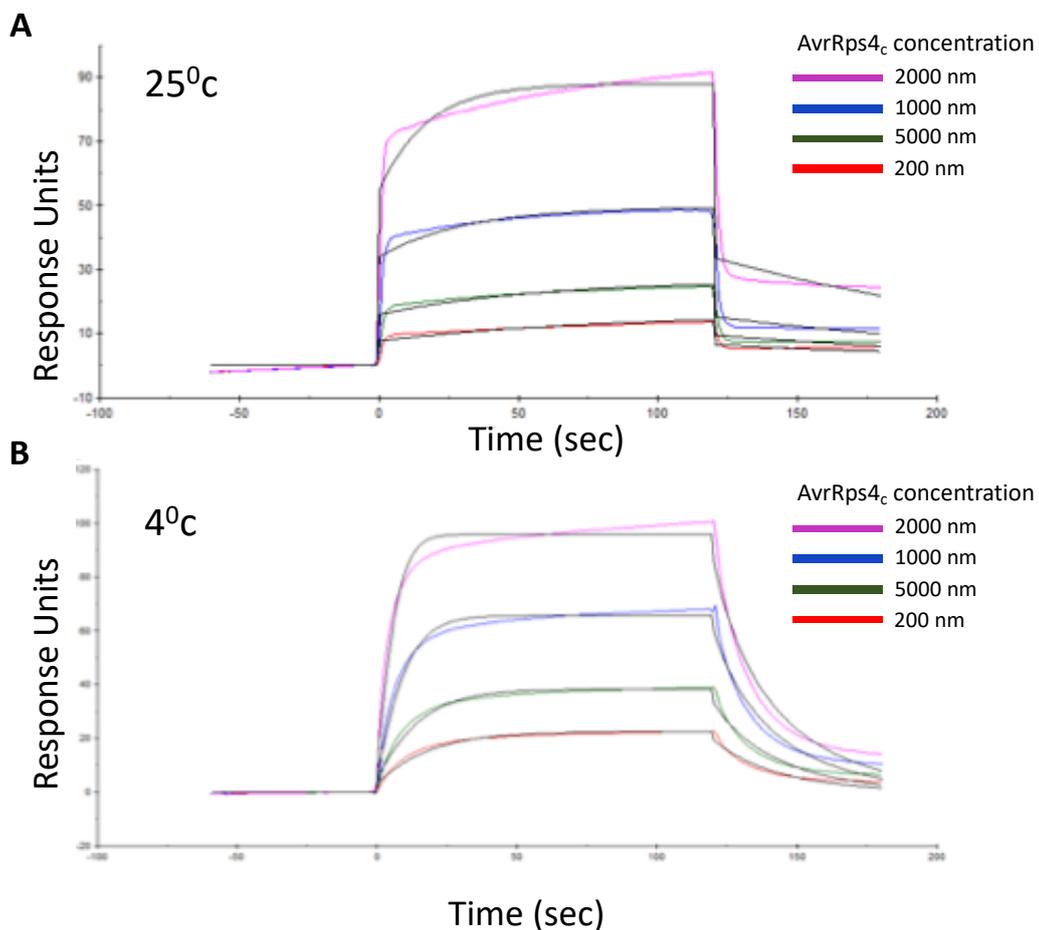
Figure 5.8 Lower temperature slowed dissociation of RRS1 WRKY Dom6SE1195-C1290 and AvrRps4c. SPR experiment conducted at **(A)** 25^{0}C and **(B)** 4^{0}C. Slow rates of dissociation observed at lower experimental condition when all other factors were kept constant AvrRps4c concentrations shown in key.

To determine the best concentrations of analyte to collect R_{max} data with, I conducted an initial experiment with a large range of RRS1 WRKY Dom6S$_{E1195-C1290}$ concentrations from 50-3000 nM, **Figure 5.9A**. From this data I decided to conduct R_{max} experiments at concentrations of 500 nM and 3000 nM RRS1 WRKY Dom6S$_{E1195-C1290}$, as the %R_{max} values for these analyte concentrations fell in the desired ranges of ~80% and ~20% respectively. R_{max} experiments were then carried out at each of these concentrations with three binding level replicates taken for each analyte concentration with each AvrRps4$_c$ effector. The experiment was repeated twice with results from the two data sets combined to produce the %R_{max} data shown in **Figure 5.9B**. The %R_{max} plot showed a clear increased affinity of RRS1 WRKY Dom6S$_{E1195-C1290}$ for wildtype AvrRps4$_c$ compared to E187A and E187A/E175A mutants, indicating the importance of these glutamic acid residues in binding RRS1.

In addition to representing ligand and analyte binding as % R_{max}, SPR can also give steady state affinity information on a binding event. One such parameter is the equilibrium dissociation constant, K_D, which reflects the point at which the rate of complex dissociation equals the rate of association. This is calculated using a titration of analyte concentrations and fitting the experimental data responses to a steady-state affinity model. I looked to gain steady state kinetic information on the RRS1 WRKY Dom6S$_{E1195-C1290}$ -AvrRps4$_c$ binding event. However, I was unable to determine K_D values for this interaction as, despite optimisation work, the data failed quality tests used by the Biacore software when fitting to steady-state affinity models. The issues leading to this failure were predominantly a high bulk effect and a very fast-on fast-off binding event, which seems to occur between RRRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$.

Figure 5.9 AvrRps4$_c$ residues E187 and E175 play important role in RRS1 WRKY Dom6S$_{E1195-C1290}$ binding of AvrRps4$_c$ (A) To identify RRS1 WRKY Dom6S$_{E1195-C1290}$ concentrations to conduct %Rmax analysis at, initial binding of RRS1 WRKY Dom6S$_{E1195-C1290}$ at concentrations from 0-3000 nM to AvrRps4$_c$ WT, E187A and E187A/E175A was investigated. **(B)** %Rmax of RRS1 WRKY Dom6S$_{E1195-C1290}$ bound to wildtype and mutant AvrRps4$_c$ assuming a 1:1 binding model for RS1 WRKY Dom6S: AvrRps4. The left and right panel were conducted with 500 nM and 3000 nM RRS1 WRKY Dom6S$_{E1195-C1290}$ respectively. The median is represented in the centre line of the box, and the upper and lower quartile by the box limits. Whiskers extend from smallest value in Q_1 - 1.5x interquartile range (IQR) (cont) and the largest value within Q_3 + 1.5x IQR. Individual data points from each biological repeat

are represented by a black square or circle, repeats are seen to cluster with values from alike repeats. This is likely due to differences in concentration of AvrRps4 protein used in each repeat as due to a lack of aromatic residues AvrRps4 protein concentration is hard to measure. Alternatively, the clustering might reflect differences in loading of RRS1 WRKY protein on chip between repeats which though kept in a set range did differ between repeats. Plots were produced using the ggplot2 package[171] in R.

5.2.4 Purifying/understanding the structural basis of RRS1 WRKY Dom6S-AvrRps4$_c$ complex

In addition to gaining biochemical insights into the association between RRS1 WRKY Dom6S and AvrRps4$_c$, a key development in understanding this interaction would be in attaining an atomic structure of this protein complex. Understanding the structural basis of NLR-effector interactions is a key goal in the field of NLR biology, and presents a major bottleneck in efforts towards engineering NLRs with expanded pathogen recognition capabilities. To understand the structural basis of RRS1 WRKY Dom6S recognition of AvrRps4, a diverse range of crystallisation screens were setup with purified RRS1 WRKY Dom6S$_{E1195-C1290}$ -AvrRps4$_c$ complex.

RRS1 WRKY Dom6S$_{E1195-C1290}$ -AvrRps4$_c$ protein complex was produced using purification 'method 2' described in 2.6.3. This was due to the fact that the growth of cultures expressing RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ required different induction levels of IPTG induction for high yielding protein expression, 0.4 mM and 1 mM respectively. As such cultures expressing RRS1 WRKY Dom6S$_{E1195-C1290}$ in pOPINS3C and cultures expressing untagged AvrRps4$_c$ in pOPINA were grown separately, pelleted and resuspended in lysis buffer. Following thawing, the lysate of the two cultures were mixed prior to sonication. The rationale for this strategy is that during IMAC purification the His-SUMO tagged RRS1 WRKY Dom6S$_{E1195-C1290}$ would interact and co-elute with untagged-AvrRps4$_c$ if the proteins were forming a complex. Following IMAC purification and gel filtration, the Hisx6-SUMO tag was cleaved from the RRS1 WRKY Dom6S$_{E1195-C1290}$ protein using 3C protease and removed by IMAC purification. The protein complex then underwent a final round of purification by gel filtration, an example trace of which can be seen in **Figure 5.10A.** The presence of both proteins in the peak trace was confirmed by SDS-PAGE. Following concentration of protein and prior to crystallisation screen setup, the presence and integrity of RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ was confirmed by mass spectrometry analysis, **Figure 5. 10 B**, where the predicted molecular weights for each protein, 11137.65 Da and 9929.0 Da

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

respectively, were observed to match the mass spectrometry spectra peaks of 11137.75 Da
and 9929.20 Da indicating the proteins had not degraded during purification.

RRS1 WRKY Dom6S$_{E1195-C1290}$-AvrRps4$_c$ complex was then used in a large range of
crystallisation screens designed to test diverse regions of crystallisation space to identify
conditions which may support crystal formation. 96-well commercial screens were setup in
sitting-drop diffusion experiments with two wells setup for each buffer condition with X
and ½X protein concentration.
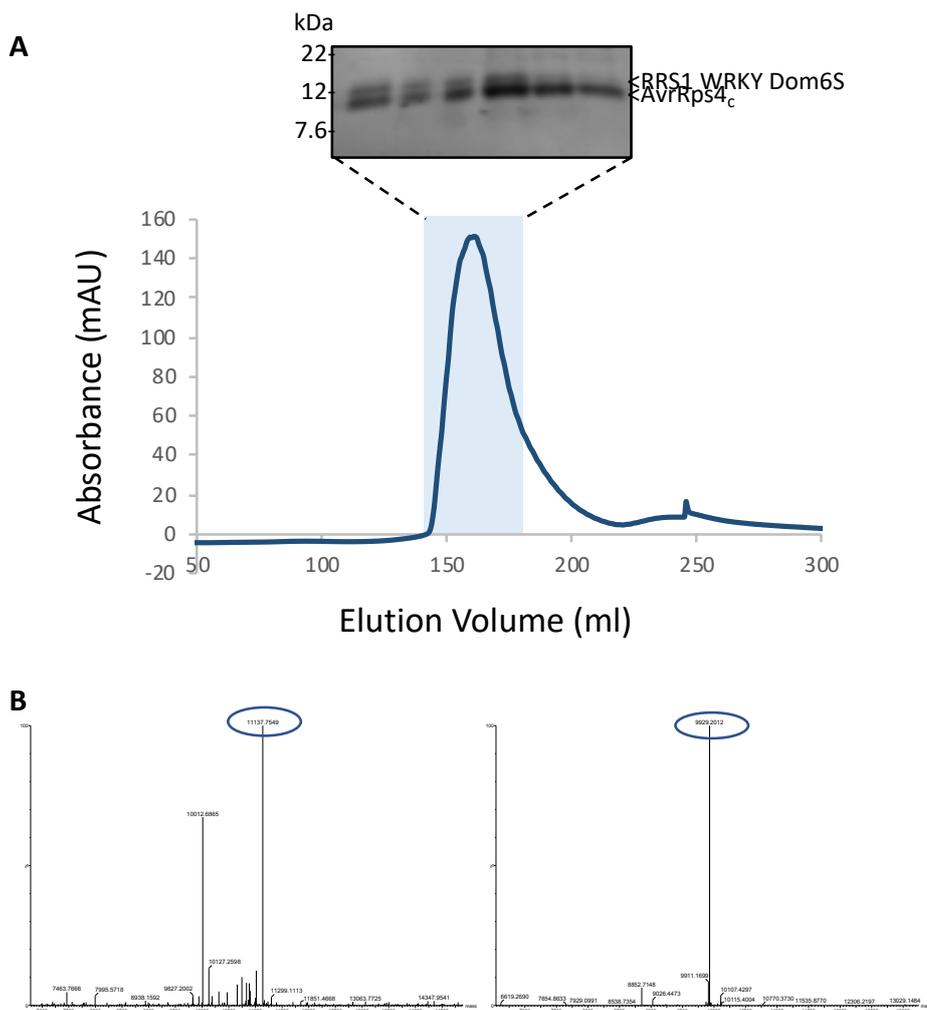
Figure 5.10 Purification of RRS1 WRKY Dom6S (E1195-C1290) and AvrRps4c (G134-Q221) complex.
(A) Gel filtration trace of RRS1 WRKY Dom6S$_{E1196-C1290}$ –AvrRps4$_c$ complex eluting from a Superdex 75
26/600 gel filtration column peaking at 162 ml following 3C protease cleavage of RRS1 WRKY
Dom6S$_{E1196-C1290}$ Hisx6-SUMO tag, accompanying SDS-PAGE gel above trace shows fractions from
peak highlighted in blue. **(B)** Intact mass spectrometry analysis of purified RRS1 WRKY Dom6S –
AvrRps4$_c$ complex. Spectra show different species identified by peaks in intact mass spectrometry
analysis, peak of interest for RRS1 WRKY Dom6S$_{E1196-C1290}$ (11137.7549 Da) and AvrRps4$_c$ is circled
(9929.2012 Da), exact mass (in Daltons) is labelled above each peak. Experiments, except for mass
spectrometry, were repeated 3 times with similar results found across repeats.

Screens with RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ complex were setup at ~12 mg/ml and 6 mg/ml with protein concentrations measured prior to each screen. Screens were setup at 20^{0}C in six commercial screens; JCSG-plus™, Morpheus®, MIDAS™, ProPlex™, PACT premier™ and Structure (all Molecular-Dimensions). Additionally, the complex was also screened in a custom screen designed by Clare Stevenson and Dave Lawson at JIC crystallography platform known as the KISS screen. However, none of these screens highlighted any conditions in which crystal formation of the complex was supported. There are many parameters which can affect the nucleation and growth of protein crystals, such as the composition and pH of buffer reservoir solution, protein concentration and temperature. The parameters in which protein crystallisation will occur are not predictable. As such, in addition to using multiple sparse matrix crystallization screens, which vary in their buffer composition, precipitant mix and pH, I wanted to investigate the effect of temperature on crystalizing this complex. As SPR R$_{max}$ experiments suggested that the RRS1 WRKY Dom6S$_{E1195-C1290}$ -AvrRps4 complex appeared to be more stable at lower temperatures of 4-8^{0}C, I hypothesised that lower temperatures may aid crystallisation of the complex. Therefore, I setup several commercial 96-well sitting drop screens at 4^{0}C to see if this would support crystallization. Screens setup at this temperature were at a protein complex concentration of 12 mg/ml and 6 mg/ml. The screens trialled at 4^{0}C were JCSG-plus™, Morpheus®, PACT premier™ and ProPlex™. At 4^{0}C protein crystals were observed in well C5 of JCSG-plus™ using protein at a measured concentration of 12 mg/ml, **Figure 5.11**. The conditions for this well were 0.8 M Sodium phosphate monobasic monohydrate, 0.8 M Potassium phosphate monobasic, 0.1 M Sodium HEPES, pH 7.5 with no Precipitant. Whilst these crystals were not of a high enough quality to collect X-ray diffraction data, work is ongoing to optimise these conditions to obtain crystals suitable for X-diffraction data collection.
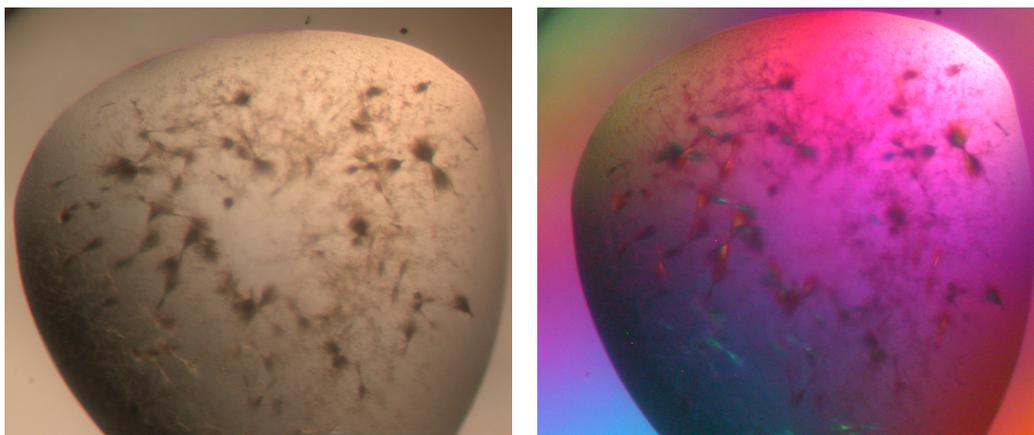
Figure 5.11 Images of crystals seen during RRS1 WRKY Dom6S$_{E1195-C1290}$ -AvrRps4c trials.
Crystals observed in well C5 of JCSG-plus™ screen with 12 mg/ml protein at 4^{0}C, buffer
conditions 0.8 M Sodium phosphate monobasic monohydrate, 0.8 M Potassium phosphate
monobasic, 0.1 M Sodium HEPES, pH 7.5 with no Precipitant. Left image shows crystals in
brightfield light and right in polarised light.

I hypothesised that issues with co-crystallising the RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$
complex may be due to dissociation of the complex during crystallisation, as observed with
RRS1 WRKY$_{E1195-T1273}$ and AvrRps4$_c$ in **Figure 5.16**. To aid the stability of the RRS1 WRKY
Dom6S$_{E1195-C1290}$ -AvrRps4$_c$ complex, I decided to try expressing a construct of RRS1 WRKY
Dom6S which was linked with a short peptide linker to AvrRps4$_c$. Similar methods have
been previously employed in gaining structures of protein complexes, for example the RRS1
TIR-RPS4 TIR complex structure in which the two TIR domains were linked with a 5 residue
'GSGGS' linker. Co-crystallising proteins using a linker comes with the caveat that linking
the two proteins may impose artificial interactions or steric hinderance between the two
proteins. In order to minimise such artificial interactions, I made use of the 6 amino acid
residue 'NAAIRS' linker sequence. This sequence is described as having a neutral structure
and can adopt both β-strand and α-helical secondary structures[241], therefore minimising
the formation of any forced secondary structure between two proteins. The protein-protein
interaction model in **Figure 5.7** was used to choose which protein termini to link, choosing
the two termini which, in the model, were the closest. For the purposes of this study, I
decided to link the C-terminal end of RRS1 WRKY Dom6S$_{S1184-C1290}$, a construct used in the
lab prior to the switch to RRS1 WRKY Dom6S$_{E1195-C1290}$, to the N-terminal end of AvrRps4$_c$
using a 'GSNAAIRSNAAIRSGS' sequence. Using PCR primers with extensions to introduce
the NAAIRS linker sequence, a Golden Gate Level 0 pICSL01005 construct with RRS1 WRKY
Dom6S$_{S1184-C1290}$ -GSNAAIRSNAAIRSGS- AvrRps4$_c$ was generated. This was then used as a PCR
template to clone RRS1 WRKY Dom6S$_{S1184-C1290}$ -GSNAAIRSNAAIRSGS- AvrRps4$_c$ into pOPINE

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

with a 3C protease cleavage site added to the C-terminus of the construct as a primer extension, **Figure 5.12**. This construct was expression tested in Rosetta™ 2 (DE3), BL21(DE3) and SHuffle® T7 cell lines but proved insoluble expression in all lines.

Figure 5.12 Linked construct of RRS1 WRKY Dom6S and AvrRps4. Protein has a C-terminal 3C protease cleavable 6xHis tag. Amino acid residue number are shown above protein.

5.3   RRS1 WRKY directly interacts with AvrRps4$_c$ in vitro

I hypothesised that the lack of success in crystallising RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ complex may be due in part to the predicted unstructured secondary structure of RRS1's Dom6S (secondary structure and disorder prediction conducted using Phyre2[164] and RONN software[162]) which can impede crystallization of a protein. As such, I looked to expand my work to focus on AvrRps4 interaction with the RRS1 WRKY domain without Dom6. As pervious *in planta* work has shown Dom6S of RRS1 is not required for recognition of AvrRps4[152], all functionally important interactions between AvrRps4 and RRS1 WRKY must be facilitated without Dom6S. Therefore, information on RRS1 WRKY-AvrRps4 interaction is highly relevant for informing mechanistic studies of this NLR.

5.3.1   Purifying RRS1 WRKY and AvrRps4$_c$

To examine the interaction of RRS1 WRKY with AvrRps4, I used a construct with the same boundaries utilised by Zhang et al, 2017[71] which expressed RRS1-R E1195-T1273. Having established this construct expressed well in pOPINS3C in Rosetta™ 2 (DE3) cells, **Figure 5.3B**, purification was scaled up to 6-8 L cultures. Following cleavage of 6xHis-SUMO affinity tag with 3C protease, **Figure 5.13B**, RRS1 WRKY$_{E1195-T1273}$ protein was reloaded on a Superdex™ 75 16/600 column for a final round of gel filtration purification. The presence and purity of the protein in each peak fraction following gel filtration was confirmed by SDS-PAGE, **Figure 5.13A**. The presence, and integrity, of final concentrated protein was confirmed by mass spectrometry analysis prior to downstream experiments which found the predicted molecular weight for this protein, 9344.62 Da matched the mass spectrometry spectra peak of 9344.79 Da. AvrRps4$_c$ protein used in interaction studies with RRS1 WRKY$_{E1195-T1273}$ was produced as described in 5.2.1.

The complex of RRS1 WRKY$_{E1195-T1273}$ and AvrRps4$_c$ for crystallisation trials was produced using 'method 2' described in 2.6.3 and 5.2.4. Due to differences in IPTG induction concentration required for these two proteins to produce high yields, 0.4 mM and 1 mM respectively, cultures growing 6xHis-SUMO tagged RRS1 WRKY$_{E1195-T1273}$ expressed in pOPINS3C and untagged AvrRps4$_c$ expressed in pOPINA were grown separately with cell lysates mixed prior to sonication. Untagged AvrRps4$_c$ in complex with RRS1 WRKY$_{E1195-T1273}$, co-eluted with 6xHis-SUMO tagged RRS1 WRKY$_{E1195-T1273}$ during IMAC purification. Following gel filtration, 6xHis-SUMO tag was cleaved from RRS1 WRKY and removed by IMAC purification, **Figure 5.14**. The RRS1 WRKY$_{E1195-T1273}$ -AvrRps4$_c$ complex then underwent a second gel filtration down a Superdex™ 75 16/600 column. The presence of both proteins under the gel filtration trace peak was confirmed by SDS-PAGE, **Figure 5.14**.

A

B

C
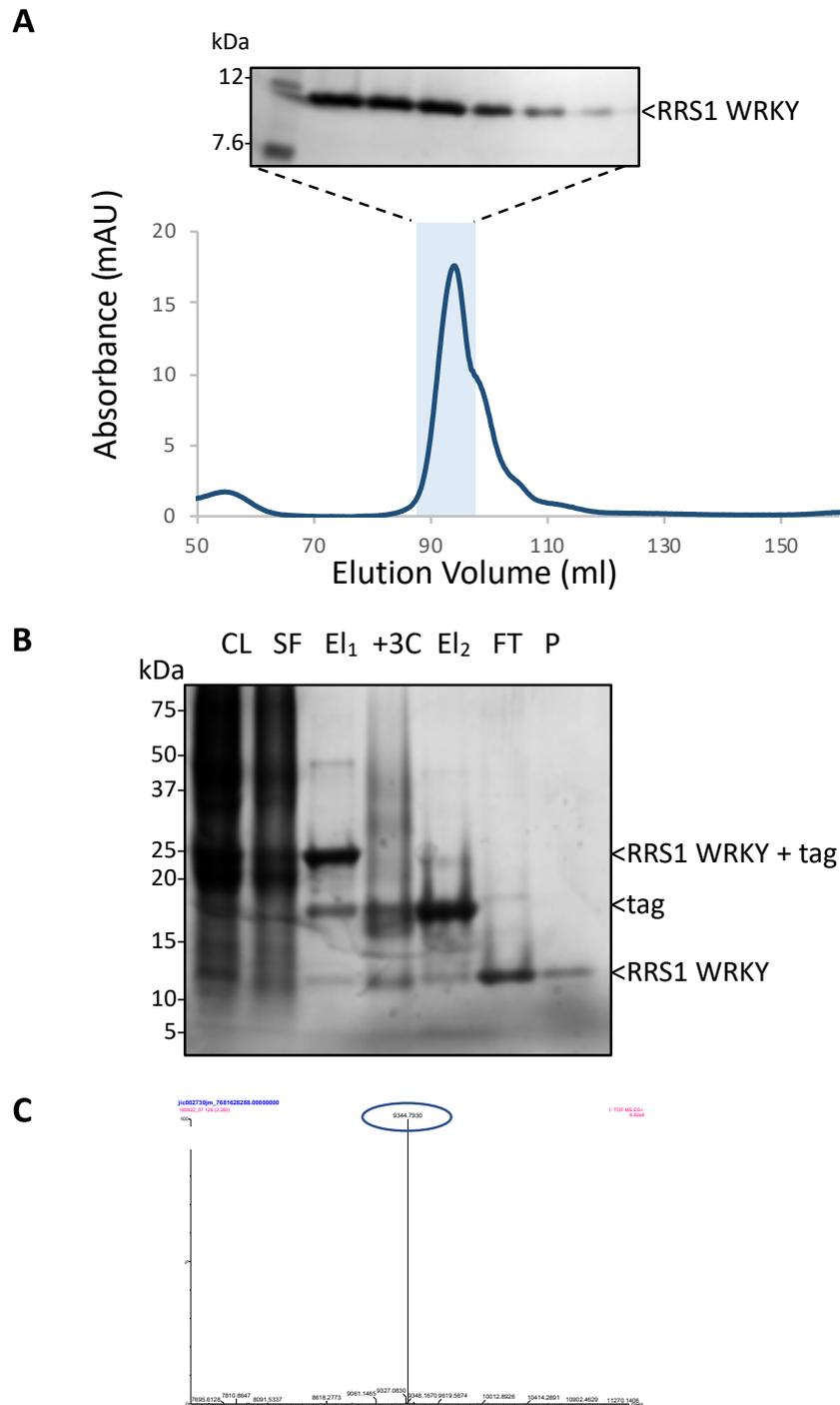
Figure 5.13 Purification of RRS1 WRKY (E1195-T1273) in pOPINS3C. (A) Gel filtration trace of RRS1 WRKY$_{E1195-T1273}$ eluting from a Superdex 75 26/600 gel filtration column peaking at 95 ml following 3C protease cleavage of Hisx6-SUMO tag, accompanying SDS-PAGE gel above trace shows fractions from peak highlighted in blue. **(B)** SDS-PAGE gel showing samples taken throughout the purification of RRS1 WRKY$_{E1195-T1273}$ from crude lysate to purified protein. Fraction codes are as follows;
CL-crude lysate, soluble fraction, El1 -Concentrated protein following IMAC purification and first gel filtration, +3C-Post overnight 3C protease cleavage, FT-Flow through after second IMAC purification, El2 -Elution from second IMAC, P- Purified protein after second Gel filtration. **(C)** Intact mass

(cont) spectrometry analysis of purified RRS1 WRKY, spectra showing different species identified by peaks in intact mass spectrometry analysis, peak of interest for RRS1 $WRKY_{E1195-T1273}$ is circled (9344.7930 Da), exact mass (in Daltons) is labelled above each peak. Experiments, except for mass spectrometry, were repeated 3 times with similar results found across repeats.
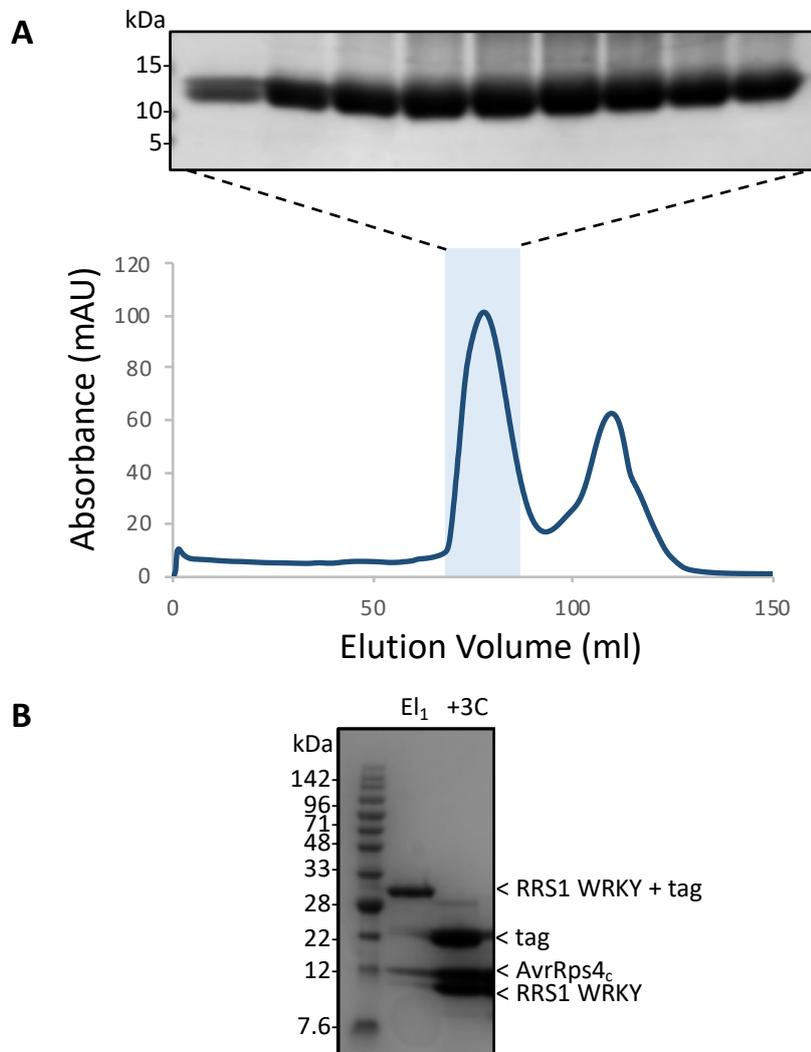
Figure 5.14 Purification of RRS1 WRKY$_{E1195-T1273}$ and AvrRps4c $_{G134-Q221}$ complex. (A) Gel filtration trace of RRS1 WRKY$_{E1195-T1273}$ –AvrRps4$_c$ complex eluting from a Superdex 75 16/600 gel filtration column peaking at 79 ml following 3C protease cleavage of RRS1 WRKY$_{E1195-T1273}$ Hisx6-SUMO tag, accompanying SDS-PAGE gel above trace shows fractions from peak highlighted in blue. **(B)** SDS-PAGE gel of RRS1 WRKY$_{E1195-T1273}$ + Hisx6-SUMO tag and AvrRps4$_c$ after first IMAC purification and gel filtration (El$_1$) and after 3C protease cleavage of RRS1 WRKY$_{E1195-T1273}$ Hisx6-SUMO tag (+3C). Experiment repeated 3 times with similar results found across repeats.

5.3.2   RRS1 WRKY and AvrRps4$_c$ form a complex in vitro

As with RRS1 WRKY Dom6S, I wanted to confirm RRS1 WRKY alone could interact with AvrRps4$_c$ in vitro. I looked to first confirm this using the qualitative method of analytical gel filtration. As descried in 5.2.2, RRS1 WRKY$_{E1195-T1273}$ and AvrRps4$_c$ were first run as single proteins on a Superdex™ 75 10/300 GL column analytical gel filtration column. As single proteins RRS1 WRKY$_{E1195-C1290}$ and AvrRps4$_c$ eluted at 15.2 ml and 13.0 ml respectively, **Figure 5.15**. However, when the two proteins were mixed in a molar ratio of 1:1 and left to incubate on ice for one hour before being run on a Superdex™ 75 10/300 GL column, the two proteins eluted in a single peak together at 12.6 ml indicating that a complex had been formed. Conversely when RRS1 WRKY$_{E1195-C1290}$ was incubated with double mutant AvrRps4$_c$ E187A/E175A, no peak shift was observed, and the two proteins eluted as separate peaks at 13.4 ml and 15.1 ml indicating no complex was being formed, **Figure 5.15**. SDS-PAGE analysis confirmed the presence of the relevant proteins under the gel filtration peaks, **Figure 5.15B**.

Having confirmed the ability of RRS1 WRKY to directly bind AvrRps4 qualitatively, I aimed to gain a quantitative insight on this binding event using SPR. As described in 5.3.2, SPR R$_{max}$ experiments were setup with AvrRps4$_c$, wild type and negative control mutants E187A and E187A/E175A, used as the ligand bound to the surface of the Ni-NTA sensor chip, and RRS1 WRKY$_{E1195-T1273}$ as the analyte. Similar to RRS1 WRKY Dom6S$_{E1195-C1290}$, I found that RRS1 WRKY$_{E1195-T1273}$ was prone to bind non-specifically to the surface of the sensor chip. I conducted extensive optimisation experiments to identify suitable conditions for conducting R$_{max}$ analysis including varying the temperature of the experiment from 4-25^{0}C, concentration of NaCl in the SPR running buffer, and the flow rate at which the analyte RRS1 WRKY was applied to sensor chip. The optimal conditions identified through these approaches were to use 8^{0}C, with 500 mM NaCl in the SPR running buffer, and an analyte flow rate of 30µl/min for 120 seconds contact time and 60 seconds dissociation time. R$_{max}$ data was collected for the interaction between RRS1 WRKY$_{E1195-T1273}$ and AvrRps4$_c$ at the analyte concentrations of 500 nM and 3000 nM. Two biological repeats were conducted each containing three technical repeats of each investigated interaction. The data from the two biological replicates was combined to give %R$_{max}$ Figures for RRS1 WRKY$_{E1195-T1273}$ interaction with wild-type, E187A and E187A/E175A AvrRps4$_c$, **Figure 5.16**.

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

Despite the extensive optimisation work described above, I was unable to generate equilibrium dissociation constant (K_D) values for RRS1 WRKY$_{E1195-T1273}$ and AvrRps4$_c$ binding. This was because the data failed the quality tests used by the Biacore software when attempting to fit the data to a steady-state affinity model, largely due to high bulk effects.
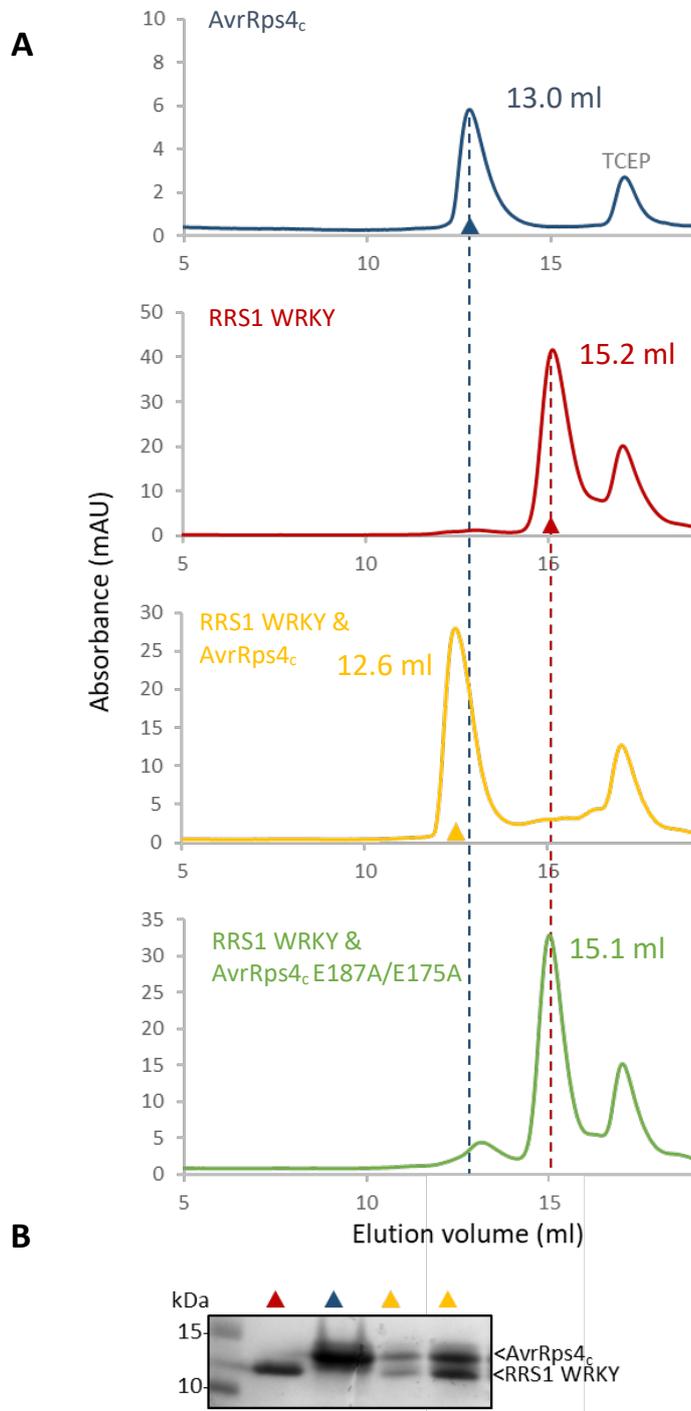
Figure 5.15 RRS1 WRKY$_{E1195-T1273}$ and AvrRps4$_c$ form a complex in vitro (A) Analytical gel filtration traces showing elution volume of AvrRps4$_c$ (blue), RRS1 WRKY$_{E1195-T1273}$ (red), RRS1 WRKY$_{E1195-T1273}$ /AvrRps4$_c$ complex (yellow) and mixing of RRS1 WRKY$_{E1195-T1273}$ and AvrRps4$_c$ E187A/E175A (green).

(B) SDS-PAGE analysis shows the fractions taken from volumes on trace indicated by coloured triangles. Peak volumes shown indicate elution volume. Baselines corrected to zero. The peak observed in all traces at ~17 ml is caused by the presence of the reducing agent TCEP in the sample buffer. Experiment repeated 3 times with similar results found across repeats.
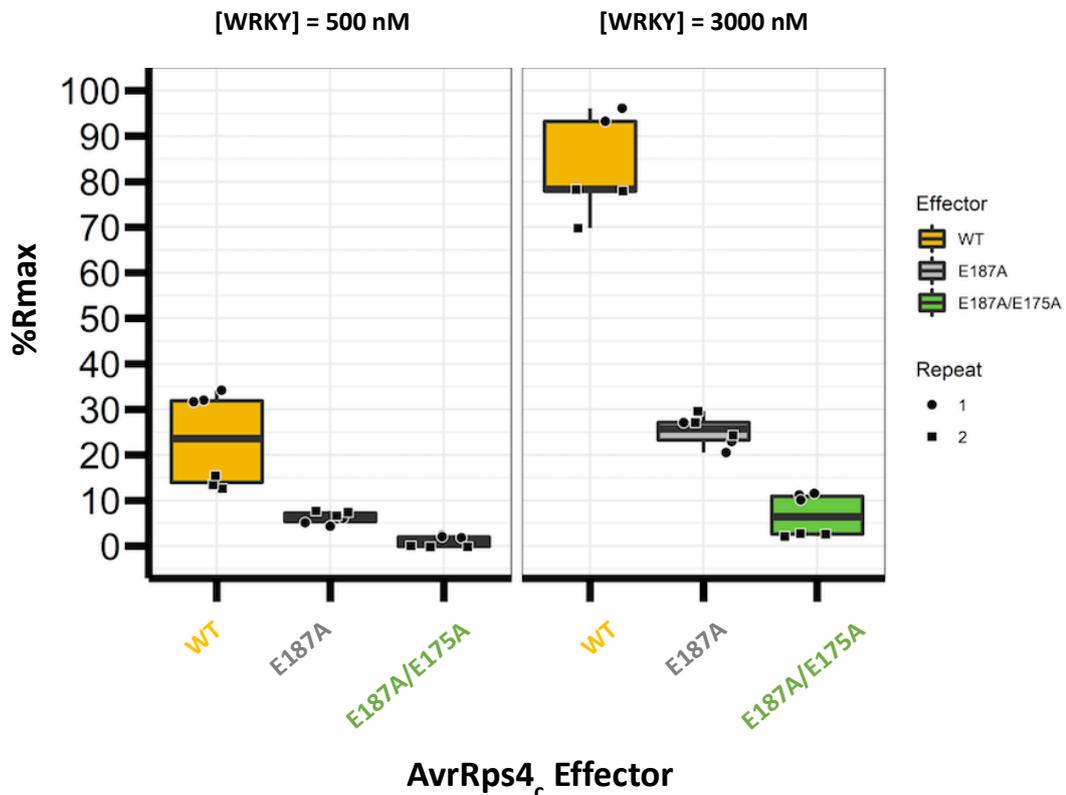
Figure 5.16 AvrRps4$_c$ residues E187 and E175 play important role in RRS1 WRKY$_{E1195-T1273}$ binding of AvrRps4$_c$. %Rmax of RRS1 WRKY$_{E1195-T1273}$ bound to wild type and mutant AvrRps4$_c$ assuming a 1:1 binding model for RS1 WRKY Dom6S: AvrRps4. The left and right panel were conducted with 500 nM and 3000 nM RRS1 WRKY$_{E1195-T1273}$ respectively. The median is represented in the centre line of the box, and the upper and lower quartile by the box limits. Whiskers extend from smallest value in Q_1 - 1.5x interquartile range (IQR) and the largest value within Q_3 + 1.5x IQR. Individual data points from each biological repeat are represented by a black square or circle. Plots were produced using the ggplot2 package[171] in R.

5.3.3   Crystalizing RRS1 WRKY and AvrRps4$_c$ complex

To identify conditions for crystallization of the RRS1 WRKY$_{E1195-T1273}$-AvrRps4$_c$ complex, commercial 96-well sitting drop vapour diffusion screens were setup. Initially, sparse-matrix screens used protein concentrations of 11.2 mg/ml and 5.6 mg/ml in JCSG-plus™ and MIDAS™ (Molecular Dimensions). At these concentrations I observed a high level of precipitation in wells containing the higher 11.2 mg/ml of protein but not 5.6 mg/ml. As such, I decided to lower the protein concertation to 6.3 mg/ml and ~3.15 mg/ml for future

screening. Based on these concentrations, the KISS, Morpheus® and ProPlex™ crystallisation screens were setup at 25^0C (ProPlex™ is a crystallisation screen especially formulated for the crystallisation of protein complexes). Two conditions in these screens supported the growth of crystals, **Figure 5.17**. Firstly, in well A1 of the KISS screen setup with 6.3 mg/ml of protein complex with well conditions of 0.1M sodium acetate trihydrate, 10% PEG 3350, pH 4, 0.2M ammonium sulfate. Secondly, in well C2 of the ProPlex™ screen also setup with 6.3 mg/ml of protein complex with buffer conditions 0.1M Sodium citrate, pH 4.5 and 20% PEG4000. However, both these crystals were hypothesised to comprise only AvrRps4$_c$ as they did not fluoresce in UV light, **Figure 5.17**, which is characteristic of AvrRps4$_c$ protein which has a very low extinction coefficient at 280 nM due to a lack of aromatic residues. If these crystals contained RRS1 WRKY$_{E1195-T1273}$ they would be expected to glow in UV light as RRS1 WRKY$_{E1195-T1273}$ has a high extinction coefficient. The conditions above were also very similar to those in which AvrRps4$_c$ was originally crystallized[166], 14–16% vol/vol 2-methyl-2, 4-pentandiol buffered with 0.1 M sodium acetate, pH 5.1–5.3. Therefore, these crystals were not taken further for analysis by X-ray diffraction.

Having observed the increased stability of the RRS1 WRKY$_{E1195-T1273}$-AvrRps4$_c$ complex at lower temperatures of 4-8^{0}C in analytical gel filtration and SPR experiments, **Figure 5.8**, I setup crystallisation trials at 4^{0}C to investigate whether a lower temperature would help crystallization. I setup the following 96-well sparse matrix commercial crystallisation screens at 6.1 mg/ml and 3 mg/ml; Structure, JCSG-plus™, PGA™ and PACT premier™ (Molecular Dimensions) but no crystal formation was observed.

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

A

0.1 M sodium acetate
trihydrate, pH 4,0.2 M
ammonium sulfate,
10% PEG 3350
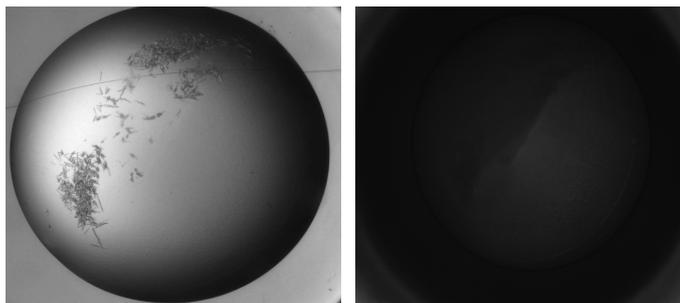
B

0.1 M Sodium citrate,
pH 4.5, 20% PEG4000
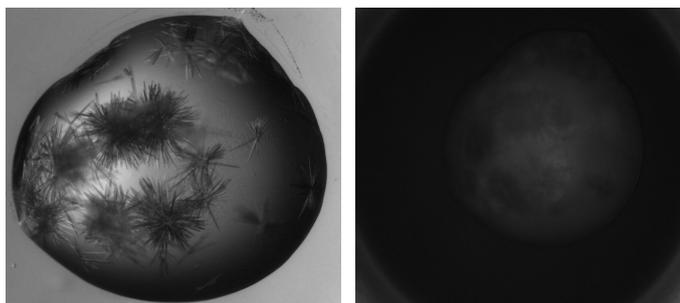
Figure 5.17 Images of crystals seen during RRS1 WRKY$_{E1195-I273}$ -AvrRps4$_c$ crystallization trials predicted to be AvrRps4$_c$ only. Images on left are brightfield view images on right taken in UV light. (cont) **(A)** Crystals found well A1 of KISS screen with buffer conditions 0.1M sodium acetate, pH 4.0, 0.2M ammonium sulfate, 10% PEG 3350. **(B)** Crystals found in well C2 of commercial screen ProPlex™, buffer conditions 0.1M, sodium citrate, pH 4.5, 20% PEG4000. Crystals in both wells were found not to glow in UV light and were therefore predicted to be AvrRps4$_c$ only.

5.4   Investigating interactions of RRS1B WRKY and AvrRps4$_c$

In addition to gaining biochemical and structural insights into the binding event between RRS1 WRKY and AvrRps4, I also expanded the study to examine the interaction of the RRS1B WRKY with AvrRps4. This was based on the rationale that should a structure of this interface be obtained, it would provide an excellent opportunity to compare how a single effector, AvrRps4, is recognised by different NLRs when compared to a structure for RRS1 WRKY-AvrRps4. Given the sequence similarity of only 56% across the WRKY domain between RRS1 and RRS1B, understanding this interface at the structural level could provide interesting insights into the requirements and flexibilities of this NLR-effector interaction.

To generate an expression construct of RRS1B WRKY with similar boundaries to RRS1 WRKY$_{E1195-T1273}$, the WRKY domain sequences of RRS1-R and RRS1B were aligned using Clustal Omega[242]. This identified RRS1B residues N1163-H1237 as the equivalent construct, **Figure 5.18A**. RRS1B WRKY$_{N1163-H1237}$ was cloned into pOPINS3C as for the RRS1 WRKY construct. Small-scale expression tests of RRS1B WRKY$_{N1163-H1237}$ in pOPINS3C were

conducted in Rosetta™ 2 (DE3), BL21(DE3), SHuffle® T7 and Lemo21(DE3) cell lines, with expression levels judged by SDS-PAGE of the soluble fraction of the total cell lysate, **Figure 5.18B**. The expression tests found that similar to RRS1 WRKY, RRS1B WRKY showed greatest yields of soluble expression in Rosetta™ 2 (DE3) grown in LB media with 0.4 mM IPTG induction.

Having optimised conditions for soluble expression of RRS1B WRKY$_{N1163-H1237}$, expression was scaled up to 6-8 L of culture, followed by protein purification. Following IMAC purification and gel filtration, the 6xHis-SUMO tag was cleaved from RRS1B WRKY$_{N1163-H1237}$ with 3C protease, and the protein subjected to a final round of gel filtration purification on a Superdex™ 75 26/600 column. The presence of protein was confirmed by SDS-PAGE of the fractions comprising the peak, **Figure 5.19**.

A

RRS1-R WRKY$_{E1195}$

```
EKKHRESKVKKVV--SIPAIDEGDLWTWRKYGQKDILGSRFPRGYYRCAYKFTHGCKATK   1247
--RRKNNKEKRVVCVVDRGSRSSDLWVWRKYGQKPIKSSPYPRSYYRCAS--SKGCFARK   1215
:::::.* *:**    .  ..***.****** * .* :**.*****   ::** * *
```

RRS1B WRKY$_{N1163}$

```
QVQRSETDSNMLAITYLSEHNH                                       1267
QVERSRTDPNVSVITYISEHNH                                       1247
**:**.** *: .***:*****
```

B

Figure 5.18 Expression trialling of RRS1B WRKY$_{N1163-H1237}$ (A) Sequence alignment of RRS1-R WRKY and RRS1B WRKY amino acid sequence, alignment constructed using Clustal Omega[242], construct used for expression testing is highlighted in orange with core 'WRKYGQK' motif highlighted in blue, and amino acid number listed on the right of the sequence. **(B)** Small-scale expression trial results for expressing RRS1B WRKY$_{N1163-H1237}$ in pOPINS3C in a variety of cell lines listed below each gel. 16% SDS-PAGE were stained with Coomassie dye to visualise protein. RRS1B WRKY protein bands of interest are highlighted with an arrow to the right of gel. CL: crude lysate, SF: Soluble fraction.

To investigate AvrRps4 recognition by RRS1B I produced complex of RRS1B WRKY$_{N1163-H1237}$ - AvrRps4$_c$ utilising method 3, described in 2.6.3. In this method, AvrRps4$_c$ in pOPINF, and RRS1B WRKY$_{N1163-H1237}$ in pOPINS3C were grown and purified by IMAC and gel filtration separately. Both proteins then underwent overnight cleavage of 6xHis and 6xHis-SUMO tags respectively with 3C protease. Following protease cleavage, cleaved affinity tag and tagged protein was removed by IMAC purification. During this second IMAC purification untagged RRS1B WRKY$_{N1163-H1237}$ and AvrRps4$_c$ flow directly through the IMAC column and

were eluted into the same vessel. The untagged proteins were incubated on ice for an hour to allow the proteins to interact and form a complex, as previously demonstrated in analytical gel filtration studies. The proteins then underwent a final round of gel filtration purification on a Superdex™ 75 16/600. During this step, the proteins eluted in a single peak form the column, and the presence of both proteins was confirmed by SDS-PAGE, indicating that RRS1B WRKY$_{N1163-H1237}$ and AvrRps4$_c$ could form a complex in vitro, **Figure 5.20**. I found that yields of RRS1B WRKY$_{N1163-H1237}$-AvrRps4$_c$ complex were higher when the proteins were produced by this method rather than mixing the cell lysates of cultures of AvrRps4$_c$ and RRS1B WRKY$_{N1163-H1237}$ as was used for RRS1 WRKY$_{E1195-T1273}$-AvrRps4$_c$ complex purification, the reasons for this are not fully understood.

Figure 5.21 Purification of RRS1B WRKY (N1163-H1237) in pOPINS3C. Gel filtration trace of RRS1B WRKY eluting from a Superdex 75 26/600 gel filtration column peaking at 251 ml following 3C protease cleavage of SUMO tag, accompanying SDS-PAGE gel above trace shows fractions from peak highlighted in blue. The trace blip peak at ~257 ml was caused by a changeover in the AKTA fraction collection plate. Experiment repeated 3 times with similar results found across repeats.
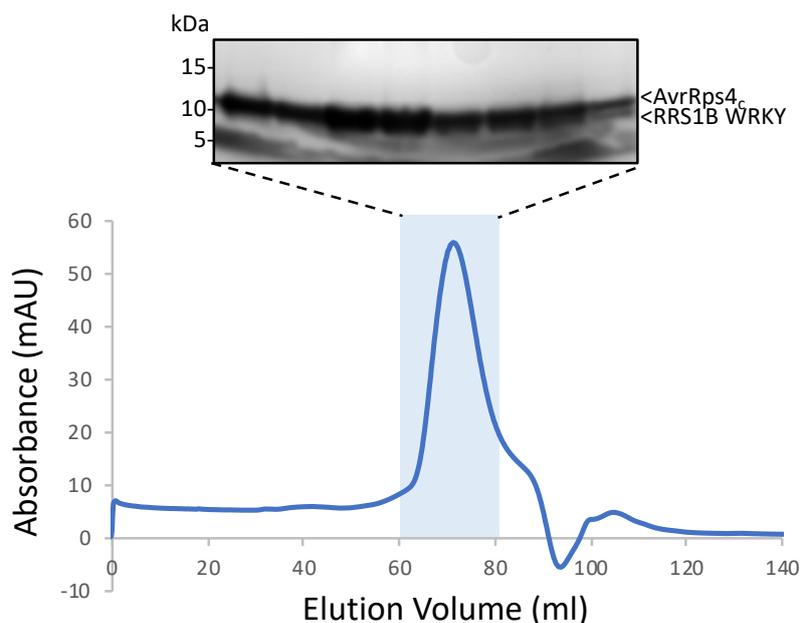
Figure 5.24 Purification of RRS1B WRKY (N1163-H1237) and AvrRps4c(G134-Q221) complex. Gel filtration trace of RRS1B WRKY$_{N1163-H1237}$ –AvRps4$_c$ complex eluting from a Superdex 75 16/600 gel filtration column peaking at 72 ml following 3C protease cleavage of RRS1 RRS1B WRKY$_{N1163-H1237}$ Hisx6-SUMO tag, accompanying SDS-PAGE gel above trace shows fractions from peak highlighted in blue. Experiment repeated 3 times with similar results found across repeats.

In addition to direct purification of the complex, I investigated whether RRS1B WRKY$_{N1163-H1237}$ and AvrRps4$_c$ could form a complex by mixing the separately purified components followed by analytical gel filtration. For this, RRS1B WRKY$_{N1163-H1237}$ and AvrRps4$_c$ were separately purified and run individually to monitor their elution volumes. Subsequent to this, the two proteins were mixed together in a 1:1 molar ratio and left to incubate on ice for four hours. However, unlike RRS1 WRKY$_{E1195-T1273}$, RRS1B WRKY$_{N1163-H1237}$ did not form a complex with AvrRps4$_c$ under these conditions, as no peak shift was observed **Figure 5.21**. Why RRS1B WRKY$_{N1163-H1237}$ and AvrRps4$_c$ did not form a complex when mixed but did when co-purified is not fully understood and further work will be needed to identify binding conditions to support this interaction with separately purified proteins.
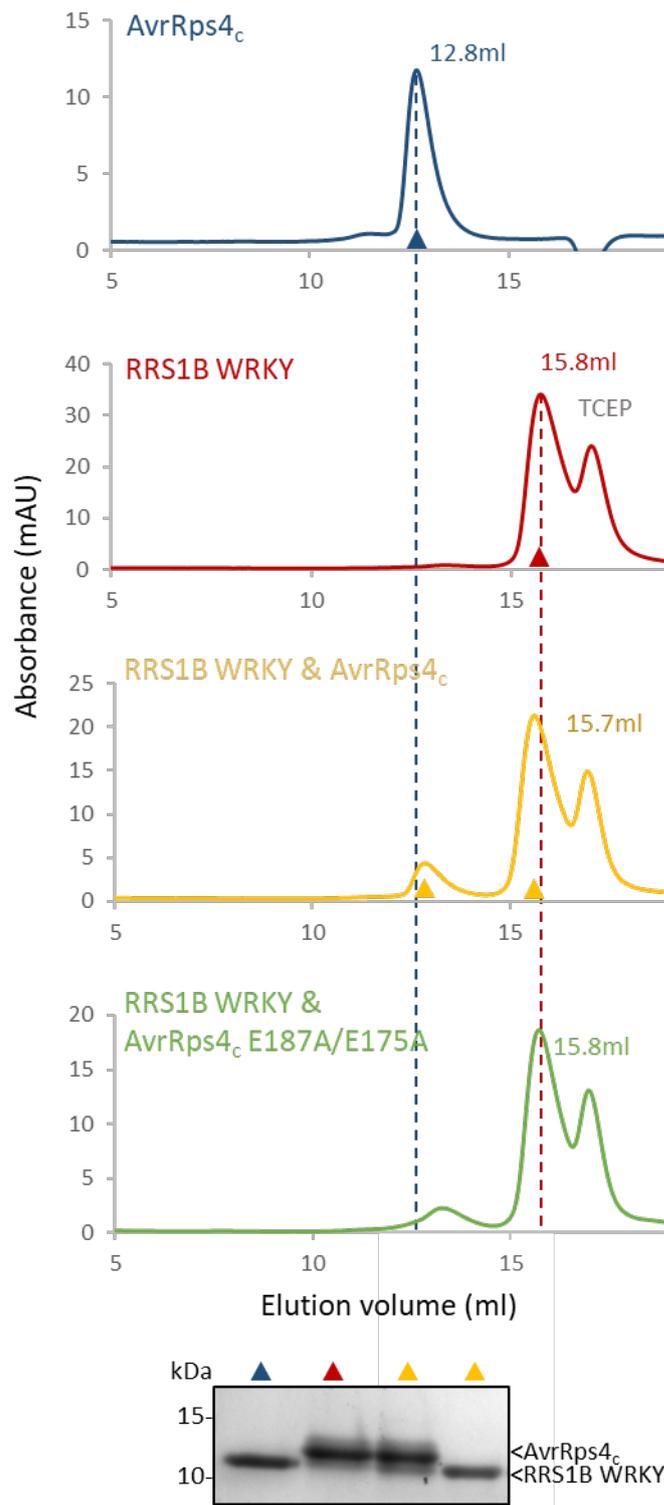
Figure 5.27 Qualitative binding analysis of RRS1B WRKY and AvrRps4$_c$. (A) Analytical gel filtration traces showing elution volume of AvrRps4c (blue), RRS1B WRKY (red), mixing of RRS1B WRKY and AvrRps4$_c$ (yellow) and mixing of RRS1B WRKY$_{N1163-H1237}$ and AvrRps4$_c$ E187A/E175A (green). **(B)** SDS-PAGE analysis shows the fractions taken from volumes on trace indicated by coloured triangles. Peak volumes shown indicate elution volume. Baselines corrected to zero. The peak observed in all traces at ~17 ml is caused by the presence of the reducing agent TCEP in the sample buffer. Experiment repeated twice with similar results found across repeats.

5.5   Non-WRKY domain RRS1-AvrRps4 interactions

There is some evidence in the literature for effectors interacting with integrated domain NLRs outside of an NLR's integrated domain. For example the *Magnaporthe oryzae* effector AVR-Pia has been observed to interact with *Oryza sativa* NLR RGA5 at sites outside of RGA5's integrated HMA/RATX1 domain[72]. Following from this observation, unpublished *in planta* co-immunoprecipitation from Dr Sung Huh and Dr Yan Ma in the Jones laboratory suggested that the TIR domain of RRS1 may also interact with AvrRps4 and RRS1 WRKY Dom6S. The data showed the ability of RRS1'S TIR domain to interact constitutively with RRS1 WRKY Dom6S and that this interaction was enhanced in the presence of AvrRps4, but not PopP2. To test if this interaction could be recapitulated in vitro, and allow for further biochemical study of this interaction, I conducted a qualitative binding experiment by analytical gel filtration.

RRS1 TIR protein was produced using the construct and conditions developed by Williams et al, 2014[114] which expressed RRS1 K6-Y153 in pMCSG7 with a TEV protease cleavable 6xHis tag in Rosetta™ 2 (DE3) cells. RRS1 TIR$_{K6-G153}$ protein was expressed and purified by IMAC purification before gel filtration. The 6xHis tag was cleaved by overnight incubation with TEV protease before un cleaved protein and cleaved tag was removed by IMAC. The protein then underwent a final round of purification by gel filtration on a Superdex™ 75 26/600 column and the presence of RRS1 TIR$_{K6-G153}$ was confirmed within the gel filtration peak by SDS-PAGE, **Figure 5.22**.
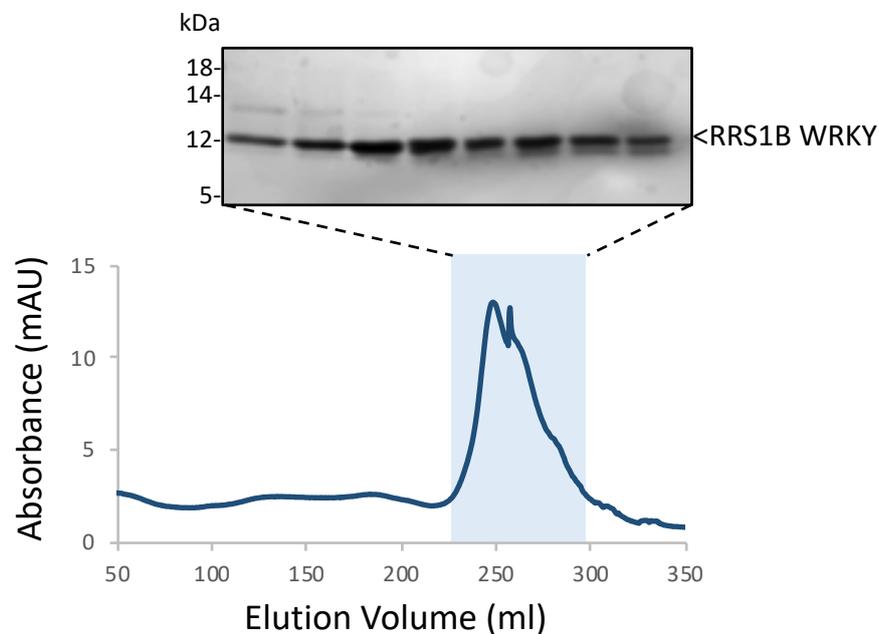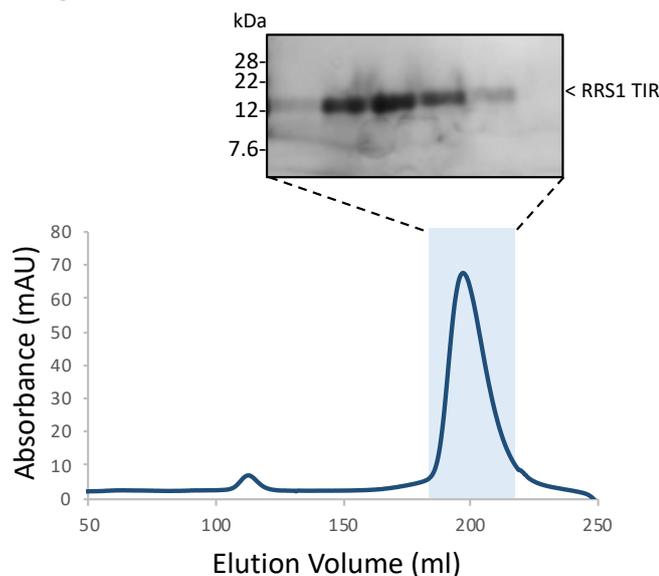

Figure 5.30 Purification of RRS1 TIR (K6-G153) in pMCSG7. Gel filtration trace of RRS1 TIR eluting from a Superdex 75 26/600 gel filtration column peaking at 199 ml following 3C protease cleavage of tag, accompanying SDS-PAGE gel above trace shows fractions from peak highlighted in blue.

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

To investigate the binding of RRS1 TIR$_{K6-G153}$, RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ qualitative analytical gel filtration analysis was conducted. The proteins were first run separately on a Superdex™ 75 10/300 GL column, eluting at 12.8 ml 14.1 ml and 12.7 ml respectively, **Figure 5.23**. As RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ had previously been observed to interact, a complex of RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ was also run to enable comparison of peak shifts in the presence of RRS1 TIR, **Figure 5.23**. The RRS1 TIR$_{K6-G153}$, RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ were then mixed in a molar ration of 1:1:1 and incubated on ice overnight. The protein mixture was then run on the analytical gel filtration column. However, no shift in trace peak, indicative of complex formation, was observed. Only peaks for RRS1 TIR$_{K6-G153}$, and a peak for complex of RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$, were seen at 11.9 ml and 13.1 ml, **Figure 5.23**. The lack of interaction between RRS1 TIR-RRS1 WRKY and AvrRps4$_c$ in vitro is not fully understood. A hypothesis could be that this interaction in facilitated by another interaction partner present in the *in planta* coimmunoprecipitation assays, but not when the proteins are mixed in isolation. Alternatively, the binding conditions provided in this in vitro assay could be inappropriate for supporting this binding event. It has been previously observed for example certain protein complexes are better supported in in vivo environments such as through co-expression in the same cell strain than in vitro. Conversely, the observed *in planta* RRS1 TIR, RRS1 WRKY and AvrRps4$_c$ interaction could be an artefact potentially due to protein overexpression levels. Should this interaction be proved true in future work, it could provide an interesting avenue of investigation for future work into understanding this NLR-effector interface.
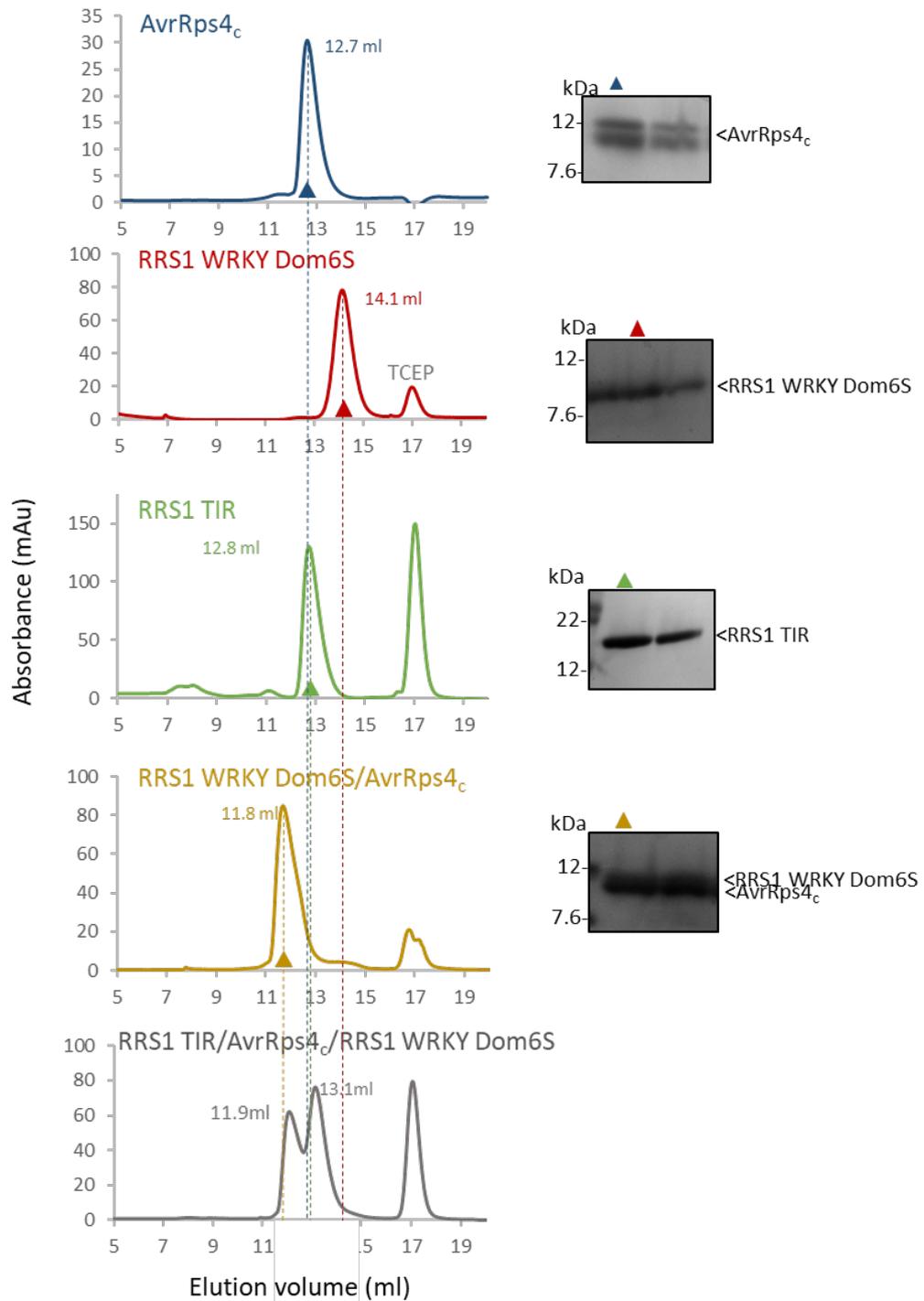
Figure 5.33 Qualitative binding analysis of RRS1 WRKY Dom6S, RRS1 TIR and AvrRps4$_c$. Analytical gel filtration traces showing elution volume of AvrRps4$_c$ (blue), RRS1 WRKY Dom6S$_{E1195-C1290}$ (red), RRS1 TIR$_{K6-G153}$ (green), RRS1 WRKY Dom6S$_{E1195-C1290}$ /AvrRps4$_c$ complex (yellow) and mixing of RRS1 WRKY Dom6S$_{E1195-C1290}$, RRS1 TIR$_{K6-G153}$, AvrRps4$_c$ (grey). SDS-PAGE analysis shows the fractions taken from volumes on trace indicated by coloured triangles. Peak volumes shown indicate elution volume. Baselines corrected to zero. The peak observed in all traces at ~17 ml is caused by the presence of the reducing agent TCEP in the sample buffer. Experiment repeated twice with similar results found across repeats.

5.6   Exploring RRS1 WRKY interactions with W-box DNA

In addition to understanding the structural basis of AvrRps4 recognition by RRS1/RRSB WRKY, I wanted to examine the effect AvrRps4$_c$ has on the ability of the RRS1 WRKY domain to bind to W-box DNA. For this I used a specialised SPR technique called ReDCaT (Re-usable DNA Capture Technique), specialised for examining protein-DNA interactions[172]. As a negative control for wild type W-box DNA sequence (TTGACCG)[243] I trialled the use three W-box mutant variants. These were: Mutant 1 which was the W-box mutant sequence used in previous EMSA (electrophoretic mobility shift assay) studies with RRS1[65], Mutant 2 which was generated through combining W-box mutations used by other non-RRS1 publications[174,175] and Mutant 3 which was generated by scrambling the wild-type W-box using an online scrambling tool from GenScript. Nucleotide deviations of these mutants from the wild type W-box sequence are highlighted in **Figure 5.25**. The wild-type and mutant sequences of W-box DNA were trialled as both single and triple repeats of the sequence, the details are shown in **Table 2.8**.

The SPR ReDCaT technique involves immobilizing biotinylated single stranded ReDCaT linker DNA sequences to a streptavidin coated SPR sensor chip. The DNA sequences for binding testing are then generated as two separate oligonucleotides, forward and reverse. The first strand contains only the forward sequence of the DNA sequence to be tested (W-box), whilst the second strand contains the reverse sequence of your testing DNA sequence and the complementary sequence to the ReDCaT linker sequence. These two oligonucleotides are annealed by heating and injected over the ReDCaT linker immobilized chip. The annealed oligonucleotides bind to the ReDCaT chip. A pictorial representation of this experimental setup is shown in **Figure 2.2**. This allows you to generate a SPR sensor chip displaying your DNA sequence of choice to proteins which are then flowed over as analyte.

Initial experiments used a construct of RRS1 WRKY Dom6S generated prior to switching to the RRS1 WRKY Dom6S$_{E1195-C1290}$ described in 5.2. This construct expressed RRS1 residues S1183-C1290 in pOPINF in SHuffle® T7 cell grown in LB media. Following 3C protease cleavage on 6xHis tag the protein underwent a final round of purification by gel filtration in which protein presence in trace peak was confirmed by SDS-PAGE, see **Figure 5.24.**

These early experiments showed that the RRS1 WRKY Dom6S protein was interacting non-specifically with the ReDCaT chip. Subsequent trialling of this system varying NaCl

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

concentration in the SPR buffer and analyte flow rate established conditions in which this non-specific interaction was reduced. These optimised conditions used a buffer containing 300 mM NaCl, and an analyte concentration of 500 nM and analyte flow rate of 30 µl/min with 60 seconds contact time and 60 seconds dissociation time.
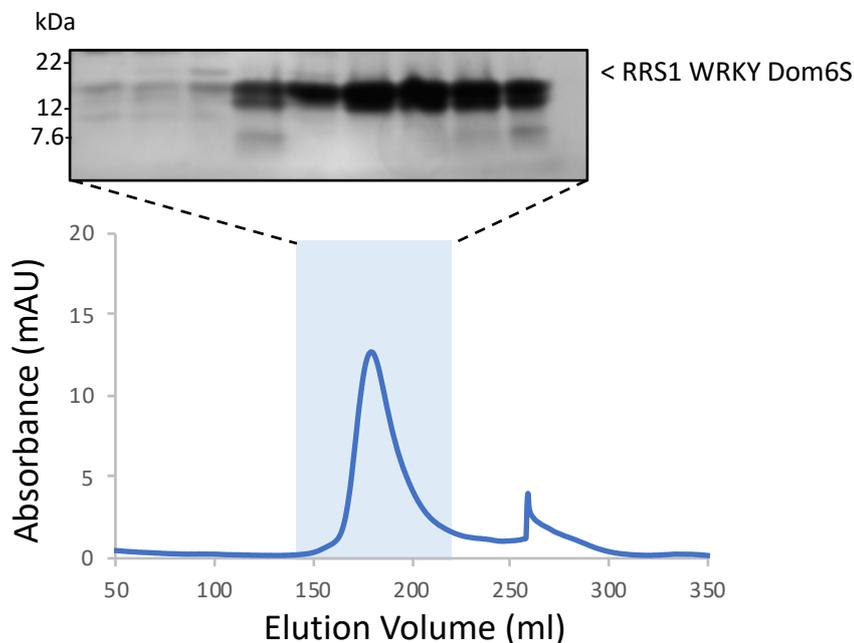
Figure 5.35 Purification of RRS1 WRKY Dom6S (S1184-C1290) in pOPINF. Gel filtration trace of RRS1 WRKY Dom6S$_{S1184-C1290}$ eluting from a Superdex 75 26/600 gel filtration column peaking at 182 ml following 3C protease cleavage of Hisx6 tag, accompanying SDS-PAGE gel above trace shows fractions from peak highlighted in blue. Experiment repeated 3 times with similar results found across repeats.

Under these conditions RRS1 WRKY Dom6S$_{S1184-C1290}$ showed a clear increased affinity for wild type W-box DNA over mutant 1, 2, or 3 W-box DNA. This increased affinity of RRS1 WRKY Dom6S$_{S1184-C1290}$ for wild type W-box DNA was observed for both single and triple DNA sequence repeats, **Figure 5.18B**.

Having established that RRS1 WRKY Dom6S$_{S1184-C1290}$ was capable of binding W-box DNA in the SPR ReDCaT experiment, I investigated whether AvrRps4 could disrupt this interaction. The experimental plan was to conduct both displacement and competition assays. In the displacement setup, W-box DNA would first be immobilized to the ReDCaT chip. Following this, RRS1 WRKY would then be flowed over the chip as the analyte and bind the W-box DNA. Next, AvrRps4$_c$ would then be flowed over the RRS1 WRKY bound W-box DNA chip and a response for RRS1 WRKY loss of DNA binding monitored. In the competition assay,

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

RRS1 WRKY would be mixed with AvrRps4$_c$ prior to flowing over the W-box DNA immobilized chip and affinity for DNA compared to RRS1 WRKY alone binding to W-box DNA.

A

Wild Type W-box	T	T	G	A	C	C	G
Mutant 1 (EMSA)	T	A	G	A	C	G	G
Mutant 2 (Literature)	T	T	G	C	A	C	G
Mutant 3 (Scrambled)	T	G	A	T	C	G	C

B
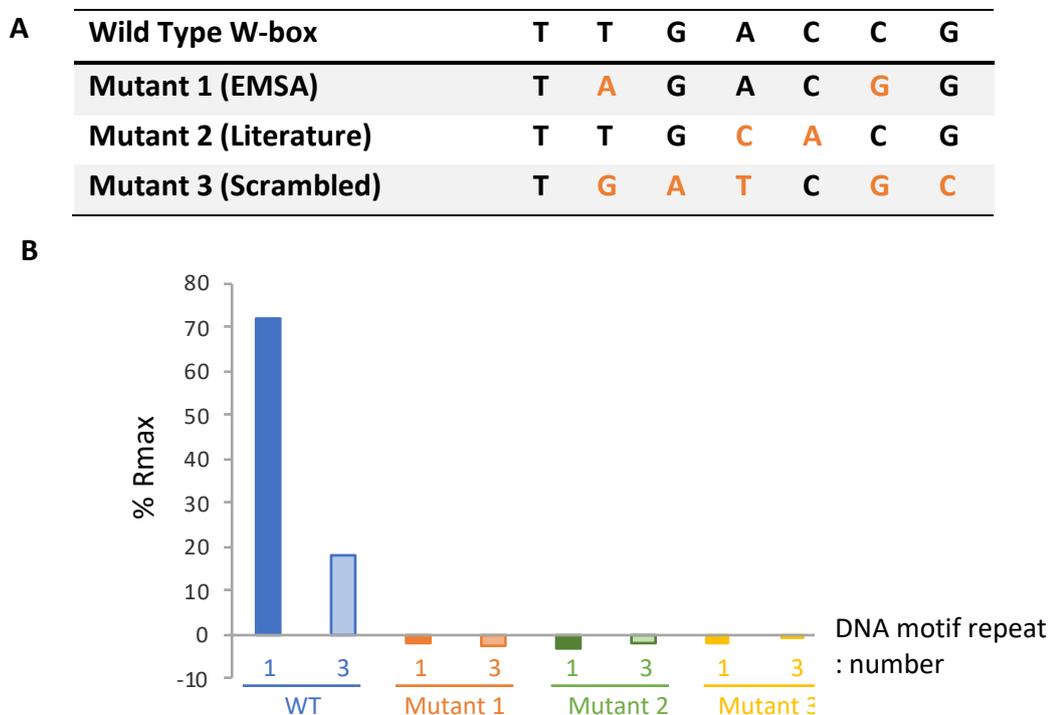
Figure 5.39 RRS1 WRKY Dom6S$_{S1184-C1290}$ binds WT but not mutant W-box DNA in vitro. (A) W-box DNA sequences used in SPR ReDCaT chip experiment. 3 different W-box DNA mutants were compared to wild type, Mutant 1 is the sequence previously used in RRS1 WRKY EMSA study[65] (EMSA mutant), Mutant 2 is based on sequences used in the literature and Mutant 3 was generated by scrambling the wild-type W-box sequence using a GenScript online scrambling tool. Nucleotide deviations from wild type sequence are highlighted in orange. **(B)** %Rmax plots of RRS1 WRKY Dom6S$_{S1184-C1290}$ binding to WT and mutant W-box DNA immobilised on a ReDCaT SPR chip. RRS1 WRKY Dom6S$_{S1184-C1290}$ binding was tested against single and 3 repeats of each DNA sequence.

 I setup this experiment using the RRS1 WRKY$_{E1195-T1273}$ construct described in 5.3.1 to complement structural work with this construct. However, I was unable to collect this data due to issues with non-specific binding of RRS1 WRKY$_{E1195-T1273}$ to the surface of the sensor chip. Conditions were setup as for the RRS1 WRKY Dom6S$_{S1184-C1290}$ using SPR buffer containing 300 mM NaCl and a flow rate of 30 µl/min (60 seconds contact time, 60 seconds dissociation time). Various optimization experiments were tried to reduce the non-specific binding effect, including varying the RRS1 WRKY$_{E1195-T1273}$ analyte concentration from 10-500 nM and decreasing the analysis temperature to 8^{0}C, as had worked for R$_{max}$ studies, but no

conditions rescued the non-specific binding issues sufficiently. I also tested various SPR running buffers. This included increasing the NaCl concentration in the SPR buffer to 860 mM and increasing the surfactant Tween20 from 0.05% to 0.1%, but again no significant reduction in RRS1 WRKY$_{E1195-T1273}$ non-specific binding was achieved. This experiment was repeated with RRS1B WRKY$_{N1163-H1237}$, but similar to RRS1 WRKY$_{E1195-T1273}$, issues with non-specific binding to the sensor chip prevented data being collected. Work is now ongoing to optimise conditions for completing the analysis of how AvrRps4$_c$ might affect the ability of RRS1 and RRS1B WRKY domains to bind W-box DNA.

5.7   Discussion

The objective of this chapter was to use structural biology and biochemical techniques to explore the molecular and structural basis of AvrRps4 recognition by RRS1. Using a range of allelic variants of RRS1, and the paralogous RRS1B, I investigated whether inherent flexibilities of the RRS1 system enable this NLR to recognise multiple effectors, **Table 5.1**. It was hoped that through gaining structural information on the RRS1 WRKY-AvrRps4 interface we could complement previous *in planta* work to further dissect the mechanism of RRS1/RPS4 activation. Through gaining the structure of RRS1 WRKY-AvrRps4, the aim was to compare this with the published structure of RRS1 WRKY-PopP2[71] and investigate how a single NLR can bind structurally and mechanistically distinct effectors. Should a structure of RRS1B WRKY-AvrRps4$_c$ also be achieved, this would then provide interesting insights into how a single effector, AvrRps4, binds two NLR WRKY domains which only share 56% sequence identity[141].

Initial expression trials identified conditions in which I could express RRS1 WRKY Dom6S (E1195-C1290) and WRKY (E1195-T1273). Despite trials in multiple cell strains I was unbale to produce soluble RRS1 WRKY Dom6R$_{E1195-Y1373}$. As described in chapter 3, additional trialling to identify soluble constructs of RRS1 WRKY Dom6R, which looked at a variety of N-terminal expression boundaries (RRS1-R S1184, R1194, E1195 and E1209) and expression in multiple expression vectors with various soluble tags (pOPINF, pOPINS3C and pOPINM) also failed to identified conditions for soluble expression of this protein in *E. coli*. I hypothesise this may be due to the inherent predicted disorder of the Dom6R extension, which may hinder expression of this protein in *E. coli*. The constructs I carried forward for quantitative and qualitative binding experiments mirrored the N-terminus of the construct used by

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

Zhang et al, 2017[71], such that direct comparison could be conducted in the future should a structure of RRS1 WRKY$_{E1195-T1273}$-AvrRps4 be achieved.

Table 5.1 Summary of successful expression of RRS1 WRKY variations and AvrRps4c with experimental update.

Protein	Amino acid	Expression conditions	Experiment update
RRS1 WRKY	E1195-T1273	pOPINS3C, Rosetta™ 2 (DE3) *E. coli*	Interacts with AvrRps4c by analytical gel filtration and SPR. AvrRps4 residues E187 and E175 important for facilitating this interaction.
RRS1 WRKY Dom6S	E1195-C1290	pOPINS3C, Rosetta™ 2 (DE3) *E. coli*	
RRS1B WRKY	N1163- H1237	pOPINS3C, Rosetta™ 2 (DE3) *E. coli*	Only interacts with AvrRps4c immediately following SUMO tag cleavage cell lysate stage of purification.
AvrRps4c	G134-Q221	pOPINF, SHuffle® T7 *E. coli*	
RRS1 WRKY & AvrRps4c	RRS1: E1195-T1273, AvrRps4: G134-Q221	Cultures of RRS1 WRKY and AvrRps4 grown separately and cell lysates mixed after thawing	Crystallisation efforts continue. Only observed AvrRps4c crystals so far.
RRS1 WRKY Dom6S & AvrRps4c	RRS1: E1195-C1290, AvrRps4: G134-Q221	Cultures of RRS1 WRKY Dom6S and AvrRps4 grown separately and cell lysates mixed after thawing	Crystallisation conditions for optimisation identified.
RRS1B WRKY & AvrRps4c	RRS1B: N11630-H1237, AvrRps4: G134-Q221	RRS1B WRKY only interacts with AvrRps4c immediately following SUMO tag cleavage cell lysate stage of purification	Shown proteins interact via analytical gel filtration. Crystallisation efforts continue.

Having identified conditions for the soluble expression of RRS1 WRKY$_{E1195-T1273}$ and WRKY Dom6S$_{E1195-C1290}$ I confirmed these proteins could both directly bind AvrRps4$_c$ in vitro by analytical gel filtration, **Table 5.1**. Previous *in planta* work had shown that AvrRps4 E187A /E175A mutants were incapable of activating RRS1/RPS4 dependent HR. These glutamic acid residues are located in a patch of negative surface charge on AvrRps4$_c$ which ClusPro 2.0 protein-protein interaction modelling predicts interact with key lysine residue (K$_2$) in RRS1's 'WRKYGQ**K**' motif, **Figure 5.7**. To help validate this model, I investigated whether the inability of AvrRps4 E187A /E175A to induce HR was due to a loss/reduction in binding affinity with RRS1 WRKY domain. Using analytical gel filtration, I showed that both AvrRps4$_c$ E187A and E187A /E175A were unable to bind RRS1 WRKY or WRKY Dom6S protein. This observation supports the hypothesis that these residues form important interactions in this

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

NLR-effector interface, and that loss of this interaction prevents AvrRps4 from binding RRS1 WRKY domain and activating the RRS1/RPS4 complex.

SPR R_{max} analysis showed that whilst wild type AvrRps4$_c$ could bind RRS1 WRKY$_{E1195-T1273}$ or WRKY Dom6S$_{E1195-C1290}$, the binding of single mutant AvrRps4$_c$ E187A and double mutant E187A/E175A was significantly reduced. Interestingly, AvrRps4$_c$ E187A/E175A showed lower affinity for RRS1 WRKY$_{E1995-T1273}$ than single mutant E187A in R_{max} experiments. This mirrors *in planta* coimmunoprecipitation data which showed a reduced and abolished ability of the E187A and E187A/E175A mutants, receptively, to disrupt RRS1 Dom4 and WRKY Dom6R interactions[153]. The quantitative R_{max} data from this study therefore supports the previously proposed model that binding of AvrRps4 to the WRKY domain of RRS1 disrupts interactions between RRS1 Dom4 and WRKY Dom6R. The data therefore supports that the inability of AvrRps4 E187A/E175A to activate RRS1/RPS4 caused by a failure to interact with the RRS1 WRKY domain and disrupt interaction between the WRKY and Dom4 region of RRS1.

Unfortunately, due to issues with data quality I was unable to attain K_D values for the interactions between RRS1 WRKY$_{E1195-T1273}$ /WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$. The fast on-off interaction between RRS1 WRKY$_{E1195-T1273}$ /WRKY Dom6S$_{E1195-C1290}$ protein and AvrRps4$_c$ was intriguing, and not expected given the stability of the complex as observed by analytical gel filtration. This difference is likely due to the setup of these two experimental assays. The protein complex is formed for analytical gel filtration analysis by incubating the proteins together for an extended period of time on ice, compared to SPR where the contact time is only seconds as the analyte is flowed over the ligand chip. Additionally, the buffer conditions in which the SPR and analytical gel filtration experiments for RRS1 WRKY Dom6S$_{E1195-C1290}$ interactions were conducted differ due to issues with non-specific binding. SPR experiments were conducted in a higher salt buffer containing 860 mM NaCl whilst analytical gel filtration buffer contained 500 mM NaCl. NaCl concentration for RRS1 WRKY$_{E1195-T1273}$ studies however was 500 mM NaCl in both SPR and analytical gel filtration studies. It is therefore likely that the observed differences in RRS1 WRKY-AvrRps4 complex stability are due to innate differences between the SPR and analytical gel filtration techniques.

The virulence mechanism of AvrRps4 in *A. thaliana* is not fully understood. Given the lack of apparent enzymatic activity of AvrRps4 it is hypothesised the effector may act as a

structural block in WRKY transcription factor DNA binding. Given the observation of RRS1 WRKY inability to bind AvrRps4's E187A or E187A/E175A it will be interesting for future work to understand if these glutamic acid residues play a role in the virulence AvrRps4 virulence function, for example in disrupting the ability of WRKY transcription factors to bind W-box DNA. I started to investigate this using ReDCaT SPR analysis. ReDCaT SPR allows you to gain quantitative binding data on a protein-DNA interaction. The aim of using this technique was to investigate the ability of AvrRps4 to disrupt RRS1 WRKY interactions with W-box DNA. Previous EMSA data suggested that AvrRps4 did not affect the ability of RRS1 to bind W-box DNA, but the requirement of the RRS1-R K1221 residue for both binding AvrRps4 and W-box DNA called this in to question. I hypothesised that the qualitative EMSA study may not have been sensitive enough to monitor transient or weaker changes AvrRps4 may have on the binding of DNA by RRS1. The work in this chapter has now identified conditions in which this assay may now be conducted. It will be interesting to see in future work if this assay does reveal DNA binding disruption abilities of AvrRps4, and how this is affected in E187A/E175A mutants.

The ultimate goal for the work in this chapter was to gain an atomic structure of the complex of RRS1 WRKY and AvrRps4$_c$ to allow further dissection of the structural basis of effector recognition. To this end, extensive crystallisation trials were setup with RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4$_c$ complex at 25^0C. The screens trialled were largely of the "sparse matrix" variety, to cover a large area of crystallisation space[244], and included screens such as ProPlex™, which have been specially formulated for crystallisation of protein complexes. However, these screens did not yield conditions in which crystals formed. As I had observed an increased stability in RRS1 WRKY Dom6S$_{E1195-C1290}$ and AvrRps4 complex formation during SPR experiments at lower temperatures, I decided to setup further crystallisation screens at 4^0C. From these lower temperature trials, I identified conditions in which small fine bundle like crystals formed. As these crystals were unsuitable for the collection of X-ray diffraction data, work is ongoing to optimise these conditions to produce larger crystals. As the Dom6S region of RRS1 is predicted to contain large regions of disordered structure, I hypothesised that removing these potentially destabilising regions to express RRS1 WRKY alone may improve our ability to gain a structure of the RRS1 WRKY-AvrRps4 interface. Further to this, *in planta* data has shown the Dom6S is not required for AvrRps4 binding and RRS1/RPS4 induction of ETI, so all functionally important interactions between AvrRps4 and the C-terminal region of RRS1 must be contained in the

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

WRKY domain alone. However, despite setting up multiple screens at both 25^0C and 4^0C, the RRS1 WRKY$_{E1195-T1273}$-AvrRps4$_c$ complex proved recalcitrant to crystallisation.

One hypothesis to explain why we have been unable to get crystals of the complex is that crystallisation conditions may be leading to dissociation of the complex. For example, as seen in RRS1 WRKY$_{E1195-T1273}$ -AvrRps4$_c$ complex screens where crystals of AvrRps4$_c$ were observed to be forming. One way to prevent this complex dissociation during crystallisation is to express the two interactors linked with a flexible peptide sequence as one protein construct. This technique was used to obtain structures of the RRS1 TIR-RPS4 TIR complex[87]. Based on this, I expressed a construct in which WRKY Dom6S$_{S1184-C1290}$ and AvrRps4$_c$ were linked using a flexible 'NAAIRS' sequence[241]. Unfortunately, despite trialling in multiple cell lines, soluble expression of this construct was not observed. Producing linked RRS1 WRKY and AvrRps4 still provides a promising avenue for future work to gain a complex of this NLR-effector interface, although it may require further extensive trialling with multiple solubility tags or linkers to identify suitable conditions.

In addition to trialling linked constructs of RRS1 WRKY and AvrRps4, further work could be done to identify suitable crystallisation conditions for this complex. For example, a thermal shift assay could be conducted with these proteins to identify buffer conditions which increase the stability of this NLR-effector complex. It has been shown that raising the melting point of a protein through the addition of buffer components which stabilise proteins increase the likelihood of protein crystal formation[245,246]. Using assay screens such as differential scanning fluorimetry for RRS1 WRKY and AvrRps4$_c$ protein could therefore identify buffer conditions to improve the stability of the complex and increase the likelihood of crystallisation. Alternatively, other methods of crystallisation could be also conducted such as hanging drop, microbatch or microdialysis, which may provide more suitable conditions for crystallizing this complex[247]. Beyond crystallography, small angle X-ray scattering (SAXS) analysis could be conducted with RRS1 WRKY-AvrRps4 complex. This would provide a molecular envelope for the complex in to which the published crystal structures of RRS1 WRKY[71] and AvrRps4$_c$[166] could be docked. This would allow us to gain structural information on this binding event without the need for crystallisation.

Given that more than 70% of WRKY transcription factors are implicated in defence[61,62], this family of proteins represents a likely target hub for effectors looking to manipulate host responses to pathogens. As such, understanding how different WRKY integrated domain

Gaining biochemical and structural insights into AvrRps4 recognition by RRS1

NLRs bind and activate upon effector recognition could provide vital insights for future engineering of synthetic NLRs which bind previously unrecognized WRKY targeting effectors. Previous *in planta* work suggests that RRS1 and RRS1B may activate their respective NLR immune receptor complexes by distinct mechanisms. Domain switching of WRKY domains between the RRS1 and RRS1B results in RPS4/B dependent autoactivity demonstrating the WRKY domains of these NLRs are not interchangeable. In addition, AvrRps4 recognition by RRS1B requires the WRKY and C-terminal Dom6B whilst RRS1 only requires the WRKY domain. Furthermore, truncation of both the WRKY and Dom6 region of RRS1 results in an RPS4 dependent autoactivity in the case of RRS1, but not RRS1B[153] (and unpublished data from Dr Yan Ma). These observations suggest that the two NLRs are undergoing different inter- and intra-molecular interaction changes upon binding AvrRps4. Therefore, should we be able to gain a structure of RRS1 WRKY and RRS1B WRKY in complex with AvrRps4$_c$ we could begin to look at the conserved features and, indeed differences, which enable these RRS1/B WRKY domains to perceive AvrRps4 despite only sharing 56% sequence identity. Structural information of these binding events would also aid our mechanistic understanding of how different integrated WRKY domain NLRs are activated and could help inform the future generation of synthetic NLRs. Having established a pipeline for purifying RRS1B WRKY$_{N1163-H1237}$ singularly and in complex with AvrRps4$_c$, we are now in a position to start working towards attaining these structures either through crystallography of SAXS.

It is important to have a thorough understanding of all the interaction surfaces between a given NLR and an effector to support and guide future engineering of synthetic NLRs with expanded effector recognition. Previous *in planta* work had observed that AvrRps4 could associate with protein from exons 1-5 of RRS1 (M1-K1189) which did not contain WRKY Dom6, though this interaction was significantly weaker than with the WRKY Dom6 region of RRS1[65]. This suggested that AvrRps4 could interact with RRS1 outside of the integrated WRKY domain. Follow up work suggested that the TIR domain of RRS1 could interact with RRS1 WRKY Dom6S, and that binding of AvrRps4 enhanced this interaction. However, the biological function of this interaction is still not understood (unpublished work of Dr Yan Ma and Dr Sung Huh). This observation is similar to work in the literature demonstrating the ability of AvrPia to bind to the *O. sativa* NLR RGA5 at sites additional to the integrated HMA/RATX1 domain. This led me to investigate whether I could recapitulate the binding event between RRS1 WRKY Dom6S, TIR and AvrRps4$_c$ in vitro such that this potential RRS1-AvrRps4 interaction surface could be further explored biochemically. However, I was not

able to recapitulate this binding event by mixing and incubating the proteins on ice, with AvrRps4$_c$ only observed to bind RRS1 WRKY Dom6S$_{E1195-C1290}$. As discussed in 5.5, the reason for the inability to recapitulate this binding event is not known understood. Hypothesis include a lack of an interaction partner either another domain of RRS1/RPS4 or external protein found in partner or that the conditions in the experimental setup were not appropriate for supporting binding. Future work to optimise binding conditions may lead to the ability to produce and study this multi domain complex in vitro and provide insights on a role for this interaction surface in RRS1/RPS4 activation. The observation that effectors may bind NLRs at multiple locations is also interesting from an evolutionary stand point and may be important to understand for the future engineering of synthetic NLRs.

General discussion & future perspectives

6

Crop losses from pest and disease pose a major threat to our global food security. With pressures on global agricultural outputs likely to rise due to climatic change and increasing demands of a growing population, plant pathogen-derived yield losses are unsustainable[248,249]. Only through understanding the complex interplay between plants and their pathogens can we learn to manipulate the interactions between these systems, and help tip the evolutionary balance towards crops.

The apparent convergent downstream signalling of NLRs across plant species highlight these receptors as a prime candidate for use in transgenic crops[250]. The power NLRs have to elevate disease resistance in transgenic crops is exemplified by the deployment of *Rpi-vnt1* from the wild relative of potato *Solanum venturii* in potato (*Solanum tuberosum*). *Rpi-vnt1* in transgenic lines conferred resistance to the oomycete pathogen *P. infestans,* the causal agent of late-blight disease, through recognition of the effector Avr-vnt1[251]. In addition to interfamily transfer of existing cloned NLRs[252,253], in principle there is huge potential for improving crop disease resistance through engineering synthetic NLRs capable of perceiving previously unrecognised pathogen effectors.

Integrated domain-containing NLRs are a prime candidate scaffold for engineering efforts. Using the design structure of integrated domain NLRs, receptors could be engineered to recognise pathogen effectors based on effector host targets. This strategy could confer recognition to pathogens which currently evade molecular detection by the plant immune system. However, such engineering strategies will not be as straightforward as simply swapping the integrated domains of NLRs for known effector targets. This is demonstrated by the autoactive phenotype when the WRKY domain of RRS1 is switched to other WRKY transcription factor domains[152]. Moreover, extensive understanding of the structural basis of NLR recognition of effectors and the NLR intra- and inter-molecular interactions which subsequently activate ETI will be needed to engineer NLRs successfully. In this manner we are now entering the structural age of NLR biology, where major breakthroughs in our

understanding and ability to translate laboratory results into agricultural outputs will likely rely on gaining structural insights into the functioning of NLRs.

6.1 Investigating expression systems for NLR protein production

Until recently, available structures of plant NLRs were limited to singular domains, primarily of the N-terminal CC or TIR domains and integrated domains, with prediction of full-length NLRs relying on inference from mammalian NLR structures. With the publishing of the CNL ZAR1 structure, we now have for the first time an atomic-level model for a full-length plant NLR. Whilst this structure provides significant insights into the activation and autoinhibition mechanism of NLRs, discussed in 1.4.1, ZAR1 only represents one method of effector perception, utilising an indirect-decoy effector recognition strategy. There is therefore a great incentive in the field for gaining structures of TNLs and CNLs which utilise various methods of effector detection.

A major bottleneck in attaining the structure of plant NLRs is the associated difficulties with expressing full-length NLR protein to yields appropriate for structural studies. This barrier has plagued both the animal and plant NLR field for decades. This study therefore set out to evaluate a diverse array of expression systems for the production of the NLRs RRS1 and RPS4, in order to identify conditions which supported soluble full-length expression of these plant NLRs. I evaluated various heterologous host species including prokaryotic *E. coli*, eukaryotic Sf9 insect cells as well as various plant-based expression systems including transgenic *A. thaliana* and agroinfiltrated *N. benthamiana*. In addition to trialling *in vivo* expression systems, cell-free wheat germ extract was also investigated for RRS1 expression capabilities. A comprehensive list of soluble expression results from the trials in this study can be found in **Appendix 5.**

These trials demonstrated that soluble expression of multidomain RRS1/RPS4 protein was reliant upon a plant-based folding environment. The exact capabilities which make plant-based expression systems suitable for soluble RRS1 expression but not other eukaryotes such as insect cells is not fully understood. Soluble expression conceivably requires a chaperone repertoire found only in plants or post-translational modifications which are specifically supported in plant systems. Interestingly unlike RRS1, expression of full-length ZAR1 was supported in Sf21 insect cells[108,109]. This suggests that different NLRs may have distinct folding requirements given the expression of NLRs varies between expressions

systems with varying protein folding capacities e.g. insect cell v plant expression systems. This result is particularly interesting in the case of RRS1 and ZAR1 given both these NLRs are found in *A. thaliana*. The differences in folding requirements might reflect the evolution of the two NLRs. It could be hypothesised that more ancient NLRs might require simplified folding machinery compared to more recently evolved NLRs. The basis of this hypothesis is that ancient NLRs will have been maintained in the genome through a longer period of evolution, which may select for less reliance on chaperones for proper protein folding. A criterion for identifying potential ancient NLRs is the presence of NLR orthologs across multiple genera. ZAR1 for example is one of the few Arabidopsis NLRs where a *N. benthamiana* orthologue has been defined suggesting it is one of the more ancient NLRs in the Arabidopsis NLR repertoire[254]. This hypothesis might therefore explain the ability to express soluble ZAR1 in insect cells but not RRS1.

Whilst expression of full-length RRS1 was observed in wheat germ cell-free extract and agroinfiltrated *N. benthamiana* leaves, the trials in this study highlighted transgenic *A. thaliana* plants as the most viable source of NLR protein in terms of scalability and cost. Using *A. thaliana* plants, I was able to purify RRS1 to quantities suitable for analytical gel filtration analysis, **Figure 6.1** & **Figure 4.10**. This is a major breakthrough in purifying RRS1 protein and signifies that with further scale up, quantities of RRS1 could be purified suitable for structural studies and other quantitative biochemical techniques such as SPR.

Current success rates for crystallography structure determination are still low with structural genomics projects estimating that less than 4% of expression targets result in a defined structure[255,256]. Given this low success rate and the high protein concentrations required for crystallography, successful structure determination of RRS1, and plant NLRs is general, is more likely to come from utilising techniques such as SAXS and cryo-electron microscopy. A crystalline sample is not required for SAXS analysis and provides the molecular envelope of a protein into which structures can then be docked. Using the known structures of RRS1's WRKY[71] and TIR domain[114] and modelling from ZAR1, a structure of RRS1 should be attainable in this manner. Cryo-electron microscopy has been utilised to solve the structure of the mammalian NLR NLRC4-NAIP inflammasome and recently ZAR1. Similarly to SAXS, by not relying on a crystalline sample cryo-electron microscopy overcomes a major hurdle in structure determination by traditional X-ray crystallography. Historically, cryo-electron microscopy use was restricted to macromolecules with a molecular weight in excess of 500 kDa and a resolution limit of ~5 Å. However, in recent

years the size limit of single-particle cryo-electron microscopy has been significantly lowered to ~65 kDa with a resolution of ~3 Å through the application of technologies which can enhance the image contrast for small protein complexes[257]. Given that 155 kDa RRS1 has been observed in this study and others to form a multimeric complex[218,233], determination of the structure of RRS1 should therefore be technically feasible using this method.

Whilst this study tested a diverse array of expression systems for RRS1 and RPS4 protein production, a number of alternatives still remain untested. In addition to plant cell cultures discussed in 4.5.1, other heterologous systems which could be investigated include prokaryotes such as *Bacillus subtilis*[258] and *Ralstonia eutropha*[259], and eukaryotic systems such as yeast[260], micro-algae[261] and mammalian cells[262]. Given that the results in this study suggest that multi-domain expression of RRS1 and RPS4 is not well supported in prokaryotes, future work would benefit from focussing on eukaryotic systems. In particular use of algae expression systems such as micro-algae *Chlamydomonas reinhardtii* hold strong potential for plant NLR production. As a plant-based system, although significantly more distantly related to *A. thaliana* than wheat or *N. benthamiana*, it could be hypothesised that micro-algae possess the appropriate chaperones and cellular machinery for plant NLR production. Micro-algae also offer several benefits over transgenic plants including a faster transformation to scale up production timeline, and the ability to grow the cells in a bioreactor[261]. As a single cell type, micro-algae also offers less variation in recombinant protein production compared to transgenic plants which could benefit downstream processing uniformity[261]. Similarly, as discussed in 4.5.1, plant cell cultures such as *N. tabacum* BY-2 or NT-1 cells[226] would offer similar benefits of batch consistency and control of protein production pipelines and would be worth exploring further for NLR protein production.

The insights presented in this extensive multi-system expression screen study should provide a strong foundation for future studies looking to express plant NLRs, **Figure 6.1** and **Appendix 5**. This study highlights the importance of protein folding environment for plant NLRs. The results in this study suggest that whilst some NLRs may be correctly folded in insect cells[108,109], others may strictly require a plant-based system for multi-domain expression. Transgenic plants were particularly highlighted as the most viable source of NLR protein for the case of RRS1 and RPS4 and likely other NLRs given the scalability, relative low costs and folding capabilities this system provides. However, what future studies are

likely to highlight is there is no single 'silver bullet' for expression of these receptors given the difficulties the field has faced in gaining structures of NLRs in recent decades. Nonetheless, with improvements in high-throughput expression testing strategies a technical advancement in structure resolution techniques such as cryo-electron microscopy, barriers to the determination of full-length NLR structures continue to fall.

Negative
- Low to no expression outside of RRS1 WRKY. No soluble multidomain constructs.
- Low to no soluble expression of multidomain constructs. Difficult scale up setup.

Positive
- Good expression of truncations focused on RRS1 WRKY and TIR domains
- Low-Medium expression of RRS1 and RPS4 truncations of domains of interest e.g. RPS4 CTD

E. coli · Sf9 cells

Non-plant based Heterologous expression systems

RRS1-R TIR — NB-ARC — LRR — Dom4 — WRKY — Dom6R

RPS4 TIR — NB-ARC — LRR — CTD

Plant based Heterologous expression systems

Wheat germ cell free · Transgenic *A. thaliana* · Transient expression *N. benthamiana*

- Expression of full length RRS1
- Good expression of full length RRS1 and RPS4
- Expression of full length RRS1 and RPS4 and RRS1B and RPS4B. Excellent affinity tag trailing capabilities

- Very low protein yields. High costs. Requires cDNA.
- Inefficient purification with anti-FLAG beads. Long plant generation pipeline for new lines.
- Labor intensive infiltration process

Figure 6.1 Summary of protein expression screening of RRS1 and RPS4 in this study. Figure outlines the main positives and negative outcomes of a range of heterologous and plant-based expression systems trialled for expression of RRS1 and RPS4 in this study

6.2 Understanding the structural basis of effector recognition by integrated domain NLRs

Whilst full-length structures are an important ultimate goal for the field of NLR biology, great value is still held in gaining the structure of NLR subdomains and particularly in understanding NLR-effector interfaces. Understanding the structural basis of NLR-effector binding events is key to deciphering the mechanism through which NLRs are activated and regulate plant defences. Such insights will provide crucial information for deciphering plant-pathogen interactions and help guide future engineering efforts to generate NLRs with expanded recognition capabilities.

The power structural biology brings to engineering integrated domain containing NLRs with expanded pathogen recognition capacity, has recently been demonstrated in a proof-of-concept study[263]. The study focussed on the recognition of *M. oryzae* AVR-Pik effector variants by the HMA integrated domain containing *O. sativa* NLR Pikp-1 and Pikm-1 alleles. Whilst the allele Pikm-1 confers recognition in its NLR pair to AVR-PikD/E/A variants, Pikp-1 only confers resistance to AVR-PikD. Through gaining the structure of Pikp-1 HMA domain complexed with AVR-PikD, and Pikm-1 HMA with AVR-PikD/E/A, the authors were able to highlight two key residues in Pikm-1 which facilitated recognition of AVR-PikE and AVR-PikA[70,181,263]. Mutation of these residues in Pikp-1 to match those found in Pikm-1, conferred Pikp-1 extended recognition of AVR-PikE and Avr-PikA[263]. This is a key proof-of-concept example of how structure-guided engineering can extend the effector recognition profile of an NLR.

The Pik/AVR-Pik systems provides insights into understanding the structural basis of NLR recognition of multiple allelic variations of a single effector. Effector recognition by RRS1 expands on this concept by providing a system to investigate how a single NLR binds structurally and mechanistically distinct effectors. To enable this investigation, a structure of RRS1 WRKY complexed with AvrRps4 is required to conduct comparative studies with the RRS1 WRKY-PopP2 complex[71]. The work presented in this study has now developed the tools to gain such a structure. In this project I have developed pipelines for the production of both RRS1 WRKY Dom6S$_{E1195-C1290}$ and RRS1 WRKY$_{E1195-T1273}$ protein. These proteins have now been shown to directly interact in vitro with AvrRps4$_c$, both qualitatively by analytical gel filtration analysis and in quantitative SPR studies. Through these binding investigations, I have begun to validate a protein-protein interaction model for the complex of RRS1 WRKY and AvrRps4$_c$ which highlighted two AvrRps4 residues, E175 and E187, as playing an

important role in facilitating recognition by the RRS1 WRKY domain. Future work continuing crystallisation trials or conducting SAXS analysis with these protein complexes will hopefully lead to the structural resolution of this NLR-effector interface.

Attaining a RRS1 WRKY-AvrRps4$_c$ structure would enable interesting comparative work with the RRS1 WRKY-PopP2 structure[71]. Given the apparent differences in the virulence mechanism of AvrRps4 and PopP2, it will be interesting to see what are the conserved and distinct interfaces these effectors establish with the RRS1 WRKY domain. From an evolutionary viewpoint, comparisons of the two different interfaces will provide insights into how the emergence of new plant NLR effector perception specificities arise. In addition to comparative work with PopP2, understanding the structural basis of AvrRps4 binding to RRS1 may shed light on questions arising from *in planta* work. For example, why does the mutant AvrRps4$_{KRVY/AAAA}$ retain the ability to bind RRS1 but not activate RRS1/RPS4 defence. As the KRVY motif region of AvrRps4 (K135-Y138) is not in the electron dense region of the AvrRps4$_c$ structure, we are unable to predict the interactions involved in this region[150]. However, when complexed with the RRS1 WRKY domain, this region of AvrRps4 may be stabilised and allow us to observe the intricacies of the interaction involving this motif.

Whilst the structure of RRS1 WRKY$_{E1195-T1273}$ in complex with PopP2 highlights a number of key binding interfaces[71], *in planta* work has shown that in the context of full-length RRS1, Dom6R is required to translate PopP2 binding into RRS1/RPS4 complex activation. It is vital therefore that we understand the interactions involving Dom6R with PopP2 and other domains within the RRS1/RPS4 complex. This will allow us to investigate how multiple pathogen effector recognition is facilitated by RRS1-R. As RRS1 Dom6R shows little homology to any other known protein domain, we are unable to conduct accurate protein-protein interaction modelling studies with RRS1 WRKY Dom6R. Therefore, work to understand the intricacies of interactions with Dom6R will require attaining a structure of this protein domain. Identification in this study of a construct and condition in which RRS1 WRKY Dom6R$_{E1209-Y1373}$ protein can be produced in insect cells, therefore opens an exciting new branch of RRS1-effector interaction investigations. Given the mechanistic importance of Dom6R in PopP2 activation of the RRS1/RPS4 complex, attaining structural information of the C-terminal extension could provide key insights into the mechanism which transduces PopP2 binding in to RPS4 dependent defence activation of RRS1-R but not RRS1-S. Understanding how different activation mechanisms are coordinated and maintained in

an NLR would provide invaluable understanding as to how we could synthetically design NLRs with multiple recognition specificities in the future.

Whilst comparing the structures of RRS1 WRKY with AvrRps4$_c$ and PopP2 enables investigation into how a single NLR binds distinct effectors, comparing the RRS1 and RRS1B WRKY interface with AvrRps4$_c$ allows us to question how a single effector is recognised by different NLRs. RRS1 and RRS1B share only 56% and 58% sequence identity across their WRKY and Dom6 regions respectively[141]. Attaining a structure of both RRS1 and RRS1B WRKY bound to AvrRps4 would therefore enable a comparison of how these distinct WRKY domains bind AvrRps4. Moreover, given that the AvrRps4 induced activation mechanisms of RRS1 and RRS1B appear to differ significantly[152], it will be interesting to see in future work if these differences are reflected in distinct interactions with AvrRps4. In this manner, comparison of these two systems would enable the investigation of how evolution has driven the emergence of distinct activation strategies for the same effector. Work in this study has developed the tools to start investigating these questions with the identification of soluble expression conditions of RRS1B WRKY$_{N1163-H1237}$. I have shown in this study that RRS1B WRKY$_{N1163-H1237}$ can directly interact with AvrRps4$_c$ in vitro and identified a pipeline for the production of RRS1B WRKY$_{N1163-H1273}$ -AvrRps4$_c$ complex for use in structural studies. With such tools in place, investigations can continue to understand AvrRps4$_c$ recognition by RRS1B at the atomic level.

WRKY transcription factors are widely involved in host immune responses[61,62]. Consequently, they represent a likely target for pathogen effectors to attenuate host defences. AvrRps4 and PopP2 are currently the only effectors identified to target WRKY transcription factors. However, the observation that WRKY domains are one of the mostly commonly found integrated domains in NLRs, implies there is a strong selection force driving integration of WRKY domains into NLRs. This is likely the result of effector manipulation of WRKY transcription factors. Should further WRKY transcription factor-targeting effectors be identified, attaining structural information of these effectors' interactions with WRKY transcription factors should be prioritised. Through such structures, we may be able to utilise structure-guided engineering strategies to expand the recognition profile of WRKY integrated domain containing NLRs such as RRS1. In doing so this work could be used to confer resistance to previously unrecognised pathogens.

The work in this study has developed key tools to begin to understand fundamental questions of NLR effector perception, namely how a single NLR can recognise multiple distinct effectors, and how a single effector is recognised by different NLRs. This work will contribute to a growing body of research into effector perception by NLRs which aims to utilise molecular understanding of plant-pathogen interactions to strengthen food security.

6.3   Summary and outlook

The work presented in this thesis aimed to develop tools and insights into the structural functioning of NLRs. Through extensive expression trials, pipelines have been identified for the production of RRS1 protein suitable for structural and biochemical studies. These trials additionally provide a valuable foundation for future studies looking to purify plant NLRs, **Figure 6.1** and **Appendix 5**. Furthermore, strategies have also been developed for the production of RRS1 WRKY, WRKY Dom6S and WRKY Dom6R as well as paralogous RRS1B WRKY domain protein, **Table 5.1**. As demonstrated in this study, this material has already begun to shed light on our understanding of effector perception by integrated domain containing NLRs. Future work is now in a position to utilise these tools to further tease apart the intricacies of multiple effector recognition by RRS1.

Whilst our models for plant NLR activation are constantly developing, key questions still remain; what are the intra- and inter-molecular interactions involved in NLR auto-regulation? How does effector binding relieve NLR autoinhibition? What role does nucleotide hydrolysis play in NLR activation? What is the significance of NLR oligomerisation in instigation of ETI, and how does this differ between the NLR N-terminal domain classes? Our ability to engineer synthetic NLRs to target previously unrecognized effectors, relies on gaining an in-depth understanding of molecular intricacies of NLR activity and deciphering the black boxes in our current thinking. Of course, whilst scientific generation of transgenic crops with an expanded NLR repertoire is scientifically possible, deployment of such plants for agricultural purposes still faces significant political and social hurdles. Therefore, until public opinion and government legislation towards the use of genetically engineered crops changes, the potential of utilising our understanding of NLRs will be significantly constrained.

It will be exciting to see how our understanding of NLRs evolves through future structural and biochemical studies. In gaining such knowledge we will take significant steps forward in

strengthening plant disease resistance with the potential to revolutionize modern crop disease control.

Appendix 1: List of primers used in this study

Name	Description	Cloning method	Sequence
E. coli and Sf9 screening			
HB 1	Forward RRS1-R TIR for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGACCAATTGTGAAAAGGATGAGG
HB 2	Reverse RRS1-R TIR for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCATCCAATTCGTCCAACATAAAAGTGC
HB 3	Forward RRS1-R NB-ARC for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGATCTATTCGAAGCTGCTGG
HB 4	Reverse RRS1-R NB-ARC for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCAGGTTCCACCCAAAGTTTCTG
HB 5	Forward RRS1-R LRR for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGAAGAACCTGGAGATGTTGAGGACG
HB 6	Reverse RRS1-R LRR for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCAAAGCTTCTCCGAGTCAGAACC
HB 7	Forward RRS1-R Dom4 for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGCCTATGCATTACAAGTTCAAC
HB 8	Reverse RRS1-R Dom 4 for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCACTTTTTGGTACGTCAGACAAATTCTCGG
HB 28	Forward RRS1-R Leucine zipper for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGTTGAGAGTCAGCTATGATGATTTACAGG
HB 9	Forward RRS1-R WRKY (E1209) for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGGAGGGGAGATCTATGGACTTGGC
YM261	Forward RRS1-R WRKY (S1184) for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGAGCGAAAGTAGGGTAAAG
HB 125	Forward RRS1-R WRKY (R1194) for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGCGCGAAAGTAAGGTAAAGAAAGTGG
HB 10	Reverse RRS1-R WRKY for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCAATGGTTATGCTCAGATAGGTAAGTAATAGC
YM262	Reverse RRS1-R Dom6S for _E. coli_ Sf9 screen cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCAGCAGATGGAGGAGGAAG

Name	Description	Cloning method	Sequence
YM259	Reverse RRS1-R Dom6R for *E. coli* Sf9 screen cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCAATAATCGAAGAATGTTGACCA
HB 77	Forward cCFP for *E. coli* Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGCCATGACAAGCAGAAGAACGG
HB 78	Reverse nVenus for *E. coli* Sf9 screen cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCACCCTGATGTTGTGGCGGATC
HB 14	Forward RPS4 TIR for *E. coli* Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGATGGAGACATCATCTATTTCC
HB 15	Reverse RPS4 TIR for *E. coli* Sf9 screen cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTTATCCCTCCGGTGGTATTCC
HB 16	Forward RPS4 NB-ARC for *E. coli* Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGAGTCACAACGCCGTCGTGGG
HB 17	Reverse RPS4 NB-ARC for *E. coli* Sf9 screen cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCAGTGTGTCCTTGTCACCCTCC
HB 18	Forward RPS4 LRR for *E. coli* Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGCCATGCTTAAGGTGGGTCGA
HB 19	Reverse RPS4 LRR for *E. coli* Sf9 screen cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCATGGAACTGATGTAAGACTCG
HB 20	Forward RPS4 CTD for *E. coli* Sf9 screen cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGGAGTTTCCACCAAATCTTCAGTGC
HB 132	Forward RPS4 CTD L954 cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGCTTGTTTCAGAATCTTTGTTCAGC
HB 21	Reverse RPS4 CTD for *E. coli* Sf9 screen	Infusion	ATGGTCTAGAAAGCTTCAGAAATTCTTAACCGTGTGCATGATCACTG
HB 133	Reverse RPS4 CTD L1120 cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTTAAAGAGCACTATTTTTGTCTTTATCC
T7	Forward T7 promoter pOPIN sequencing	Infusion	TAATACGACTCACTATAGGG
HB 56	Reverse pOPIN sequencing	Sequencing	TGATAGGCAGCCTGCACC
HB 42	Forward RRS1 Dom4 sequencing	Sequencing	TCTTAGCCGATGTGTCTCTC
RRS1 WRKY-NAAIRS linker-AvrRps4 construct generation			
HB 94	Forward RRS1 WRKY for cloning into Level 0 vector	Golden gate	AGGAAGACAGAATGTCTGACGTACCAAAAAAGGAGAAGAAACATCG

Name	Description	Cloning method	Sequence
HB 95	Reverse RRS1 Dom6S 'GSNAAIRS' for cloning into Level 0 vector	Golden gate	AGGAAGACAGTACTCCGGATTGCTGCGTTCGAACCGCAGATGGAGGAGG
HB 96	Forward 'NAAIRSGS' AvrRps4 CTD for cloning into Level 0 vector	Golden gate	AGGAAGACAGAGTAACGCAGCAATCCGGAGTGGTTCGGGTAAACGAGTCTATC
HB 97	Reverse AvrRps4 CTD for cloning into Level 0 vector	Golden gate	AGGAAGACAGCGAATTGGTTGATTCGCGGTCTTCG
AvrRps4 cloning			
HB 88	Forward AvrRps4 CTD cloning into pOPINA	Infusion	AGGAGATATACCATGGAGTTTCCACCAAATCTTCAGTGC
HB 89	Reverse AvrRps4 CTD cloning into pOPINA	Infusion	GTGGTGGTGGTGTTTGAAATTCTTAACCGTGCATGATCACTG
HB 113	Forward AvrRps4 CTD cloning into pOPINE	Infusion	AGGAGATATACCATGGTAAACGAGTCTATCAAATTGG
HB 114	Reverse AvrRps4 CTD cloning into pOPINE	Infusion	GTGATGGTGATGTTTTGGTTGATTCTGCGGTCTTCG
HB 115	Forward AvrRps4 CTD cloning into pOPINF	Infusion	AAGTTCTGTTTCAGGGCCCGGGTAAACGAGTCTATCAAATTGG
HB 116	Reverse AvrRps4 CTD cloning into pOPINF	Infusion	ATGGTCTAGAAAGCTTTATTGGTTGATTCTGCGGTCTTCG
RRS1 & RPS4 cDNA generation			
HB 23	Forward RRS1 TIR cloning into Level 0 vector	Golden gate	AGGAAGACAA AATG ACCAATTGTGAAAAGGATGAGGAATTCGTGTGC
HB 25	Reverse RRS1 Dom6R cloning into Level 0 vector	Golden gate	AGGAAGACAA CGAA CCATAATCGAAGAATGTTGACCAAGGG
HB 26	Forward RPS4 TIR cloning into Level 0 vector	Golden gate	AGGAAGACAA AATGGAGACATCATCTATTTCCACTGTGGAGG
HB 27	Reverse RPS4 CTD cloning into Level 0 vector	Golden gate	AGGAAGACAA CGAA CCGAAATTCTTAACCGTGTGCATGATCAC
ZD 581	Forward RPS4 exon 1 cloning into Level 0 vector	Golden gate	AGGAAGACAA AATG GAGACATCATCTATTTCCACTGTG
ZD 581	Reverse RPS4 exon 1 cloning into Level 0 vector	Golden gate	AGGAAGACAA CTTC CCAACAACTCAATGATACGAG
ZD	Forward RPS4 exon 2 cloning into Level 0 vector	Golden gate	AGGAAGACAA AAAC TGCCCGGAATTGGTAAAACCACAC
ZD	Reverse RPS4 exon 2 cloning into Level 0 vector	Golden gate	AGGAAGACAA TTTC ATTTTATTTTGCAATACATTG
ZD	Forward RPS4 exon 3 cloning into Level 0 vector	Golden gate	AGGAAGACAA GAAA GCTGCCAATGTTAGAGG
ZD 722	Reverse RPS4 exon 3 cloning into Level 0 vector	Golden gate	AGGAAGACAA TCCT TGTCACCCTCCCAAAGTTG
ZD 723	Forward RPS4 exon 4 cloning into Level 0 vector	Golden gate	AGGAAGACAA AGGA CACACCATGCTTAAGGTG
ZD 713	Reverse RPS4 exon 4 cloning into Level 0 vector	Golden gate	AGGAAGACAA GACC CCCATTGTAGCGTTTCCGA

Name	Description	Cloning method	Sequence
ZD 714	Forward RPS4 exon 5 cloning into Level 0 vector	Golden gate	AGGAAGACAA GGTC TTGTTTCAGAATCTTTGTTC
ZD 715	Reverse RPS4 exon 5 cloning into Level 0 vector	Golden gate	AGGAAGACAA CGAA CCGAAATTCTTAACCGTGTGCATG
Wheat germ expression vectors			
YM 159	Forward RRS1 TIR cloning into pEU	Golden Gate	CACCGGTCTCTAATGACCAATTGTGAAAAGGATG
HB 31	Forward RRS1 Dom4 cloning into pEU	Golden Gate	CACCGGTCTCTAATGTTGAATGCACATGG
HB 33	Forward RRS1 WRKY cloning into pEU	Golden Gate	CACGGTCTCTAATGGACGAGGGAGATCTATGG
YM 170	Reverse RRS1 Dom6R cloning into pEU	Golden Gate	TAGGTCTCACGAACCATAATCGAAGAATGTTGACCAAGG
HB 35	Reverse RRS1 Dom6S cloning into pEU	Golden Gate	TAGGTCTCACGAACCGCAGATGGAGGAGGAAGTGG
HB 44	Forward pEU vector sequencing	Sequencing	CTAACCACCTATCTACATCACC
HB 45	Reverse pEU vector sequencing	Sequencing	CGACCTGAGGTAATTATAACCC
SPR ReDCaT			
HB 57	SPR ReDCaT Chip Forward WT W-box 1 repeat	SPR	CGTTGACCG
HB 58	SPR ReDCaT Chip Reverse WT W-box 1 repeat	SPR	CGGTCAAGCCTACCCTAGTCCTCCTGC
HB 59	SPR ReDCaT Chip Forward EMSA mutant W-box 1 repeat	SPR	CGTAGACGG
HB 60	SPR ReDCaT Chip Reverse EMSA mutant W-box 1 repeat	SPR	CCGTCTACGCCTACCCTAGTCCTCCTGC
HB 61	SPR ReDCaT Chip Forward Literature mutant W-box 1 repeat	SPR	CGTTGCACGG
HB 62	SPR ReDCaT Chip Reverse Literature mutant W-box 1 repeat	SPR	CCGTGCAAGCCTACCCTACGTCCTCCTGC
HB 63	SPR ReDCaT Chip Forward Scrambled mutant W-box 1 repeat	SPR	TGATCGC
HB 64	SPR ReDCaT Chip Reverse Scrambled mutant W-box 1 repeat	SPR	GCGATCACCTACCCTAGTCCTCCTGC
HB 52	SPR ReDCaT Chip Forward WT W-box 3 repeats	SPR	CGTTGACCGTTGACCGAGTTGACTTTTA
HB 53	SPR ReDCaT Chip Reverse WT W-box 3 repeats	SPR	TAAAAAGTCAACTCGGTCAACGGTCAACGCTACCCTAGCGTCCTCCTGC
HB 54	SPR ReDCaT Chip Forward EMSA mutant W-box 3 repeats	SPR	CGTAGACGGTAGACGGAGTAGACGTTTTA
HB 55	SPR ReDCaT Chip Reverse EMSA mutant W-box 3 repeats	SPR	TAAAACGTCTACTCCGTCTACGCTCACGCTACCCTAGCGTCCTCCTGC

Name	Description	Cloning method	Sequence
HB 67	SPR ReDCaT Chip Forward Literature mutant W-box 3 repeats	SPR	CGTTGCACGGTTGCACGGTTGCACGG
HB 68	SPR ReDCaT Chip Reverse Literature mutant W-box 3 repeats	SPR	CCGTGCAACCGTGCAACCGTGCAACGCCTACCCTACGTCCTCCTGC
HB 65	SPR ReDCaT Chip Forward Scrambled mutant W-box 3 repeats	SPR	ATGCTTAACGTTCTACGTGTGCGTTACGT
HB 66	SPR ReDCaT Chip Reverse Scrambled mutant W-box 3 repeats	SPR	ACGTAACGCACACGTAGAACGTTAAGCATCCTACCCTACGTCCTCCTGC
RRS1 & RRS1B WRKY expression constructs			
HB 117	Forward RRS1B WRKY cloning into pOPINS3C/F/M	Infusion	AAGTTCTGTTTCAGGGCCCGAACAACAAGGAAAAGAGAGAGTGG
HB 118	Reverse RRS1B WRKY cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTTAATGGTTATGCTCAGAGATG
HB 108	Forward RRS1 WRKY E1195 cloning into pOPINS3C/F/M NP	Infusion	AAGTTCTGTTTCAGGGCCCGGAAAGTAAGGTAAAGAAAGTGG
HB 110	Reverse RRS1 WRKY cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCAAGTGGGCCGTGGATGGTTATGC
HB 47	Forward RRS1 WRKY S1184 cloning into pOPINS3C/F/M C8	Infusion	AAGTTCTGTTTCAGGGCCCGTCTGACGTACCAAAAAGG
HB 48	Reverse RRS1 WRKY cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCAAGCCTTGCGTTTAGTGG
YM 259	Reverse RRS1-R Dom6R for *E. coli* Sf9 screen cloning into pOPINS3C/F/M	Infusion	ATGGTCTAGAAAGCTTCAATAATCGAAGAATGTTGACCA

Appendix 2: Heterologous Protein Expression Vector Maps

pOPINF: Carbenicillin- Cleavable His6 tag

AmpR

lef2/ORF 603

CMV Enhancer

pOPINF
5531 bp

Chicken ß-Actin Promote

pUC origin of replication

lac Operator

T7 Promoter

p10 Promoter+5UTR

N-His-(partial) 3C tag

ORF 1629

lacZ promoter and gene insert

T7 Terminator

Rabbit ß-Globin Poly A Site

*Msc*I

*Kpn*I

*Hind*III
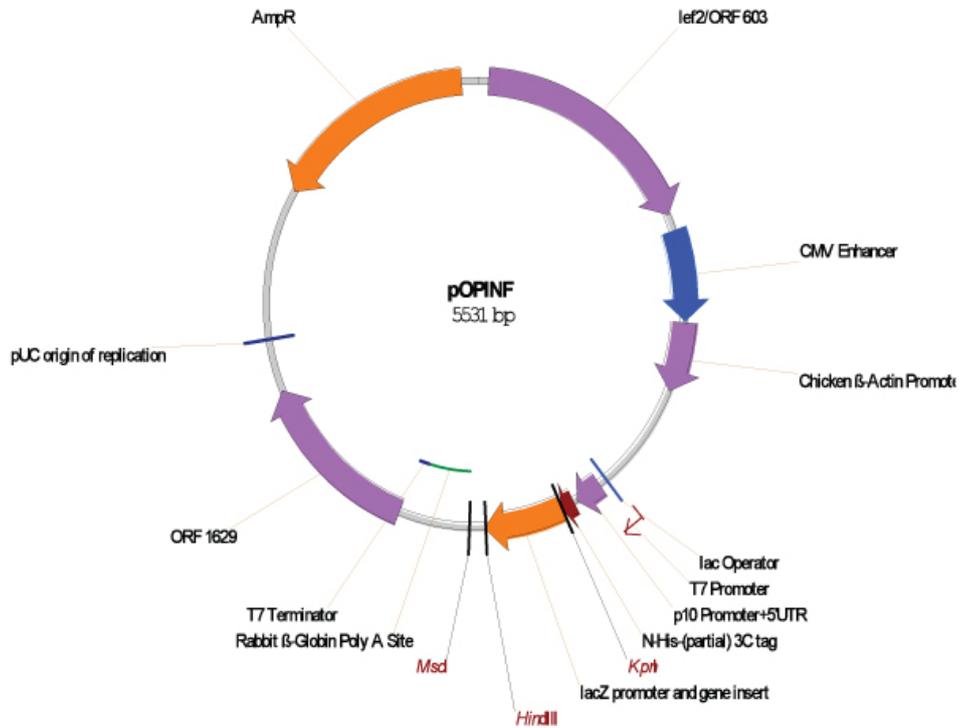
pOPINM: Carbenicillin- Cleavable His6-MBP

lef2, ORF603

CMV Enhancer

Chicken ß-Actin Promo

pUC origin

pOPINM
6665 bp

lac Operator

T7 Promoter

p10 Promoter plus 5'UTR

ORF 1629

T7 Terminator

Rabbit ß-Globin polyA site

*Msc*I (3925)

*Pme*I (3850)

*Hind*III (3832)

lacZ promoter and gene insert

*Kpn*I (3499)

MAHHHHHHSSG-MBP-SSG-3C tag
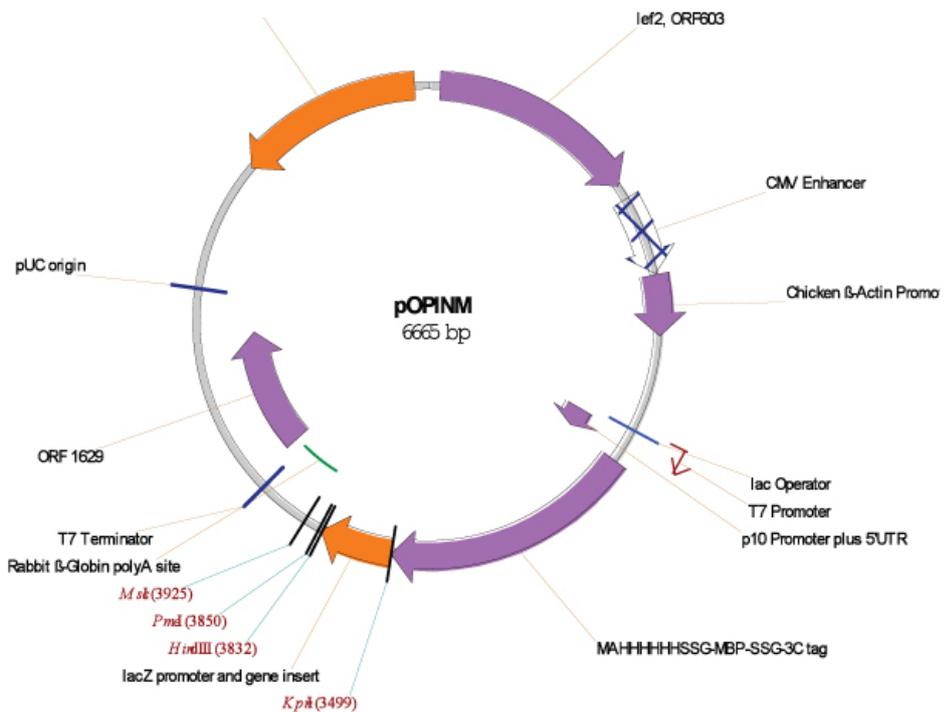
pOPINS3C: Carbenicillin- Cleavable His6-SUMO tag
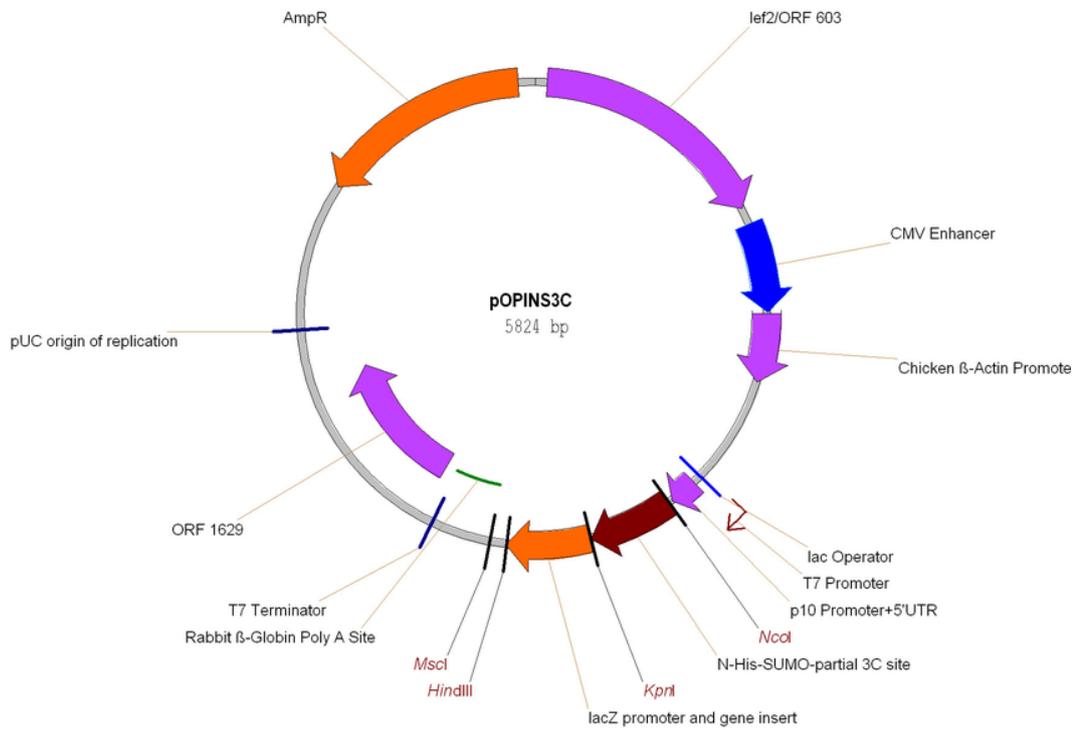
pOPINA: Kanamycin- Untagged (reliant on plasmid digestion)
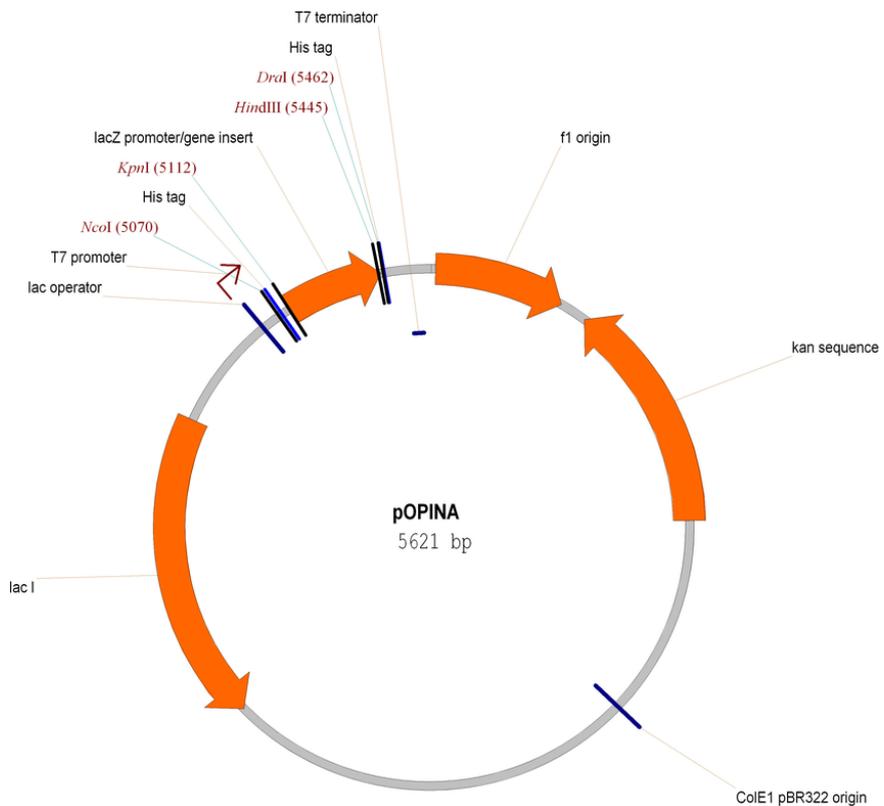
pOPINE: Carbenicillin- Non-cleavable His6 tag
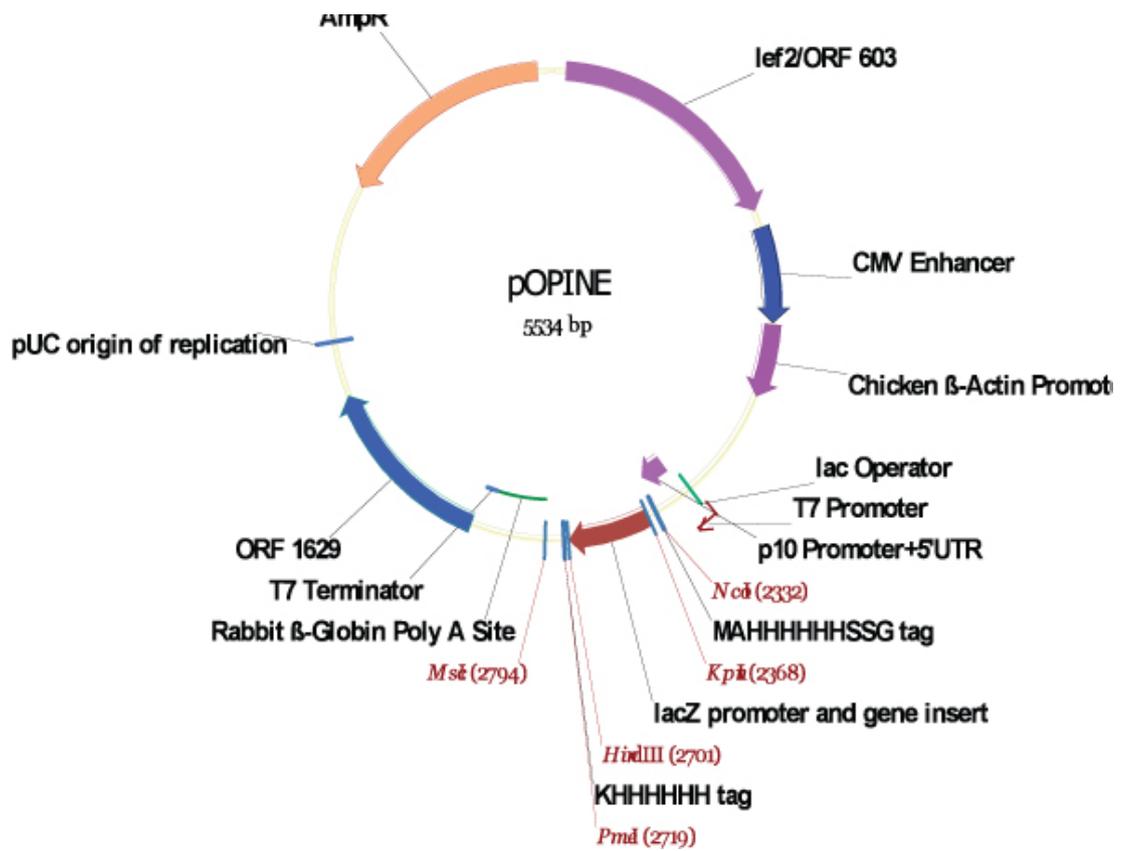
pEU-E01: Carbenicillin- Wheat Germ expression vector
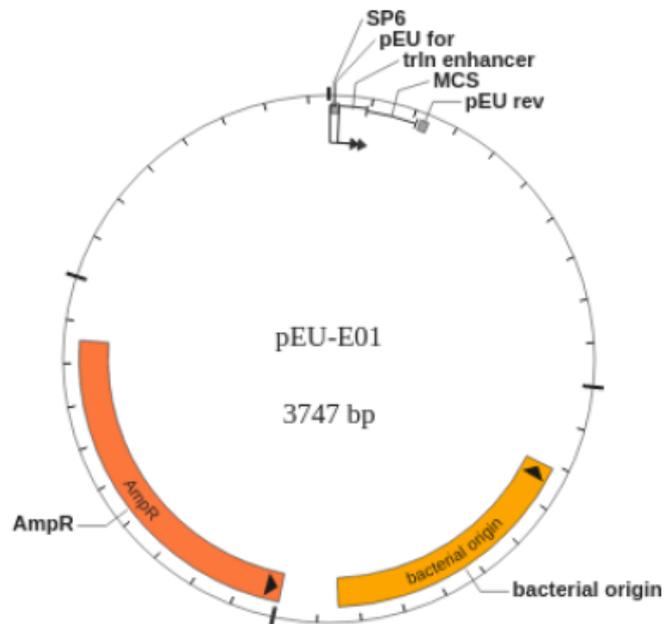
RRS1-HF cDNA pICSL86977: Kanamycin- plant expression vector
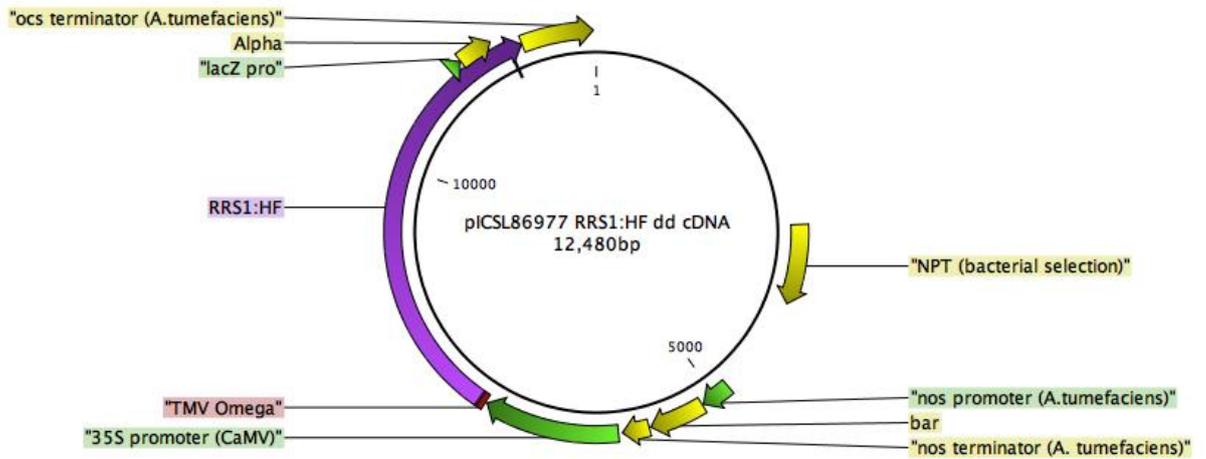
pMCSG7: Carbenicillin- *E. coli* expression vector TEV cleavable His6 tag expression
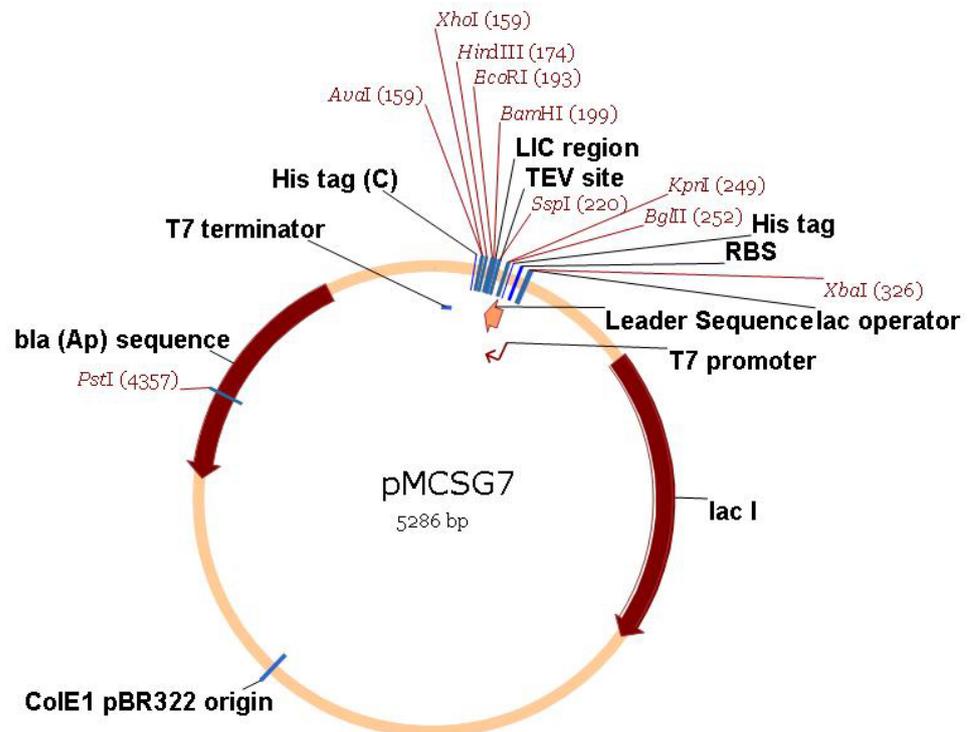
pGreenII-0229: Kanamycin- plant expression vector, used for 35S::RPS4-HS
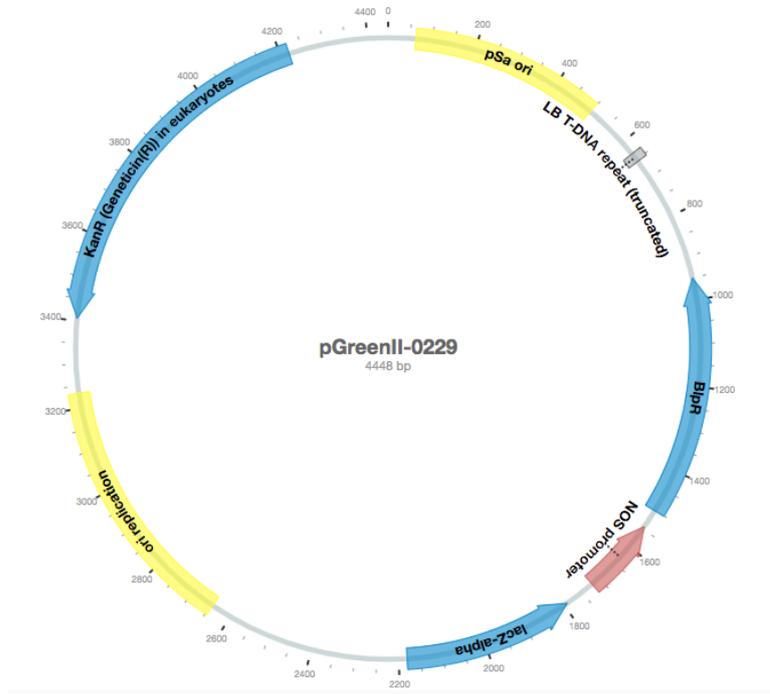
pICH86988: Kanamycin- plant expression vector, used for 35S::RRS1-HF
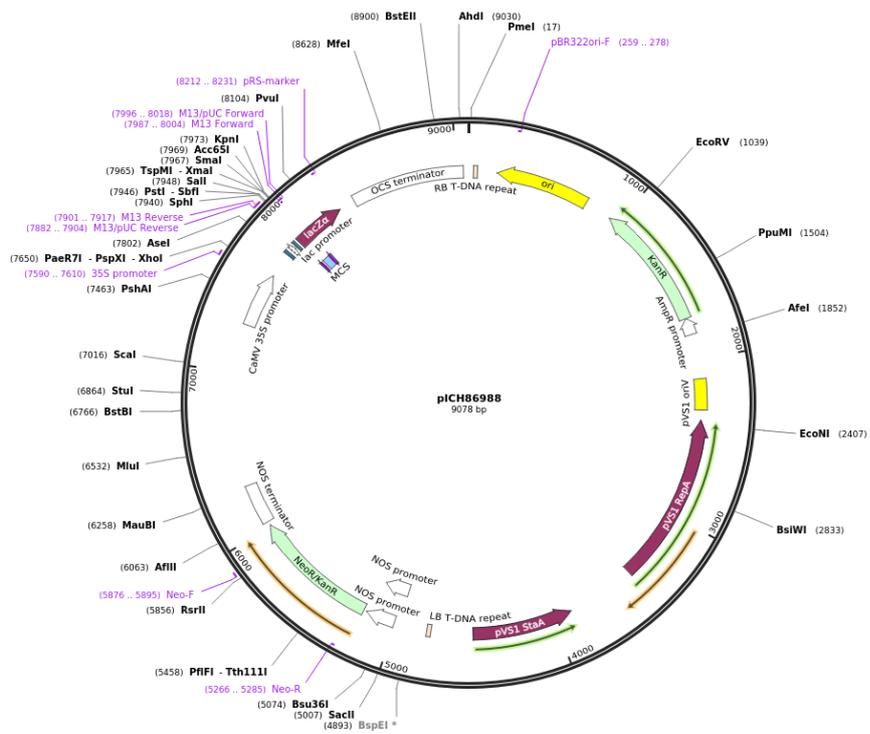
pICH86988: Carbenicillin- plant expression vector, used for pAt3::RPS4-HA/

pAT2::RRS1^{Ws-2}-HF LexA::AvrRps4-mNeon
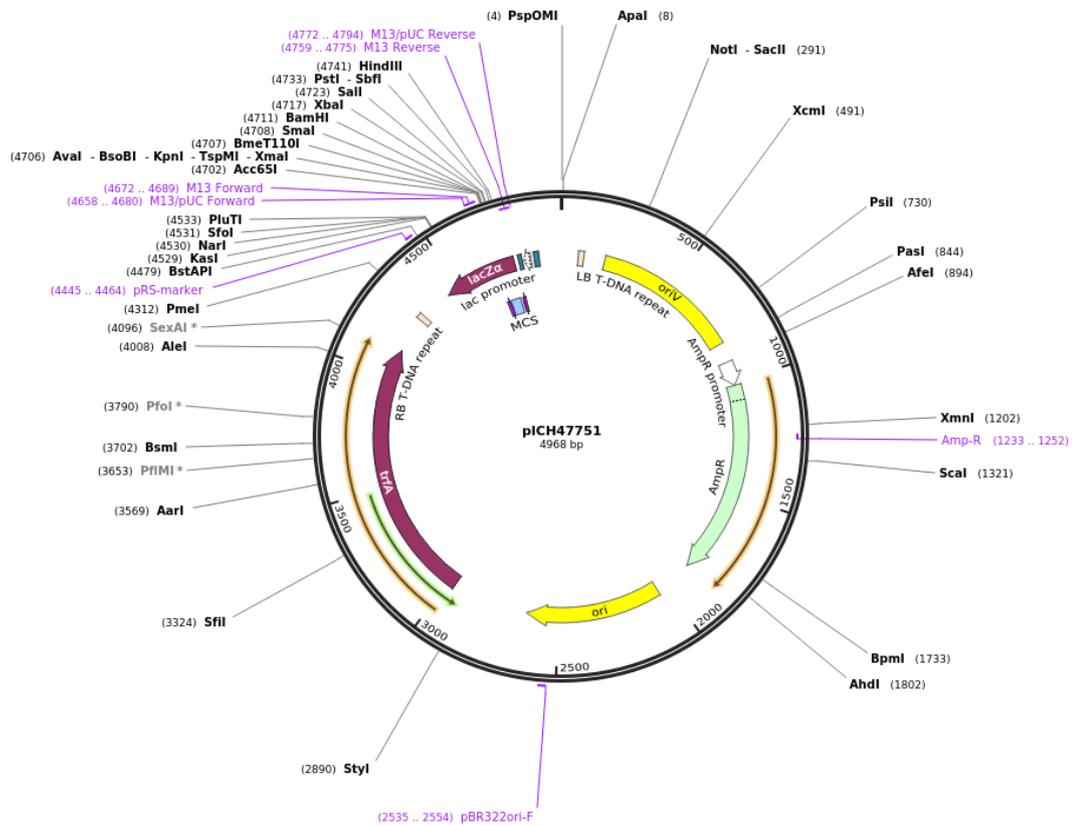
Appendix 3: Analytical Gel filtration calibration curves

Molecular Weight Calibration curve: HiLoad Superdex 75 26/60

$$logMW= (-1.241 \times Ve)/110.7 + 6.3908$$

$$y = -1.241x + 6.3908$$
$$R^2 = 0.9955$$
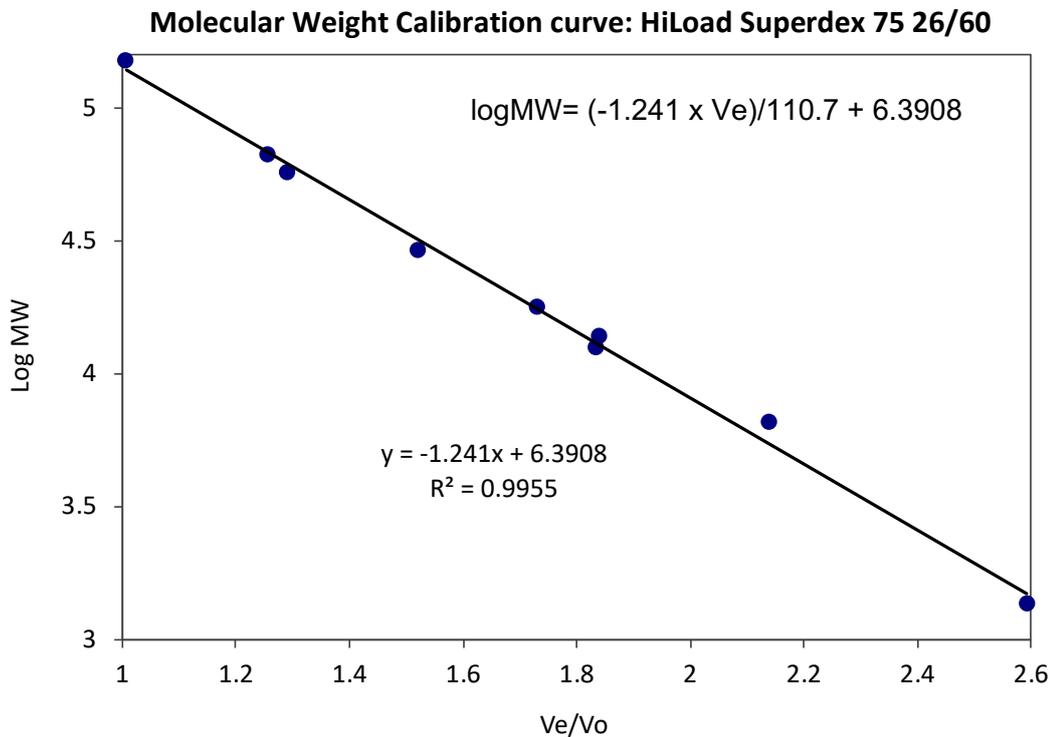
Curve generated by Dr John Steele. Protein run were: Vitamin B12 (1.4 kDa), Aprotinin (6.5 kDa), Cytochrome C (12.4 kDa), RNAse A (13.7 kDa), Myoglobin (17.6 kDa), Carbonic anhydrase (29 kDa), HPLF+7 monomer (56.9 kDa) , BSA (66 kDa), Alcohol dehydrogenases (150 kDa)

Molecular Weight Calibration curve: Superose™ 6 Increase column

$$Log\ MW = (-1.5737 \times Ve)/6.73 + 9.1513$$
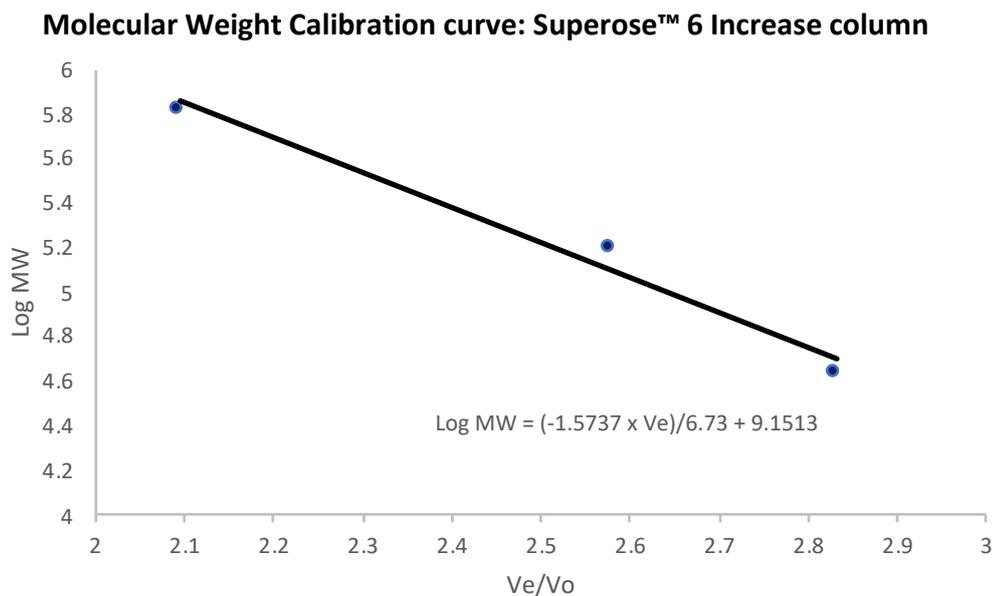
Curve generated by Dr Abbas Maqbool. Protein run were: Ovalbumin (42.7 kDa), Aldolase (156.8 kDa) and Thyroglobulin (660 kDa)

Appendix 4: Table of cloned plasmids used in this study

Gene	Name	Amino Acids	Vector	Purpose	Origin/Cloned by	Reference
RRS1	RRS1 TIR-Dom6R	M1-Y1373	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 NB-ARC	I154-T595	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 NB-ARC-Dom6S	I154-C1290	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 NB-ARC-Dom6S	I154-C1290	pOPINS3C	OPPF expression screening	Hannah Brown	
RRS1	RRS1 NB-ARC-Dom6R	I154-Y1373	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 NB-ARC-Dom6R	I154-Y1373	pOPIS3C	OPPF expression screening	Hannah Brown	
RRS1	RRS1 NB-ARC-LRR	I154-T595	pOPINS3C	OPPF expression screening	Hannah Brown	
RRS1	RRS1 LRR	K596-L867	pOPINS3C	OPPF expression screening	Hannah Brown	
RRS1	RRS1 LRR-Dom4	K596-K1189	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 LRR-Dom6R	K596-Y1373	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 LRR-Dom6R	K596-Y1373	pOPINS3C	OPPF expression screening	Hannah Brown	
RRS1	RRS1 Leucine Zipper motif-Dom6S	L1089-C1290	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 Leucine Zipper motif-Dom6R	L1089-Y1373	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 Leucine Zipper motif-Dom6R	L1089-Y1373	pOPINS3C	OPPF expression screening	Hannah Brown	
RRS1	RRS1 Dom4	P868-K1189	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 Dom4	P868-K1189	pOPINS3C	OPPF expression screening	Hannah Brown	

Gene	Name	Amino Acids	Vector	Purpose	Origin/Cloned by	Reference
RRS1	RRS1 TIR-Dom6R	M1-Y1373	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 NB-ARC	I154-T595	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 NB-ARC-Dom6S	I154-C1290	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 NB-ARC-Dom6S	I154-C1290	pOPINS3C	OPPF expression screening	Hannah Brown	
RRS1	RRS1 NB-ARC-Dom6R	I154-Y1373	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 NB-ARC-Dom6R	I154-Y1373	pOPIS3C	OPPF expression screening	Hannah Brown	
RRS1	RRS1 NB-ARC-LRR	I154-T595	pOPINS3C	OPPF expression screening	Hannah Brown	
RRS1	RRS1 LRR	K596-L867	pOPINS3C	OPPF expression screening	Hannah Brown	
RRS1	RRS1 LRR-Dom4	K596-K1189	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 LRR-Dom6R	K596-Y1373	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 LRR-Dom6R	K596-Y1373	pOPINS3C	OPPF expression screening	Hannah Brown	
RRS1	RRS1 Leucine Zipper motif-Dom6S	L1089-C1290	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 Leucine Zipper motif-Dom6R	L1089-Y1373	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 Leucine Zipper motif-Dom6R	L1089-Y1373	pOPINS3C	OPPF expression screening	Hannah Brown	
RRS1	RRS1 Dom4	P868-K1189	pOPINF	OPPF expression screening	Hannah Brown	
RRS1	RRS1 Dom4	P868-K1189	pOPINS3C	OPPF expression screening	Hannah Brown	

Gene	Name	Amino Acids	Vector	Purpose	Origin/Cloned by	Reference
RRS1 *slh1*	RRS1 slh1 WRKY-Dom6S	E1209-C1290	pOPINF	OPPF expression screening	Hannah Brown	
RPS4	RPS4 TIR-LRR	M1-P879	pOPINF	OPPF expression screening	Hannah Brown	
RPS4	RPS4 TIR-LRR	M1-P879	pOPINS3C	OPPF expression screening	Hannah Brown	
RPS4	RPS4 NB-ARC	S477-T659	pOPINF	OPPF expression screening	Hannah Brown	
RPS4	RPS4 NB-ARC	S477-T659	pOPINS3C	OPPF expression screening	Hannah Brown	
RPS4	RPS4 LRR-CTD	P660-F1217	pOPINS3C	OPPF expression screening	Hannah Brown	
RPS4	RPS4 CTD	E880-F1217	pOPINF	OPPF expression screening	Hannah Brown	
RPS4	RPS4 CTD	L954-L1120	pOPINF	OPPF expression screening	Lennart Wirthmueller	
AtWRKY 41	AtWRKY 41 WRKY	L121-E208	pOPINF	OPPF expression screening	Lennart Wirthmueller	
AtWRKY 18	AtWRKY 18 WRKY	A157-E240	pOPINF	OPPF expression screening	Lennart Wirthmueller	
AtWRKY 40	AtWRKY 40 WRKY	V124-N213	pOPINF	OPPF expression screening	Lennart Wirthmueller	
RRS1B	RRS1B WRKY	N1163- H1237	pOPINF	Production of RRS1B WRKY in *E. coli*	Hannah Brown	
RRS1B	RRS1B WRKY	N1163- H1237	pOPINM	Production of RRS1B WRKY in *E. coli*	Hannah Brown	
RRS1B	RRS1B WRKY	N1163- H1237	pOPINS3C	Production of RRS1B WRKY in *E. coli*	Hannah Brown	
RRS1	RRS1 WRKY	E1195-T1273	pOPINS3C	Production of RRS1 WRKY in *E. coli*	Hannah Brown	
RRS1	RRS1 WRKY Dom6S	E1195-C1290	pOPINS3C	Production of RRS1 WRKY Dom6S in *E. coli*	Hannah Brown	

Gene	Name	Amino Acids	Vector	Purpose	Origin/Cloned by	Reference
RRS1	RRS1 WRKY	E1195-T1273	pOPINE	Production of RRS1 WRKY in *E. coli* for SPR	Zhang et al, 2017 (recloned by Hannah Brown)	Zhang et al, 2017
RRS1	RRS1 WRKY Dom6S	E1195-C1290	pOPINE	Production of RRS1 WRKY Dom6S in *E. coli* for SPR	Hannah Brown	
AvrRps4	AvrRps4c	G134-Q221	pOPINF	Production of AvrRps4 in *E. coli*	Richard Hughes	Sohn et al, 2012
AvrRps4 E187A	AvrRps4c E187A	G134-Q221	pOPINF	Production of AvrRps4 in *E. coli*	Hannah Brown	
AvrRps4 E187A E175A	AvrRps4c E187A E175A	G134-Q221	pOPINF	Production of AvrRps4 in *E. coli*	Hannah Brown	
AvrRps4	AvrRps4c	G134-Q221	pOPINA	Production of AvrRps4 in *E. coli*	Lennart Wirthmueller	
RPS4	RPS4 gDNA	M1-F1217	pICSL86977	Expression of RPS4 protein in planta and cDNA generation	Yan Ma	Sarris et al, 2015
RRS1-R	RRS1-R gDNA	M1-Y1373	pICSL86977	Expression of RRS1-R protein in planta and cDNA generation	Yan Ma	Sarris et al, 2015
RRS1-R	RRS1-R cDNA	M1-Y1374	pICSL86977	Expression of RRS1-R protein in planta	Hannah Brown	
RRS1-S	RRS1-S gDNA	M1-C1288	pICSL86977	Expression of RRS1-S protein in planta	Yan Ma	Sarris et al, 2015
RRS1-S	RRS1-S cDNA	M1-C1288	pICSL86977	Expression of RRS1-S protein in planta	Hannah Brown	
RRS1B	RRS1B gDNA	M1-R1372	pICSL86977	Expression of RRS1B protein in planta	Yan Ma	
RRS1B	RRS1B cDNA	M1-R1372	pICSL86977	Expression of RRS1B protein in planta	Hannah Brown	

Gene	Name	Amino Acids	Vector	Purpose	Origin/Cloned by	Reference
RRS1/RPS4/AvrRps4	pAt3::RPS4-HA/ pAT2::RRS1 $^{Ws-2-}$HF/LexA::AvrRps4-mNeon	RRS1: M1-Y1374, RPS4: M1-F1217, AvrRps4: M1-Q221	pICH47751	Generation of stable transformant of inducible AvrRps4 A. thaliana lines	Pingtao Ding/ Bruno Ngou/ Billy Tasker Brown	Ngou et al, 2020
RRS1/RPS4/AvrRps4	pAt3::RPS4-HA/ pAT2::RRS1 Ws-2-HF/LexA::AvrRps4$_{KRVY135-138AAA}$-mNeon	RRS1: M1-Y1374, RPS4: M1-F1217, AvrRps4: M1-Q221	pICH47751	Generation of stable transformant of inducible AvrRps4 A. thaliana lines	Pingtao Ding/ Bruno Ngou/ Billy Tasker Brown	Ngou et al, 2021
RRS1/RPS4/AvrRps4	35S::RPS4-HS/ 35S::RRS1 $^{Ws-2}$-HF	RRS1: M1-Y1374, RPS4: M1-F1217	35S::RPS4-HS vector pGreenII 0229, 35S::RRS1 Ws-2-HF vector pICH86988	Generation of stable transformant of A. thaliana lines overexpressing RRS1 and RPS4	Pingtao Ding	Huh et al, 2017
RRS1	RRS1 TIR	K6-G153	pMCSG7	Production of RRS1 TIR in E. coli	Williams et al, 2014	Williams et al, 2014

Appendix 5: Results of protein expression trials

Table lists succesfulf soluble expresion attempts of RRS1 and RPS4 protein in all expression systems trialled in this study

E. coli

Gene	Protein/Domain	Amino acids	Vector	Media	Expression line	Expression Level (Expression scale)	Comments
RRS1	NB-ARC	I154-T595	pOPINF	AIM	Rosetta	Very low (small scale)	OPPF trials
	NB-ARC-LRR	I154-T595	pOPINS3C	AIM	Rosetta	Very Low (small scale)	OPPF trials: Only ~36KDa truncation expressed
	Leucine zipper motif-Dom6R	L1089-Y1373	pOPINF	AIM	Rosetta	Very low (small scale)	OPPF trials
	Dom4	P868-K1189	pOPINF	AIM	Rosetta	High (small scale)	OPPF trials: Only ~29KDa truncation expressed. Failed to scale up well.
	WRKY-Dom6R	E1209-Y1373	pOPINM	AIM	Lemo	Very low (small scale)	OPPF trials
RRS1	RRS1 WRKY	E1195-T1273	pOPINF	LB	Rosetta	Very High	Suitable for scale up production
RRS1	RRS1 WRKY Dom6S	E1195-C1290	pOPINF	LB	Rosetta	Very High	Suitable for scale up production
RRS1B	RRS1B WRKY	N1163-H1237	pOPINF	LB	Rosetta	Very High	Suitable for scale up production

Gene	Protein/Domain	Amino acids	Vector	Virus Titre	Expression Level	Comment
Sf9 Insect cell						
RRS1	NB-ARC-LRR	I154-T595	pOPINS3C	3uL	Low	OPPF trials
RRS1	Dom4	P868-K1189	pOPINF	30uL	Low	OPPF trials: ~28KDa truncation, mass spec predicts P868-K1101
RRS1	Dom4	P868-K1189	pOPINS3C	3uL	Low	OPPF trials: Mass spectrometry full peptide coverage
RRS1	WRKY-D6R	E1209-Y1373	pOPINM	3 & 30uL	Medium	OPPF trials
RPS4	CTD	L954-L1120	pOPINF	3uL	Medium	OPPF trials: Near full mass spec coverage by mass spec
AtWRKY 18	WRKY	A157-E240	pOPINF	3uL	High	OPPF trials: Mass spectrometry full peptide coverage
AtWRKY 40	WRKY	V124-N213	pOPINF	3uL	High	OPPF trials: Mass spectrometry full peptide coverage

Transgenic *A. thaliana*

Gene	Protein/Domain	Amino acids	Vector	Plant host species /Genotype	Expression Level	Comment
RRS1	RRS1:HF	M1-Y1373	35S::RPS4-HS vector pGreenII 0229, 35S::RRS1 Ws-2-HF vector pICH86988	35S::RPS4-HS/ 35S::RRS1 Ws-2-HF (A. thaliana)	High	Yield of proteins was limited by use of anti-FLAG/HA affinity beads which due to their non-rechargeable nature, place a cost limit on protein yield. Future work should investigate use of other purification pipelines in transgenic *A. thaliana*
RPS4	RPS4: HS	M1-F1217	35S::RPS4-HS vector pGreenII 0229, 35S::RRS1 Ws-2-HF vector pICH86988	35S::RPS4-HS/ 35S::RRS1 Ws-2-HF (A. thaliana)	High	
RRS1	RRS1:HF	M1-Y1373	pICH47751	pAt3::RPS4-HA/ pAT2::RRS1 Ws-2-HF/LexA::AvrRps4-mNeon (A. thaliana)	High	
RPS4	RPS4: HA	M1-F1217	pICH47751	pAt3::RPS4-HA/ pAT2::RRS1 Ws-2-HF/LexA::AvrRps4-mNeon (A. thaliana)	High	

Agrobacterium mediated transient expression in *N. benthamiana*

Gene	Protein/Domain	Amino acids	Vector	Plant host species /Genotype	Expression Level	Comment
RRS1	RRS1:HF	M1-Y1373	pICSL86977	*N. benthamiana*	Medium	
RPS4	RPS4: HA	M1-F1217	pICSL86977	*N. benthamiana*	Medium	
RRS1	RRS1 Dom4 WRKY Dom6S	P868-C1290	pICSL86977	*N. benthamiana*	Medium	
RRS1	RRS1 Dom4 WRKY Dom6R	P868-Y1373	pICSL86977	*N. benthamiana*	Medium	
RPS4	CTD	E880-F1217	pICSL86977	*N. benthamiana*	Medium	

Gene	Protein/ Domain	Amino acids	Vector	Plant host species /Genotype	Expression Level	
Wheat Germ Cell Free						
RRS1	RRS1:HF	M1-Y1374	pEU-GG	Wheat germ Cell free	Medium (very small scale)	Expression was observed but at such low levels that scaling the system up to quantities appropriate for biochemical and structural study is not within the capacity of this study.
RRS1	RRS1WRKY Dom6S	E1209-C1290	pEU-GG	Wheat germ Cell free	Medium (very small scale)	
RRS1	RRS1 Dom4 WRKY Dom6R	P868-Y1373	pEU-GG	Wheat germ Cell free	Medium (very small scale)	

Bibliography

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

2. Dodds, P. N. & Rathjen, J. P. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* **11,** 539–548 (2010).

3. Couto, D. & Zipfel, C. Regulation of pattern recognition receptor signalling in plants. *Nat. Rev. Immunol.* **16,** 537 (2016).

4. Cui, H., Tsuda, K. & Parker, J. E. Effector-Triggered Immunity : From Pathogen Perception to Robust Defense. *Annu. Rev. Plant Biol* **66625,** 1–25 (2015).

5. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444,** 323–329 (2006).

6. Cook, D. E., Mesarich, C. H. & Thomma, B. P. H. J. Understanding Plant Immunity as a Surveillance System to Detect Invasion. *Annu. Rev. Phytopathol* (2015). doi:10.1146/annurev-phyto-080614-120114

7. Thomma, B. P. H. J., Nu, T. & Joosten, M. H. A. J. Of PAMPs and Effectors : The Blurred PTI-ETI Dichotomy. *Plant Cell* **23,** 4–15 (2011).

8. Jones, D. A., Thomas, C. M., Hammond-Kosack, K. E., Balint-Kurti, P. J. & Jones, J. D. Isolation of the tomato Cf-9 gene for resistance to Cladosporium fulvum by transposon tagging. *Science (80-.).* **266,** 789–793 (1994).

9. Romeis, T. *et al.* Rapid Avr9-and Cf-9–dependent activation of MAP kinases in tobacco cell cultures and leaves: Convergence of resistance gene, elicitor, wound, and salicylate responses. *Plant Cell* **11,** 273–287 (1999).

10. Burgh, A. M. Van Der & Joosten, M. H. A. J. Plant Immunity : Thinking Outside and Inside the Box. *Trends Plant Sci.* **24,** 587–601 (2019).

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

12. Gust, A. A. *et al.* Bacteria-derived Peptidoglycans Constitute Pathogen-associated Molecular Patterns Triggering Innate Immunity in Arabidopsis * □. *J. Biol. Chem.* **282,** 32338–32348 (2007).

13. Kunze, G. *et al.* The N Terminus of Bacterial Elongation Factor Tu Elicits Innate Immunity in Arabidopsis Plants. *Plant Cell* **16,** 3496–3507 (2004).

14. Kaku, H. *et al.* Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor. *Proc. Natl. Acad. Sci.* 1–6 (2006).

15. Fliegmann, J., Mitho, A. & Wanner, G. An Ancient Enzyme Domain Hidden in the Putative ^N_L -Glucan Elicitor Receptor of Soybean May Play an Active Part in the Perception of Pathogen-associated Molecular Patterns during Broad Host Resistance *. *J. Biol. Chem.* **279,** 1132–1140 (2004).

16. Ferrari, S. *et al.* Oligogalacturonides : plant damage-associated molecular patterns and regulators of growth and development. *Front. Plant Sci.* **4,** 49 (2013).

17. Brutus, A., Sicilia, F., Macone, A., Cervone, F. & Lorenzo, G. De. A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. *Proc. Natl. Acad. Sci.* **1,** (2010).

18. Zipfel, C. Early molecular events in PAMP-triggered immunity. *Curr. Opin. Plant Biol.* **12,** 414–420 (2009).

19. Kadota, Y. *et al.* Direct regulation of the NADPH oxidase RBOHD by the PRR-associated kinase BIK1 during plant immunity. *Mol. Cell* **54,** 43–55 (2014).

20. Li, L. *et al.* The FLS2-associated kinase BIK1 directly phosphorylates the NADPH oxidase RbohD to control plant immunity. *Cell Host Microbe* **15,** 329–338 (2014).

21. Geesey, G. G. *et al.* Type III Secretion Machines : Bacterial Devices for Protein Delivery into Host Cells. *Science (80-.).* **284,** (1999).

22. Whisson, S. C. *et al.* A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* **450,** 115 (2007).

23. Davis, E. L., Hussey, R. S. & Baum, T. J. Getting to the roots of parasitism by nematodes. *Trends Parasitol.* **20,** 134–141 (2004).

24. Bolton, M. D., Yadeta, K. A., Baarlen, P. Van, Boeren, S. & Vervoort, J. The Cladosporium fulvum Virulence Protein Avr2 Inhibits Host Proteases Required for Basal Defense. *Plant Cell* **20,** 1948–1963 (2008).

25. Mukhtar, M. S. *et al.* in a Plant Immune System Network. *Science (80-.).* **333,** 596–601 (2012).

26. Xiang, T., Zong, N., Zou, Y. & Wu, Y. Pseudomonas syringae Effector AvrPto Blocks Innate Immunity by Targeting Receptor Kinases. *Curr. Biol.* 74–80 (2008). doi:10.1016/j.cub.2007.12.020

27. Shan, L. *et al.* Bacterial Effectors Target the Common Signaling Partner BAK1 to Disrupt Multiple MAMP Receptor- Signaling Complexes and Impede Plant Immunity. *Cell Host Microbe* 17–27 (2008). doi:10.1016/j.chom.2008.05.017

28.   Zhang, J. *et al.* Receptor-like Cytoplasmic Kinases Integrate Signaling from Multiple Plant Immune Receptors and Are Targeted by a Pseudomonas syringae Effector. *Cell Host Microbe* **7,** 290–301 (2010).

29.   Meyers, B. C., Kozik, A., Griego, A., Kuang, H. & Michelmore, R. W. Genome-Wide Analysis of NBS-LRR – Encoding Genes in Arabidopsis. *Plant Cell* **15,** 809–834 (2003).

30.   Qi, D., Innes, R. W. & Li, X. Recent advances in plant NLR structure, function, localization, and signaling. *Front. Immunol.* **4,** 1–10 (2013).

31.   Monteiro, F. & Nishimura, M. T. Structural, functional, and genomic diversity of plant NLR proteins: an evolved resource for rational engineering of plant immunity. *Annu. Rev. Phytopathol.* **56,** 243–267 (2018).

32.   Guo, Y. *et al.* Genome-Wide Comparison of Nucleotide-Binding Site-Leucine-Rich Repeat-Encoding Genes. *Plant Physiol.* **157,** 757–769 (2011).

33.   Jones, J. D., Vance, R. E., & Dangl, J. L. Intracellular innate immune surveillance devices in plants and animals. *Science (80-.).* **354,** aaf6395 (2016).

34.   Elinav, E., Strowig, T., Henao-Mejia, J. & Flavell, R. A. Regulation of the antimicrobial response by NLR proteins. *Immunity* **34,** 665–679 (2011).

35.   Zhang, L., Chen, S., Ruan, J., Wu, J., Tong, A.B., Yin, Q., Li, Y., David, L., Lu, A., Wang, W.L. and Marks, C. Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Innate Immun.* **350,** (2015).

36.   Bentham, A., Burdett, H., Anderson, P. A., Williams, S. J. & Kobe, B. Animal NLRs provide structural insights into plant NLR function. *Ann. Bot.* (2016). doi:10.1093/aob/mcw171

37.   Flor, H. H. Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* **9,** 275–296 (1971).

38.   Keen, N. T. Gene-for-gene complementarity in plant-pathogen interactions. *Annu. Rev. Genet.* **24,** 447–463 (1990).

39.   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).

40.   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.* (2006).

41.   Ravensdale, M. *et al.* Intramolecular Interaction Influences Binding of the Flax L5 and L6 Resistance Proteins to their AvrL567 Ligands. *PLOS Pathog.* **8,** (2012).

42. Axtell, M. J. & Staskawicz, B. J. Initiation of RPS2 -Specified Disease Resistance in Arabidopsis Is Coupled to the AvrRpt2-Directed Elimination of RIN4. *Cell* **112,** 369–377 (2003).

43. Axtell, M. J., Chisholm, S. T., Dahlbeck, D. & Staskawicz, B. J. Genetic and molecular evidence that the Pseudomonas syringae type III effector protein AvrRpt2 is a cysteine protease. *Mol. Microbiol.* **49,** 1537–1546 (2003).

44. Mackey, D. *et al.* RIN4 Interacts with Pseudomonas syringae Type III Effector Molecules and Is Required for RPM1-Mediated Resistance in Arabidopsis. *Cell* **108,** 743–754 (2002).

45. Chung, E. *et al.* Article 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 (2011).

46. Chung, E., El-kasmi, F., He, Y., Loehr, A. & Dangl, J. L. Article 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 (2014).

47. Hoorn, R. A. L. Van Der & Kamoun, S. From Guard to Decoy : A New Model for Perception of Plant Pathogen Effectors. *Plant Cell* **20,** 2009–2017 (2008).

48. Xing, W. *et al.* The structural basis for activation of plant immunity. *Nature* 243–247 (2007). doi:10.1038/nature06109

49. Ntoukakis, V., Saur, I. M. L., Conlan, B. & Rathjen, J. P. ScienceDirect The changing of the guard : the Pto / Prf receptor complex of tomato and pathogen recognition. *Curr. Opin. Plant Biol.* **20,** 69–74 (2014).

50. Marone, D., Russo, M. A., Laidò, G., Leonardis, A. M. De & Mastrangelo, A. M. Plant Nucleotide Binding Site – Leucine-Rich Repeat (NBS-LRR) Genes : Active Guardians in Host Defense Responses. *Int. J. Mol. Sci.* 7302–7326 (2013). doi:10.3390/ijms14047302

51. Deyoung, B. J., Qi, D., Kim, S., Burke, T. P. & Innes, R. W. Activation of a plant nucleotide binding-leucine rich repeat disease resistance protein by a modified self protein. *Cell Microbiol.* **14,** 1071–1084 (2013).

52. Sarris, P. F., Cevik, V., Dagdas, G., Jones, J. D. G. & Krasileva, K. V. Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biol.* (2016). doi:10.1186/s12915-016-0228-7

53. Kroj, T., Chanclud, E., Michel-romiti, C., Grand, X. & Morel, J. Integration of decoy

domains derived from protein targets of pathogen effectors into plant immune receptors is widespread. *New Phytol.* (2016).

54. Cesari, S. *et al.* A novel conserved mechanism for plant NLR protein pairs : the ' integrated decoy ' hypothesis. *Front. Plant Sci.* **5,** 1–10 (2014).

55. Germain, H. & Se, A. Tansley review Innate immunity : has poplar made its. *New Phytol.* 678–687 (2011).

56. Wu, C., Krasileva, K. V, Banfield, M. J. & Terauchi, R. The ' sensor domains ' of plant NLR proteins : more than decoys ? *Front. Plant Sci.* **6,** 5–7 (2015).

57. Parker, J. E., Deslandes, L., Parker, J. E. & Deslandes, L. Article A Receptor Pair with an Integrated Decoy Converts Pathogen Disabling of Transcription Factors to Article A Receptor Pair with an Integrated Decoy Converts Pathogen Disabling. 1074–1088 (2015). doi:10.1016/j.cell.2015.04.025

58. Sarris, P. F. *et al.* A Plant Immune Receptor Detects Pathogen Effectors that Target WRKY Transcription Factors Article A Plant Immune Receptor Detects Pathogen Effectors that Target WRKY Transcription Factors. *Cell* **161,** 1089–1100 (2015).

59. Wu, C. H., Belhaj, K., Bozkurt, T. O., & 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.* **15479,** (2015).

60. Schwessinger, B., Bart, R., Krasileva, K. V & Coaker, G. Focus issue on plant immunity : from model systems to crop species. *Front. Plant Sci.* **6,** 2014–2016 (2015).

61. Dong, J., Chen, C. & Chen, Z. Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. *Plant Mol. Biol.* **51,** 21–37 (2003).

62. Chi, Y. *et al.* Protein – Protein Interactions in the Regulation of WRKY Transcription Factors. *Mol. Plant* **6,** 287–300 (2013).

63. Macho, A. P. & Zipfel, C. Plant PRRs and the Activation of Innate Immune Signaling. *Mol. Cell* **54,** 263–272 (2014).

64. Wang, X. *et al.* The rpg4 -Mediated Resistance to Wheat Stem Rust (Puccinia graminis) in Barley (Hordeum vulgare) Requires Rpg5 , a Second NBS-LRR Gene , and an Actin Depolymerization Factor. *Mol. plant-microbe Interact.* **26,** 407–418 (2013).

65. Sarris, P. F. *et al.* A Plant Immune Receptor Detects Pathogen Effectors that Target WRKY Transcription Factors. *Cell* **161,** 1089–1100 (2015).

66. Le Roux, C., Huet, G., Jauneau, A., Camborde, L., Trémousaygue, D., Kraut, A., Zhou, B., Levaillant, M., Adachi, H., Yoshioka, H. and Raffaele, S. A Receptor Pair with an

Integrated Decoy Converts Pathogen Disabling of Transcription Factors to Article A Receptor Pair with an Integrated Decoy Converts Pathogen Disabling. *Cell* **161,** 1074–1088 (2015).

67. Yoshimura, Satomi, Utako Yamanouchi, Yuichi Katayose, Seiichi Toki, Zi-Xuan Wang, Izumi Kono, Nori Kurata, Masahiro Yano, Nobuo Iwata, and T. S. Expression of Xa1 , a bacterial blight-resistance gene in rice , is. *Proc. Natl. Acad. Sci.* **95,** 1663–1668 (1998).

68. Ashikawa, I. *et al.* Two Adjacent Nucleotide-Binding Site–Leucine-Rich Repeat Class Genes Are Required to Confer Pikm -Specific Rice Blast Resistance. *Genet. Soc. Am.* **2,** (2008).

69. 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).

70. Maqbool, A. *et al.* Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *Elife* (2015). doi:10.7554/eLife.08709

71. Zhang, Z. *et al.* Mechanism of host substrate acetylation by a YopJ family effector. *Nat. plants* **3,** 17115 (2017).

72. Ortiz, D., Guillen, K. De, Cesari, S., Kroj, T. & Structurale, C. D. B. Recognition of the Magnaporthe oryzae Effector AVR-Pia by the Decoy Domain of the Rice NLR Immune Receptor RGA5. *Plant Cell* (2017). doi:10.1105/tpc.16.00435

73. Sukarta, O. C. A., Slootweg, E. J., & Goverse, A. Structure informed insights for NLR fucnitoning in plant immunity. *Semin. Cell Dev. Biol.* (2016). doi:10.1016/j.semcdb.2016.05.012

74. Takken, F. L. W. & Goverse, A. How to build a pathogen detector : structural basis of NB-LRR function. *Curr. Opin. Plant Biol.* 1–10 (2012). doi:10.1016/j.pbi.2012.05.001

75. Takken, F. L. W., & Tameling, W. I. L. To Nibble at Plant Resistance Proteins. *Science (80-.).* **324,** 744–747 (2009).

76. Nishimura, M. T. *et al.* TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in Arabidopsis. *Proc. Natl. Acad. Sci.* (2017). doi:10.1073/pnas.1620973114

77. Pan, Q., Wendel, J. & Fluhr, R. Divergent Evolution of Plant NBS-LRR Resistance Gene Homologues in Dicot and Cereal Genomes. *J. Mol. Evol.* 203–213 (2000). doi:10.1007/s002399910023

78. Collier, S. M., Hamel, L. & Moffett, P. Cell Death Mediated by the N-Terminal Domains of a Unique and Highly Conserved Class of NB-LRR Protein. *Mol. plant-microbe Interact.* **24,** 918–931 (2011).

79. Andolfo, G. *et al.* Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. *BMC Biol.* 1–12 (2014).

80. Cannon, S. B. *et al.* Diversity , Distribution , and Ancient Taxonomic Relationships Within the TIR and Non-TIR NBS-LRR Resistance Gene Subfamilies. *J. Mol. Evol.* **54,** 548–562 (2002).

81. Williams, S. J. *et al.* Supplementary Materials for Structural Basis for Assembly and Function of a Heterodimeric Plant Immune Receptor. *Science (80-.).* (2014). doi:10.1126/science.1247357

82. Schreiber, K. J., Bentham, A., Williams, S. J. & Kobe, B. Multiple Domain Associations within the Arabidopsis Immune Receptor RPP1 Regulate the Activation of Programmed Cell Death. *PLOS Pathog.* **12,** e1005769 (2016).

83. Hyun, K., Lee, Y., Yoon, J., Yi, H. & Song, J. Crystal structure of Arabidopsis thaliana SNC1 TIR domain. *Biochem. Biophys. Res. Commun.* **481,** 146–152 (2016).

84. Zhang, X. *et al.* Multiple functional self-association interfaces in plant TIR domains. **114,** E2046–E2052 (2017).

85. Nimma, S., Ve, T., Williams, S. J. & Kobe, B. Towards the structure of the TIR-domain signalosome. *Curr. Opin. Struct. Biol.* **43,** 122–130 (2017).

86. Mestre, P. & Baulcombe, D. C. Elicitor-Mediated Oligomerization of the Tobacco N Disease Resistance Protein. *Plant Cell* **18,** 491–501 (2006).

87. Bernoux, M. *et al.* Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. *Cell Host Microbe* (2011). doi:10.1016/j.chom.2011.02.009

88. Swiderski, M. R., Birker, D., Jones, J. D. G., Centre, J. I. & Colney, P. The TIR Domain of TIR-NB-LRR Resistance Proteins Is a Signaling Domain Involved in Cell Death Induction. *Mol. plant-microbe Interact.* **22,** 157–165 (2009).

89. Frost, D. *et al.* Tobacco Transgenic for the Flax Rust Resistance Gene L Expresses Allele-Specific Activation of Defense Responses. *Mol. plant-microbe Interact.* **17,** 224–232 (2004).

90. Rairdan, G. J. *et al.* The Coiled-Coil and Nucleotide Binding Domains of the Potato Rx Disease Resistance Protein Function in Pathogen Recognition and Signaling. *Plant Cell* **20,** 739–751 (2008).

91. Ade, J., Deyoung, B. J., Golstein, C. & Innes, R. W. Indirect activation of a plant nucleotide binding site – leucine-rich repeat protein by a bacterial protease. *Proc. Natl. Acad. Sci.* **104,** (2007).

92. Horsefield, S. *et al.* NAD+ cleavage activity by animal and plant TIR domains in cell

death pathways. *Science (80-.).* **365,** 793–799 (2019).

93. Wan, L. *et al.* TIR domains of plant immune receptors are NAD + -cleaving enzymes that promote cell death. *Science (80-.).* **365,** 799–803 (2019).

94. Hao, W., Collier, S. M., Moffett, P. & Chai, J. Structural Basis for the Interaction between the Potato Virus X Resistance Protein (Rx) and Its Cofactor Ran GTPase-activating Protein 2 (RanGAP2) *. *J. Biol. Chem.* **288,** 35868–35876 (2013).

95. Casey, L. W. *et al.* The CC domain structure from the wheat stem rust resistance protein Sr33 challenges paradigms for dimerization in plant NLR proteins. *Proc. Natl. Acad. Sci.* **113,** 12856–12861 (2016).

96. El, F. & Nishimura, M. T. Structural insights into plant NLR immune receptor function. *Proc. Natl. Acad. Sci.* 3–5 (2016). doi:10.1073/pnas.1615933113

97. Yan, N. *et al.* Structure of the CED-4 – CED-9 complex provides insights into programmed cell death in Caenorhabditis elegans. *Nature* **437,** (2005).

98. Riedl, S. J., Li, W., Chao, Y., Schwarzenbacher, R., & Shi, Y. Structure of the apoptotic protease-activating factor 1 bound to ADP. *Nature* **434,** 926–933 (2005).

99. Leipe, D. D., Koonin, E. V & Aravind, L. 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.* 1–28 (2004). doi:10.1016/j.jmb.2004.08.023

100. Takken, F. L. W., Albrecht, M. & Tameling, W. I. L. Resistance proteins : molecular switches of plant defence. *Curr. Opin. Plant Biol.* 383–390 (2006). doi:10.1016/j.pbi.2006.05.009

101. Hu, Z., Yan, C., Liu, P., Huang, Z., Ma, R., Zhang, C., Wang, R., Zhang, Y., Martinon, F., Miao, D. and Deng, H. Crystal Structure of NLRC4 Reveals Its Autoinhibition Mechanism. *Science (80-.).* **341,** 172–175 (2013).

102. Hu, Z., Zhou, Q., Zhang, C., Fan, S., Cheng, W., Zhao, Y., Shao, F., Wang, H.W., Sui, S.F. and Chai, J. Structural and biochemical basis for induced self-propagation of NLRC4. *Innate Immun.* **350,** (2015).

103. Tameling, W.I., Vossen, J.H., Albrecht, M., Lengauer, T., Berden, J.A., Haring, M.A., Cornelissen, B.J. and Takken, F. . Mutations in the NB-ARC Domain of I-2 That Impair ATP Hydrolysis Cause Autoactivation. *Plant Physiol.* **140,** 1233–1245 (2006).

104. Tameling, W. I., Elzinga, S. D., Darmin, P. S., Vossen, J. H., Takken, F. L., Haring, M. A., & Cornelissen, B. J. The Tomato R Gene Products I-2 and Mi-1 Are Functional ATP Binding Proteins with ATPase Activity. *Plant Cell* **14,** 2929–2939 (2002).

105. Dinesh-Kumar, S. P., Tham, W.-H. & Baker, B. J. Structure–function analysis of the

tobacco mosaic virus resistance gene N. *Proc. Natl. Acad. Sci.* **97,** 14789–14794 (2000).

106. Bernoux, M., Burdett, H., Williams, S. J., Zhang, X., Chen, C., Newell, K., ... & Dodds, P. N. Comparative analysis of the flax immune receptors L6 and L7 suggests an equilibrium-based switch activation model. *Plant Cell* (2016). doi:10.1105/tpc.15.00303

107. Inoue, H. *et al.* Blast resistance of CC-NB-LRR protein Pb1 is mediated by WRKY45 through protein – protein interaction. *Proc. Natl. Acad. Sci.* **110,** 9577–9582 (2013).

108. Wang, J. *et al.* Ligand-triggered allosteric ADP release primes a plant NLR complex. *Science (80-.).* **364,** eaav5868 (2019).

109. Wang, J. *et al.* Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science (80-.).* **364,** eaav5870 (2019).

110. Wang, G. *et al.* 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 (2015).

111. Bertin, J. & DiStefano, P. S. The PYRIN domain: a novel motif found in apoptosis and inflammation proteins. *Cell Death Differ.* **7,** 1273 (2000).

112. Inohara, N. *et al.* Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-κB. *J. Biol. Chem.* **274,** 14560–14567 (1999).

113. Ogura, T. & Wilkinson, A. J. AAA+ superfamily ATPases: common structure–diverse function. *Genes to Cells* **6,** 575–597 (2001).

114. Williams, S. J. *et al.* Structural Basis for Assembly and Function of a Heterodimeric Plant Immune Receptor Simon. *Science (80-.).* **344,** 299–303 (2014).

115. Vanaja, S. K., Rathinam, V. A. K. & Fitzgerald, K. A. Mechanisms of inflammasome activation : recent advances and novel insights. *Trends Cell Biol.* **25,** 308–315 (2015).

116. Elmore, J. M., Lin, Z. D. & Coaker, G. Plant NB-LRR signaling: Upstreams and Downstreams. *Curr. Opin. Plant Biol.* **14,** 365–371 (2012).

117. Cui, H. *et al.* A core function of EDS1 with PAD4 is to protect the salicylic acid defense sector in Arabidopsis immunity. *New Phytol.* **213,** 1802–1817 (2017).

118. Wagner, S. *et al.* Article Structural Basis for Signaling by Exclusive EDS1 Heteromeric Complexes with SAG101 or PAD4 in Plant Innate Immunity. *Cell Host Microbe* **14,** 619–630 (2013).

119. Xing, D. & Chen, Z. Effects of mutations and constitutive overexpression of EDS1 and PAD4 on plant resistance to different types of microbial pathogens. *Plant Sci.* **171,** 251–262 (2006).

120.   Heidrich, K. *et al.* Arabidopsis EDS1 connects pathogen effector recognition to cell compartment–specific immune responses. *Science (80-.).* **334,** 1401–1404 (2011).

121.   Bhattacharjee, S., Halane, M. K., Kim, S. H. & Gassmann, W. Pathogen effectors target Arabidopsis EDS1 and alter its interactions with immune regulators. *Science (80-.).* **334,** 1405–1408 (2011).

122.   Peart, J. R., Mestre, P., Lu, R., Malcuit, I. & Baulcombe, D. C. NRG1, a CC-NB-LRR, together with N , a TIR-NB-LRR Protein , Mediates Resistance against Tobacco Mosaic Virus. *Curr. Biol.* **15,** 968–973 (2005).

123.   Castel, B. *et al.* Diverse NLR immune receptors activate defence via the RPW 8-NLR NRG 1. *New Phytol.* **222,** 966–980 (2019).

124.   Qi, T., Schultink, A., Cho, M.J., Pham, J. and Staskawicz, B. J. NRG1 is required for the function of the TIR-NLR immune receptors Roq1 and RPP1 in Nicotiana benthamiana. *bioRxiv* 284471 (2018).

125.   Wu, C., Abd-el-haliem, A., Bozkurt, T. O. & Belhaj, K. NLR signaling network mediates immunity to diverse plant pathogens. *bioRxiv* 90449 (2016).

126.   Wu, C. *et al.* NLR network mediates immunity to diverse plant pathogens. *Proc. Natl. Acad. Sci.* **114,** 8113–8118 (2017).

127.   Xu, F. *et al.* NLR-Associating Transcription Factor bHLH84 and Its Paralogs Function Redundantly in Plant Immunity. *PLOS Pathog.* **10,** (2014).

128.   Padmanabhan, M. S. *et al.* Novel Positive Regulatory Role for the SPL6 Transcription Factor in the N TIR-NB-LRR Receptor-Mediated Plant Innate Immunity. *PLOS Pathog.* **9,** (2013).

129.   Fenyk, S. *et al.* The Potato Nucleotide-binding Leucine-rich Repeat (NLR) Immune Receptor Rx1 Is a Pathogen-dependent DNA-deforming Protein *. *Publ. JBC Pap. Press* (2015). doi:10.1074/jbc.M115.672121

130.   Coll, N. S. *et al.* Arabidopsis Type I Metacaspases Control Cell Death. *Science (80-.).* (2010).

131.   Bendahmane, A., Kanyuka, K. & Baulcombe, D. C. The Rx gene from potato controls separate virus resistance and cell death responses. *Plant Cell* **11,** 781–791 (1999).

132.   Coll, N. S., Epple, P. & Dangl, J. L. Programmed cell death in the plant immune system. *Cell Death Differ.* **18,** 1247–1256 (2011).

133.   Su, J. *et al.* Active photosynthetic inhibition mediated by MPK3 / MPK6 is critical to effector-triggered immunity. *PLOS Biol.* **16,** 1–29 (2018).

134.   Narusaka, M. *et al.* RRS1 and RPS4 provide a dual Resistance-gene system against fungal and bacterial pathogens. *Plant J.* (2009). doi:10.1111/j.1365-

313X.2009.03949.x

135. 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.* 467–479 (2011). doi:10.1111/j.1365-313X.2011.04502.x

136. Césari, S. *et al.* The NB-LRR proteins RGA 4 and RGA5 interact functionally and physically to confer disease resistance. *EMBO J.* **33,** 1941–1959 (2014).

137. Gassmann, W., Hinsch, M. E. & Staskawicz, B. J. The Arabidopsis RPS4 bacterial-resistance gene is a member of the TIR-NBS-LRR family of disease-resistance genes. *plant J.* **20,** 265–277 (1999).

138. Cesari, S. *et al.* 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 (2013).

139. Narusaka, M., Toyoda, K., Shiraishi, T., Iuchi, S. & Takano, Y. Leucine zipper motif in RRS1 is crucial for the regulation of Arabidopsis dual resistance protein. *Sci. Rep.* **6,** 18702 (2016).

140. Xu, X., Chen, C., Fan, B. & Chen, Z. Physical and functional interactions between pathogen-induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors. *Plant Cell* **18,** 1310–1326 (2006).

141. Saucet, S. B. *et al.* Two linked pairs of Arabidopsis TNL resistance genes independently confer recognition of bacterial effector AvrRps4. *Nat. Commun.* **6,** (2015).

142. Prior, P., Allen, C. & Elphinstone, J. *Bacterial wilt disease: molecular and ecological aspects*. (Springer Science & Business Media, 2013).

143. Xin, X.-F., Kvitko, B. & He, S. Y. Pseudomonas syringae: what it takes to be a pathogen. *Nat. Rev. Microbiol.* **16,** 316 (2018).

144. Yaqin, Y. *et al.* Colletotrichum higginsianum as a Model for Understanding Host – Pathogen Interactions : A Review. *Int. J. Mol. Sci.* **19,** 2142 (2018).

145. Tasset, C., Bernoux, M., Jauneau, A., Pouzet, C., Brière, C., Kieffer-Jacquinod, S., Rivas, S., Marco, Y. and Deslandes, L. Autoacetylation of the Ralstonia solanacearum Effector PopP2 Targets a Lysine Residue Essential for RRS1-R- Mediated Immunity in Arabidopsis. *PLOS Pathog.* **6,** e1001202 (2010).

146. Birker, D. *et al.* A locus conferring resistance to Colletotrichum higginsianum is shared by four geographically distinct Arabidopsis accessions. *Plant J.* **60,** 602–613 (2009).

147. Sohn, K. H., Zhang, Y. & Jones, J. D. G. The Pseudomonas syringae effector protein ,

AvrRPS4 , requires in planta processing and the KRVY domain to function. *Plant J.* 1079–1091 (2009). doi:10.1111/j.1365-313X.2008.03751.x

148. Li, G. *et al.* Distinct Pseudomonas type-III effectors use a cleavable transit peptide to target chloroplasts. *plant J.* **77,** 310–321 (2014).

149. Halane, M. K. *et al.* The bacterial type III-secreted protein AvrRps4 is a bipartite effector. *PLOS Pathog.* **14,** e1006984 (2018).

150. Hoon, K. *et al.* Distinct regions of the Pseudomonas syringae coiled-coil effector AvrRps4 are required for activation of immunity. *Proc. Natl. Acad. Sci.* **109,** 16371–16376 (2012).

151. Lewis, J. D. *et al.* The YopJ superfamily in plant-associated bacteria. *Mol. Plant Pathol.* **12,** 928–937 (2011).

152. Ma, Y. *et al.* Distinct modes of derepression of an Arabidopsis immune receptor complex by two different bacterial effectors. *Proc. Natl. Acad. Sci.* **115,** 10218–10227 (2018).

153. Ma.Y; Guo.H; Hu.L; Pons Martinez. P; Moschou. P.N; Cevik. V; Ding.P; Duxbury.Z; Sarris.P.F; Jones.J.D.G. Distinct modes of derepression of an immune receptor complex by two different effectors. *Proc. Natl. Acad. Sci.* **Submitted,** (2012).

154. Essuman, K., Summers, D.W., Sasaki, Y., Mao, X., DiAntonio, A. and Milbrandt, J. The SARM1 Toll/Interleukin-1 Receptor Domain Possesses Intrinsic NAD+ Cleavage Activity that Promotes Pathological Axonal Degeneration Highlights. *Neuron* **93,** 1334–1343 (2017).

155. Murashige, T. & Skoog, F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **15,** 473–497 (1962).

156. Berrow, N. S. *et al.* A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35,** e45 (2007).

157. Bird, L. E. High throughput construction and small scale expression screening of multi-tag vectors in Escherichia coli. *Methods* **55,** 29–37 (2011).

158. Stols, L. *et al.* A New Vector for High-Throughput , Ligation-Independent Cloning Encoding a Tobacco Etch Virus Protease Cleavage Site A New Vector for High-Throughput , Ligation-Independent Cloning Encoding a Tobacco Etch Virus Protease Cleavage Site. *Protein Expr. Purif.* **25,** 8–15 (2002).

159. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot , One Step , Precision Cloning Method with High Throughput Capability. *PLoS One* **3,** e3647 (2008).

160. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A Modular Cloning

System for Standardized Assembly of Multigene Constructs. *PLoS One* **6,** e16765 (2011).

161. Sohn, K.H., Segonzac, C., Rallapalli, G., Sarris, P.F., Woo, J.Y., Williams, S.J., Newman, T.E., Paek, K.H., Kobe, B. and Jones, J. D. The Nuclear Immune Receptor RPS4 Is Required for RRS1 SLH1 -Dependent Constitutive Defense Activation in Arabidopsis thaliana. *PLOS Genet.* **10,** (2014).

162. Yang, Z. R., Thomson, R., Mcneil, P. & Esnouf, R. M. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinforma. Orig. Pap.* **21,** 3369–337610 (2005).

163. Mészáros, B., Erdős, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46,** 329–337 (2018).

164. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10,** 845–858 (2015).

165. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2006).

166. Sohn, K. H., Hughes, R. K., Piquerez, S. J., Jones, J. D. G. & Banfield, M. J. Distinct regions of the Pseudomonas syringae coiled-coil effector AvrRps4 are required for activation of immunity. *Proc. Natl. Acad. Sci.* **109,** 16371–16376 (2012).

167. Gasteiger, E. *et al.* in *The Proteomics protocol handbook* 571–608 (2005).

168. Bender, K. W. *et al.* Autophosphorylation-based Calcium (Ca 2+) Sensitivity Priming and Ca 2+/ Calmodulin Inhibition of Arabidopsis thaliana Ca 2+ -dependent Protein Kinase 28 (CPK28) * □. *J. Biol. Chem.* **292,** 3988–4002 (2017).

169. Sawasaki, T., Hasegawa, Y., Tsuchimochi, M. & Kamura, N. A bilayer cell-free protein synthesis system for high-throughput screening of gene products. *FEBS Lett.* **514,** 102–105 (2002).

170. Harbers, M. Wheat germ systems for cell-free protein expression. *FEBS Lett.* **588,** 2762–2773 (2014).

171. Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, 2016).

172. Stevenson, C. E. M. *et al.* Investigation of DNA sequence recognition by a streptomycete MarR family transcriptional regulator through surface plasmon resonance and X-ray crystallography. *Nucleic Acids Res.* (2013). doi:10.1093/nar/gkt523

173. Maeo, K., Hayashi, S., Kojima-Suzuki, H., Morikami, A. and Nakamura, K. Role of

conserved residues of the WRKY domain in the DNA-binding of tobacco WRKY family proteins. *Biosci. Biotechnol. Biochem.* **65,** 2428–2436 (2001).

174.   Wang, Z. *et al.* A WRKY transcription factor participates in dehydration tolerance in Boea hygrometrica by binding to the W-box elements of the galactinol synthase (BhGolS1) promoter. *Planta* **230,** 1155–1166 (2009).

175.   Zhou, Q. *et al.* Soybean WRKY-type transcription factor genes , differential tolerance to abiotic stresses in transgenic Arabidopsis plants. *Plant Biotechnol. J.* **6,** 486–503 (2008).

176.   Burdett, H., Kobe, B. & Anderson, P. A. Animal NLRs continue to inform plant NLR structure and function. *Arch. Biochem. Biophys.* (2019). doi:10.1016/j.abb.2019.05.001

177.   Williams, S. J. *et al.* 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 (2011).

178.   Maekawa, T., Cheng, W., Spiridon, L.N., Töller, A., Lukasik, E., Saijo, Y., Liu, P., Shen, Q.H., Micluta, M.A., Somssich, I.E. and Takken, F. L. Coiled-Coil Domain-Dependent Homodimerization of Intracellular Barley Immune Receptors Defines a Minimal Functional Module for Triggering Cell Death Takaki. *Cell* **9,** 187–199 (2011).

179.   Zhang, X. *et al.* Multiple functional self-association interfaces in plant TIR domains. *Proc. Natl. Acad. Sci.* E2046-2052 (2017). doi:10.1073/pnas.1621248114

180.   Steele, J. F. C., Hughes, R. K. & Banfield, M. J. Structural and biochemical studies of an NB-ARC domain from a plant NLR immune receptor. *bioRxiv* 557280 (2019).

181.   De la Concepcion, J. C. *et al.* Polymorphic residues in rice NLRs expand binding and response to effectors of the blast pathogen. *Nat. Plants* **4,** 576 (2018).

182.   Guo, L. *et al.* Specific recognition of two MAX effectors by integrated HMA domains in plant immune receptors involves distinct binding surfaces. *Proc. Natl. Acad. Sci.* **115,** 11637–11642 (2018).

183.   Qi, Y., Gao, N., Wang, H., Zhou, J. & Chai, J. Ligand-triggered allosteric ADP release primes a plant NLR complex Jizong.

184.   Kane, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coil. *Curr. Opin. Biotechnol.* **6,** 494–500 (1995).

185.   Gustafsson, C., Govindarajan, S. & Minshull, J. Codon bias and heterologous protein expression. *Trends Biotechnol.* **22,** 346–353 (2004).

186.   Baneyx, F. & Mujacic, M. Recombinant protein folding and misfolding in Escherichia coli. *Nat. Biotechnol.* **22,** 1399–1408 (2004).

187. Kaur, J., Kumar, A. & Kaur, J. International Journal of Biological Macromolecules Strategies for optimization of heterologous protein expression in E . coli : Roadblocks and reinforcements. *Int. J. Biol. Macromol.* **106,** 803–822 (2018).

188. GOMES, A. R., MUNIVENKATAPPA, S., BYREGOWDA, VEEREGOWDA2, B. M. & BALAMURUGAN3, V. An Overview of Heterologous Expression Host Systems for the Production of Recombinant Proteins. **4,** 346–356 (2016).

189. Kapila, J., Rycke, R. De & Montagu, M. Van. An Agrobacterium -mediated transient gene expression system for intact leaves. *Plant Sci.* **122,** 101–108 (1997).

190. Edavettal, S. C., Hunter, M. J. & Swanson, R. V. in *Structure-Based Drug Discovery* 29–47 (Springer, 2012).

191. Esposito, D. & Chatterjee, D. K. Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **17,** 353–358 (2006).

192. Butt, T. R., Edavettal, S. C., Hall, J. P. & Mattern, M. R. SUMO fusion technology for difficult-to-express proteins. *Protein Expr. Purif.* **43,** 1–9 (2005).

193. Fox, J. & Waugh, D. *Maltose-Binding Protein as a Solubility Enhancer*. *E. coli gene expression protocols* (Humana Press, 2003).

194. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.* **5,** 172 (2014).

195. Correa, A. & Oppezzo, P. Tuning different expression parameters to achieve soluble recombinant proteins in E . coli : Advantages of high-throughput screening. *Biotechnol. J.* **6,** 715–730 (2011).

196. Jia, B. & Jeon, C. O. High-throughput recombinant protein expression in Escherichia coli : current status and future perspectives. *Open Biol.* **6,** 160196 (2016).

197. Imasaki, T., Wenzel, S., Yamada, K., Bryant, M. L. & Takagi, Y. Titer estimation for quality control (TEQC) method : A practical approach for optimal production of protein complexes using the baculovirus expression vector system. *PLoS One* **13,** 1–21 (2018).

198. Fuoco, N. *et al.* Factors influencing recombinant protein yields in an insect cell-bacuiovirus expression system: multiplicity of infection and intracellular protein degradation. *Antiviral Res.* **149,** 89–94 (2018).

199. Ooijen, G. Van *et al.* The small heat shock protein 20 RSI2 interacts with and is required for stability and function of tomato resistance protein I-2. *plant J.* **63,** 563–572 (2010).

200. Shirasu, K. The HSP90-SGT1 Chaperone Complex for NLR Immune Sensors. *Annu. Rev. Plant Biol* **60,** 139–164 (2009).

201. Oh, E. *et al.* Selective Ribosome Profiling Reveals the Cotranslational Chaperone Action of Trigger Factor In Vivo. *Cell* **147,** 1295–1308 (2011).

202. Aigner, H. *et al.* Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2. *Science (80-.).* **358,** 2–9 (2017).

203. Mühlmann, M., Forsten, E., Noack, S. & Büchs, J. Optimizing recombinant protein expression via automated induction profiling in microtiter plates at different temperatures. *Microb. Cell Fact.* **16,** 1–12 (2017).

204. Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H. & Kompala, D. . Plasmid-encoded protein: The principal factor in the 'metabolic burden' associated with recombinant bacteria. *Biotechnol. Bioeng.* **35,** 668–681 (1990).

205. Schein, C. H. & Noteborn, M. H. M. Formation of Soluble Recombinant Proteins in Escherichia Coli is Favored by Lower Growth Temperature. *Bio/Technology* **6,** 291–294 (1988).

206. Vera, A., González-Montalbán, N., Arís, A. & Antonio, V. The Conformational Quality of Insoluble Recombinant Proteins Is Enhanced at Low Growth Temperatures. *Biotechnol. Bioeng.* **96,** 1101–1106 (2007).

207. Harper, S. & Speicher, D. W. Purification of proteins fused to glutathione S-transferase. *Methods Mol. Biol.* **681,** 259–280 (2011).

208. LaVallie, E. R., DiBlasio-Smith, E. A., Collins-Racie, L. A., Lu, Z. & McCoy, J. M. in *E. coli gene expression protocols* (ed. Vaillancourt, P. E.) 119–140 (Humana Press, 2003). doi:10.1385/1-59259-301-1:119

209. Li, K. *et al.* Escherichia coli transcription termination factor NusA: heat-induced oligomerization and chaperone activity. *Sci. Rep.* **6,** 2347 (2013).

210. Derman, A., Prinz, W. A., Belin, D. & Beckwith, J. Mutations That Allow Disulfide Bond Formation in the Cytoplasm of Escherichia coli. *Science (80-.).* **262,** 1744–1747 (1993).

211. Georgiou, G. & Valaxt, P. Expression of correctly folded proteins in Escherichia coil. *Curr. Opin. Biotechnol.* **7,** 190–197 (1996).

212. Figler, R. A., Omote, H., Nakamoto, R. K. & Al-Shawi, M. K. Use of Chemical Chaperones in the Yeast Saccharomyces cerevisiae to Enhance Heterologous Membrane Protein Expression : High-Yield Expression and Purification. **376,** 34–46 (2000).

213. Tatzelt, J., Prusinerl, S. B. & Welcha, W. J. Chemical chaperones interfere with the formation of scrapie prion protein. *EMBO J.* **15,** 6363–6373 (1996).

214. London, W. N., Back, J. F., Oakenfull, D. & Smith, M. B. Increased Thermal Stability of

Proteins in the Presence of Sugars and Polyols1. *Biochemistry* **18,** 5191–5196 (1979).

215.   Yumerefendi, H., Desravines, D. C. & Hart, D. J. Library-based methods for identification of soluble expression constructs. *Methods* **55,** 38–43 (2011).

216.   Mas, P. J. & Darren, J. H. *Heterologous Gene Expression in E. coli: Methods and Protocols*. **1586,** (Humana Press, New York, NY, 2017).

217.   Steele, J. F. C. Molecular recognition in plant immunity. (University of East Anglia, 2016).

218.   Pok, B. *et al.* Estradiol-inducible AvrRps4 expression reveals distinct properties of TIR-NLR-mediated effector-triggered immunity. *bioRxiv* 701359 (2019).

219.   Fischer, R., Stoger, E., Schillberg, S., Christou, P. & Twyman, R. M. Plant-based production of biopharmaceuticals. *Curr. Opin. Plant Biol.* **7,** 152–158 (2004).

220.   Lomonossoff, G. P. & D'Aoust, M.-A. Plant-produced biopharmaceuticals: A case of technical developments driving clinical deployment. *Science (80-.).* **353,** 1237–1240 (2016).

221.   Hiatt, A., Cafferkey, R. & Bowdish, K. Production of antibodies in transgenic plants. *Nature* **342,** 76–78 (1989).

222.   Tacket, C. O. *et al.* Immunogenicity in humans of a recombinant bacterial antigen delivered in a transgenic potato 0. *Nat. Med.* **4,** 607–609 (1998).

223.   Kapila, J., Rycke, R. De, Montagu, M. Van & Angenon, G. An Agrobacterium-mediated transient gene expression intact leaves system for intact leaves. *Plant Sci.* **122,** 101–108 (1997).

224.   Montomoli, E., Landry, N., Ward, B. J. & Tre, S. Preclinical and Clinical Development of Plant-Made Virus-Like Particle Vaccine against Avian H5N1 Influenza. *PLoS One* **5,** e15559 (2010).

225.   Mccormick, A. A. *et al.* Plant-produced idiotype vaccines for the treatment of non-Hodgkin ' s lymphoma : Safety and immunogenicity in a phase I clinical study. *Proc. Natl. Acad. Sci.* **105,** 10131–10136 (2008).

226.   Hellwig, S., Drossard, J., Twyman, R. M. & Fischer, R. Plant cell cultures for the production of recombinant proteins. *Nat. Biotechnol.* **22,** 1415–1422 (2004).

227.   Endo, Y. & Sawasaki, T. Cell-free expression systems for eukaryotic protein production. *Curr. Opin. Biotechnol.* **17,** 373–380 (2006).

228.   Zemella, A., Thoring, L., Hoffmeister, C. & Kubick, S. Cell-Free Protein Synthesis : Pros and Cons of Prokaryotic and Eukaryotic Systems. *ChemBioChem* **16,** 2420–2431 (2015).

229.   Heard, W., Sklena, J., Tome, D. F. A., Robatzek, S. & Jones, A. M. E. Identification of

Regulatory and Cargo Proteins of Endosomal and Secretory Pathways in Arabidopsis thaliana by Proteomic Dissection. *Mol. Cell. proteomics* **2,** 1796–1813 (2015).

230.   Gaberc-Porekar, V. & Menart, V. Perspectives of immobilized-metal affinity chromatography. *J. Biochem. Biophys. Methods* **49,** 335–360 (2001).

231.   Dangl, J. L. & Jones, J. D. G. A pentangular plant inflammasome. *Science (80-.).* **364,** 31–32 (2019).

232.   Mermigka, G. & Sarris, P. F. The Rise of Plant Resistosomes. *Trends Immunol.* **40,** 670–673 (2018).

233.   Huh, S. U. *et al.* Protein-protein interactions in the RPS4 / RRS1 immune receptor complex. *PLOS Pathog.* **13,** e1006376 (2017).

234.   Zuo, J., Niu, Q. & Chua, N. An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. *plant J.* **24,** 265–273 (2000).

235.   Bernoux, M., Ellis, J. G. & Dodds, P. N. New insights in plant immunity signaling activation. *Curr. Opin. Plant Biol.* **14,** 512–518 (2011).

236.   Wiermer, M., Palma, K., Zhang, Y. & Li, X. Should I stay or should I go ? Nucleocytoplasmic trafficking in plant innate immunity. *Cell. Microbiol.* **9,** 1880–1890 (2007).

237.   Shen, Q. *et al.* Nuclear Activity of MLA Immune Receptors Links Isolate-Specific and Basal Disease-Resistance Responses. *Science (80-.).* **315,** 1098–1104 (2007).

238.   Voss, S. & Skerra, A. Mutagenesis of a flexible loop in streptavidin leads to higher affinity for the Strep-tag II peptide and improved performance in recombinant protein purification. *Protein Eng.* **10,** 975–982 (1997).

239.   Duan, M. *et al.* DNA binding mechanism revealed by high resolution crystal structure of Arabidopsis thaliana WRKY1 protein. *Nucleic Acids Res.* **35,** 1145–1154 (2007).

240.   Patching, S. G. Surface plasmon resonance spectroscopy for characterisation of membrane protein–ligand interactions and its potential for drug discovery ☆. *Biochim. Biophys. Acta (BBA)-Biomembranes* **1838,** 43–55 (2014).

241.   Wilson, I. A. N. A. *et al.* Identical short peptide sequences in unrelated proteins can have different conformations : A testing ground for theories of immune recognition. *Proc. Natl. Acad. Sci.* **82,** 5255–5259 (1985).

242.   Sievers, F. & Higgins, D. G. in *Multiple Sequence Alignment Methods* (ed. Russell, D. J.) 105–116 (Humana Press, 2014). doi:10.1007/978-1-62703-646-7_6

243.   Eulgem, T., Rushton, P. J., Schmelzer, E., Hahlbrock, K. & Somssich, I. E. Early nuclear events in plant defence signalling : rapid gene activation by WRKY transcription

factors. *EMBO J.* **18,** 4689–4699 (1999).

244. Jancarik, J. & Kim, S.-H. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Crystallogr.* **24,** 409–411 (1991).

245. Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357,** 289–298 (2006).

246. Dupeux, F., Rower, M., Seroul, G., Blot, D. & Marquez, J. A. A thermal stability assay can help to estimate the crystallization likelihood of biological samples research papers. *Acta Crystallogr. Sect. D Biol. Crystallogrpahy* **67,** 915–919 (2011).

247. Weber, P. C. *Overview of Protein Crystallization Methods. Methods in Enzymology* **276,** (Academic Press, 1997).

248. Strange, R. N. & Scott, P. R. Plant disease: a threat to global food security. *Annu. Rev. Phytopathol.* **43,** 83–116 (2005).

249. Chakraborty, S. & Newton, A. C. Climate change, plant diseases and food security: an overview. *Plant Pathol.* **60,** 2–14 (2011).

250. Jacob, F., Vernaldi, S. & Maekawa, T. Evolution and conservation of plant NLR functions. *Front. Immunol.* **4,** 1–16 (2013).

251. Jones, J. D. G. *et al.* Elevating crop disease resistance with cloned genes. *Philos. Trans. R. Soc. B Biol. Sci.* **369,** 20130087 (2014).

252. Narusaka, M. *et al.* Interfamily transfer of dual NB-LRR genes confers resistance to multiple pathogens. *PLoS One* **8,** e55954 (2013).

253. Díaz-Tatis, P. A. *et al.* Interfamily transfer of Bs2 from pepper to cassava (Manihot esculenta Crantz). *Trop. Plant Pathol.* **44,** 225–237 (2019).

254. Schultink, A., Qi, T., Bally, J. & Staskawicz, B. Using forward genetics in Nicotiana benthamiana to uncover the immune signaling pathway mediating recognition of the Xanthomonas perforans effector XopJ4. *New Phytol.* **221,** 1001–1009 (2019).

255. Chayen, N. E. & Saridakis, E. Protein crystallization : from purified protein to diffraction-quality crystal. *Nat. Methods* **5,** 147–153 (2008).

256. Thornton, J. Structural genomics takes off. *Trends Biochem. Sci.* **26,** 88–89 (2001).

257. Dubochet, J., Frank, J. & Henderson, R. Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat. Publ. Gr.* **17,** 471–492 (2018).

258. Cui, W. *et al.* Exploitation of Bacillus subtilis as a robust workhorse for production of heterologous proteins and beyond. *World J. Microbiol. Biotechnol.* **34,** 145 (2018).

259. Srinivasan, S., Barnard, G. C. & Gerngross, T. U. A novel high-cell-density protein expression system based on Ralstonia eutropha. *Appl. Environ. Microbiol.* **68,** 5925–

5932 (2002).

260.   Macauley-Patrick, S., Fazenda, M. L., McNeil, B. & Harvey, L. M. Heterologous protein production using the Pichia pastoris expression system. *Yeast* **22,** 249–270 (2005).

261.   Miyake-stoner, E. S. S. & Mayfield, S. Micro-algae come of age as a platform for recombinant protein production. *Biotechnol. Lett.* **32,** 1373–1383 (2010).

262.   Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.* **22,** 1393 (2004).

263.   De la Concepcion, J. C., Franceschetti, M., Terauchi, R., Kamoun, S. & Banfield, M. J. Protein engineering expands the effector recognition profile of a rice NLR immune receptor. *bioRxiv* 611152 (2019).

264.   Ooijen, G. Van *et al.* Structure – function analysis of the NB-ARC domain of plant disease resistance proteins. *J. Exp. Bot.* **59,** 1383–1397 (2008).

265.   Comeau, S. R., Gatchell, D. W., Vajda, S. & Camacho, C. J. ClusPro : an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **20,** 45–50 (2004).

266.   Ma, Y. Molecular mechanisms of NLR pair-mediated immunity in Arabidopsis. (University of East Anglia, 2016).