Molecular recognition in plant immunity

John Francis Chamberlain Steele

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

Banfield Lab

Department of Biological Chemistry

John Innes Centre

Norwich

March 2016

Abstract

Plant pathogens constitute a major threat to global food security. The use of naturally resistant crop varieties can limit crop losses, however new races of pathogen can arise that are able to overcome these defences. Plant breeding for race-specific resistance typically relies on disease-resistance genes, which generally encode proteins with nucleotide-binding and leucine-rich repeat domains (NB-LRRs). NB-LRRs are a large of proteins found in both plants and animals, with plant NB-LRRs further classified by the presence of N-terminal coiled-coil or toll-interleukin receptor domains.

Although qualitative models exist to describe R-protein regulation and activation, these are predominantly based on genetic and molecular studies. Biochemical investigations into R-protein function have been hampered by difficulties obtaining sufficient yields of material. When suitable material has been identified, biochemical studies have been used to complement well-established *in planta* assays to validate numerous hypotheses.

This work describes the screening processes undertaken to obtain R-protein domains suitable for downstream experiments. Using *E. coli* for high-throughput screening of constructs from a single R-protein, traditional construct design to investigate multiple R-protein domains and expanding our expression hosts to eukaryotic systems we successfully purified four coiled-coil domains and a single NBARC domain for use in downstream experiments.

Characterisation of this NBARC domain by circular dichroism and small-angle X-ray scattering indicates that the protein is well-folded and stable in solution, allowing *in vitro* investigations. In testing models for R-protein regulation we were able to confirm previous findings, such as low levels of ATPase activity, however we were unable to find evidence for a commonly cited method of signal repression. A preliminary crystal structure of the NBARC domain shows significant similarity to Apaf-1, and highlights the importance of conserved motifs in NBARC architecture.

The tools presented here should prove a valuable resource to complement existing models to better understand the structure, biochemistry, and ultimately regulation of plant R-proteins.

Acknowledgements

As has been said many times, and will doubtlessly be said many times more, PhDs are hard. Fortunately they are not done in isolation, and the support provided by friends, family, and colleagues is invaluable. It is hard to overstate the effect a supervisor can have on your success as a student, and I am fortunate to have had Mark to advise me throughout this process. It is difficult to be supportive of a student's research without micro-managing and putting a student under too much pressure, however this is something Mark has done admirably. His calm manner, advice when things didn't make sense (often), and support when things didn't work (even more often) have been invaluable over the past 4 years.

The whole Banfield group, both past and present, has been a great source of support: from giving technical advice, to providing a great and friendly environment, they have made this experience much more enjoyable. In particular I would like to thank both Richard Hughes and Marina Franceschetti for both their direct contributions to this project, and their patience and help as I have come to grips with life in a research environment. The department of Biological chemistry are as welcoming and friendly as anyone could hope for, and I would like to thank everyone for making Chatt such a good place to work.

I have been very fortunate to have a fantastic group of friends going through this experience together. From the overachievers who (unknowingly) have pushed me to work harder, to those who have shared their disaster stories, I've had a great support network full of advice, sympathy and drinking buddies. In particular Jo for putting up with my ramblings in the office, and my mild obsession with crosswords, Mike for his huge grin, enthusiastic advice and his inability to drink rum, and Richard also for his sage advice and our exchanges of lab-based frustrations.

Outside of the PhD bubble I'd like to thank Hannah Morris for being there when I needed support, and helping avert a fair few meltdowns. Johnny Grist and Andy Hearse need a mention here, as their friendship, and frequent hospitality, have kept Norwich from becoming all-consuming. Two former PhD students deserve recognition. Stuart King for all of his help dealing with issues both in and out of the lab, and Dan Tromans for helping me feel at home in Norwich and the John Innes. When moving somewhere so new, surrounded by so many brilliant people and so far from everyone you know, it can be an especially tough time, and I would like to thank Jenny Backeberg for listening quietly, patiently and (I hope) non-judgementally as I tried to get my life in order. Sometimes just having someone who knows what is going on is enough.

Friends are undoubtedly important, but without my family none of this would have been possible (obviously). I have been very lucky to have had such loving, and supportive parents. From sacrificing Sunday mornings to stand in the rain watching under 16s lose more football matches than they win, to listening to awful renditions of rock music, to supporting the choice to spend another 4 years as a student, they have given me freedom to explore the world. My sisters have been a great influence to me, probably more than I let on, and I am grateful for their belief in my abilities, whilst simultaneously knocking me down a peg when they thought I needed it. You have both been great examples to me. I also thank Alfie Steele for his boundless energy and quiet friendship over the years.

Finally I would like to thank Annis Richardson whose kindness, patience and intelligence makes me want to be a better person.

Table of contents

Abstract 2

Acknowledgements 3

Table of contents 5

List of Figures 12

List of Tables 16

List of abbreviations 17

1 General Introduction 19

1.1 Plant immunity 20

1.2 PAMPs/MAMPs and PTI 23

1.3 Effectors and ETS 24

1.3.1 Apoplastic effectors 25

1.3.2 Translocated effectors 26

1.4 R-proteins and ETI 29

1.5 Models for AVR recognition 30

1.5.1 Direct interactions (canonical NB-LRRs) 30

1.5.2 Guard hypothesis 30

1.5.3 The decoy model 31

1.5.4 Integrated decoys 32

1.6 R-protein structure 34

1.6.1 N-terminal TIR or CC domains and signalling 34

1.6.2 LRR domains as protein-interaction and recognition interfaces 35

1.6.3 The NBARC as a molecular switch 37

1.7 Models for R-protein regulation 39

1.8 Aims of this project 41

2 Materials and methods 42

2.1 General chemicals and reagents 43

2.1.1 Chemicals 43

2.1.2 Antibiotics 43

2.1.3 Bacterial media- LB 43
2.1.4 Bacterial media- Power broth™ 43
2.1.5 Bacterial media- AIM 43
2.1.6 Yeast non-selective media- YPAD 44
2.1.7 Yeast selective media 44
2.1.8 Protein purification buffers 44
2.2 DNA techniques 45
2.2.1 Plant and Bacterial expression vectors 45
2.2.2 Cloning 46
2.2.3 PCR 47
2.2.4 DNA oligonucleotides 47
2.2.5 Plasmid purification 47
2.2.6 Gel electrophoresis 48
2.2.7 Purification of DNA from agarose gels 48
2.2.8 Sequencing 48
2.3 Bacterial techniques 48
2.3.1 Bacterial strains used 48
2.3.2 Preparation of chemically competent *E. coli* 49
2.3.3 Transformation of chemically competent *E. coli* 49
2.3.4 Transformation of electro-competent *A. tumefaciens* 50
2.4 High-throughput screening of R3a constructs 50
2.4.1 DNA digestion 50
2.4.2 DNA library purification 51
2.4.3 Bacterial transformation and picking 51
2.4.4 High-throughput protein expression 51
2.4.5 Cell lysis 52
2.4.6 Western blotting 52
2.4.7 Protein expression of best constructs 52
2.4.8 Low-stringency purification and western blot of best constructs 53

2.5 Rational design of protein constructs 53
2.6 Medium-throughput *E. coli* expression screening 54
2.7 Transient transformation of *N. benthamiana* 55
2.8 Protein purification 55
2.8.1 Purification of soluble protein from *E. coli* 55
2.8.2 Generating SeMet protein from E. coli 56
2.8.3 Purification of proteins from Insect cells 56
2.8.4 Protein purification from *N. benthamiana* 57
2.8.5 SDS-PAGE 58
2.8.6 Staining SDS-PAGE gels 59
2.8.7 Protein concentration measurement 59
2.8.8 Protein concentration 59
2.9 Dynamic light scattering 59
2.10 Yeast 2-hybrid 60
2.10.1 Yeast transformation 60
2.10.2 Yeast 2-hybrid assays 61
2.11 Analytical gel filtration 61
2.12 Circular dichroism 61
2.13 SAXS 62
2.14 Antibody techniques 62
2.14.1 Conjugation of anti-R3a antibodies to Dynabeads 62
2.14.2 Protein immunoprecipitation 62
2.15 Western blotting 63
2.15.1 Protein transfer 63
2.15.2 Antibody probing 63
2.15.3 Ponceau staining of membranes 64
2.16 Crystallographic techniques 64
2.16.1 Initial crystallisation screens 64
2.16.2 Crystal optimisation 65

2.16.3 Microseeding 66
2.16.4 Additive screening 67
2.16.5 Cryoprotection of crystals and heavy atom soaking for data collection 67
2.16.6 Data collection 68
2.16.7 Data processing 68
2.16.8 Obtaining phases 69
2.17 ATPase assays 71
2.17.1 Linked enzymatic assay 71
2.17.2 Malachite green 72
2.18 DNA binding assays 72
2.18.1 SPR 72
2.18.2 Fluorescence Anisotropy 73
3 Strategies for obtaining soluble CNL R-proteins and domains 74
3.1 Introduction- Biochemical investigations into R-protein function 75
3.2 High-throughput screening of a high-profile R-protein 77
3.2.1 Rationale and theory 77
3.3 Results and Discussion 79
3.3.1 Library screening 82
3.3.2 Large-scale expression of single, N-terminally tagged R3a fragments 90
3.4 Rational design of R3a constructs 98
3.4.1 Defining construct boundaries 98
3.4.2 Cloning and expression screening 99
3.5 Medium-throughput screening of Solanaceous coiled-coil R-proteins 102
3.5.1 Construct design 102
3.5.2 Initial Screens (performed at OPPF by R. Hughes) 104
3.5.3 Large-scale expression and purification 105
3.6 Troubleshooting NRC1 NBARC point mutant expression 106
3.7 *In planta* expression of R3a domains 108
3.8 Discussion 110

4 Probing for R3a interactors 114
4.1 Introduction- INF1, AVR3a and R3a, PTI and ETI 115
4.1.1 INF and the *Phytophthora-Solanaceae* pathosystem 115
4.1.2 AVR3a and effector-triggered susceptibility 116
4.1.3 R3a and effector-triggered immunity 117
4.1.4 Using R3a to investigate R-protein signalling 117
4.2 Results 118
4.2.1 Purification of R3a CC domain 118
4.2.2 Generation of anti-R3a-CC antibodies 120
4.2.3 Immunoprecipitation of untagged R3a to probe for interactors 122
4.2.4 Optimising R3a immunoprecipitation conditions 123
4.2.5 Probing for R3a-AVR3a interactions 133
4.2.6 Using R3a antibodies to R-protein levels *in planta* 135
4.3 Discussion 137
5 Structural characterisation of NRC1 NBARC domain 139
5.1 Introduction 140
5.2 Results and discussion 141
5.2.1 Expression and purification of recombinant NRC1 NBARC domain in *E. coli* and Insect cells 141
5.2.2 Assessment of folding and thermal stability by circular dichroism 145
5.2.3 Preliminary small-angle X-ray scattering experiments suggest a rigid, globular fold 148
5.2.4 Investigating NRC1 NBARC oligomeric state *in vitro* 151
5.2.5 Crystallisation of NRC1 NBARC domain 152
5.2.6 Data collection 159
5.2.7 Data processing and structure solution 159
5.2.8 Preliminary structure of the NRC1 NBARC domain 161
5.2.9 Validating experimental phasing solutions 161
5.2.10 Investigation of preliminary refined structure and maps 163

5.2.11 Uninterpretable electron density prevents completion of NRC1 NBARC structure 168

5.2.12 Using the preliminary structure of NRC1 NBARC 171

5.3 Discussion 172

5.3.1 Heterologous expression of NRC1 NBARC in *E. coli* results in well-folded protein 172

5.3.2 Preliminary crystal structure of NRC1 reveals co-purification with ADP and highlights central role for conserved Histidine 173

5.3.3 Designing structure-based mutants to investigate ligand-dependent regulation 174

6 Functional characterisation of NRC1 NBARC domain 175

6.1 Introduction 176

6.2 Results 177

6.2.1 Purification of NRC1 N-terminal domains for *in vitro* assays 177

6.2.2 Assessment of point mutant protein folding by circular dichroism 179

6.2.3 NRC1 coiled-coil and NBARC domains do not form a stable complex *in vitro* 180

6.2.4 Yeast 2-hybrid is unable to identify interactions between N-terminal NRC1 domains 182

6.2.5 NRC1 NBARC domain is a functional ATPase 184

6.2.6 Investigating ligand-dependent oligomeric changes in NRC1 NBARC domain 187

6.2.7 Aspartate to Valine mutation in NRC1's conserved 'MHD' motif leads to increased ATPase activity 193

6.2.8 NRC1 NBARC does not interact strongly with DNA 194

6.3 Discussion 202

6.3.1 NRC1 exhibits ATPase activity consistent with predictions. An autoactive mutant displays higher rates of ATP hydrolysis, contrary to expectations 202

6.3.2 Mechanistic basis for increased ATPase activity in NRC1 NBARC^{D481V} 203

6.3.3 NRC1 signalling repression cannot be explained simply by NBARC sequestration of the coiled-coil domain 204

6.3.4 NBARC-DNA interactions 205
7 General discussion 208
7.1 Development of tools for future experiments 209
7.1.1 Purification of soluble R-protein domains 209
7.1.2 Generation of R3a-detecting antibodies 210
7.2 Preliminary structural characterisation of NRC1 NBARC domain 211
7.3 Putative roles for the MHD motif 213
7.4 Models for R-protein regulation 213
7.5 Future perspectives 215
8 References 216
9 Appendix ..... **Error! Bookmark not defined.**

List of Figures

Figure 1.3.1 Delivery of pathogenic proteins into plant cells 27
Figure 1.5.1 Schematic of how host proteins influence disease outcomes 33
Figure 1.5.2 Zig-zag model of plant immunity 34
Figure 1.7.1 Schematic of R-protein regulation and changes upon effector detection 40
Figure 2.10.1 Schematic of protein interaction-dependent activation of gene expression in yeast 2-hybrid 60
Figure 2.16.1 Schematic of a sitting drop vapour-diffusion experiment 65
Figure 2.16.2 Demonstration of structural bias introduced by phases 70
Figure 3.2.1 Schematic demonstrating the use of dual-tagged pESPRIT vector for solubility screening 79
Figure 3.3.1 Translated sequences of N-terminal (left) and C-terminal (right) halves of R3a used for library creation 80
Figure 3.3.2 DNA smear resulting from time-course digestion of the N-terminal R3a libraries 81
Figure 3.3.3 In situ western blot showing anti-histidine and streptavidin signal of bacterial colonies expressing recombinant R3a fragments 83
Figure 3.3.4 Western blot analysis of small-scale growth and expression of the top 95 colonies identified in the fluorometric screen 84
Figure 3.3.5 SDS-PAGE and coomassie staining of small-scale expression screens to assess solubility of positive candidates from low-stringency IMAC and western blotting 85
Figure 3.3.6 Typical layout of a 24-well plate used to optimize soluble expression of recombinant R3a fragments 86
Figure 3.3.7 Coomassie stained SDS-PAGE gel showing the boost in soluble expression by the addition of 0.5% DMSO 87
Figure 3.3.8 Cartoon representation of the regions covered by solubly-expressing R3a fragments 89
Figure 3.3.9 Size-exclusion chromatography purification of construct #1 following nickel-affinity IMAC 91
Figure 3.3.10 SDS-PAGE analysis of R3a construct #1 stability over time 95
Figure 3.3.11 R3a construct #66 purification 96
Figure 3.3.12 gel filtration trace of R3a construct #15$^{\Delta BAP}$ 97
Figure 3.4.1 Total soluble protein from R3a CCF-1 expression trials 99

Figure 3.4.2 Gel filtration trace and collected fractions of R3a$^{20-160}$ after his-tag removal..... 101
Figure 3.5.1 Coomassie stained SDS-PAGE of E. coli expressed R-proteins and their domains in pOPIN-F and pOPIN-S3*C*..... 105
Figure 3.13.2 Purification of the coiled-coil domains of solanaceous CNL coiled-coils 106
Figure 3.6.1 Low-stringency purification of NRC1 point mutants expressed in Sf9 cells..... 107
Figure 3.14.2 SDS-PAGE gel of purified NRC1 NBARC and point mutant from E. coli and insect cell expression..... 108
Figure 3.7.1 Low-stringency purification of R3a domains from N. benthamiana ... 110
Figure 4.2.1 progression of R3a CC expression and initial purification..... 119
Figure 4.2.2 Final purification of R3a coiled-coil domain 120
Figure 4.2.3 Pre- and post- immunisation serum from rabbit 1 and rabbit 2 121
Figure 4.2.4 Immunoprecipitation of untagged R3a from stable transgenic N. benthamiana..... 123
Figure 4.2.5 Western blot probing the effects of changing pH and salt concentration of R3a immunoprecipitation efficiency..... 124
Figure 4.2.6 Comparing the effects of antibody conjugation to magnetic beads on the ability to immunoprecipitate R3a. 126
Figure 4.2.7 Western blot to assess ability of different polyclonal anti-R3a antibody batches to immunoprecipitate R3a..... 128
Figure 4.2.8 Effect of buffer composition on R3a extraction and immunoprecipitation 130
Figure 4.2.9 Immunoprecipitation of native, denatured transgenic R3a and purified R3a coiled-coil domain..... 132
Figure 4.2.10 Western blot of AVR3a and R3a co-immunoprecipitation..... 133
Figure 4.2.11 Western blot of R3a and FLAG-tagged AVR3a alleles when expressed individually and extracts mixed 134
Figure 4.2.12 Time-course of R3a expression in N. benthamiana..... 136
Figure 4.2.13 Detection of I-2 by anti-R3a antibodies 136
Figure 5.2.1 Purification of SUMO-tagged NRC1 NBARC..... 142
Figure 5.2.2 Final purificaton of the NRC1 NBARC domain 143
Figure 5.2.3 SDS-PAGE of His-tagged NRC1 NBARC domain purified from insect cells..... 144
Figure 5.2.4 Assessment of wild-type protein purity from insect cells and E. coli . 144
Figure 5.2.5 Circular dichroism spectrum of NRC1 NBARC..... 146

Figure 5.2.6 Thermal melt curves for NRC1 NBARC 147
Figure 5.2.7 Scattering intensity-by-angle plots for NRC1 NBARC 148
Figure 5.2.8 Normalized Kratky plot of NRC1 NBARC SAXS 149
Figure 5.2.9 Porod-Debye plot of NRC1 NBARC SAXS 150
Figure 5.2.10 Analytical gel filtration for wild-type NRC1 NBARC 152
Figure 5.2.11 PACT screen crystal hits 153
Figure 5.2.12 Schematic of crystallisation experiments 155
Figure 5.2.13 Crystals of NRC1 NBARC domain purified from insect cells 158
Figure 5.2.14 Density-modified electron density map generated by experimental phasing 161
Figure 5.2.15 Identification of distinct non-main chain electron density and fitting of ADP 162
Figure 5.2.16 Comparison of the partial structure of NRC1 NBARC domain and the crystal structure of Apaf-1 NBARC domain 163
Figure 5.2.17 Identification of residues interacting with bound ADP 164
Figure 5.2.18 Interactions between Walker-A lysine and ADP 165
Figure 5.2.19 Location of conserved Walker-B residues 166
Figure 5.2.20 Proximity of conserved "MHD" motif to bound ADP 167
Figure 5.2.21 Histidine-ligand interaction in context of NBARC subdomains 168
Figure 5.2.22 Problematic regions of electron density preventing model completion 169
Figure 6.2.1 Purification of SUMO-tagged NRC1 coiled-coil 177
Figure 6.2.2 Final purification of the NRC1 coiled-coil domain 178
Figure 6.2.3 Circular dichroism spectra of wild-type NRC1 NBARC and generated point mutants 179
Figure 6.2.4 Analytical gel filtration of NRC1 coiled-coil and NBARC domains 181
Figure 6.2.5 Yeast 2-Hybrid for NBARC dimerisation and coiled-coil interactions 183
Figure 6.2.6 Schematic of linked ATPase assay 185
Figure 6.2.7 PK/LDH linked ATPase assay of NRC1 NBARC domain 186
Figure 6.2.8 Malachite green assay comparing BSA and NRC1 NBARC-mediated phosphate release 187
Figure 6.2.9 Spiking linked PK/LDH assay with non-hydrolysable ATP analogue. 188
Figure 6.2.10 Analytical gel filtration of NRC1 NBARC domain incubated with either ADP or ATP-γ -S 190
Figure 6.2.11 Analytical gel filtration of NRC1 Walker-B mutant and wild-type protein derived from insect cell expression 191
Figure 6.2.12 Analytical gel filtration of NRC1 NBARC^{D481V} 192

Figure 6.2.13 Phosphate releasing activity of NRC1 NBARC mutants in malachite green ATPase assay 194
Figure 6.2.14 Schematic of an SPR experiment 195
Figure 6.2.15 SPR of NRC1 NBARC interaction with double-stranded DNA..... 197
Figure 6.2.16 Schematic of fluorescence anisotropy experiments..... 198
Figure 6.2.17 Fluorescence anisotropy of labelled single-stranded DNA in response to BSA and NRC1 NBARC 199
Figure 6.2.18 Binding curve for NRC1 NBARC and single-stranded DNA..... 200
Figure 6.2.19 Saturated binding curve for NRC1 NBARC double-stranded DNA interactions 201

List of Tables

Table 2.2.1 Vectors used for heterologous expression in this project 45
Table 3.3.1 Summary of conditions required for soluble expression of R3a constructs 88
Table 3.11.2 Effect of protein additives on particle radius by DLS 92
Table 3.3.3 pH and salt screen to identify suitable buffer conditions for non-aggregated R3a construct #1 93
Table 3.3.4 Effect of increasing salt concentration on size of R3a construct #1 aggregates 94
Table 3.4.1 Summary of R3a domain boundaries 98
Table 3.5.1 Summary of domain predictions and domain boundaries used for R-protein cloning 103
Table 5.2.1 Estimation of NRC1 NBARC particle size in solution 151
Table 5.2.2 Conditions resulting in NRC1 crystal growth from seeding 157
Table 5.2.3 Data collection statistics for NRC1 NBARC 160
Table 5.2.4 Bromide and Iodide SAD data collection 160
Table 5.2.5 Data collection for NRC1 NBARC crystals from Insect cell expression 170
Table 5.2.6 Structure-guided mutants of NRC1 NBARC 171
Table 6.2.1 Estimated NRC1 NBARC particle size in response to treatment with either ADP or ATP-γ-S by DLS 189
Table 6.2.2 Summary of NRC1 elution peaks and corresponding molecular weights 193

List of abbreviations

This thesis uses standard abbreviations for nucleic acids (one letter code) and amino acid (one and three letter codes). Standard SI units are employed, unless otherwise stated.

Abbreviation	Meaning
3-AT	3-amino-1,2,4-triazole
A_{280}	Absorbance at 280nm
A_{340}	Absorbance at 340nm
A_{620}	Absorbance at nm
Apaf-1	Apoptotic protease activating factor 1
ATP-γ-S	Adenosine 5′-[γ-thio]triphosphate
AVR	Avirulence protein
BAK1	BRI1-associated kinase 1
BAP	Biotin-acceptor peptide
BCIP	5-bromo-4-chloro-3-indolyl phosphate
Bis-Tris	2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol
BLAST	Basic local alignment search tool
BRI1	Brassinosteroid insensitive 1
BTP	Bis-Tris-propane
CC	Coiled-coil
CD	Circular dichroism
CED-4	C. elegans death protein 4
CNL	Coiled-coil NBARC Leucine-rich-repeat protein
DAMP	Damage-associated molecular pattern
DLS	Dynamic light scattering
DMSO	Dimethyl sulphoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ETI	Effector-triggered immunity
flg22	Short flagellin-derived peptide
FLS2	Flagellin sensitive 2
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HMFM	Hogness modified freezing media
HR	Hypersensitive response (cell death)
IMAC	Immobilised metal ion affinity chromatography
IP	Immunoprecipitation
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
LRR	Leucine rich repeat domain
LysM	Lysine-rich motif (chitin binding)
MAMP	Microbe-associated molecular pattern
MAPK	Mitogen-activated protein kinase

MBP	Maltose binding protein
MES	2-(N-morpholino)ethanesulfonic acid
MMT	MES, malic acid, Tris buffer
MS	Mass spectrometry
MW	Molecular weight
NBARC	Nucleotide binding with similarty to Apaf-1, R-proteins and Ced-4
NB-LRR	Nucleotide binding Leucine-rich-repeat protein
NBT	Nitro blue tetrazolium chloride
Ni^{2+}-NTA	Nickel- nitrilotriacetic acid column
NLR	NOD-like receptor/NBARC-LRR receptor
OD_{600}	Optical density at 600nm
OPPF	Oxford protein production facility
ORF	Open reading frame
PAMP	Pathogen-associated molecular pattern
PB	Power broth
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PFAM	Protein family database
PK/LDH	Pyruvate kinase/lactate dehydrogenase
PRR	Pattern recognition receptor
PTI	PAMP-triggered immunity
R-gene	Disease resistance gene
R-protein	Disease resistance protein
Rx	Resistance to potato virus x protein
SAD	Single wavelength anomalous dispersion
SAXS	Small-angle X-ray scattering
SC-LY	Synthetic complement media lacking leu and trp
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate- polyacrylamide gel electrophoresis
SUMO	Small ubiquitin-like modifier
T3SS	Type 3 secretion system
TB	Terrific broth
TBS	Tris buffered saline
TCEP	3,3′,3″-Phosphanetriyltripropanoic acid
TIR	Toll-interleukin receptor
TNL	Toll-interleukin receptor NBARC Leucine-rich-repeat protein
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
Y2H	Yeast 2-hybrid

1

General Introduction

When faced with infection by invading microbes, animals use a complex series of immune responses arising from both the adaptive and innate immune systems to prevent colonisation and disease progression. The adaptive immune response relies on specialised circulatory cells (T- and B- lymphocytes), which are able to detect self and non-self-derived molecules (antigens). This immune response is responsive to new threats, with the ability to "remember" past pathogens, and forms the basis of modern immunizations. Plants on the other hand, lack specialised circulatory immune cells. In order to respond to invasion by a pathogen, plants rely on genetically determined innate immunity to detect foreign bodies and respond appropriately. Although passive defences exist in the form of structural polymers and constitutively expressed defensive proteins, plants are also able to actively detect and, for the most part, rebuff potential invaders. Without specialised cells to monitor for foreign bodies, individual plant cells are capable of mounting an appropriate response to invaders, and to do so use a complex series of intracellular and extracellular receptors.

The focus of this project has been to understand the structural and biochemical properties of plant disease-resistance proteins (R-proteins). R-proteins reside in the plant cytoplasm are capable of mounting rapid and effective immune responses to microbial pathogens in a race-cultivar dependent manner. In order to understand the significance of results generated, and rationale for the approaches taken in this project, it is necessary to have a sound understanding of plant immunity. The purpose of this chapter is to give a broad introduction to the key concepts in plant immunity, with emphasis on prior work done with R-proteins.

1.1 Plant immunity

Pre-harvest losses caused by plant diseases are a major constraint to achieving food security. These losses have wide-reaching effects, ranging from globalised price increases, to more localised and dramatic effects for subsistence and small-scale farmers who are more susceptible to price fluctuations and losses in productivity.

Such losses can be reduced through a variety of techniques. Crop rotation relies on the fact that many pathogens typically have limited host ranges. Replacing a susceptible crop with a non-susceptible one for a number of growing seasons ideally results in the population of the disease-causing pathogen falling to a point whereby the original crop can be replanted at this site. This method can be effective

for reducing crop losses and therefore boost per-acre yields. However there are limitations to the efficacy of this method. Some pathogens have large host ranges, such as *Phytophthora ramorum*, making rotation to a non-host crop more difficult. Alternatively, pathogens such as *Phytophthora infestans* are able to produce long-lived structures, such as oospores, which may persist in the soil for several years[3].

Practices like crop rotation can be supplemented with chemical controls, such as the application of pesticides and fungicides. Excessive use of chemical agents is undesirable for multiple reasons: many pesticides may have off-targets, causing harm to ecologically important organisms, exposure to such chemicals can be potentially harmful to humans, and organisms can evolve resistance to such chemicals over time.

Alternatively, the spread of a disease may be reduced by the use of specific cultivars which are resistant to colonisation by the pathogen. This resistance is genetically determined, and dependent on both the genotype of the host and the genotype of the pathogen[4, 5]. The resulting resistance can be narrow- if either host or pathogen does not express the correct proteins the result for the plant will be disease susceptibility. Effector-triggered immunity (Section 1.4) is an example of such gene-for-gene resistance. In ETI, a pathogenic protein termed an effector, is recognised by a host disease-resistance protein, R-protein (encoded by an R-gene), which typically results in a rapid and localised cell death response. As biotrophic and hemi-biotrophic pathogens require living cells to survive, this cell death response is an effective way to prevent colonisation and disease progression.

Historically, the genes responsible for these resistance phenotypes have been introgressed from a related non-crop species into the target crop over the course of several years. The gene can then be crossed into an elite variety, with several rounds of backcrossing, to add the new resistance phenotype to varieties with desirable properties, such as high yield or similar consumer-desired traits.

With new technologies it is possible to transfer the genes responsible for resistance phenotypes, once they have been identified and cloned, directly into certain crops. Although this certainly reduces the time required to produce new resistant crop varieties, this approach has the same limitations as traditional breeding techniques. If a new variety is successful and becomes widely used, it places a selective pressure on the pathogen to overcome this new resistance. The end result being that in some instances new resistance genes are rapidly overcome (or “broken”) in the field.

There are a number of key methods that may overcome this shortcoming. Elite varieties are genetically homogenous, meaning that large areas of farmland can be susceptible to infection by a single virulent pathogen. Using more genetically diverse crop varieties may reduce susceptibility to such infections, however this may lead to decreased crop yields in the short-term as elite varieties are produced specifically for their favourable characteristics. A second approach would be to "stack" multiple resistance genes in a single variety, or into several high-yielding varieties. Rather than relying on a single resistance gene to give resistance, using multiple R-genes, which recognise multiple pathogenic effector proteins, decreases the chances of the pathogen evolving to overcome these combined resistances simultaneously. Without a good understanding of the signalling pathways used by these proteins however, stacking R-proteins may have limited success if similar signal transduction machinery is used. Furthermore certain combinations of R-genes are incompatible, leading to inviable embryos or aberrant developmental phenotypes[6], which may restrict the repertoires of resistance genes that can be deploy in parallel.

Our increasing understanding of how R-proteins work is opening new avenues into producing synthetic resistance proteins. The aim of such work is to mutate a single R-protein to allow it to respond to a broader range of effectors, giving resistance to multiple pathogen races. So far work with the *Solanum tuberosum* resistance protein R3a has demonstrated that such approaches show merit with work by Segretin *et al* and Chapman *et al* producing mutant variants of R3a that are able to respond to both their normal targets, and to effectors that would normally be undetected[7, 8]. Further work by Harris *et al* with the resistance protein Rx demonstrates that it is possible to produce mutants that not only respond to a broader range of pathogenic proteins, but are able to translate this recognition to disease resistance[9]. The methods used to generate these synthetic proteins are time consuming and labour-intensive, typically relying on random mutagenesis followed by phenotyping many hundreds of individuals. In the case of Rx a second round of labour-intensive screening was required to obtain a desirable phenotype.

To understand the importance of R-proteins and the key concepts surrounding their function, it is important to put these specific immune receptors in the broader context of plant immunity.

1.2 PAMPs/MAMPs and PTI

Plants mount responses against attempted infection by detecting highly conserved "molecular patterns". Depending on the context these patterns are referred to as microbe-associate molecular patterns, MAMPs, PAMPs with respect to pathogens, or when referring to general damage often associated with invasion, DAMPs. These "molecular patterns" may be proteins, such as bacterial Ef-Tu and bacterial flagellin, carbohydrates, such as fungal chitin, or host-derived molecules released from neighbouring cells[10-13]. One thing that these microbially-derived patterns have in common is that they each show a high level of conservation throughout diverse species, both with regards to sequence and prevalence. Significant modifications to these molecules is suggested to lead to decreased fitness, and so these patterns provide a reliable platform for microbial detection. Interestingly the mammalian immune system, despite significant differences from the plant immune system, also targets these infection-related molecules to aid immunity[14].

Plant proteins responsible for detecting PAMPs are known as pattern-recognition receptors (PRRs) and the resulting immune response termed PAMP-triggered immunity (PTI). PRRs are transmembrane proteins with apoplastic leucine-rich repeat regions (LRRs), and can be divided into sub-classes on the basis of their cytoplasmic domains: receptor-like kinases have a cytoplasmic kinase domain that allows signal transduction when activated. On the other hand, receptor-like proteins (RLPs) lack a predicted kinase domain, and so presumably require additional components to activate a signalling cascade.

One of the best studied PRRs is FLS2 (Flagellin-sensing 2), the transmembrane receptor from *Arabidopsis thaliana* required to detect a conserved region of bacterial flagellin, known as flg22[15]. FLS2 is an apoplastic LRR with a cytoplasmic serine/threonine protein kinase domain, and its activation requires the co-receptor BAK1[16]. Investigations of FLS2 have helped our understanding not only of the specific recognition of an important MAMP, but has also revealed common themes in PAMP-triggered immunity (PTI). For example investigations into other PRRs suggest that BAK1 is required for not only FLS2 function, but also responses to elf18 (an N-terminal region of EF-Tu) and pep1, an *A. thaliana* DAMP[17]. Many of the functional readouts for FLS2 activation have been used to screen for activity of other PRRs. Upon flg22 detection, cells produce a burst of reactive oxygen species (ROS), increase the concentration of cytoplasmic calcium, deposit the structural polymer callose, and undergo transcriptional changes leading to upregulation of a

number of defence-related proteins such as PR1[18]. These broad responses are also seen upon perception of elf18 and chitin, and can be used to screen for novel elicitors. Together these responses are capable of preventing poorly adapted phytopathogens from colonising plants and causing disease symptoms.

Alongside these defensive readouts, studies into FLS2 may also be revealing common themes for PRR behaviour post-activation. When an active FLS2-BAK1 complex forms it undergoes endocytosis[19], which is emerging as a potential common behaviour of many PRRs[20].

Of particular interest to this project, the crystal structure of the FLS2 ectodomain was recently solved, revealing a corkscrew-like arrangement of LRR units and highlighting the mechanistic basis for activation by flg22 and BAK1 dimerisation[21]. Binding of flg22 occurs by binding along a major groove in the protein's surface, forming an extensive interaction from LRR 3 to LRR 16, and forming a partially shared interface with BAK1. This binding interface is consistent with previous work indicating that LRRs 7-10 are important for high-affinity binding[22]. Interestingly, Sun *et al* suggest that ligand binding does not seem to require or lead to significant conformational changes in this receptor. Structural similarities between FLS2 and BRI1 (a plant hormone receptor) leads to the possibility that other apoplastic LRRs may share the same general global fold and thus have common activation mechanisms.

Given that plants appear to have a repertoire of PRRs capable of detecting essential microbial features and mounting effective defences, the question arises of how plant diseases can occur. As PAMPs cannot be lost or significantly modified without fitness cost to the microbe, selective pressures result in the evolution of populations of microbe with proteins capable of subverting PRR recognition and responses. These proteins are collectively referred to as effectors.

1.3 Effectors and ETS

The nomenclature used to describe the proteins produced by pathogens to enhance virulence has changed over the years. Many of these proteins are termed avirulence proteins, or AVRs, as they were initially identified by their effect of limiting the range of cultivars a given pathogen could infect. Avirulence proteins are typically only detrimental to a pathogen in specific contexts, i.e. in the presence of a resistance protein capable of detecting the AVR. For example the effector AVR3a from *P. infestans* was originally identified by the inability of *P. infestans* race 3 to grow on

plants harbouring the R3 locus. In this context Avr3a is an avirulence protein, however in the absence of the R3 locus (specifically the R3a gene) Avr3a boosts the virulence of *P. infestans* by disrupting basal immunity. Effectors, or effector proteins, is a more broad term used to denote proteins produced by pathogens during the course of infection to enhance pathogen success. This term encompasses proteins that are known to be recognised (AVRs), and those for which this response has not been identified. Therefore "effector" will be used to discuss both AVR and non AVR-proteins.

Effector recognition leads to an immune response, decreasing the fitness of the pathogen. Despite this, many effectors are maintained in natural populations, suggesting that in the absence of cultivar-specific R-genes, effectors bestow an advantage to the pathogen. With access to cheaper and faster sequencing technology, improvements to map-based cloning, and bioinformatics tools the number of identified and cloned effectors has increased hugely in the past two decades. With increased examples of effectors to work with researchers have been able to dissect the activities of multiple effectors to determine how these proteins can be beneficial to pathogens. A key theme in research into the function of effectors has been the identification of proteins capable of subverting plant immunity. This subversion can act at multiple levels of plant immunity, and can occur in two distinct compartments: the apoplast, or the host cytoplasm.

1.3.1 Apoplastic effectors

When filamentous pathogens invade prospective hosts they form intimate associations with cells using specialised feeding structures called haustoria. This host-pathogen interface is one of the first sites for both potential detection and defence, making it an ideal place for effectors to act to subvert immunity.

In order to protect against fungal pathogens plants produce a number of chitinases capable of degrading fungal cell walls. These proteins are able to reduce pathogen fitness by directly attacking the microbe, but can also act to trigger further induced responses by releasing chitin monomers to be detected by PRRs. To combat this fungal pathogens produce chitinase inhibitors to protect against attack, and chitin-binding proteins to reduce PAMP-triggered responses, presumably by preventing chitin detection by PRRs. Diverse fungal pathogens from *Magnaporthe oryzae* to *Cladosporium fulvum* have evolved similar molecular machinery to sequester chitin oligomers, producing the lysine-motif (LsyM domain)- containing proteins SLP1 and Ecp6 respectively[23-25]

Plants produce not only chitinases, but also a suite of proteases that aid defence against filamentous pathogens. In the same way that the selective pressure of chitinase production leads to the evolution of chitinase inhibitors, filamentous pathogens secrete a number of serine, aspartate and cysteine protease inhibitors to boost pathogenicity and facilitate growth in the apoplast. Whilst fungal pathogens have evolved LsyM-containing proteins to bind self-derived chitin oligomers, protease inhibitors have evolved not only to target diverse proteases, but have also convergently evolved in distinct kingdoms (fungi and oomycetes) to target the same host protease[26]. Such points of functional convergence are interesting as they may highlight important immune proteins and guide future work into improving disease resistance.

Apoplastic effectors play an important role in determining the outcome of many plant-pathogen interactions, however they are not the only tool at the microbe's disposal. Many microbes supplement their apoplastic arsenals with effectors that act in the host cytoplasm.

1.3.2 Translocated effectors

Getting an effector into the apoplast is relatively straightforward to conceptualise, after all it is in the microbe's interest to be able to deliver these proteins, and so transportation and secretion can be actively supported. Plant cells on the other hand have an incentive not to take up foreign proteins during infection that could subvert normal functions. How then can pathogenic proteins enter and become cytoplasmic, rather than apoplastic effectors? For bacteria this issue can be addressed directly. The type 3 secretion system (T3SS) can act like a molecular syringe. It spans from the microbial cytoplasm, through the apoplast until it reaches and penetrates the host's plasma membrane. This needle-like structure can then be used to inject a number of bacterial effectors directly into the host cytoplasm (Figure 1.3.1).

Fungi and oomycetes lack an analogous T3SS. A large number of oomycete effectors are characterised by an N-terminal RXLR (arginine, any amino acid, leucine, arginine) motif[27, 28]. This motif occurs downstream of canonical secretion signals, indicating that the protein is first secreted out of the pathogen, and does not appear to be required for effector activity within the cytoplasm. Mutations to the RXLR motif of AVR3a results in non-uptake of the protein into plant cells, highlighting the role of this motif in translocation[28]. However the exact function of this motif remains controversial. Contradicting reports about whether the RXLR motif alone is capable of mediating pathogen-independent protein uptake[29-31], and

disputes over the potential for this motif to bind to cell membranes[32, 33] to facilitate entry into host cells make it difficult to draw meaningful conclusions about exactly how the RXLR motif functions.

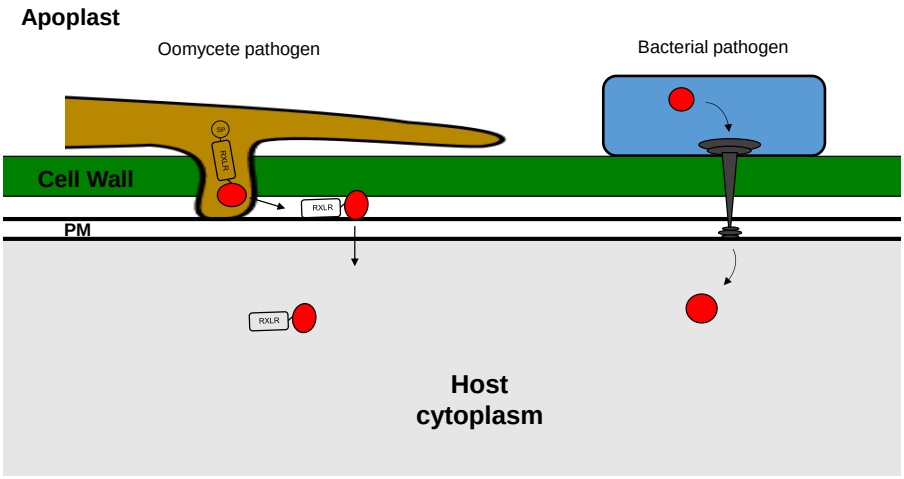


Figure 1.3.1 Delivery of pathogenic proteins into plant cells

Bacterial pathogens (right) are able to use the type 3 secretion system (T3SS) to directly transfer proteins into the host's cytoplasm. Oomycete pathogens (left) secrete effectors with N-terminal RXLR effectors into the host cytoplasm. The most likely site for this delivery is the haustorium, a specialised feeding structure that forms intimate associations with the host plasma membrane. Once in the apoplast, these RXLR effectors then translocate into the host cell through some unknown mechanism.

Regardless of the exact mechanism of entry, once inside the host cytoplasm effectors are able to carry out a wide range of activities. This project has focussed on their role in both subverting and triggering immune responses, but this is not the extent of effector function. Manipulation of the host encompasses not only immune-related activities, but also influences cell growth and metabolism to facilitate pathogenicity. Subversion of immune responses can act at many levels- from preventing signal initiation by a PRR to preventing transcriptional changes with multiple points for interference between.

If an effector is able to stop a PRR from initiating a signalling cascade in response to a PAMP the plant can be rendered effectively blind, analogous to apoplastic chitin binding proteins sequestering chitooligosaccharides to prevent PRR activation. *Pseudomonas syringae* for example has evolved the 18kDa cytoplasmic effector AvrPto that is capable of interacting with FLS2, and other defence-related kinase domains, and suppresses PTI by interfering with kinase activity[34, 35]. This inhibitory activity is not general to all kinases and the effector is itself the target of additional host kinases to promote virulence[36], part of an emerging theme of effector "helpers".

In addition to trying to prevent immune signalling at the source, effectors can target downstream components of the signalling cascade to prevent resistance. In the case of *P. syringae*, inhibition of FLS2 function by AvrPto can be supplemented or replaced by the activity of an additional effector, HopA1. This translocated effector has been shown to interact with two mitogen-activated protein kinases (MAPKs), and prevents flg22-induced oxidative bursts and full immunity through its phospholyase activity[37], an activity that seems to be conserved in bacteria infecting both plants and animals[38, 39].

Just as diverse pathogens have produced apoplastic effectors to target the same classes of defensive hydrolytic enzymes, targeting of related signalling pathways, namely MAPKs, is a common theme in cytoplasmic host manipulation. PexRD2 is a 121 amino acid protein produced by *P. infestans* comprising an N-terminal secretion signal, an RXLR motif and a C-terminal effector domain of approximately 64 residues forming a WY-fold. The effector domain of this protein alone is sufficient for virulence and avirulence functions, and acts as a dimer *in vivo* and *in vitro*[40, 41]. Once in the plant cell, this homodimer interacts with MAPKs (specifically *S. tuberosum* MAPKKKε) and interferes with cell death by preventing activities of the protein[41], analogous to *P. syringae* HopA1 activity.

Effectors are not only able to interfere with the signalling events leading to immunity, but can play a more direct role in preventing appropriate responses. The bacterial pathogen *Xanthomonas campestris* is able to use the T3SS to translocate the effector AVRBs3 into host cells, where it then uses host machinery to accumulate in nucleus[42, 43]. Once in the nucleus AVRBs3 is able to manipulate defence-related transcription through its DNA binding and transcription-activation domains[43, 44]. This DNA-binding and transcriptional control is not limited to *X. campestris*, but can be found in a wide range of *Xanthamonas* species[45], and so may be a general method to enhance pathogeneicity.

Understanding how effectors function not only illuminates the specific activities of individual proteins, but gives a greater insight into host manipulation, important features of the immune system, pathogen evolution and limitations, and allows the development of new and useful tools. For example knowledge of the *Phytophthora* RXLR motif has allowed bioinformatic screening to identify new candidate effectors, opening new avenues for investigation[46,47], and investigating *X. campestris* effectors has resulted in the development of new tools for DNA manipulation[48].

Identification of the WY-domain has led to new ideas regarding effector evolution, and the relationship between sequence identity and structural similarity. WY-fold proteins are a significant subset of the large RXLR-class of oomycete effector proteins, of which there are over 500 predicted in the *P. infestans* genome[27]. Despite low sequence identity, these WY-fold effectors are predicted to share a similar core fold which appears to be adaptable to a variety of oligomeric states and repeat lengths [40, 49-51]. Despite structural similarities, these proteins display a wide range of biochemical properties, leading to the hypothesis that this fold has evolved as a scaffold onto which new functions can be added.

With multiple proteins seemingly capable of overcoming basal resistance mechanisms, plant disease should be rampant. Fortunately there is a fitness cost associated with effector expression in specific contexts. That fitness cost comes in the form of cultivar-specific plant disease-resistance proteins (R-proteins) that are capable of detecting the presence of certain effectors and triggering rapid immune responses.

1.4 R-proteins and ETI

Although the term R-protein refers to any plant protein capable of initiating a gene-for-gene style immune response, the focus of this project has been typical NB-LRR (in particular CC-NB-LRR or CNL) proteins. These proteins share a conserved nucleotide-binding domain (NB or NBARC domain) and C-terminal leucine-rich repeat (LRR) domain, and may have N-terminal Toll-interleukin receptor-like (TIR) or coiled-coil (CC) domains (Section 1.6). However the ability to trigger specific resistance is not restricted to this family of proteins[52].

A key function of an R-protein is to detect the presence of a pathogen effector. This typically results in a rapid cell death response (termed a hypersensitive response, or HR), which is particularly effective at stopping further infection by biotrophic or hemi-biotrophic microbes. Just as pathogens use multiple methods to interfere with basal immunity, plants deploy R-proteins that use diverse methods to detect cytoplasmic effectors.

1.5 Models for AVR recognition

1.5.1 Direct interactions (canonical NB-LRRs)

The most straightforward model to describe how R-proteins are able to detect the presence of pathogenic effectors is that the effector directly interacts with the R-protein, resulting in activation of the protein. This is the case for Flax (*Linum usitatissimum*) resistance proteins L5 and L6 which were found to interact directly with the corresponding AVRL567 alleles from flax rust (*Melampsora lini)*. Recognition specificity is determined in this case by the ability to interact with the various alleles[53]. From mutagenesis and domain swap experiments it is possible to assign this interaction activity to the LRR domains of these proteins[54]. Where examples of direct interactions can be found between canonical NB-LRR proteins and effectors, binding appears to be mediated predominantly by LRR domains[55, 56].

Although direct interactions are able to explain the recognition of certain effectors, a lack of evidence for direct interactions in other pathosystems has led to hypotheses regarding indirect activation. Indirect activation covers a wide range of potential mechanisms, from recognition of host protein-effector complexes to recognition of modified host proteins. There are two main models for indirect R-protein activation depending on the effect that interactions with host proteins has on pathogen virulence in the absence of an R-protein: the guard hypothesis and the decoy model.

1.5.2 Guard hypothesis

In the guard hypothesis an effector targets a specific host protein to enhance virulence of the pathogen. In the presence of the cognate R-protein, interaction between the effector and host protein (or guardee) results in R-protein activation and HR[57]. In this instance silencing or removal of the host target will result in a loss of effector-dependent enhanced virulence.

AvrPto is not only able to interact with the kinase domains of FLS2 and related PRR proteins, but also with a 321 amino-acid cytoplasmic kinase, Pto[58 59]. Pto forms large complexes with the NB-LRR, Prf, through multiple interactions with the poorly characterised N-terminal region of the R-protein[60-62]. When AvrPto interacts with Pto, it stimulates phosphorylation of Pto in a Prf-dependent manner, leading to HR and limiting infection[63-65]. In this example of guarding, it is the effector-dependent change in Pto that activates Prf, with the loss of either the R-protein (Prf) or the guardee (Pto) leading to a loss of immune response.

Similarly, the *A. thaliana* host protein Rin4 (RPM1-interacting protein 4) interacts with, and is required for, the activation of the NB-LRR, RPM1[66]. Unlike Pto Rin4 is targeted by multiple pathogen effectors, however interaction with two (AvrRpm1 and AvrB) also leads to phosphorylation of the guardee and immune activation[66]. Interactions between RIN4 and a third effector AvrRpt2 leads to activation of a separate R-protein, RPS2[67]. Although triggered by a similar guarding mechanism to RIN4-dependent RPM1 activation, RPS2 activation occurs instead by cleavage and degradation of RIN4[68, 69].

One potential advantage for indirect activation of R-proteins is that it may make it more difficult for effectors to evolve to avoid recognition. In the case of direct interactions, provided that the R-protein recognition interface is distinct from the functional region of the effector, surface mutations may lead to escape of recognition without affecting fitness. If R-protein activation is based not on direct interactions, but the function of an effector then these simple surface mutations will not result in evasion by the effector- the only ways to prevent recognition would be to either prevent effector function, presumably with an associated fitness cost, or by evolving separate proteins capable of mitigating the effects of this active R-protein.

Although this system of recognition may lead to more stable resistance for the plant, the unequal prevalence of guards and effectors in potential host and pathogen populations may result in an unstable situation for the host target, leading to additional hypotheses about the nature of effector targets in pathology.

1.5.3 The decoy model

At its most simple, the decoy model put forth by van der Hoorn and Kamoun states that plants will evolve host proteins that are able to mimic effector targets, but are only involved in effector perception[70]. The authors state that in the absence of an R-protein capable of detecting effector-mediated modification, there will be a pressure for the host target to evolve to escape interactions that aid virulence. Conversely in the presence of a cognate R-protein, there will be a pressure for the host protein to evolve a higher affinity for the effector to facilitate recognition. The decoy model solves this unstable situation by the introduction of a second protein, which can evolve a high affinity for the effector to aid recognition in the presence of the cognate R-protein, but the manipulation of which does not lead to enhanced virulence.

In this paper the authors cite a number of examples to back this idea. AvrBS3 (introduced in Section 1.3.2) was identified through its avirulence function.

Alongside upregulating promoters that enhance pathogenicity, this protein also binds to the promoter of the resistance protein BS3 and induces expression. This in turn triggers an immune response to limit infection. A second example is the case of Pto (Section 1.5.2, Guard hypothesis). *P. syringae* expressing AvrPto have enhanced fitness on plants even in the absence of Pto, however this advantage is lost on plants that do not express FLS2, suggesting that FLS2 is the bona fide effector target, with Pto acting as a decoy which mimics this protein.

More recently Wang *et al* demonstrated that a host protein PBL2 was a functional decoy for its paralog BIK1 to allow detection of the *X. campestris* effector AvrAC[71]. AvrAC interacts with, and modifies, BIK1 to prevent its activation via phosphorylation, and thus suppresses PTI[72]. AvrAC also triggers an immune response in plants that express the NLR ZAR1, and an additional host protein, RKS1. The key finding for this most recent publication was the discovery that although PBL2 was modified in the same way as BIK1, the loss of PBL2 did not affect AvrAC virulence[73].

Recently researchers have discovered novel NB-LRRs that appear to take the use of decoy molecules to a new level, directly linking effector activity to R-protein activation.

1.5.4 Integrated decoys

R-protein architecture shows some degrees of plasticity. Although most common are CNL or TNL proteins, there are R-proteins with additional domains, such as leucine-zipper-like domains or the N-terminal solanaceous domain. Recently research into predicted R-proteins in rice has led to the discovery of R-proteins with domains that directly tie effector function to R-protein activation.

The atypical R-protein Pikp-1, which has a novel small heavy-metal associated domain (HMA domain) fused between the classical coiled-coil and NBARC domains, works in concert with the genetically linked canonical CNL Pikp-2 to mediate an immune response to *Magnaporthe oryzae* AVR-Pik proteins, via direct interactions with the effector[74]. Unlike the rice R-protein Pi-ta, which directly interacts with its cognate effector through its LRR domain, Pikp-1 interacts with AVR-Pik proteins through its HMA domain to mediate recognition[75]. Likewise, the rice R-protein pair RGA4 and RGA5 utilise a similar domain (this time fused to the C-terminus of RGA5) to detect the presence of *M. oryzae* effectors AVR-Pia and CO39[76, 77]. The inclusion of non-canonical domains into R-proteins is not restricted

to rice, but can also be seen in the model plant *A. thaliana*. The *A. thaliana* Resistance to Ralstonia solanacearum (RRS1) protein belongs to the TNL class of R-proteins, and works in conjunction with another TNL, genetically linked, called RPS4. Like RGA4 RRS1 also has a novel C-terminal fusion, however this fusion encodes a WRKY transcription factor domain[78]. This incorporated WRKY domain is essential to detect the *P. syringae* effectors AvrRPS4 and the acetyltransferase-like effector Pop2, which interact with multiple WRKY transcription factors[79, 80]. Together these examples have led to the formulation of an integrated decoy model. These fused domains act analogously to molecular decoys described in Section 1.5.3, however incorporation into R-proteins leads to a direct link between effector function and immune activation[81].

Host-microbe interactions and the behaviour of individual components can be complex. Figure 1.5.1 summarises how translocated effectors may aid or hinder pathogenicity depending on the genotype of the host.

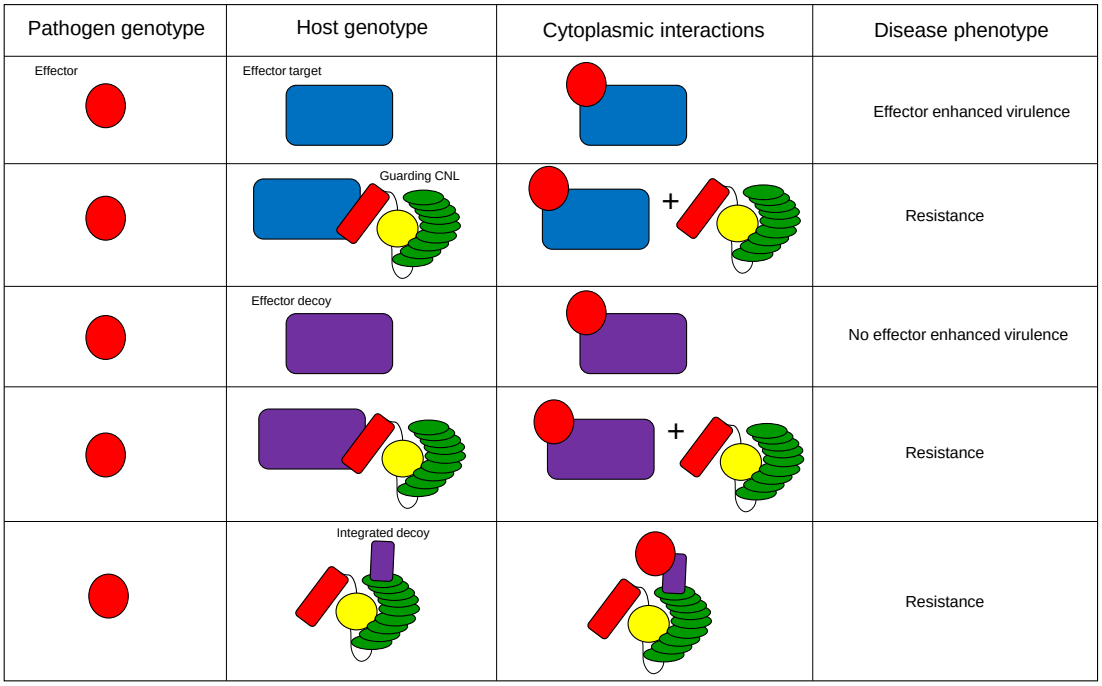


Figure 1.5.1 Schematic of how host proteins influence disease outcomes

When pathogens translocate effectors (red circles) into host cytoplasm, they can interact with effector targets, decoys, or intergrated decoy regions. Depending on whether the host produces a cognate resistance protein, these pathogenic effectors can either aid, or hinder infection.

The general outcome of microbe-plant interactions can be represented more simply by a zig-zag model describing how PAMPs, effectors, PRRs and R-proteins contribute to infection and resistance (Figure 1.5.2).

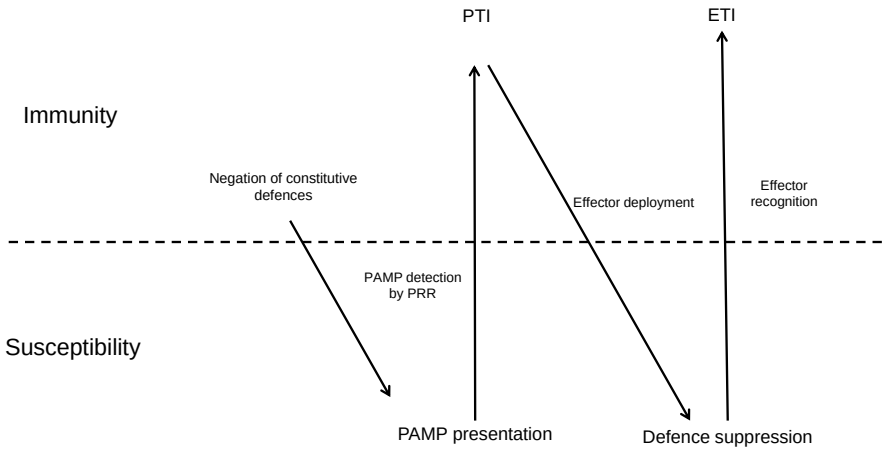


Figure 1.5.2 Zig-zag model of plant immunity

Whether or not a plant-pathogen interaction results in disease or resistance depends on the interactions of host-microbe proteins. Degradation of structural polymers such as cellulose, or the deployment of effectors reduce the plant's ability to defend itself and leads to susceptibility. Detection of PAMPs, or effectors leads to recapitulation of immune responses and prevents disease progression. Adapted from Jones and Dangl[2].

1.6 R-protein structure

The majority of R-proteins identified to date have shown very similar domain organisations- a central NBARC domain, a C-terminal LRR domain and most commonly either a Toll-interleukin receptor-like (TIR) domain or a coiled-coil (CC) domain at the N-terminus. The modular nature of R-proteins has led to hypotheses that the individual domains are responsible for different aspects of R-protein behaviour. For R-proteins to be useful in plant immunity they must carry out certain functions. They must be able to detect the presence of an effector (recognition), translate that detection to an appropriate response (signalling), and remain in a signalling-competent, inactive state prior to effector recognition (inhibition). Investigations into the behaviour of R-protein domains suggest that such functional division is possible.

1.6.1 N-terminal TIR or CC domains and signalling

The two largest families of R-proteins have either N-terminal TIR (Toll-interleukin receptor-like) domains or predicted coiled-coil (CC) domains, abbreviated to TNL or

CNL (TIR/CC-NBARC-LRR) proteins. TIR domains have been implicated in signal transduction in animal immune receptors[82] and overexpression of the TIR domains of the plant TNLs L6, RPS4 and RPP1A lead to effector-independent cell death[83, 84]. TIR domains have a 5-stranded parallel β-sheet surrounded by 5 α-helices[82], with multiple homo- and hetero- dimerization interfaces that are important for signalling and regulation[84, 85]. Furthermore structural investigations into the role of homo- vs. hetero- dimers in the RPS4 RRS1 system highlighted that activation leads to a change in hetero- to homo- dimer formation to enable signalling, rather than depending solely on intra-molecular rearrangements[85] (Section 1.7).

Consistent with observations from TNLs, the overexpression of the coiled-coil domain of ML10 (flax) and Rp1-D (maize) alone are able to trigger a cell death response[86, 87], as does the CC domain of *A. thaliana* RPS5 when fused to its NBARC domain.

Although these data suggests that CNLs use their N-terminal domains to initiate cell-death signalling, the functions of coiled-coil domains are not limited to triggering HR. Overexpression of the CC of Rx for example does not lead to cell-death. Instead it appears to be involved in interactions with a host protein, either for localisation or as a guardee[88], similar to the guarding of Pto by Prf, which is also mediated by the N-terminus of the NB-LRR[61]. Additionally the coiled-coil domain of a number of plant resistance proteins appear to be targets for post-translational modification and localisation[89-91].

Recently, the crystal structure of two unrelated R-protein coiled-coil domains have also been solved. Unlike the broad similarities seen between the TIR domains of RPS4, RRS1 and L6, these two coiled-coil domains show dramatically different topologies. MLA10 coiled-coil domain forms homodimers, as demonstrated for various TIR domains, and forms an elongated structure with two interlocking coiled-coils resembling links in a chain[86]. The CC domain of Rx on the other hand shows a more globular fold overall, with a more typical stacked 4-helix bundle[92]. Whether these differences in structure reflect the functional differences of these proteins (MLA10 CC triggering cell death, and the CC of Rx interacting with RanGAP2), or whether CC domains as a whole will show diverse structures is yet to be demonstrated.

1.6.2 LRR domains as protein-interaction and recognition interfaces

R-proteins possess a leucine-rich repeat region (LRR) C-terminal to the highly-conserved NBARC domain, which show high sequence diversity. LRRs have been

implicated in mediating protein-protein interactions, and have been suggested to facilitate protein recognition[93]. Confirmation of this idea comes from plant PRRs such as FLS2 using its LRR domain to recognise flg22, as well as direct interactions seen in flax rust and rice blast resistance to mediate recognition of pathogenic proteins. This leads to the hypothesis that molecular recognition is a generic function of R-protein LRRs.

The LRR domains of plant R-proteins are predicted to not only detect the presence of cognate effectors, but are implicated in the autoinhibition of R-proteins in the absence of a pathogen. Expression of the CC-NBARC domain of RPS5 leads to effector-independent cell-death, which can be abolished by the inclusion of a short region of the LRR domain[89]. Similarly, step-wise truncations of Rx indicate that the LRR has repressive functions in this protein too[94]. Co-expression of the Rx CC-NBARC domains with its LRR *in trans* results in a stable complex. This complex is disrupted by the expression of the potato virus X (PVX) coat protein, which normally triggers an Rx-mediated response[95]. Domain swaps between homologous proteins further support the role of the LRR as a signal inhibitor, with specific LRR and ARC combinations required to prevent autoactivity[96, 97].

Although no examples of cytoplasmic plant LRR structures are available, the functional similarities between cytoplasmic and the apoplastic LRR domains of PRRs allows homology modelling, and inference of mechanistic basis of R-protein LRR function. The crystal structure of the LRR domains of FLS2 and BRI1 have been solved, and despite detecting very different ligands (a microbe-derived peptide and a host-derived hormone respectively), the two protein domains show a high degree of structural similarity. This may suggest that their super-helical structure may be a common feature of apoplastic LRR domains. Whether this is true of their cytoplasmic counterparts remains to be demonstrated. Unlike PRR ectodomains, LRR domains of cytoplasmic R-proteins are suggested to have two functions (both detection and repression), which may not be supported by the same structural features shown by these apoplastic receptors, particularly given the suggested lack of conformational change in FLS2 in response to flg22 binding. Recently, the solution of the mouse NBARC-containing protein NLRC4 was solved to give greater insight into the inactivation of these proteins by extensive LRR-NBARC interactions[98]. Given that this represents a cytoplasmic, immune-based NBARC-LRR protein, it would be interesting to see whether plant cytoplasmic LRRs favour a conformation similar to structures already seen within the plant kingdom, or whether

functional similarities from distantly related proteins will be reflected in their structures.

1.6.3 The NBARC as a molecular switch

Most R-proteins identified to date are predicted to have a central nucleotide binding domain with sequence similarity to that seen in the mammalian apoptotic protease activating factor 1 protein (Apaf-1) and *C. elegans* death protein 4 (Ced-4). In light of this predicted conservation, this domain is referred to as the NBARC (Nucleotide Binding domain found in Apaf-1 R-proteins and Ced-4), or sometimes the NBS (Nucleotide Binding Site), however to emphasise the putative functional differences in the sub-regions of this module, it will be referred to here as the NBARC domain. Sequence analysis of plant NBARC domains identifies a conserved Walker-A (or P-loop) motif seen in many AAA proteins (ATPases associated with diverse cellular activities), which is known to be required for nucleotide binding, and also a conserved Walker-B (or kinase-2) motif- an essential feature for hydrolysis of nucleotide tri-phosphates to nucleotide di-phosphates (such as the conversion of ATP to ADP).

If initiation of cell death signalling is performed by the CC domain, and the LRR is responsible for both signal repression and effector detection, what is the purpose of this additional, highly conserved domain? Initial suggestions regarding the function of this region comes from the similarity shared between this protein and the animal proteins Apaf-1 and CED-4. Both of these proteins are involved in programmed cell death, however the stimuli they respond to are very different. Apaf-1 is a modular protein that responds to cell stresses. Like CNLs it has an N-terminal helical domain, this time a helical bundle termed a CARD domain (Caspase activation and recruitment domain). C-terminal to this is the NBARC domain, followed by a region of WD40 repeats, which are implicated in mediating protein-protein interactions. Like Apaf-1, Ced-4 is modular and shares an N-terminal CARD domain and NBARC domain. Unlike Apaf-1 and R-proteins, this protein lacks C-terminal extensions typically associated with protein-ligand interactions. A potential explanation for this absence is that the Ced-4 protein does not respond to stress stimuli as Apaf-1 and R-proteins do, but is developmentally regulated, and therefore may be under different regulatory pressures. Despite responding to different stimuli, the function of activated Apaf-1 and Ced-4 are very similar- both proteins form large oligomeric complexes that interact with downstream procaspases to trigger caspase

maturation. These matured cysteine-aspartatc proteases go on to cleave a number of downstream targets resulting in apoptosis.

How do these two proteins contribute to our ideas about R-protein regulation? Apaf-1 is able to remain in an apparently auto-inactive state until required to trigger a response, much like R-proteins. The switch from inactive to an active state requires the binding of cytochrome-c[99] (released from mitochondria under stress conditions), and dATP[100]. Although not requiring cytochrome-c release to activate cell death, Ced-4 also appears to require the ability to bind to dATP for activation[101]. Both Ced-4 and Apaf-1 nucleotide binding is attributable to their NBARC domains, as highlighted by the presence of conserved ATP-binding motifs (P-loop or Walker-A motifs), which when perturbed prevent cell-death initiation. Given the presence of conserved Walker-A motifs in many R-proteins, their apparent ability to self-regulate and the sequence similarity between the NBARC domains of many R-proteins and these animal homologs (>30%) suggests that nucleotide binding is required for R-protein function.

Indeed, mutations in the Walker-A motif of several R-proteins has been linked to loss-of-function. Absence of canonical GTPase motifs in R-protein sequences led to early speculation that these proteins would be specific for ATP, with adenosine substrate specificity subsequently confirmed experimentally[102]. Furthermore, the N-terminal domains of I-2 and Mi-1 refolded from inclusion bodies have both been shown to be catalytically active, capable of hydrolysing ATP to ADP, with reports that additional phosphatase activity may be possible to produce AMP[102, 103]. Additional reports of ATP binding and hydrolysis from the soluble expression of the tobacco mosaic virus (TMV) resistance protein N, point to this being a shared activity of both CNL and TNL NBARC domains[104].

Given that NBARC domains require nucleotide binding to function, and that distinct R-proteins are capable of hydrolysing ATP, hypotheses were put forth regarding the role of ATP in R-protein function. Two main possibilities are that the hydrolysis of ATP provides energy to drive conformational changes in the protein, giving rise to activation. Alternatively, the binding of ATP alone may be sufficient to lead to protein activation. Researchers used knowledge of related STAND proteins to target a second ATPase-related motif, the Walker-B motif (also known as the kinase-2 motif) for mutation and *in planta* investigations. Mutations in this motif allow nucleotide binding, however ATPase activity is significantly reduced by disrupting interactions with catalytic metal ions. When expressed *in planta*, R-proteins with reduced ATPase activity display effector-independent activation, suggesting that ATP

binding is sufficient for activation, and conversely that ATP hydrolysis is required for proper regulation[105].

A further common feature of Ced-4 and Apaf-1 is their ability to form large multimeric complexes (termed apoptosomes) once activated[106]. The NBARC domains of these proteins were originally called nucleotide-binding oligomerization domains (NOD) [107]. These large, ring-like structures are seen in other cell-death initiating Ced-4 homologs, such as the *Drosophila* protein DARK, with inter-molecular interactions mediated by the NBARC domain[108]. This leads the possibility that R-proteins may also form the large apoptosomes upon activation, however it may be that these structures are specific to caspase-interacting proteins.

More recently, the crystal structure of NLRC4 from *Mus musculus*, another Apaf-1-like protein, has been solved. This structure represents a more comprehensive view of the autoinhibitory activities of NBARC-containing proteins as the crystal structure entails not only the NBARC domain of NLRC4, but also the C-terminal LRR repeats. In this structure extensive interactions between the LRR domain and the NBARC domain suggests that this C-terminal extension acts to prevent nucleotide exchange.

Homology modelling of R-proteins onto the structure of ADP-bound Apaf-1 and ATP bound Ced-4 suggests that nucleotide exchange may be associated with conformational changes in the NBARC domain[109]. These conformational changes are then predicted to drive larger conformational changes in the protein resulting in an active R-protein. Similar homology modelling combined with sequence alignments have suggested that R-protein NBARC domains structurally resemble Apaf-1, and point to an invariant histidine in the “MHD” motif of the NBARC playing a role in interacting with bound ADP[110].

1.7 Models for R-protein regulation

These studies indicating functional divisions between R-protein domains and the roles of intramolecular interactions have led to the formation of models for R-protein regulation such as the one presented in Figure 1.7.1.

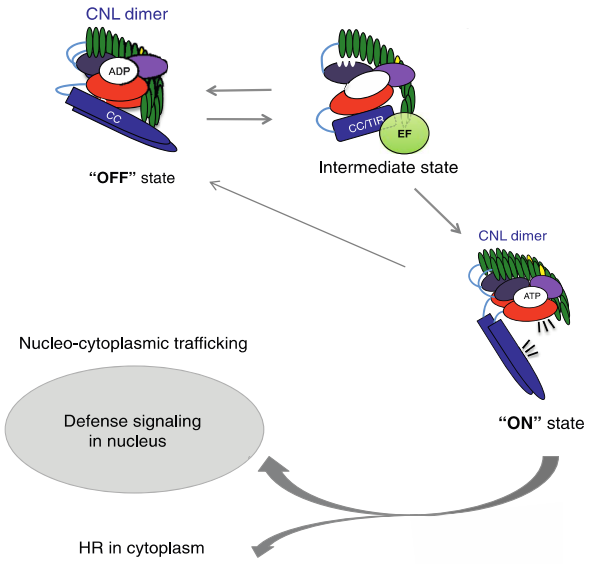


Figure 1.7.1 Schematic of R-protein regulation and changes upon effector detection

In this model for R-protein regulation, the domains of the protein form extensive intra-molecular interactions, with ADP bound at the centre of the NBARC domain. Upon effector detection the protein undergoes conformational changes to allow ADP release. The CNL then undergoes ATP binding and further conformational changes, such as the release of the CC domain, into an "on" state to trigger HR. Adapted from Takken and Goverse[1].

Although models like this give a good conceptual framework to understand how R-proteins function in general, there are a number of R-proteins whose behaviour cannot be predicted in this way. Overexpression of the NB domain of Rx, and not the CC domain leads to effector-independent cell death, and fusion of the CC domain of Rp1-D21 to its NBARC domain leads to repression of normal cell death responses. Likewise evidence suggests that the coiled-coil domain has a role in protein localisation[89, 90] and is not simply an interaction interface for signal transduction. Although the idea of the LRR as an effector sensor and auto-inhibitory domain is well evidenced, exceptions such as the case of Rp1-D should be acknowledged and kept in mind when working with a novel R-protein.

Furthermore these models are useful for qualitative rather than quantitative predictions. Given that much of the work underlying these models was performed *in planta*, biochemical and mechanistic details such as interaction affinities are lacking. Biochemical investigations into R-protein activity have been severely hampered by the lack of sufficient yields of suitable protein for sensitive investigations. Production of materials for such studies has been a central part of this project, to allow detailed investigations into R-protein regulation.

1.8 Aims of this project

This project has four key aims. Firstly to develop tools, in the form of solubly-expressing proteins, that will be of interest to the field. Secondly to use these derived tools to investigate the biochemical basis for R-protein regulation and auto-inhibition in the absence of an effector. Thirdly to improve our understanding of R-proteins and their domains at the structural level, allowing comparison to other characterised systems, and furthering our mechanistic understanding of R-protein regulation. Finally this project will attempt to develop tools to further our knowledge of the inter- and intra- molecular interactions that are important for R-protein function. A fully descriptive model for R-protein regulation, with empirical data describing how biochemical and structural features influence regulation, the mechanistic basis for signal repression, and high-resolution data regarding important intramolecular interactions may allow us to fine-tune R-protein responsiveness. Combining this information with structural data regarding the specifics of effector detection may open up possibilities for rational design of R-proteins with targeted responses to a wide variety of pathogen, and ultimately durable resistance in the field.

2

Materials and methods

2.1 General chemicals and reagents

2.1.1 Chemicals

All chemicals were supplied by Sigma-Aldrich Ltd, ThermoFisher Scientific Inc. or VWR International, unless explicitly stated.

2.1.2 Antibiotics

Antibiotics were prepared by dissolving into de-ionised H_2O and filter sterilised using 0.3µm Minisart® filters before being stored at -20°C. Exceptions to this are Rifampicin, which was dissolved in 100% DMSO, and Chloramphenicol, which was dissolved in ethanol. 1000x stocks were used to supplement both liquid and solid media to the following concentrations: Carbenicillin 100µg/mL, Kanamycin 30µg/mL Rifampicin 50µg/mL, Spectinomycin 100µg/mL, Chloramphenicol 34µg/mL

2.1.3 Bacterial media- LB

Lysogeny broth (Lauria-Bertani) was purchased from Formedium and 2.5% w/v dissolved in de-ionised water before autoclaving. This broth (LB broth Millar) contains high salt concentrations (1.0% w/v tryptone, 0.5% w/v yeast extract, 1.0% w/v sodium chloride), and so for salt-sensitive antibiotics, namely spectinomycin, L broth (1.0% w/v tryptone, 0.5% yeast extract 0.5% w/v sodium chloride, 0.1% w/v glucose- Formedium) was used.

For solid LB/L agar plates 1% w/v agar (Formedium) was added prior to autoclaving.

2.1.4 Bacterial media- Power broth™

Power broth™ (Molecular dimensions) powder was purchased from molecular dimensions and prepared as per manufacturer's recommendations. 5.2% w/v Power broth was dissolved in dH_2O with 0.4% v/v glycerol, and autoclaved before use.

2.1.5 Bacterial media- AIM

Bacterial auto-induction media (not suitable for $BL21^{AI}$ protein expression) consisted of 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. Overnight cultures were centrifuged, re-suspended in an equal volume of AIM, and this stock used to

inoculate Litre-cultures so that OD_{520} = 0.05. Cultures were then allowed to grow at 18-20°C for 16-24 hours.

2.1.6 Yeast non-selective media- YPAD

Untransformed MaV203 yeast were grown overnight in liquid YPAD (1% w/v yeast extract, 2% w/v bactopeptone, 2% w/v glucose, 0.003% w/v adenine hemisulphate, pH 6.0) prior to transformation. Untransformed yeast were propagated by streaking on solid YPAD agar plates (YPAD supplemented with 2% w/v agar).

2.1.7 Yeast selective media

Yeast co-transformants were selected for by plating transformed cells on double dropout synthetic-complement (SC) media plates lacking leucine and tryptophan (0.67% w/v Yeast nitrogen base without Amino Acids, 0.16% w/v yeast synthetic drop-out medium supplements, 0.0075% w/v histidine, 0.0075% w/v uracil made up in dH_2O and pH to 5.6 prior to addition of 2% w/v agar. SC-LT was then autoclaved and 2% w/v glucose added when cooled <50°C.).

Triple-dropout plates (SC-LTH) were produced as described above, with the exclusion of histidine from the medium.

2.1.8 Protein purification buffers

2.1.8.1 Buffers for extraction of protein from bacteria

IMAC column equilibration buffer, referred to throughout as A1 buffer, consisted of 50mM Tris-HCl pH 8.0, 500mM NaCl, 50mM glycine, 20mM imidazole, and 5% v/v glycerol. When used as a bacterial lysis buffer, 50mL A1 was supplemented with 1 cOmplete™ EDTA-free protease inhibitor cocktail (Roche) tablet. IMAC column elution buffer was produced as A1, but with 500mM imidazole. Gel filtration equilibration and running buffer, referred to throughout as A4 buffer, consisted of 20mM HEPES pH 7.5 and 150mM NaCl.

2.1.8.2 Buffers for extraction of protein from insect cells

For purification of proteins from *Sf9* expression, IMAC column equilibration was performed with 50mM Tris-HCl pH 7.5, 500mM NaCl, 30mM imidazole (referred to as IC A1). Elution of proteins from Ni^{2+}-NTA columns was performed using IC A1 with 500mM imidazole. To make insect cell lysis buffer, IC A1 was supplemented with 0.2% v/v TWEEN 20 and 1 EDTA-free cOmplete™ protease inhibitor cocktail tablet per 50mL. Gel filtration was performed with a superdex™ S75 26/600 GL column pre-equilibrated with running buffer (20mM Tris pH 7.5, 200mM NaCl, 1mM TCEP, referred to as IC A4)

2.2 DNA techniques

2.2.1 Plant and Bacterial expression vectors

Vectors used for expression of proteins in *N. benthamiana*, *E. coli* and *sf9* (insect cell) heterologous systems are listed in Table 2.2.1.

Vector	Expression system	Features	Cloning method	Reference
pOPIN-S3C	Bacterial/ Insect	N-terminal histidine-tagged SUMO tag	Infusion®	OPPF
pOPIN-F	Bacterial/ Insect	N-terminal histidine tag	Infusion®	Berrow 2007[111]
pOPIN-M	Bacterial	N-terminal histidine-tagged MBP tag	Infusion®	Berrow 2007[111]
pESPRIT	Bacterial	N-terminal histidine tag, C-terminal biotin-acceptor peptide	Restriction-ligation	Yumerefendi 2010[112]
pOPIN-TRBO	*N. benthamiana*	CaMV35S promoter, pOPIN-F infusion cassette	Infusion®	Linbo 2007 (modified by R. Hughes)
pK7-FLAG (PVX derivative)	*N. benthamiana*	CaMV35S promoter, N-terminal FLAG tag	Gateway®	Oh 2009[113], Chapman 1992[114] (modified by M. Franceschetti)
pCB302-3	*N. benthamiana*	CaMV35S promoter	Gateway®	Hamada 2004[115]

Table 2.2.1 Vectors used for heterologous expression in this project

2.2.2 Cloning

Cloning was performed using two ligation-independent techniques: Gateway® (Invitrogen) and Infusion® (Clontech) reactions.

Cloning into gateway-compatible vectors was performed in a two-step process- firstly cloning purified PCR product into a DONR vector, and secondly cloning the target gene into a destination vector.

DONR vectors were created by adding 100-150ng purified PCR product with flanking *att*B sites to 100ng purified vector, containing *att*P sites, and incubating for either 1 hour at room temperature or 16 hours at 4°C with 1µL of BP clonase ® mix II (Invitrogen) in a total volume of 5µL. BP clonase® mix II contains both an integrase and integration host factor proteins to recognise the *att* sites and catalyse recombination. 1µL proteinase K was added to the mixture and incubated for 15 minutes at 37°C. The plasmid mix was cooled on ice before being transformed into DH5α *E. coli*.

Sequence-verified DONR vector was purified as described in Section 2.2.5. 100µL of purified vector with *att*L sites flanking the insert were added to 100µL purified destination vector, which carry *att*R sites, and 1µL LR clonase mix to a total volume of 5µL. Like the BP clonase mix, LR clonase mix also contains integrase and integration host factor proteins, but also includes exiscionase proteins to facilitate *att* site recognition and recombination between DONR and Destination vectors. Plasmid mix was incubated for either 1 hour at room temperature or overnight at 4°C. Reaction mix was subjected to treatment with 1µL proteinase K for 15 minutes at 37°C before chilling on ice and transformation into DH5a.

Infusion® cloning involves incubation of adaptor-sequence-flanked PCR product and linearised vector, resulting in partial 3'->5' digestion by the proof-reading activity of proteins in the Infusion® cloning reaction mix to give complementary single-strand overhangs. The overhangs between insert and linearised vector then anneal and are resistant to further exonuclease activity[116]. 15bp flanking sequences are introduced to target inserts by PCR, allowing this method of cloning to be essentially sequence independent.

Infusion® vectors pOPIN-F, pOPIN-M and pOPIN-S3C were linearised by incubation with KpnI and HindIII. When exposed to 3'->5' exonuclease activity, the KpnI overhang site is lost, whilst the HindIII overhangs are maintained. The resulting

fusion of adaptor-flanked insert into this linearised vector leads to the formation of a 3C protease cleavage site upstream of the target protein.

100ng of KpnI and HindIII linearised pOPIN vector was mixed with 50-100ng of purified PCR product (Section 2.2.3) and 1µL of Infusion® HD enzyme mix to a total volume of 5µL. Reactions were incubated at 42°C for 30-60 minutes before transforming into DH5α (Section 2.3.3). Transformants were plated on LB agar plates supplemented with carbenicillin, IPTG, and X-gal to allow blue-white selection.

2.2.3 PCR

Polymerase chain reactions (PCRs) were carried out for two purposes. Firstly to amplify target DNA sequences, and add adaptor sequences for downstream cloning. Secondly to confirm the presence of a target construct in bacterial colonies following transformation (colony PCR).

Colony PCR was performed using MYTAQ™ (Bioline) following manufacturer's recommendations. Typically reactions were performed using vector-specific forward primer and insert-specific reverse primers. Reactions consisted of 10µL MYTAQ 5x reaction buffer, 1.5U MYTAQ™ enzyme, 0.2µM of each dNTP, 0.4µM each primer, 1µL saturated bacterial culture, and made up to 50µL with dH_2O

PCR for downstream cloning used Velocity DNA Polymerase (Bioline). Reaction mixtures were as described above, however using 5x Hi-Fi reaction buffer, 2.0U Velocity enzyme and 20-50ng purified DNA template.

PCR reactions were performed in a T3000 Thermocycler (Biometra), with PCR programmes following the Polymerase manufacturers' recommendations.

2.2.4 DNA oligonucleotides

DNA oligonucleotides were synthesised to order by Integrated DNA technologies using their custom DNA oligonucleotide synthesis service. Oligonucelotides were delivered lyophilised, and so were dissolved in dH_2O to generate a stock solution of 100µM, which was stored at -20°C. Working stocks were produced by dilution with dH_2O to 20µM and also stored at -20°C.

2.2.5 Plasmid purification

Purification of plasmid DNA was performed using Bioline Isolate II plasmid mini kit, according to manufacturer's guidelines. Overnight cultures were pellet by centrifugation at 4,000*g* for 10 minutes and the supernatant removed. Pellets were

resuspended in 500µL resuspension buffer supplemented with RNAse and transferred to 2mL eppendorf tubes. 500µL alkaline lysis buffer was added and mixed by gentle inversion. Samples were left at room temperature for circa 5 minutes before 600µL neutralising buffer added. Samples were mixed by several rounds of gentle inversion and cell debris pelleted by centrifugation at 13,300*g* for 30 minutes. Supernatant was transferred to spin columns and centrifuged for 30 seconds at 10,000*g* to bind DNA. Flow through was removed and membranes washed with both wash buffer 1 and wash buffer 2. Membranes were dried to remove residual wash buffer and ethanol by centrifugation at 10,000*g* for two minutes, with the spin columns then placed into fresh 1.5mL eppendorf tubes. 30-50µL deionised water was spotted onto the centre of the membranes and incubated at room temperature for 5-10 minutes. DNA was then eluted by centrifugation at 10,000*g* for 1 minute.

2.2.6 Gel electrophoresis

DNA was separated on 1.0% w/v agarose gels in TAE (40mM Tris-acetate pH 8.0, 1.0mM EDTA) with 5µg/mL ethidium bromide. DNA samples were mixed with 4x loading buffer (12% w/v ficoll 400 and 0.25% w/v Orange G). Quick-Load® 1-kb DNA ladder (NEB) was loaded into one lane for molecular weight determination.

Separation was performed using horizontal electrophoresis performed at 80V for 10-20 in 1x TAE buffer.

2.2.7 Purification of DNA from agarose gels

Gels were visualised using a UV transluminator and appropriate bands excised with a fresh razor. DNA was then purified from the agarose using Nucleospin® Gel and PCR clean-up kit (MACHEREY-NAGEL) as per manufacturer's user manual.

2.2.8 Sequencing

Sequencing reactions were performed by Lightrun™ sequencing (GATC-biotech). Sequence quality was visually assessed using 4peaks (Nucleobytes).

2.3 Bacterial techniques

2.3.1 Bacterial strains used

Medium-throughput expression screens used several proprietary bacterial strains, predominantly based on BL21 (DE3) *E. coli*. There strains were: BL21 (New England BioLabs), BL21 Star™ (ThermoFisher, referred to as BL21*), BL21^{AI}

(ThermoFisher) SoluBL21™ (amsbio), Lemo21 (NEB), Rosetta™ 2 (Merck Millipore) and Shuffle® T7 competent cells (NEB).

Plasmid storage and propagation was performed using DH5α[117], with the exception of Gateway®-compatible empty vectors, which were stored in *ccdB* survival™ 2 competent cells (Invitrogen).

Glycerol stocks were made with 0.6mL saturated LB overnight culture plus 20% glycerol, stored in 1.5mL sterile eppendorf tubes and stored at -80°C.

Agrobacterium tumefaciens GV3101[118], which carry an inactive tumour-inducing co-plasmid pMP90[119] were used for transient transformation of *Nicotiana benthamiana* leaves. Cells were grown on L media supplemented with rifampicin and gentamycin for chromosomal and inactive pMP90 selection respectively.

2.3.2 Preparation of chemically competent *E. coli*

Chemically competent *E. coli* were transformed as described by Hanahan[117] with some modifications. A single colony of competent cells was selected from a freshly-streaked LB-agar plate and used to inoculate 10mL of LB broth, subsequently grown at 37°C overnight (~16h) with constant shaking at 200rpm. This overnight culture was then used to inoculate 200mL fresh LB broth, which was grown under conditions described above until an OD_{600} = 0.3 was reached. The cells were then chilled on ice for 10 minutes prior to being harvested by 10 minutes centrifugation at 1,500*g* at 4°C. Supernatant was removed and the pellets re-suspended in 80mL TFB1 (100mM rubidium chloride, 10mM calcium chloride, 50mM manganese chloride, 30mM potassium acetate, 15% v/v glycerol, pH 5.8). Cells were again pelleted by centrifugation, supernatant discarded and re-suspended in 8mL TFB2 (10mM MOPS pH 6.5, 10mM rubidium chloride, 75mM calcium chloride, 15% v/v glycerol). 50µL of re-suspended cells were then aliquoted into pre-chilled sterile 1.5mL Eppendorf tubes, flash-frozen in liquid nitrogen, and stored at -80°C until use.

2.3.3 Transformation of chemically competent *E. coli*

For transformations, aliquots of competent cells were thawed on ice, and 50-100ng DNA added. Cells were incubated on ice for a further 30-45 minutes before heat shock treatment of 30-60 seconds at 42°C, followed by the addition of 450µL LB and incubation on ice for one minute. Cells were left to recover for 45-60 minutes by incubation at 37°C with shaking at 180-200rpm. Typically 50µL or 100µL of recovery were plated on LB-agar supplemented with a selective antibiotic, and incubated at 37°C overnight.

2.3.4 Transformation of electro-competent *A. tumefaciens*

Electro-competent *A. tumefaciens* strain GV301 were thawed on ice for ~10 minutes before being added to a pre-chilled electroporation cuvette (GeneFlow) and 1µL purified plasmid added (30-100ng). Cells were then shocked at 2,500V, and recovered for 1 hour incubation in 450µL of LB at 28°C with constant shaking. 20µL and 40µL aliquots of recovery were plated on L or LB agar supplemented with rifampicin, and spectinomycin or kanamycin as appropriate.

2.4 High-throughput screening of R3a constructs

High-throughput screening of R3a constructs, which had undergone time-course digestion was performed as described in Yumerefendi[112] and An[120], but shall be recapped here. Generation of plasmid library was performed by Sri Purhanik. Western blotting and analysis was performed with Philippe Mas.

2.4.1 DNA digestion

Six plasmids, corresponding to the three N-terminal and 3 C-terminal designed regions of R3a, were codon optimised for *E. coli* expression any synthesised by Shinegene Molecular Biotech Inc. These plasmids were then transformed into DH5α, with single colonies used to inoculate fresh LB media and incubated overnight at 37°C. Cells were pelleted by centrifugation and DNA extracted by alkaline lysis. Cells were re-suspended in 50mM glucose, 25mM Tris pH 8.0, 10mM EDTA, to which 2 volumes lysis buffer (0.2M NaOH, and 1% w/v SDS) was added and incubated at room temperature for 5 minutes with constant rotation. Neutralisation buffer (5M potassium acetate, 13% v/v acetic acid) was added and mixed thoroughly before samples were sonicated to remove cell debris. Supernatant was added to a fresh tube on ice and DNA extracted by adding an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and incubation at room temperature for 10 minutes with constant rotation. Samples were centrifuged at 5,000*g* for 10 minutes at 4°C, and the top 15mL of aqueous phase transferred to a fresh tube. 0.1 volumes 3M sodium acetate and 2.5 volumes isopropanol was added, mixed, and incubated for 1 hour at -80°C. Supernatant was removed and 70% ethanol added to the pellet, centrifuged for 15 minutes, and supernatant removed and sample air dried. Salts and RNA were removed using a QIAprep spin miniprep kit (QIAGEN), as per manufacturer's recommendations.

Vectors were linearised by digestion with NsiI and NotI (LRR libraries), or AscI and AatII (CC-NBARC) libraries. Restriction enzymes were purchased from NEB, and

reactions carried out as per manufacturer's guidelines. DNA was extracted and purified as described in Section 2.2.7. Time-course digestion was performed with 33.3ng/µL DNA and 100U/µg DNA ExonucleaseIII. Reactions were performed with 4µg DNA in NEB reaction buffer1 with an additional 5mM NaCl at 22°C in a total volume of 120µL. 2µL of digestion mix was removed each minute and quenched in NT buffer (Nucleospin® extract kit II Machery-nagel), with quenched samples heated to 70°C for 10 minutes. Digested plasmid were then desalted using Nucleospin® extract II kit, before treatment with Mung Bean Nuclease for 30 minutes at 30°C to produce blunt ends for ligation. Blunt-ended vectors were cleaned using Nucleospin® extract kit, and treated with *pfu* for 10 minutes at 72°C.

2.4.2 DNA library purification

Polished, blunt-end plasmids were separated on a 0.5% Agarose gel (Section 2.2.6) performed at 80V for 1-2 hours at 4°C to generate library smears. High molecular-weight, medium molecular-weight and low molecular weight regions were then excised using a sterile razor, DNA extracted and purified for each sub-library separately as described in Section 2.2.7.

2.4.3 Bacterial transformation and picking

Purified libraries were ligated using a Rapid ligation kit (Roche) and transformed into 50µL of $BL21^{AI}$ cells as described in Section 2.3.3, with 10µL and 100µL aliquots of recovery plated to determine the transformation efficiency. Recovery mixture was plated on 22 x 22cm Genetix plates with LB agar with kanamycin to produce 3x oversampled libraries and incubated overnight.

Once well-defined, opaque colonies were visible they were picked using a colony-picking robot and deposited in 384-well plates containing TB media (1.2% w/v Typtone, 2.4% w/v Yeast extract, 0.4% v/v glycerol) plus Hogness modified freezing medium (HMFM- 25mM K_2HPO_4 10mM KH_2PO_4 , 2.85mM disodium citrate, and 3% glycerol, 250µM $MgSO_4$) and 0.1M phosphate buffer. Plates were then incubated at 37°C overnight with constant shaking. Once cultures had reached saturation, they were stored at -80°C as master plates.

2.4.4 High-throughput protein expression

384-well plates were thawed at room temperature before being transferred to identical plates containing only TB media, kanamycin and chloramphenicol, which were grown overnight at 37°C with constant orbital shaking.

Overnight cultures were then stamped in triplicate onto Hybond Extra nitrocellulose membranes (Amersham) placed on top of two 22 x 22cm LB-agar plates. These plates were incubated at 37°C overnight to allow visible colonies to form. Membranes were then transferred from LB-agar media plates to plates supplemented with 50μM biotin and 0.2% arabinose to induce expression. These plates were then incubated for 4-5 hours at 25°C.

2.4.5 Cell lysis

Cells were lysed by placing blotting paper that had been pre-soaked in alkaline-lysis solution (0.5M NaOH, 1.5M NaCl) directly on top of the nitrocellulose membranes and incubating for 10 minutes. Nitrocellulose membranes were then transferred to blotting paper pre-soaked with neutralising solution (1M Tris-HCl pH 7.5, 1.5M NaCl) and incubated at room temperature for 5 minutes, with this neutralisation step repeated 3-4 times. Membranes were then incubated with 2x SSC buffer (0.3M NaCl, 30mM trisodium citrate) for 15 minutes and cell debris removed by gentle scraping with a glass spreader.

2.4.6 Western blotting

Once cell debris had been removed membranes were washed with PBS + 0.1% v/v TWEEN 20 (PBS-T). Membranes were then blocked overnight at 10°C with SuperBlock™ buffer (ThermoFisher). Following blocking, membranes were washed in PBS-T, and then incubated with mouse α-his tag antibody for 1 hour at 10°C. Following three rounds of washing with PBS-T, membranes were incubated with rabbit α-mouse secondary antibody conjugated to Alexa532, and Streptavidin conjugated to Alexa488 for 1 hour at 10°C with constant rotation. Membranes were washed three times in PBS-T and once in dH_2O for 5 minutes before being loaded onto a Typhoon 9400 scanner (GE Healthcare). Data collection was optimised at excitation λ= 488nm with a 520 BP emission filter, and at excitation λ= 532nm with 555 BP emission filter to reach saturation with the highest-intensity spots.

2.4.7 Protein expression of best constructs

Once signal intensity had been recorded for each spot using Visual grid, the data were exported to Microsoft Excel for triplicate averaging and ranking. Averaged signal intensities were normalised against average histidine background signal, with constructs with normalised signal intensity ≤1 omitted from further data analysis. Biotin signals were averaged and normalised as described above, and normalised values used to rank constructs. The top 48-ranking constructs from the N-terminal

libraries and top 47 ranking constructs for the C-terminal libraries were selected for further experiments.

10µL of each top-ranking construct was transferred from thawed TB-HMFM master plates to a 96 deep-well plate with 1mL TB media supplemented with kanamycin and chloramphenicol, with MBP added as a positive control to well 96. This plate was then incubated overnight at 37°C with constant shaking, with 40µL of saturated colony transferred to 24 deep-well plates containing 4mL TB media plus kanamycin and chloramphenicol. This 96-well plated was then incubated for a further 8-10 hours at 25°C, 10% glycerol added, and stored at -80°C.

24-well plates were incubated at 37°C for 3-4 hours, until OD_{600} approached 0.6-0.8, and induced by the addition of 0.2% L-arabinose. Protein expression was performed at 25°C overnight.

2.4.8 Low-stringency purification and western blot of best constructs

Cells were pelleted by centrifugation, supernatant removed, and pellets re-suspended in 3mL re-suspension buffer (20mM Tris pH 8, 250mM NaCl, 20% sucrose). After 10 minutes resuspension by constant shaking at 10°C, re-suspended cells were supplemented with 0.25mg/mL lysozyme and incubated for 20-30 minutes at 4°C with constant shaking. Sphaeroplasts were centrifuged for 20 minutes at 2,250*g*, supernatant carefully removed, and pellets frozen at -80°C for 30 minutes.

Sphaeroplasts were lysed by thawing at room temperature with the addition of 10mM Tris-HCl pH 7.0, 0.5% Brij and 10U Benzonase and incubated for 10 minutes. Plates were centrifuged at 4°C for 30 minutes at 2,250*g* and supernatant loaded into 96-well plates containing 100µL Ni^{2+}-NTA resin (QIAGEN) which had been washed in wash solution (50mM sodium phosphate pH 7.0, 300mM NaCl, 5mM imidazole), and incubated for 30-60 minutes at 4°C. Beads were washed and bound proteins eluted (50mM sodium phosphate pH 7.0 300mM NaCl, 300mM imidazole) using a Tecan liquid handling robot. Samples were then separated by SDS-PAGE run at 75mA for 140-150 minutes, transferred to membranes for western blotting (Section 2.4.6), and probed as described in Section 2.4.6.

2.5 Rational design of protein constructs

The translated sequences for *S. tuberosum* R3a, *S. lycopersicum* I-2, *S. demissum* R2, *S. bulbocastrum* Rpi-blb2, *S. lycopersicum* NRC1 and *N. benthamiana* NRC1-like were subject to bioinformatic analysis using a suite of different online servers.

PFAM was used to identify likely boundaries for the NBARC domain. Typically the predicted C-terminal boundary occurred prior to the conserved “MHD” motif, and so this boundary was extended to include this region. The extent of this addition was determined by predicted secondary structures after this motif and predictions of disordered regions. NBARCs were typically extended to include an additional β-strand following the MHD motif and following α-helices until a region of significant disorder was identified.

Secondary structure predictions were performed by PHYRE2[121], which also generated 3D homology models. 3D structures were also used as a guide to identify self-contained structures, however when designing boundaries secondary structures took precedence, i.e. if a homology model predicts termination of a secondary structure into a disordered loop region, whereas secondary structure predictions suggest that the secondary structure element extends into this putative disordered region, the construct boundaries are extended to include this 2° structure predication.

Disorder predictions were performed by RONN[122], and likely LRR regions identified by LRRfinder (http://www.lrrfinder.com/), to supplement the information described above.

2.6 Medium-throughput *E. coli* expression screening

Medium-throughput screening in *E. coli* was performed by transferring single colonies to 24-well deep-well plates with 4mL LB and appropriate antibiotic. Plates were incubated overnight at 37°C with constant shaking, and 200-500µL saturated overnight culture transferred to a fresh 24-well deep well plate with 4mL of LB, Power broth™ or AIM. These plates were then incubated at 37°C with constant shaking for 3-6 hours, until average OD_{600} approximated 0.5-0.8. Once the target OD was reached plates were put on ice to rapidly chill. After ~5 minutes wells were supplemented with the appropriate treatment (0.4-0.5% ethanol, 0.4-0.5% DMSO, 5-10% glycerol) and induced by the addition of either 1mM IPTG or 0.2% L-arabinose.

Cells were harvested by centrifugation in a Sorval Legend centrifuge at 2,500*g* for 15 minutes at 4°C. Supernatant was discarded and pellets re-suspended in 400-800µL A1 buffer supplemented with either cOmplete EDTA-free protease inhibitor cocktail, or 0.5mM PMSF. Plates were put on ice and cells lysed by sonication for 1-2 minutes, with sonication performed for 1 second followed by 3 seconds pause.

Cell debris was removed by centrifuging plates in a Sorval Legend centrifuge for 45 minutes at 2,500*g* at 4°C.

30µL of supernatant was mixed with 10µL 4x SDS-loading buffer and heated to 95°C for 3-5 minutes, before loading onto a 17% SDS-PAGE gel. Gels were run for 50-60 minutes at 120V and stained with InstantBlue™ coomassie® stain.

2.7 Transient transformation of *N. benthamiana*

Transient transformation of *N. benthamiana* was performed as described in Kapila[123] with the following modifications: 10-20µL of *A. tumefaciens* glycerol stock derived from a single colony were plated on LB agar supplemented with the appropriate antibiotics, and allowed to grow for two days at 28°C. The resulting lawn was then transferred to sterile 5mL tubes and re-suspended in 2mL infiltration buffer (10mM MES pH 5.6, 10mM MgCl2, 150µM acetosyringone) for OD measurements. Cultures were then diluted to the appropriate OD and left for 2-5 hours before being syringe infiltrated into leaf three or four of 4-week old *Nicotiana benthamiana*.

For pOPIN-TRBO R3a domain expression and purification *A. tumefaciens* were re-suspended to an OD_{600} of 1.2. For R3a-AVR3a interaction studies, an OD_{600} of 0.3 for effectors in pK7-FLAG and 0.5 for R3a in pCB302-3 was used.

2.8 Protein purification

2.8.1 Purification of soluble protein from *E. coli*

50mL of LB broth was inoculated by scraping cells from a glycerol stock and incubated for 16h at 37°C with constant shaking at 200-230rpm. 20mL of this overnight culture was used to inoculate 1L fresh media, which was incubated as above until the target OD_{600}= 0.4-0.7 was reached. Depending on the optimal induction regime determined by small-volume screening, expression was then either initiated by addition of 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG), or IPTG addition delayed until cultures had been chilled on ice for 5-10 minutes and treated with either 0.4-0.5% dimethyl sulfoxide (DMSO), 0.4% ethanol, or 10% glycerol. For protein expression from $BL21^{AI}$ cells induction was performed in an identical manner, however with the addition of 0.2% arabinose instead of IPTG. Post-induction cultures were incubated for 16-20h at 18°C with constant shaking at 200-230rpm.

Cells were harvested by 10 minutes centrifugation at 6,000*g* in an Evolution RT floor centrifuge (Sorvall), the supernatant discarded and cell pellets stored at -80°C until

required. Pellets were re-suspended in buffer A1 (50mM Tris-HCl pH 8.0, 50mM glycine, 5% glycerol, 500mM NaCl and 20mM imidazole) plus protease inhibitor cocktail, and lysed by sonication for 1-3 minutes, depending on cell density, at 40% amplitude. After sonication the lysed cells were transferred to oakridge centrifuge tubes and centrifuged at ~37,000*g* for 30 minutes in an Evolution RC centrifuge (Sorvall) to separate insoluble cell components. The resulting supernatant was then diluted with A1 buffer plus protease inhibitor to a volume of approximately 50mL per litre original culture and flowed through nickel immobilised metal affinity column (IMAC) to collect histidine-tagged proteins.

Immobilised proteins were washed with three column volumes of A1 buffer before elution in 10mL buffer B1. The eluate was then loaded on to a HiLoad superdex 75 26/600 preparative gel filtration column, pre-equilibrated with A4 buffer (20mM HEPES pH 7.5, 150mM NaCl). 8mL fractions were collected in clean 24-deep-well plates and stored at 4°C until use.

2.8.2 Generating SeMet protein from E. coli

E. coli carrying the pOPIN-S3C:NRC1^{NBARC} were grown overnight in minimal media consisting of M9 (0.6% w/v Na_2HPO_4, 0.3% w/v KH_2PO_4 , 0.05%w/v NaCl, 0.1% w/v NH_4Cl) 0.4% w/v glucose, 2mM $MgSO_4$, 0.1mM $CaCl_2$, 0.8% w/v each Val, Phe, Ile, Leu, glu, Lys, Arg, Ser, Thr, Tyr, His, Gln, Trp, 0.001% thiamine, supplemented with Carbenicillin. Cells were pelleted by centrifugation at 5,000*g* for 10 minutes and washed twice in 20mL M9 plus carbenicillin, before being re-suspended in 20ml M9 and Carbenicillin. This was used to inoculate 2xL of minimal media (described above), and cultures incubated at 37°C for 6-8 hours with constant shaking. When cells reach an OD_{600} of 0.3, 0.01% w/v Thr, Lys, Phe, 0.005% w/v Leu, Ile Val, and 0.006% w/v selenomethionine was added. Cells were grown for a further 45 minutes at 37°C before induction overnight at 18°C was initiated by the addition of 1mM IPTG.

2.8.3 Purification of proteins from Insect cells

Proteins were purified from *Sf9* pellets as recommended by the OPPF (https://www.oppf.rc-harwell.ac.uk/OPPF/protocols/index.jsp). Frozen cell pellets were re-suspended in ~30mL lysis buffer (50mM Tris-HCl pH 7.5, 500mM NaCl, 30mM imidazole, 0.2% TWEEN20**,** and cOmplete EDTA protease inhibitor cocktail tablets) and lysed using a cell disruptor (Constant Systems Ltd) at 14,000psi. Following lysis, 10-50µL benzonase was added, and incubated at room temperature for 15 minutes. Cell debris was removed by centrifugation for 30-40 minutes at

30,000g in an Evolution RC floor centrifuge (Sorvall) at 4°C. Supernatant was passed through minisart® 0.45µM filters (Sartorius), with additional incubation with benzonase if required, before loading onto a 5mL His-trap FF column pre-equilibrated with equilibration buffer (50mM Tric-HCl pH 7.5, 500mM NaCl, and 30mM imidazole). Bound proteins were eluted by the addition of 500mM imidazole with eluate loaded directly onto a sepharose S75 gel filtration column pre-equilibrated with IC running buffer (30mM Tris-HCl pH 7.5, 200mM NaCl, 1mM TCEP).

Fractions containing the target protein were pooled and concentrated (Section 2.8.7) to <20mL before tag removal by incubation with 3C protease at 4°C overnight. Tagged protein and protease was removed by subtractive IMAC using a 5mL His-trap FF column pre-equilibrated as described above. Unbound protein and 3x column-volume washes were collected and assessed for purity by SDS-PAGE. Buffer exchange was performed by several rounds of concentration (Section 2.8.8) and dilution in A4 buffer before being concentrated to ~8mg/mL.

2.8.4 Protein purification from *N. benthamiana*

Four-to-five-week old *N. benthamiana* leaves were harvested (three-to-four days post-infiltration for transient expression), the mid-vein removed, frozen in liquid nitrogen and stored at -80°C. Tissue was then ground to a powder in liquid nitrogen, and 0.5g powder added into pre-chilled 2mL screw cap tubes. Each tube was left on ice for 1 minute before 1.0mL extraction buffer (10% glycerol, 25mM Tris-HCl pH 7.5, 1mM EDTA, 150mM NaCl, 10mM DTT, 0.1% TWEEN 20, and 0.02% w/v poly(vinylpolypyrrolidone)) was added, the tubes vortexed thoroughly and then left on ice for 30 minutes. After incubation samples were again thoroughly vortexed before centrifugation at 30,000*g* for 10 minutes to pellet cell debris. Supernatant was then transferred to fresh, chilled 1.5mL eppendorfs and centrifuged as above until no plant debris was present in the supernatant, which was then transferred to fresh Eppendorf tubes for imunnoprecipitation experiments (Section 2.14.2) or western blotting (Section 2.15).

For larger-scale purification of R3a domains, 3-4 *N. benthamiana* leaves 5 days after infiltration were harvested and treated as described as above. Typically 1.5 - 2.5g of ground tissue was transferred to chilled tubes to which 2x v/w extraction buffer (described above) was added and mixed by vortexing. After 30 minutes incubation on ice with 2-3 rounds of additional vortexing, tubes were centrifuged for 45 minutes at 4,500rpm in a Legend® RT centrifuge (Sorvall) at 5-10°C.

Supernatant was then passed through 0.45µm filters before being loaded onto a pre-equilibrated IMAC column. Low stringency purifications involved the addition of 15mM imidazole to samples prior to loading onto a 5mL His-trap FF column, high stringency purifications were performed by the addition of 30mM imidazole prior to loading on a 1mL His-Trap FF column. Columns were then washed with 3x column volume of A1 buffer before bound proteins eluted by washing with 2-3 column volumes B1 buffer.

2.8.5 SDS-PAGE

SDS-PAGE was performed as described by Laemmelli with some modifications. 12% or 17% w/v polyacrylamide resolving gels were made by diluting 30% acrylamide-bis-acrylamide (37.5:1) solution (Severn Biotech ltd) in 375mM Tris-HCl pH 8.8 with 0.1% w/v SDS. Immediately prior to pouring 0.1% w/v ammonium persulphate and 0.04% v/v N,N,N',N'-tetramethylethylenediamine were added and mixed. Resolving gels were poured between two Mini-Protean® system (BioRad) glass plates with 1mm separation to circa 2mL below the top of the shorter plate, and water-saturated butanol carefully applied to the top of the gels whilst setting to ensure a level boundary between resolving and stacking gels. Butanol was removed prior to pouring 5% stacking gels (5% acrylamide-bis-acrylamide (37.5:1) solution, 63mM Tris-HCl pH 6.8 0.1% w/v SDS, with 0.1% w/v ammonium persulphate and 0.1% v/v N,N,N',N'-tetramethylethylenediamine added immediately prior to pouring) to the top of the gels. Plastic combs were inserted into the stacking gel whilst liquid. Once set these gels were wrapped in clingfilm and stored in a humid environment at 4°C.

Protein separation was performed by placing gels in a BioRad miniPROTEAN® tetra cell system. The plastic combs were removed to allow sample loading, and chambers filled with SDS-running buffer (25mM Tris-HCl, 250mM NaCl, and 0.1% w/v SDS) ensuring that gels were covered by running buffer. Protein samples were prepared by heating to 95°C for 3-4 minutes in 4x SDS-loading buffer (final concentrations: 50mM Tris-HCl pH 6.8, 0.1M DDT, 2% w/v SDS, 0.1% bromophenol blue and 10% glycerol). To enable molecular weight determination, protein samples were loaded alongside a molecular weight ladder- either SeeBlue® Plus 2 pre-stained protein standard (ThermoFisher), or RunBlue prestained dual colour markers (Expedeon) for SDS-PAGE coomassie staining, or PageRuler™

Plus prestained protein ladder (ThermoFisher) for protein separation preceding western blotting experiments.

2.8.6 Staining SDS-PAGE gels

Gels were stained by incubation with InstantBlue™ Coomassie stain at room temperature for a minimum of 30 minutes before de-staining with de-ionised H_2O.

2.8.7 Protein concentration measurement

As standard protein concentration was determined by measuring absorbance at 280nm using a nanodrop spectrophotometer. Protein concentration of wild-type NRC1 NBARC domain and point mutants for enzymatic and binding assays was determined by infrared absorbance using a direct-detect® spectrophotometer (Merck Millipore). Protein concentrations for crystallisation experiments refer to A_{280} measurements.

2.8.8 Protein concentration

Purified protein was concentrated using Sartorius protein concentrators. 10kDa molecular-weight cut-off concentrators were used for NRC1 NBARC domain; 5kDa cut-off concentrators were used for all other proteins. Proteins were concentrated by multiple rounds of centrifugation at 7,000*g* for 15 minutes until the target concentration was achieved.

2.9 Dynamic light scattering

Dynamic light scattering (DLS) was performed using a DynaPro™ Titan dynamic light scattering instrument (Wyatt Technology). For size determination of NRC1 NBARC particles in solution, protein was diluted to 1mg/mL in A4 buffer before filtering with a 0.1µm Ultrafree® centrifugal PVDF filter (Merck Millipore) by centrifugation on a bench top centrifuge at 12,000*g* for 60 seconds.

Assessment of R3a construct #1 aggregation was performed by diluting 1 volume purified protein in 9 volumes 0.1M buffer before filtering as described above.

DLS readings were taken at 25°C with 10 readings taken every second for 10 seconds per sample. Typically laser power was reduced to 30% for consistent signal intensity. Data were assessed with DYNAMICS V6 software (Wyatt Technology), and species <0.1nm omitted from analysis.

2.10 Yeast 2-hybrid

Yeast (*Saccharomyces cerevisiae*) transformations were performed using the ProQuest® Two-Hybrid with slight modifications. The ProQuest® Y2H system uses the low-copy number (*ARS/CEN*) vectors to reduce overexpression effects, and Gateway®-compatible vectors to allow fusion of genes of interest to GAL4 DNA binding domain and the GAL4 activation domain. Interacting proteins fused to these domains recapitulate the activity of the GAL4 transcription factor. By using a yeast strain that is auxotrophic for histidine, and having the synthesis of the gene required to allow growth on media lacking this amino acid (HIS3) under the control of a GAL4-responsive promoter, interacting proteins can be identified by their ability to grow on histidine-deficient media (Figure 2.10.1).

Figure 2.10.1 Schematic of protein interaction-dependent activation of gene expression in yeast 2-hybrid

In the Proquest® yeast 2-hybrid system HIS3 gene expression is driven by GAL4-responsive promoter elements (Left). Right: GAL4 activity is only recapitulated upon the interaction of two proteins (X and Y), which are fused separately to the GAL4 DNA binding and activating domains. Taken from the ProQuest™ two-hybrid manual, Invitrogen.

2.10.1 Yeast transformation

A single colony of yeast strain MaV203 was placed into 10mL of YPAD liquid media using a flamed loop, and incubated overnight at 28°C with constant shaking at 190-220 rpm. OD_{600} of the culture was measured by adding 100µL of overnight culture to 900µL fresh YPAD, and overnight culture diluted to give 40mL of culture at OD_{600}= 0.4. This culture was split into two sterile falcon tubes, sealed with parafilm and incubated at 28°C with constant shaking for 2-3 hours. Cells were pelleted by centrifugation for 5 minutes at 1,600*g* using a Legend® RT centrifuge (Sorvall®). Supernatant was removed, and pellets washed by re-suspending in autoclaved, de-ionised H_2O. Washed cells were pelleted, with supernatant discarded as above, and cells resuspended in a total volume of 500µL yeast resuspension buffer (5mM Tris-HCl pH 7.5, 100mM Lithium acetate, 5mM EDTA). Cells were then incubated at room temperature for 10 minutes before 10µL of re-suspended cells were added to 10µL pre-prepared DNA samples.

DNA samples consisted of 10µg pre-prepared salmon sperm DNA (sheared by sonication, heat denatured and rapidly chilled on ice) mixed with 200ng of each prey and bait vector. 70µL transformation buffer (10mM Tris-HCl pH 7.5, 100mM lithium acetate, 1mM EDTA, 40% w/v PEG 3350) was added to each yeast-DNA combination and mixed by gentle pipetting. Cells were then incubated for 30-45 minutes at 28°C before the addition of DMSO, mixed by pipetting, and heat-shocked at 42°C for 7 minutes. Cells were pelleted by 1 minute centrifugation in a Sorvall® Legend RT centrifuge at 1,600*g* at room temperature. Supernatant was discarded and cells re-suspended in 100µL autoclaved de-ionised H_2O before being plated onto SC-LT selective agar plates, sealed with parafilm, and incubated at 28°C for 3-4 days.

2.10.2 Yeast 2-hybrid assays

Single colonies of co-transformed yeast were re-suspended in 100µL distilled water in microcentrifuge tubes. 2µL of re-suspended cells were spotted out onto SC-LT and SC-LTH plates supplemented with 10mM, 7.5mM or 5mM 3-amino-1,2,4-triazole (3AT). Plates were incubated overnight at 28°C and inspected for colony growth. Plates were then incubated at room temperature overnight for up to five days, or until a clear growth phenotype could be observed.

2.11 Analytical gel filtration

Analytical gel filtration was performed at 4°C by loading ~100µL of protein onto a superdex™ S75 10/300 GE sepharose column. For NRC1 NBARC size determination, protein was diluted to 50µM in A4 buffer prior to loading. Analytical gel filtration to detect NRC1 CC/NBARC complex formation was performed by diluting NRC1 NBARC to 1mg/mL in the required amount of coiled-coil to give 1:1, 1:2, 1:4 and 1:10 ratios in approximately 120µL. These proteins were then incubated on ice for >1 hour, or at 4°C overnight before loading onto the column.

Calibration curves were produced by Lennart Wirthmueller.

2.12 Circular dichroism

Circular dichroism was performed using a Chirascan™ Plus (AppliedPhotophysics), with spectra measured from 180-260nm. Protein was diluted into phosphate buffer to 0.1mg/mL and spectra gathered in 1nm steps. Readings were taken with empty sample chamber to allow absorption measurements and spectra recorded in triplicate including buffer only samples to allow buffer subtraction.

2.13 SAXS

Small-angle X-ray scattering (SAXS) experiments were performed at Diamond Light Source (Oxfordshire), beamline B21. For each sample 18 frames of 10 seconds exposure were collected. Data were collected for NRC1 NBARC at 10, 5, 2.5, 1.25, 0.625 and 0.3125mg/mL in 20mM HEPES pH7.5 with 150mM NaCl.

Data were processed using Scatter (*Rambo et al*). Frames were visual inspected for radiation damage by plotting I vs Rg and removing outlying frames from further analysis. Buffer frames were averaged and subtracted from the averaged experimental data to allow evaluation of inter-particle interactions with intensity vs scattering angle plots. Rg and porod values were determined manually as recommended in the Bioisis tutorial (http://www.bioisis.net/tutorial).

2.14 Antibody techniques

2.14.1 Conjugation of anti-R3a antibodies to Dynabeads

Sheep α-rabbit magnetic Dynabeads® (ThermoFisher) were prepared as per manufacturer's guidelines. Beads were washed and resuspended in 0.1M PBS with 1mM EDTA, and incubated with 1.2µL α-R3a antibodies per 100µL beads for 30 minutes on ice. Tubes were placed in a magnetic rack and supernatant removed. Beads were then re-suspended in conjugation buffer (20mM sodium phosphate pH 7.5, 150mM NaCl) containing 5mM BS3 (bis(sulfosuccinimidyl)suberate), and incubated at room temperature for 30 minutes with constant rotation. Cross-linking reactions were quenched by the addition of 1M Tris-HCl pH 7.5 to a final concentration of 50mM, and incubated for a further 15 minutes with constant rotation.

To compare the effects of conjugation, beads were washed and re-suspended in conjugation buffer with α-R3a antibodies as described above. Prior to separation on a magnetic rack, half of the solution was removed and placed into a fresh pre-chilled tube on ice. Following conjugation of the remaining half, both sets of beads were washed in 3x volumes of conjugation buffer.

2.14.2 Protein immunoprecipitation

For IP experiments using sheep α-rabbit beads that had not been pre-conjugated to R3a antibodies, clarified plant lysate (Section 2.8.4) was incubated with the primary antibody for 15-30 minutes on ice prior to the addition on 40-100µL of washed sheep α-rabbit Dynabeads®. For IP experiments using Dynabeads pre-conjugated

to α-R3a antibodies, this first incubation step was omitted and 40-100µL washed, conjugated beads was added to clarified plant lysate. Magnetic beads were then incubated with samples for 1-2 hours at 4°C with constant rotation.

Beads were washed by separating beads from solution with a magnetic rack and gentle pipetting off supernatant, followed by resuspension in ~3 volumes 0.1M PBS + 150mM NaCl. This process was repeated three times before bound proteins were eluted in either 50-100µL SDS-loading buffer incubated at 72°C for seven minutes with constant rotation, or 30-50µL sodium citrate pH 2.0 at room temperature for seven minutes with constant rotation. Acid eluted proteins were then neutralised by addition of 3µL 1M Tris-HCl pH 7.5 before the addition of 4xSDS-loading buffer and heating to 95°C for 3-5 minutes.

2.15 Western blotting

2.15.1 Protein transfer

Proteins were transferred to Immobilon-P® PVDF membranes (Merck Millipore) by electrophoresis using a Mini Trans-Blot® cell (Biorad). Transfers were performed in Bjerrum transfer buffer (48mM Tris pH 9.3, 39mM glycine, 20% v/v methanol) supplemented with 0.0125 w/v SDS, with membranes first equilibrated in methanol for 5 minutes. Transfers were performed at 4°C overnight at 90mA.

2.15.2 Antibody probing

Membranes were blocked by incubation with 10-20mL TBS-T with 5% w/v milk powder for 1 hour at room temperature, with blocking solution then removed. Primary antibodies (typically α-R3a antibodies) were added in a 1:8,000 dilution to fresh blocking solution, with 8mL added to each membrane. Antibodies were incubated with the membranes for an hour at room temperature before being removed, and the membranes washed three times in 15-20mL TBS-T. Following washing, secondary antibodies (horseradish peroxidase-conjugated α-rabbit antibodies) were diluted 1:6,000 in blocking buffer, 8mL of which was then added to each membrane, and incubated for an hour at room temperature. This buffer was then removed, membranes washed 3 times in TBS-T and once in TBS. Secondary antibodies were visualised by application of SuperSignal™ West Pico and Femto chemiluminescent substrate (ThermoFisher) combined in a 1:3 ratio. Chemiluminescence was detected by either autoradiogram or using an ImageQuant LAS 500 spectrophotometer (GE Healthcare).

FLAG-tagged proteins were detected using rat α-FLAG primary antibodies and alkaline phosphatase (AP) conjugated α-rat antibodies as described above. AP-conjugated antibodies were visualised by the application of 5-bromo-4-chloro-3-indolyl phosphate (BCIP), and Nitro blue tetrazolium chloride (NBT). In the presence of AP these two substrates precipitate to form insoluble purple and blue dyes respectively. Membranes were incubated with BCIP and NBT until dark bands were visible before being washed with dH_2O.

2.15.3 Ponceau staining of membranes

Staining of total proteins present on nitrocellulose membranes was performed by incubation of membranes with 15-20mL Ponceau S stain (0.1% w/v Ponceau S in 5% v/v acetic acid). Once bands were clearly visible, with minimal background staining, membranes were washed with dH_2O.

2.16 Crystallographic techniques

2.16.1 Initial crystallisation screens

In order to produce crystals suitable for X-ray diffraction experiments, proteins must form large, repeating structures, ultimately leading to the formation of macroscopic crystals. In vapour diffusion experiments these repeating structures are formed in response to increasing protein concentration as water diffuses from the drop, containing a low concentration of solutes, to the reservoir, which contains a higher concentration of solutes (Figure 2.16.1). As the drop and reservoir equilibrate protein concentration increases to the point where it precipitates out of solution. If the conditions allow this precipitation to occur in a slow, controlled manner, then precipitation can lead to crystal formation and growth. The conditions required to support such controlled crystallisation are typically narrow, and unless multiple conditions are met (for example the right precipitant, salt and pH combinations) the result will be uncontrolled precipitation of the protein.

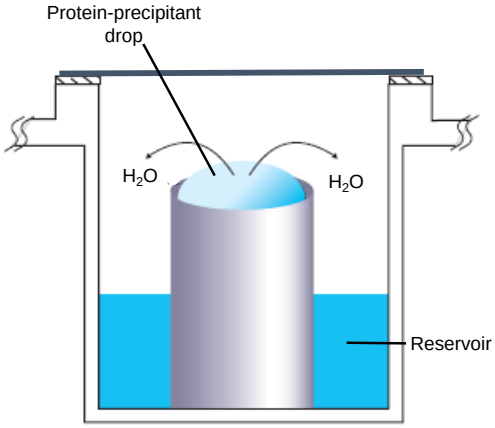


Figure 2.16.1 Schematic of a sitting drop vapour-diffusion experiment

In sitting drop vapour-diffusion experiments, a chamber is partially filled with a crystallisation solution containing, amongst other things, a precipitant. A small volume of this reservoir solution is mixed with protein, and a small drop deposited within the chamber. The chamber is then sealed, and the drop equilibrates with the reservoir conditions, gradually increasing the concentration of protein within the drop. Modified from Hampton Research.

Formation of macroscopic crystals relies on two stages: crystal nucleation and crystal growth. Given the diverse parameters that can affect the ability of proteins to crystallise, initial crystallisation experiments often focus on “sparse-matrix” screens. These screens are designed to test diverse regions of crystallisation space to find conditions that may support crystal formation.

To identify suitable conditions for the crystallisation of NRC1 NBARC domain, seven commercial 96-well screens were used for sitting-drop vapour diffusion experiments:, PEG suite, and Ammonium sulphate suite from QIAGEN; PACT, JCSG , Structure screen, Shotgun screen, and Morpheus screen from Molecular Dimensions. Sitting drop experiments were performed in MRC crystallisation plates (molecular dimensions), which contain two wells per reservoir. Each crystallisation condition was tested using a 1:1 mix (total drop volume 0.6µL) of reservoir solution, with protein at 10mg/mL and 5mg/mL. Plates were sealed and kept at 20°C to for the duration of crystallisation experiments. NRC1 NBARC domain was found to crystallise in <16 hours, typically forming small comet-like crystals in multiple conditions, over a range of pH and with a variety of molecular weight PEGs.

2.16.2 Crystal optimisation

As the conditions that allow controlled protein precipitation are specific for each protein, and influenced by many factors, sparse-matrix screens are unlikely to sample the best conditions for crystallisation. Unless crystals from initial screens are

particularly well formed and large, optimisation experiments are performed to systematically test crystallisation space around these initial conditions to improve conditions for crystal growth.

Crystal optimisations were performed in 48-well sitting-drop MRC Maxi plates (molecular dimensions) around the most promising conditions identified in the original crystallisation conditions: 0.1M MMT (MES, malic acid, Tris- QIAGEN) pH 5.0 and 25% PEG 1500. PEG by pH screens were typically performed using 2% steps in PEG concentration from 21-31%, against 0.5 step changes in pH from 4.0 to 6.0, with a 1:1 ratio of protein to reservoir condition in a total drop volume of 2µL. This optimisation indicated that crystals were improved at lower PEG concentrations. Below pH 4.5 heavy precipitate was observed and above pH 5.0 only clear drops were seen. Repeated rounds of similar optimisations were performed to identify the best crystallisation conditions as 6-12% PEG 1500, 0.1M MMT pH 4.5 (PEG optimum varying between protein purifications).

Similar optimisations were performed by varying protein concentration in 1mg/mL steps from 4-12mg/mL against pH 4.5 conditions with 6-12% PEG 1.5K, with an ideal protein concentration of 6-8mg/mL identified.

2.16.3 Microseeding

Crystal microseeding allows sampling of conditions that may be able to facilitate crystal growth, but are not suitable for nucleation- such as having too low an initial concentration of precipitant. Crystal micro-seed stock was produced from either purified NBARC domain grown in 0.1M MMT pH 4.5 with 8% PEG 1500, or NBARC domain that had been incubated with a 10-fold excess of ATP-γ-S overnight prior to growth in 0.1M MIB buffer (sodium malonate, imidazole, boric acid- QIAGEN) pH 5.0 and 25% PEG 1500. Crystals were mechanically disrupted in the growth drop before being transferred to a fresh Eppendorf. The total volume of the reservoir was transferred to this tube and crystals further disrupted by several short bursts of vortexing. This seed stock was then flash-frozen in liquid nitrogen and stored at -80°C until use.

Sparse-matrix microseeding experiments were performed with an Oryx Nano liquid handling robot as described above, with each drop consisting 0.3µL protein, 0.3µL reservoir condition and 0.1µL seed stock.

2.16.4 Additive screening

Additive screens were performed in two sets of 48-well plates. 130µL of 8% PEG 1500, 0.1M MMT pH 4.5 was added to each reservoir with 20µL of SilverBullets Bio screen (Hamilton Research). Screens were set up using an Oryx Nano liquid handling robot to deposit 1µL of protein at 8mg/mL and 1µL reservoir condition to each well. Plates were sealed and stored at 20°C for crystallisation. Heavy precipitation was seen in all crystallisation conditions in <16 hours. Screens were repeated with 5µL SilverBullet screen added to 145µL reservoir conditions, without any improvement to precipitation.

2.16.5 Cryoprotection of crystals and heavy atom soaking for data collection

Protein diffraction experiments were performed at low temperature, with crystals sitting in a stream of liquid nitrogen having been first flash-frozen in liquid nitrogen (~68K). This flash freezing can result ice crystals forming either within the crystal or in the mother liquor. Ice crystal formation within solvent channels in protein crystals or within the loop holding the crystal can lead to interfering diffraction patterns, making data processing of protein-diffraction difficult. By using cryoprotectants, the formation of ice crystals is significantly reduced. This reduction in ice formation can be performed in two ways- firstly by the replacement of water within solvent channels with a cryoprotectant agent such as ethylene glycol or glycerol, or by replacement of water surrounding the crystal with oils that do not penetrate the crystal. We used the former method to protect crystals upon freezing.

To make cryprotectants that would not stress the crystals, or dissolve them, solutions similar to the reservoir conditions supporting crystal growth were used. Typically cryoprotectants used a concentration of precipitant 2-4% higher than in the corresponding reservoir condition, and supplemented with 20% ethylene glycol. Crystals were transferred from sitting drops in which they formed to small drops of cryoprotectant where they were washed by gently moving the crystal through the solution. These crystals were then transferred to a loop-holding cassette in liquid nitrogen.

Solutions for heavy-atom soaks were based on reservoir conditions permitting crystal growth, with precipitant increased 2-4%, and 0.5M potassium iodide or potassium iodide added. Crystals were transferred to small drops of these solutions and soaked for 60-90 seconds. To protect these crystals during freezing

cryoprotectants were produced as described above, but with the inclusion of 0.5M potassium iodide or bromide.

2.16.6 Data collection

All X-ray diffraction experiments were performed at Diamond Light Source (Oxfordshire). Native and Bromide peak data were collected on beamline DLS-i02, Iodide and Selenium peak data sets were collected on beamline DLS-i04, and data collections from *Sf9*-derived protein crystals were performed on beamline DLS-i03.

Native datasets were collected at λ=0.9795Å, bromide-soaked crystals at λ= 0.9150Å, iodide peak collection at λ= 2.0000Å and selenium peak at λ=0.9797Å.

3-5 test images at a maximum resolution of 3.0-3.2Å were collected with Ω=0.5° and separated by 45° rotation to allow assessment of likely diffraction resolutions and development of data collection strategies using EDNA[124].

2.16.7 Data processing

Data were processed using the xia2 pipeline with XDS and pointless. XDS is a programme suite that will index, integrate, and scale data collected from CCD and similar devices. Firstly the program identifies background pixel intensities, and locates diffraction spots. Secondly it indexes reflections to give a spacegroup for the unit cell, and evaluates the intensity of all spots on the frame. Finally XDS scales the reflection intensities, removes outliers and generates reflection statistics. Pointless assigns likely Laue groups consistent with the crystal class determined by XDS. This is performed by assessing the unmerged data from XDS to look for potentially symmetry-related reflections and systemic absences. Pointless assess all possible Laue groups within the cell parameter constraints and scoring them on the basis of identifiable potential systemic absences.

The process of assigning unit cell parameters, crystal orientation and space group is based on predicting where in reciprocal space Bragg's law is satisfied, and assigning Miller indices (hkl) to the reflections. These parameters are refined to give the best estimate of expected *hkl* positions. By recording the intensity of reflections it is possible to assign intensities to each spot (I_{hkl}). Once the unit cell is determined it is possible to predict where reflections are expected to appear, and by identifying reflections that are not recorded (systemic absences) it is possible to assign a Laue group to the crystal.

2.16.8 Obtaining phases

The electron density within a crystal can be thought of as a complex wave. As such it can be recreated using a fourier transform, which is the sum of a serious of simple waves. Equation 1 shows how electron density at a particular position or volume of the crystal (p(x,y,z)) can be reproduced.

$$\mathrm{p}(x,y,z) = \frac{1}{V}\sum_h \sum_k \sum_l |F_{hkl}| e^{(-2\pi i(hx+ky+lz-\alpha\prime_{hkl})}$$

Equation 1 Fourier sum to generate electron density

Where V is the volume of the unit cell, $|F_{hkl}|$ is the inverse of the recorded intensities for spots at position hkl. Here we see how recorded spot positions and intensities *hkl* ($|F_{hkl}|$) contribute to the final electron density map. The term α'_{hkl} denotes the phase of the spot *hkl*, however this information is not recorded during X-ray diffraction experiments.

In order to reproduce an electron density map we need to calculate phases for each recorded reflection. This can be done in two main ways- through molecular replacement and experimental methods. Molecular replacement uses proteins believed to be structurally similar to the target protein and tries to “find” the molecule in the experimental unit cell by a series of rotation and translation functions. This is performed by generating a Patterson map (a map of the inter-atomic vectors, which does not require phase information) of the known and unknown proteins and rotating one relative to the other until they align. Once the search model (known structure) has been placed in the unit cell it is possible to calculate the phases of the experimental diffraction to recreate the electron density of the protein of interest. Phase information contributes greatly to the derived electron density map, and so molecular replacement can introduce significant bias into the resulting structure (Figure 2.16.2).

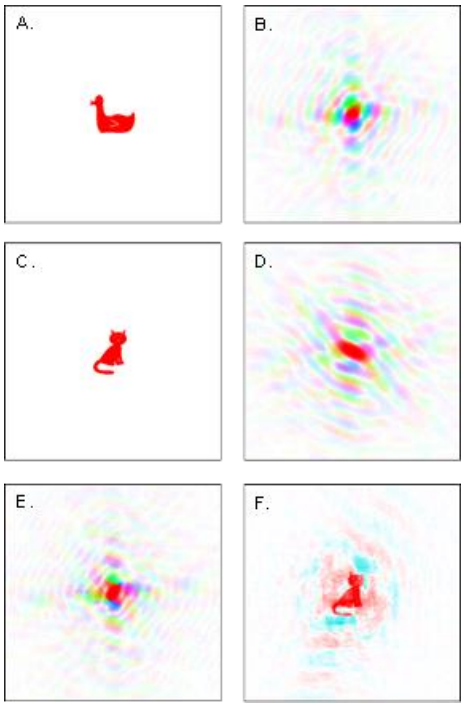


Figure 2.16.2 Demonstration of structural bias introduced by phases

The structure of a duck (A.), representing the structure of an unknown protein, and cat (C.), representing a known structure to be used as a search model. B. and C. show the fourier transform of each structure. E. shows the application of the phases from the cat transform (D.) to the amplitudes of the duck transform (B.). Transforming E. leads to the generation of a cat-like structure (F.). Taken from K. Cowtan (http://www.ysbl.york.ac.uk/~cowtan/fourier/magic.html)

Alternatively it is possible to approximate phases using experimental techniques. To do so heavy atoms (HA) are introduced into the crystal (either at the point of protein synthesis or by soaking existing crystals in a HA-containing solution). These HA-derived crystals can be used in two ways- either the effect they have on the intensity of spot intensity (isomorphous replacement), or by utilising the ability of these atoms to absorb X-rays of a particular wavelength (anomalous dispersion), which was used in this project.

To solve the structure of NRC1 NBARC domain phases were experimentally determined by single wavelength anomalous dispersion (SAD). This method relies on the ability for heavy atoms to absorb X-ray radiation at specific wavelengths, which are then reemitted with slight shifts in phase and amplitude. Under these conditions Friedl's law (stating that the intensity of two symmetry-related reflections is equal, $I_{hkl} = I_{-h-k-l}$) is no longer valid. The anomalous differences between Friedl pairs can then be used to locate the anomalous scatterers in the unit cell

(substructure solution) and calculate phases. These phases can then be "bootstrapped" to all atoms in the atoms give an initial approximation of the electron density.

For NRC1 NBARC, SAD was performed with SeMet, bromide and iodide-derived data. Structure solution was performed by PHENIX AutoSol, which uses Hybrid substructure search (Hyss) to locate anomalous scatterers. Although anomalous signal was detected for all HA derivatives, only SeMet data produced a valid solution.

To investigate the effect of growing crystals at a different pH, or from an alternative protein source (insect cell expression) molecular replacement was performed using the most complete structure of NRC1 with ADP and some regions of the model removed. As molecular replacement can introduce bias into the resulting electron density, removal of these features allows us to determine whether the resulting solution is valid. Non-valid solutions generally result in electron density identical to the existing model, whereas valid solutions usually result in additional features, such as electron density where we would expect ADP to be present- as was the case for MR into these alternative crystals.

2.17 ATPase assays

2.17.1 Linked enzymatic assay

The NBARC domain of NRC1 was screened for ATPase using either a modified Malachite green assay[125] or a linked enzymatic reaction adapted from Nørby[125].

The linked assay uses the ability of pyruvate kinase (PK) to dephosphorylate phosphoenolpyruvic acid (PEP) to pyruvate using ADP as a phosphate-acceptor. The pyruvate is then converted to lactate by lactate dehydrogenase (LDH) in a reaction which oxidises the cofactor NADH to NAD^+. Conversion of NADH to NAD^+ results in a decrease in absorption at 340nm. ATPase activity of a protein can therefore be deduced by looking at the rate of decrease in absorption at 340nm as ATP is converted to ADP.

Reactions were carried out at 25°C in flat-bottomed 96-well plates using an Ultrospec 3300 pro (GE healthcare Life Sciences) to record the absorption at 340 nm every minute for 30 minutes. Initial readings at 340nm of the primed reactions (reaction buffer, PK, LDH and NRC1 NBARC) were taken every minute for 10 minutes prior to the addition of ATP to establish a baseline absorbance

Primed reactions were made by adding 1.5µL PK/LDH solution (Sigma Aldrich), 800µM PEP, 400µM NADH and 1µM NRC1 NBARC. Reactions were then initiated by the addition of 1mM ATP.

2.17.2 Malachite green

Malachite green assays to detect release of free phosphate were performed as described in Lanzetta[125] with modifications. ATPase assays were performed in 96-well clear, flat bottomed plates. 5µL of 100µM protein in A4 buffer (20mM HEPES pH 7.5, 150mM NaCl) was added to 5µL of 2x reaction buffer (20mM HEPES, 4mM $MgCl_2$, 2mM ATP) and incubated for 1 hour at room temperature. Reactions were quenched by the addition of 90µL malachite green reagent (1M HCl, 8.5mM ammonium molybdate, 1.2mM malachite green, 0.1% Triton N101) and incubated for 15 minutes. Absorbance at 630nm was recorded using an Ultrospec 3300 pro (GE healthcare Life Sciences) plate reader.

A range of reaction buffers were used in initial screens to take into account various potential requirements for ATPase activity of NRC1 NBARC based on previous investigations into the ATPase activity of I-2, Apaf-1 and a number of other STAND proteins. Briefly the buffers varied in pH (6.5, 7.0, 7.5), buffer component (50mM Tris-HCl or 20mM HEPES), mono- and di-valent cation availibilty (5mM Mg^{2+} invariable as either $MgCl_2$ or $MgSO_4$) with or without 5mM $MnCl_2$, either 150 or 50mM NaCl or 50mM KCl), the addition of divalent anions (50mM $(NH_4)_2SO_4$), the addition of 10% glycerol, and the reducing agent dithiothreitol (DTT, 1.5mM).

None of the screened buffers significantly affected ATPase activity of WT NRC1 NBARC, and so ATPase assays for NRC1 NBARC mutants were performed in 20mM HEPES pH 7.5, 75mM NaCl, 5mM $MgCl_2$ and 2mM ATP.

2.18 DNA binding assays

2.18.1 SPR

DNA binding capabilities of purified NRC1 NBARC domain were assessed via surface plasmon resonance (SPR) using the reusable DNA capture technology (ReDCaT, Stevenson *et al*). Binding assays were performed by Clare Stevenson using a Biacore T200 system (GE Healthcare). Assays were performed at 20°C, with 50nm, 100nm and 500nm NRC1 NBARC tested against 20bp, 30bp and 40bp double-stranded DNA oligonucelotides (oligonucleotide sequences found in Appendix).

2.18.2 Fluorescence Anisotropy

Fluorescence anisotropy was performed using single-stranded 5′-fluorescein-labelled oligonucleotide (Appendix). Anisotropy was measured using a Synergy™ H4 plate reader (BioTek®). Reactions were set up with 150nM oligonucleotide and performed at room temperature. After the addition of protein into DNA-containing buffer (20mM TRIS pH 7.5, 100mM NaCl), samples were incubated for 20 minutes to allow signal to stabilise before recording polarized fluorescent signals.

3

Strategies for obtaining soluble CNL R-proteins and domains

3.1 Introduction- Biochemical investigations into R-protein function

In the two decades since the first R-protein was cloned, our understanding of their function and behaviour at the cellular and genetic level has grown dramatically. We now have numerous examples of R-genes and their specificities- from the highly specific R3a/AVR3a interaction, disrupted by changes in only two amino acids of the effector[126], to the ability of Mi-1 to detect both nematode and aphid effectors[127]. Models for effector recognition from direct interactions[128], to the guard hypothesis[57], to the more complex decoy hypothesis[70], have all been proposed on the basis of genetic analysis and cell-biology experiments. By separating R-proteins into their constituent domains and expressing these *in planta* it is possible to form hypotheses about the regulation of R-proteins post-translation.

Our understanding of the biochemistry of these proteins, an important tool for validation of many of these hypotheses, has lagged significantly behind our understanding at the genetic and cellular level. This lack of information is a large obstacle to the rational improvement of R-protein function. Models explaining R-protein regulation are qualitatively useful in making broad predictions about responses to changes in R-protein coding sequence, and can be used to rationalise a number of mutant phenotypes. However, these models lack the details required to make *de novo* predictions of R-protein behaviour in response to many specific mutations. Multiple studies have indicated that individual R-protein domains associate[89] and it is believed to be through these intramolecular interactions that R-proteins are kept in a signalling competent, but inactive, state[129]. How strong these interactions are, and how variable these interactions may be has not been demonstrated, and would be difficult to address simply through *in planta* investigations.

The strength of these intramolecular interactions has implications on the likely success of manipulating R-proteins for crop improvement- if the interactions between the domains are weak then it can be expected that the proteins are capable of spontaneously switching to an active conformation and back without detrimental effects on the plant. In such a system the R-protein would be more tolerant to mutations that affect these interactions, for example mutations in the LRR that allow recognition of additional AVR proteins but also perturb the inhibitory interactions between the LRR and NBARC domain. If, however, these interactions

are highly stable then it may be expected that only mutations which do not affect these interaction interfaces can be tolerated without additional mutations to affect other aspects of the system. Work with the CNL Rx highlights this issue- mutations introduced to expand the responsiveness of Rx lead to a creeping autoactive (cell death) phenotype, rendering the increased responsiveness of little use in the context of crop improvement. To overcome this the researchers performed a second round of random mutagenesis to identify additional changes that would allow expanded recognition without compromising regulation[9].

The primary cause of the disparity between the available data from phenotypic screens and biochemical data has been the difficulties of obtaining sufficient yields of well-expressed protein in a form that can be used in sensitive *in vitro* assays. Beyond the expression of MLA10 and Rx coiled-coil domains in *E. coli*, remaining studies have either relied on truncated domains[103], refolded protein[102], or very low levels of protein expressed in eukaryotic systems[86, 130].

For example, primary work investigating the function of predicted Walker-A and Walker-B motifs seen in the NBARC domain of R-proteins, which are also found in proteins that bind and hydrolyse ATP or GTP, relied on refolded protein purified from inclusion bodies[102]. The data indicated that although these proteins are able to facilitate conversion of ATP to ADP, this reaction is slow by enzymatic standards. As the authors note, multiple bands corresponding to either degradation products or premature termination of protein translation are present in these assay, which would make accurate protein quantification and thus rate measurements more difficult to accurately calculate. Due to the difficulties of distinguishing potential artefacts associated with refolding, we focused on identifying solubly expressed protein for our investigations.

When examples of soluble R-protein domains are identified, it can lead to novel discoveries- Fenyk *et al* investigated the biochemical function of the nucleotide binding (NB) domain of R-proteins when expressed solubly in *E. coli*. Using these materials, the researchers were able to demonstrate that these domains displayed phosphatase abilities- hydrolysing ATP not only to ADP as would be expected, but also producing AMP, something which would not be anticipated by general models based on *in vivo* data[103].

Two independent studies by Maekawa *et al*[86] and Williams *et al*[130] were recently able to demonstrate the expression of full-length MLA27 from barley in insect cells, and the resistance protein M from flax in yeast respectively. Despite the low levels

of expression both groups were able to quantify the ATP and ADP occupancy of the purified proteins, giving the first direct demonstration of the links between activation state and ligand binding. The expression of these two full-length proteins represents a huge step forward for biochemists interested in characterising intracellular plant immune receptors, however further investigations are hampered by the low yields of protein.

Traditional approaches to screen for soluble protein typically use a suite of bioinformatics programmes to look for predicted structural domains, disordered regions and secondary structure elements to try and rationalise areas of the protein that are most likely to produce well behaving constructs. It is possible to use both de novo and comparative methods to characterised proteins to aid construct design. Protein secondary structures are highly influenced by primary sequence, and so α-helices, β-strands and intrinsically disordered regions can be predicted using only primary sequence. A number of databases can be used to identify probable domain boundaries- PFAM[131] for example is a useful tool to identify common functional domains, which can be combined with secondary structure analysis to allow informed construct design. Similarly, homology modelling can be used to identify regions that are predicted to form distinct structural features and improve the chances of producing a biologically representative construct. This approach can be described as low- to medium throughput as screens are typically limited by the infrastructure available and information on which to base construct boundaries.

As obtaining soluble R-protein has proven non-trivial in the past, we took multiple approaches to getting well-folded, solubly expressed R-protein. Initially our work focussed on utilising modern techniques that may have been unavailable to previous groups interested in the biochemistry of R-proteins to identify soluble constructs for expression in *E. coli*. From there we moved on to explore traditional construct design to obtain soluble protein from *E. coli*, before expanding our expression systems to include *Sf9* expression (insect cells), and *N. benthamiana*. This chapter will discuss the results of these screening experiments, briefly outlining the basis of each method. Full methodologies can be found in Chapter 2.

3.2 High-throughput screening of a high-profile R-protein

3.2.1 Rationale and theory

We initially focussed on intensively screening a single R-protein that has been well studied historically, and continues to be the subject of active research. We used

these criteria so that any structural or biochemical data produced would be supported by previous work, and that the tools to further probe the biological relevance of any biochemical data would be well-developed. For these reasons we decided to focus on *S. tuberosum* R3a, as its function with regards to recognition of *P. infestans* AVR3a alleles has been well characterised[126, 132-134].

We hypothesised that as high-impact targets, multiple attempts to produce soluble R-proteins will already have been made using traditional methods. We took the lack of published data using soluble, heterologously expressed proteins as evidence that these approaches were insufficient, and therefore decided to collaborate with a group specialising in screening proteins for soluble expression.

Darren Hart's group at the EMBL in Grenoble have established a system allowing the high-throughput screening of a library of truncations of a gene of interest in *E. coli* using an N-terminal histidine tag and a C-terminal biotin-acceptor peptide (BAP) (Figure 3.2.1 A)[135]. The N-terminal histidine tag can be used for protein purification using standard nickel IMAC columns, and also to help assess the stability of each protein. Biotinylation of the BAP can be used to predict the solubility and stability of proteins expressed in *E. coli*; post-translational biotinylation of proteins is dependent on the ability of the modifying enzyme to access the target protein or substrate, in this case the BAP. If the protein expressed is misfolded or insoluble it will form inclusion bodies, rendering the BAP inaccessible for post-translational modification. The misfolded protein will therefore undergo little or no biotinylation, and the use of fluorophore-conjugated streptavidin can be used to measure relative biotinylation efficiency (Figure 3.2.1 B). By using time-course digest of a gene of interest, and ligating these fragments into the dual-tag vector, this system selects areas of sequence space that would not normally be sampled through traditional approaches. Colony-picking robots, growing bacteria directly onto nitrocellulose membranes and a dual-tagging system, with different excitation and emission spectra, make it possible to screen thousands of constructs in a short amount of time.

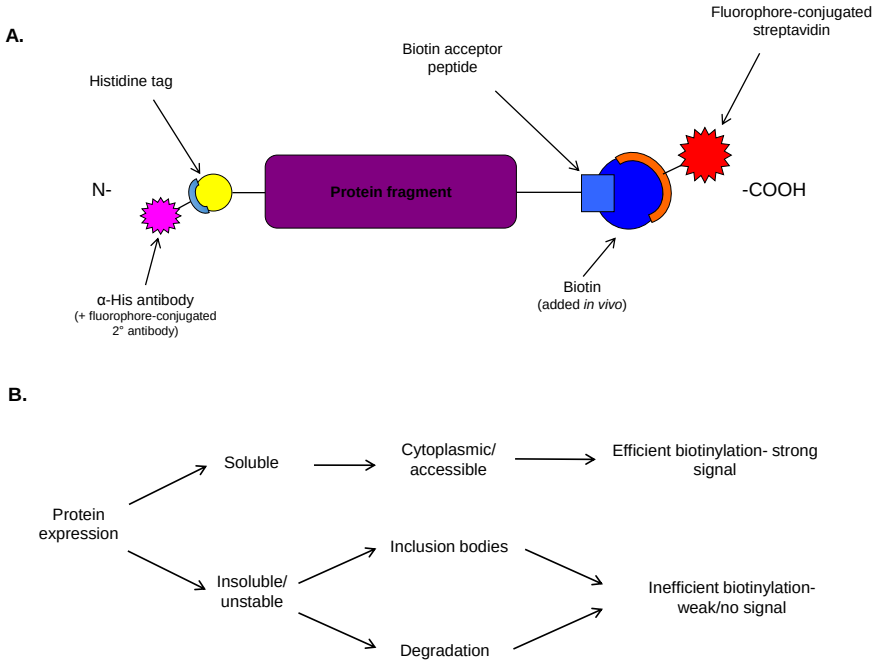


Figure 3.2.1 Schematic demonstrating the use of dual-tagged pESPRIT vector for solubility screening

(A.) Cartoon representation of the results of protein expression using the dual-tagged pESPRIT vector. Ligation of time-course digested gene into the vector results in an N-terminal histidine tag fusion and a C-terminal biotin-acceptor peptide (BAP). Biotin is added to the BAP as an endogenous post-translational modification in *E. coli*. After cell lysis this fragment can be probed for the presence of the histidine tag using specific antibodies and for biotinylation with fluorophore-conjugated streptavidin.

(B.) Representation of the fate of proteins expressed in *E. coli* and how this is used to probe for soluble expression. Soluble, stable proteins are efficiently biotinylated, resulting in a strong signal when probed with fluorophore-conjugated streptavidin. Unstable or insoluble proteins are either degraded or form inclusion bodies, which are less accessible for biotinylation. Measuring both the anti-histidine and anti-biotin signals allows determination of the fate of fusion proteins in *E. coli*.

3.3 Results and Discussion

Creation of the library

Note: Time-course digestion and ligation into the bacterial expression vector was performed by Sri Puranik prior to my arrival in Grenoble. All high-throughput screening (including both fluorometric screens and western blotting) was performed at EMBL in Grenoble with D. Hart's group. Medium-throughput confirmatory screens and downstream expression and processing was performed at JIC. Initial expression screens were performed by myself and Sri, with construct ranking and western blotting performed by myself and Philippe Mas.

In this project we used R3a coding sequence that had been codon optimised for expression in *E. coli*, and had internal NsiI and AscI sites silenced by synonymous mutation synthesised by Shinegene. To generate a non-biased library of R3a constructs we used time-course digestion of the gene of interest using DNA exonuclease III. We produced two libraries separated at the approximate boundary between the NBARC and LRR domains, corresponding roughly to the N- and C-terminal halves of the protein, and used unidirectional digests to generate the gene truncations: 5'->3' digestions for the library corresponding to the C-terminal half of the protein, and 3'->5' digests for the library corresponding to the N-terminal half of the protein. As these unidirectional digests would create libraries with constant 3' or 5' ends that could potentially affect the solubility and expression of the genes in *E. coli*, we produced three separate pESPRIT vectors with different 3' or 5' ends for each half of the gene, giving a total of six pESPRIT-R3a vectors (Figure 3.3.1).

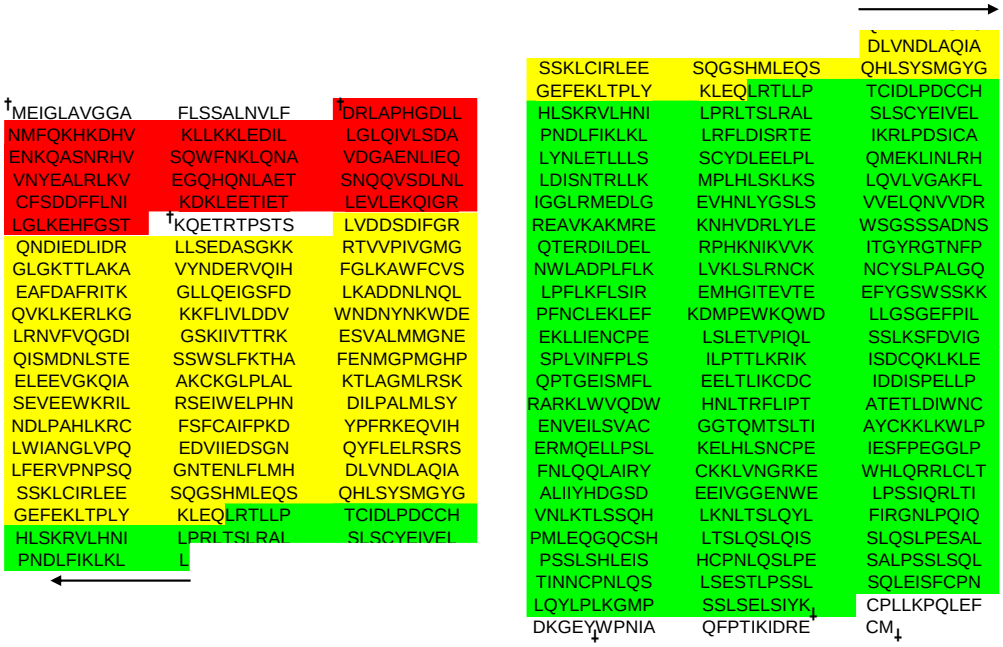


Figure 3.3.1 Translated sequences of N-terminal (left) and C-terminal (right) halves of R3a used for library creation

The R3a ORF was translated and divided into an N-terminal half (left) and C-terminal half (right). The predicted coiled-coil domain is highlighted in red, the predicted NBARC domain in yellow and putative LRR domain in green. Each half of the protein was subdivided to produce three sub-libraries, which would give fixed termini during time-course digestion. Fixed termini are denoted by †, and the direction of digestion indicated by arrows.

In order to produce unidirectional digestion libraries, N-terminal encoding and C-terminal encoding vectors were cut by restriction with AatII and AscI, and NsiI and NotI respectively to produce exonucleaseIII susceptible ends. The three N-terminal encoding and the three C-terminal encoding linearised vectors were pooled for time-course digestion. Time course digestion was performed by the addition of ExoIII to the pooled vectors, with 1µL aliquots removed each minute and quenched in 2M NaCl. Mung bean exonuclease (MBE) treatment was used to produce blunt ends for vector ligatation. The MBE treated pooled vectors were then run on a 0.5% agarose gel to produce two library smears (N- and C-terminal libraries) and subdivided into large, medium and small sub-libraries- this step is important as it allows the definition of minimum vector size to promote the selection of biologically relevant fragments, and to prevent small fragment dominance for bacterial transformation (Figure 3.3.2). The sub-library smears were excised and purified before being treated with T4 DNA ligase to circularise the vectors.

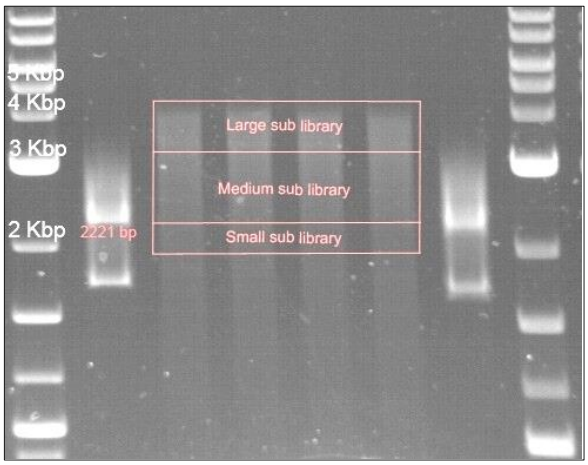


Figure 3.3.2 DNA smear resulting from time-course digestion of the N-terminal R3a libraries

Time-course digestion of the N-terminal pESPRIT-R3a vectors visualised as a smear on a 0.5% agarose gel. The three vectors corresponding to the N-terminal half of R3a with unique fixed termini were pooled and digested by exonuclease III, with aliquots removed and quenched every minute. The smear was then subdivided into large, medium and small sub-libraries (to prevent smaller fragments dominating downstream processing) and these sub-libraries excised and DNA extracted for bacterial transformations.

The circularised N-terminal and C-terminal libraries were transformed into BL21 RILAI ($BL21^{AI}$) strain *E. coli*, plated on large LB-Agar plates, supplemented with kanamycin and chloramphenicol, and left to grow overnight at 37°C to produce colonies. Colonies were picked robotically, placed in 384-well plates containing TB-HMFM media (terrific broth with hogness modified freezing media- Section 2.4.4),

and allowed to grow to saturation overnight before storage at -80°C. We selected 11,000 colonies for the N-terminal library and 12,000 colonies from the C-terminal library, oversampling to compensate for the introduction of internal stop codons and frame-shifts, producing a total of 60 384-well reference plates stored -80°C.

3.3.1 Library screening

To screen the library, reference plates were thawed at room temperature and used to inoculate fresh 384-well plates containing TB. These plates were grown overnight at 37°C with constant shaking, and overnight cultures were then stamped in a grid layout in triplicate onto nitrocellulose membranes, which were placed onto LB agar plates. These plates were incubated at 25°C overnight to allow bacterial growth before the membranes were transferred to LB agar plates supplemented with both 0.2% arabinose, to induce gene expression, and biotin to allow biotinylation of the recombinant protein. Induction was performed for 4-5 hours at 30°C.

Cells were lysed by subjecting the membranes to alkaline buffer. This allows expressed proteins to bind directly to the nitrocellulose membrane at the point where the colony was stamped, and thus allows results of expression screening to be linked directly back to the appropriate well from the reference culture plates. The membranes were washed to remove insoluble and non-proteinaceous cell lysate, and probed to detect the presence of the N- and C-terminal tags. Mouse α-histidine antibodies were added to each membrane, and incubated for one hour with constant rotation. The membranes were then washed before the addition of both the secondary antibody (alexa 532-conjugated α-mouse), and the fluorophore-conjugated streptavidin.

The stained membranes were then scanned using a Typhoon scanner to detect fluorescence, and the intensity of the signals recorded (Figure 3.3.3).

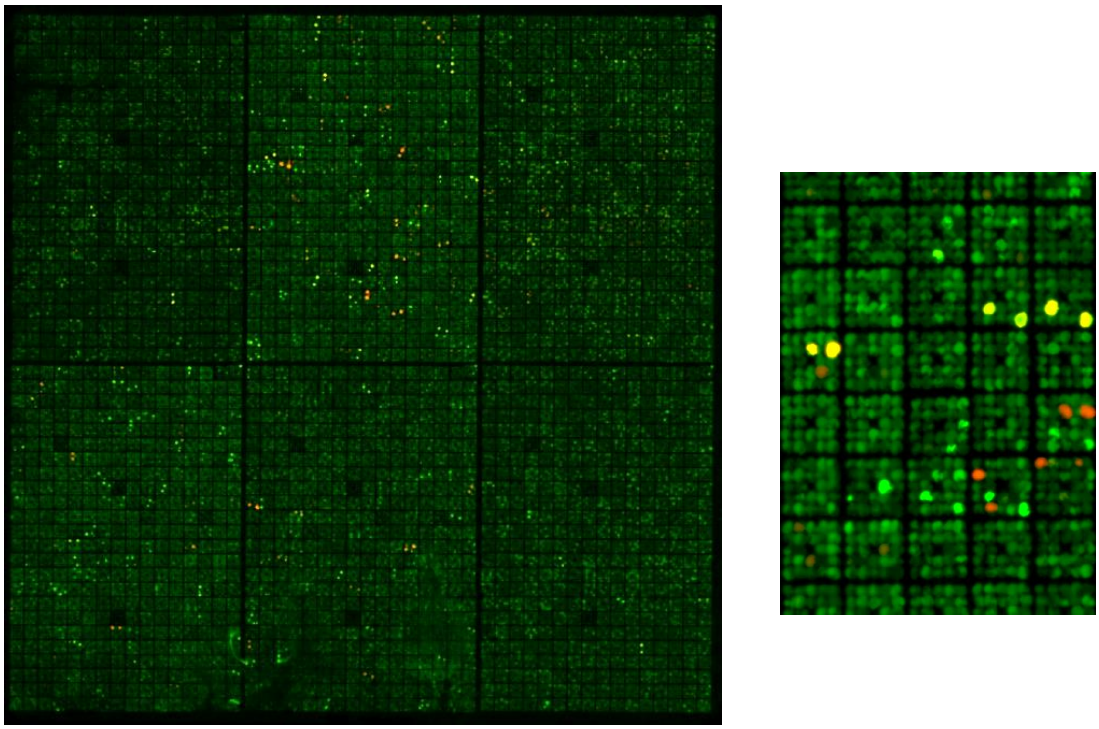


Figure 3.3.3 In situ western blot showing anti-histidine and streptavidin signal of bacterial colonies expressing recombinant R3a fragments

Left: Bacteria gridded onto six nitrocellulose membranes and incubated to form colonies, which were then induced with IPTG. Lysing the colonies directly on a nitrocellulose membrane in alkaline solution allows the proteins expressed in each colony to bind in place, and so single spots can be attributed to individual wells from the 384-well reference plates. Once proteins are bound to the membrane, anti-histidine antibodies and streptavidin can be applied, and by using fluorescent probes it is possible to identify and quantify colonies expressing one, or both tags. Right: close up of a section of nitrocellulose membrane. Green indicates the presence of biotinylated proteins, red shows histidine-tagged proteins, co-localisation appears as yellow.

Intensities of both the anti-histidine and streptavidin signals were averaged over the three replicates and the colonies were ranked by intensity of the histidine signal. The top-ranking 48 N-terminal library constructs and the top 47 C-terminal constructs were selected for further testing, with maltose-binding protein (MBP) included as a positive control. These colonies were picked from the reference plates and used to inoculate a 96-deep-well plate containing TB. This plate was incubated overnight at 37°C before 10% glycerol added and stored at -80°C for downstream screening.

To test for solubility during expression in liquid culture and identify false-positives, the 96 constructs were grown at small scale (4mL cultures). Expression was induced by the addition of arabinose, and cultures incubated at 25°C overnight. Cultures were pelleted and sphaeroplasts prepared as described in Section 2.4.5.

Sphaeroplasts were lysed by freeze-thaw, and the addition of a lysis buffer containing 0.5% Brij. Solubility was tested by low-stringency nickel purification of soluble lysate, clarified by centrifugation, followed by western blotting (Figure 3.3.4).

Left: α-histidine signal **Right**: α-biotin signal

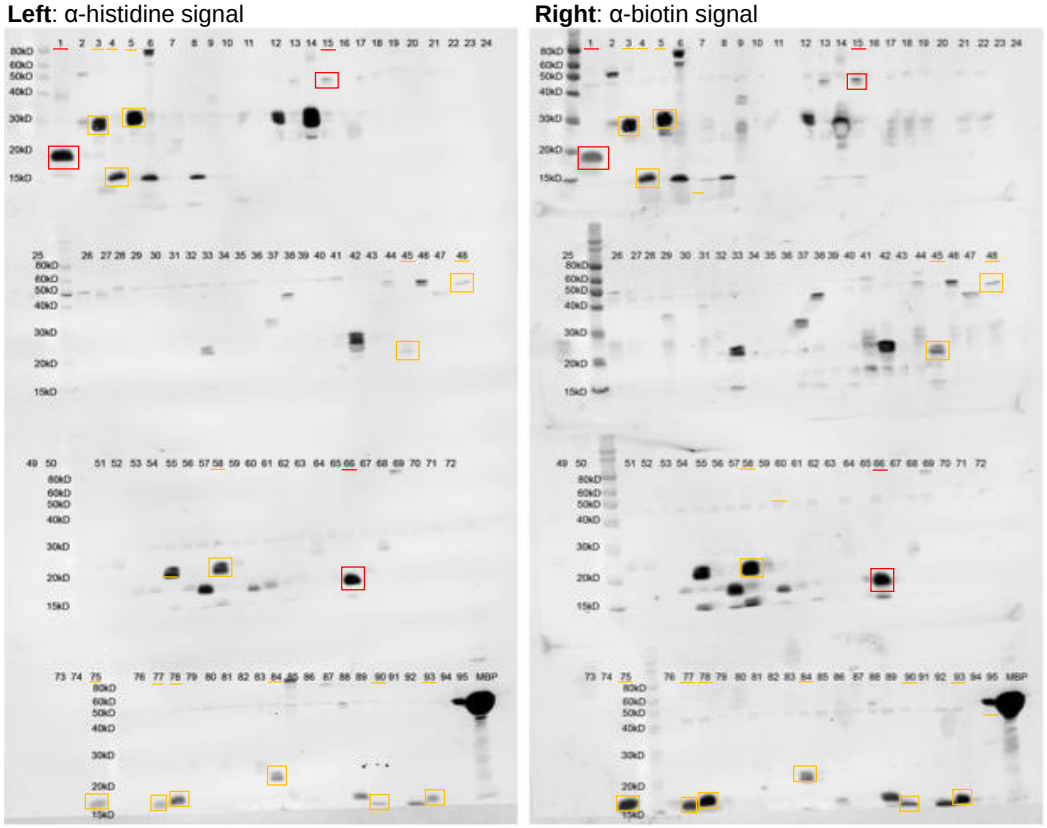


Figure 3.3.4 Western blot analysis of small-scale growth and expression of the top 95 colonies identified in the fluorometric screen

The highest ranking 48 colonies for the N-terminal half and the highest ranking 47 colonies for the C-terminal half of R3a were selected for small-scale protein expression in liquid media. After expression was induced, lysate underwent low-stringency nickel IMAC and separation by SDS-PAGE. Probing with α-histidine antibodies (left) and fluorophore-conjugated streptavidin (right) reveals a large number of colonies selected from the fluorometric screen that do not appear to produce histidine-tagged, or biotinylated proteins (such as lanes 70-74). Constructs numbered 1-48 are the ranked-values of colonies for the N-terminal R3a library, colonies 49-95 are the top-ranking colonies 47 colonies on the basis of fluorescent signal (where colony 49 represents the highest scoring colony and 95 represents the 47th highest-scoring colony). Lanes highlighted in yellow indicate constructs capable of soluble expression at small-scale tests. Lanes highlighted in red indicate constructs taken through for large-scale expression and tag removal.

As can be seen by the number of blank lanes for either the α-histidine signal or the α-BAP signal, the fluorometric screening was susceptible to false-positives, reducing the number of constructs available for downstream screening.

Positive colonies identified in the western-blotting screen were tested for reproducibility in small-scale screens, and soluble yield in the absence of IMAC purification. These colonies were grown in 24-well plates with 4mL LB overnight at 37°C with constant shaking. Overnight cultures were used to inoculate fresh media and allowed to grow for ~4 hours to reach an approximate OD_{600} of 0.6. Culture densities varied widely, typically from 0.3 to 0.9, possibly due to the effects of leaky expression of different constructs. Expression was induced by the addition of 0.2% arabinose and the cultures incubated at either 37°C for a further 4 hours or at 18°C overnight. Cells were pelleted by centrifugation and lysed in 0.4-0.8mL lysis buffer (500mM NaCl, 20mM HEPES pH 7.5, 0.5mM PMSF) by sonication. Insoluble lysates were removed by centrifugation and clarified lysate loaded onto 17% SDS-PAGE gels to check expression by coomassie staining (Figure 3.3.5).

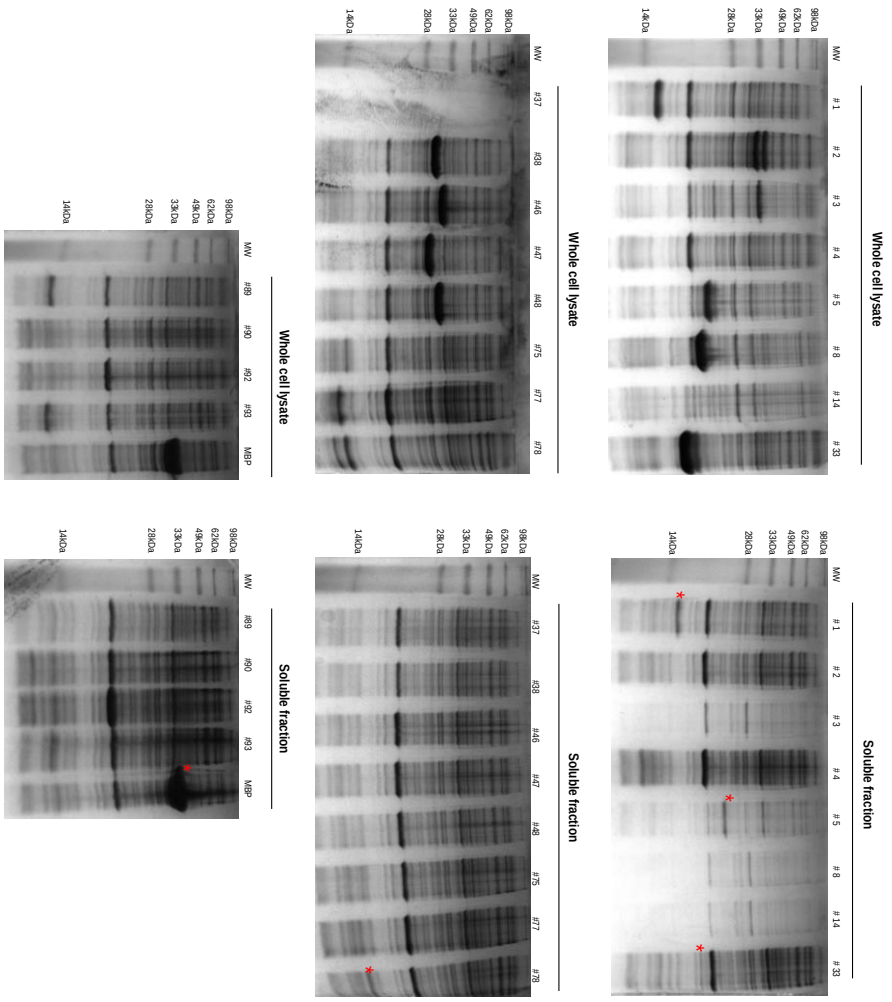


Figure 3.3.5 SDS-PAGE and coomassie staining of small-scale expression screens to assess solubility of positive candidates from low-stringency IMAC and western blotting

Colonies were grown in 24-well blocks at 37°C and induced by the addition of 0.2% arabinose once average OD_{600} reached 0.6. Whole-cell lysates show accumulation of unique bands for most constructs tested, however comparison with total soluble protein shows very little enrichment for the majority of constructs. Bands corresponding to solubly expressing recombinant proteins are indicated by *. Lane numbers refer to colony ranked values from fluorometric screening. Performing induction at 18°C (not shown) resulted in improved solubility for very few constructs.

We were able to identify few well-expressing constructs under these conditions, and so performed screens to optimise expression conditions. Constructs were transformed into SoluBL21, BL21* and C41 strains of *E. coli* to look for strain-dependent expression effects. SoluBL21 and BL21* were both chosen due to their purported efficacy for producing soluble protein, with the inclusion of C41 cells as these cells are able to express many toxic proteins.

Alongside screening different bacterial strains we investigated the effect of different additives on protein solubility. Strain-by-additive screens were performed by growing individual constructs in 24-well plates in LB overnight, and using these cultures to inoculate fresh 24-well plates, typically screening two constructs in three expression strains with wells set up in quadruplicate. When cultures had reached OD_{600} of 0.6 +/- 0.3 they were rapidly chilled on ice and subjected to the appropriate additive (0.4% ethanol, 0.5% DMSO, 10% glycerol, or 0.5mL LB) immediately prior to induction with 1mM isopropyl β-D-1-thiogalactopyranoside (IPTG), or 0.2% arabinose for $BL21^{AI}$ strain. A typical 24-well block layout is described in Figure 3.3.6

	1	**2**	**3**		**4**	**5**	**6**
A	LB	LB	LB		LB	LB	LB
B	DMSO	DMSO	DMSO		DMSO	DMSO	DMSO
C	Ethanol	Ethanol	Ethanol		Ethanol	Ethanol	Ethanol
D	Glycerol	Glycerol	Glycerol		Glycerol	Glycerol	Glycerol
	BL21	SoluBL21	BL21*		BL21	SoluBL21	BL21*
Construct 1					**Construct 2**		

Figure 3.3.6 Typical layout of a 24-well plate used to optimize soluble expression of recombinant R3a fragments

To allow medium-throughput screening of strain and additive conditions that may boost soluble expression of constructs, bacterial growth and expression was performed in 24-well blocks. These blocks were typically divided to allow testing of two constructs, each in three expression strains- for example columns 1, 2 and 3 may contain soluBL21, BL21* and C41 transformed with construct 1. These bacteria were grown overnight and used to inoculate fresh media in quadruplicate (lanes A, B, C and D) and allowed to grow to an OD_{600} of 0.6. Once the target OD was reached, blocks were chilled on ice and supplemented with either 0.5% DMSO, 0.4% ethanol, 10% glycerol or 0.5mL LB immediately prior to the addition of 1mM IPTG. After incubation for a further 16 hours at 18°C cells were pelleted by centrifugation in the plates and soluble expression assessed by SDS-PAGE.

Cells were left at 18°C with constant shaking for protein expression to continue for 16 hours. Plates were then processed as described in Section 2.6. Using this method it was possible to identify expression conditions that appeared to give a significant boost to soluble expression of certain constructs (Figure 3.3.7). The conditions allowing soluble expression of R3a constructs are summarised in Table 3.3.1.

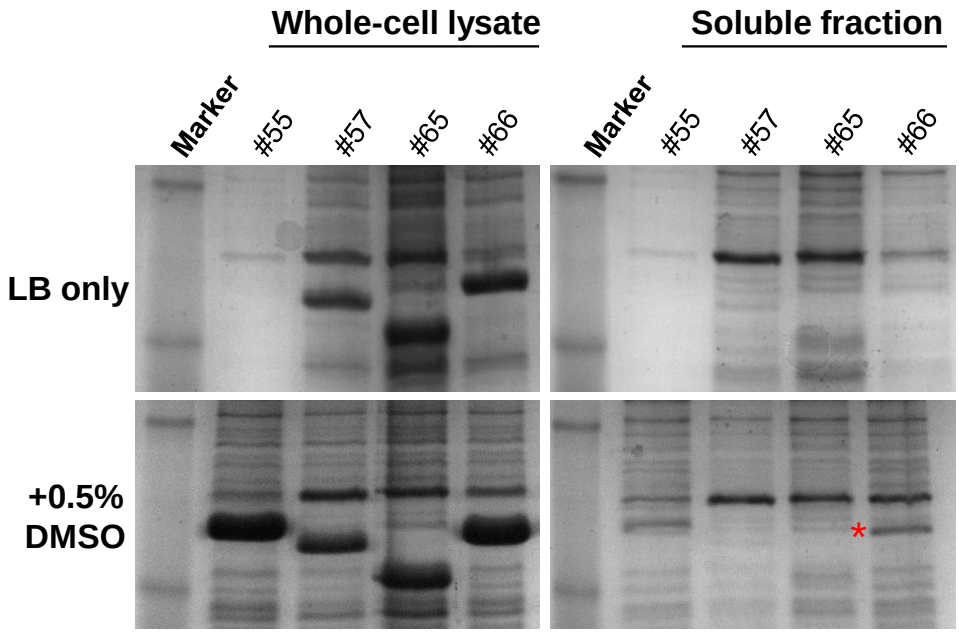


Figure 3.3.7 Coomassie stained SDS-PAGE gel showing the boost in soluble expression by the addition of 0.5% DMSO

Additive screens were able to identify specific conditions that could lead to dramatic increases in soluble protein expression. In this example supplementing the growth media of construct #66 in $BL21^{AI}$ with 0.5% DMSO lead to a much stronger band in the total soluble protein extract (bottom left panel) than when expression was performed in LB alone (Top panels). The band predicted to correspond to soluble construct #66 is highlighted by *.

These constructs, named as their ranked values from the fluorometric solubility screens, were sent for sequencing to prioritise those most likely to yield biologically interesting protein. The regions of R3a covered by these constructs are shown schematically in Figure 3.3.8. Purified plasmid for constructs #1, #15 and #55 were treated with BspE1 to cleave the BAP and re-ligated to give R3a fragments fused to a single, N-terminal, histidine tag that could be removed post-translationally by TEV protease. As the BAP is a small tag predicted to have little effect on the solubility of the fusion proteins, these constructs were then used for large-scale protein purification after sequencing confirmed BAP removal and proper re-ligation.

Construct	Strain	Media	Additive
#1$^{(21\text{-}120)}$	BL21^{AI}	LB	-
	soluBL21	LB	DMSO
#3$^{(161\text{-}589)}$	SoluBL21	LB	DMSO
#4$^{(161\text{-}228)}$	BL21^{AI}	LB	-
#5$^{(21\text{-}224)}$	soluBL21	LB	DMSO
	BL21*	LB	-
#15$^{(161\text{-}410)}$	soluBL21	LB	DMSO
	soluBL21	AIM	-
#48 (Frame shift)	soluBL21	AIM	DMSO
	soluBL21	LB	DMSO
#55$^{(1143\text{-}1282)}$	BL21^{AI}	LB	DMSO
	BL21^{AI}	LB	Glycerol
#58$^{(1141\text{-}1282)}$	BL21*	LB	-
#66$^{(1180\text{-}1282)}$	BL21^{AI}	LB	DMSO
#75$^{(1181\text{-}1282)}$	BL21^{AI}	LB	-
#77$^{(1227\text{-}1282)}$	BL21^{AI}	LB	37°C
#78$^{(1150\text{-}1282)}$	BL21^{AI}	LB	37°C
#84$^{(1185\text{-}1282)}$	BL21^{AI}	LB	DMSO
	BL21^{AI}	LB	Glycerol
#90$^{(1186\text{-}1282)}$	BL21^{AI}	LB	37°C
#93$^{(21\text{-}343)}$	BL21^{AI}	LB	37°C

Table 3.3.1 Summary of conditions required for soluble expression of R3a constructs

Screening multiple strain and additive combinations allowed the soluble expression of several R3a fragments. Construct numbers relate to their rankings from fluorometric screening and numbers in parentheses denote the amino-acids of R3a these constructs span. All protein inductions for growth in LB media were performed by rapidly chilling cultures prior to addition of 0.2% arabinose (for BL21^{AI}) or 1mM IPTG. Protein expression was performed for 16 hours at 18°C as standard- constructs that expressed solubly only when grown at 37°C for 4 hours are highlighted in the "additive" column. DMSO indicates constructs that required the addition of 0.5% DMSO prior to induction for soluble expression. Those labelled glycerol required the addition of 10% glycerol, and those marked with "-" required no additive to produce soluble protein. Growth and expression in auto-induction media (AIM) was performed by inoculating fresh AIM with saturated overnight cultures to a target OD_{520} of 0.05, and expression performed for 16 hours at 18°C.

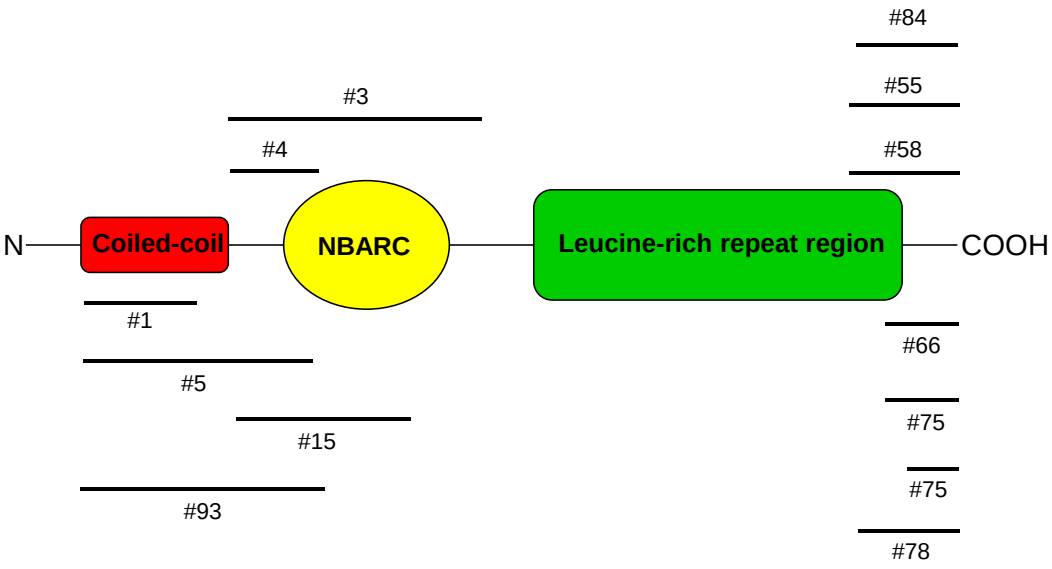


Figure 3.3.8 Cartoon representation of the regions covered by solubly-expressing R3a fragments

Constructs encoding R3a fragments that could be expressed solubly were sent for sequencing to prioritise those of most biological interest. The sections of R3a each of these construct spans is approximated by the bar length and positioning (for example construct #1 spans ~ ¾ of R3a's predicted coiled-coil domain, whereas construct #5 extends to include a small section of the NBARC domain). No soluble constructs starting from the native N-terminus were identified, however soluble constructs were found to begin at both the start of the predicted CC domain, and the end of this region. Only LRR constructs using the endogenous C-terminus were identified as expressing solubly, with no examples of soluble expression from either alternative C-terminal regions.

3.3.2 Large-scale expression of single, N-terminally tagged R3a fragments

Removal of the biotin-acceptor tag for the pESPRIT vector and re-ligation allows expression of individual constructs with only an N-terminal histidine tag that can be removed proteolytically. We removed the C-terminal BAP from several constructs that were identified as screening solubly and re-screened these for soluble expression. Three of the resulting constructs were identified as expressing solubly in *E. coli* and so were used for large-scale expression and purification.

3.3.2.1 R3a construct #1 (coiled-coil region)

Construct #1$^{\Delta BAP}$ (R3a$^{21\text{-}120}$) was expressed in 8xL of LB in soluBL21 with the addition of 0.5% DMSO immediately prior to induction with 1mM IPTG and protein expression performed for 16 hours at 18°C. Cells were pelleted by centrifugation, re-suspended in A1 buffer (500mM NaCl, 20mM HEPES pH 7.5, 10% glycerol) with protease inhibitor cocktail, and cells lysed by sonication. Insoluble cell debris were removed by centrifugation and the clarified lysate underwent a two-step purification programme- nickel-affinity IMAC on a pre-equilibrated 5mL column, followed by preparative gel filtration using a sepharose S75 superdex column. Analysis of the fractions by SDS-PAGE indicated that the construct was still expressed solubly following BAP removal, and could be purified with few contaminants (Figure 3.3.9).

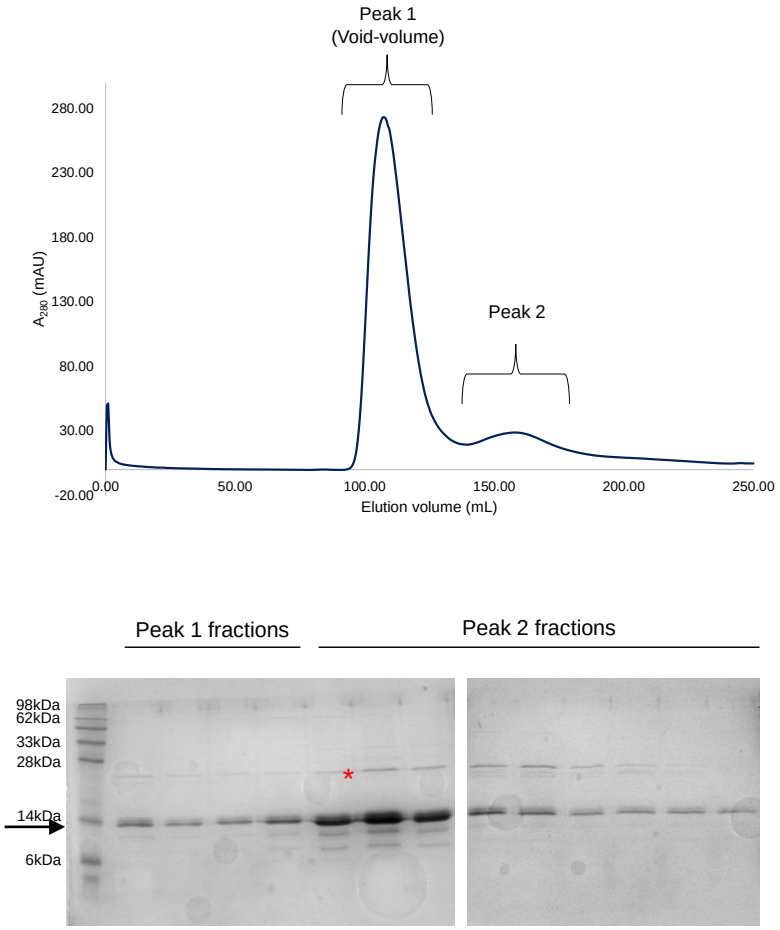


Figure 3.3.9 Size-exclusion chromatography purification of construct #1 following nickel-affinity IMAC

Size-exclusion chromatography of soluble protein from 8L expression of construct #1 without the C-terminal biotin acceptor peptide (BAP) after total soluble protein has been purified by nickel-affinity IMAC to enrich histidine-tagged protein. Gel filtration trace (top) shows the protein eluting as two major peaks- one corresponding to the void-volume of the column (peak 1, ~110mL), and a second peak at approximately 160mL. Void-volume elution is indicative of either insoluble protein, or the formation of large (>75kDa) complexes or aggregates. SDS-PAGE analysis of both peaks (bottom) shows a band of the expected molecular mass for this construct, indicated by arrow, in both the void-volume, and non-void-volume peaks, with few contaminants. The main contaminant co-eluting across all fractions is highlighted by *; the absence of further contaminants under peak 1 may be indicative of construct #1 aggregation, rather than forming a complex with contaminating proteins.

When the protein was concentrated it was discovered that above ~0.5mg/mL it precipitated. In order to overcome this issue, dynamic light scattering (DLS) was used to identify buffer conditions in which the protein may be less likely to form insoluble aggregates. Protein concentrated to 0.5mg/mL was diluted 1:1 with the appropriate buffer conditions and incubated for 30 minutes prior to DLS measurements.

Initial screens using additives frequently used to decrease the polydispersity of proteins prior to setting up crystallisation trials were unable to identify single components capable of significantly decreasing the average particle size in solution (Table 3.3.2).

Treatment	R (nm)	MW (kDa)
KCl	31.7	10913
$CaCl_2$	47.2	27752
NH_4Cl	31.1	10504
$(CH_3)_2AsO_2H$	29.6	9310
Na_2SO_4	39.1	17925
LiCl	32.2	11325
NaCl	29.5	9217

Table 3.3.2 Effect of protein additives on particle radius by DLS

DLS measurements of the average particle size of R3a construct #1 and corresponding predicted molecular weight. Dilution of construct #1 into 0.1M solutions containing either cacodylic acid or NaCl resulted in a small decrease in molecular mass, although well above the predicted molecular mass of monomeric protein (~14.5kDa).

Although certain additives can prevent aggregation of specific proteins, this phenomenon is difficult to generalise, and DLS is not suitable for high-throughput screens. As none of these commonly used additives had a significant effect on the average particle size, we decided to take a more systematic approach using a more generalised method. A key determinant of protein-protein interactions is the buffer pH, as this can have a dramatic effect on side-chain charges and thus the properties of interaction interfaces. Additionally, the presence of salts such as sodium chloride can stabilise proteins in solution, or disrupt salt bridges. We therefore decided to perform a pH and NaCl screen to identify improved buffer conditions, around which further optimisations could be performed (Table 3.3.3).

pH	NaCl (mM)	R (nm)	MW (kDa)
6.5	0	50.6	32628
6.5	500	109.1	197185
7	0	54.6	39073
7	500	38.2	16925
7.5	0	16	2213
7.5	500	3.2	51
8	0	6.7	292
8	500	41.3	20297
8.5	0	5.6	192
8.5	500	34.4	13297

Table 3.3.3 pH and salt screen to identify suitable buffer conditions for non-aggregated R3a construct #1

DLS screening of purified R3a construct #1 against different pH Bis Tris propane (BTP) buffer with and without 500mM NaCl found that diluting the protein in buffer at pH 7.5 with 500mM NaCl led to a dramatic decrease in particle size and corresponding predicted molecular weight. Interestingly the addition of NaCl did not have a predictable effect on particle size, with added salt leading to an increase in particle size at pH 6.5, 8.0 and 8.5.

Incubation of the protein at pH 7.5, with the addition of 500mM NaCl led to a significant decrease in particle size, and corresponding decrease in predicted molecular weight in solution. As the addition of sodium chloride did not have a predictable effect on the behaviour of this protein, in many cases the addition of the salt led to an increase in particle size, we decided to assay the effect of increasing salt concentration on particle size at pH 7.5 (Table 3.3.4).

NaCl (mM)	R (nm)	MW (kDa)
0	15.2	1949
100	4.7	125
200	5.2	159
300	5.5	181
400	ND	ND
500	1.4	8

Table 3.3.4 Effect of increasing salt concentration on size of R3a construct #1 aggregates

DLS measurements of particle size of R3a construct #1 diluted in BTP pH 7.5 supplemented with increasing concentrations of NaCl showed erratic responses to salt, however generally the addition of NaCl decreased average particle size. No condition was identified that corresponded to a predicted molecular mass of monomeric (14.5kDa) or dimeric protein (29.0kDa). The last two conditions indicate a molecular mass smaller than that of monomeric protein, and more importantly, smaller than identical conditions from previous experiments (pH 7.5 + 500mM NaCl with a particle radius of 50nm, Table 3). ND- not determined.

Measurements of particle size appeared to indicate a dramatic decrease in response to increasing salt concentration, which would suggest that under these conditions the protein is less likely to form aggregates. The last two conditions however, suggested that the protein had formed particles with a molecular mass significantly less than that of a monomer. Furthermore, comparison of the particle size at pH 7.5 with 500mM NaCl from this assay with identical conditions performed at an earlier time point (Table 3.3.3) showed a significant difference in predicted molecular mass.

We suspected that this dramatic change in molecular mass was related not to reduction of aggregation, but by the protein undergoing degradation. We tested this by running SDS-PAGE of samples of protein collected over multiple time points and found that the protein was degrading over time (Figure 3.3.10). This degradation meant that we were unable to identify conditions reducing protein aggregation, and thus were unable to further purify this construct.

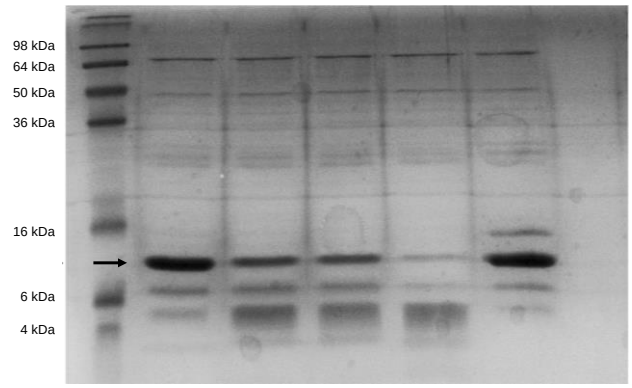


Figure 3.3.10 SDS-PAGE analysis of R3a construct #1 stability over time

Coomassie stained 17% SDS-PAGE gel of construct #1 samples taken from 3, 5, 7 and 14 days storage at 4°C. The end lane shows protein sample flash frozen in liquid nitrogen and stored at -80°C after 24 hours storage at 4°C. Arrow indicates predicted size of construct, migration to lower molecular mass, indicative of degradation may indicate loss of N-terminal histidine tag.

3.3.2.2 R3a construct # 66 (C-terminal LRR region)

Construct #66$^{\Delta BAP}$ (R3a$^{1180\text{-}1282}$) in BL21^{AI} was grown in 8L of LB and induced with arabinose with 0.5% DMSO added immediately before induction. Protein expression continued for 16 hours at 18°C. Cell harvesting, and lysis were performed in the same manner as construct #1. Protein was purified by nickel IMAC followed by gel filtration on a superdex sepharose S75 column (Figure 3.3.11).

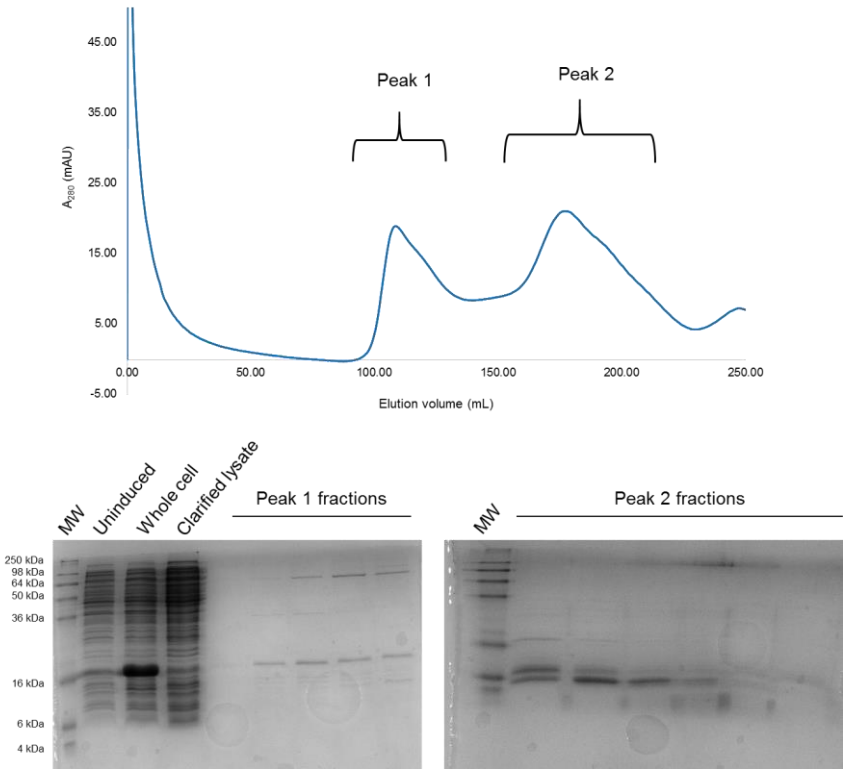


Figure 3.3.11 R3a construct #66 purification

Top: preparative gel filtration of R3a #66 following IMAC. A significant proportion of protein is present in the column void-volume (~110mL), suggesting large molecular weight or insolubility.

Bottom: SDS-PAGE and coomassie staining of fractions eluting from the column. Peak 1 fractions consist of a protein of the expected molecular weight, with a high-molecular weight contaminant co-eluting. Peak 2, distinct from the void-volume, is populated by proteins with a lower than anticipated molecular mass, indicating either a contaminant, or degradation of the target protein.

A large proportion of this protein eluted in the void-volume of the column, indicating either a complex with high molecular-weight contaminants, or formation of large aggregates. SDS-PAGE reveals that protein eluting away from the void volume (peak 2) migrates as lower molecular mass protein than expected. This lower molecular mass species could be either degradation of the target protein, or may be an undesired contaminant co-purifying with this construct. Low yields of the protein meant that we were unable to perform N-terminal tag cleavage and a second round of purification with this construct.

3.3.2.3 R3a construct # 15 (N-terminal NBARC region)

The yield and purity of construct #15 was investigated by purification of this protein from 8L of soluBL21 *E. coli*. At the point of induction with 1mM IPTG, cultures were supplemented with 0.5% DMSO, and protein expression performed for 16 hours at 18°C. Soluble protein was purified by IMAC and gel filtration (Figure 3.3.12).

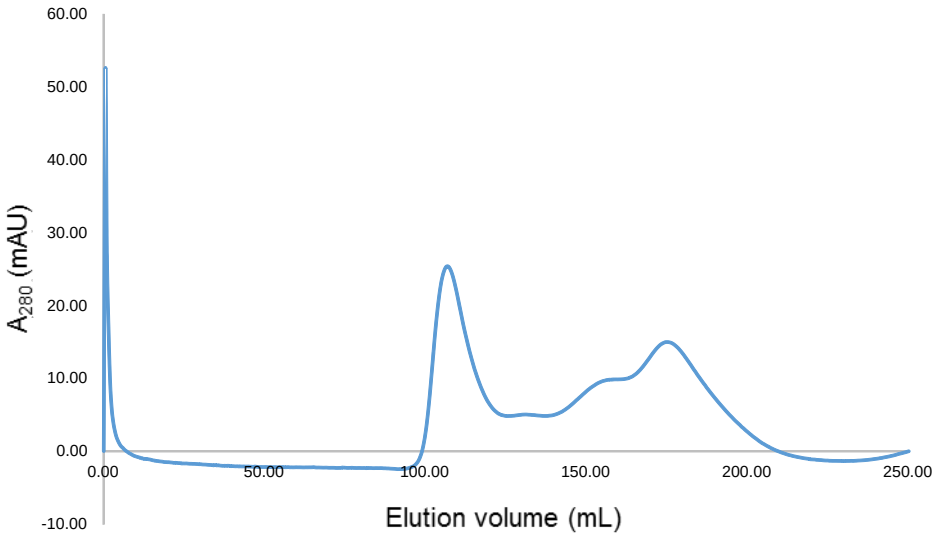


Figure 3.3.12 gel filtration trace of R3a construct #15$^{\Delta BAP}$

Purification of R3a construct #15$^{\Delta BAP}$ after nickel-affinity IMAC suggests a large proportion of the protein is insoluble or in large molecular-weight complexes, with a relatively large elution peak at 110mL (column void-volume). A second major peak can be identified at approximately 175mL with a minor peak that partially overlaps, at approximately 160mL. These additional peaks may indicate different oligomeric states, co-purification with contaminants, or degradation of target protein during purification.

The low A_{280} of the protein fractions, combined with the presence of multiple peaks (seen as “shoulders” where they overlap) indicate that despite the large cultures used, total soluble protein yields remain very low, and unlikely to be high-purity.

In a further attempt to improve yields of purified protein, these three constructs were cloned into the pOPIN-F vector, which uses 3C protease to post-translationally cleave the N-terminal histidine tag. However, this change in vector did not lead to a significant increase in protein yield or purity (data not shown).

3.4 Rational design of R3a constructs

3.4.1 Defining construct boundaries

Alongside the assessment of the constructs from the high-throughput screening, a number of R3a constructs were designed using traditional bioinformatics-led approaches. Scorer2.0 (http://coiledcoils.chm.bris.ac.uk/Scorer/), PFAM (http://pfam.xfam.org/) and LRRfinder (http://www.lrrfinder.com/lrrfinder.php) were used to identify putative boundaries for the coiled-coil, NBARC and LRR domains respectively. This information was supplemented with RONN disorder analysis (https://www.strubi.ox.ac.uk/RONN) to try and limit the inclusion of N- or C- terminal disordered regions in the constructs. As PFAM predictions of the NBARC domain did not extend to cover the MHD motif, a highly conserved motif shown to be involved in the regulation and activation of several R-proteins, the NBARC construct was extended to include this region. Secondary structure prediction by PHYRE2 was used to prevent construct termination mid-secondary structure, which may destabilise expressed proteins. A summary of these predictions and the eventual domain boundaries used is presented in Table 3.4.1.

	Prediction	Notes	Boundaries used	Rationale
CC	**29-119**	**Start mid-helix**	**20-160**	**Includes additional helix (124-150)**
NBARC (PFAM)	**180-463**	**Excludes MHD motif**	**161-583**	**Includes MHD and surrounding SS elements**
LRR (PFAM)	**585-1238**	**Terminates in disordered region**	**584-1249**	**Extends to end of β-strand**
LRR (LRR FINDER)	**412-1282**	**Start within predicted NBARC**		
N-terminal disorder	**-**			
C-terminal disorder	**1153-1238**			

Table 3.4.1 Summary of R3a domain boundaries

R3a domain boundaries were predicted using multiple bioinformatic servers. The CC domain was extended to include an N-terminal helix omitted from the Scorer2.0 CC prediction. PFAM prediction of the NBARC domain terminated before the highly conserved MHD motif, and so this construct was extended to include this motif, as well as the neighbouring secondary structure elements to increase stability and preserve biological context for this region. As LRR domains are highly variable and poorly characterised, both PFAM and LRRfinder were used to determine the start of this region. Given the difficulties in confidently predicting this structural unit, the LRR was designed to begin at the point of NBARC termination. Disorder prediction by RONN identified no significant disorder at the N-terminus of R3a and an island of disorder towards the C-terminus, however the very C-terminal region was not predicted to be intrinsically disordered. Additional information regarding the predicted secondary structure elements was provided by PHYRE2 and used to limit premature termination within structured regions.

3.4.2 Cloning and expression screening

These domains were cloned into the pOPIN-F (hist-tag:ORF) and pOPIN-M (his-tag:MBP:ORF) vectors and screened for soluble expression as described in Section 2.8.1. A construct spanning the coiled-coil domain of R3a in pOPIN-F ($R3a^{20-160}$) appeared to express well in small-scale screens in soluBL21 (Figure 3.4.1), and so was selected for large-scale expression. We were unable to identify conditions that resulted in soluble expression of either the NBARC or LRR domain when assessed by SDS-PAGE and coomassie staining.

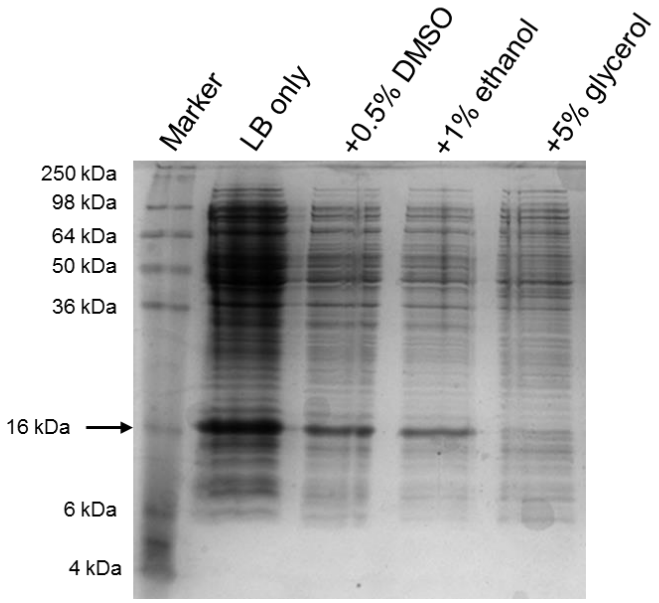


Figure 3.4.1 Total soluble protein from R3a CCF-1 expression trials

Coomassie stained SDS-PAGE gel of expression trials of the coiled-coil domain of R3a in soluBL21 show good levels of soluble protein under three of four expression conditions. Bacterial cultures were incubated at 18°C with constant shaking after induction with 1mM IPTG. Arrow indicates band migration consistent with the predicted molecular weight of the construct.

250mL of LB was used to grow R3a$^{20\text{-}160}$ overnight at 37°C with constant shaking. This overnight culture was then used to inoculate 4L of LB, which were grown to an OD_{600} 0.6 before being rapidly chilled on ice. Protein expression was induced by the addition of 1mM IPTG, and cultures incubated at 18°C for 16 hours. Clarified lysate was produced as described in Section 2.8.1, and purified in a two-step procedure; nickel IMAC followed by gel filtration on a superdex S75 sepharose column.

SDS-PAGE of the fractions following gel filtration showed a number of contaminants co-eluting with the target protein, however the large amounts of protein of the correct approximate molecular mass indicated that a second round of purification may remove these contaminants while retaining sufficient yields of target protein. Fractions were pooled and concentrated to ~10mL, with no signs of aggregation, and the N-terminal histidine tag removed by incubation with 3C protease overnight at 4°C. Cleaved protein was purified by nickel-affinity IMAC, with flow-through and washes collected and pooled to <5mL. This protein solution was then loaded onto a superdex S75 column for buffer exchange, further purification, and to check that the protein eluted away from void-volume, which would indicate protein aggregation or insoluble protein (Figure 3.4.2).

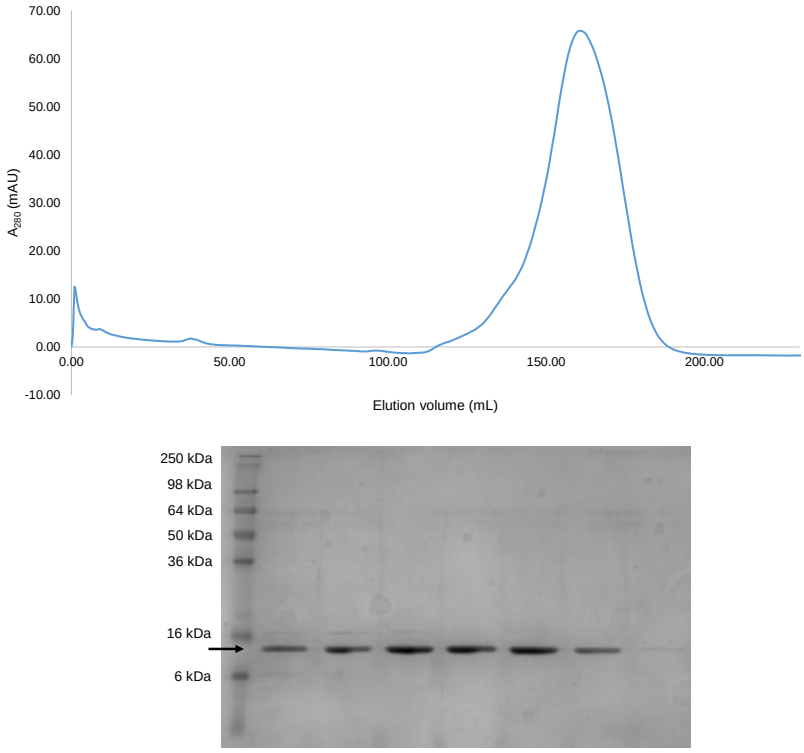


Figure 3.4.2 Gel filtration trace and collected fractions of R3a$^{20\text{-}160}$ after his-tag removal

R3a$^{20\text{-}160}$ treated overnight with 3C protease was purified by nickel IMAC to remove residual tagged protein, 3C protease and negatively charged contaminants. Flow-through and washed were pooled and concentrated to <5mL to improve resolution of size-exclusion chromatography. Top, separation of proteins on a superdex S75 column shows a predominant peak at ~160mL with a minor, less well defined peak at ~120mL. Bottom, SDS-PAGE of fractions collected from the major peak shows a single band of approximately the correct molecular mass with few contaminants or degradation products.

Gel-filtration traces of the construct following histidine tag removal showed that the protein eluted as a single, broad peak with a maximum absorbance at 160.7mL, which correlates to a protein with molecular mass of approximately 38.8kDa, which does not correspond well to either dimeric (~32.2kDa) or trimeric (~48.3kDa) protein. However, this method is not very appropriate for accurate protein mass determination in this instance, as a key assumption if that the proteins are roughly globular. Coiled-coil proteins such as MLA10 may take on a much more elongated structure[86], altering the way in which they migrate through the column. SDS-PAGE analysis of the fractions (Figure 3.4.2, bottom) showed that the protein eluted with very few contaminants, and so these fractions were pooled and concentrated. This protein was expected to migrate with the lysozyme MW marker band (16kDa), however the band corresponding to the target protein has an apparent lower molecular mass. Native protein was sent for intact mass spectrometry (MS), and a

single peak corresponding to a mass of 16192.0Da identified (Appendix), which is within the margin of error for the predicted mass 16191.39Da.

Despite being very similar to construct #1 identified in the high-throughput screen (the difference being a C-terminal extension of 40 amino acids), this coiled-coil construct showed none of the issues with aggregation or stability, and so was determined to be appropriate for down-stream experiments.

3.5 Medium-throughput screening of Solanaceous coiled-coil R-proteins

Note: Cloning was performed by Richard Hughes (JIC). Initial screens in E. coli were performed by Richard Hughes and Heather Rada at Oxford Protein Production Facility. Screening in insect cells was performed by Heather Rada.

Having demonstrated that rational design of R-protein constructs was able to yield soluble protein from *E. coli* expression, we decided to investigate whether other R-protein domains could be expressed solubly, as the differences in protein sequences between these proteins may result in differences in expression in this heterologous system. We additionally broadened our expression systems to include both *E. coli* and insect cells as we hypothesised that the complex eukaryotic cell environment may be more suitable for the expression of these proteins.

3.5.1 Construct design

Constructs spanning the full-length ORF, predicted coiled-coil, NBARC and LRR domains were designed for *Solanum tuberosum* R3a, *S. demissum* R2, *S. lycopersicum* I-2, *S. bulbocastrum* Rpi-blb2, *S. lycopersicum* NRC1 and *N.* benthamiana NRC1-like following a similar workflow as described for R3a in Section 3.4.1. The results from these prediction servers are summarized in Table 3.5.1. As we were able to express the coiled-coil domain of R3a, we also used sequence alignments between these R-proteins and R3a to help define the boundaries for the coiled-coil domains. Each construct was cloned into pOPIN-F and pOPIN-S3C (his:SUMO:ORF) and transformed into the *E. coli* strains LEMO and Rossetta 2, as well as used in baculovirus-mediated transformation of insect-cell cultures.

R-protein	Predicted CC	Predicted NBARC	LRR start (PFAM)	LRR start (LRRfinder)	CC used	NBARC used	LRR used
R3a	29-119 ; 124-157	180-463	585	586-	1-160	161-554	555-1282
I-2	29-189	173-457	601	578-	1-150	151-548	549-1266
R2	6-56 ; 96-133	165-447	595	(399-527) ; 528-	1-156	157-499	500-845
Rpi-blb2	1-94 ; 322-386	536-815	1175	939-	1-512	513-866	867-1267
NRC1	1-141	160-443	-	(15-769) ; 770-	1-149	150-494	495-888
NRC1-like	1-136	159-440	-	741-	1-149	150-490	491-881

The full ORFs of R3a, I-2, R2, Rpi-blb2, NRC1 and NRC1-like were translated and submitted to multiple bioinformatic servers to predict likely domain boundaries. This information was used alongside sequence alignments to the R3a$^{(20\text{-}160)}$ construct, which can be expressed solubly in *E. coli*, as well as secondary and tertiary structure predictions by PHYRE2 to try and maximise stability and maintain structural units.

Table 3.5.1 Summary of domain predictions and domain boundaries used for R-protein cloning

3.5.2 Initial Screens (performed at OPPF by R. Hughes)

Transformants were placed in 96-deep-well plates and grown in PowerBroth (PB) overnight at 37°C with constant shaking. Overnight cultures were used to inoculate 3mL of fresh PB in four 24-well blocks, and incubated until an average OD_{600} of 0.5 was reached. Cultures were cooled to approximately 20°C before gene expression induced by the addition of 1mM IPTG, with expression continuing at 18°C overnight. 1mL of overnight culture was transferred to a 96-deep-well block and cells pelleted by centrifugation. After centrifugation, media was removed and cells frozen at -80°C for over 20 minutes. Cells were lysed by thawing at room temperature and resuspension in buffer containing detergent (TWEEN 20) and lysis clarified by centrifugation. 1mL soluble extract was added to a 96-well flat bottomed plate containing washed Ni-NTA magnetic beads and incubated for 30 minutes. Beads were removed magnetically, washed, and bound proteins eluted before loading onto SDS-PAGE gels for separation and visualising by coomassie staining (Figure 3.5.1).

Medium-throughput screens of *E. coli* expression were able to identify soluble expression of the coiled-coil domains of R3a, I-2, R2, NRC1 and NRC1-like, and the NBARC domain of NRC1 all in pOPIN-S3C in BL21 LEMO cells. Interestingly the results of expression trials in insect cells were unable to identify any constructs expressing solubly that were not found in the *E. coli* screen (data not shown).

We then tested these constructs in-house to see if this soluble expression was reproducible, and whether soluble expression could be seen without the initial low-stringency nickel IMAC and detergent. These constructs were grown in LB overnight at 37°C, and overnight cultures was used to inoculate 2L fresh PB. Once cells had reached an OD_{600} of 0.6, cultures were rapidly chilled, induced by the addition of 1mM IPTG and incubated at 18°C for 16 hours. Cell pellets were re-suspended in A1 buffer with protease inhibitor, lysed by sonication and processed as described in Section 2.8.1. Clarified lysate was purified by nickel IMAC followed by gel filtration to assess solubility. All six constructs appeared to express solubly under these conditions, although yields were not very high.

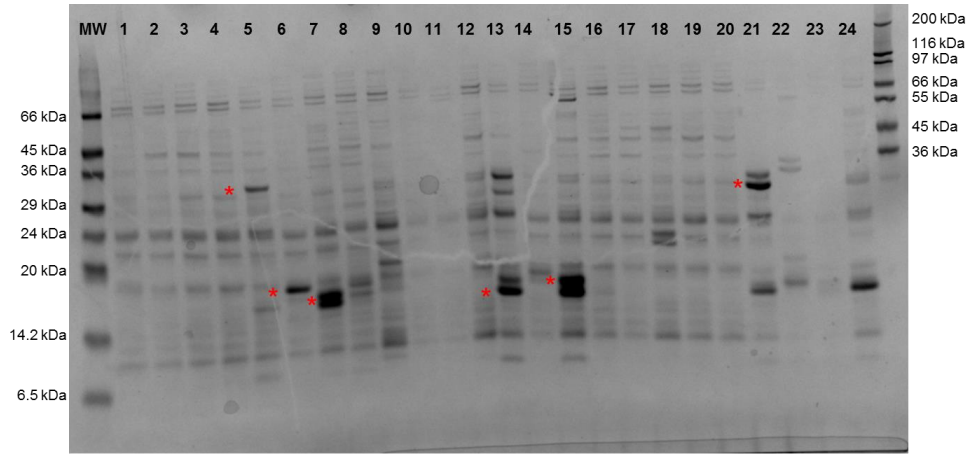


Figure 3.5.1 Coomassie stained SDS-PAGE of E. coli expressed R-proteins and their domains in pOPIN-F and pOPIN-S3C

SDS-PAGE soluble extract from *E. coli* expressing CNL R-proteins and their domains cloned into pOPIN-F and pOPIN-S3C following nickel IMAC shows a number of soluble constructs. Positive constructs are clearly identifiable in lanes: 5 (R3a^{CC}-F); 6 (R3a^{LRR}-S3C); 7 (R3a^{FL}-S3C); 13 (R2^{CC}-S3C); 15 (R2^{CC}-S3C); 21 (I-2^{CC}-S3C). Gel from Richard Hughes.

3.5.3 Large-scale expression and purification

Having identified conditions allowing for soluble expression, we decided to test large-scale expression of these constructs to determine approximate protein yields and purity.

4L of PB was inoculated with overnight cultures for each construct and induced by addition of 1mM IPTG after rapid chilling on ice. Cultures were grown for a further 16 hours at 18°C, and harvested and processed as described in Section 2.8.1. Clarified lysates were loaded onto pre-equilibrated Ni-NTA columns and bound proteins eluted directly onto a superdex S75 column. Fractions were collected only when a minimum A_{280} threshold had been reached, with the exception of NRC1 coiled-coil domain for which all fractions were collected, as this construct lacks tryptophan residues required for efficient absorbance at this wavelength. The coiled-coil domain appeared to express solubly, however low yield meant that the A_{280} did not reach the threshold for collection (20mAU), and so purity could not be assessed. The coiled-coil domains of I-2, NRC1 and NRC1-like, as well as the NBARC domain from NRC1 all expressed reasonably well. Gel filtration fractions for these proteins were pooled individually, and the N-terminal tag removed by overnight incubation with 3C protease at 4°C. The coiled-coil domain of R3a expressed solubly in this

instance, however yields were too low for the protein to undergo tag removal, and given the identification of a well-expressing construct this protein was not prioritised for purification. Purification of coiled-coil domains of R2, I-2, NRC1, and NRC1-like can be seen in Figure 3.5.2

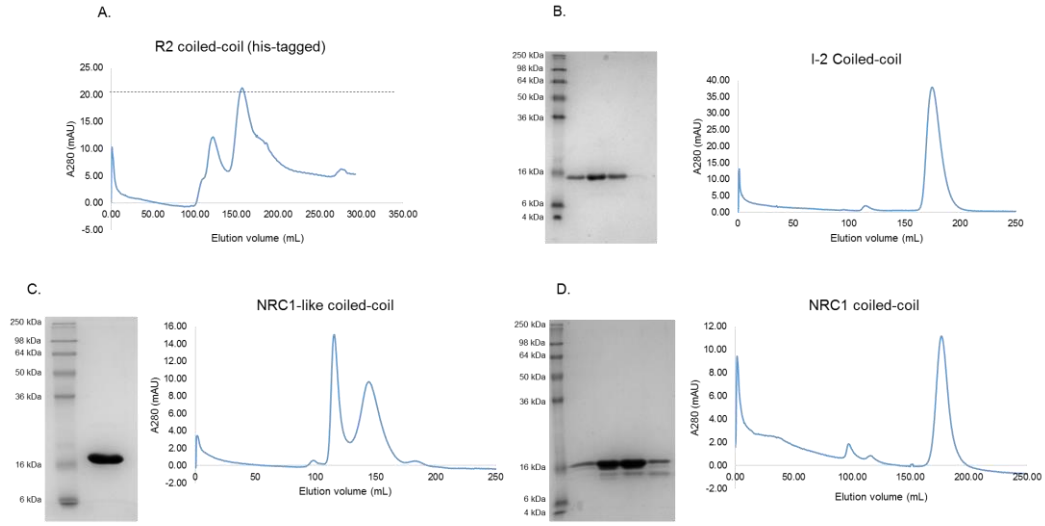


Figure 3.5.2 Purification of the coiled-coil domains of solanaceous CNL coiled-coils

Two-step purification of coiled-coil domains of CNLs in pOPIN-S3C by nickel IMAC followed by gel filtration. **A**.) The coiled-coil domain of R2 appears to show some soluble expression, however due to low yields (minimum A_{280} for collection of 20mAU not exceeded) fractions could not be collected for SDS-PAGE. **B**). SDS-PAGE and gel filtration of I-2 coiled-coil domain following tag removal. **C**.) SDS-PAGE of pooled and concentrated NRC1-like coiled-coil following tag removal and gel filtration (trace shown to the right) **D**.) SDS-PAGE and gel filtration trace for NRC1 coiled-coil following tag removal. Protein yields for this construct were reasonable- low A_{280} recorded results from a lack of absorbing residues in this protein.

3.6 Troubleshooting NRC1 NBARC point mutant expression

The NBARC domain of NRC1 has been a major part of the work performed in this project. A detailed explanation of the work performed with this construct is presented in Chapter 6. Generation of mutants in this protein has been of high importance, and multiple approaches taken to obtain material. As such we present here a short explanation of the insect cell expression screens performed to generate material.

Point mutants in the NBARC domain of NRC1 could not be expressed solubly in *E. coli* despite extensive screening efforts. As the wild-type protein could be expressed

in insect cells with only an N-terminal histidine tag, we screened the point mutants in both pOPIN-F and pOPIN-S3C for soluble expression in this expression system. Surprisingly we were unable to detect expression of four of the seven point mutants in insect cells regardless of the tag used. Only one mutant appeared to express to a similar level as the wild-type protein, with marginal expression for the remaining two (Figure 3.6.1).

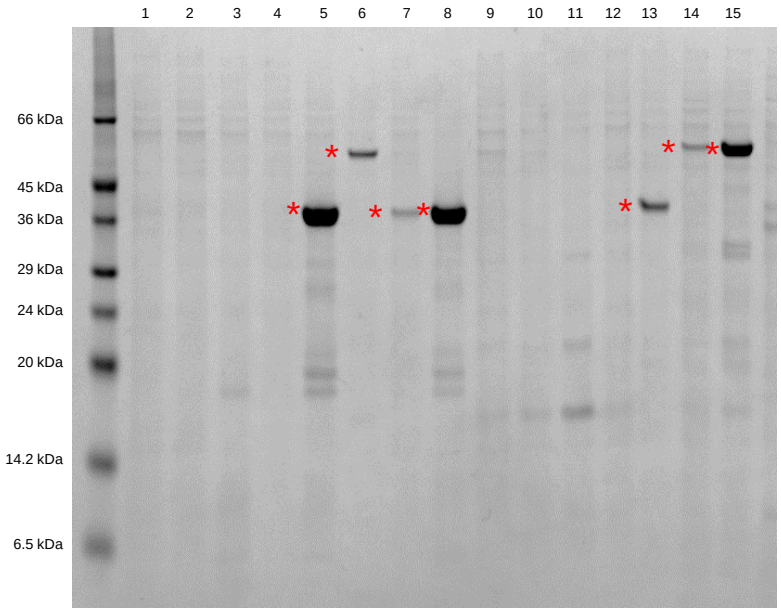


Figure 3.6.1 Low-stringency purification of NRC1 point mutants expressed in Sf9 cells

SDS-PAGE of baculovirus-mediated expression of NRC1 NBARC point mutants in insect cells. 1) pOPIN-F:K191A, 2) pOPIN-F:D481V 3) pOPIN-F:H480F , 4) pOPIN-F:H480A 5) pOPIN-F:D268N 6) pOPIN-S3C:D267N 7) pOPIN-F:D267N_D268N 8) pOPIN-F:WT 9)pOPIN-S3C:K191A 10) pOPIN-S3C:D481V 11) pOPIN-S3C:H480F 12) pOPIN-S3C:H480A 13) pOPIN-F:D267N 14) pOPIN-S3C:D267N_D268N 15) pOPIN-S3C:WT. Intensity of bands suggests that NRC1^{D268N} in pOPIN-F may express to a similar level to wild-type protein, with NRC1^{D267N} and NRC1^{D267N_D268N} also expressing solubly, however to a much lower level.

These three constructs (wild-type, D268N and D268N_D267N) were then scaled up to 2.5L insect cell expression, pelleted and frozen at -80°C. Pellets were thawed in lysis buffer and lysed by cell disruption (Constant Cell disruption systems). Cell lysate was clarified by centrifugation. Benzonase was added to the supernatant and incubated at room temperature for 15-30 minutes before passing the supernatant through 0.45µm filters. Proteins were then purified in a two-step process, similar to that of *E. coli*-derived proteins, however using buffers optimised for insect-cell purification. IMAC was performed using Ni^{2+}-NTA matrices equilibrated with insect cell A1 (Section 2.1.8.2) and gel filtration performed with a superdex S75 column

pre-equilibrated with insect cell A4 running buffer. After gel filtration, fractions containing the target protein were pooled and concentrated to <10mL for removal of the N-terminal histidine tag by incubation with 3C protease at 4°C overnight. NRC1 D268N_D268N did not reach a suitable level of purity or yield for tag cleavage. The cleaved proteins were then purified by IMAC to remove tagged protein and protease, and assessment of purity by SDS-PAGE indicated that these proteins had reached sufficient purity for downstream experiments without requiring an additional round of gel filtration (Figure 3.6.2).

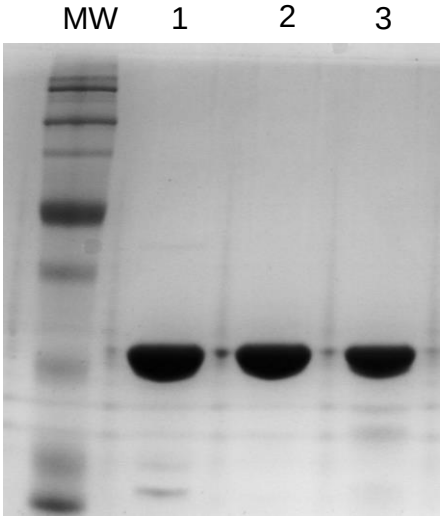


Figure 3.6.2 SDS-PAGE gel of purified NRC1 NBARC and point mutant from E. coli and insect cell expression

Final purified NRC1 NBARC domain used in downstream experiments. 1) WT protein derived from *E. coli* expression, 2) WT protein purified from insect cells, 3) NRC1 NBARC^{D268N}, purified from insect cells.

3.7 *In planta* expression of R3a domains

One potential reason for the difficulties of expressing R-proteins in *E. coli* is the lack of plant protein-chaperones that may have co-evolved to aid R-protein folding, such as the dependency of I-2 on the presence of the chaperone RSI2[136], or the more general reliance on the HSP90-RAR1-SGT1 co-chaperones[137, 138]. To address this, the NBARC and LRR domains of R3a identified in Section 3.4.1 were cloned into a modified version of the pTRBO vector. This vector, which uses the CaMV 35S promoter to drive expression, was modified by Richard Hughes to include the N-terminal histidine tag and restriction sites from the pOPIN-F vector. The same PCR products used to clone the domains into the *E. coli* expression vectors pOPIN-F and

pOPIN-S3C were used for the cloning reaction. Cloning was performed as described in Section 2.2.2, and 5µL of cloning reaction was transformed into DH5α. Transformants were plated on LB agar supplemented with kanamycin and incubated at 37°C overnight. Single colonies were selected and grown in LB liquid media with kanamycin overnight at 37°C. These liquid cultures were pelleted for DNA purification (Section 2.2.5) and sent for sequencing using the Cauliflower Mosaic virus (CMV) 35S promoter primer (Appendix).

Positive plasmids were transformed into *Agrobacterium tumefaciens* strain GV301.3 by electroporation, plated on L agar and grown for 48 hours at 28°C. Single colonies were selected and screened for the presence of the insert by PCR using the CMV 35S promoter primer and gene-specific reverse primer. Positive colonies were grown as liquid cultures for two days to be used for transient transformation of *N. benthamiana*. *A. tumefaciens* cells were pelleted by centrifugation and re-suspended in infiltration buffer (10mM MES pH 5.6, 10mM $MgCl_2$, and 150µM acetosyringone) to an OD_{600} of 1.2-1.5. After 2-4 hours incubation re-suspended *A. tumefaciens* were pressure infiltrated into five four-week old *N. benthamiana* leaves.

After 5 days, leaves were harvested, frozen in liquid nitrogen and stored at -80°C. To isolate soluble protein, leaves were ground in liquid nitrogen with a chilled mortar and pestle, and 0.5mg powdered leaf aliquoted into 2mL screw-top tubes on dry ice. Cells were lysed by thawing in 2x v/w lysis buffer for 30 minutes on ice with several rounds of vortexing. Cell debris was removed by centrifugation and the clarified lysate pooled for all replicates. Clarified lysates were filtered with 0.45µm filters before being loaded onto a pre-equilibrated nickel column. For low-stringency purification 15mM imidazole was added to supernatant prior to loading onto a 5mL HisTrap FF column. Columns were then washed with three column-volumes of wash buffer (50mM Tris-HCl pH 8.0, 500mM NaCl, 10mM NaCl, 5% v/v glycerol) before eluting in 3 column-volumes of B1 buffer (Figure 3.7.1).

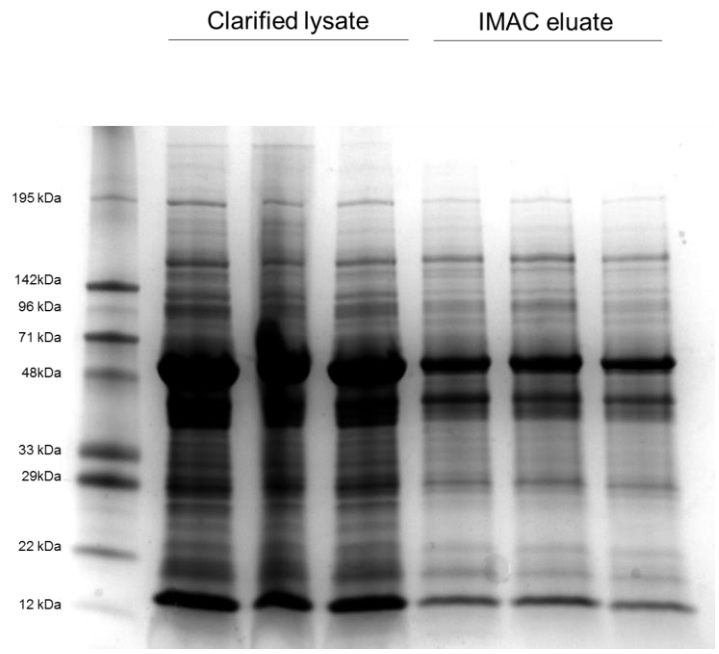


Figure 3.7.1 Low-stringency purification of R3a domains from N. benthamiana

Low stringency purification of soluble proteins from *N. benthamiana* does not result in accumulation of protein bands of the correct size, which are not found in wild-type tissue extract.

We were unable to detect significant accumulation of any of the constructs by SDS-PAGE, although a number of eluting proteins are visible. High-stringency purifications, with 30mM imidazole added to samples prior to loading onto a 1mL HisTrap FF column did not result in accumulation of protein bands at the appropriate molecular weight. It is possible that these protein domains are subject to post-translational control to prevent appreciable accumulation *in planta* despite the use of high-expression promoters.

3.8 Discussion

Prior to this project, we suspected that the structure and *in vitro* biochemistry of R-proteins has been the subject of far fewer published investigations than their behaviour at the genetic level and cellular level due to difficulties associated with obtaining sufficient yields of suitable protein. We have undertaken extensive screens to try and obtain R-protein constructs that would express solubly in *E. coli*, insect cells and *N. benthamiana*. A large portion of our efforts have been focussed on *E. coli* as our expression system of choice. We focussed on this expression

system for a number of reasons: the cells are easily transformable, they grow rapidly on simple media, there exist multiple strains and expression vectors, medium/high-throughput screens are relatively straight forward and less time intensive than other systems, scale-up of expression is relatively straightforward requiring little specialist equipment, and once appropriate conditions are identified, they are often able to produce large amount of protein. There are significant drawbacks to using such a simple prokaryotic system- most *E. coli* strains are unable to catalyse disulphide bridges, they are unable to apply many post-translational modifications, such as glycosylation that may aid stability or solubility, and they lack some of the more complex chaperones that may have co-evolved with the target protein and may be required for correct folding.

By screening constructs against a wide range of experimental conditions we have been able to produce soluble R-protein domains for downstream experiments, suggesting that at least some of the difficulties associated with expression of these protein domains may be overcome by the right combination of construct, strain and expression conditions. As we have not been able to rationalise why certain constructs expressed in specific conditions, groups wishing to investigate these proteins *in vitro* will most likely have to undertake similar, intensive approaches to obtain the raw material for downstream investigations.

In addition to the two published examples of soluble coiled-coils we have been able to identify a further four solanaceous coiled-coil domains that can be purified from *E. coli*, those of R3a, I-2, NRC1 and NRC1-like. There are many questions regarding the behaviour and structure of coiled-coil domains which have yet to be resolved through *in vivo* methods- whether there is a representative fold similar to MLA10 or Rx, how these coiled-coils interact with downstream signalling partners, and how these signalling domains are kept in a signalling-competent, inactive state. By increasing the number of examples of R-protein coiled-coil domains available for study, we hope to be able to understand what generalisations can be made about the behaviour of these proteins, and better understand how they affect R-protein function.

Additionally we have been able to produce, to our knowledge for the first time, a soluble NBARC domain from a plant R-protein with relatively high yields and purity. With this new material we hope to better probe this domain, pivotal to the proper regulation of many R-proteins. Producing protein solubly will allow us to deconvolute some of the issues regarding the homogeneity of refolded protein, and may lead to

novel insights about the behaviour of this protein domain *in planta*. Alongside exploring the biochemistry of this domain, the reasonable yield and reproducibility of the purifications allows for structural investigations using techniques such as x-ray crystallography and small-angle x-ray scattering. It is hoped that high-quality, high resolution data about the structure of the NBARC domain will give us insight into the evolution of this domain, allow an understanding of the subtle changes that may distinguish the “on” state with “off” state in the context of the full-length protein, and explain the mechanistic basis for the importance of certain conserved motifs.

Although we were able to express a number of constructs in *E. coli*, this approach has a number of limitations. Most surprising was the inability for *E. coli* to express many point mutants in the NBARC domain of NRC1 solubly. This inability to produce soluble protein is difficult to explain- the point mutations were predicted to have very different effects on the biochemistry of this protein, whilst minimising any effects on the overall structure. From work trying to produce apo-NBARC domain we may have expected that mutations preventing nucleotide binding, such as the P-loop K191A mutation, would result in unstable, and therefore insoluble protein. What we did not anticipate is that mutations that were predicted to have little effect on the ligand occupancy of the protein would also lead to issues with expression.

As we were unable to overcome these issues with expression in *E. coli*, we chose to explore alternative systems to produce these proteins- primarily insect cell expression. Our initial expectations were that this system would give lower overall yields than proteins expressed in *E. coli*, but that it would still be possible to purify the majority of these proteins using a complex eukaryotic environment. Surprisingly we were unable to produce most point mutants in this system- wild-type protein and the point mutant D268N both expressed very well, and the double mutant D267N_D268N appeared to be solubly expressed, however to such low yields that we were unable to purify this protein sufficiently for use in our assays.

Throughout this project we have been unable to produce and purify solubly expressing R-protein LRR domains, regardless of the expression system used. Plant cytoplasmic LRR domains as a whole remain uncharacterised and so predictions rely on either mammalian LRR domains, or more frequently, extracellular plant LRR domains such as that of the brassinosteroid receptor BRI1, FLS2, or immune receptors from animal systems. In the context of R-protein biology, the LRR domain is of great interest. Just as the NBARC domain defines the activation state of the protein, the LRR is predicted to be responsible for effector

detection and auto-inhibition. Studies wishing to expand recognition of have relied on random mutagenesis and labour-intensive screens of many hundreds of variations to try and identify desired phenotypes. Structural and biochemical analyses of this domain are unlikely to be a silver bullet allowing rationalised mutant design *de novo*, but may allow more targeted mutagenesis, making such screens more manageable. Beyond simply allowing rational manipulation of R-protein recognition, the LRR is believed to be key to auto-inhibition of R-proteins, and so our description of the behaviour of these proteins will remain incomplete until we are able to investigate the behaviour of this domain, its affinities and plasticity, more thoroughly.

The soluble expression and purification of full-length R-proteins remains an unsolved issue for studying these proteins biochemically. Regardless of the expression system used we were unable to see significant accumulation of soluble protein in either the simple prokaryotic *E. coli* expression system, or in more complex eukaryotic environments. This study, although screening intensively for soluble protein, was not exhaustive and there remain a number of alternative expression systems to be explored by interested groups: yeast expression systems have been used previously to produce small amounts of soluble protein, and it may be that a similar medium-throughput screen of multiple R-proteins will result in increased soluble yields of different domains, or even full-length protein. As the mechanism which prevents significant accumulation of R-proteins *in planta* is not yet fully understood, there remains the potential to use specific plant hosts for heterologous expression of R-proteins- it may be that using a host separated by sufficient evolutionary time from the target R-protein may bypass such regulatory mechanisms.

4

Probing for R3a interactors

4.1 Introduction- INF1, AVR3a and R3a, PTI and ETI

Plant immunity can be thought of as several layers of increasingly specific plant responses to microbes. Initial detection of the presence of broad classes of microbe is mediated by the detection of conserved, essential molecular motifs (PAMPs), and leads to an effective immune response termed PAMP-triggered immunity (PTI). As these PAMPS cannot be lost or significantly altered without a fitness cost, populations of the pathogen arise expressing specific effectors to supress recognition, or negate the responses that these PAMPs trigger. These effectors may act in the apoplast to inhibit defensive proteins or block molecule perception, or may act in the plant's cytoplasm to disrupt cell signalling, alter transcription or manipulate key host proteins (Section 1.3.2).

In response to the susceptibility brought about by these effectors, plants can evolve specific R-proteins, capable of detecting these effectors and restoring an immune response (Section 1.4). Once activated by the presence of an effector, R-proteins must generate a signal for an immune response to be mounted at the cellular level. Although examples of signal transduction proteins have been identified for certain aspects of plant immunity (for example mitogen-activated protein kinases, MAPKs, involved in PTI), R-protein-specific signalling networks remain very poorly characterised.

The interactions between *P. infestans* and solanaceous plants are a well-studied example of host-pathogen co-evolution and so we used this pathosystem to explore how a particular R-protein (R3a) might be detecting its cognate effector, and to look for interacting signalling components.

4.1.1 INF and the *Phytophthora-Solanaceae* pathosystem

Phytophthora spp. produce a family of small secreted proteins known as elicitins that induce defence responses in a number of plant species[139]. INF1 is a 10kDa cysteine-rich elicitin produced by *P. infestans*, the expression of which shows developmental regulation[140], and when recognised causes rapid immune responses in *Nicotiana spp.* in an HSP90/HSP70-dependent manner[141]. In *N. benthamiana* recognition of INF1 is mediated by direct interaction with the lectin-like transmembrane protein NbLRK1[142] and requires the co-receptor BAK1[143].

Recognition leads to localized cell-death, a more extreme response than those typically associated with PTI, which is very effective at preventing the spread of the hemi-biotroph *P. infestans*.

The role of INF1 in *P. infestans* as a PAMP is an interesting one. Recognition by plant species limits host range and therefore incurs a fitness cost, yet the protein is maintained in all populations as is the case for canonical PAMPs like flagellin and fungal chitin. Expression of the protein however does not seem to be essential to the microbe. Isolates recovered from East Germany (at the time the German Democratic Republic) in 1976-77 showed a loss of INF1 production[144]. These isolates were still able to complete their lifecycle, and later experimental silencing of the INF1 gene led to an ability of *P. infestans* to colonise a normally non-host plant (*N. benthamiana*)[145]. The fact that isolates without INF1 expression were both geographically and temporally restricted (isolates from later time periods in the same region having regained INF1 expression) suggests that these populations were less fit, despite the protein apparently being non-essential for development or infection under these conditions. In order to remain virulent whilst mainlining INF1 production, *P. infestans* populations have evolved proteins to manipulate prospective hosts.

4.1.2 AVR3a and effector-triggered susceptibility

Similar to the production of AvrPto to allow *P. syringae* growth without triggering flg22-mediated resistance, *P. infestans* has evolved cytoplasmic effectors that can supress INF1-induced cell death (ICD). AVR3a is one such effector found in all *P. infestans* isolates to date, with very few polymorphisms. European isolates of *P. infestans* show two predominant alleles of AVR3a, with only three positions showing non-synonymous mutations. Prior to the introgression of R3a from *S. demissum* into the crop plant *S. tuberosum* (Section 4.1.3), the predominant AVR3a allele was $AVR3a^{C19\_K80\_I103}$, however the introduction of R3a led to the emergence of a virulent allele $AVR3a^{S19\_E80\_M103}$, which cannot be recognised by R3a[132]. Of the three polymorphic sites only two, position 80 and 103, are found in the functional effector domain downstream of the RXLR translocation motif, and so these two alleles will be referred to as $AVR3a^{KI}$ and $AVR3a^{EM}$ respectively. The introduction of R3a into commercial varieties of potato led to a dramatic decrease in the prevalence of $AVR3a^{KI}$, however the allele still persists in some natural populations[146]. A possible explanation for this persistence is due to the different extents to which AVR3a alleles are able to suppress ICD. $AVR3a^{KI}$ is able to almost completely abolish the

INF1-mediated immune response, whereas AVR3a^{EM} is much less effective at preventing this cell death[126], which reduces the infectivity of the pathogen. The availability of hosts lacking R3a may therefore determine whether an ability to efficiently suppress ICD is favoured over the ability to evade recognition by R3a. Interestingly the ability of AVR3a to supress ICD appears to be mediated by a different region of the protein than those responsible for R3a recognition[132].

4.1.3 R3a and effector-triggered immunity

The locus giving resistance to *P. infestans* race 3, termed the R3 locus, was first mapped to a distal position on *S. tuberosum* chromosome 11[147], with subsequent work identifying this as a complex locus consisting of at least two functional R-genes[148]. The cloning of the R3a gene, giving resistance to *P. infestans* strains expressing the virulence factor AVR3a, was aided by the synteny between this region of *S. tuberosum* chromosome 11 and the SL8 region of *S. lycopersicum*, known to harbour the resistance gene I-2[149]. This co-linearity, along with knowledge of I2C-1 coding sequence, R3a-associated markers, and the use of bacterial artificial chromosomes ultimately led to the cloning of a 4176bp region encoding a 1282 amino acid predicted CNL, which was able to confer resistance to *P. infestans* races expressing AVR3a.

When R3a becomes activated in response to the presence of AVR3a^{KI}, the result is a rapid, localised HR that can be reconstituted in the model plant *N. benthamiana*, enabling a wide range of functional studies to be performed more easily. Mutagenic screens indicate that R3a LRR is important for recognition of AVR3a, and that specific mutations in this region allows R3a to respond to both AVR3a^{KI} and the normally avirulent AVR3a^{EM} allele[7, 8]. Cell localisation experiments suggest that recognition of AVR3a^{KI} may lead to re-localisation of R3a to endosomal compartments, contrasting with the re-localisation of the TNL RPS4 and CNL MLA10, both of which require nuclear localisation for an immune repsonse[150, 151]. Interestingly, in this instance the autoactive R3a^{D501V} mutant that triggers rapid cell death in the absence of AVR3a does not show endosomal localization.

4.1.4 Using R3a to investigate R-protein signalling

The signalling events downstream of effector perception by R-proteins, particularly CNLs, are largely uncharacterised. There are a number of possible explanations for the apparent difficulties in elucidating signalling components, however in the case of R3a one of the most significant obstacles is that the addition of commonly used tags

to R3a results in a loss of AVR3a-dependent HR[152]. Although we cannot be certain that the reason for this inactivation is due to interference with required protein-protein interactions, it is possible that fusion of R3a to epitope tags results in steric interference with normal inter- and intra- molecular associations.

Overexpression of the coiled-coil domain of MLA10 leads to effector-independent cell death[153], suggesting that this region of the protein is responsible for interacting with downstream signalling partners to trigger HR. The coiled-coil domain of Rx has been shown to interact with the host protein RanGap2, which is required for effective immune responses to Potato virus X[88]. We hypothesised that the coiled-coil domain of R3a may therefore be involved in either triggering signalling cascades leading to cell death, or may associate with required host proteins, and so investigations into this domain may reveal novel R-protein interactors. In order to purify and detect plant-expressed R3a in an active state we generated R3a-specific antibodies raised against *E. coli*-derived $R3a^{CC}$ for immunoprecipitation (IP) experiments.

4.2 Results

Note: All western blot and pulldown experiments were performed with Marina Franceschetti. Figure 1.2.12 was produced by M. Franceschetti and N. benthamiana tissue expressing S. lycopersicum I-2 for figure 1.2.13 was kindly provided by A. Giannakopoulou.

4.2.1 Purification of R3a CC domain

Bacterial expression of the coiled-coil domain of R3a was performed as identified in Section 3.4.2. 6L of *E. coli* carrying the pOPIN-F:$R3a^{CC}$ vector were grown in LB at 37°C to an OD_{600} of 0.6, chilled on ice, and induced with 1mM IPTG. Following 16 hours of protein expression at 18°C, cells were harvested and lysed by sonication. Cell debris was removed by centrifugation, and the clarified lysate loaded onto a pre-equilibrated nickel affinity column. Bound proteins were eluted in a single step with 500mM imidazole and loaded onto a superdex S75 26/300 sepharose column. Fractions eluting from the column were assessed by SDS-PAGE to identify those containing the protein of interest with minimal contamination (Figure 4.2.1).

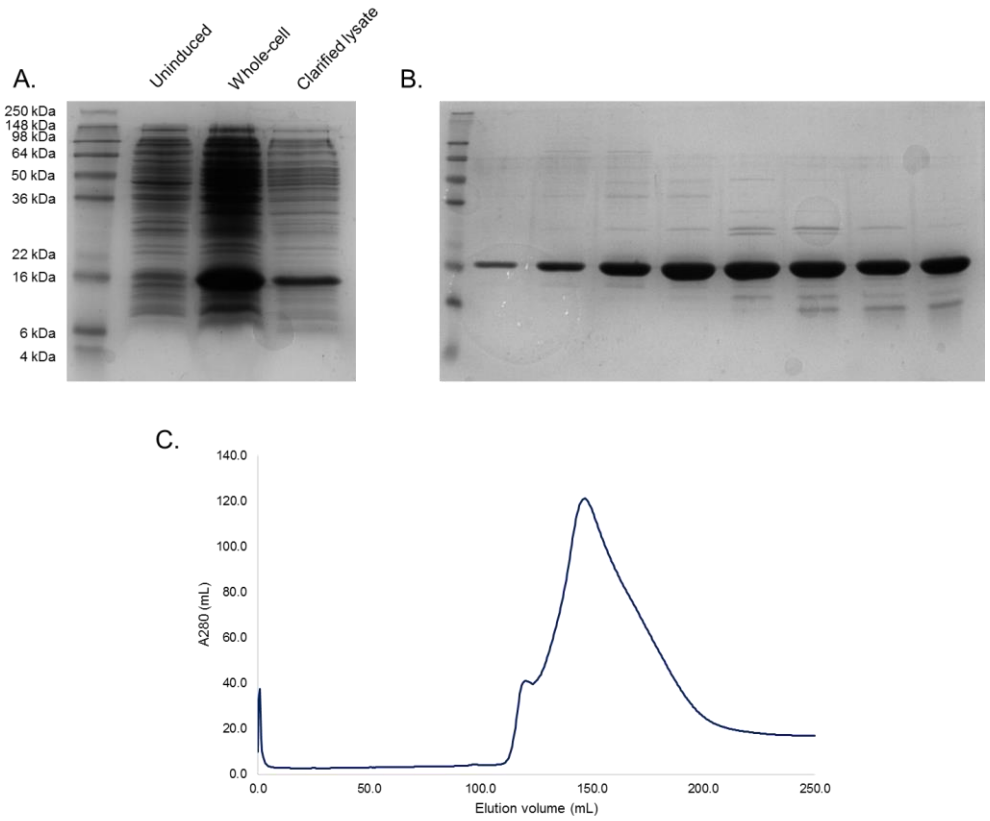


Figure 4.2.1 progression of R3a CC expression and initial purification

A) From left to right. Molecular weight marker, uninduced, whole cell and soluble protein samples of pOPIN-F:R3aCC *E. coli*. Induction was performed by adding 1mM IPTG to litre cultures at OD_{600} which had been chilled to below 25C.
B) SDS-PAGE of pOPIN-F:R3aCC soluble extract eluting from an S75 preparative gel filtration column. The first two lanes correspond to the void-volume of the column, indicating aggregation of the protein.
C) A_{280} trace of pOPIN-F:R3aCC preparative gel filtration. Protein eluted over a large volume, which may be indicative of multiple oligomeric states or protein-matrix interactions.

Histidine-tagged R3a coiled-coil domain eluted from the gel filtration column in a broad peak of over 50mL. Fractions containing the protein were pooled and concentrated before addition of 3C protease to remove the N-terminal histidine tag. Cleaved histidine tag and protease were removed by subtractive IMAC through a pre-equilibrated nickel affinity column. The flow-through and washes were concentrated to <5mL and loaded onto an S75 column for SEC. This second round of SEC resulted in a tighter elution peak, consisting of the R3a coiled-coil domain with few visible contaminants (Figure 4.2.2).

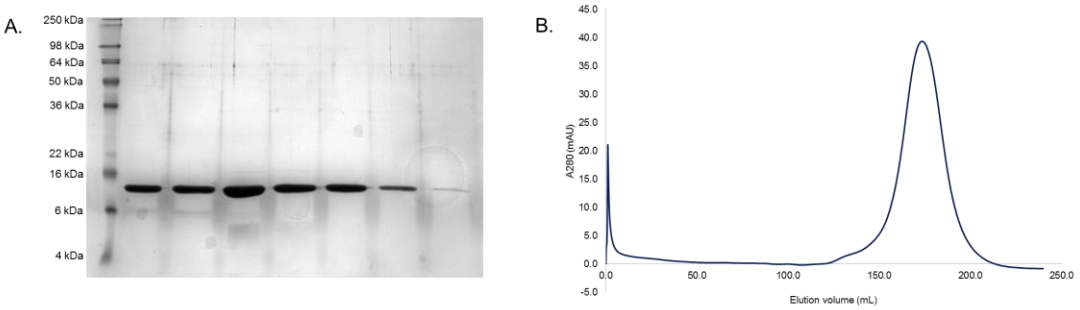


Figure 4.2.2 Final purification of R3a coiled-coil domain

A). SDS-PAGE of fractions from second round of preparative gel filtration following tag removal.

B). Gel filtration trace of R3aCC eluting from a superdex S75 column. Protein elutes much tighter after tag removal, peaking at ~160mL, with very little to no protein in the column void-volume.

Unlike the partial coiled-coil construct identified in Section 3.3.2.1, which differs from this by the omission of ~20 C-terminal residues, this purified protein showed no issues during concentration, and remained stable when stored at 4°C for over 4 weeks.

4.2.2 Generation of anti-R3a-CC antibodies

To allow monitoring of untagged, functional R3a we used the purified CC domain to generate R3a-specific antibodies. 4mg of purified R3a coiled-coil domain was sent to Eurogentec for production of α-R3a antibodies in rabbit. Two rabbits (referred to here as rabbit 1 and rabbit 2) were used for antibody generation, and serum samples used to assess background signal (Figure 4.2.3, left). Neither rabbit's pre-immune serum showed a signal at ~150kDa, which would mask signal for R3a detection, and so both underwent a three-month course of antibody production by repeated exposure to purified R3a CC domain. Following this course of antibody production we tested sera from both rabbits to assess both background signal and ability to detect R3a expressed under the CaMV 35S promoter in stable transgenic *N. benthamiana* (R3a transgenics, or R3aTG) (Figure 4.2.3, Right).

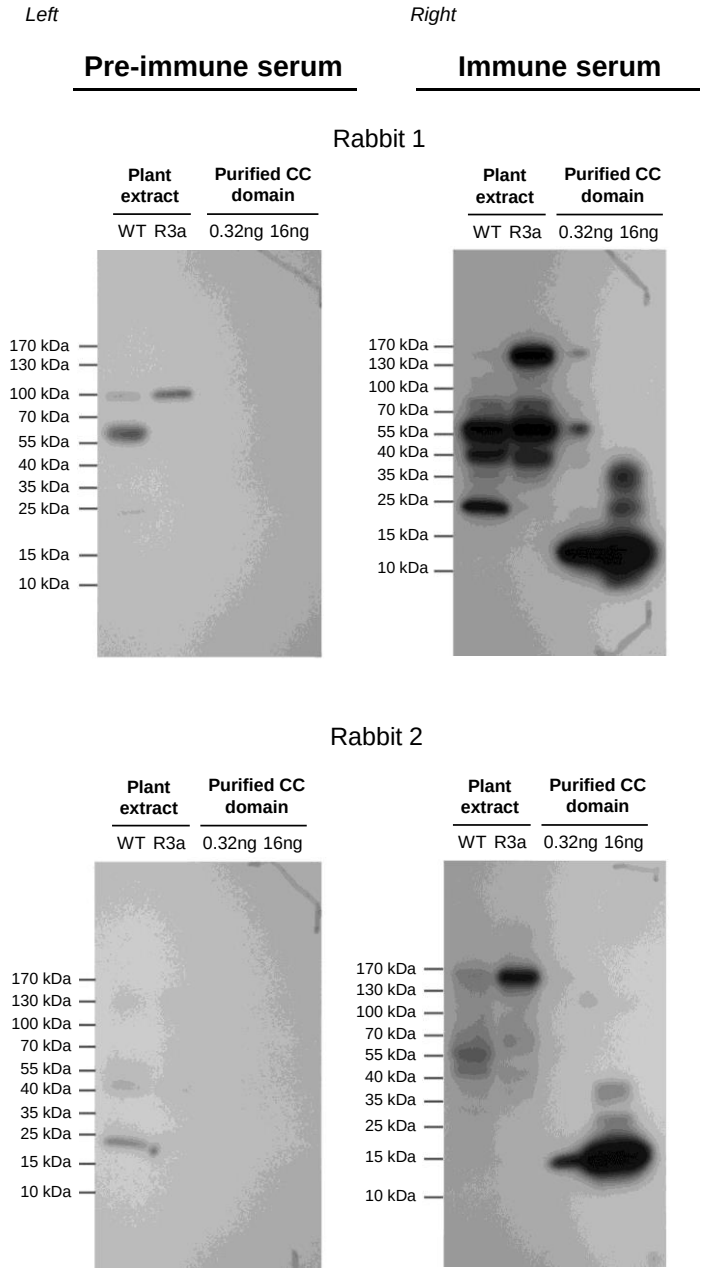


Figure 4.2.3 Pre- and post- immunisation serum from rabbit 1 and rabbit 2

Left. Pre-immune serum from rabbit 1 (top) and rabbit 2(bottom). Both rabbits show some degree non-specific binding to *N. benthamiana* extract prior to challenge with R3a^{CC} domain, however neither show a response to the purified protein.

Right. Following a course of R3a^{CC} immunisation both rabbits show strong signal for both R3a expressed *in planta* and the purified coiled-coil domain. Rabbit 2 appears to give lower overall background signal relative to R3a-specific signal, indicating fewer non-specific interactions.

Both sera showed a clear signal at ~150kDa against R3aTG extract, however there was also a signal detected that corresponds to a slightly higher molecular mass (~160-170kDa) when these antibodies were used against wild-type *N. benthamiana*

soluble extract, which was most noticeable for rabbit 2. A key aim of this project was to use these antibodies to immunoprecipitate R3a from *N. benthamiana* and identify proteins R3a interacts with in the absence of AVR3a. As the high background signal from Rabbit 1 serum would lead to an increase in false-positives, by identifying these off-target proteins and their potential interactors, we decided to focus on Rabbit 2 antibodies.

Although the total background signal from rabbit 2 was significantly lower than that of rabbit 1, we believed that the high-molecular weight non-specific signal was fainter in rabbit 1, which may make discriminating between R3a and this similar mass, non-specific band easier. We therefore had total IgG purified from both rabbits, which was stored in 10µL aliquots at -80°C until required.

4.2.3 Immunoprecipitation of untagged R3a to probe for interactors

Given the limited knowledge regarding R-protein interactions with host proteins, and the loss of HR in when R3a is tagged, we used the generated α-R3a antibodies to immunoprecipitate (IP) R3a from *N. benthamiana* tissue for IP-mass spectrometry experiments to identify interacting proteins.

4-5 week old *N. benthamiana* leaves, stably transformed to express untagged R3a under the CaMV35S promoter (R3aTG), were harvested, mid-veins removed, and frozen in liquid nitrogen. Tissue was ground with a pestle and mortar, pre-chilled with liquid nitrogen, and ground tissue transferred to chilled tubes, flash frozen, and stored at -80°C until use.

GTEN extraction buffer (10% glycerol, 25mM Tris pH 7.5, 1mM EDTA, 150mM NaCl, 10mM DTT, 0.1% TWEEN 20, 0.02% w/v poly(vinylpolypyrrolidone)) was added to tissue in a 2:1 v/w ratio on ice, and vortexed to mix. Samples were incubated on ice for 30 minutes with periodic vortexing. Cell debris was removed by several rounds of centrifugation in a bench-top centrifuge at 13,300*g* for 10 minutes, with supernatant transferred to fresh, pre-chilled tubes each round. Dynabeads® (ThermoFisher) sheep α-rabbit magnetic beads were prepared as described in Section 2.14.1. Beads were washed, re-suspended and 0.6µL rabbit α-R3a antibodies added. Beads were incubated with the antibodies on ice for 30 minutes. 30µL of prepared beads were added to clarified plant lysate and incubated at 4°C for 1-2 hours with constant rotation. Samples were placed in a magnetic rack to collect beads, and supernatant removed by gentle pipetting. An aliquot of supernatant was used to monitor unbound proteins. After washing, bound proteins

were eluted from magnetic beads by incubating beads with 1xSDS protein loading buffer (Section 2.8.5) at 72°C for 7 minutes with constant agitation. Input, unbound and eluted proteins were then separated on 4-20% polyacrylamide tris-glycine mini-PROTEAN™ gels (Bio-Rad). Separated proteins were then transferred to PVDF membranes overnight to be probed with α-R3a antibodies (Figure 4.2.4).

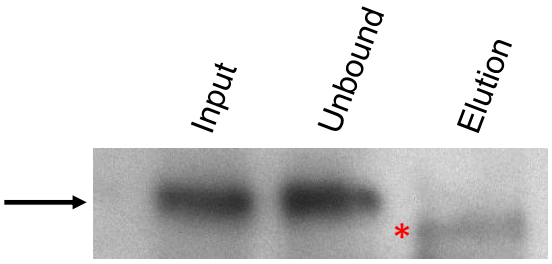


Figure 4.2.4 Immunoprecipitation of untagged R3a from stable transgenic N. benthamiana

R3a extracted from R3aTG tissue is clearly visible in the input lane (expected size indicated by arrow). Incubation with rabbit α-R3a antibodies and sheep α-rabbit magnetic Dynabeads® did not result in accumulation of the protein in eluate samples. Strong signal in unbound protein lanes indicates that the α-R3a antibodies are unable to immunoprecipitate the protein in these conditions. A non-specific band in the eluate is marked with by *.

As the volume used for elution was ~20 times less than the original volume of supernatant we would expect a significant concentration of R3a, and thus a much stronger signal in the elution lanes relative to the input. The inability to detect R3a in the eluate suggests inefficient pulldown by the α-R3a antibodies and magnetic beads, and so these conditions are not appropriate to investigate R3a interactions. We therefore performed a series of experiments to identify conditions more suitable for R3a immunoprecipitation.

4.2.4 Optimising R3a immunoprecipitation conditions

Immunoprecipitation of a protein relies on strong protein-protein interactions between antibodies and the target protein or tag. These interactions can be influenced by a number of factors, depending on the nature of the interaction interface. Charge-charge interactions are responsive to changes in salt concentration, which may interfere with salt-bridge formation. Similarly pH can affect side-chain charges and thus disrupt interaction interfaces. Although the original IP

conditions are used as they are close to physiological conditions, it is possible that these are sub-optimal conditions for IP of R3a. We therefore monitored the effect that changing these two factors had on R3a immunoprecipitation.

4.2.4.1 Optimising bead conjugation-salt vs pH

We focused our initial efforts on optimising antibody binding by screening for pH-dependent effects, and determining whether high salt concentrations were inhibiting R3a binding. Transgenic R3a *N. benthamiana* tissue was ground and divided into six 0.5g aliquots. To each aliquot 1mL GTEN buffer was added, containing either 150mM or 75mM NaCl, buffered to either pH 7.0, 7.5 or 8.0. Tissue was thoroughly vortexed and incubated on ice for 30 minutes. Unbound protein was removed by placing tubes on a magnetic rack and removing supernatant. After washing beads were incubated with 1xSDS loading buffer for seven minutes at 72°C to elute bound protein. Samples were then run on a 4-20% polyacrylamide gel, transferred to PVDF membranes and probed with α-R3a antibodies (Figure 4.2.5).

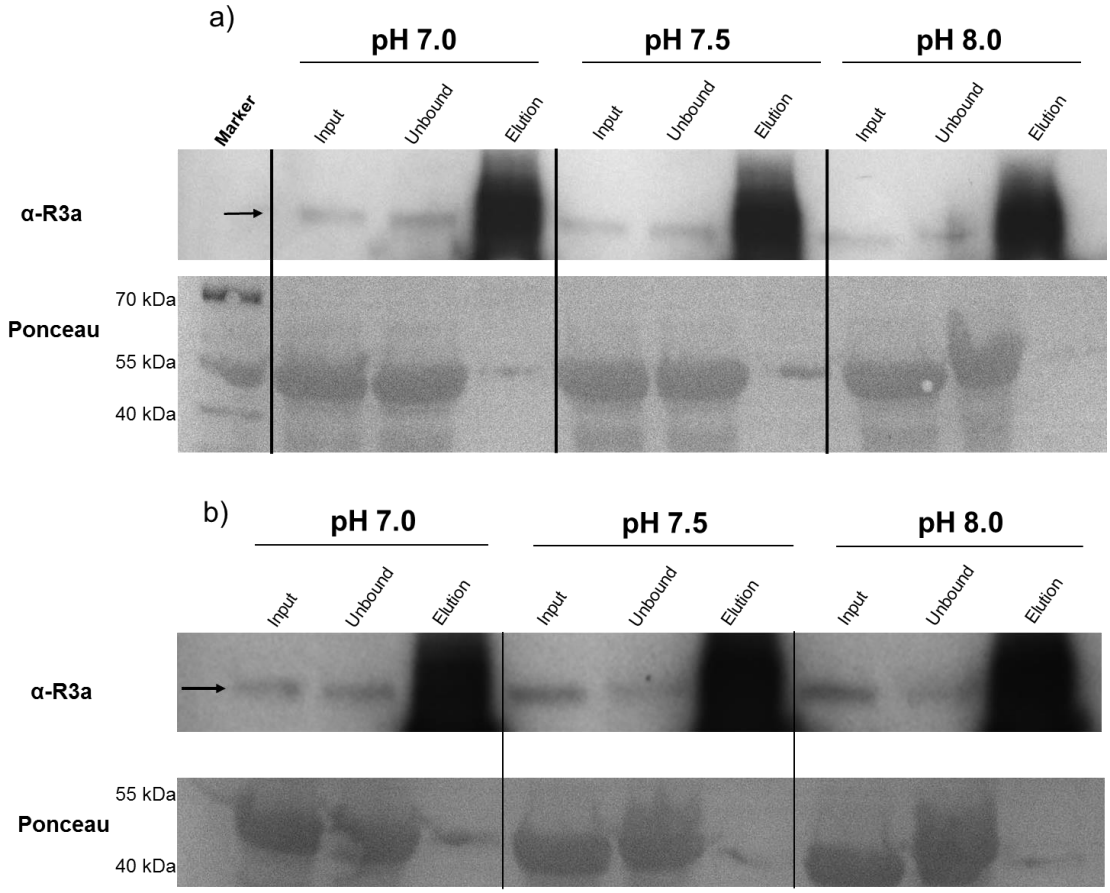


Figure 4.2.5 Western blot probing the effects of changing pH and salt concentration of R3a immunoprecipitation efficiency

A.) Protein extraction and immunoprecipitation experiments performed in high salt (150mM NaCl conditions) in extraction buffer at pH 7.0, 7.5 and 8.0. pH appears to have little effect on the efficiency of protein extraction as input signal intensity is roughly equal across all input lanes. High non-specific signal in elution lanes makes determining the presence of an R3a-specific signal difficult. Equal intensity of R3a-specific signal in both unbound protein lanes and input lanes indicates that almost all R3a remains unbound to α-R3a beads at each pH.
B.) Protein extraction and immunoprecipitation performed under low salt (75mM NaCl) conditions. High background from anti-body dissociation during protein elution results in high background.

Although the elution lanes for this experiment are predominated by non-R3a signal, making identification of individual bands difficult, the relative intensity of input to unbound sample lanes for all conditions indicates that immunoprecipiatation of R3a is not improved under these conditions.

The smearing seen in the elution lanes was most likely due to dissociation of antibody heavy chain during protein elution (Full blots can be seen in Appendix). We therefore investigated whether conjugation of α-R3a antibodies to the sheep α-rabbit antibodies of the beads would decrease background signal, and improve IP and detection of R3a.

4.2.4.2 Effect of bead conjugation on pulldown efficiency

Sheep α-rabbit magnetic Dynabeads were prepared as described in Section 2.14.1. Beads were washed and re-suspended in PBS with EDTA. 1.2µL α-R3a antibodies from rabbit 2 were added to beads, and incubated on ice for 15-30 minutes. Half of the beads were removed and added to a fresh Eppendorf on ice to monitor immunoprecipitation in the absence of chemical cross-linking to the Dynabeads (referred to throughout as unconjugated beads). The other half were chemically conjugated to the α-R3a antibodies by treatment with BS^3 (bis(sulfosuccinimidyl)suberate) as described in Section 2.14.1. Both sets of beads were then washed in 400µL Tris and split into a total of four eppendorf tubes: WT tissue + conjugated beads, R3aTG + conjugated beads, WT tissue + unconjugated beads, and R3TG + unconjugated beads. Total soluble protein from 1g of both wild-type and R3aTG *N. benthamiana* was prepared as described in Section (2.8.4). A 12µL aliquot was removed from each tissue extract and boiled with 4xSDS loading buffer as input controls. 800µL of R3aTG soluble extract was then added to the corresponding α-R3a bead sample. These mixtures were incubated for 1-2 hours at

4°C with constant rotation. Following incubation beads were separated using a magnetic rack and the supernatant removed- a 12µL sample of this was boiled in 4xSDS-loading buffer to monitor the amount of unbound protein. After washing, bound protein was eluted in citrate buffer pH 2.0 and 8µL of eluate boiled in SDS-loading buffer. Samples were then run on a 4-20% polyacrylamide tris-glycine gel and protein transferred to a PVDF membrane overnight. This membrane was probed with α-R3a antibodies to determine what effect conjugation had on the efficiency of immunoprecipitating R3a (Figure 4.2.6).

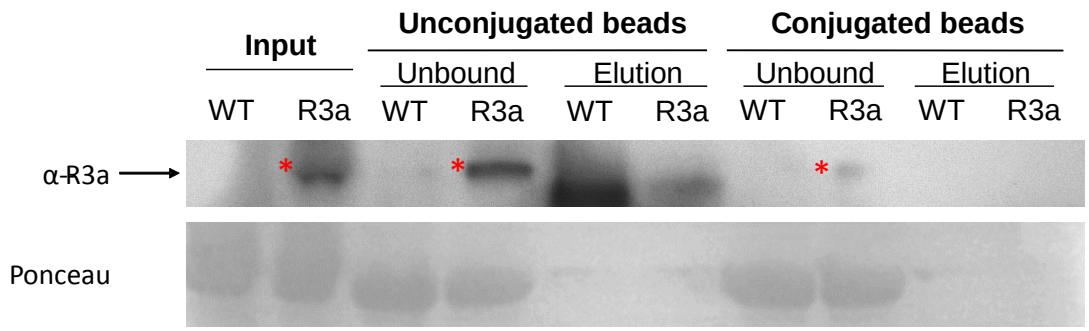


Figure 4.2.6 Comparing the effects of antibody conjugation to magnetic beads on the ability to immunoprecipitate R3a.

R3a signal comparing effect of conjugating α-R3a antibodies to magnetic beads on the ability to immunoprecipitate the protein from stable transgenic *N. benthamiana*. Top panel shows α-R3a probing, with R3a signal indicated by *. R3a-specific signal in the input loading, with an equivalent band in the unbound protein sample when antibodies are pre-conjugated to the beads. A faint α-R3a signal can be seen in the unbound protein fraction when antibodies are conjugated. Eluted protein samples do not show a specific R3a signal consistent with the mass from the input, despite the low signal in the conjugated bead unbound protein fraction. Bottom panel shows ponceau staining of the membrane of a ~50kDa band as a loading control.

R3a-specific signal is present in the input control, however no corresponding band is detectable in the elution lanes for either conjugated or unconjugated beads. When examining the unbound protein sample however, it is apparent that almost no R3a protein has been bound by the antibodies when left unconjugated to the beads. A band in the unbound sample lane corresponding to the R3a input signal with almost no change in intensity relative to the input control indicates equal amounts of protein in both the input and unbound samples. For the samples which had the α-R3a antibodies conjugated to the beads prior to incubation, we can see a faint band in the unbound protein sample corresponding with the input R3a signal, with no signs of R3a in the eluate. Chemically conjugating α-R3a antibodies to these magnetic

beads did not appear to hinder immunoprecipitation, and appeared to decrease non-specific signal following elution.

We had prioritised rabbit 2 α-R3a antibodies as the lower background signal seen in Figure 4.2.3 would result in cleaner immunoprecipitations, and more accurate MS data. Having been unable to optimise conditions with this batch of polyclonal antibodies we explored the use of rabbit 1 α-R3a antibodies to IP R3a.

4.2.4.3 Testing antibody-specific effects

As we had been unable to produce clear enrichment of R3a using antibodies derived from rabbit 2, selected for the low overall background signal in serum tests, we looked for differences between the two sets of polyclonal antibodies to see whether the issues described above were antibody-specific.

As we were unsure how the antibodies from rabbit 1 would respond to conjugation we took a systematic approach comparing the effects of conjugation to magnetic beads for both antibodies. Samples were eluted by acid elution (Figure 4.2.7, A) and then by boiling the beads in SDS-loading buffer (Figure 4.2.7, B). Acid elution is a preferable method for removing bound proteins as it is less likely to strip the light-chain and heavy-chain of the antibodies, which increase the background signal from secondary antibodies for western blotting. We included a subsequent round of elution by boiling in SDS loading buffer to ensure low signal was not due to inefficient dissociation of bound proteins. Transgenic R3a tissue was prepared as described in Section 2.8.4 and supernatant separated into four aliquots, to which were added either magnetic beads conjugated to α-R3a antibodies from rabbit 1 or rabbit 2, or washed magnetic beads and 1μL of α-R3a antibodies from each rabbit.

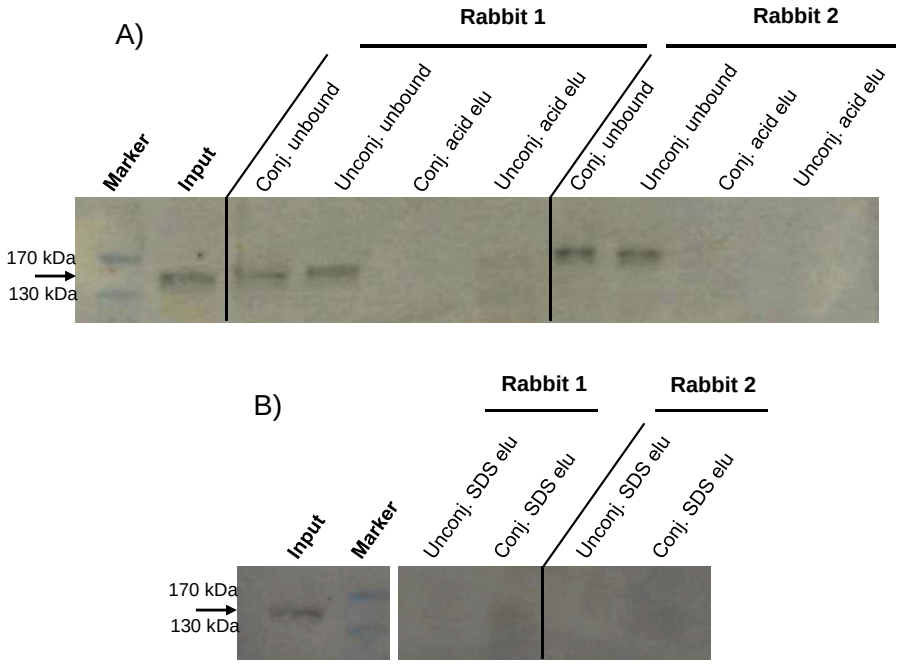


Figure 4.2.7 Western blot to assess ability of different polyclonal α-R3a antibody batches to immunoprecipitate R3a

Protein extracted from R3aTG plants added to rabbit 1 or rabbit 2 antibodies that were either conjugated to magnetic beads, or mixed with magnetic beads but without undergoing the chemical conjugation reaction. Proteins were separated by SDS-PAGE and probed using α-R3a antibodies, R3a specific band indicated by arrow.

A) Input control showing 15µL aliquot from the total protein extract, which was then divided and added to each of the four experimental condition (1mL each). Extract was then added to α-R3a antibodies either conjugated or left unconjugated with magnetic beads. Unbound protein samples represent the proteins present in supernatant after 2 hours incubation with α-R3a antibodies. Beads were washed and treated with 50µL citrate buffer pH 2.0 to elute bound proteins.

B) Proteins eluting from beads after boiling in 50µL 1xSDS-loading buffer. Following acid elution beads were treated with 1xSDS loading buffer and incubated at 72°C for 7 minutes to test whether absence of acid elution signal is not due to inefficient dissociation of R3a from the beads. Input control is the same as in a).

Both antibodies that were conjugated to magnetic beads, and those that were not were unable to show significant enrichment of R3a. The signal detected in the unbound samples was roughly equal to the intensity of the signal detected in the input lane- that is the aliquot taken from the total soluble protein from R3aTG tissue. There were no obvious antibody-dependent effects as both rabbit 1 and rabbit 2 derived antibodies showed s amounts of R3a in the unbound samples.

This presumed inability to bind to R3a is further confirmed by the absence of signal in eluate of either conjugated or unconjugated samples for both antibodies, despite

a predicted concentration effect of approximately 20 fold. The lack of signal in the eluate cannot be explained by poor dissociation by acid treatment, as additional elution performed by boiling beads in SDS-loading buffer did not lead to an increase in signal.

As we could not explain pulldown inefficiency due to antibody-specific effects we investigated whether the protein extraction and binding conditions may have a detrimental effect on R3a enrichment.

4.2.4.4 Protein extraction buffer

It is possible that a component of the buffers used to extract and bind R3a was interfering with R3a-antibody interactions, and so we tested whether a different protein extraction buffer would lead to improved R3a purification and enrichment. Transgenic R3a tissue was ground in liquid nitrogen and 0.5g transferred to four chilled 2mL screw-top tubes. 1mL GTEN extraction buffer (buffer 1) was added to two tubes (one with pre-conjugated beads, one with unconjugated beads incubated with α-R3a antibodies), and 1mL new extraction buffer adapted from Franceschetti *et al* (2011)[154] (50mM Tris pH 7.5, 5mM $MgCl_2$ 10% glycerol, 1% triton X-100, 1mM PMSF, 2% PIC) added to the other two tubes. Samples were vortexed and incubated on ice for 30 minutes to liberate soluble protein. Cell debris was removed by centrifugation and the supernatant added to tubes containing magnetic beads conjugated to α-R3a antibodies derived from either rabbit 1 or rabbit 2 (Figure 4.2.8).

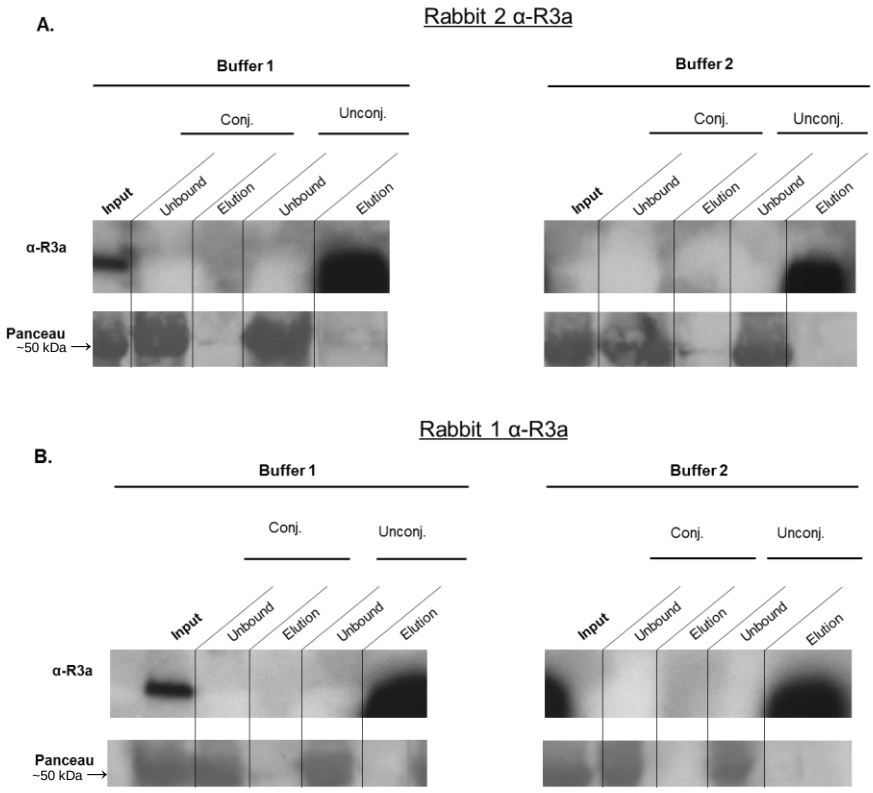


Figure 4.2.8 Effect of buffer composition on R3a extraction and immunoprecipitation

A. Comparison of two buffers on the ability to liberate soluble R3a and the effect of buffer composition on IP efficiency. Extracted soluble protein, shown in "input" lanes, was added to beads that had either previously been conjugated to α-R3a antibodies from rabbit 2, or washed and left unconjugated with antibodies added prior to sample mixing. Buffer 1 shows good R3a signal in the input, but no defined signal for eluting R3a either with conjugated or unconjugated beads. Buffer 2 does not seem to be able to liberate soluble R3a.
B. As in A., but with immunoprecipitation performed using α-R3a antibodies from rabbit 1. Both membranes were probed with antibodies from rabbit 2.

Buffer 1 and 2 were tested and run on the same blot. They have been separated here to aid interpretation. Set A. and B. represent separate blots.

Replacement of GTEN-based buffer with an alternative buffer resulted in a loss of R3a signal in sample input lanes, despite using the same tissue and equivalent total protein extraction as determined by ponceau staining. We concluded that replacement of GTEN buffer is unlikely to result in dramatic improvements in R3a extraction and immunoprecipitation. Finally we decided to test whether the inability to immunopreciptate the coiled-coil domain was an intrinsic property of these antibodies, rather than inappropriate experimental conditions.

4.2.4.5 Immunoprecipitation of purified coiled-coil domain

Having been unable to optimise pulldown conditions we proposed two hypothesise to explain the inability of the polyclonal antibodies to immunoprecipitate to R3a in these experiments: either the biology of native R3a does not permit binding, such as by burying the coiled-coil domain in a stable interaction, or these antibodies are for some technical reason unable to interact with both R3a, and the beads required to immunoprecipitate the protein.

To distinguish between these two scenarios we used α-R3a antibodies to pulldown R3a from transgenic tissue, heat-treated R3a tissue extract (to test whether low pulldown efficiency is due to specific native conformation), purified R3a coiled-coil domain (to test whether the conjugated beads are capable of immunoprecipitating isolated protein), and purified coiled-coil domain added to *N. benthamiana* leaf extract (to test whether the low immunoprecipitation efficacy is due to sequestration of the coiled-coil domain by interacting proteins).

R3aTG tissue was ground and soluble protein extracted as described above. Soluble extract was then divided into two aliquots, one of which underwent heat treatment by incubation at 95°C for seven minutes, before being cooled on ice. An aliquot of frozen R3a coiled-coil domain (Section 3.4.2) was thawed on ice and 10ng added to 1mL of GTEN extraction buffer. To test whether the coiled-coil domain was being sequestered by host proteins and thus unavailable for immunoprecipitation, 0.5g of wild-type *N. benthamiana* leaf tissue was ground in liquid nitrogen and added to a chilled 2mL screw-top tube with 1mL GTEN extraction buffer. Soluble protein was prepared in the same way as R3aTG tissue, and 10ng purified coiled-coil domain added to the wild-type extract, and extraction buffer only. 30µL of magnetic beads conjugated to α-R3a antibodies from rabbit 2 were added to each sample and incubated for 1-2 hours at 4°C with constant end-over-end rotation. Unbound protein was removed by separating beads with a magnetic rack and removing supernatant. Beads were then washed and eluted by the addition of 0.1M citrate buffer pH 2.0 incubated at room temperature for 10 minutes. Samples were boiled in SDS-loading buffer and loaded onto a 4-20% polyacrylamide gel (Figure 4.2.9).

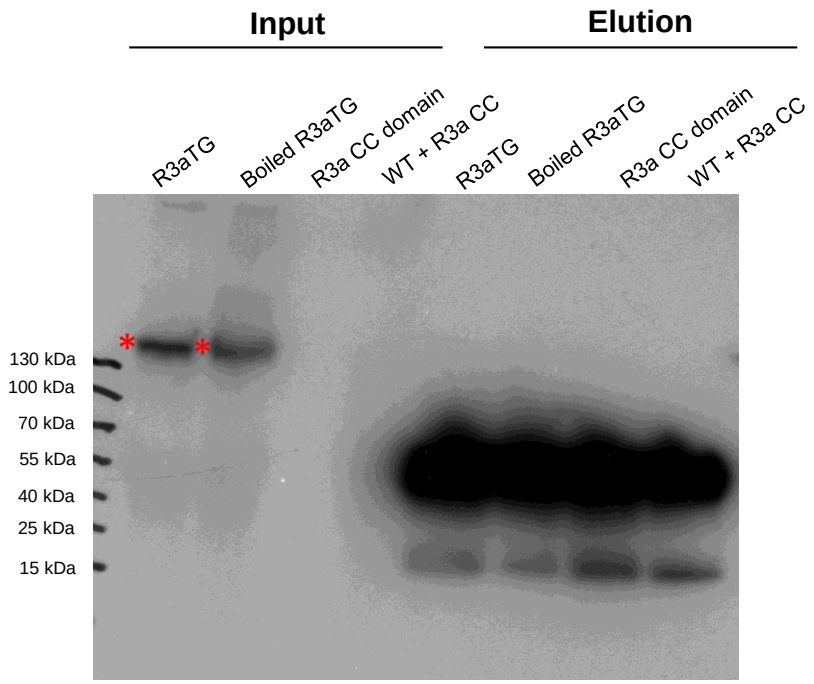


Figure 4.2.9 Immunoprecipitation of native, denatured transgenic R3a and purified R3a coiled-coil domain

Samples of soluble extract from transgenic R3a *N. benthamiana* tissue extract, the same extract after boiling to denature proteins, R3a CC domain purified from *E. coli*, and wild-type *N. benthamiana* extract with purified R3a CC domain. R3a signal is indicated by *. Although unable to detect the purified coiled-coil domain in the input, concentration by eluting in a smaller volume relative to input samples should allow this protein to be detected even with low-efficiency IP.

Despite the ability to detect R3a in the input for both native and heat-denatured R3a extract, we were unable to detect appropriate signal following immunoprecipitation. This suggests that an inability to IP the protein is not a result of native R3a conformations. There is no signal for purified coiled-coil following acid elution in ~40 fold smaller volume than the input, indicating that there has been no appreciable concentration of the protein in these conditions.

We have been unable to explain the lack of R3a-enrichment following immunoprecipitation by antibody batch-specific effects, effect of magnetic bead treatment, buffer components, interference by high salt concentration, or pH-dependent effects. This final experiment leads us to conclude that the inability to IP R3a is not due to the native conformation of the protein, indicated by the negative result with heat treated R3a extract. Nor is it due to protein-protein interactions in plant cell extract as we were unable to IP coiled-coil domain at >90% purity in the absence of host proteins. We concluded therefore that these antibodies are not

suitable for co-IP and mass-spectrometry experiments to determine interacting partners in this manner. However, as these antibodies are able to detect R3a, they can be used for targeted co-immunoprecipitation experiments.

4.2.5 Probing for R3a-AVR3a interactions

Multiple methods for AVR detection by R-proteins have been suggested using two key mechanisms: direct and indirect perception. Direct perception is the simplest mechanism to explain R-protein activation. The effector physically associates with an R-protein leading to activation. In this system R-protein binding of the effector allows nucleotide exchange, leading to conformational changes in the NBARC domain and release of the coiled-coil domain. The CC domain is then in a conformation that allows interaction with downstream signalling components to trigger a response. We therefore decided to use co-immunoprecipitation (co-IP) experiments test whether R3a perceives AVR3a by forming a stable complex.

FLAG-tagged AVR3a^{KI} allele was transiently co-expressed with untagged, and therefore functional R3a via agrobacterium-mediated transformation of wild-type *N. benthamiana*. Leaf samples were collected two days post-infiltration (prior to the onset of visible HR), frozen in liquid nitrogen and stored at -80°C. Soluble protein was extracted in GTEN buffer as described above. Clarified lysate was added to α-FLAG magnetic beads and incubated for two hours at 4°C. Following washing, bound proteins were eluted by heating samples to 72°C in 1xSDS buffer and separated by SDS-PAGE on a 4-20% polyacrylamide tris-glycine gel. Proteins were then transferred to PVDF membranes overnight, to be probed using appropriate antibodies (Figure 4.2.10).

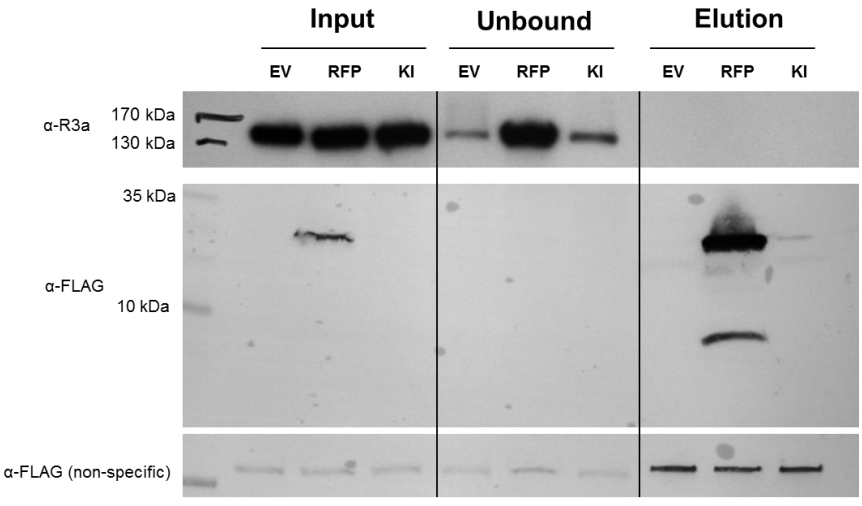


Figure 4.2.10 Western blot of AVR3a and R3a co-immunoprecipitation

Pulldown of FLAG-tagged proteins co-expressed with R3a in *N. benthamiana* does not show signs of complex formation. Top panel shows R3a present in all samples, with a large amount of R3a in the unbound protein fraction. No signs of R3a in the eluate is indicative the FLAG-tagged proteins are unable to co-IP R3a by forming a stable complex. Middle panel shows α-FLAG signal for these samples. Signal is only detectable for FLAG:RFP in either input or elution samples, suggesting an absence of FLAG:AVR3a^{KI}. Bottom panel shows non-specific FLAG signal as loading control.

Probing with α-FLAG antibody shows no signal for FLAG-tagged effector in either the input or elution lanes, despite FLAG-tagged RFP being clearly visible. We are unable therefore to determine whether complex formation occurs. We hypothesised that the most likely reason for the absence of the effector in these samples is due to its recognition by R3a leading to HR as the protein accumulates in the cytoplasm.

To overcome this issue we individually expressed the effector and R3a, extracted soluble protein, mixed the two extracts to allow complex formation, and then co-IP from this mixture. Leaf samples were taken from transgenic R3a plants and from wild-type *N. benthamiana* transiently transformed to express FLAG-tagged AVR3a^{KI}, AVR3a^{EM}, or RFP three days post infiltration. This tissue was flash frozen in liquid nitrogen and soluble protein extracted as described in Section 2.8.4. 1mL of R3aTG soluble extract was added to 1mL of each FLAG-tagged construct, and incubated on ice before the addition of α-FLAG magnetic beads. Beads were incubated in the plant extract for 1-2 hours at 4°C with constant rotation before being separated from the solution using a magnetic rack. Beads were then washed, and the bound proteins eluted by incubating beads in citrate buffer pH 2.0 for 5-10 minutes (Figure 4.2.11).

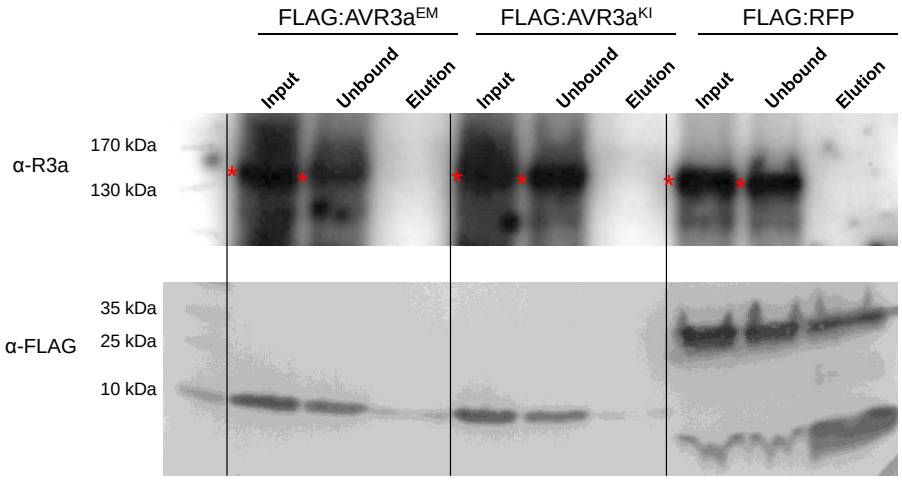


Figure 4.2.11 Western blot of R3a and FLAG-tagged AVR3a alleles when expressed individually and extracts mixed

Soluble protein extract from transgenic R3a *N. benthamiana* mixed with the soluble extract from *N. benthamiana* transiently transformed to express FLAG-tagged AVR3a^{KI}, AVR3a^{EM}, or RFP. FLAG-tagged proteins are detectable in the input and unbound lanes, indicating that these proteins are unable to form a stable complex with R3a under these conditions

We were unable to detect any signs of co-immunoprecipitation of R3a with FLAG-tagged effectors. Although α-FLAG signal is low in elution lanes, the complete absence of α-R3a signal despite over-exposing the autoradiogram (as seen by background spots between 170- and 130kDa markers) leads us to suppose that under these conditions the two proteins do not co-immunoprecipitate. As we have been unable to immunoprecipitate R3a, we are unable to test this by reciprocal co-IP experiments. Having been unable to use these antibodies to positively identify R3a-interacting proteins by immunoprecipitation, we examined the potential for alternate uses of these antibodies.

4.2.6 Using R3a antibodies to R-protein levels *in planta*

The use of these antibodies is not limited to identification of interacting partners, and can be used to assess R3a levels *in planta* and the stability of proteins in response to specific mutations.

Time course experiments showing the accumulation of R3a under the Rpi-vnt1 promoter were performed by Marina Franceschetti, and can be used to show that R3a expression is undetectable until two days post infiltration, with maximum levels reached approximately 5 days post infiltration (Figure 4.2.12). This work also demonstrates that under this promoter R3a levels remain significantly lower than R3aTG plants, where expression is driven by the CaMV 35S promoter. The significant increase in protein levels driven by the 35S promoter should be considered when testing recognition-induced cell death. High protein levels may lead to cell death responses that are not consistent with the behaviour of protein under native (or native-like) promoters which accumulate to lower levels.

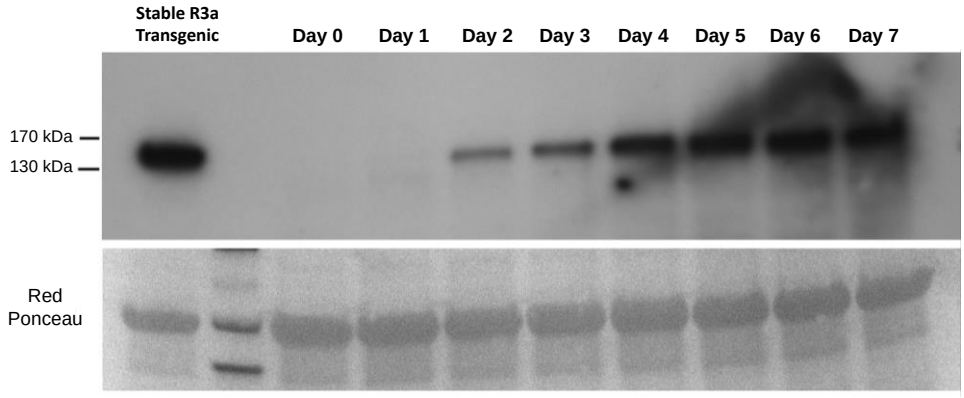


Figure 4.2.12 Time-course of R3a expression in N. benthamiana

Wild-type *N. benthamiana* transiently transformed by *A. tumefaciens* carrying the R3a gene in the pCBNptII vector under control by the Rpi-vnt1.1 promoter. Leaf tissue was harvested at time points indicated and stored at -80°C until use. 40µg of soluble protein was loaded to monitor accumulation of protein over time. 20µg of soluble protein extract from stable transgenic R3a *N. benthamiana* leaf tissue was loaded (far left lane) and allows comparison of relative protein levels. Courtesy of M. Franceschetti.

Recent work with R3a showing the possibility to expand responses to effectors has triggered an interest in whether these changes can be transferred to similar R-proteins. We tested whether these antibodies were also able to detect the homologous protein I-2 from *S. lycopersicum*, which is able to respond to the *Fusarium oxysporum* effector AVR2. Western blots of plant extract indicate that these antibodies are able to detect I-2 comparably to R3a (Figure 4.2.13). These antibodies have since been used to confirm that changes in responses of *S. lycopersicum* I-2 point mutants to different effectors are not due to significant changes in protein levels, but most likely reflect bona fide changes in protein behaviour[155].

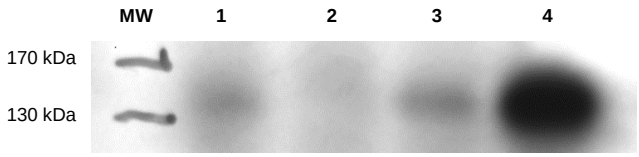


Figure 4.2.13 Detection of I-2 by α-R3a antibodies

Lane 1- *N. benthamiana* transiently expressing *S. lycopersicum* I-2 three days post infiltration, detected by anti-R3a antibodies. Lane 2- Empty vector control. Lane 3- *N. benthamiana* transiently expressing R3a. Lane 4- Stable transgenic *N. benthamiana* expressing R3a.

4.3 Discussion

At the point of writing, little is known about the signalling cascades involved in triggering R-protein mediated HR following effector detection, with traditional mutagenesis screens yielding little data. Alternative approaches, such as silencing screens used to identify NRC1 (Section 5.1), have been able to add to our knowledge of HR signalling pathways, indicating that alternative screening methods may make significant contributions to the field.

We decided to use direct methods to try and identify proteins involved in R3a-specific HR. Co-immunoprecipitation experiments (co-IP) combined with mass-spectrometry is a powerful way to identify protein-protein interactions, however in this approach has been of limited use for understanding R3a function. The reason for this is the observation that tagging R3a leads to inactivation and a loss of HR[152]. Although we cannot be certain that the reason for this inactivation is due to interference with signalling cascades, the lack of activity indicates that data collected in this method may not give a full description of R-protein behaviour *in planta*.

We decided to circumvent the issues of inactivation when R3a is tagged by producing specific antibodies against the R3a coiled-coil domain and to use these antibodies to investigate the interactions of untagged and therefore functional R3a for the first time. We were able to produce sufficient purified coiled-coil domain to generate antibodies, which are able to detect both full-length protein and the isolated coiled-coil domain by western blotting. Despite extensive efforts to optimise pulldown conditions, we have been unable to use these antibodies to efficiently immunoprecipitate R3a from plant extract. We hypothesised that there would be two key reasons for this issue- a biological reason, that the native coiled-coil domain is inaccessible *in planta* due to inter- or intra-molecular interactions, and a technical reason, that these antibodies are unable to simultaneously bind to and pulldown native R3a. Experiments using heat-treated R3a from plant leaves and soluble, purified coiled-coil domain expressed in *E. coli* indicate that the inability of these antibodies to immunoprecipitate these proteins is not due to unsuitable conformations or masking of interacting interfaces, but may be an intrinsic property of these antibodies.

Despite this limitation these antibodies may still be a valuable resource. Their ability to detect full-length I-2 *in planta* has been used to demonstrate that changes in behaviour of a series of point mutants are not related to issues with stability. Such

studies would typically rely on inactive, tagged, protein and thus only be a proxy measurement of functional protein behaviour.

Nor does the use of these antibodies to identify R3a (or I-2) interacting partners need to be restricted to immunoprecipitation experiments. Native PAGE is a method that allows the separation of protein complexes on the basis of mass and isoelectric point. It is possible that these antibodies could be used in native PAGE-mass spectrometry experiments to detect complexes containing R3a, and the components of such complexes. Alternatively these antibodies can be useful for validation of candidate interactors. Proteins such as AVR3a can be tagged without affecting their detection by R3a, and so it may be possible to tag other proteins predicted to be involved in R3a activation and signalling without loss of function. If such candidates can be identified then co-IP experiments using the tagged candidate and un-tagged R3a are a realistic way to validate interactions *in planta*.

5

Structural characterisation of NRC1 NBARC domain

5.1 Introduction

The modular nature of R-proteins is believed to allow multiple functions to be incorporated into a single protein. Essential functions for R-proteins are the ability to sense the presence of a pathogenic effector, the ability to transduce this signal to allow the plant cell to respond appropriately, and to remain in a signalling-competent, inactive state. This project focussed on Solanaceous coiled-coil R-proteins from various hosts- namely *S. tuberosum, S. lycopersicum, N. benthamiana, and S. bulbocastanum*. These proteins consist of an N-terminal coiled-coil domain (CC), a nucleotide-binding domain with similarity to Apaf-1, various R-proteins and Ced-4 (NBARC domain) and a C-terminal leucine-rich repeat region (LRR), each believed to be responsible for different aspects of R-protein function (Section 1.6).

In this generalised model of R-protein regulation the NBARC plays a pivotal role acting as a molecular switch, controlling whether the protein is in an active or inactive state based on the identity bound nucleotide. To date researchers investigating the mechanistic basis for mutant phenotypes in R-proteins in this domain have had to rely primarily on homology modelling. For R-proteins the two most likely candidate templates for homology modelling are Apaf-1 and CED-4. These two proteins share low sequence identity with many R-proteins, typically <25%, which reflects the evolutionary diversity of this protein family. Although confidence in homology modelling is improved with higher sequence identity, the environmental and functional constraints of these proteins is assumed to limit structural diversity for these proteins, allowing inference of mechanistic properties in the absence of structural studies using more similar proteins.

NRC1 was first identified in a silencing screen as an important down-stream component for multiple plant immune pathways[156]. Although positioned functionally downstream of canonical R-proteins, NRC1 shares a typical CC-NBARC-LRR domain arrangement, and has conserved motifs: notably Walker-A, Walker-B, and MHD (actually VHD in this protein) motifs. The behaviour of NRC1 reflects other CNL proteins *in planta*: aspartate to valine mutations in the MHD motif lead to autoactivation, and loss of the highly conserved Walker-A lysine leads to complete inactivation[157]. On the basis of shared architecture, and predictable responses to well-characterised mutations, we believe that the fundamental biochemical and structural properties of NRC1 will be applicable to other CNLs.

In this project we set out to use the unique tools developed in Chapter 3 to structurally characterise NRC1 NBARC. With previously characterised NBARC domains all originating from animal homologs (Apaf-1, CED-4, DARK, NLRC4) an understanding of a plant-derived NBARC domain can allow investigation of structural conservation from a more diverse context. Additionally an experimentally-derived structure of an NBARC domain from an R-protein may help better interpret the importance of conserved motifs, mutant phenotypes and act as a more appropriate template for homology models.

5.2 Results and discussion

5.2.1 Expression and purification of recombinant NRC1 NBARC domain in *E. coli* and Insect cells

Note: Cloning of NRC1 NBARC domain was performed by Richard Hughes. Insect cell expression was performed by Heater Rata and Joanne Nettleship

Heterologous expression of pOPIN-S3C:NRC1-NBARC was performed following the conditions identified in Section 3.5.3. *E. coli* carrying the recombinant plasmid were grown in PowerBroth media at 37°C with constant shaking until cultures reached an OD_{600} of approximately 0.6. Cultures were then rapidly chilled on ice and protein expression induced by addition of 1mM IPTG. Protein expression was allowed to continue for 16 hours at 18°C before the cells were harvested by centrifugation, re-suspended, and lysed by sonication. Insoluble cell lysates were removed by centrifugation and the soluble tagged-protein purified by immobilised metal affinity chromatography (IMAC) using a 5mL pre-equilibrated nickel affinity column. Further purification was performed by size-exclusion chromatography (SEC) following single-step elution of bound proteins from the IMAC column. The protein eluted from the column as a broad A_{280} peak, with a number of contaminants co-eluting, as determined by SDS-PAGE (Figure 5.2.1 B). Fractions predicted to contain the Sumo-tagged NBARC domain with minimal contaminants were pooled and concentrated to 10-20mL; tag removal was performed by incubation with 3C protease at 4°C overnight.

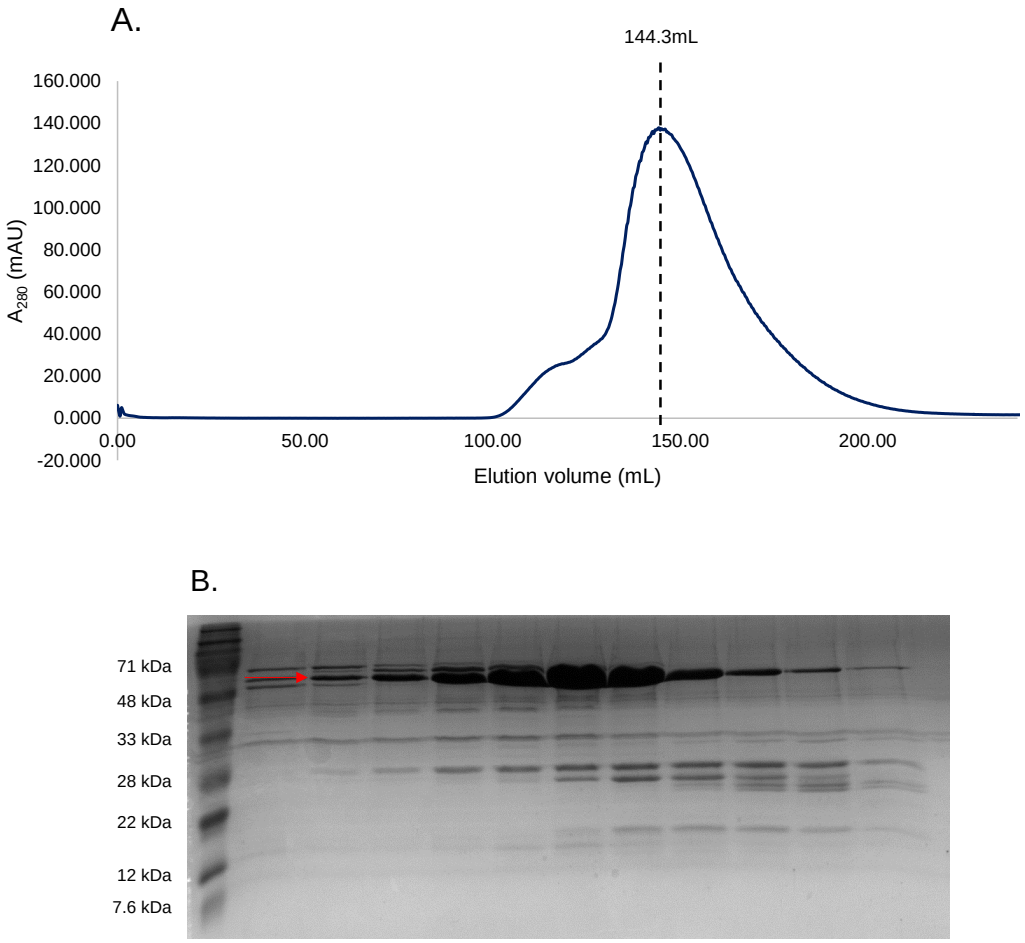


Figure 5.2.1 Purification of SUMO-tagged NRC1 NBARC

A.) Gel filtration trace showing protein eluting from superdex sepharose S75 column. Protein elution peaks at ~144mL with total protein elution spanning ~80mL. Void volume of the column is ~110mL. A minor elution peak (seen here as a shoulder) at a smaller volume indicates a fraction of the protein eluting as a large complex.

B.) SDS-PAGE of fractions eluting from A. Red arrow indicates SUMO-tagged NRC1 NBARC. Although NRC1 NBARC is the most abundant single protein in these samples, multiple bands indicate low purity and the need for further purification.

Cleaved protein was purified by subtractive IMAC, flow through and unbound washes collected and concentrated to <5mL before a final purification step and buffer exchange by SEC. The resulting peak was typically tighter than seen with SUMO-tagged protein, and interestingly either eluted at a similar or smaller volume as SUMO-tagged protein. Typically decreasing the molecular mass of a protein by ~20% (removal of a 13kDa tag from a ~50kDa fusion) would be expected to result in an observable increase in elution time, as the smaller molecule is able to explore more of the column matrix. SDS-PAGE was used to determine the fractions

containing NRC1 NBARC domain at over 90% purity, which were subsequently pooled and concentrated for downstream experiments (Figure 5.2.2).

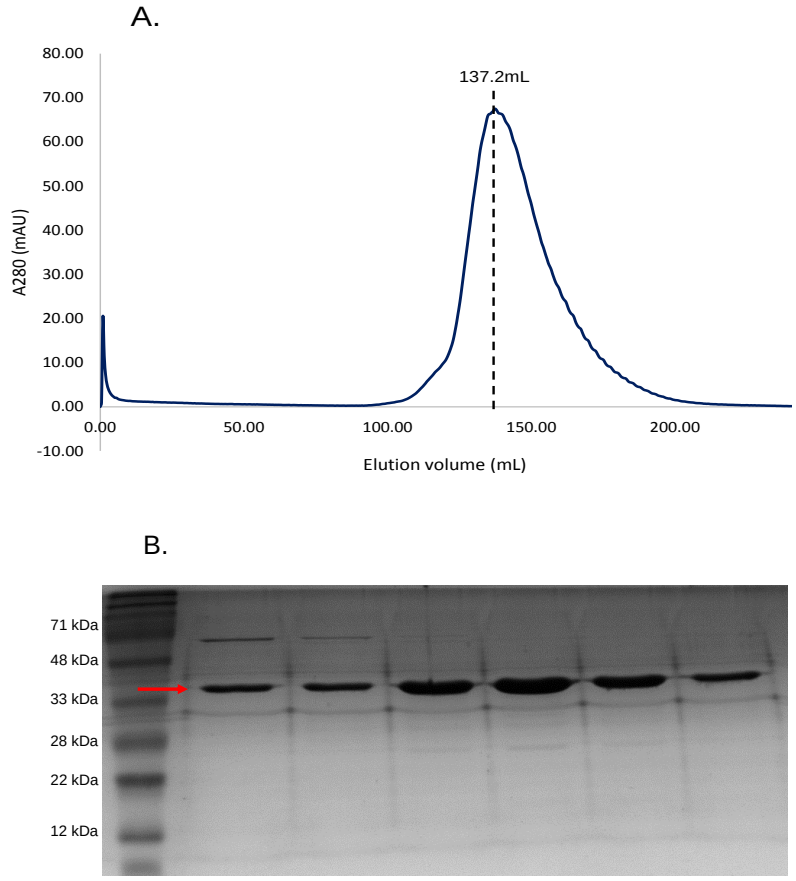


Figure 5.2.2 Final purificaton of the NRC1 NBARC domain

A.) Gel filtration trace for the purification of NRC1 NBARC domain following cleavage of N-terminal SUMO tag. Protein elutes clearly from the void-volume (110mL) in a single peak centred around 137mL.

B.) SDS-PAGE of fractions eluting from S75 column. Tag cleavage, IMAC and gel filtration results in high-purity protein. In this example the first two lanes are discarded due to the presence of a high-molecular weight contaminant.

Through these two rounds of purification we were able to produce reasonable yields (0.5-1mg protein per litre bacterial culture) of protein at high purity. The observation that tag cleavage often led to either no change in elution volume, or a decrease in elution volume, suggests that the protein may not be monomeric under these conditions, however these data are collected from preparative gel filtration and so this is explored further in Section 5.2.4.

Insect cell expression to obtain protein for crystallographic studies, and experiments described in Chapter 6, was performed as described in Section 2.8.2. Pelleted cells

were re-suspended in ~30mL of lysis buffer and lysed by cell disruption at 14 kpsi. Lysates were clarified by centrifugation and passed through a 0.45µM filter. Soluble protein was purified by IMAC followed by gel filtration. SDS-PAGE of fractions following SEC showed that fractions containing the target protein contained few contaminants (Figure 5.2.3).

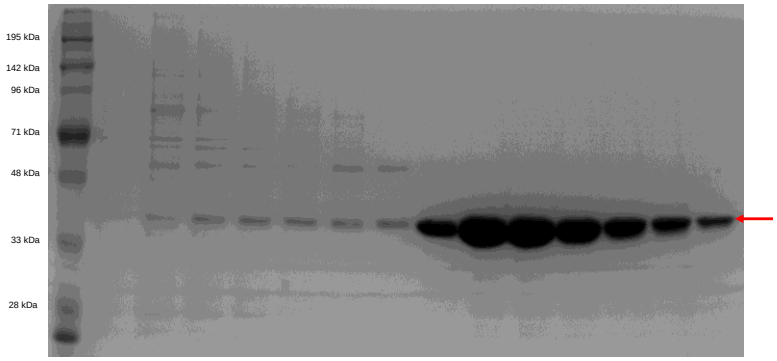


Figure 5.2.3 SDS-PAGE of His-tagged NRC1 NBARC domain purified from insect cells

4-20% Tris-glycine gel stained with coomassie shows few contaminants co-eluting with his-tagged NRC1 NBARC domain (indicated by red arrow).

After fractions were pooled and concentrated to <10mL the N-terminal tag was removed by incubation with 3C protease at 4°C overnight. Any residual tagged proteins, including the 3C protease, were removed by subtractive IMAC. Flow-through and washes were concentrated and purity confirmed by SDS-PAGE (Figure 5.2.4).

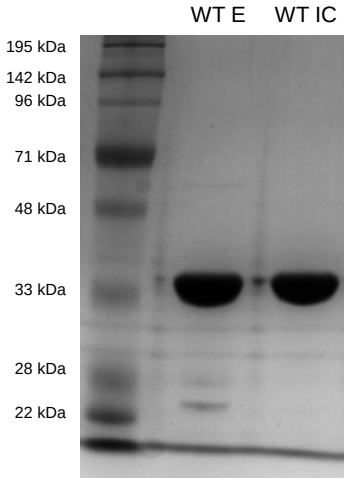


Figure 5.2.4 Assessment of wild-type protein purity from insect cells and E. coli

Wild-type protein derived from either *E. coli* (WT E) or insect cells (WT IC) separated on a 12% SDS- polyacrylamide gel, visualised by coomassie staining.

Despite undergoing only a single round of SEC, the purity of insect-cell derived wild-type protein shows a high level of purity, comparable to that of *E. coli* derived protein following two rounds of IMAC/SEC.

5.2.2 Assessment of folding and thermal stability by circular dichroism

Separation of the purified proteins by SEC indicated that the protein had a distinct elution volume, clearly distinguishable from the void-volume of the column. This is indicative that the protein does not form high-order oligomers or aggregates, and suggests that NRC1 NBARC is able to take on a stable fold when purified from heterologous over-expression in *E. coli*. Circular Dichroism is a method that records the spectral absorption of circularly polarized light as it passes through a protein solution. Specific wavelengths are differentially absorbed by protein secondary structure elements; alpha-helices show a peak in absorption at 208nm with a second peak at 222nm, whereas beta-strands typically absorb most strongly at 218nm and display a positive peak at 196nm. By comparing the transmission spectra of proteins of unknown fold with those of known proteins it is possible to calculate the secondary structure elements present in a protein. Secondly CD can be used to determine whether the introduction of specific point mutations has a significant impact on a protein's secondary structure through comparison with wild-type protein. Modern CD machines may also be able to measure and regulate sample temperature, making it possible to determine the thermal stability of proteins by measuring the loss of secondary structures over increasing temperatures.

NRC1 NBARC protein was prepared as described in section 5.2.1 and concentrated to 10mg/mL, as determined by A_{280}. As high salt concentrations and certain buffer components (such as HEPES) interfere with CD spectra, the protein was diluted in phosphate buffer to 0.1mg/mL to reduce these buffer effects. Spectra were collected from 180 to 260nm, with three measurements taken per wavelength per sample. Initial characterisation of the secondary structure elements were performed by measuring circular dichroism from 185 to 260nm at 20°C (Figure 5.2.5).

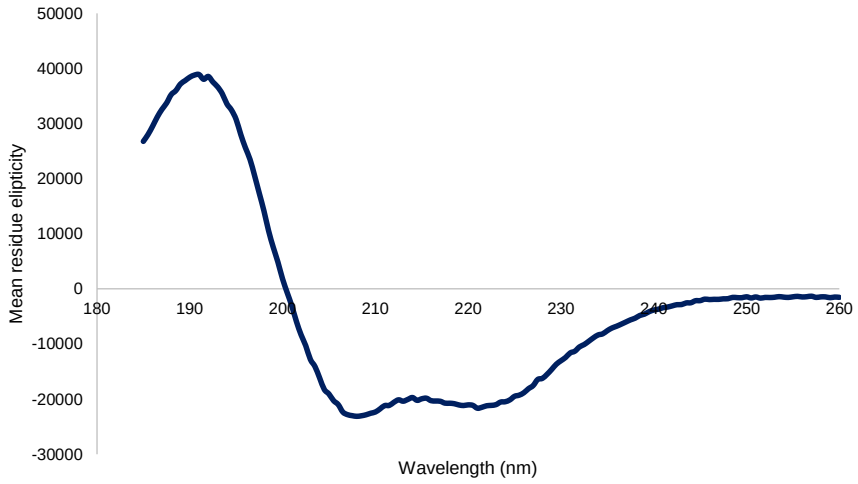
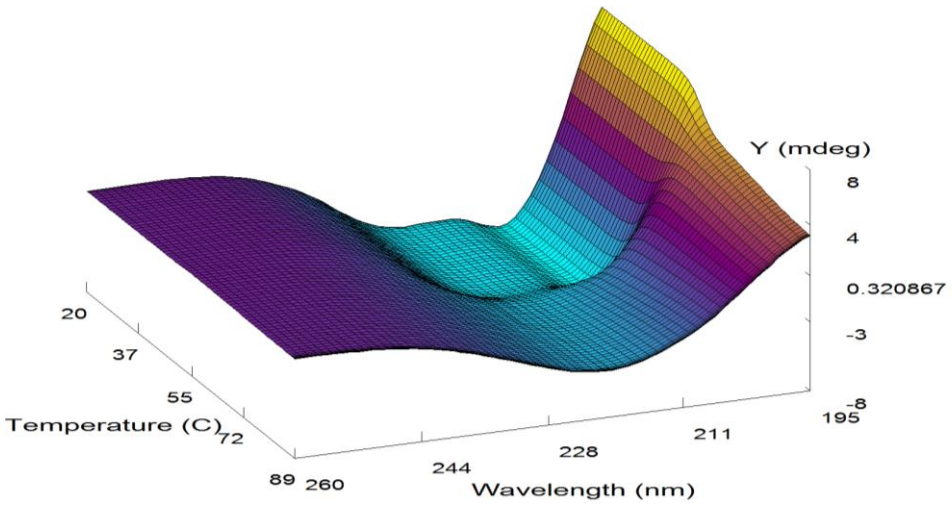


Figure 5.2.5 Circular dichroism spectrum of NRC1 NBARC

Average and buffer subtracted CD spectrum for NRC1 NBARC domain. Data were collected in 0.5nm steps in triplicate with buffer signal subtracted to give the final trace.

The curve produced shows a predominantly alpha-helix structure, as determined by the negative peaks at 222 and 208nm, with a positive peak at 196nm indicating the presence of β-strands. We also used CD spectroscopy to look at the effects of increasing temperature on the structure of WT NRC1 NBARC. Along with gaining information about the thermal tolerance of the protein (a useful proxy for stability), different transitions in unfolding can be used to infer structural information, such as whether there are defined substructures which may be more susceptible to unfolding. Thermal ramping experiments were performed as described above, with temperature increases of 1°C each minute from 20°C to 89°C. Analysis by Global 3™ (AppliedPhotosystems) of CD spectra identifies a single melting transition of approximately 54.5°C (Figure 5.2.6)

A.



B.

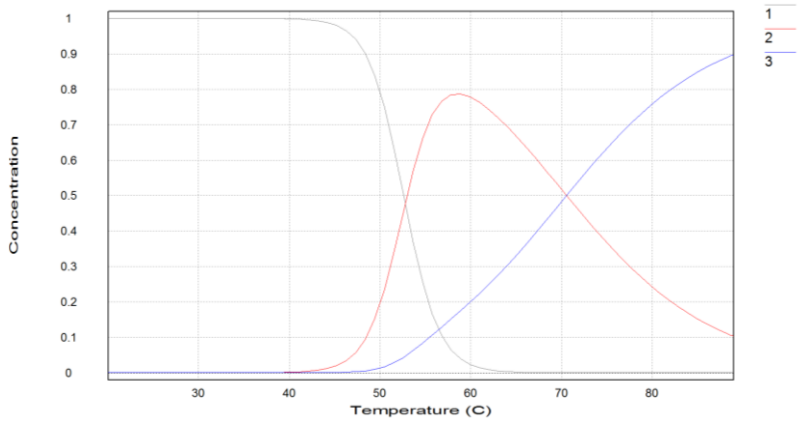


Figure 5.2.6 Thermal melt curves for NRC1 NBARC

A.) Three-dimension graph showing circular dichroism (measured in millidegrees, mdeg) plotted against wavelength (nm) and temperature (°C). Major deformation in the curve at approximately 50°C indicates loss of secondary structure and protein melting.

B.) Curve monitoring the relative concentration of species in solution against temperature. Trace 1 denotes folded NRC1 NBARC, with trace 3 representing melted protein. Trace 2 is indicative of an unfolding intermediate. Analysis of data indicates a melting temperature of 54°C for NRC1 NBARC.

With no examples of solubly expressed R-protein NBARC domains published to guide the design of well-behaved constructs, we used circular dichroism to determine that NRC1 NBARC expressed solubly in *E. coli* could produce expected secondary structures. Furthermore, thermal ramping CD experiments indicate that the protein has a reasonably high melting temperature, suggesting that the protein is stable in solution and suitable for downstream experiments.

5.2.3 Preliminary small-angle X-ray scattering experiments suggest a rigid, globular fold

SAXS uses the interactions between proteins in solution and X-rays to determine a number of protein characteristics, as such it can supplement other spectroscopic methods to assess protein folding and flexibility. Wild-type NRC1 NBARC domain was purified as described above and prepared at 10, 5, 2.5, 1.25, 0.625 and 0.3125 mg/mL for data collection at Diamond light source (Oxfordshire) beamline B21. Buffer samples were taken for subtraction against 18 experimental frames, with averaging and subtraction performed by Scatter. Data collected at 1.25mg/mL and 0.625mg/mL was analysed using the DANESSA automated data processing pipeline. Additional manual analysis was performed in Scatter. Manual data processing focussed on data collected at 1.25mg/mL (Figure 5.2.7).

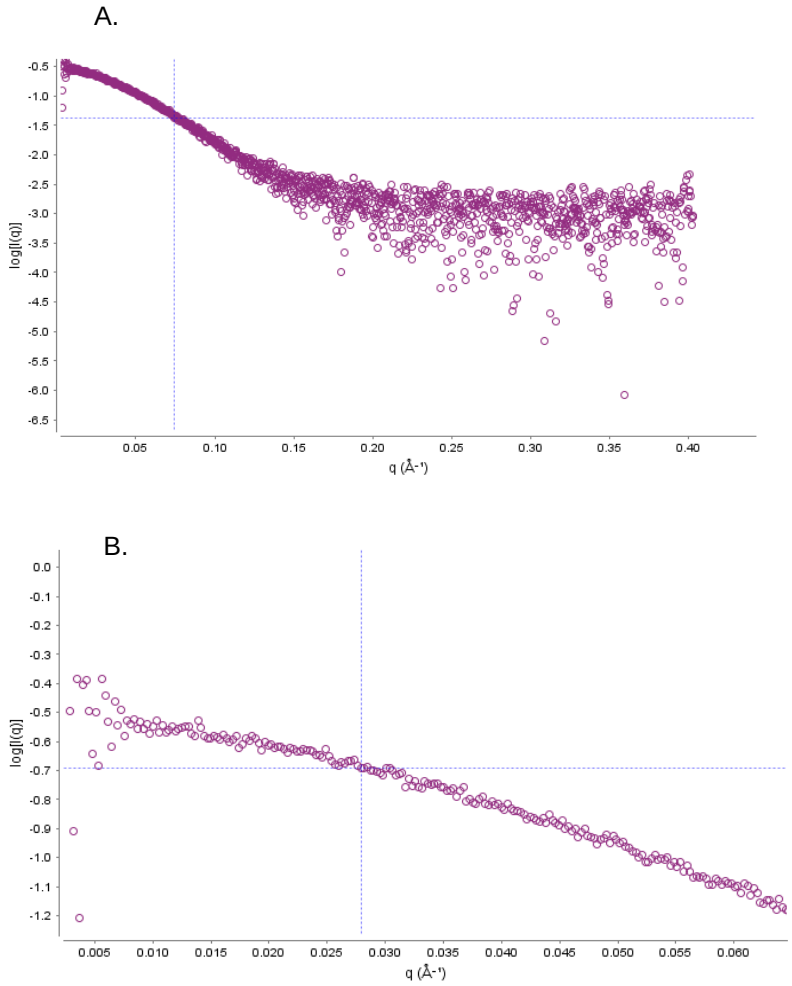


Figure 5.2.7 Scattering intensity-by-angle plots for NRC1 NBARC

A. Plot of all scattering data collected for NRC1 NBARC at 1.25mg/mL. Y-axis denotes scattering intensity (log scale) with x-axis showing angle of scattering
B. Close-up of scattering intensity at low scattering angles. Low scattering angles are more susceptible to effects of aggregation or inter-particle repulsion, which would be seen as dramatic changes in intensity over increasing scattering angles. The smooth progression seen here indicates that there are few inter-particle interactions at this concentration.

Data collected at 1.25mg/mL were used as the basis for these experiments as at this concentration showed no signs of inter-particle interactions, a fundamental requirement for SAXS experiments. Having established the suitability of this data for SAXS analysis we generated normalized Kratky plots to determine the folding and symmetry of the protein in solution (Figure 5.2.8).

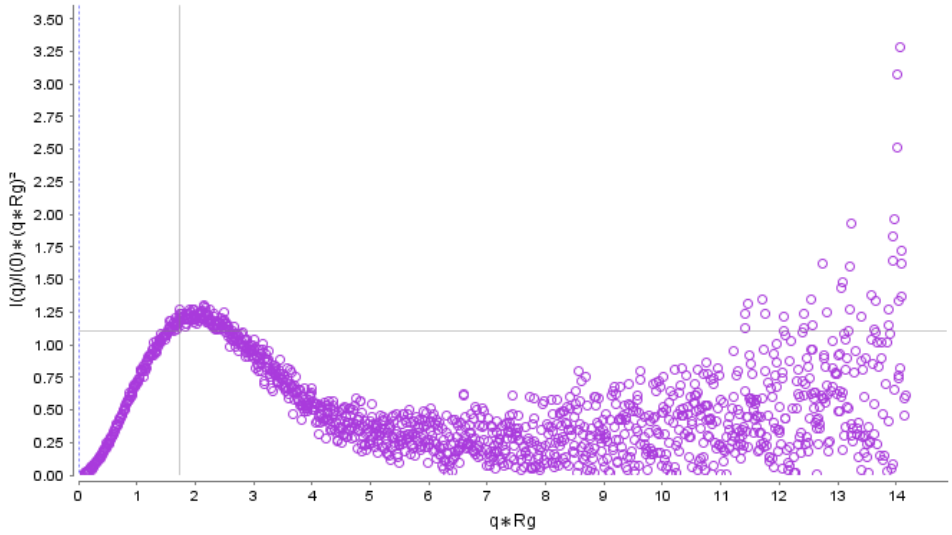


Figure 5.2.8 Normalized Kratky plot of NRC1 NBARC SAXS

The dimensionless (normalized) Kratky plot of NRC1 NBARC X-ray scattering. Perfectly symmetrical, species show a hyperbolic curve with a peak at √3. NRC1 NBARC shows the hyperbolic trace consistent with folded proteins and peaks slightly to the right of √3 indicating a slight deviation from absolute symmetry.

From the normalised Kratky plot we are to determine that the protein is folded in solution, in concordance with CD data indicating the presence of secondary structures (Section 5.2.2). A peak marginally to the right of √3 suggests that the protein is largely symmetrical in solution. Given the qualitative nature of this plot and the incomplete nature of the crystal structure of NRC1 it is not possible to determine whether conformational changes have occurred between solution and crystal structure.

Given our hypothesis regarding the histidine of the MHD motif acting as a locking mechanism (Section 7.4) we were interested to see whether NRC1 NBARC domain showed rigidity in solution, or whether there was the potential for spontaneous dissociation of the ARC2 domain to give a more open conformation, which we would predict would lead to spontaneous activation *in planta*. One way to explore this for isolated proteins in solution is to look for the presence of a Porod plateau when data is presented in a Porod-Debye plot (Figure 5.2.9).

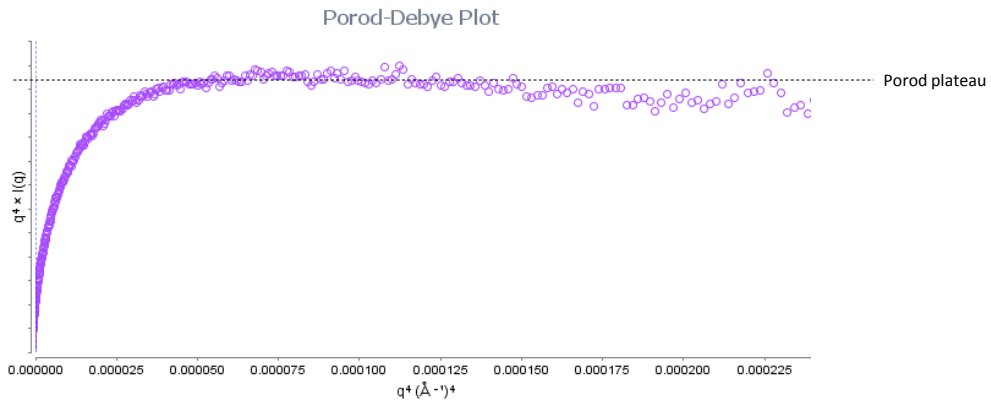


Figure 5.2.9 Porod-Debye plot of NRC1 NBARC SAXS

Plotting SAXS data collected for NRC1 NBARC domain at 1.25mg/mL transformed as $q^4 \cdot I(q)$ vs q^4 (where q is scattering angle in A^{-1} and I(q) is the intensity of scattering at q) allows identification of Porod-plateau for compact proteins. Proteins will only show a Porod-plateau in this plot if they fulfil the assumptions of defined homogenous contrast in electron density between particle and solvent.

Plateauing of the Porod-Debye plot indicates that there exists a sharp, well-defined difference in electron density between NRC1 NBARC and solution. This in turn is a qualitative measure of how compact a protein is; flexible proteins show less discrete contrast in electron density and so cannot fulfil the assumptions of the Porod-Debye fourth power law to give a plateau when scattering data are transformed in this manner. Instead, more flexible proteins show a continued increase in q^4 x I(q) over increasing q^4.

Preliminary SAXS data analysis presented here support the conclusions drawn from preparative gel filtration and circular dichroism experiments. The NBARC domain produced from *E. coli* appears to be well-folded, stable in solution and therefore suitable for further characterisation.

5.2.4 Investigating NRC1 NBARC oligomeric state *in vitro*

Removal of the SUMO tag during purification does not lead to an increase in elution volume of the NBARC domain, which would normally be expected when the molecular mass of the protein is decreased by ~13kDa. Dynamic light scattering (DLS), a method that can be used to approximate the size of a species in solution based on its ability to scatter light, indicated that purified NBARC formed species with an approximate molecular mass 81kDa in solution (Table 5.2.1).

Collection	Peak	R (nm)	MW (kDa)	% Mass
1	1	3.7	72	97.2
	2	16.2	2266	2.7
2	1	3.7	70	99.2
	2	20.3	3861	0.8
3	1	4.3	102	99.4
	2	27.6	7902	0.5
4	1	4	87	97
	2	17.5	2714	2.9

Table 5.2.1 Estimation of NRC1 NBARC particle size in solution

Dynamic light scattering experiments of NRC1 NBARC domain in 20mM HEPES pH 7.5 and 150mM NaCl shows that the majority of species in solution have an estimates molecular weight of 70-102kDa. Averaging data from the most biologically-relevant size range gives an approximate mass of 81kDa $\pm$ 15.5.

Although DLS can be a useful technique for investigating qualitative differences in protein behaviour (for example reduction of aggregation in different buffers), it is not sensitive enough to infer accurate molecular masses with confidence. In order to more accurately determine the size of the protein in solution, we employed analytical gel filtration (aGF) using a sepharose 10/300 GE column. Gel filtration is a method for separating out proteins based on molecular size and shape by the volume required for the protein of interest to pass through a column. In gel filtration small molecules are able to permeate the entire column, and are thus able to "explore" the inter-pore and inter-particle (in this case sepharose beads) volume, leading to a large elution volume. Large molecules are unable to penetrate the pores and so can only explore the inter-particle volume, thus eluting in a smaller

volume. By comparison against the elution of known molecular weight standards it is possible to calculate the molecular mass of an unknown species in solution.

NRC1 NBARC consistently eluted with a single peak at 10.3mL, corresponding to a species with approximate molecular mass of 63kDa (Figure 5.2.10), giving a mass intermediate of expect monomer (~40kDa) and dimer (~80kDa).

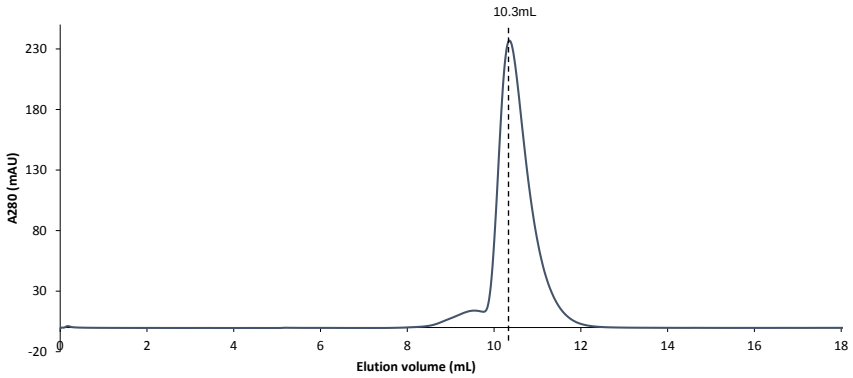


Figure 5.2.10 Analytical gel filtration for wild-type NRC1 NBARC

Analytical gel filtration of wild-type NRC1 NBARC domain shows a major elution peak at 10.3mL. A second minor elution peak is identifiable at 9.546mL, corresponding to molecular weights of 63 and 90kDa respectively

Analytical gel filtration requires two main criteria for accurate mass determination- that the protein be approximately globular, and that the protein does not interact with the column matrix. Experiments with NRC1 give somewhat ambiguous results as neither identifiable peak aligns well to an expected molecular weight for NRC1 NBARC. It is possible that these two peaks correspond to NRC1 NBARC as a monomer (minor peak) and dimer (major peak) or dimer and trimer, with delayed elution due to protein-matrix interactions. Data from DLS suggesting that the major species in solution is ~73kDa would favour the latter interpretation, however this is not unequivocal.

5.2.5 Crystallisation of NRC1 NBARC domain

5.2.5.1 Sparse-matrix screening and optimisation

Purified NRC1 NBARC domain was prepared as described in Section 5.2.1, with fractions showing >90% purity as determined by SDS-PAGE being pooled and concentrated to 10mg/mL measured by A_{280}. Commercial sparse-matrix crystallisation screens were used to identify appropriate conditions for the crystallisation of NRC1 NBARC. Seven different 96-well screens were used (JCSG, Structure screen 2, PACT, and Morpheus screens from Molecular dimensions,

Ammonium sulphate and PEG suite screens from Qiagen) to maximise the sampling of crystallisation space. Screens were set up in 96-well, sitting drop plates with two protein concentrations per reservoir (10mg/mL and 5mg/mL) and wells dispensed using an OryxNano (Douglas instruments) micro-volume liquid handling robot. From these initial screens several conditions capable of initialising crystal growth were identified (e.g. Figure 5.2.11).

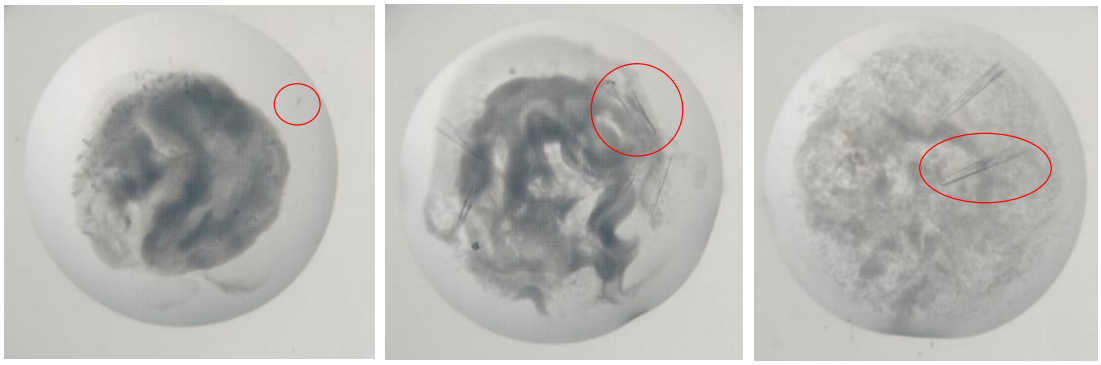


Figure 5.2.11 PACT screen crystal hits

Crystals formed from 96-well, sitting drop vapour diffusion experiments. Protein was mixed 1:1 with reservoir conditions to a total volume of 0.6uL. Drops show crystal growth in 0.1M MES buffer pH 5 and 25% PEG 6000 supplemented with (from left to right) 0.1M NH_4Cl, 0.1M LiCl, 0.1M NaCl. (Note, left and centre drops formed with initial protein concentration 10mg/mL, right drop used initial protein concentration of 5mg/mL). Red ovals highlight crystals

Many crystals from these initial sparse-matrix screens were too small, showed defects such as splits, or were surrounded by heavy precipitation, and so optimisation screens were performed to improve these crystals. Optimisations typically take the conditions identified in sparse-matrix screens and systematically vary conditions in an attempt to produce larger or better formed crystals. Optimisation screens performed here were typically in the form of precipitant by pH matrices, however changes in ratio of reservoir condition-to-protein volumes in the drops, and varying initial protein concentration were also investigated

The crystals pictured in Figure 5.2.11 were not reproducible in optimisation screens, and so alternative conditions were identified from the PACT screen, with 25% PEG 1500, MMT buffer pH 4.0, 5.0 or 6.0. Precipitant vs pH screens in 48-well, sitting drop screens showed that crystallisation could occur at much lower PEG and protein concentrations. For native NBARC domain an optimal pH of 4.5-4.6 was identified for these conditions, with lower pH leading to rapid and uncontrolled

precipitation and higher pH leading to clear drops. Interestingly, despite thorough screening it was not possible to produce NRC1 NBARC crystals without considerable precipitate also forming.

We were not certain that existing crystal structures would be able to be used for molecular replacement (MR) for solving the phases for the crystal structure of NRC1 NBARC, and so prioritised production of protein incorporating the heavy atom (HA) selenium (as selenomethioine, SeMet) for crystallisation. Introduction of heavy atoms into crystals either incorporated into the protein during synthesis, or by soaking crystals in solutions with high concentrations of heavy atoms such as bromide or iodide allows us to approximate phases using two methods. Either comparisons can be made between the intensity of diffracted spots from HA-derived crystals and native crystals (isomorphous replacement), or by using X-rays with a wavelength corresponding to the absorption maximum of inner-shell electrons in the heavy atoms (SAD). Structure solution by SAD relies on the violation of Friedl's law: that the intensities of symmetry-related reflection are equal, and can be used to locate anomalous scatterers (heavy atoms) within the unit cell. These experimental methods were preferable not only because we were unsure about the suitability of existing structures for molecular replacement, but also as they are less-susceptible to bias than MR.

SeMet NRC1 NBARC was produced as described in Section 2.8.2 and crystallised following the same method as above. The best SeMet crystals were found to form in 0.1M MMT buffer pH 5.0 with 15% PEG 1500. Crystals were then washed in cryoprotectant and frozen in liquid nitrogen before being transported for X-ray diffraction experiments.

5.2.5.2 Screening for alternative crystal forms- seeding and additive screens

Due to difficulties completing the structure of NRC1 (see Section 5.2.11), crystallisation screens were set up to identify new crystal forms using two methods: additive screening and seeding. Additive screening typically attempts to improve crystals by the use of various ions and small molecules to affect existing crystallisation and packing. As the crystals grew most reliable and consistently at pH 4.5 in 8% PEG 1.5K, two 48-well plates were set up with 130µL of these conditions in each reservoir. To each reservoir 20µL of silver bullet additive screen (Hampton research corp.) were added and plates monitored for crystallisation. Drops were set up as for typical optimisations that had proven to be reproducible, 1µL of protein at 8mg/mL mixed with 1µL reservoir condition. Heavy precipitate was seen in all

conditions, and so the concentration of additive was decreased to a one in 15 dilution. However, this did not lead to any improvement to the rapid and uncontrolled precipitation in the experiment.

In micro-seeding experiments, fragments of existing crystals to allow crystal growth to occur in conditions that do not necessary promote crystal nucleation (Figure 5.2.12). We used seeding for two main experiments: improve existing crystals by iterative seeding, and using seed stocks in sparse-matrix screens to identify new crystal forms. To create a seed stock, crystals grown in 11% PEG 1500 at pH 4.5 were mechanically broken *in situ*, and then diluted in ~150µL of reservoir solution. Crystals were further fragmented by several short bursts of vortexing before being rapidly frozen in liquid nitrogen and stored at -80°C until required.

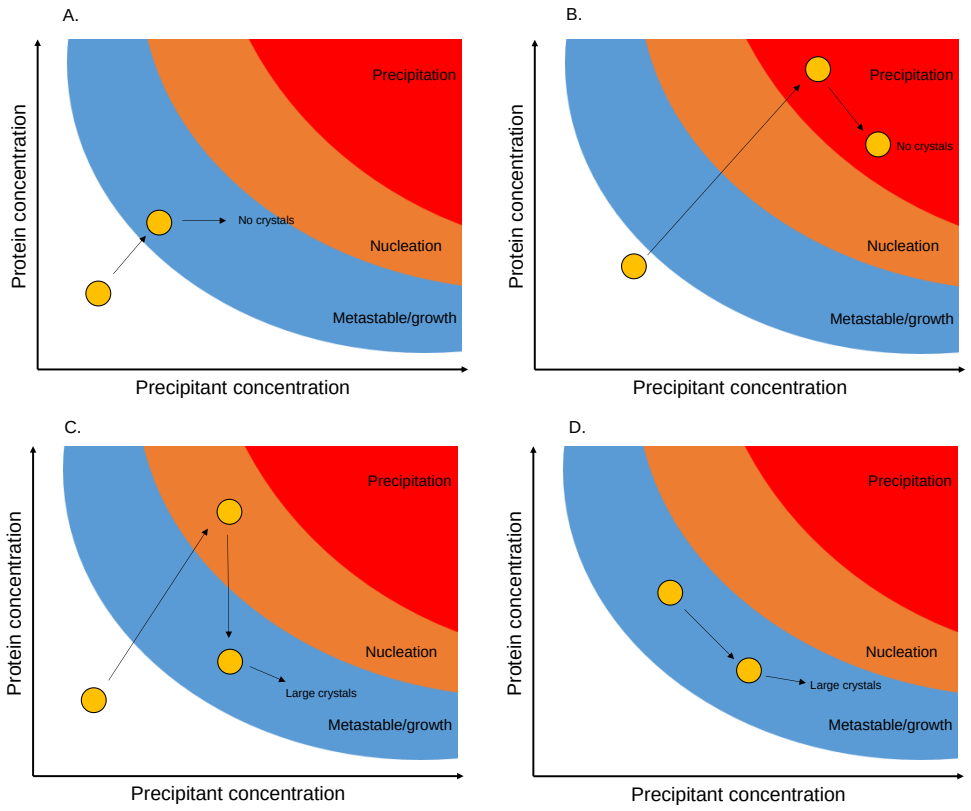


Figure 5.2.12 Schematic of crystallisation experiments

The outcome of protein crystallisation experiments is highly sensitive to (amongst other things) the concentration of protein and precipitant used. The majority of sparse-matrix screens will typically result in clear drops (scenario A.) or uncontrolled precipitation (scenario B.). In A. the concentration of protein or precipitant does not reach the threshold to initiate crystal formation. In B. initial protein or precipitant concentration results in uncontrolled precipitation of the protein before ordered nucleation sites can form. C. shows an ideal condition. The as the drop and reservoir equilibrate the protein and precipitant concentrations in the drop increases to the point where crystal nucleation sites for. As these nucleation sites form, the concentration of protein decreases until nucleation can no longer occur and the drop reaches a metastable condition suitable for growth of existing micro crystals. Small crystals such as seen in Figure 5.2.11 left-hand panel arise when metastable conditions are not reached before depletion of the protein and so significant crystal growth cannot occur. D. shows a micro-seeding experiment. Ground crystals act as nucleation sites, bypassing the need for the delicate balance between scenario A and B.

We hypothesised that protein precipitation seen in NRC1 NBARC crystallisations may be limiting the material available for crystal growth once nucleation has occurred. We therefore performed iterative seeding experiments, introducing nucleation sites before precipitation could occur. Although seeding was able to recapitulate and facilitate crystal growth in the existing pH 4.5, PEG 1.5K conditions, including reducing the minimum protein and PEG concentrations needed to form crystals, there remained a high level of precipitation in conditions supporting crystallisation. Iterative rounds of seeding, whereby the best crystals were used to produce fresh seed stock to re-seed around the same conditions, resulted in almost identical crystals, tapered or split rods with medium- to high levels of precipitation.

Alternatively sparse-matrix screening using seed stock identified multiple additional crystallisation conditions (Table 5.2.2).

None of the identified conditions yielded better looking crystals, or ones with an obviously different morphology. Crystals with different morphologies were desirable as they are less likely to share the same pathologies that may have affected electron density in existing crystals. Three examples of crystals grown in different conditions were flash-frozen and sent to the Diamond light source synchrotron for data collection, however no protein diffraction was seen for any of these new forms

pH	Precipitant	Concentration	Additive
5.5	PEG 3K	20.0%	-
5.5	PEG 10K	17.0%	0.1M Ammonium acetate
5.5	PEG 3350	25.0%	0.2M Ammonium sulphate
7	PEG 3350	15.0%	Succinic acid
6.5	P550MME_P20K	30.0%	Carboxylic acids
7.5	P550MME_P20K	30.0%	Amino acids
6.5	EDO_P8K	30.0%	Divalents
6.5	MPD_P1K_P3350	37.5%	Ethylene glycols
6.5	MPD_P1K_P3350	37.5%	Carboxylic acids
7.5	MPD_P1K_P3350	37.5%	Amino acids
6.5	P550MME_P20K	30.0%	Monosaccharides
5	PEG 1500	25.0%	MIB buffer
6	PEG 1500	25.0%	MIB buffer

Table 5.2.2 Conditions resulting in NRC1 crystal growth from seeding

Summary table for conditions giving rise to the most promising crystals (based on morphology) from seeding. PEG: polyethylene glycol. MPD_P1K_P3350: (RS)-2-methyl-2,4-pentanediol with PEG 1K and polyethylene glycol monomethyl ether 550. EDO_P8K: ethylene glycol and PEG 8K. P550MME_P20K: polyethylene glycol monomethyl ether 500 and PEG 20K.

5.2.5.3 Screening for improved crystals- alternative protein source

Having tried several methods to improve crystal packing and identify new crystal morphologies, we investigated whether protein derived from a different expression system may have subtle differences that may lead to changes in crystal packing. Wild-type NBARC domain was expressed in Sf9 insect cells by J. Nettleship and H. Rada at Oxford Protein Production Facility. Protein was purified from insect cells as

described in Section 2.8.2. Purified protein was then used in PACT 96-well screen to identify conditions suitable for crystallisation. Wild-type insect-cell derived protein (WTIC) crystallised in similar conditions to the *E. coli* derived protein, with a generally similar morphology: hexagonal rods, with some tapering and splits. On the whole, however, these crystals appeared better formed, with less precipitation, fewer splits, and a greater proportion growing to form large rods (Figure 5.2.13). The most promising crystals were identified in 0.1M HEPES pH 7.5 with 25% PEG 6K and 0.2M NaCl, and so these conditions were used for optimisations. Initially these crystals could not be reproduced, however using a seed stock from *E. coli* derived protein co-crystallised with ATP-γ-S (chosen due to the large number of small crystals with very little precipitation), resulted in recapitulation of crystal growth. Optimisation of seed stock dilutions (screening undiluted, one in 10 and one in 100 dilution in a small PEG-by-pH screen) was able to produce drops containing approximately a dozen, large and well-formed crystals when the most dilute seed was used. As these crystals appeared well formed and with less precipitate, co-crystallisations of WTIC NBARC with either ADP or ATP-γ-S using the same protein and seed stock were set up. Crystals were formed with both ADP and ATP-γ-S present, with the ATP analogue co-crystallisations resulting in smaller, rod-like clusters with some individuals for harvesting.

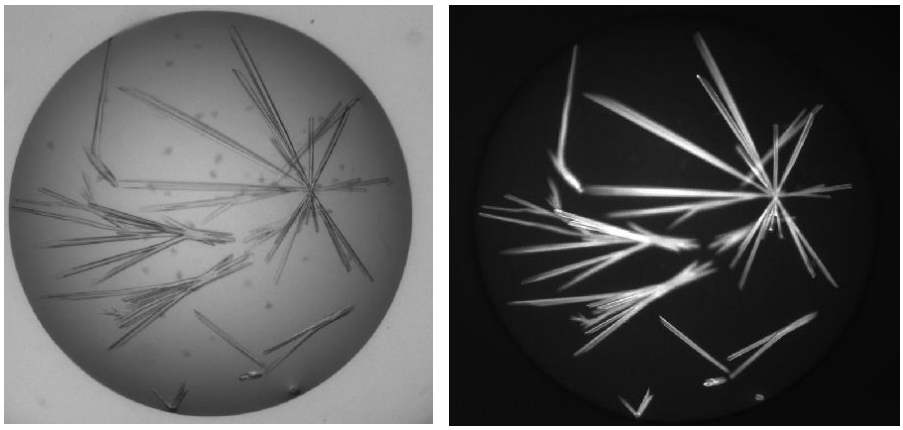


Figure 5.2.13 Crystals of NRC1 NBARC domain purified from insect cells

Wild-type NRC1 NBARC domain purified from insect cells crystallised from sparse-matrix screens at pH 7.0 in 20% PEG 6K with 0.2M NaCl. Left shows visible light image-heavy precipitation makes individual crystal hard to distinguish. Right image shows same drop imaged under UV light. Strong fluorescence of protein crystals allows easier identification of crystals, and discriminates between protein (fluorescent) and salt (non-fluorescent) crystals.

Crystals identified from this 96-well screen were typically larger than *E. coli* derived counterparts, with more examples of solid, well-formed crystals. Despite these general improvements, many crystals displayed splits or multiple crystals forming from a single point.

5.2.6 Data collection

Native and SeMet crystals were harvested, washed in cryo-protectant (as described in Section 2.16.5) and flash frozen in liquid nitrogen prior to shipping to the Diamond Light Source in Oxfordshire for data collection. For bromide and iodide SAD collections crystals were first soaked for 1 minute in a solution containing 0.5M potassium iodide or potassium bromide before washing in cryo-protectant also containing KI of KBr (Section 2.16.5). Data collection strategies were developed using EDNA following 3-5 test shots (Ω = 0.2°). Native and SeMet datasets were collected at λ= 0.9797Å to a resolution of 2.5Å and 3.2Å respectively. Iodide datasets were collected at λ= 2.0Å to a resolution of 3.24Å and bromide data sets collected at λ= 0.9200Å to a maximum resolution of 3.1Å.

1250 images were collected for native data sets with a rotation angle (Ω) of 0.15°. For selenomethionine data collections 3000 images were collected with Ω = 0.15°. Bromide and Iodide datasets consisted of 2150 and 2400 images respectively.

5.2.7 Data processing and structure solution

Data were processed using the XIA2 pipeline[158] with XDS[159] and Pointless[160, 161], integrating over all images in each dataset. Processed data used for initial structure solutions are presented in Table 5.2.3.

	NRC1 NBARC Selenium peak	NRC1 NBARC native
Data collection		
Instrumentation	DLS-i04	DLS-i02
Wavelength (Å)	0.9797	0.9795
Resolution range (Å)	53.42 - 3.22	55.26 - 2.5
Space group	$P6_4$	$P6_4$
Unit cell parameters	a = b =110.53Å c = 53.32Å α = β = 90° γ = 120°	a = b = 110.53Å c = 53.32Å α = β = 90° γ = 120°
Completeness (%)	100 (100)	100 (100)
Unique reflections	6159	13066
Redundancy	21.7 (23)	10.4 (10.3)
Anomalous redundancy	11.1 (11.6)	-
R_{merge}	0.092 (0.815)	0.051 (0.785)
I/σ(I)	21.2 (5.1)	26.7 (3.2)
Number SAD sites*	5	-

Table 5.2.3 Data collection statistics for NRC1 NBARC

Data used for experimental phasing and refinement following processing using the Xia2 automated pipeline. Data presented in parentheses indicate statistics for high-resolution shell. *number of SAD sites determined by PHENIX AutoSol.

Experimental phasing was performed using PHENIX Autosol[162] using single wavelength anomalous diffraction (SAD) to give a solution with a figure of merit (FOM), a measure of initial phase accuracy, of 0.346. FOM values range from 0 to 1 with scores above 0.45 classed as very good and between 0.25 and 0.45 as marginal.

Initial phases calculated to 3.2Å were applied to high-resolution data (2.5Å) prior to model building. Refinement was performed using Refmac5[161, 163] followed by manual building in Coot[164].

As the minimum R_{free} attainable was greater than 0.3 (the typical cut-off for confident interpretation) SAD was performed using iodide- and bromide- soak derived datasets to try and improve initial phases (Table 5.2.4).

	NRC1 NBARC Bromide peak	NRC1 NBARC Iodide peak
Data collection		
Instrumentation	DLS-i02	DLS-i04
Wavelength (Å)	0.9150	2.0000
Resolution range (Å)	95.51 - 3.10	95.21 – 3.24
Space group	$P6_4$	$P6_4$
Unit cell parameters	a = b =110.28Å c = 53.35Å α = β = 90° γ = 120°	a = b = 109.94Å c = 53.64Å α = β = 90° γ = 120°
Completeness (%)	99.9 (99.5)	100 (99.7)
Unique reflections	6886	6036
Redundancy	22.1 (22.6)	19.4 (20.0)
Anomalous redundancy	11.3 (11.5)	10.1 (9.7)
R_{merge}	0.151 (0.940)	0.126 (0.771)
I/σ(I)	24.6 (3.9)	21.4 (4.7)
Number SAD sites*	3	10

Table 5.2.4 Bromide and Iodide SAD data collection

Processing of data collected from NRC1 NBARC crystals following heavy atom soaking. Numbers in parentheses indicate statistics for highest-resolution shell. *SAD sites determined by PHENIX AutoSol HySS.

Phasing using PHENIX Autosol was unable to produce valid phase solutions using either of these datasets, and combining phase information from multiple heavy atom

derivatives did not improve initial phases, possibly due to non-isomorphisms between crystals.

5.2.8 Preliminary structure of the NRC1 NBARC domain

Note: The structure of NRC1 NBARC domain remains incomplete. A key statistic, Rf_{ree}, which gives an indication of a model's ability to explain experimental data remains at 0.333. Data models with R_{free} above 0.3 should be interpreted with caution.

5.2.9 Validating experimental phasing solutions

Phase approximation by SAD using SeMet datasets and phase extension to 2.5Å native data in PHENIX AutoSol gave a figure of merit (FOM) of 0.346, indicative that a reasonable solution has been achieved. A more definite test of solution quality is investigation of the resulting electron density map. Good solutions to the phase problem should give rise to regions of continuous electron density, with clear solvent channel through the crystal. The resulting experimental map for NRC1 NBARC domain shows both clear solvent channels and interpretable electron density (Figure 5.2.14).

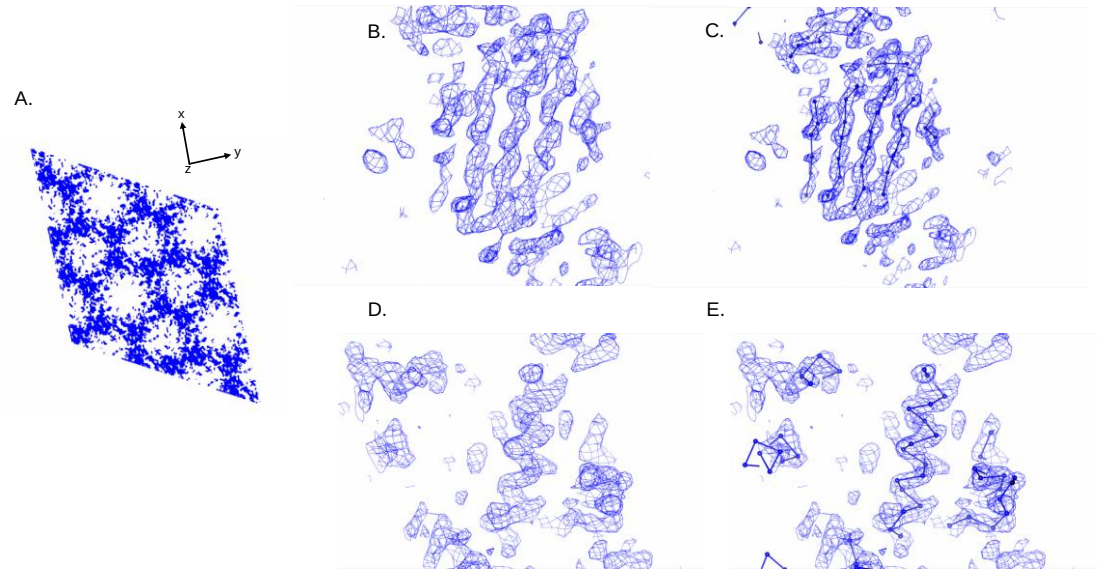


Figure 5.2.14 Density-modified electron density map generated by experimental phasing

A. AB face of NRC1 NBARC unit cell solved by experimental phasing. Blue regions indicate electron density (2.5σ), separated by clear solvent channels. **B**. Electron density (1.5σ) consistent with β-strands. **C**. Interpretation of density in (B.) Spheres correspond to Cα atoms joined by peptide bonds. **D**. Distinct density at 1.5σ corresponding to an α-helix. **E**. Placement of protein backbone into density of D.

The clear and interpretable electron density generated by experimental phasing strongly suggests that a valid solution for phases has been calculated. Interestingly we were able to identify a continuous region of electron density in the putative ligand binding pocket that was distinct from protein backbone in this region. Preliminary analysis indicated that this density was consistent with ADP (Figure 5.2.15).

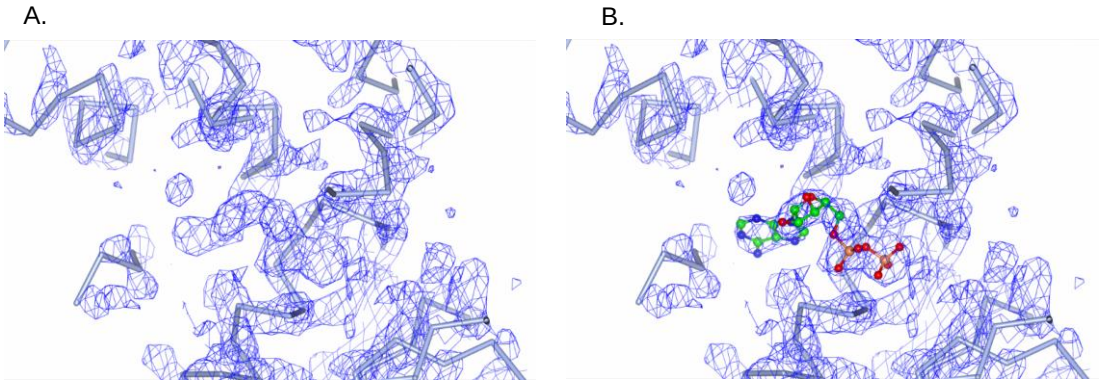


Figure 5.2.15 Identification of distinct non-main chain electron density and fitting of ADP

A. Building protein backbone (blue cylinders) into experimentally derived electron density at 1.75σ shows a region of electron density that cannot be explained by protein main chain or side-chains. B. Placement of a ball-and-stick representation of ADP into this region of electron density shows a good fit to existing density. Red spheres indicate oxygen atoms, phosphate atoms are orange, carbons green and nitrogen atoms blue.

ADP was not added during purification nor was it present in crystallisation conditions, and therefore the bound ADP identified in the electron density must have co-purified with the protein. This suggests a high affinity for the ligand, as purification was performed over two days and included two rounds of preparative-scale gel filtration. Co-purification of nucleotides has been observed for the NBARC domains of Apaf-1 and CED-4, as well as the full-length NLRC4 protein, indicating high nucleotide affinity as a common feature of NBARC domains.

Having determined that we have been able to obtain a valid solution to generate electron density of the crystal structure of NRC1, with clear electron density around the ligand binding site and ARC2 subdomain, we performed several rounds of manual model building and refinement. Although refinement typically leads to a comprehensive structural model that adequately describes experimental data, refinement can introduce biases into electron density maps and should be considered when investigating structures with poor associated statistics.

5.2.10 Investigation of preliminary refined structure and maps

As part of the PHENIX AutoSol pipeline, a preliminary model interpreting the electron density was produced by PHENIX AutoBuild. This preliminary model did not fully explain the existing electron density, and was unable to assign amino-acid identities to many residues. Model improvement was performed in COOT by building Cα traces in obvious regions of density. Identified heavy atom sites were used to assign identities to many residues, and were of particular use around the putative ligand binding site, where strong electron density made interpretation more difficult. After manual building phase refinement was performed using Refmac5. The output from this refinement was then used to identify regions with negative electron density in the 2Fo-Fc map and Cα traces extended where appropriate. Several rounds of refinement were performed in this manner, comparing refined maps against the original density-modified maps from PHENIX AutoSol to ensure that refined maps did not significantly diverge from the experimentally-derived electron density map.

5.2.10.1 Electron density interpretation results in a structure closely resembling Apaf-1

Although model building remains incomplete due to regions of missing or poorly resolved electron density (Section 5.2.11), it has been possible to interpret large regions of the protein to give an idea of the general domain arrangement (Figure 5.2.16).

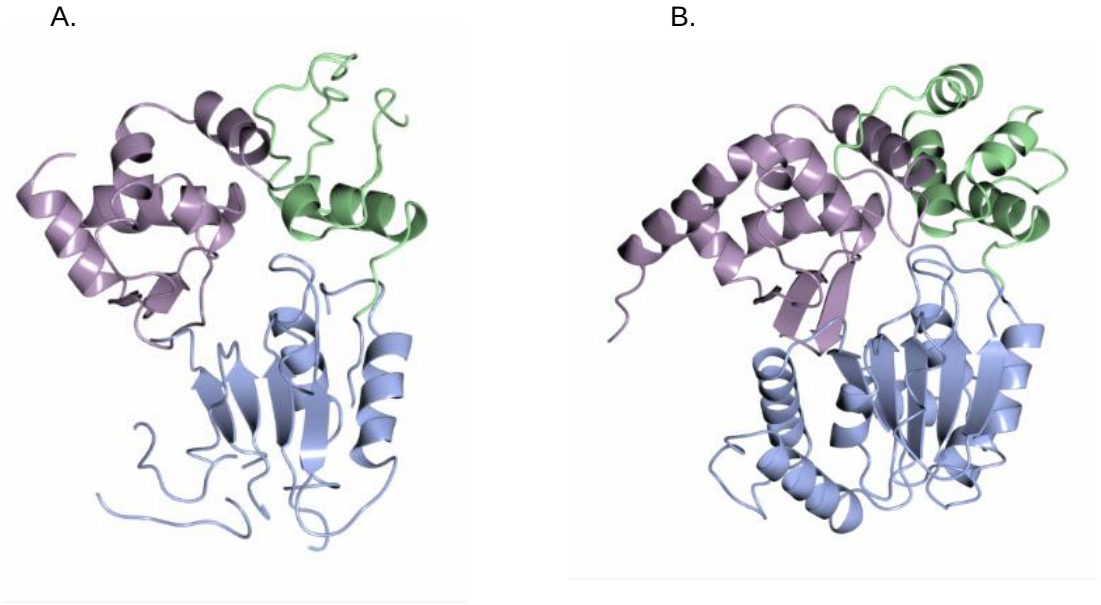


Figure 5.2.16 Comparison of the partial structure of NRC1 NBARC domain (A.) and the crystal structure of Apaf-1 NBARC domain (B.).

A. Preliminary crystal structure of NRC1 NBARC domain. B. Crystal structure of Apaf-1 NRC1 NBARC domain. NB domain coloured in blue, ARC1 in green and ARC2 subdomain in purple. Structures can be superimposed in CCP4MG with an rms of 2.11 over 213 residues.

This preliminary structure shows four β-strands corresponding to the Rossmann fold found in multiple NTPases, an α-helical "bundle" of the ARC1 subdomain and the winged-helix fold of the ARC2 subdomain. Comparison with Apaf-1 NBARC shows high similarity in domain composition and organisation. Given the preliminary nature of the NRC1 NBARC crystal structure, and the observation that electron density outside of the ligand binding site and ARC2 domains is less well defined, detailed comparisons of these regions may not be appropriate. In general terms however, the overall fold of these two domains is broadly very similar, in agreement with predictions from homology modelling.

Electron density in the refined maps is weakest in peripheral regions of the protein, particularly the ARC1 domain and α-helices around the NB domain, however given the quality of electron density around the ligand binding pocket, and positioning of SeMet residues we have been able to confidently interpret this region and draw some preliminary conclusions.

5.2.10.2 Identification of ligand-interacting residues

By interpreting the experimental electron density map, aided by the positioning of selenomethionine residues, we were able to identify key residues interacting with the bound ADP (Figure 5.2.17).

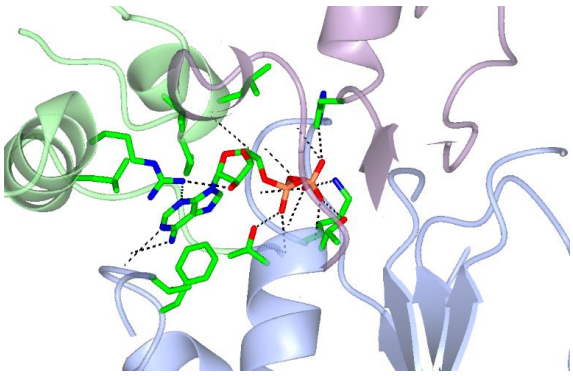


Figure 5.2.17 Identification of residues interacting with bound ADP

ADP is involved in multiple interactions between both side-chain and main-chain atoms of NRC1 NBARC domain. All three subdomains (NB, ARC1 and ARC2) contribute a number of potential interacting partners

Given the incomplete nature of the structure presented here, and the statistics associated with model accuracy, inferences from such incomplete representations should be made with caution. Despite this the preliminary structure shows good electron density in certain regions, notably in the ligand-binding and ARC2 domain, and structural features give us greater confidence in our interpretation of the electron density. These structural features relate to three known functional motifs: the Walker-A motif, Walker-B motif and MHD motif.

5.2.10.3 Positioning of the Walker-A lysine

The conserved lysine in the Walker-A motif of plant NBARC domains is known to be essential for function *in planta*. Furthermore, work with various NTPases highlights the role of this residue in nucleotide binding. Mutations to this residue frequently result in a loss of binding, with structural examples demonstrating the direct role for this residue in protein-ligand interactions. The crystal structure of NRC1 shows this lysine positioned in close proximity to the β-phosphate of bound ADP, forming charge-charge interactions (Figure 5.2.18).

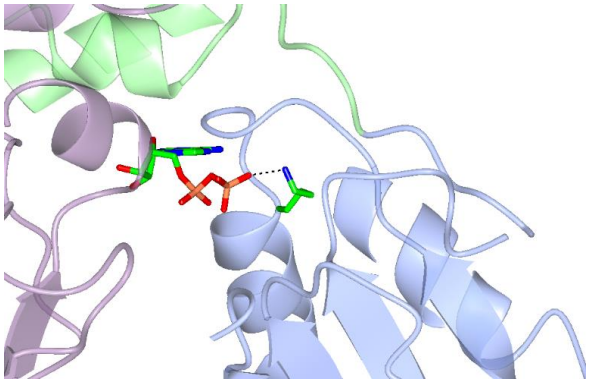


Figure 5.2.18 Interactions between Walker-A lysine and ADP

Lysine 191 of the conserved Walker-A motif is positioned within the ligand binding site, <3.3Å from the β-phosphate of ADP, allowing bond formation between both ADP and ATP.

The positioning of this lysine within the ligand-binding pocket, and its proximity to the bound ligand is in line with our expectations on the basis of prior knowledge regarding NTPase-ligand interactions.

5.2.10.4 Conserved aspartate residues face the ligand-binding pocket

The NBARC domain of two R-proteins have been demonstrated to catalyse the hydrolysis of ATP. This activity is normally attributed to the conserved Walker-B motif, which contains a catalytic aspartate residue. This residue directly or indirectly stabilises divalent cations (normally Mg^{2+}), which are required for bond cleavage. The putatively catalytic aspartate of NRC1 is shown in Figure 5.2.19

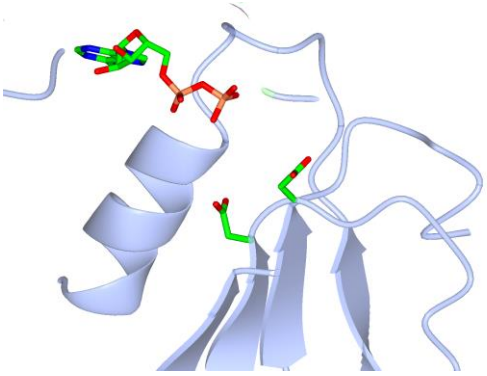


Figure 5.2.19 Location of conserved Walker-B residues

Both the predicted catalytic aspartate residue D268 and neighbouring D267 of the Walker-B motif face into the ligand-binding pocket of NRC1, in proximity to β-phosphate of bound ADP and predicted β-γ bond of hypothetic ATP molecules.

As we would expect for functional ATPases, the conserved and putatively catalytic aspartate of NRC1 NBARC is located within the ligand-binding pocket, facing the β-phosphate of bound ADP, and presumably any potential ATP β-γ phosphate bonds. Interestingly the second conserved aspartate in this Walker-B motif also faces the ligand-binding site, which may explain residual activity seen in Section 6.2.7.

5.2.10.5 Involvement of the MHD motif in ligand interactions

We were particularly interested to find the MHD motif located in very close proximity to the bound ligand. The high conservation of this motif, and the invariability of the central histidine in particular suggested that this motif is functionally important.

Mutations in this motif lead to autoactivation, suggesting a role in regulation, and experiments linking this autoactivation to changes in ligand occupancy point to potential changes in protein-ligand interactions. In this crystal structure we see the central histidine <4Å from the β-phosphate of ADP, and therefore close enough to form salt bridges with the ligand (Figure 5.2.20).

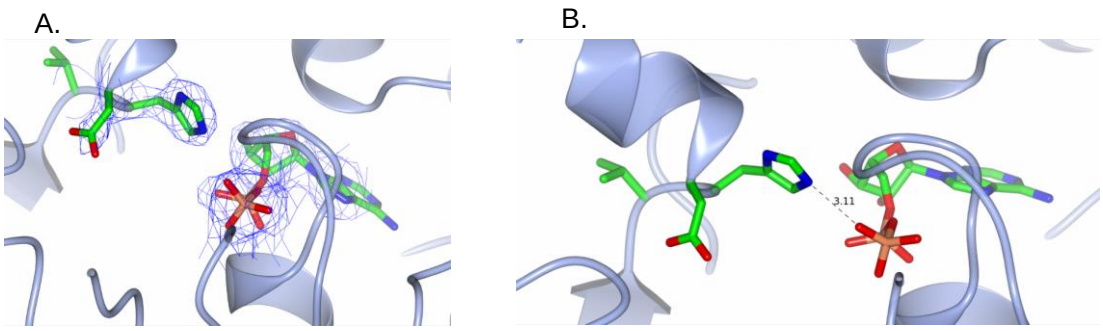


Figure 5.2.20 Proximity of conserved "MHD" motif to bound ADP

A. Electron density around the “MHD” (in this case VHD) motif of NRC1 NBARC domain and ADP at 1.5σ shows good fit for these two regions. Density for C-terminal aspartate is present, although not as well defined.
B. Atomic distances of 3.11Å between histidine and ADP β-phosphate is within the limits for salt-bridge formation (<4.0Å).

This proximity also suggests that an alternative conformation is required when the protein is bound to ATP, as the conformation presented here would lead to significant steric clashes between the histidine and γ-phosphate. This interaction may also play a role in preventing spontaneous activation, which is discussed later (Section 7.4).

5.2.10.6 Global view of MHD and ADP

Using this partial structure and the well-resolved data around the ligand-binding site of NRC1 NBARC we are able to hypothesise about the role of the MHD motif in protein regulation. Our data indicates that this “closed” conformation in the presence of ADP is unfeasible in the presence of ATP, given the positioning of the histidine central to the MHD motif. Looking at this interaction in the context of the full domain can allow us to infer the potential for ATP binding to alter NBARC conformation (Figure 5.2.21).

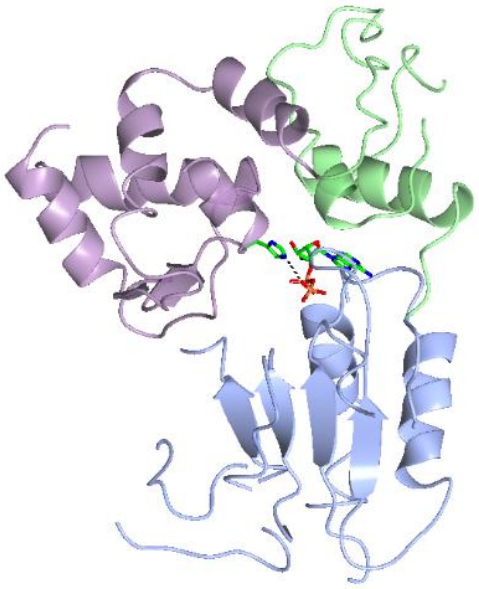


Figure 5.2.21 Histidine-ligand interaction in context of NBARC subdomains

Cartoon representation of NRC1 NBARC domain (blue) with the side chain of the conserved histidine of the MHD motif and bound ADP represented as cylinders. Dashed line indicates potential salt bridge over a distance of 3.1Å.

From this view we hypothesise that in the presence of the γ-phosphate of ATP, the NBARC domain will take on a more open conformation, with the ARC2 domain swinging away from the NB/ARC1 domains. This change is likely to be facilitated by the loop region connecting the first two helices in the ARC2 domain (lilac), giving the potential for flexible rearrangements between these regions.

5.2.11 Uninterpretable electron density prevents completion of NRC1 NBARC structure

This crystal structure of NRC1 NBARC bound to ADP remains incomplete, despite significant efforts to improve model quality and electron density maps. There are two major regions of poor electron density that prevent model completion- the region corresponding to the ARC1 domain, and the regions expected to form connecting helices between the β-strands of the Rossmann fold (Figure 5.2.22).

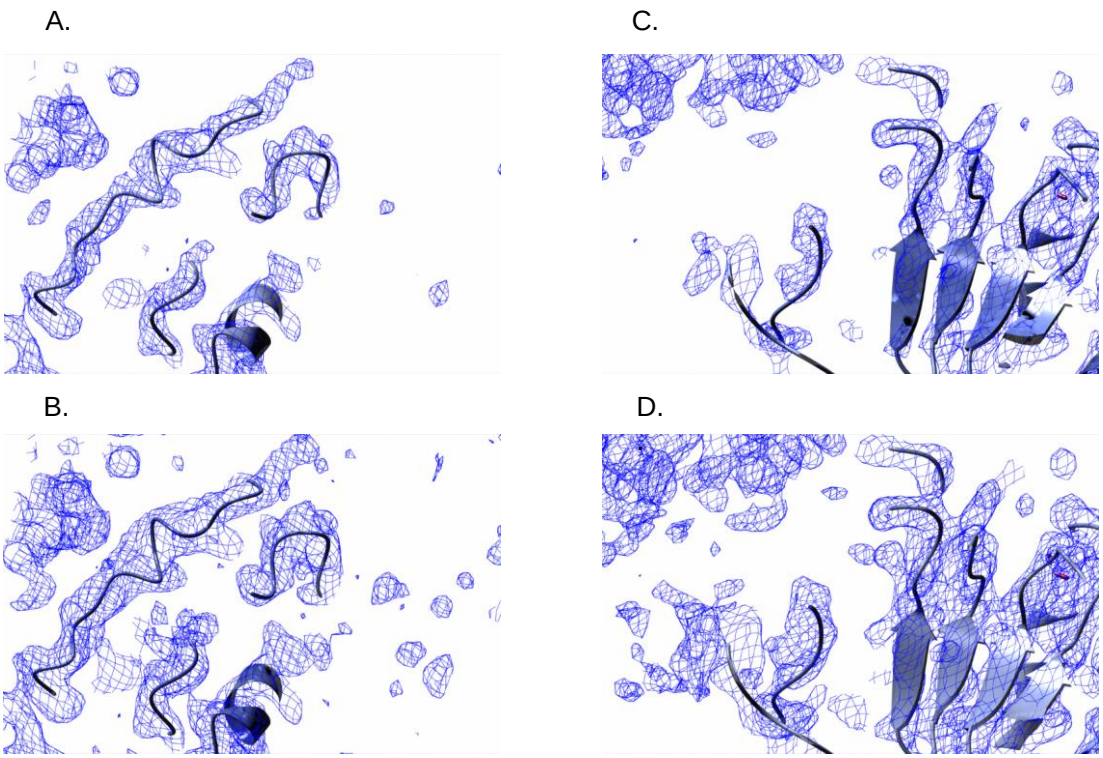


Figure 5.2.22 Problematic regions of electron density preventing model completion

Electron density and ribbon interpretation of problematic regions of the NRC1 NBARC crystal structure. A. and B. show the same region in the ARC1 domain at 1.5σ and 1.0σ respectively. This region is predicted to consist of a 4-helix bundle, however discontinuity of electron density prevents building of coherent structures. C and D show the same region at the top of the Rossmann fold β-strands at 1.5σ and 1.0σ respectively. Predictions based on other NTPases would indicate that each strand should lead into an α-helix, however density in this region is patchy at best.

Electron density can often be poorly resolved in highly-mobile protein regions, such as in loops or mobile domains. For the NBARC domain of R-proteins, no regions are predicted to be inherently flexible. The ARC1 domain has been suggested to act as a hinge region, and it may be that in an ADP-bound conformation, or in the absence of the CC/LRR domains, two helices are more flexible than the others and thus unresolvable. However the ADP bound structure of Apaf-1 gives no indications of this issue, and characterisation of the protein by SAXS suggests that the protein is rigid in solution. The Rossmann fold is not predicted to be flexible, and has been well resolved in other crystal structures. Given the large spatial separation of these two regions, and the well-resolved density of the ARC2 domain, which we hypothesise should be able to take on different respective conformations, it is unlikely that these pathologies are due to biological properties of the protein.

An alternative explanation is that the data we have collected have an explicit pathology that manifests as disorder, for example the data may be twinned. Twinning tests performed by XTRIAGE do not clearly indicate twinning for the native or SeMet data sets, although some tests do give ambiguous results.

We have screened multiple crystals from *E.* coli-derived protein to collect datasets without this potential pathology. SeMet-derived phases have been extended to several crystals from *E. coli* produced protein (resolutions ranging from 2.9-3.3Å), from both similar crystallisation conditions, and from conditions at significantly higher pH. Despite these screens the electron density from these datasets demonstrated the same issues in these regions of the protein.

Native datasets from insect cell-derived protein crystals were collected (Table 5.2.5) and phases solved by molecular replacement using the most complete structure of NRC1 NBARC from the process described in Section 5.2.10. The resulting electron density maps showed little improvement in quality at problematic regions.

	NRC1 NBARC WTIC
Data collection	
Instrumentation	DLS-i03
Wavelength (Å)	0.9795
Resolution range (Å)	55.4 - 2.88
Space group	$P6_4$
Unit cell parameters	a = b = 110.81Å c = 53.39Å α = β = 90° γ = 120°
Completeness (%)	100 (100)
Unique reflections	8660
Redundancy	19.9 (20.8)
Anomalous redundancy	-
R_{merge}	0.051 (0.785)
CC half	0.999 (0.552)
I/σ(I)	22.1 (2.1)
Number SAD sites*	-

Table 5.2.5 Data collection for NRC1 NBARC crystals from Insect cell expression

Data collections looking for changes in crystal packing, or improvements in derived electron density maps by using an alternate protein source. Although able to produce reasonable electron density maps following MR with the incomplete model of NRC1 NBARC domain, overall electron density is not improved in problematic regions. Data were analysed automatically by the xia2 pipeline at Diamond Light Source

5.2.12 Using the preliminary structure of NRC1 NBARC

The mechanistic basis for R-protein regulation has yet to be fully characterised structurally and biochemically. Despite the limitations in the structure described above we have been able to design a series of mutants to allow us to probe the biochemistry of NRC1 NBARC domain (Table 5.2.6).

Mutation	Expected biochemical effect	Potential physiological effect
K191A	Loss of nucleotide binding	Inactive R-protein
D481V	Unknown- potential increased ATP affinity or deficient ATPase activity	Highly auto-active
H480A	Unknown- potential increased ligand exchange, or impaired ligand discrimination	Mild auto-activation or constitutively inactive protein
H480F	Potential increased ligand exchange	Autoactivation, mimicking active (ATP-bound) state
D268N	Significant reduction/loss of ATPase activity	Auto-activation
D267N	Potential reduced ATPase activity	None/mild autoactivation
D268N_D267N	Abolition of ATPase activity	Autoactivation

Table 5.2.6 Structure-guided mutants of NRC1 NBARC

Mutants were selected for either their known effect on R-protein biochemistry and regulation *in planta* or to investigate the mechanistic basis for R-protein regulation. The crystal structure of NRC1 NBARC domain reveals an aspartate residue (D267) in close proximity to the predicted catalytic aspartate of the Walker-B motif (D268). To account for potential misjudgements in assigning residue identities in this region, and to prevent compensatory effects in response to either of these residues a double mutant (D268N_D267N) was included.

Specifically targeting residues that we see involved in different ligand interactions should allow us to better characterise how ligand binding, discrimination and hydrolysis influence the structure and activity of plant NBARC domains. By linking knowledge of the roles of motifs conserved in other proteins (such as the Walker-B motif), and phenotypes seen in *in planta* experiments we are able to suggest the resulting effect of different mutations on the biochemistry and *in planta* activity of NRC1 NBARC domain. Where these predictions are shown to be wrong will allow

us to better assess the predictive capabilities of general models for R-protein regulation.

5.3 Discussion

We have used a range of structural techniques to investigate the physical properties of NRC1 NBARC domain and generate tools that will allow us to more fully explore the structural and biochemical basis for NRC1 regulation. Our data are consistent with a number of predictions, such as the structural similarity between Apaf-1 and NRC1 NBARC domains, and can partially validate hypotheses about R-protein structure-function relationships generated from indirect methods. Ambiguity regarding certain aspects of NRC1 NBARC behaviour in solution suggest that further experiments are required for full characterisation of this protein domain.

5.3.1 Heterologous expression of NRC1 NBARC in *E. coli* results in well-folded protein

We aimed to use a range of physical techniques to understand and characterise the basis of R-protein activation by investigating the NBARC domain of NRC1. At the time of writing there are no examples of the purification of solubly-expressed plant NBARC domains in the literature, and so the purification of soluble NBARC domain from NRC1 gives us a unique opportunity to explore the behaviour of this regulatory domain. The first notable aspect of this protein's behaviour is that it does not appear to exist as a simple monomer in solution. This contrasts with our prediction that the NBARC domain would exist as a monomer in the inactive state, with the potential to oligomerise in response to activation as seen in the apoptosomes formed by APAF-1 and CED-4. Ambiguity in analytical gel filtration data over the precise molecular mass of this species in solution could be resolved using additional experiments such as size-exclusion chromatography with multi-angle light-scattering (SEC-MALS), or analytical ultracentrifugation, which is not liable to artefacts from protein-matrix interactions.

Furthermore the design of a solubly-expressing R3a coiled-coil construct was used to inform construct design to screen expression of multiple R-proteins and their domains in *E. coli* and insect cells (Section 3.5.1). This method allowed us to identify a further three coiled-coil domains that could be expressed solubly in *E. coli*. Our characterisation here of a well-behaving, solubly expressed R-protein NBARC domain may be used by future researchers to expand our repertoire of biochemically characterised R-proteins in a similar manner.

SAXS can be a very useful tool to complement many physical and biochemical techniques. Although X-ray crystallography is the gold standard for producing atomic details about protein structure, the need to crystallise proteins may lead to artefacts in conformation or oligomeric state. SAXS in contrast is unable to produce the high-resolution data of X-ray crystallography. The major benefit of SAXS is that the data are collected from experiments performed in solution. This can allow comparison of crystallographic conformations and solution envelopes to validate crystallographic structures. Beyond this SAXS can also be used to investigate specific protein properties. Calculations of Rg values and Porod volumes can give an idea of oligomeric state in the absence of high-resolution envelopes, and investigation of derived plots such as the normalised-Kratky plot can give information regarding the symmetry of a protein in solution. Of particular interest here is the ability to determine the flexibility of a protein by generating Porod-Debye plots. Having established these plots for wild-type protein, presumably in an ADP-bound state, future work can build on this to compare the effects of ligand binding, or the effects of mutation on protein symmetry, size and flexibility.

5.3.2 Preliminary crystal structure of NRC1 reveals co-purification with ADP and highlights central role for conserved Histidine

We undertook a series of experiments to determine the crystal structure of NRC1 NBARC domain with the view to better understand the structural basis for its conserved switch mechanism. We have solved the crystal structure of NRC1 NBARC domain to 2.5Å, revealing that the protein co-purifies with ADP, and the molecular basis for this binding. Furthermore, we identified interactions between the β-phosphate of ADP and a histidine that we believe plays a pivotal role in ligand discrimination, determining the conformation of the active and inactive states. This histidine forms the centre of the widely conserved MHD motif, and its role in interacting with ADP, and potentially stabilising the ARC2 domain, may explain the high level of conservation of this motif across a wide range of R-proteins.

Complete structural characterisation of this domain has been unsuccessful due to limitations in resolving and interpreting regions of electron density in our experimentally-derived maps. Despite this, the electron density maps, and use of identified SeMet sites, allowed sequence placement and gives us confidence in our interpretation of the ligand binding site and surrounding areas. Refinement of

phases has allowed us to identify the amino acid side-chains involved in co-ordinating protein-nucleotide interactions, and has allowed us to design a series of structure-guided mutants to probe different aspects of ligand discrimination and auto-inhibition.

5.3.3 Designing structure-based mutants to investigate ligand-dependent regulation

With experimentally derived structural information about the interactions between NRC1 NBARC and ADP we have been able to design a number of point mutants to explore regulation of the protein. Some of these mutants have been explored in similar proteins, such as the Walker-A mutation leading to loss-of-binding in multiple NTPases, however the effect these changes have on R-protein dynamics in solution has yet to be explored. We have been able to rationalise the phenotype of previously generated mutations that have no experimentally determined basis, such as the commonly used $MH^{D->V}$ mutation to cause autoactivation. Furthermore as we were able to express the NBARC domain of NRC1 solubly in *E. coli* we had hoped to use these mutants to provide detailed analysis of the mechanisms of ligand discrimination and regulation through investigations into the biochemistry and dynamics of mutations to the conserved MHD motif in isolation. The biochemistry of both wild-type and point-mutants are explored in Chapter 6 “Functional characterisation of NRC1 NBARC domain”.

6

Functional characterisation of NRC1 NBARC domain

6.1 Introduction

Models for R-protein regulation have been typically derived from experiments performed either *in planta* or using refolded protein. Such models allow us to hypothesise about the behaviour of the N-terminal domains of NRC1 in isolation and conversely allows us to test and validate such generalised models. For example signalling repression is suggested to occur by stable interactions between the NBARC and coiled-coil domains[1, 153, 165]. These *in planta* experiments are qualitatively very useful, however due to the complex environments in which they are performed it is not possible to derive quantitative data regarding binding affinities and stability. Additionally the complex cellular environment means that such interactions may be indirect, mediated by additional host proteins, or may require post-translational modification of the domains to facilitate such interactions. Purification of proteins allows the removal of potentially confounding host proteins, and allows the use of sensitive techniques to generate quantitative data regarding their biochemical properties.

Biochemical investigations into refolded N-terminal R-protein domains suggest that these proteins are catalytically active[102, 103], however the rates of activity appear to be very low. Indeed some researchers questioned whether the domains were true nucleotide binding sites with bona fide enzymatic activity[166]. Soluble expression of these domains from homologous R-proteins should allow comparison to these previously published data, and exploration of whether this low activity may be due to the effects of refolding, or whether slow ATP hydrolysis is an intrinsic property of these proteins. Better understanding of rates of ATP hydrolysis may have implications for maintaining active states *in planta*, and the tolerance these proteins will have to mutations that may inadvertently affect protein-nucleotide interactions.

Here we test three key putative functions of R-protein NBARC domains that are believed to influence the regulation of R-proteins *in planta*. We investigate the possibility of signal repression by CC/NBARC interactions, the ability to revert from “on” to “off” states by ATP hydrolysis, and the possibility of the NBARC domain to interact with DNA.

6.2 Results

Note: DNA fluorescence anisotropy assays were performed with the Cann group (Durham University). SPR was performed by Clare Stevenson (JIC).

6.2.1 Purification of NRC1 N-terminal domains for *in vitro* assays

Both NRC1 NBARC domain and coiled-coil domains were expressed using conditions identified in Section 3.5.2. 6L of pOPIN-S3C:NRC1^{CC} in BL21 Lemo were grown in PB media and induced at an OD_{600} of 0.6 by the addition of 1mM IPTG after cultures had been rapidly cooled on ice. Cells were pelleted by centrifugation and lysed by sonication. Cell debris was removed by centrifugation and clarified lysate loaded onto a pre-equilibrated 5mL Ni^{2+}-NTA column. Bound protein was eluted by single step addition of 500mM imidazole and loaded onto a superdex sepharose S75 column. Given the low predicted absorption of light at 280nm by the protein, all fractions were collected from the column as the protein was unlikely to produce a significant peak (Figure 6.2.1).

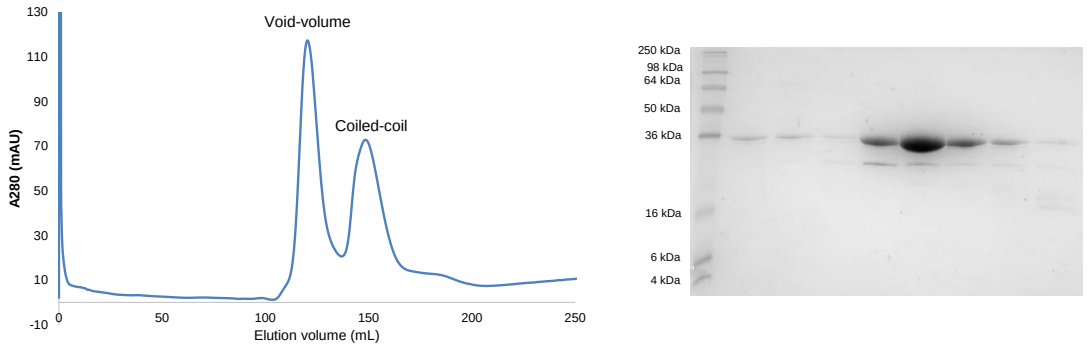


Figure 6.2.1 Purification of SUMO-tagged NRC1 coiled-coil

Soluble *E. coli* extract expressing SUMO-tagged NRC1 coiled-coil domain loaded onto a sepharose S75 superdex column following nickel-affinity IMAC. Left. Gel filtration trace of eluting proteins. Major peak at approximately 110mL is indicative of the presence of proteins larger than ~70kDa eluting from the IMAC column. Right. SDS-PAGE of fractions eluting from the SEC column indicates that protein consistent with the mass of SUMO-tagged NRC1^{CC} elutes in both the void-volume (first 3 lanes) of the column and at approximately 150mL with few contaminants.

SUMO-tagged NRC1^{CC} was found to elute at both 150mL and ~110mL (void-volume of the column), indicating that a fraction of the protein is either misfolded or forming large complexes. Coomassie staining of SDS-PAGE separated proteins show very few contaminants in the void volume (Figure 6.2.1, Right, first three lanes), making self-association or aggregation a likely explanation for this elution peak. The lower

molecular-weight bands seen in the coomassie stained SDS-PAGE gels could be either contamination or degradation products. As our intention was to purify soluble, well-folded protein only fractions containing the SUMO-tagged protein that eluted away from the column void-volume were used for further purification. Selected fractions were pooled and concentrated to <20mL prior to the addition of 3C protease and incubated overnight at 4°C. Residual tagged protein and protease were removed by subtractive IMAC. Flow-through and washes were collected and concentrated to <5mL and loaded onto a superdex sepharose S75 column for final purification and buffer exchange (6.2.2).

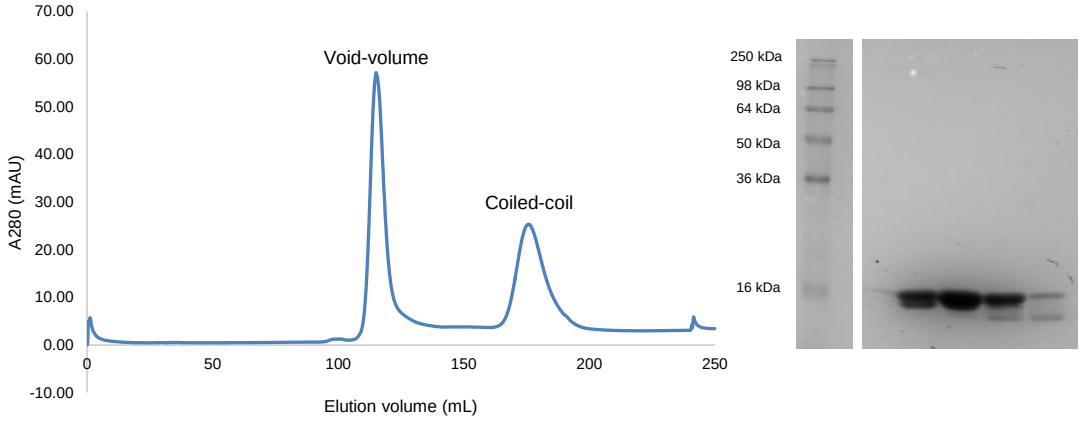


Figure 6.2.2 Final purification of the NRC1 coiled-coil domain

Purification of NRC1^{CC} following SUMO-tag removal and subtractive IMAC shows two well-defined peaks eluting from a superdex S75 SEC column (Left). The first peak is consistent with the void-volume of the column suggesting a species of >70kDa. The second peak shows clear separation from the void-volume indicating well folded protein. Right. SDS-PAGE and coomassie staining of fractions from the second elution peak. Migration of the protein is consistent with the expected size of the coiled-coil domain, however multiple bands may suggest degradation of the protein.

We were able to detect NRC1^{CC} eluting in both the void-volume of the column, and at ~170mL. This was surprising as only SUMO-tagged protein eluting away from the void volume, indicative of well-folded protein, was used for tag removal. The presence of NRC1^{CC} in the void-volume of this second purification may be indicative of either instability of the protein in solution leading to aggregation, or an equilibrium between monomeric and multimeric coiled-coil domain. SDS-PAGE and coomassie staining shows that purified protein contains no visible high molecular-mass contaminants, however the presence of lower molecular-mass bands indicates some protein degradation. This was confirmed by the observation that the protein was unstable in solution, with significant degradation apparent after ~3 days.

Protein eluting away from the void-volume was pooled and concentrated for downstream experiments, with fresh protein produced for each use.

NRC1 NBARC domain was typically purified from 4L of PB induced at OD_{600} = 0.6 with 1mM IPTG after rapid chilling of cultures on ice. Protein expression was performed at 18°C for 16h with constant shaking. Purification of soluble protein was performed as described in Section 5.2.1 to obtain untagged, soluble NRC1 NBARC domain.

6.2.2 Assessment of point mutant protein folding by circular dichroism

The introduction of point mutations may lead to global changes in protein fold, significantly affecting structure and therefore function. To test that the point mutants we were able to generate did not have significant changes in folding we performed circular dichroism experiments using wild-type NRC1 NBARC domain from both *E. coli* and insect cells, and used these spectra to investigate the effects of D268N and D481V mutations on the secondary structure of purified NBARC domain. Proteins were diluted >20 fold to a final concentration of 0.1mg/mL in phosphate buffer and spectra recorded from 180-260nm in 1nm steps at 25°C (Figure 6.2.3).

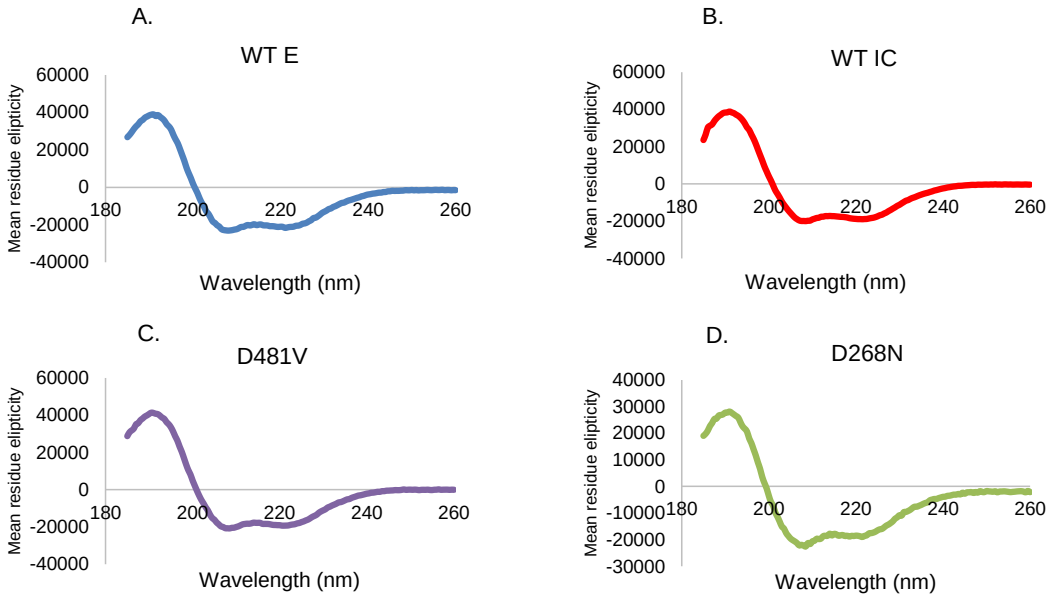


Figure 6.2.3 Circular dichroism spectra of wild-type NRC1 NBARC and generated point mutants

A. Wild-type NBARC domain purified from *E. coli*. B. Wild-type NBARC domain purified from insect cells. C. NRC1 NBARC^{D481V} purified from *E. coli*. D. NRC1 NBARC^{D268N} purified from insect cells.

Purified proteins were diluted >20 fold into phosphate buffer to minimise buffer interference. Spectra were recorded at 25°C in 1nm steps, with three readings recorded per wavelength. Plots represent averaged, buffer subtracted spectra.

Comparison of CD spectra shows no apparent change in secondary structure between heterologous expression system, and indicates that the introduction of point mutations of aspartate at position 481 and 268 to valine and asparagine respectively do not lead to significant changes in secondary structure. We were therefore confident that changes in biochemical activity of these proteins is not likely to be due to substantial changes in protein folding.

6.2.3 NRC1 coiled-coil and NBARC domains do not form a stable complex *in vitro*

A potential explanation for signal repression in the absence of a pathogen is that the CC domain is sequestered by the NBARC domain bound to ADP, and thus unable to interact with downstream signalling components. In this model displacement of ADP for ATP results in a conformational change in the NBARC domain, which releases the CC domain, allowing it to interact with essential signal transduction factors. We tested whether this was a feasible method for regulation of NRC1 activity by investigating whether the CC and NBARC domains were able to interact in the absence of any additional factors.

Analytical gel filtration was performed with the coiled-coil domain and NBARC domains to determine their individual elution volumes. The two domains were then mixed and incubated on ice for 2-6 hours or 4°C overnight and the protein mixture elution volumes measured by analytical gel filtration. Gel filtration separates proteins based on their size and conformation; larger proteins elute in a smaller volume than proteins with a lower molecular mass. We therefore focussed on the formation of a new A_{280} peak in an elution volume smaller than that seen for NRC1 NBARC domain of circa 10.3mL (Figure 6.2.4).

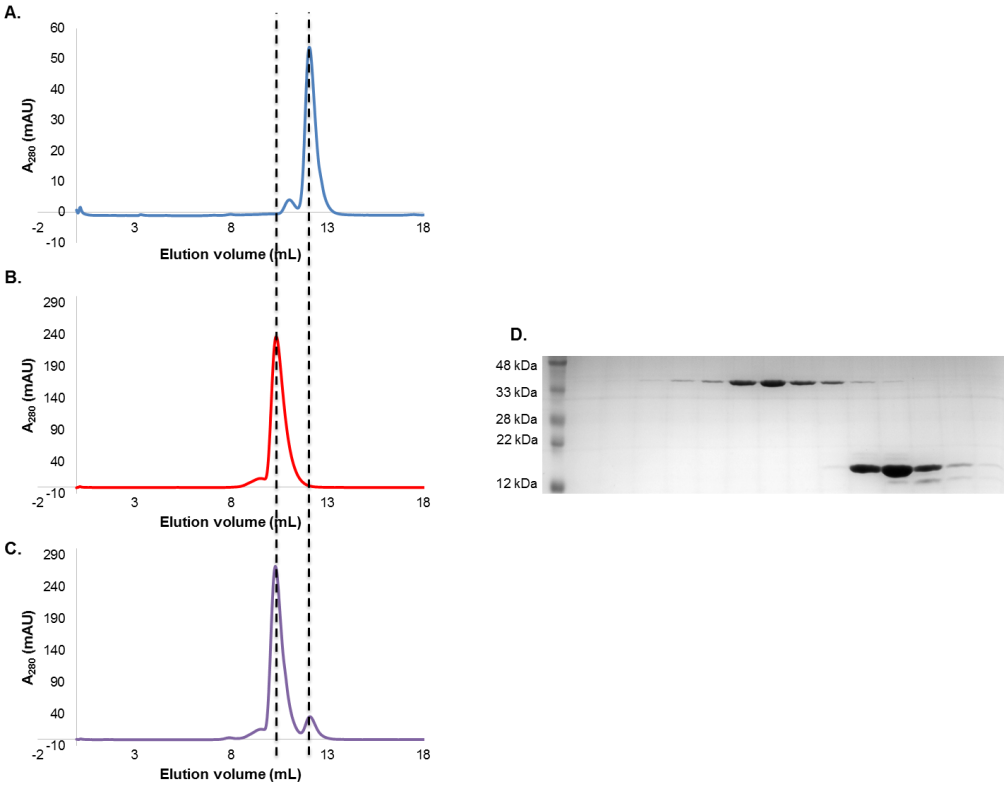


Figure 6.2.4 Analytical gel filtration of NRC1 coiled-coil and NBARC domains

A. Elution of NRC1 coiled-coil domain shows a single major peak at approximately 12.1mL.
B. NRC1 NBARC domain elutes as a single peak at approximately 10.4mL
C. Incubation of these two domains at a 4:1 molar ratio (coiled-coil to NBARC) shows no significant change in elution volume and no formation of additional higher-molecular mass species. $NRC1^{CC}$ has very low absorption at 280nm and so produces a much smaller elution peak than the NBARC domain despite the high molar excess
D. SDS-PAGE and coomassie staining of fractions eluting from C. The lack of both proteins present in the same fractions confirms the results from C.

NBARC and CC domains were mixed at 1:1, 1:2 and 1:4 molar ratios, using three separate protein preparations for the NBARC domain and two independent purifications of the CC domain. No signs of complex formation were seen during analytical gel filtration even with high molar excess of CC domain. This may indicate that the proteins require either additional host proteins or the LRR domain of the protein to stabilise weak interactions, or require post-translational modifications in order to form a stable complex *in planta*.

6.2.4 Yeast 2-hybrid is unable to identify interactions between N-terminal NRC1 domains

Although heterologous protein expression in *E. coli* allows high yields of highly purified proteins, the simple prokaryotic environment may not be able to recreate the folds and modifications of proteins expressed in eukaryotic cells. We used a eukaryotic-based *in vivo* assay to confirm two key *in vitro* results regarding NRC1 NBARC interactions- the inability of the coiled-coil domain to form a complex with the NBARC domain, and the observation that the NBARC domain appears to purify as a homodimer. We used a yeast 2-hybrid (Y2H) system to test for interactions for two reasons. Firstly the eukaryotic environment may produce NRC1 domains that better reflect the state of the protein *in planta*, and secondly Y2H can be more sensitive to transient interactions than exclusion chromatography and immunoprecipitation experiments.

ProQuest yeast 2-hybrid system uses the separation of the GAL4 transcription factor into separate DNA binding (DBD) and activation domains (AD) to test whether two proteins interact when fused independently to the DBD and AD domains. If the two proteins interact, the DBD and AD come into proximity and are able to drive transcription of a target gene. Reconstitution of the transcription factor drives expression of HIS3, which is required for growth of the yeast strain MaV203 on media lacking histidine. Once yeast cells have been co-transformed with vectors carrying the genes of interest fused N-terminally to the activation and binding domains, they are spotted out onto histidine-deficient media, and only if the two proteins interact are the cells able to survive and form colonies. In order to fine-tune this selection and reduce false-positives by leaky expression of HIS3, the media is also supplemented with a repressor of this gene expression, 3-amino-1,2,4-triazole (3AT).

DNA encoding the coiled-coil domain of NRC1 was cloned into the pDEST32 vector as described in Section 2.2.2. The region encoding the NBARC domain was cloned into both pDEST22 and pDEST32 vectors to allow testing for dimerization. Yeast were co-transformed with plasmids encoding pDEST22:NRC1^{CC}, pDEST32:NRC1^{NBARC}, pDEST22:NRC1^{NBARC}, pDEST22:empty vector, or pDEST32:empty vector to test for interactions and baseline activity. Empty vector controls are required to determine whether fusion of the vector to the gene of interest leads to spontaneous activation and gene expression- particularly important for investigations into the activity of R-protein NBARC domains in light of recent publications highlighting the potential for DNA binding[167](Section 6.2.8). Following

transformation yeast were plated on media lacking leucine and tryptophan (SC-LW media) to select for co-transformants, as only cells carrying both pDEST22 and pDEST32 vectors are able to grow in the absence of these two amino acids. After four days at 28°C growth individual colonies were selected, re-suspended in water and spotted on to media lacking leucine, tryptophan and histidine (SC-LTH). Plates were incubated at 28°C and imaged every day for four days (Figure 6.2.5).

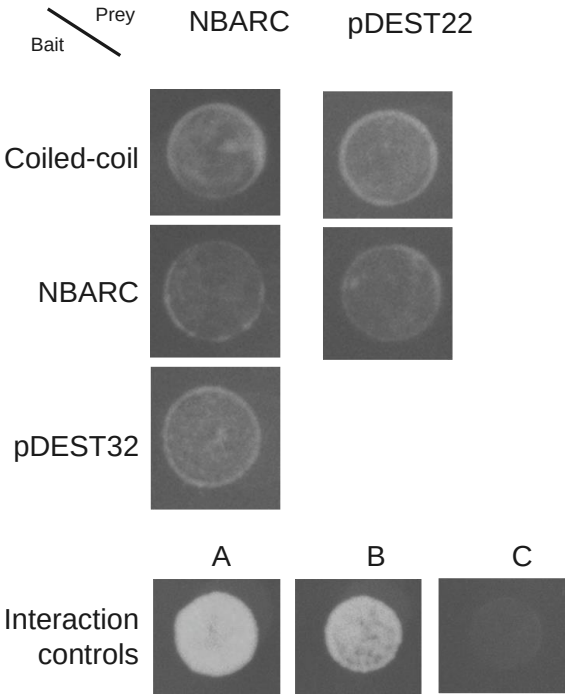


Figure 6.2.5 Yeast 2-Hybrid for NBARC dimerisation and coiled-coil interactions

MaV203 yeast, auxotrophic for leu, trp and his, co-transformed with pDEST22 and pDEST32 vectors encoding NRC1^{CC} and NRC1^{NBARC} domains. Image shows colonies after four days growth at 28°C on SC-LYH media, supplemented with 10mM 3-AT. Three transformants from two independent transformations were selected and spotted out with representative colonies shown above. All colonies show intermediate phenotypes between the “weak interaction” control C, and the “moderate interaction control” B. Colony growth when individual constructs are co-transformed with empty vectors indicates mild auto-activation, with experimental colonies not growing significantly more than these negative controls.

Although there appears to be moderate growth for yeast cells expressing both the coiled-coil and NBARC domain of NRC1, this growth is comparable to yeast carrying the individual domains and empty vector controls (pDEST22:EV with pDEST32:NRC1^{CC}, and pDEST22:NRC1^{NBARC} with pDEST32:EV). Growing yeast in lower concentrations of the inhibitor 3AT (7.5mM and 5mM) affected all co-

transformants equally, and so did not lead to a difference in growth phenotype between experimental and negative control transformants.

Contrary to *in vitro* data, there are no indications of strong NBARC-NBARC interactions in this system. One of the first indications that the protein may be forming oligomers was the observation that removal of the ~13kDa N-terminal SUMO tag lead to a decrease in elution volume from a preparative gel filtration column (Section 5.2.1). This may suggest that the N-terminal tag prevents normal protein-protein interactions, which can only occur after cleavage. In this yeast 2-hybrid system cloning into both the bait and prey vectors leads to an N-terminal protein fusion. It is possible that fusion to the DBD and AD acts analogously to the presence of the N-terminal SUMO tag in preventing NBARC dimerization.

6.2.5 NRC1 NBARC domain is a functional ATPase

Given the presence of conserved Walker-A and Walker-B motifs in NRC1 NBARC domain, we hypothesised that this protein should be functional ATPase, as is the case for I-2 and Mi-1. In the model described in Section 1.7, ATPase activity may be an important characteristic of R-proteins, allowing them to reset after activation by hydrolysing ATP to ADP with associated conformational changes. In this model ATPase activity and nucleotide exchange are intrinsic functions of NBARC domains, requiring no external activation as is the case for many GTPases, but instead require the removal of the inhibitory LRR domain.

Purified NRC1 NBARC domain should therefore be able to facilitate the turnover of ATP in the absence of the LRR domain. We took two approaches to test this- one assay measuring the release of ADP as it is displaced by ATP, the second measuring the release of free phosphate (Pi).

ADP release was measured using a linked enzymatic assay. ADP is used as a phosphate acceptor in the conversion of phosphoenolpyruvate to pyruvate by pyruvate kinase (PK). Pyruvate is then used by lactic dehydrogenase (LDH) to produce lactate in a reaction that results in the oxidation of NADH. NADH absorbs light at 340nm to a much greater extent than NAD^+ and so the conversion of NADH to NAD^+ can be measured spectroscopically as a decrease in A_{340}, allowing indirect measurement of the amount of ADP released by an ATPase (Figure 6.2.6).

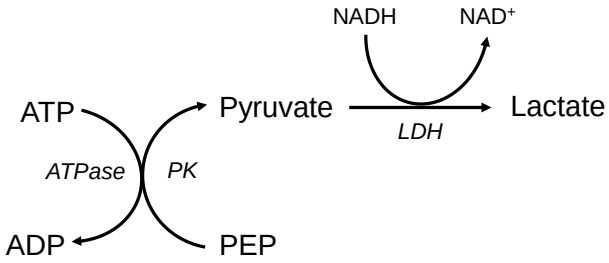


Figure 6.2.6 Schematic of linked ATPase assay

This linked pyruvate kinase (PK) lactate dehydrogenase (LDH) assay allows indirect measurement of ADP and thus ATP hydrolysis performed by a putative ATPase. ADP is used by PK to produce pyruvate from phosphoenolpyruvate (PEP). This pyruvate is then a substrate for the NADH-dependent formation of lactate by LDH. The assay measures the oxidation of NADH, which absorbs light at 340nm strongly, to NAD^+ which does not. By using an excess of PEP and NADH in reaction conditions, and the high K_{cat} of both PK and LDH it is possible to measure the rate of ATP hydrolysis by a putative ATPase by observing change in A_{340} over time.

The ATPase activity of NRC1 NBARC domain was assessed using this linked PK/LDH assay, with the rate of decrease in A_{340} being proportional to the rate of ADP release. Initial measurements were made in the absence of NRC1 NBARC domain every minute for 10 minutes to assess the background change in A_{340} and readings taken every minute after the addition of NRC1 NBARC domain (or BSA as a negative control) to measure the average rate of A_{340} decrease. Also included were samples that contained NRC1 NBARC domain, but no ATP. Structural studies indicate that the protein co-purifies with ADP and so this control was used to distinguish between ATP hydrolysis and subsequent ADP release, and stochastic dissociation of ADP from the protein (Figure 6.2.7).

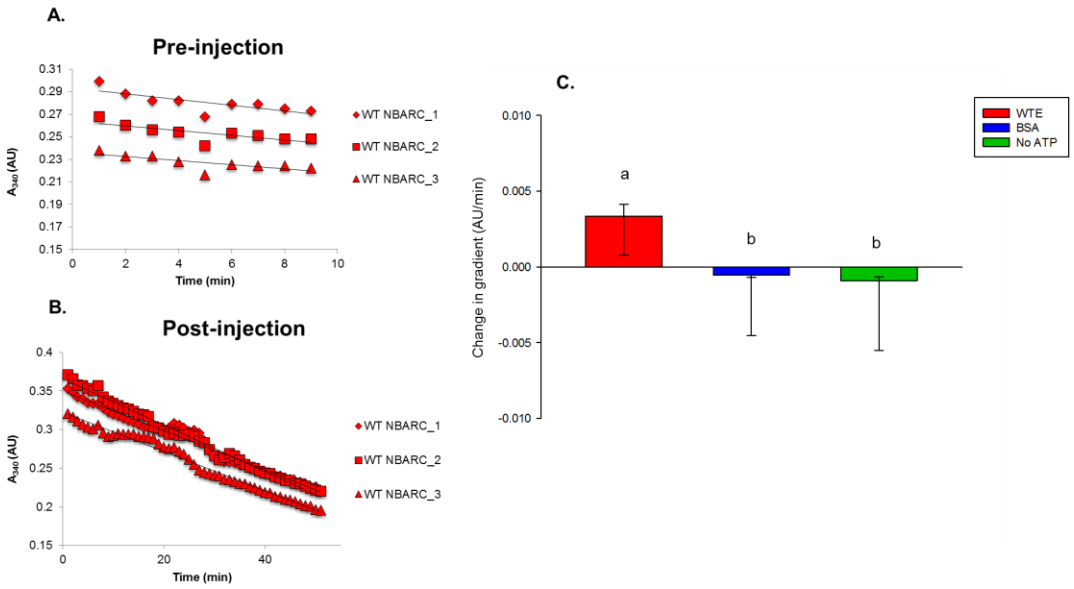


Figure 6.2.7 PK/LDH linked ATPase assay of NRC1 NBARC domain

A. Plot showing three technical replicates for sample absorbance at 340nm over time prior to the addition of NRC1 NBARC domain.
B. The same samples as in A. following addition of 50µM NRC1 NBARC domain.
C. Comparison of rate of A_{340} decrease between samples before and after addition of protein. Positive values indicate an increase in rate of NADH oxidation and therefore more ADP available for linked reactions. No ATP control indicates the rate of ADP release from WT NBARC domain in the absence of ATP. Columns indicate median values, with error bars showing SE. Data from three repeats were analysed using Sigma plot. Shapiro-Wilk test indicates that data are not normally distributed and so analysis was performed as pairwise one-way ANOVA on ranks. Significance was determined using Dunn's method. For WT NBARC domain n=8, for BSA and no ATP controls n=9.

In these assays we are able to see a clear difference between NRC1 NBARC and the BSA control, indicating that the NRC1 NBARC generates ADP for use in the linked assay. Additionally the difference between NRC1 NBARC domain in full assays and assay conditions lacking ATP suggests that this ADP is not the result of stochastic dissociation of the co-purifying ADP. It is possible that the ADP measured is not newly formed by the hydrolysis of ATP, but rather the presence of ATP displaces pre-existing ADP. Continuing these assays for a longer time period, say until the depletion of PEP would demonstrate the bona fide ATP turnover. Alternatively the use of a second, independent ATPase assay can be used to demonstrate ATPase activity.

The second method we used to test the ATPase activity of NRC1 NBARC measures the release of free phosphate (Pi) using malachite green[125, 168]. This method

requires fewer additional inputs, and is preferable to the linked assay as free phosphate is unlikely to co-purify with the protein, a complication for the linked assays. In this method Pi forms a complex with molybdenum which absorbs light at 630nm. The reactions were performed using 2mM $MgCl_2$ and 1mM ATP in HEPES pH7.5. Reactions were stopped by quenching with malachite green reagent and quenched reactions left for 20 minutes at room temperature to stabilise before readings were taken. As the solution continues to increase in A_{630} over time, BSA was included as an internal control in all runs, allowing background subtraction (Figure 6.2.8).

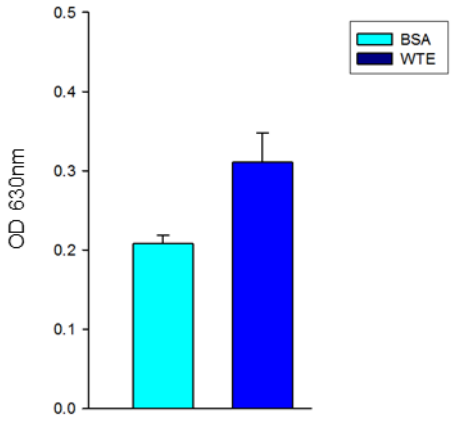


Figure 6.2.8 Malachite green assay comparing BSA and NRC1 NBARC-mediated phosphate release

ATPase assays performed with either BSA (light blue) or WT NRC1 NBARC domain purified from *E. coli*. OD_{630} increases with phosphate concentration. Bars represent mean OD values with error bars showing SD. Reactions performed in triplicate. Data presented here are from a full assay shown in Figure 6.2.13.

6.2.6 Investigating ligand-dependent oligomeric changes in NRC1 NBARC domain

A number of NBARC domain containing proteins are known to undergo changes in oligomeric state when activated- Apaf-1, Ced-4 and DARK all form large apoptosomes when activated, with interactions mediated by their NBARC domains. We therefore wanted to investigate whether the NBARC domain of NRC1 showed changes in oligomeric state in response to activation- either by the binding of a non-hydrolyzable ATP analogue (ATP-γ-S), or the introduction of autoactivating mutations.

To monitor the ability of ATP-γ-S to bind to NRC1 NBARC domain we used the linked PK/LDH assay, as described in Section 6.2.5. Full ATPase assays were set

up, and the rate of ADP release/ ATP hydrolysis measured for 10 minutes. After this time period, 5mM ATP-γ-S was added to the reactions and change in A_{340} decrease measured (Figure 6.2.9).

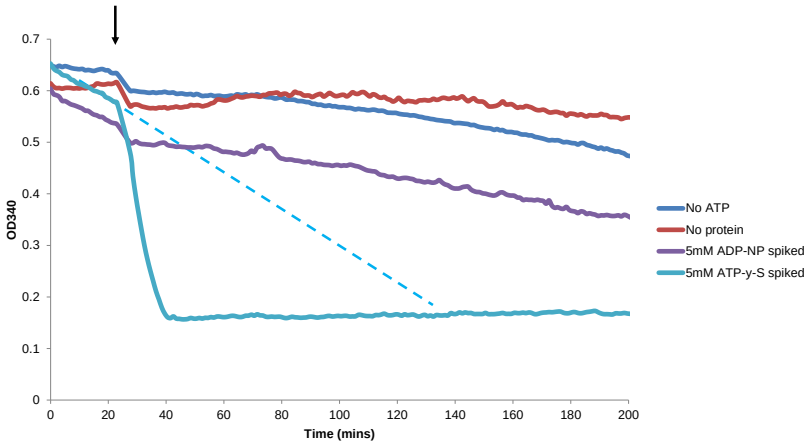


Figure 6.2.9 Spiking linked PK/LDH assay with non-hydrolysable ATP analogue

Averaged A_{340} for three technical replicates of linked PK/LDH assay with 50µM WT NRC1 NBARC acting as a source of ADP. Arrow indicates time point for addition of 5mM ATP-γ-S, ADP-NP or A4 buffer for the two negative controls. Dotted line is linear extrapolation from original rate of reaction for ATP-γ-S to show expected behaviour of NRC1 NBARC in the absence of non-hydrolysable analogue. A_{340} shows a sharp decrease following addition of ATP-γ-S, indicating rapid release of ADP. Plateau at ~40 minutes suggests that NRC1 NBARC is no longer able to bind to hydrolysable ATP and is thus fully occupied by the non-hydrolysable analogue.

From this preliminary experiment it appears that complete binding of ATP-γ-S to the NBARC domain is achieved approximately 20-30 minutes after addition of the substrate analogue. We used this to establish a minimum incubation period for NRC1 NBARC and ATP-γ-S before performing additional experiments.

In Section 5.2.4 we demonstrated that the elution volume of NRC1 NBARC domain from a superdex 10/300 GE column is approximately 10.3mL. As the NBARC domain was found to co-purify with ADP (Section5.2.8), we believed that this would represent this domain in the “off” state, and so activation via nucleotide exchange could result in either higher order oligomers as is the case for Apaf-1, or could result in dissociation of the complex leaving monomeric protein. Preliminary experiments were performed by DLS (dynamic light scattering) using 1mg/mL NRC1 NBARC domain incubated with either ADP or ATP-γ-S overnight at 4°C (Table 6.2.1).

	R (nm)	%Pd	MW-R (kDa)	%Mass
ADP treated	3.6	13	67	97.8
ATP-γ-S treated	3.2	19.5	51	97.4

Table 6.2.1 Estimated NRC1 NBARC particle size in response to treatment with either ADP or ATP-γ-S by DLS

NRC1 NBARC domain incubated with either 1mM ADP or 1mM ATP-y-S overnight at 4C. Ten readings were taken at 25°C per sample and data interpreted using DYNAMICS® v6 software (Wyatt technology) assuming particles as rayleigh spheres. R (nm) denotes calculated particle radius, %Pd denotes polydispersity of samples, MW-R indicates the calculated molecular weight of particles based on radial measurements. Addition of the non-hydrolsable ATP analogue ATP-γ-S appears to result in a decrease in particle mass of approximately 16kDa.

This preliminary result suggests a change in particle dimension in response to treatment with an ATP analogue. To confirm this result, and get more accurate data regarding protein mass, we performed analytical gel filtration with fresh protein that was incubated with either ADP or ATP-γ-S. Protein was incubated for 6-16 hours at 4°C, with one set incubated for 4 hours at 4°C followed by 1 hour at room temperature, before being loaded onto a superdex™ S75 10/300 GL column. Figure 6.2.10 shows the resulting elution traces of the NBARC domain in the presence of these two ligands.

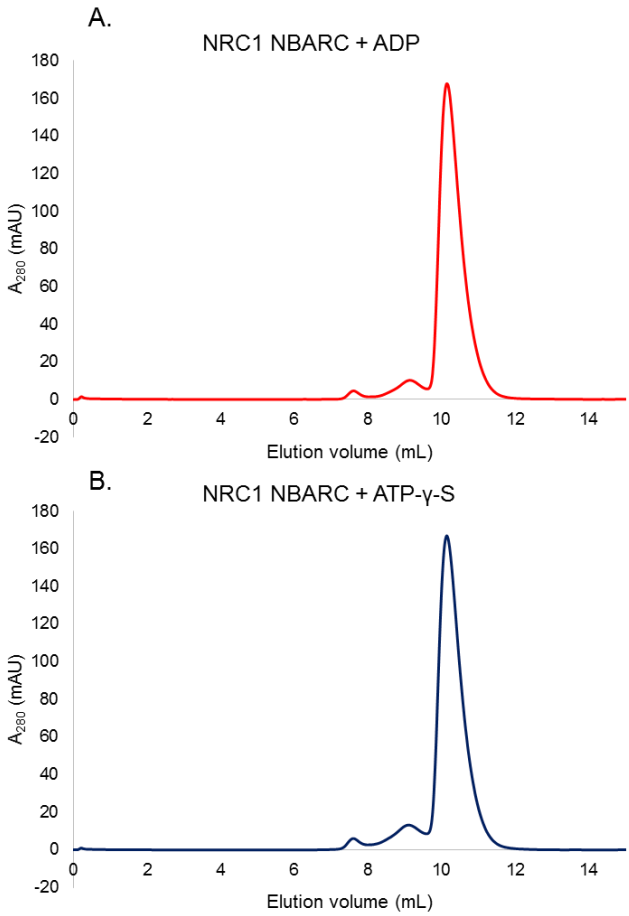


Figure 6.2.10 Analytical gel filtration of NRC1 NBARC domain incubated with either ADP or ATP-γ -S

Analytical gel filtration of NRC1 NBARC domain purified from *E. coli* incubated with either 1mM ADP (A.) or 1mM ATP-γ-S (B.). Elution profiles were consistently similar over multiple experiments, indicating that the size of NRC1 NBARC complexes does not change in the presence of these two ligands under these conditions.

The elution profiles we see for NRC1 NBARC domain incubated with either ADP or ATP-γ-S shows no significant changes in response to the different ligands. Although we were unable to characterise the bound ligand prior to these assays, the linked enzymatic assay demonstrating ADP release suggests that nucleotide exchange should occur during this time frame, and given the excess of ATP-γ-S we should expect at least partial occupancy with this ligand. However as the substrate trapping experiment (Figure 6.2.9) was performed only once, it may be that this rate of nucleotide exchange is not representative of the behaviour of NRC1 NBARC in solution, and so it is possible that when incubated at 4°C the protein has undergone

little nucleotide exchange. This would mean that the gel filtration traces in Figure 6.2.10 are representative only of ADP-bound protein. To overcome this issue we tested the oligomeric state of the NRC1 NBARC Walker-B mutant, which we predicted would have impaired ATPase activity and therefore be more likely to co-purify with ATP (Figure 6.2.11).

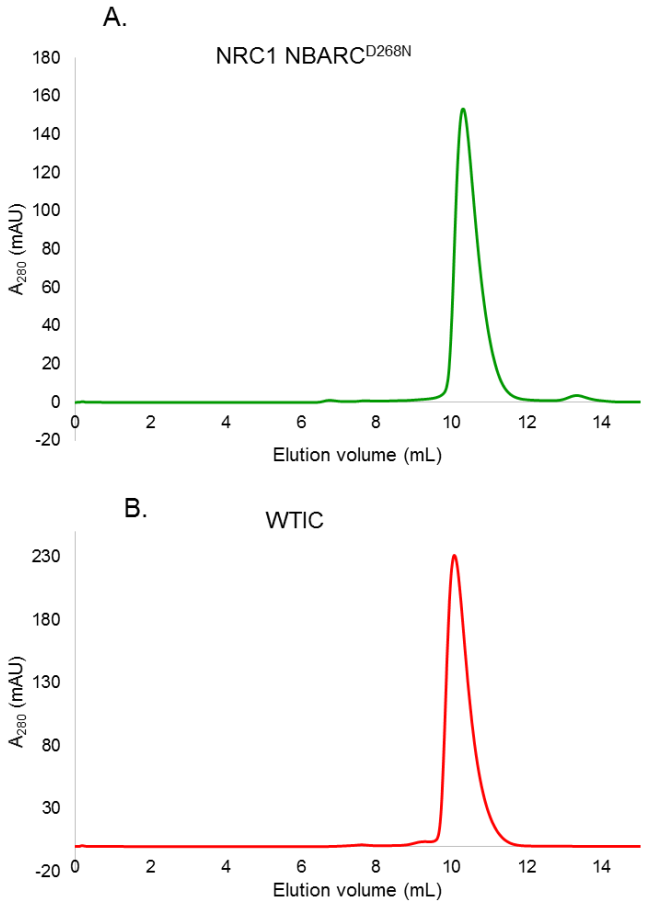


Figure 6.2.11 Analytical gel filtration of NRC1 Walker-B mutant and wild-type protein derived from insect cell expression

A. Analytical gel filtration of NRC1 NBARC^{D268N} mutant predicted to be deficient in ATPase activity purified from insect cells. Protein shows a single, well-defined peak at approximately 10.2mL.

B. Analytical gel filtration of wild-type NRC1 NBARC domain purified from insect cells. Protein elutes identically to NRC1 NBARC^{D268N}, and consistent with the major elution peak of *E. coli*-derived protein.

Mutation to the conserved aspartate in the Walker-B motif of NRC1 NBARC domain does not result in a dramatic change in oligomeric state when measured by analytical gel filtration. We predicted that this protein should be more likely to co-purify with ATP due to deficient ATPase activity, and that this trace would therefore

represent the oligomeric state of the protein bound to ATP. As we see in Section 6.2.7, this Walker-B mutation does not lead to complete abolition of ATPase activity and so further work is required to characterise the bound ligand to determine whether this trace is representative of ATP-bound protein.

Finally we tested whether MH$^{D->V}$ mutations cause activation by altering protein oligomeric state, potentially mimicking activated or ATP-bound protein. As shown in Figure 6.2.12, the introduction of this mutation in NRC1 NBARC domain did not result in changes in oligomeric state. Analytical gel filtration of this protein shows additional minor elution peaks. The elution volume of each peak and corresponding calculated molecular mass is summarised in Table 6.2.2 , with values for *E. coli*-derived wild-type protein elution peaks for comparison.

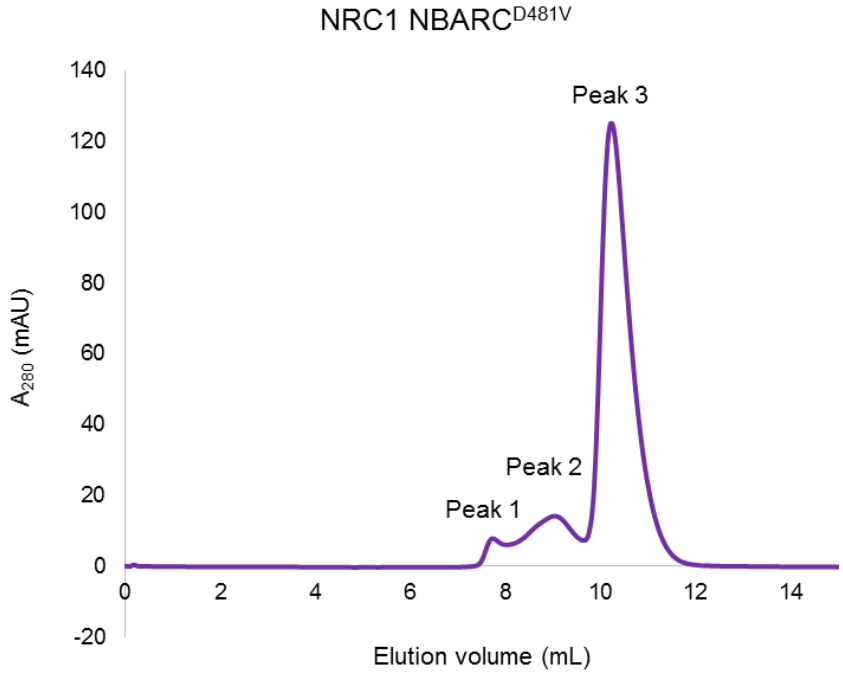


Figure 6.2.12 Analytical gel filtration of NRC1 NBARC^{D481V}

NRC1 NBARC^{D481V} mutant elutes with a predominant peak at 10.2mL, consistent with the behaviour of wild-type protein, presumably indicative of this domain in its inactive state. Additional peaks eluting at smaller volumes is indicative of multiple oligomeric states or conformations.

	Peak	Elution volume (mL)	MW (kDa)
NRC1	**1**	7.73	171.88

NBARC^{D481V}	**2**	9.044	94.99
	3	10.22	56.06
WT NRC1 NBARC	**1**	9.55	75.79
	2	10.35	52.90

Table 6.2.2 Summary of NRC1 elution peaks and corresponding molecular weights

Summary of the peaks identified for analytical gel filtration of NRC1 NBARC^{D481V} and their corresponding molecular mass. The peaks seen for WT NRC1 NBARC is included to give a comparison of potential oligomeric states.

Analytical gel filtration of NRC1 NBARC^{D481V} shows that the majority of the protein elutes in the same way as wild-type protein. We are able to identify two minor elution peaks and have calculated their approximate molecular masses, which appear to be distinct from wild-type elution. This gel filtration has been performed with a single protein purification and so we cannot confidently state that these peaks will be a consistent feature of this purified protein.

6.2.7 Aspartate to Valine mutation in NRC1's conserved 'MHD' motif leads to increased ATPase activity

As we were unable to attribute autoactivation by the MH$^{D->V}$ substitution to constitutive changes in oligomeric state, we decided to explore other aspects of predicted NBARC biochemistry to explain this dysregulation. The observation that this substitution results in preferential binding of ATP over ADP in the flax disease resistance protein M led authors to suggest that this mutation either leads to a greater affinity of ATP in these mutants, or that these proteins are deficient in ATPase activity.

We investigated whether the MH$^{D->V}$ mutation in NRC1 NBARC domain resulted in decreased ATPase activity of NRC1 NBARC^{D481V}. Malachite green assays were performed as described in Section 2.17.2 with the results summarised in Figure 6.2.13 Phosphate releasing activity of NRC1 NBARC mutants in malachite green ATPase assay .

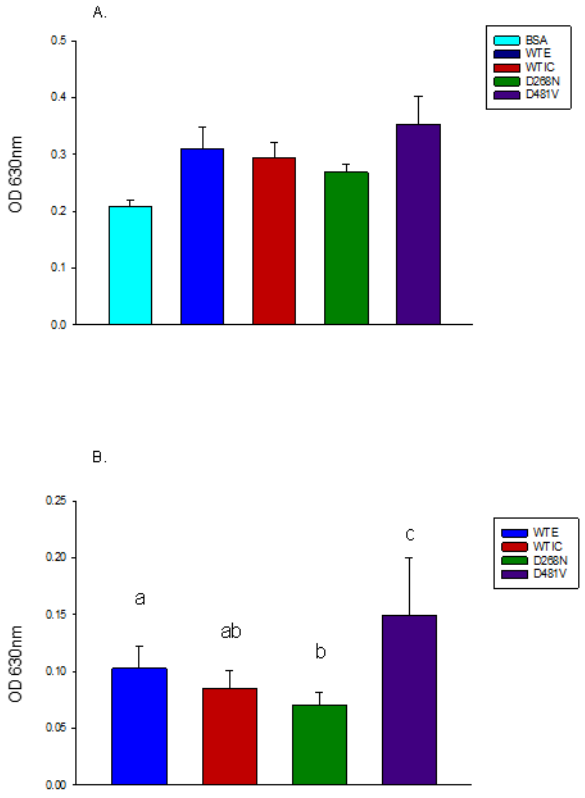


Figure 6.2.13 Phosphate releasing activity of NRC1 NBARC mutants in malachite green ATPase assay

A. Mean OD_{630} after 1 hour incubation of ATP reaction buffer with 50µM protein followed by quenching for 20 minutes with malachite green reagent. Error bars indicate standard deviation of the mean. Malachite green ATPase assays were performed twice for NRC1 $NBARC^{D481V}$ samples and three times for all other treatments.
B. As in A. but with average BSA reading subtracted for each run. A_{630} decreases over time and so subtraction of BSA background readings allows better comparison between assays. Data were analysed in Sigma Plot. Significance was determined using holm-sidak method for pairwise comparisons of the mean values. Mean phosphate release is approximately 12.11µM for *E. coli*-derived wild-type protein, 5.85uM for insect-cell derived wild-type protein, 0.519µM for $NBARC^{D268N}$ (insect-cell) and 29.89µM for $NBARC^{D481V}$ (*E. coli*).

The preliminary results from this assay indicate that rather than a decrease in ATPase activity, this mutation results in a significant increase in ATPase activity, however given the low signal in these assays replicates would increase the statistical robustness of these experiments and our confidence in this interpretation.

6.2.8 NRC1 NBARC does not interact strongly with DNA

The difficulties to date in positively identifying the signalling components of many R-proteins has led some researchers to suggest that this indicates a very short signalling cascade, making it less likely to be disrupted by random mutagenesis screens. Recently the Cann group at Durham University have published evidence

suggesting that NBARC domain of Rx interacts non-specifically with DNA[167], and that this activity may be conserved for the NBARC domains of multiple R-proteins. The source of protein for these experiments included the coiled-coil domain as well as the NBARC domain, and required the liberation of protein from inclusion bodies.

We decided to test whether solubly expressed NRC1 NBARC domain also showed an ability to interact non-specifically with DNA. We initially used surface plasmon resonance (SPR) to look for non-specific NBARC-DNA interactions.

SPR can be used to detect a wide range of macromolecular interactions. One interacting partner (bait) is attached to a metal chip and a laser aimed at an angle at the other side of the chip. A detector measures the intensity of reflected light, which is sensitive to the properties of the interface between the chip and the medium (in this case a buffer solution). Potential interacting partners can be washed over the surface of the chip, where they can bind to the bait. Binding changes the refractive index of the chip and results in a decrease in intensity of reflected light, measured in response units (RU) (Figure 6.2.14).

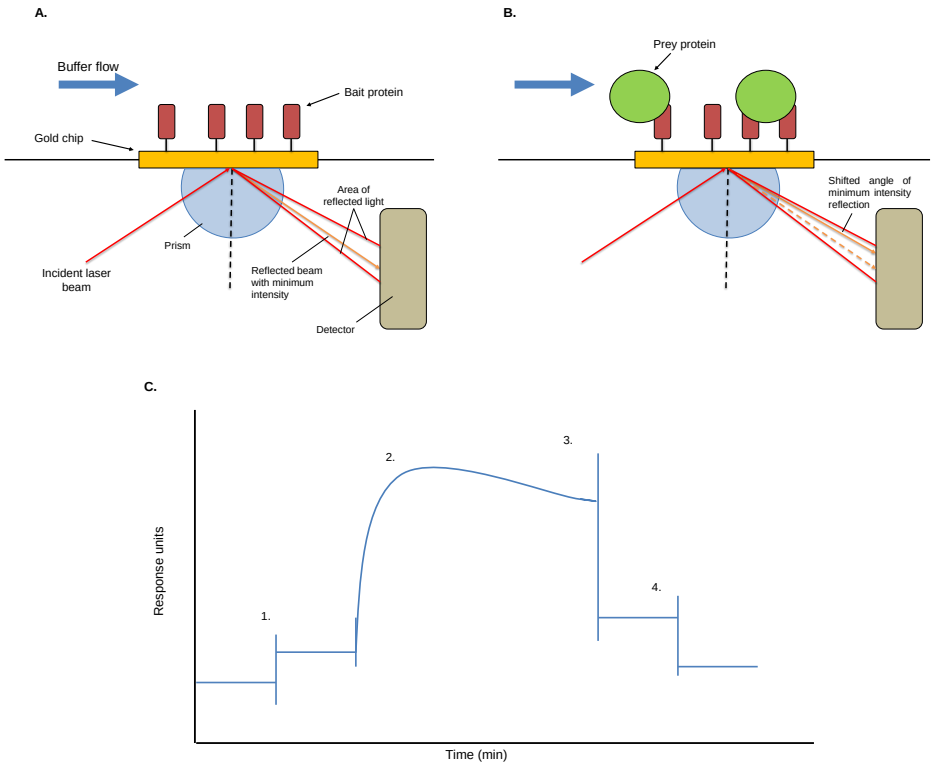


Figure 6.2.14 Schematic of an SPR experiment

A. Typical SPR set up. A flow cell containing a gold chip to which one protein-interacting partner is adhered to has buffer passed over it. A laser is shone at the other side of the chip at angles that allows total internal reflection. Using these conditions an evanescent wave is produced, which can be absorbed by free electrons to generate surface plasmons, resulting in decreased intensity of the reflected light at a specific angle. The intensity of reflected light is measured by a detector and the angle of minimum reflection can be recorded.

B. Addition of a species capable of binding to the anchored interacting partner changes the refractive index of the chip, affecting the angle at which light is maximally absorbed by the production of surface plasmons. This change in maximum absorbance is measured by the detector, which records a new angle of minimum reflection. This shift in angle of minimum light intensity is recorded in response units (RU).

C. A typical SPR sensogram.1. Loading of a potential interacting partner onto a chip leads to an increase in response units. 2. Flowing over a species capable of binding to this interacting partner gives a large increase in response units. 3. Removal of the binding species, for example by high-salt washes. 4. Stripping the chip of the affixed interacting partner to allow re-use of a single chip.

Standard SPR setup can be modified to allow easy testing of protein interactions against multiple DNA sequences. ReDCaT (Re-usable DNA capture technique) uses a short, generic DNA oligo conjugated to biotin[169]. This DNA oligo is then be bound to a streptavidin-coated SPR chip. Target DNA sequences can then be used in SPR experiments by including a short section of DNA that is complementary to the immobilised sequence, allowing these two sections to anneal. Once the test DNA has annealed to the generic DNA oligo adaptor the target protein can be injected into the flow cell and the response measured. Once an experiment is complete the cell can be washed in high-salt conditions to remove the bound protein and annealed DNA, leaving only the biotin-conjugated adaptor sequence.

Using this method we tested the ability of NRC1 NBARC domain to interact with five randomly selected double-stranded oligonucleoutides. Protein was tested at 50, 100, 500 and 1000µM.The sensogram (Figure 6.2.15) shows no signs of interaction between NRC1 NBARC domain and this selection of oligonucleotides.

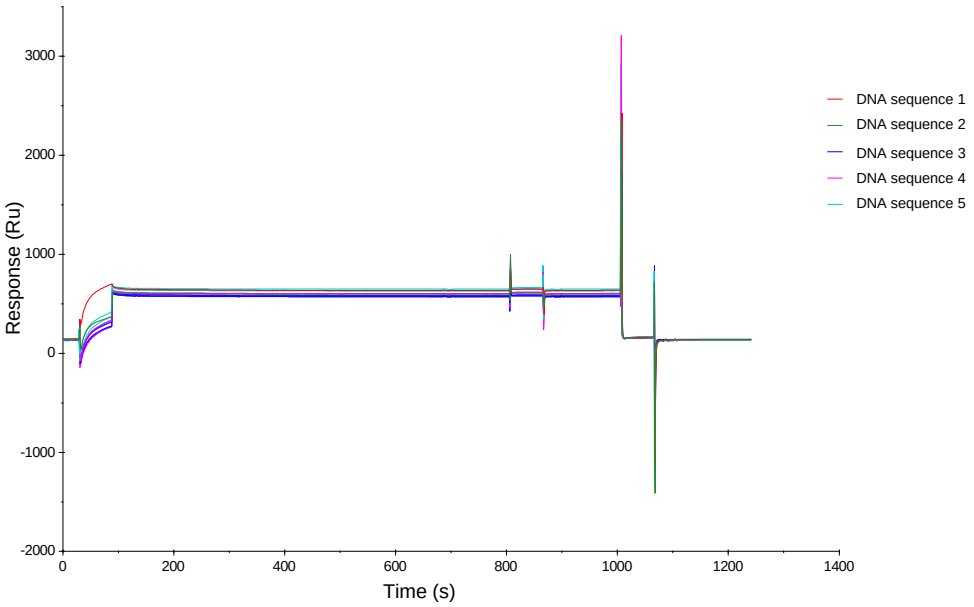


Figure 6.2.15 SPR of NRC1 NBARC interaction with double-stranded DNA

Introduction of NRC1 NBARC to flow cells containing affixed dsDNA does not lead to an increase in response units. Five separate DNA sequences ranging from 20 to 40 nucleotides in length were tested against 50, 100 and 500nM NRC1 NBARC domain. After initial binding of DNA to the SPR chip (an increase of ~500 RU) traces remain unresponsive to both the addition of protein and wash steps, indicating no binding events occur.

It is possible to calculate Rmax, the maximum expected increase in response units for saturation binding for the DNA oligos used to give a sense of scale expected for interactions between this protein and DNA. For the shortest DNA oligo against monomeric NRC1 NBARC is approximately 3300RU, and for the longest oligo 1600RU. Even factoring the potential for sub-optimal concentrations of NRC1 NBARC the lack of any signs of response strongly suggests that this protein does not interact with DNA as even partial occupancy should be registered in this system.

As we were unable to detect interactions by SPR we established a collaboration with Martin Cann's group at Durham University to test NRC1 NBARC domain using techniques they have employed to investigate Rx-DNA interactions. To investigate NRC1 NBARC interactions with DNA we used fluorescence anisotropy with a fluorescently labelled DNA oligo demonstrated to interact with Rx. Fluorescence anisotropy is based on the ability of fluorophores in solution to depolarize light. Highly mobile fluorophores are able to re-orientate rapidly in solution and so when subjected to polarized lights, the light emitted is relatively depolarised. Larger, less mobile fluorophores such as those attached to large proteins are less free to tumble in solution and so when these are subjected to polarized light, fewer are able to re-

orientate prior to photon emission and thus emit light polarized parallel to the excitation light (Figure 6.2.16).

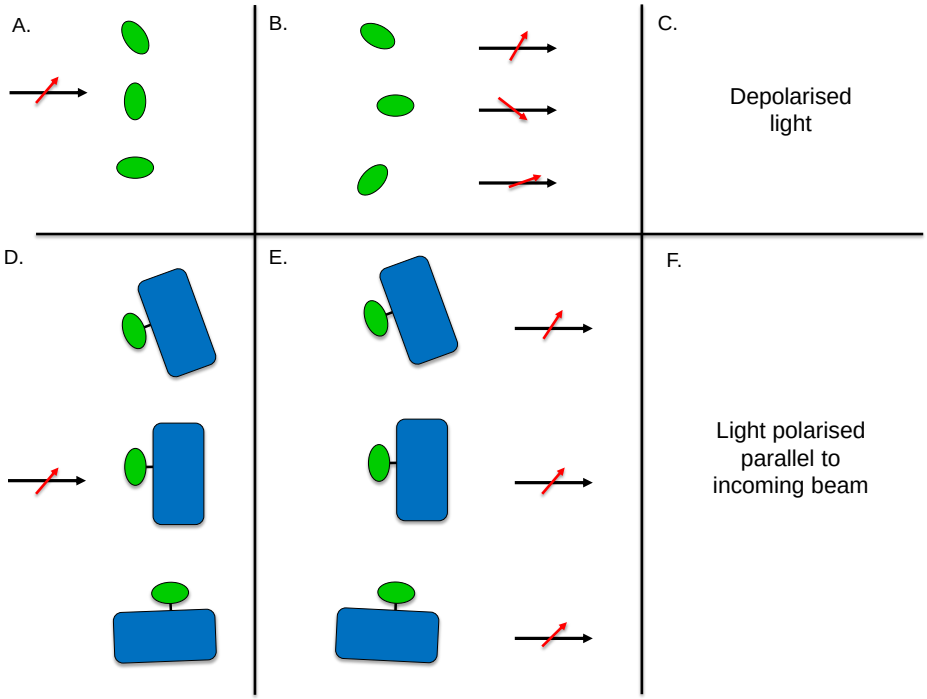


Figure 6.2.16 Schematic of fluorescence anisotropy experiments

A. Incoming light (black arrow) with specified polarisation (red arrow) is absorbed by fluorophores in solution (green ovals). **B**. Light is emitted from fluorophore after time tau (fluorescence lifetime. If rotational lifetime of the molecule (theta) is less than that of the fluorescence lifetime the molecule will emit light with a new polarisation. **C**. Measurement of polarisation of light emitted parallel and perpendicular to excitation beam determines whether depolarisation has occurred. **D**. Incoming light is absorbed by a fluorophore conjugated to a large macromolecule. **E**. Larger complexes have a longer rotational lifetime and so are less likely to re-orientate prior to photon emission. **F**. Detected light is recorded as being primarily parallel to the excitation beam and so emitted light is described as anisotropic.

If an additional protein is added to D that is capable of binding the conjugated macromolecule, rotational lifetime will increase, resulting in less depolarisation of light. If the added protein is unable to interact with the conjugated macromolecule, no change in anisotropy is recorded.

We used untagged, purified NRC1 NBARC domain to investigate its ability to prevent the depolarization of light emitted from a fluorescently labelled oligonucleotide. Initial tests indicated that NRC1 NBARC was able to reduce depolarization of emitted light significantly more than BSA (Figure 6.2.17), with no dramatic difference between interactions with single-stranded and double-stranded DNA (data not shown).

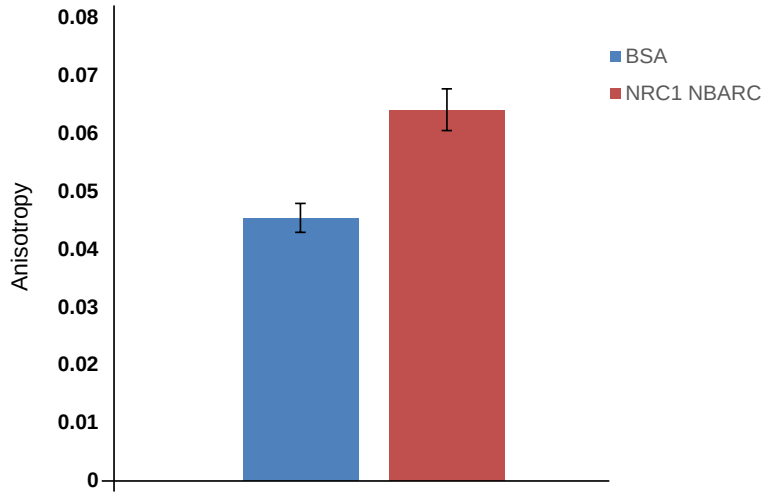
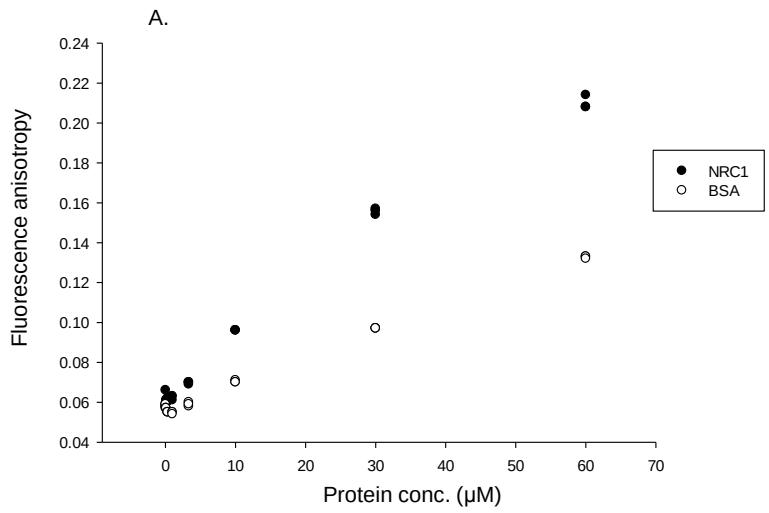


Figure 6.2.17 Fluorescence anisotropy of labelled single-stranded DNA in response to BSA and NRC1 NBARC

Single-stranded fluorescently labelled DNA treated with 10uM BSA or NRC1 NBARC. After 20 minutes incubation at room temperature in a 96-well plate samples were subjected to polarized excitation light and the polarization of the emitted light recorded. Increased anisotropy in the presence of NRC1 NBARC suggests binding to the DNA to increase rotational lifetime. Graph shows mean values and SD.

In order to try to determine the binding affinity of this protein for DNA we produced a binding curve using increasing concentrations of NRC1 NBARC against single-stranded DNA oligonucleotide (Figure 6.2.18).



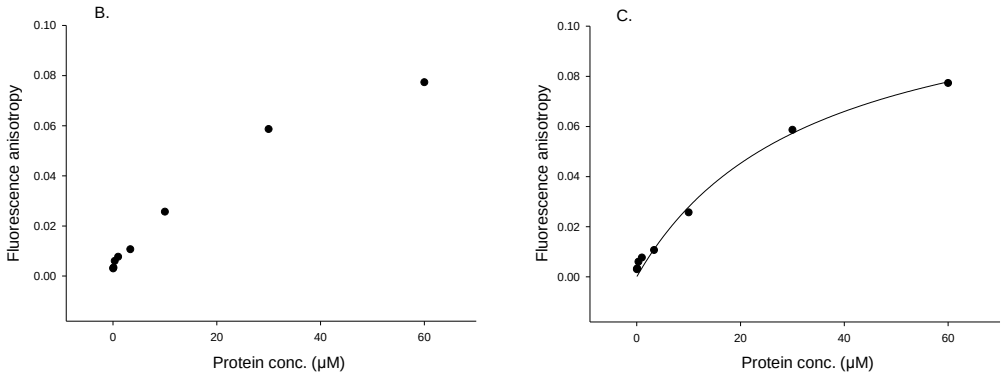


Figure 6.2.18 Binding curve for NRC1 NBARC and single-stranded DNA

A. Raw data showing changes in fluorescence anisotropy of fluorescently labelled DNA in response to increasing concentrations of NRC1 NBARC or BSA. Protein concentrations range from 0.03μM to 60μM with each concentration tested with triplicate technical replicates.
B. Averaged and BSA-subtracted fluorescence anisotropy of NRC1 NBARC interacting with single-stranded DNA from A.
C. Binding curve calculated using SigmaPlot global curve analysis, selecting for single-site binding kinetics. Derived curve gives a Kd = 33.54 +/- 8.22, and R^2 0.9886.

In this experiment we were able to see greater anisotropy in the presence of NRC1 NBARC than BSA, indicating that NRC1 NBARC may be interacting with the single-stranded DNA in this system. Subtraction of BSA signal produces a binding curve that does not reach saturation point, but shows a dose-dependency. Although it is possible to derive a logarithmic curve, the lack of saturation means that we cannot be confident that a linear interpolation may not better represent binding interactions. As the binding curve did not reach saturation we were not confident in the derived Kd for this interaction. In order to saturate the system we increased the concentration of NRC1 NBARC to 160μM. As there appeared to be negligible binding at the lowest concentrations used, we increased the starting concentration of NRC1 NBARC to 0.3μM (Figure 6.2.19).

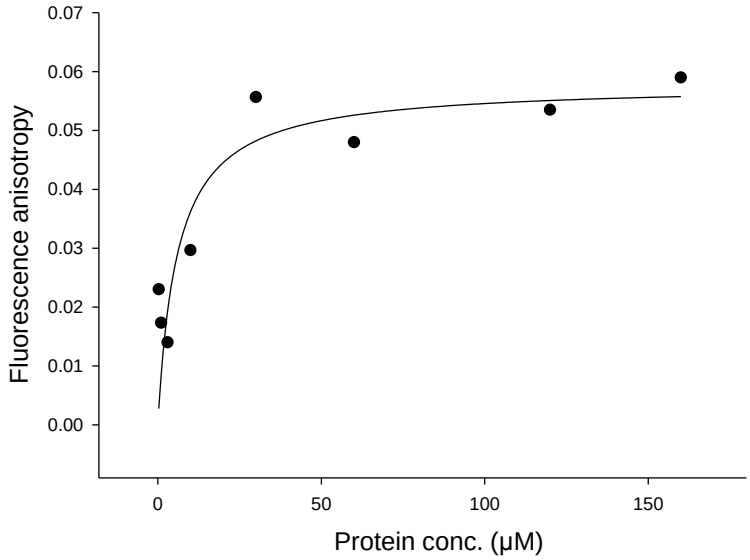


Figure 6.2.19 Saturated binding curve for NRC1 NBARC double-stranded DNA interactions

Averaged and BSA-subtracted fluorescence anisotropy in response to increasing NRC1 NBARC concentrations. Protein concentration ranges from 0.3uM to 160uM. Curve was fitted using SigmaPlot global curve analysis to give a Kd = 5.9692 +/- 3.70, with R^2 0.7271

The ability to fit a hyperbolic binding curve suggests that NRC1 NBARC binds to DNA in a single-site interaction, and we can derive a Kd of approximately 6µM in this instance. The nature of these experiments was exploratory, and so the above figures represent single experiments with technical replicates. In order to make more definitive judgements regarding binding affinities and significance of the data more replicates are required, particularly given the low signal from these experiments.

Disparity between the apparent Kd from the two binding curves makes interpretation of a likely Kd problematic. Interestingly the apparent maximum response from our saturated binding curve is approximately 0.06, which is surpassed in our previous curve without signs of plateauing- the significance of which cannot be determined without further experiments.

6.3 Discussion

6.3.1 NRC1 exhibits ATPase activity consistent with predictions. An autoactive mutant displays higher rates of ATP hydrolysis, contrary to expectations

We have used two independent methods to investigate the ability of NRC1 NBARC domain in isolation to facilitate ATP hydrolysis. Although turnover is significantly higher than BSA controls, the slow rate of hydrolysis or exchange raises the question of biological relevance. R-proteins are known to respond very rapidly to the presence of effectors, and so how likely is it that the ability to hydrolyse ATP to ADP over the course of hours acts as a bona fide regulatory mechanism? Despite low enzymatic activity relative to proteins like catalase, *in planta* investigations indicate that R-proteins require the ability to hydrolyse ATP: mutating residues in the R-protein I-2 that result in decreased rates of ATPase activity causes effector-independent cell-death[105]. This suggests that ATP hydrolysis, even with very low turnover rates, is required for R-protein regulation. Analysis of the data from linked PK/LDH assays suggests that the data are not normally distributed and so post-hoc analysis was performed to account for this non-normal distribution. Although this analysis stated that wild-type protein displays significantly enhances activity relative to negative controls, this non-normal distribution, variability in the data and low signal make a compelling case for increased replicates to improve confidence in this result.

Low ATPase activities are not entirely unexpected. I-2 and Mi-1 also display low levels of ATP turnover, as does the animal homologue Apaf-1[102, 170]. When the full-length resistance protein M was purified from yeast, bound ligand could be quantified, but the authors were unable to confidently detect ATPase activity[130]. There are multiple potential reasons for this low ATPase activity: these reactions are performed isolated from other host proteins, and so it may be that additional proteins may be required to stimulate ATPase activity, similar to GTPase regulation. Alternatively it may be that in the absence of the CC or LRR domains, ATPase activity is reduced- for example the LRR domain may facilitate nucleotide exchange. Or it may be that these proteins require ATPase activity, but in order to balance inhibition in the absence of a pathogen, with rapid responses to invasion, this activity must be low (further discussed in Section 7.2).

When investigating the basis for autoactivity in the flax resistance protein M, the authors suggest that the higher occupancy of ATP relative to ADP may be due to either increased affinity for this ligand, or a deficiency in ATPase activity. The idea

that MH$^{D->V}$ mutations may reduce the ability of R-proteins to hydrolyse ATP is consistent with the observation that functional Walker-B motifs are required to prevent autoactivation. However two observations may indicate a different role for the MHD motif in R-protein regulation: protein modelling typically places the MHD motif distal from the Walker-B motif, with numerous examples of ATPases that do not possess MHD-like motifs, and *in planta* assays of R-protein responses to mutations indicate that MHD mutations result in far more dramatic auto-activation than Walker-B mutations alone, however this does not rule out multiple roles for the MHD motif.

We have demonstrated that the isolated NBARC domain of NRC1^{D481V} displays higher rates of ATP hydrolysis than wild-type protein. This result was unexpected, and may be the result of two main factors. Firstly it is possible that the R-protein NRC1 behaves, and is regulated differently to the resistance protein M found in flax. NRC1 appears to be a downstream signalling component rather than a canonical R-protein, and so there may be differences in the way that this protein is kept in a signalling-competent "off" state. The high conservation of the MHD motif in R-proteins, particularly in the domain showing the highest level of conservation argues against this point, as presumably the motif is conserved on the basis of conserved function. A second possibility is that the observation of this autoactive mutant's ability to hydrolyse ATP comes from an isolated domain, whereas experiments into the occupancy of the flax-rust protein M comes from full-length protein. In the context of full-length protein it is possible that mutating this aspartate to valine results in changes in domain orientation or affinity. Increased ATP occupancy could be due to increase nucleotide release- the observation that the NBARC domain co-purifies with ADP suggests a high affinity, or very slow ligand release, and the increase in ATPase activity, although significantly greater than wild-type, leads to less than a 2-fold increase. If the rate of ADP release in MH$^{D->V}$ mutants is increased more than the rate of ATPase activity then we would expect to see a greater proportion of R-protein bound to ATP than wild-type, with this disparity decreasing over time.

6.3.2 Mechanistic basis for increased ATPase activity in NRC1 NBARC^{D481V}

There are multiple ways in which changing the aspartate residue near the ligand binding pocket may affect the rate of ATP hydrolysis. One potential explanation is that this mutation leads to destabilisation of the NBARC●ADP complex, resulting in more rapid nucleotide exchange. Similarly, the replacement of the hydrophilic

aspartate with a hydrophobic valine may destabilise the interactions between the ARC2 and NB domain, resulting in a more open conformation of the ligand binding pocket, again facilitating nucleotide exchange. Alternatively the $MH^{D->V}$ mutation may cause an increase in ATPase activity indirectly, by its effect on the neighbouring histidine. From the preliminary crystal structure of NRC1 NBARC we hypothesise that the histidine of the MHD motif is likely to play two main roles in regulating NBARC activity. Firstly, its size and proximity to the β-phosphate of ADP suggest that this conformation is unfeasible in the presence of ATP due to steric clashes, and so one role of this histidine may be in ligand discrimination, and driving conformational changes of the protein. Secondly both the size of the residue and its charge suggests it may be required to act as a locking mechanism. If the difference between "on" state NBARC domain and "off" state is the positioning of the ARC2 domain relative to the ARC1/NB domain, then it may be that the charge-charge interactions between the histidine and β-phosphate of ADP are required to prevent the stochastic movements of the ARC2 domain, which may mimic the presence of bound ATP.

The absolute conservation of this central histidine supports the idea of a locking mechanism, as ligand discrimination by steric clashes could equally be achieved by the side chain of phenylalanine. If this is the case, then the role of the negatively charged aspartate may be to keep histidine positioned in a way that allows the histidine-B-phosphate interaction to occur- for example by stabilising a dis-favoured rotamer by charge-charge interactions. Investigations by van Ooijen *et al* suggest that the role of this aspartate may be to stabilise the helix on which the central histidine sits, therefore facilitating the histidine-beta-phosphate interactions[110], which may be the case for NRC1. Homology modelling of I-2 NBARC domain on Apaf-1 in this paper positions the MHD aspartate sidechain away from the central histidine, where it is unlikely to interact directly with the side chain of Histidine. The preliminary crystal structure of NRC1 NBARC suggests that these two side-chains may be positioned more closely, supporting a more direct role for this residue, however we cannot rule out the effect that this substitution may have on secondary structure elements in this region.

6.3.3 NRC1 signalling repression cannot be explained simply by NBARC sequestration of the coiled-coil domain

The simplest explanation to explain repression of R-protein signalling in the absence of a pathogen is the idea that the signalling component (the coiled-coil domain) is normally held in a conformation that does not allow interaction with the

necessary components to trigger a response. In the model that we have been exploring this is performed by a complex forming between the coiled-coil and NBARC domains. Investigations using purified protein in the absence of an additional factors indicate that the CC and NBARC domains or NRC1 do not form a stable interaction, leading to a few potential explanations. Firstly it is possible that complex formation requires the LRR of the protein. As we were unable to produce soluble LRR in this study, we cannot explore whether the addition of the LRR would lead to complex formation. From an evolutionary perspective it should be possible to test this theory, if enough NRC1 homologs can be found. Complex formation requires compatible protein-protein interactions, and for these interactions to remain stable there should be indications of either stabilising selection in specific regions, or co-evolution between the three domain interaction interfaces. Early indications of the feasibility of this interaction arrangement could come in the form of experiments into the presence of transient interactions between the CC and NBARC domains using sensitive methods such as SPR or isothermal calorimetry.

It is possible that these two domains do form a complex *in planta* without requiring stabilisation by the LRR, but requiring an additional bridging factor. We have yet to test this hypothesis- expression of the two domains *in planta* followed by co-IP and mass spectrometry may be able to identify such a bridging element. Common false-positive results from co-IP experiments include molecular chaperones, particularly if expression is driven by a high-expression constitutive promoter such as the CaMV 35S promoter. However, it is entirely feasible that such a chaperone may be the required stabilising factor- thus discriminating between expression artefact and legitimate interactor may be difficult.

Sequestration of the CC domain of NRC1 may not be required to repress the signalling activities of NRC1. Work by Sueldo *et al* indicates that overexpression of the NBARC domain, not the coiled-coil domain, of NRC1 leads to cell death. This is similar to Rx, where it was found that the NBARC domain also causes cell death when expressed alone. Therefore it may be that in the case of NRC1 there is no requirement to form a CC-NBARC complex to prevent CC interactions with downstream signalling components.

6.3.4 NBARC-DNA interactions

Recent publication by Fenyik *et al* describing the DNA-binding capabilities of Rx led us to speculate whether such activity could be observed in the solubly expressed NBARC domain of NRC1. Initial experiments using SPR was unable to detect any

signs of binding, despite the high sensitivity of this method and range of oligonucleotides used. This lack of apparent interaction could be due to multiple experimental factors, and so we worked with the Cann group using their established protocols to explore whether NRC1 NBARC interacts with DNA in their system.

Our preliminary data suggest that NRC1 NBARC is able to bind to single-stranded DNA (ssDNA) as measured by the changes in the ability for this labelled DNA to depolarise light when excited by a linearly polarised source. We were able to derive two apparent Kd values for single stranded DNA from an unsaturated and saturated binding curve, with approximately five-fold difference between these two values. If the interaction between NRC1 and DNA is as weak as suggested by our first binding curve it may explain why we were unable to see an interaction with SPR. Although SPR is a very sensitive method to detect interactions the best results are typically obtained using a protein concertation range 0.1 to 10 fold higher than the predicted Kd. As this would mean using purified protein at a concentration approaching 300µM, or approximately 6000 fold more than was used, it may not be surprising that we were unable to detect significant signal. Despite this, the large increase in RU predicted from NRC1 binding would suggest that weak interactions should be detectable. Future experiments using this system should therefore use increased NBARC concentrations for optimal data collection. Using such high concentrations of protein for SPR may lead to non-specific interactions between protein and the chip, and so non-specific binding must first be tested to determine the validity of such an experiment.

Although our fluorescence anisotropy experiments suggest NBARC-DNA interactions the sensitivity of this technique to factors in addition to bone fide interactions means that validation with complimentary techniques is advised. For example increasing protein concentration can affect the turbidity of solution, which will slow the rate of molecular tumbling and increase anisotropy. This factor can be compensated for by the use of appropriate controls, such as our use of BSA, which also showed increasing responses with concentration. Our subsequent investigations into the oligomeric state of NRC1 NBARC suggests that such experiments may require additional controls to improve confidence in interpretation. Changes in fluorescence anisotropy are affected by the molecular mass of the interacting partner, with the binding of larger macromolecules having a greater effect on anisotropy than smaller molecules. At the time of these experiments we believed NRC1 NBARC to act as a monomer in solution, giving it a molecular mass of approximately 40kDa. In this case the use of BSA (MW 66kDa) would give a

conservative binding affinity once this control was subtracted from experimental signal. If NRC1 NBARC is acting as a dimer in solution then the subtraction of non-specific BSA control of 66kDa may overestimate the effect of specific binding of an 80kDa dimer. Ideally such experiments should be performed with point mutants that are defective in DNA binding. As protein-DNA interfaces can be extensive, a simple single point mutation may not suffice, and so considerable work would be required to identify the smallest number of mutations abolishing DNA interactions without affecting protein stability. Such an extensive screen is beyond the scope of this project, yet may be essential to fully characterise NBARC-DNA interactions to supplement existing data regarding this novel R-protein activity.

7

General discussion

The rapid increase in confirmed or putative R-genes in plant genomes, facilitated by cheaper, faster sequencing, and tools such as RenSeq[171], is providing us with new resources to tackle virulent phytopathogens. As these new R-genes are cloned and sequenced they also allow us to evaluate our models for R-protein behaviour, and open avenues for crop improvement. For example identification of non-canonical NB-LRRs has let us explore the role of decoy modules, potentially raising the possibility for "plug and play" R-proteins with directed specificity. Likewise, by studying the function of individual domains in the context of R-protein biology we can expand our models for R-protein behaviour to inform deployment in the field and rational improvements to individual proteins.

Models from *in planta* and genetic investigations allow us to explore many aspects of R-protein biology. However, complementing these investigations with biochemical studies *in vitro* can give us a more detailed view of the mechanistic basis for regulation and signalling of these agriculturally important proteins.

This project aimed to increase our understanding of some of the biochemical and structural properties of solanaceous CNL R-proteins, however given the limited published resources for such studies a key focus was development of tools to allow such experiments to be performed.

7.1 Development of tools for future experiments

At the time of writing there were very few examples of R-proteins or their domains suitable for biochemical investigations. Difficulties in expressing protein domains, either solubly or refolded from inclusion bodies, combined with the difficulties in detecting enzymatic activity has resulted in our understanding of the biochemistry of these proteins resting on very few examples, from a diverse family. Furthermore commonly used assays, such as filter-binding assays that are required to monitor very low levels of activity are susceptible to influences of poorly folded or unstable proteins, and thus need to be complemented with additional biochemical techniques.

7.1.1 Purification of soluble R-protein domains

Central to this project was the development of material to allow further characterisation of R-proteins and their subdomains *in vitro*. Biochemical investigations of R-proteins, from oligomeric state to enzymatic activities, and putative interactions have been limited by the difficulties of obtaining sufficient yields

of suitable protein for these types of experiment. We pursued extensive screening efforts to obtain suitable material from *E. coli* and eukaryotic expression systems. We utilised both advanced methods to screen constructs that would not necessarily be included in traditional screening methods, and traditional construct design with medium-throughput screens of expression conditions, both in house and with the Oxford Protein Production Facility (OPPF).

We have not been able to rationalise some of our successful expression screens, for example why specific constructs seem to have an absolute expression strain requirement, distinct from similar constructs. Thus researchers wishing to express new R-protein domains in *E. coli* will most likely need to carry out similar expression screens. However we have been able to increase the number of solubly expressing R-protein domains, which may aid future construct design.

To date our ideas regarding the structure and behaviour of R-protein CC domains *in vitro* rely on two examples- MLA10 and Rx. To these published proteins we have been able to add four new examples of solanaceous CC domains for future work: R3a, I-2, NRC1-like and NRC1.

More significantly we have been able to demonstrate the soluble expression of the NBARC domain of the plant CNL NRC1. This provides a unique tool for investigating the behaviour of this domain in isolation. Just as we hope that by increasing the number of coiled-coils that can be expressed in *E. coli* may aid future research, this new construct may help inform design of additional R-protein NBARC constructs to allow characterisation of a diverse range of these regulatory domains. The ability to produce wild-type and specific point mutants in this protein can allow future investigations into different aspects of R-protein behaviour and regulation.

7.1.2 Generation of R3a-detecting antibodies

One of the more frustrating aspect of this project has been the inability to confidently immunoprecipitate untagged R3a from plant material. We had hoped that the production of specific antibodies would allow us to probe proteins interacting with signalling-competent R3a to help elucidate important host proteins involved in R-protein activation and signalling. However this has not been possible with the techniques used here, in spite of a series of screens designed to improve experimental conditions.

Despite this we have demonstrated that these antibodies can detect the coiled-coil domain of R3a when expressed as part of the full-length protein *in planta,* and when

purified as a single domain from *E. coli*. These antibodies have further been used to show mutant stability and accumulation for the homologous protein I-2, demonstrating that the use of these antibodies is not restricted to immunoprecipitation experiments.

As the R3a-AVR3a pathosystem is an area of ongoing interest we hope that these antibodies can be of use to researchers in this field as a tool to further elucidate the regulation of R3a-dependent HR signalling, through techniques such as native-PAGE, far-western blotting, or to identify co-purification of native R3a with tagged putative interactors.

7.2 Preliminary structural characterisation of NRC1 NBARC domain

Understanding the mechanistic basis for R-protein regulation and the effect of mutations to conserved motifs in the NBARC domain has typically relied on homology modelling from animal homologues, which may demonstrate difference in regulatory mechanisms- for example NBARC-mediated apoptosomes formation seen in many animal NBARC proteins has yet to be demonstrated as a feature of the NBARC domains of R-proteins.

Despite these potential differences the preliminary structure of NRC1 NBARC domain presented here indicates significant structural similarity to the ADP-bound form of Apaf-1 NBARC domain, as opposed to His-Asp phosphotransferases[166] thus validating the assumptions made in prior papers about this structural conservation for example[109, 110, 134, 172]. If a complete structure of this NBARC domain can be produced in future it can form a more reliable template for future R-protein homology models, particularly those for which experimental support is not available. I-2 and Mi-1 both have experimentally determined ATPase activity, and thus functional conservation can be used to supplement low sequence conservation. This activity has not been validated outside these two examples, and so it is possible that there exist subfamilies of NBARC proteins resembling and behaving like CED-4, which does not have demonstrated ATPase activity, and those acting like Apaf-1. Due to the higher sequence identity between plant NBARC domains, a completed structure may increase the accuracy of models of these domains, particularly those that deviate from characterised examples[77, 110]. Even if this preliminary ADP-bound structure remains incomplete it can be used to identify gross conformational changes if researchers are able to solve the structure of an ATP-

bound or active plant NBARC domain, to better understand the switch mechanism of these proteins.

Of more direct importance is the identification co-purifying ADP, and the NBARC residues involved in this interaction. Co-purification of this ligand indicates a high affinity of plant NBARC domains for adenosine nucleotides, as has been demonstrated for animal homologs and full-length R-proteins[98, 130, 173, 174]. Such a high affinity suggests that the low hydrolysis rate of ATP *in vitro* may not be the result of a long-lived active structure (ATP-bound NBARC), but due to low ligand turnover, however further experiments are required to validate this hypothesis. If true this may explain the apparent positive role of the LRR in HR-signalling- if the LRR forms interactions with the ARC2 subdomain as has been suggested, then activation may result in dissociation of the LRR from this domain, not simply to passively allow nucleotide exchange, but to actively "pull" the ARC2 subdomain away from the ligand binding site. This would then allow ATP binding, which could stabilise the resulting "on" conformation.

Alternatively it may be that slow ATP hydrolysis is a result of low enzymatic activity, and not limited by nucleotide exchange. ATPase activity has been demonstrated to be required for proper regulation, as reduced ATPase activity has been linked to autoactivity *in planta*. There are two possible reasons why this activity is required, and required to be slow. Firstly as the protein is being synthesised it is likely to bind to ATP, given the high ATP to ADP ratio in cells, although data regarding relative affinities is necessary for greater confidence in this interpretation. This would result in R-proteins synthesised in an "on" conformation, which would require conversion to an "off" state to prevent aberrant cell death. Secondly little is known about the protein dynamics involved in cell-death, though these processes are often depicted as single-molecule events. If cell-death initiation requires multiple copies of an R-protein to be in an "on" state (analogous to the requirements for apoptosomes formation in animals), then rapid hydrolysis of ATP may reduce the likelihood of this minimum activation concentration being reached when required. This may explain dose-dependent effects seen with the overexpression of certain R-proteins. Under a constitutively active promoter such as the CaMV35S protein synthesis occurs more quickly, and so if ATP hydrolysis is low, as suggested here, then enough nascent R-proteins are formed bound to ATP before hydrolysis can re-set the proteins, leading to cell death.

7.3 Putative roles for the MHD motif

Work with the flax rust resistance protein M expressed in yeast indicated that introduction of $MH^{D->V}$ mutations in this protein leads to higher occupancy of ATP relative to ADP when compared with wild-type protein[130]. In this work the authors suggest two explanations for this: either this mutation results in decreased ATPase activity, or an increased affinity for ATP, however this question remains unsolved due to issues demonstrating ATPase activity. Preliminary data we present here suggests that rather than decreasing ATPase activity, mutations to the aspartate of the MHD motif of NRC1 may result in increased ATP hydrolysis.

We hypothesise that this increase in ATPase activity is not due to increased ATP hydrolysis, but is due to disruption of the interactions between the conserved histidine and the β-phosphate of ADP. This could prevent the "locking" of the ARC2 domain in place, and allow increased nucleotide exchange. Indirect evidence supports the idea that the MHD motif is not directly involved in hydrolysis. Firstly, double mutants with reduced ATP hydrolysis and MHD mutations do not see an additive effect[175]. Secondly the crystal structure of NRC1 suggests that this aspartate is unlikely to be close enough to interfere with the Walker-B motif, known to be required for ATP hydrolysis. Finally many ATPases exist without this MHD motif, suggesting that this motif is dispensable for ATP hydrolysis in STAND proteins. In reference to the work performed with the TNL protein M, increased nucleotide exchange could result in greater ATP occupancy if the increased rate of exchange is greater than the rate of ATP hydrolysis, particularly if a higher affinity for ATP over ADP can be demonstrated, as has been suggested for Apaf-1[176].

7.4 Models for R-protein regulation

Together the functional and structural studies we have performed favour the interpretation put forward by Chattopadhyaya *et al*[109] and van Ooijen *et al*[175] amongst others, that the NBARC domain in the off state is bound to ADP and resembles ADP-bound Apaf-1. The positioning of the β-phosphate and MHD motif supports the interpretation that in the "on" state the ARC2 domain is pushed away from the NB-ARC1 domains, forming a more elongated structure.

The crystal structure of NRC1 NBARC domain also supports bioinformatic analyses suggesting that the invariant histidine is the most important residue in the

MHD motif, given its direct interactions with the ligand. Furthermore we suggest that this histidine has two functions in R-protein regulation- both as a ligand sensor based on steric constraints, and as a locking mechanism as a simple sensor activity indicated by the crystal structure of NRC1 could be performed by uncharged species such as phenylalanine.

We do not see evidence for sequestration of the CC by the NBARC as a general mechanism for signal repression. It is possible that the methods we employed are not sensitive enough to detect interactions, and could be tested by more sensitive isothermal calorimetry or SPR experiments. However such transient interactions are unlikely to be sufficient to prevent aberrant cell death. It may be that additional factors are required to form a stable complex, such as the LRR domain, or some other interacting partner, however we have not been able to generate the materials required to test this.

Alternative activities for the CC domain have been proposed- for example the CC of NRC1 may be involved in controlling localisation, as has been shown for a number of CC domains[90] rather than activating signalling, and so does not require intramolecular interactions to repress cell death. This is supported by work from Sueldo *et al* suggesting that NRC1 may function in a similar way to Rx, with the NB domain and not the CC domain responsible for signalling cell death[177]. For NRC1 an alternative model whereby the CC controls localisation, either through post-translational modification or interactions with host proteins may be more appropriate. In this model activation may result in re-localisation, and recognition of the NBARC in an open conformation, stabilised by ATP by downstream signalling components may explain the behaviours we see.

Whether a lack of direct CC-NBARC interactions is a general feature in multiple R-protein or is specific to NRC1 is not known. It may be that this alternative model reflects the behaviour of separate classes of R-protein (those using the N-terminus for signalling and those using the N-terminus for localisation or regulation), or may be a feature unique to NRC1-like proteins. NRC1 does not act as a traditional R-protein, but as a downstream signal protein, and so may undergo fundamentally different regulation. The activity of NRC1 downstream of other immune receptors does not preclude the activities identified here from being generally applicable to “true” R-proteins. NRC1 shares the same domain architecture as canonical CNLs with the same conserved motifs identified in bona fide R-proteins, and so is unlikely to behave in a completely unrelated way.

7.5 Future perspectives

There are many fundamental questions regarding the structure and behaviour of CNLs that remain unanswered. The role of the CC domain in controlling oligomeric state remains contentious as the two published examples give contradictory results, with Rx CC acting as a monomer and MLA10 CC forming a dimer in solution[86, 92]. The materials we have produced may give insight into whether CC monomers or dimers are more common for solanaceous R-proteins. Furthermore structural characterisation of these domains could help elucidate whether there is a typical CC fold, and the role of known motifs in CC activity.

From our medium-throughput screens we were able to solubly express three of five solanaceous CC domains, when based on a construct known to express solubly in *E. coli*. Using a similar approach the construct boundaries from NRC1 NBARC expression may likewise improve the ability to produce solubly-expressing NBARC domains from other R-proteins, allowing investigations into the variability of R-protein behaviour *in vitro* such as rates of ATP hydrolysis, and potential protein-protein interactions.

The ability to produce the NRC1 NBARC $MH^{D->V}$ mutant has led to an unexpected finding, which should be further characterised. Replication of existing assays is required to determine if the observed increase in ATPase activity is significant. The role of this motif in controlling enzymatic activity, ligand binding and protein conformation has yet to be elucidated. A combination of biochemical assays and spectroscopic methods, such as SAXS to compare protein flexibility, should allow discrimination between changes in nucleotide exchange and ATP hydrolysis.

Although we have not been able to comprehensively describe the NBARC domain of NRC1, or the full behaviour of CNL CC domains *in vitro*, this project has yielded valuable resources that can directly or indirectly improve our understanding of these important and robustly regulated proteins.

8 References

[1] Takken, F. L. W., and Goverse, A. (2012) How to build a pathogen detector: structural basis of NB-LRR function, *Curr Opin Plant Biol 15*, 375-384.

[2] Jones, J. D., and Dangl, J. L. (2006) The plant immune system, *Nature 444*, 323-329.

[3] Turkensteen, L. J., Flier, W. G., Wanningen, R., and Mulder, A. (2000) Production, survival and infectivity of oospores of Phytophthora infestans, *Plant Pathology 49*, 688-696.

[4] Flor, H. H. (1942) Inheritance of pathogenicity in Melampsora lini, *Phytopathology 32*, 653-669.

[5] Person, C., Samborski, D. J., and Rohringer, R. (1962) The gene-for-gene concept, *Nature 194*, 561-562.

[6] Alcazar, R., von Reth, M., Bautor, J., Chae, E., Weigel, D., Koornneef, M., and Parker, J. E. (2014) Analysis of a plant complex resistance gene locus underlying immune-related hybrid incompatibility and its occurrence in nature, *PLoS Genet 10*, e1004848.

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

[8] Chapman, S., Stevens, L. J., Boevink, P. C., Engelhardt, S., Alexander, C. J., Harrower, B., Champouret, N., McGeachy, K., Van Weymers, P. S., Chen, X., Birch, P. R., and Hein, I. (2014) Detection of the virulent form of AVR3a from Phytophthora infestans following artificial evolution of potato resistance gene R3a, *PLoS One 9*, e110158.

[9] Harris, C. J., Slootweg, E. J., Goverse, A., and Baulcombe, D. C. (2013) Stepwise artificial evolution of a plant disease resistance gene, *Proc Natl Acad Sci U S A 110*, 21189-21194.

[10] Ferrari, S., Savatin, D. V., Sicilia, F., Gramegna, G., Cervone, F., and Lorenzo, G. D. (2013) Oligogalacturonides: plant damage-associated molecular patterns and regulators of growth and development, *Front Plant Sci 4*, 49.

[11] Kunze, G., Zipfel, C., Robatzek, S., Niehaus, K., Boller, T., and Felix, G. (2004) The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants, *Plant Cell 16*, 3496-3507.

[12] Felix, G., Duran, J. D., Volko, S., and Boller, T. (1999) Plants have a sensitive perception system for the most conserved domain of bacterial flagellin, *Plant J 18*, 265-276.

[13] BARBER, M. S. B., R. E.; RIDE, J.P. (1989) Chitin oligosaccharides elicit lignification in wounded wheat leaves, *Physiological and molecular Plant Pathology 34*, 3-12.

[14] Hayashi, F., Smith, K. D., Ozinsky, A., Hawn, T. R., Yi, E. C., Goodlett, D. R., Eng, J. K., Akira, S., Underhill, D. M., and Aderem, A. (2001)

The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5, *Nature 410*, 1099-1103.
[15] Gomez-Gomez, L., and Boller, T. (2000) FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis, *Mol Cell 5*, 1003-1011.
[16] Gomez-Gomez, L., Bauer, Z., and Boller, T. (2001) Both the extracellular leucine-rich repeat domain and the kinase activity of FSL2 are required for flagellin binding and signaling in Arabidopsis, *Plant Cell 13*, 1155-1163.
[17] Schulze, B., Mentzel, T., Jehle, A. K., Mueller, K., Beeler, S., Boller, T., Felix, G., and Chinchilla, D. (2010) Rapid heteromerization and phosphorylation of ligand-activated plant transmembrane receptors and their associated kinase BAK1, *J Biol Chem 285*, 9444-9451.
[18] Gomez-Gomez, L., Felix, G., and Boller, T. (1999) A single locus determines sensitivity to bacterial flagellin in Arabidopsis thaliana, *Plant J 18*, 277-284.
[19] Spallek, T., Beck, M., Ben Khaled, S., Salomon, S., Bourdais, G., Schellmann, S., and Robatzek, S. (2013) ESCRT-I mediates FLS2 endosomal sorting and plant immunity, *PLoS Genet 9*, e1004035.
[20] Frescatada-Rosa, M., Robatzek, S., and Kuhn, H. (2015) Should I stay or should I go? Traffic control for plant pattern recognition receptors, *Current Opinion in Plant Biology 28*, 23-29.
[21] Sun, Y., Li, L., Macho, A. P., Han, Z., Hu, Z., Zipfel, C., Zhou, J. M., and Chai, J. (2013) Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex, *Science 342*, 624-628.
[22] Mueller, K., Bittel, P., Chinchilla, D., Jehle, A. K., Albert, M., Boller, T., and Felix, G. (2012) Chimeric FLS2 receptors reveal the basis for differential flagellin perception in Arabidopsis and tomato, *Plant Cell 24*, 2213-2224.
[23] Mentlak, T. A., Kombrink, A., Shinya, T., Ryder, L. S., Otomo, I., Saitoh, H., Terauchi, R., Nishizawa, Y., Shibuya, N., Thomma, B. P., and Talbot, N. J. (2012) Effector-mediated suppression of chitin-triggered immunity by magnaporthe oryzae is necessary for rice blast disease, *Plant Cell 24*, 322-335.
[24] Bolton, M. D., van Esse, H. P., Vossen, J. H., de Jonge, R., Stergiopoulos, I., Stulemeijer, I. J., van den Berg, G. C., Borras-Hidalgo, O., Dekker, H. L., de Koster, C. G., de Wit, P. J., Joosten, M. H., and Thomma, B. P. (2008) The novel Cladosporium fulvum lysin motif effector Ecp6 is a virulence factor with orthologues in other fungal species, *Mol Microbiol 69*, 119-136.
[25] Sanchez-Vallet, A., Saleem-Batcha, R., Kombrink, A., Hansen, G., Valkenburg, D. J., Thomma, B. P., and Mesters, J. R. (2013) Fungal effector Ecp6 outcompetes host immune receptor for chitin binding through intrachain LysM dimerization, *Elife 2*, e00790.
[26] Song, J., Win, J., Tian, M., Schornack, S., Kaschani, F., Ilyas, M., van der Hoorn, R. A., and Kamoun, S. (2009) Apoplastic effectors secreted by two unrelated eukaryotic plant pathogens target the tomato defense protease Rcr3, *Proc Natl Acad Sci U S A 106*, 1654-1659.

[27] Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I., Bulone, V., Cai, G., Cakir, C., Carrington, J. C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M. A., Fugelstad, J., Gilroy, E. M., Gnerre, S., Green, P. J., Grenville-Briggs, L. J., Griffith, J., Grunwald, N. J., Horn, K., Horner, N. R., Hu, C. H., Huitema, E., Jeong, D. H., Jones, A. M., Jones, J. D., Jones, R. W., Karlsson, E. K., Kunjeti, S. G., Lamour, K., Liu, Z., Ma, L., Maclean, D., Chibucos, M. C., McDonald, H., McWalters, J., Meijer, H. J., Morgan, W., Morris, P. F., Munro, C. A., O'Neill, K., Ospina-Giraldo, M., Pinzon, A., Pritchard, L., Ramsahoye, B., Ren, Q., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D. C., Schumann, U. D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., van de Vondervoort, P. J., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S., Fry, W., Meyers, B. C., van West, P., Ristaino, J., Govers, F., Birch, P. R., Whisson, S. C., Judelson, H. S., and Nusbaum, C. (2009) Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans, *Nature 461*, 393-398.

[28] Whisson, S. C., Boevink, P. C., Moleleki, L., Avrova, A. O., Morales, J. G., Gilroy, E. M., Armstrong, M. R., Grouffaud, S., van West, P., Chapman, S., Hein, I., Toth, I. K., Pritchard, L., and Birch, P. R. J. (2007) A translocation signal for delivery of oomycete effector proteins into host plant cells, *Nature 450*, 115-+.

[29] Tyler, B. M., Kale, S. D., Wang, Q., Tao, K., Clark, H. R., Drews, K., Antignani, V., Rumore, A., Hayes, T., Plett, J. M., Fudal, I., Gu, B., Chen, Q., Affeldt, K. J., Berthier, E., Fischer, G. J., Dou, D., Shan, W., Keller, N. P., Martin, F., Rouxel, T., and Lawrence, C. B. (2013) Microbe-independent entry of oomycete RxLR effectors and fungal RxLR-like effectors into plant and animal cells is specific and reproducible, *MPMI 26*, 611-616.

[30] Wawra, S., Djamei, A., Albert, I., Nurnberger, T., Kahmann, R., and van West, P. (2013) In vitro translocation experiments with RxLR-reporter fusion proteins of Avr1b from Phytophthora sojae and AVR3a from Phytophthora infestans fail to demonstrate specific autonomous uptake in plant and animal cells, *MPMI 26*, 528-536.

[31] Dou, D., Kale, S. D., Wang, X., Jiang, R. H., Bruce, N. A., Arredondo, F. D., Zhang, X., and Tyler, B. M. (2008) RXLR-mediated entry of Phytophthora sojae effector Avr1b into soybean cells does not require pathogen-encoded machinery, *Plant Cell 20*, 1930-1947.

[32] Yaeno, T., and Shirasu, K. (2013) The RXLR motif of oomycete effectors is not a sufficient element for binding to phosphatidylinositol monophosphates, *Plant Signal Behav 8*, e23865.

[33] Yaeno, T., Li, H., Chaparro-Garcia, A., Schornack, S., Koshiba, S., Watanabe, S., Kigawa, T., Kamoun, S., and Shirasu, K. (2011) Phosphatidylinositol monophosphate-binding interface in the oomycete RXLR effector AVR3a is required for its stability in host cells

to modulate plant immunity, *Proc Natl Acad Sci U S A 108*, 14682-14687.
[34] Xiang, T., Zong, N., Zou, Y., Wu, Y., Zhang, J., Xing, W., Li, Y., Tang, X., Zhu, L., Chai, J., and Zhou, J. M. (2008) Pseudomonas syringae effector AvrPto blocks innate immunity by targeting receptor kinases, *Curr Biol 18*, 74-80.
[35] Xing, W., Zou, Y., Liu, Q., Liu, J. N., Luo, X., Huang, Q. Q., Chen, S., Zhu, L. H., Bi, R. C., Hao, Q., Wu, J. W., Zhou, J. M., and Chai, J. J. (2007) The structural basis for activation of plant immunity by bacterial effector protein AvrPto, *Nature 449*, 243-U211.
[36] Anderson, J. C., Pascuzzi, P. E., Xiao, F., Sessa, G., and Martin, G. B. (2006) Host-mediated phosphorylation of type III effector AvrPto promotes Pseudomonas virulence and avirulence in tomato, *Plant Cell 18*, 502-514.
[37] Zhang, J., Shao, F., Li, Y., Cui, H., Chen, L., Li, H., Zou, Y., Long, C., Lan, L., Chai, J., Chen, S., Tang, X., and Zhou, J. M. (2007) A Pseudomonas syringae effector inactivates MAPKs to suppress PAMP-induced immunity in plants, *Cell Host Microbe 1*, 175-185.
[38] Li, H., Xu, H., Zhou, Y., Zhang, J., Long, C., Li, S., Chen, S., Zhou, J. M., and Shao, F. (2007) The phosphothreonine lyase activity of a bacterial type III effector family, *Science 315*, 1000-1003.
[39] Li, X., Lin, H., Zhang, W., Zou, Y., Zhang, J., Tang, X., and Zhou, J. M. (2005) Flagellin induces innate immunity in nonhost interactions that is suppressed by Pseudomonas syringae effectors, *Proc Natl Acad Sci U S A 102*, 12990-12995.
[40] Boutemy, L. S., King, S. R., Win, J., Hughes, R. K., Clarke, T. A., Blumenschein, T. M., Kamoun, S., and Banfield, M. J. (2011) Structures of Phytophthora RXLR effector proteins: a conserved but adaptable fold underpins functional diversity, *J Biol Chem 286*, 35834-35842.
[41] King, S. R., McLellan, H., Boevink, P. C., Armstrong, M. R., Bukharova, T., Sukarta, O., Win, J., Kamoun, S., Birch, P. R., and Banfield, M. J. (2014) Phytophthora infestans RXLR effector PexRD2 interacts with host MAPKKK epsilon to suppress plant immune signaling, *Plant Cell 26*, 1345-1359.
[42] Szurek, B., Rossier, O., Hause, G., and Bonas, U. (2002) Type III-dependent translocation of the Xanthomonas AvrBs3 protein into the plant cell, *Mol Microbiol 46*, 13-23.
[43] Szurek, B., Marois, E., Bonas, U., and Van den Ackerveken, G. (2001) Eukaryotic features of the Xanthomonas type III effector AvrBs3: protein domains involved in transcriptional activation and the interaction with nuclear import receptors from pepper, *Plant J 26*, 523-534.
[44] Fujikawa, T., Ishihara, H., Leach, J. E., and Tsuyumu, S. (2006) Suppression of defense response in plants by the avrBs3/pthA gene family of Xanthomonas spp, *MPMI 19*, 342-349.
[45] Boch, J., and Bonas, U. (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function, *Annu Rev Phytopathol 48*, 419-436.
[46] Pedersen, C., Ver Loren van Themaat, E., McGuffin, L. J., Abbott, J. C., Burgis, T. A., Barton, G., Bindschedler, L. V., Lu, X., Maekawa, T.,

Wessling, R., Cramer, R., Thordal-Christensen, H., Panstruga, R., and Spanu, P. D. (2012) Structure and evolution of barley powdery mildew effector candidates, *BMC Genomics 13*, 694.
[47] Goss, E. M., Press, C. M., and Grunwald, N. J. (2013) Evolution of RXLR-class effectors in the oomycete plant pathogen Phytophthora ramorum, *PLoS One 8*, e79347.
[48] Morbitzer, R., Elsaesser, J., Hausner, J., and Lahaye, T. (2011) Assembly of custom TALE-type DNA binding domains by modular cloning, *Nucleic Acids Res 39*, 5790-5799.
[49] Win, J., Krasileva, K. V., Kamoun, S., Shirasu, K., Staskawicz, B. J., and Banfield, M. J. (2012) Sequence divergent RXLR effectors share a structural fold conserved across plant pathogenic oomycete species, *PLoS Pathog 8*, e1002400.
[50] Chen, X. R., Zhang, B. Y., Xing, Y. P., Li, Q. Y., Li, Y. P., Tong, Y. H., and Xu, J. Y. (2014) Transcriptomic analysis of the phytopathogenic oomycete Phytophthora cactorum provides insights into infection-related effectors, *BMC Genomics 15*, 980.
[51] Dou, D., Kale, S. D., Wang, X., Chen, Y., Wang, Q., Wang, X., Jiang, R. H., Arredondo, F. D., Anderson, R. G., Thakur, P. B., McDowell, J. M., Wang, Y., and Tyler, B. M. (2008) Conserved C-terminal motifs required for avirulence and suppression of cell death by Phytophthora sojae effector Avr1b, *Plant Cell 20*, 1118-1133.
[52] Johal, G. S., and Briggs, S. P. (1992) Reductase activity encoded by the HM1 disease resistance gene in maize, *Science 258*, 985-987.
[53] Dodds, P. N., Lawrence, G. J., Catanzariti, A. M., Teh, T., Wang, C. I., Ayliffe, M. A., Kobe, B., and Ellis, J. G. (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes, *Proc Natl Acad Sci U S A 103*, 8888-8893.
[54] Ravensdale, M., Bernoux, M., Ve, T., Kobe, B., Thrall, P. H., Ellis, J. G., and Dodds, P. N. (2012) Intramolecular interaction influences binding of the Flax L5 and L6 resistance proteins to their AvrL567 ligands, *PLoS Pathog 8*, e1003004.
[55] Jia, Y., McAdams, S. A., Bryan, G. T., Hershey, H. P., and Valent, B. (2000) Direct interaction of resistance gene and avirulence gene products confers rice blast resistance, *EMBO J 19*, 4004-4014.
[56] Krasileva, K. V., Dahlbeck, D., and Staskawicz, B. J. (2010) Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector, *Plant Cell 22*, 2444-2458.
[57] Van der Biezen, E. A., and Jones, J. D. (1998) Plant disease-resistance proteins and the gene-for-gene concept, *Trends Biochem Sci 23*, 454-456.
[58] Martin, G. B., Brommonschenkel, S. H., Chunwongse, J., Frary, A., Ganal, M. W., Spivey, R., Wu, T., Earle, E. D., and Tanksley, S. D. (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato, *Science 262*, 1432-1436.
[59] Loh, Y. T., and Martin, G. B. (1995) The disease-resistance gene Pto and the fenthion-sensitivity gene fen encode closely related functional protein kinases, *Proc Natl Acad Sci U S A 92*, 4181-4184.

[60] Salmeron, J. M., Oldroyd, G. E., Rommens, C. M., Scofield, S. R., Kim, H. S., Lavelle, D. T., Dahlbeck, D., and Staskawicz, B. J. (1996) Tomato Prf is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the Pto kinase gene cluster, *Cell 86*, 123-133.
[61] Saur, I. M., Conlan, B. F., and Rathjen, J. P. (2015) The N-terminal domain of the tomato immune protein Prf contains multiple homotypic and Pto kinase interaction sites, *J Biol Chem 290*, 11258-11267.
[62] Gutierrez, J. R., Balmuth, A. L., Ntoukakis, V., Mucyn, T. S., Gimenez-Ibanez, S., Jones, A. M., and Rathjen, J. P. (2010) Prf immune complexes of tomato are oligomeric and contain multiple Pto-like kinases that diversify effector recognition, *Plant J 61*, 507-518.
[63] Ntoukakis, V., Balmuth, A. L., Mucyn, T. S., Gutierrez, J. R., Jones, A. M., and Rathjen, J. P. (2013) The tomato Prf complex is a molecular trap for bacterial effectors based on Pto transphosphorylation, *PLoS Pathog 9*, e1003123.
[64] Oldroyd, G. E., and Staskawicz, B. J. (1998) Genetically engineered broad-spectrum disease resistance in tomato, *Proc Natl Acad Sci U S A 95*, 10300-10305.
[65] Lin, N. C., and Martin, G. B. (2007) Pto- and Prf-mediated recognition of AvrPto and AvrPtoB restricts the ability of diverse pseudomonas syringae pathovars to infect tomato, *Mol Plant Microbe Interact 20*, 806-815.
[66] Mackey, D., Holt, B. F., 3rd, Wiig, A., and Dangl, J. L. (2002) RIN4 interacts with Pseudomonas syringae type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis, *Cell 108*, 743-754.
[67] Mackey, D., Belkhadir, Y., Alonso, J. M., Ecker, J. R., and Dangl, J. L. (2003) Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2-mediated resistance, *Cell 112*, 379-389.
[68] Chisholm, S. T., Dahlbeck, D., Krishnamurthy, N., Day, B., Sjolander, K., and Staskawicz, B. J. (2005) Molecular characterization of proteolytic cleavage sites of the Pseudomonas syringae effector AvrRpt2, *Proc Natl Acad Sci U S A 102*, 2087-2092.
[69] Axtell, M. J., and Staskawicz, B. J. (2003) Initiation of RPS2-specified disease resistance in Arabidopsis is coupled to the AvrRpt2-directed elimination of RIN4, *Cell 112*, 369-377.
[70] van der Hoorn, R. A., and Kamoun, S. (2008) From Guard to Decoy: a new model for perception of plant pathogen effectors, *Plant Cell 20*, 2009-2017.
[71] Wang, G., Roux, B., Feng, F., Guy, E., Li, L., Li, N., Zhang, X., Lautier, M., Jardinaud, M. F., Chabannes, M., Arlat, M., Chen, S., He, C., Noel, L. D., and Zhou, J. M. (2015) The Decoy Substrate of a Pathogen Effector and a Pseudokinase Specify Pathogen-Induced Modified-Self Recognition and Immunity in Plants, *Cell Host Microbe 18*, 285-295.
[72] Feng, F., Yang, F., Rong, W., Wu, X., Zhang, J., Chen, S., He, C., and Zhou, J. M. (2012) A Xanthomonas uridine 5'-monophosphate transferase inhibits plant immune kinases, *Nature 485*, 114-118.

[73] Guy, E., Lautier, M., Chabannes, M., Roux, B., Lauber, E., Arlat, M., and Noel, L. D. (2013) xopAC-triggered immunity against Xanthomonas depends on Arabidopsis receptor-like cytoplasmic kinase genes PBL2 and RIPK, *PLoS One 8*, e73469.
[74] Kanzaki, H., Yoshida, K., Saitoh, H., Fujisaki, K., Hirabuchi, A., Alaux, L., Fournier, E., Tharreau, D., and Terauchi, R. (2012) Arms race co-evolution of Magnaporthe oryzae AVR-Pik and rice Pik genes driven by their physical interactions, *Plant J 72*, 894-907.
[75] Maqbool, A., Saitoh, H., Franceschetti, M., Stevenson, C., Uemura, A., Kanzaki, H., Kamoun, S., Terauchi, R., and Banfield, M. J. (2015) Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor, *Elife 4*.
[76] Cesari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A., Rivas, S., Alaux, L., Kanzaki, H., Okuyama, Y., Morel, J. B., Fournier, E., Tharreau, D., Terauchi, R., and Kroj, T. (2013) The rice resistance protein pair RGA4/RGA5 recognizes the Magnaporthe oryzae effectors AVR-Pia and AVR1-CO39 by direct binding, *Plant Cell 25*, 1463-1481.
[77] Cesari, S., Kanzaki, H., Fujiwara, T., Bernoux, M., Chalvon, V., Kawano, Y., Shimamoto, K., Dodds, P., Terauchi, R., and Kroj, T. (2014) The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance, *EMBO J 33*, 1941-1959.
[78] Deslandes, L., Olivier, J., Theulieres, F., Hirsch, J., Feng, D. X., Bittner-Eddy, P., Beynon, J., and Marco, Y. (2002) Resistance to Ralstonia solanacearum in Arabidopsis thaliana is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes, *Proc Natl Acad Sci U S A 99*, 2404-2409.
[79] Sarris, P. F., Duxbury, Z., Huh, S. U., Ma, Y., Segonzac, C., Sklenar, J., Derbyshire, P., Cevik, V., Rallapalli, G., Saucet, S. B., Wirthmueller, L., Menke, F. L., Sohn, K. H., and Jones, J. D. (2015) A Plant Immune Receptor Detects Pathogen Effectors that Target WRKY Transcription Factors, *Cell 161*, 1089-1100.
[80] Le Roux, C., Huet, G., Jauneau, A., Camborde, L., Tremousaygue, D., Kraut, A., Zhou, B., Levaillant, M., Adachi, H., Yoshioka, H., Raffaele, S., Berthome, R., Coute, Y., Parker, J. E., and Deslandes, L. (2015) A receptor pair with an integrated decoy converts pathogen disabling of transcription factors to immunity, *Cell 161*, 1074-1088.
[81] Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T., and Dodds, P. N. (2014) A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis, *Front Plant Sci 5*, 606.
[82] Xu, Y., Tao, X., Shen, B., Horng, T., Medzhitov, R., Manley, J. L., and Tong, L. (2000) Structural basis for signal transduction by the Toll/interleukin-1 receptor domains, *Nature 408*, 111-115.
[83] Swiderski, M. R., Birker, D., and Jones, J. D. (2009) 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.
[84] Bernoux, M., Ve, T., Williams, S., Warren, C., Hatters, D., Valkov, E., Zhang, X., Ellis, J. G., Kobe, B., and Dodds, P. N. (2011) Structural and functional analysis of a plant resistance protein TIR domain

reveals interfaces for self-association, signaling, and autoregulation, *Cell Host Microbe 9*, 200-211.

[85] Williams, S. J., Sohn, K. H., Wan, L., Bernoux, M., Sarris, P. F., Segonzac, C., Ve, T., Ma, Y., Saucet, S. B., Ericsson, D. J., Casey, L. W., Lonhienne, T., Winzor, D. J., Zhang, X. X., Coerdt, A., Parker, J. E., Dodds, P. N., Kobe, B., and Jones, J. D. G. (2014) Structural Basis for Assembly and Function of a Heterodimeric Plant Immune Receptor, *Science 344*, 299-303.

[86] Maekawa, T., Cheng, W., Spiridon, L. N., Toller, A., Lukasik, E., Saijo, Y., Liu, P., Shen, Q. H., Micluta, M. A., Somssich, I. E., Takken, F. L., Petrescu, A. J., Chai, J., and Schulze-Lefert, P. (2011) Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death, *Cell Host Microbe 9*, 187-199.

[87] Wang, G. F., Ji, J., El-Kasmi, F., Dangl, J. L., Johal, G., and Balint-Kurti, P. J. (2015) Molecular and functional analyses of a maize autoactive NB-LRR protein identify precise structural requirements for activity, *PLoS Pathog 11*, e1004674.

[88] Tameling, W. I., and Baulcombe, D. C. (2007) Physical association of the NB-LRR resistance protein Rx with a Ran GTPase-activating protein is required for extreme resistance to Potato virus X, *Plant Cell 19*, 1682-1694.

[89] Qi, D., DeYoung, B. J., and Innes, R. W. (2012) Structure-Function Analysis of the Coiled-Coil and Leucine-Rich Repeat Domains of the RPS5 Disease Resistance Protein, *Plant Physiol 158*, 1819-1832.

[90] Takemoto, D., Rafiqi, M., Hurley, U., Lawrence, G. J., Bernoux, M., Hardham, A. R., Ellis, J. G., Dodds, P. N., and Jones, D. A. (2012) N-terminal motifs in some plant disease resistance proteins function in membrane attachment and contribute to disease resistance, *MPMI 25*, 379-392.

[91] Slootweg, E., Roosien, J., Spiridon, L. N., Petrescu, A. J., Tameling, W., Joosten, M., Pomp, R., van Schaik, C., Dees, R., Borst, J. W., Smant, G., Schots, A., Bakker, J., and Goverse, A. (2010) Nucleocytoplasmic Distribution Is Required for Activation of Resistance by the Potato NB-LRR Receptor Rx1 and Is Balanced by Its Functional Domains, *Plant Cell 22*, 4195-4215.

[92] Hao, W., Collier, S. M., Moffett, P., and Chai, J. J. (2013) Structural Basis for the Interaction between the Potato Virus X Resistance Protein (Rx) and Its Cofactor Ran GTPase-activating Protein 2 (RanGAP2), *Journal of Biological Chemistry 288*, 35868-35876.

[93] Kobe, B., and Kajava, A. V. (2001) The leucine-rich repeat as a protein recognition motif, *Curr Opin Struct Biol 11*, 725-732.

[94] Bendahmane, A., Farnham, G., Moffett, P., and Baulcombe, D. C. (2002) Constitutive gain-of-function mutants in a nucleotide binding site-leucine rich repeat protein encoded at the Rx locus of potato, *Plant J 32*, 195-204.

[95] Moffett, P., Farnham, G., Peart, J., and Baulcombe, D. C. (2002) Interaction between domains of a plant NBS-LRR protein in disease resistance-related cell death, *EMBO J 21*, 4511-4519.

[96] Hwang, C. F., Bhakta, A. V., Truesdell, G. M., Pudlo, W. M., and Williamson, V. M. (2000) Evidence for a role of the N terminus and leucine-rich repeat region of the Mi gene product in regulation of localized cell death, *Plant Cell 12*, 1319-1329.
[97] Rairdan, G. J., and Moffett, P. (2006) Distinct domains in the ARC region of the potato resistance protein Rx mediate LRR binding and inhibition of activation, *Plant Cell 18*, 2082-2093.
[98] Hu, Z., Yan, C., Liu, P., Huang, Z., Ma, R., Zhang, C., Wang, R., Zhang, Y., Martinon, F., Miao, D., Deng, H., Wang, J., Chang, J., and Chai, J. (2013) Crystal structure of NLRC4 reveals its autoinhibition mechanism, *Science 341*, 172-175.
[99] Zou, H., Henzel, W. J., Liu, X. S., Lutschg, A., and Wang, X. D. (1997) Apaf-1, a human protein homologous to C-elegans CED-4, participates in cytochrome c-dependent activation of caspase-3, *Cell 90*, 405-413.
[100] Li, P., Nijhawan, D., Budihardjo, I., Srinivasula, S. M., Ahmad, M., Alnemri, E. S., and Wang, X. D. (1997) Cytochrome c and dATP-dependent formation of Apaf-1/caspase-9 complex initiates an apoptotic protease cascade, *Cell 91*, 479-489.
[101] Chinnaiyan, A. M., Chaudhary, D., Orourke, K., Koonin, E. V., and Dixit, V. M. (1997) Role of CED-4 in the activation of CED-3, *Nature 388*, 728-729.
[102] Tameling, W. I., Elzinga, S. D., Darmin, P. S., Vossen, J. H., Takken, F. L., Haring, M. A., and Cornelissen, B. J. (2002) The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity, *Plant Cell 14*, 2929-2939.
[103] Fenyk, S., Campillo Ade, S., Pohl, E., Hussey, P. J., and Cann, M. J. (2012) A nucleotide phosphatase activity in the nucleotide binding domain of an orphan resistance protein from rice, *J Biol Chem 287*, 4023-4032.
[104] Ueda, H., Yamaguchi, Y., and Sano, H. (2006) Direct interaction between the tobacco mosaic virus helicase domain and the ATP-bound resistance protein, N factor during the hypersensitive response in tobacco plants, *Plant Mol Biol 61*, 31-45.
[105] Tameling, W. I., Vossen, J. H., Albrecht, M., Lengauer, T., Berden, J. A., Haring, M. A., Cornelissen, B. J., and Takken, F. L. (2006) Mutations in the NB-ARC domain of I-2 that impair ATP hydrolysis cause autoactivation, *Plant Physiol 140*, 1233-1245.
[106] Yuan, S., and Akey, C. W. (2013) Apoptosome structure, assembly, and procaspase activation, *Structure 21*, 501-515.
[107] Inohara, N., and Nunez, G. (2001) The NOD: a signaling module that regulates apoptosis and host defense against pathogens, *Oncogene 20*, 6473-6481.
[108] Yu, X. C., Wang, L., Acehan, D., Wang, X. D., and Akey, C. W. (2006) Three-dimensional structure of a double apoptosome formed by the Drosophila Apaf-1 related killer, *J Mol Biol 355*, 577-589.
[109] Chattopadhyaya, R., and Pal, A. (2008) Three-dimensional models of NB-ARC domains of disease resistance proteins in tomato, arabidopsis, and flax, *J Biomol Struct Dyn 25*, 357-371.

[110] Van Ooijen, G., Mayr, G., Kasiem, M. M. A., Albrecht, M., Cornelissen, B. J. C., and Takken, F. L. W. (2008) Structure-function analysis of the NB-ARC domain of plant disease resistance proteins, *J Exp Bot 59*, 1383-1397.
[111] Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I., and Owens, R. J. (2007) A versatile ligation-independent cloning method suitable for high-throughput expression screening applications, *Nucleic Acids Res 35*, e45.
[112] Yumerefendi, H., Tarendeau, F., Mas, P. J., and Hart, D. J. (2010) ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets, *J Struct Biol 172*, 66-74.
[113] Oh, S. K., Young, C., Lee, M., Oliva, R., Bozkurt, T. O., Cano, L. M., Win, J., Bos, J. I., Liu, H. Y., van Damme, M., Morgan, W., Choi, D., Van der Vossen, E. A., Vleeshouwers, V. G., and Kamoun, S. (2009) In planta expression screens of Phytophthora infestans RXLR effectors reveal diverse phenotypes, including activation of the Solanum bulbocastanum disease resistance protein Rpi-blb2, *Plant Cell 21*, 2928-2947.
[114] Chapman, S., Kavanagh, T., and Baulcombe, D. (1992) Potato virus X as a vector for gene expression in plants, *Plant J 2*, 549-557.
[115] Hamada, W. K., S. (2004) Cloning vector pCB301-I1M.
[116] Hamilton, M. D., Nuara, A. A., Gammon, D. B., Buller, R. M., and Evans, D. H. (2007) Duplex strand joining reactions catalyzed by vaccinia virus DNA polymerase, *Nucleic Acids Res 35*, 143-151.
[117] Hanahan, D. (1983) Studies on transformation of Escherichia coli with plasmids, *J Mol Biol 166*, 557-580.
[118] Van Larebeke, N., Engler, G., Holsters, M., Van den Elsacker, S., Zaenen, I., Schilperoort, R. A., and Schell, J. (1974) Large plasmid in Agrobacterium tumefaciens essential for crown gall-inducing ability, *Nature 252*, 169-170.
[119] Koncz, C., and Schell, J. (1986) The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector, *Molec Gen Genet 204*, 383-396.
[120] An, Y., Meresse, P., Mas, P. J., and Hart, D. J. (2011) CoESPRIT: a library-based construct screening method for identification and expression of soluble protein complexes, *PLoS One 6*, e16261.
[121] Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis, *Nat Protoc 10*, 845-858.
[122] Yang, Z. R., Thomson, R., McNeil, P., and Esnouf, R. M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins, *Bioinformatics 21*, 3369-3376.
[123] Kapila, J., De Rycke, R., Van Montagu, M., and Angenon, G. (1997) An Agrobacterium-mediated transient gene expression system for intact leaves, *Plant Science 122*, 101-108.
[124] Incardona, M. F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N., and Svensson, O. (2009) EDNA: a framework for plugin-based

applications applied to X-ray experiment online data analysis, *J Synchrotron Radiat 16*, 872-879.
[125] Lanzetta, P. A., Alvarez, L. J., Reinach, P. S., and Candia, O. A. (1979) An improved assay for nanomole amounts of inorganic phosphate, *Anal Biochem 100*, 95-97.
[126] Bos, J. I. B., Kanneganti, T. D., Young, C., Cakir, C., Huitema, E., Win, J., Armstrong, M. R., Birch, P. R. J., and Kamoun, S. (2006) The C-terminal half of Phytophthora infestans RXLR effector AVR3a is sufficient to trigger R3a-mediated hypersensitivity and suppress INF1-induced cell death in Nicotiana benthamiana, *Plant Journal 48*, 165-176.
[127] Vos, P., Simons, G., Jesse, T., Wijbrandi, J., Heinen, L., Hogers, R., Frijters, A., Groenendijk, J., Diergaarde, P., Reijans, M., Fierens-Onstenk, J., de Both, M., Peleman, J., Liharska, T., Hontelez, J., and Zabeau, M. (1998) The tomato Mi-1 gene confers resistance to both root-knot nematodes and potato aphids, *Nat Biotechnol 16*, 1365-1369.
[128] Dodds, P. N., Lawrence, G. J., Catanzariti, A. M., Teh, T., Wang, C. I. A., Ayliffe, M. A., Kobe, B., and Ellis, J. G. (2006) Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes, *Proc Natl Acad Sci U S A 103*, 8888-8893.
[129] Takken, F. L. W., Albrecht, M., and Tameling, W. I. L. (2006) Resistance proteins: molecular switches of plant defence, *Curr Opin Plant Biol 9*, 383-390.
[130] Williams, S. J., Sornaraj, P., deCourcy-Ireland, E., Menz, R. I., Kobe, B., Ellis, J. G., Dodds, P. N., and Anderson, P. A. (2011) An Autoactive Mutant of the M Flax Rust Resistance Protein Has a Preference for Binding ATP, Whereas Wild-Type M Protein Binds ADP, *MPMI 24*, 897-906.
[131] Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L., Tate, J., and Punta, M. (2014) Pfam: the protein families database, *Nucleic Acids Res 42*, D222-230.
[132] Bos, J. I. B., Chaparro-Garcia, A., Quesada-Ocampo, L. M., Gardener, B. B. M., and Kamoun, S. (2009) Distinct Amino Acids of the Phytophthora infestans Effector AVR3a Condition Activation of R3a Hypersensitivity and Suppression of Cell Death, *MPMI 22*, 269-281.
[133] Engelhardt, S., Boevink, P. C., Armstrong, M. R., Ramos, M. B., Hein, I., and Birch, P. R. J. (2012) Relocalization of Late Blight Resistance Protein R3a to Endosomal Compartments Is Associated with Effector Recognition and Required for the Immune Response, *Plant Cell 24*, 5142-5158.
[134] Segretin, M. E., Pais, M., Franceschetti, M., Chaparro-Garcia, A., Bos, J. I. B., Banfield, M. J., and Kamoun, S. (2014) Single Amino Acid Mutations in the Potato Immune Receptor R3a Expand Response to Phytophthora Effectors, *MPMI 27*, 624-637.
[135] Yumerefendi, H., Desravines, D. C., and Hart, D. J. (2011) Library-based methods for identification of soluble expression constructs, *Methods 55*, 38-43.

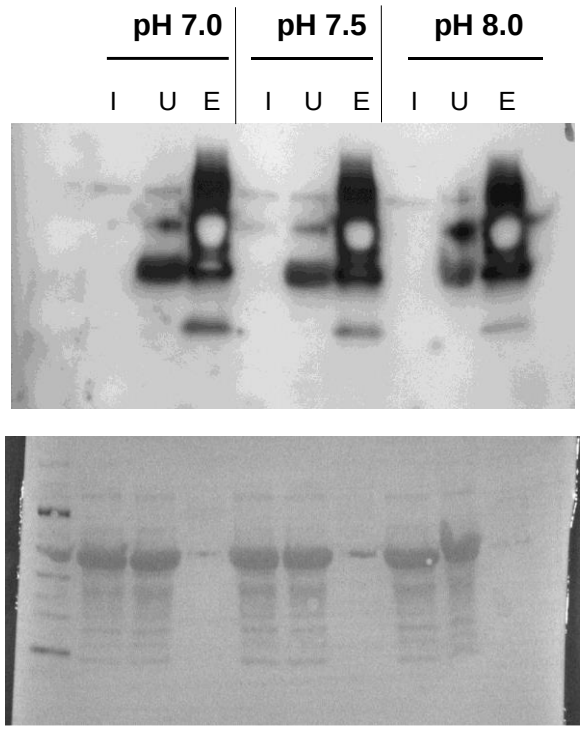
[136] Van Ooijen, G., Lukasik, E., Van Den Burg, H. A., Vossen, J. H., Cornelissen, B. J., and Takken, F. L. (2010) The small heat shock protein 20 RSI2 interacts with and is required for stability and function of tomato resistance protein I-2, *Plant J 63*, 563-572.
[137] Azevedo, C., Sadanandom, A., Kitagawa, K., Freialdenhoven, A., Shirasu, K., and Schulze-Lefert, P. (2002) The RAR1 interactor SGT1, an essential component of R gene-triggered disease resistance, *Science 295*, 2073-2076.
[138] Liu, Y., Burch-Smith, T., Schiff, M., Feng, S., and Dinesh-Kumar, S. P. (2004) Molecular chaperone Hsp90 associates with resistance protein N and its signaling proteins SGT1 and Rar1 to modulate an innate immune response in plants, *J Biol Chem 279*, 2101-2108.
[139] Yu, L. M. (1995) Elicitins from Phytophthora and basic resistance in tobacco, *Proc Natl Acad Sci U S A 92*, 4088-4094.
[140] Kamoun, S., van West, P., de Jong, A. J., de Groot, K. E., Vleeshouwers, V. G., and Govers, F. (1997) A gene encoding a protein elicitor of Phytophthora infestans is down-regulated during infection of potato, *MPMI 10*, 13-20.
[141] Kanzaki, H., Saitoh, H., Ito, A., Fujisawa, S., Kamoun, S., Katou, S., Yoshioka, H., and Terauchi, R. (2003) Cytosolic HSP90 and HSP70 are essential components of INF1-mediated hypersensitive response and non-host resistance to Pseudomonas cichorii in Nicotiana benthamiana, *Mol Plant Pathol 4*, 383-391.
[142] Kanzaki, H., Saitoh, H., Takahashi, Y., Berberich, T., Ito, A., Kamoun, S., and Terauchi, R. (2008) NbLRK1, a lectin-like receptor kinase protein of Nicotiana benthamiana, interacts with Phytophthora infestans INF1 elicitin and mediates INF1-induced cell death, *Planta 228*, 977-987.
[143] Chaparro-Garcia, A., Wilkinson, R. C., Gimenez-Ibanez, S., Findlay, K., Coffey, M. D., Zipfel, C., Rathjen, J. P., Kamoun, S., and Schornack, S. (2011) The receptor-like kinase SERK3/BAK1 is required for basal resistance against the late blight pathogen phytophthora infestans in Nicotiana benthamiana, *PLoS One 6*, e16608.
[144] Kamoun, S., van der Lee, T., van den Berg-Velthuis, G., de Groot, K. E., and Govers, F. (1998) Loss of Production of the Elicitor Protein INF1 in the Clonal Lineage US-1 of Phytophthora infestans, *Phytopathology 88*, 1315-1323.
[145] Kamoun, S., van West, P., Vleeshouwers, V. G., de Groot, K. E., and Govers, F. (1998) Resistance of nicotiana benthamiana to phytophthora infestans is mediated by the recognition of the elicitor protein INF1, *Plant Cell 10*, 1413-1426.
[146] Yoshida, K., Schuenemann, V. J., Cano, L. M., Pais, M., Mishra, B., Sharma, R., Lanz, C., Martin, F. N., Kamoun, S., Krause, J., Thines, M., Weigel, D., and Burbano, H. A. (2013) The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine, *Elife 2*, e00731.
[147] el-Kharbotly, A., Leonards-Schippers, C., Huigen, D. J., Jacobsen, E., Pereira, A., Stiekema, W. J., Salamini, F., and Gebhardt, C. (1994) Segregation analysis and RFLP mapping of the R1 and R3 alleles

conferring race-specific resistance to Phytophthora infestans in progeny of dihaploid potato parents, *Mol Gen Genet 242*, 749-754.
[148] Huang, S., Vleeshouwers, V. G., Werij, J. S., Hutten, R. C., van Eck, H. J., Visser, R. G., and Jacobsen, E. (2004) The R3 resistance to Phytophthora infestans in potato is conferred by two closely linked R genes with distinct specificities, *MPMI 17*, 428-435.
[149] Huang, S., van der Vossen, E. A., Kuang, H., Vleeshouwers, V. G., Zhang, N., Borm, T. J., van Eck, H. J., Baker, B., Jacobsen, E., and Visser, R. G. (2005) Comparative genomics enabled the isolation of the R3a late blight resistance gene in potato, *Plant J 42*, 251-261.
[150] Wirthmueller, L., Zhang, Y., Jones, J. D., and Parker, J. E. (2007) Nuclear accumulation of the Arabidopsis immune receptor RPS4 is necessary for triggering EDS1-dependent defense, *Curr Biol 17*, 2023-2029.
[151] Shen, Q. H., Saijo, Y., Mauch, S., Biskup, C., Bieri, S., Keller, B., Seki, H., Ulker, B., Somssich, I. E., and Schulze-Lefert, P. (2007) Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses, *Science 315*, 1098-1103.
[152] Engelhardt, S., Boevink, P. C., Armstrong, M. R., Ramos, M. B., Hein, I., and Birch, P. R. (2012) Relocalization of late blight resistance protein R3a to endosomal compartments is associated with effector recognition and required for the immune response, *Plant Cell 24*, 5142-5158.
[153] Bai, S., Liu, J., Chang, C., Zhang, L., Maekawa, T., Wang, Q., Xiao, W., Liu, Y., Chai, J., Takken, F. L., Schulze-Lefert, P., and Shen, Q. H. (2012) Structure-function analysis of barley NLR immune receptor MLA10 reveals its cell compartment specific activity in cell death and disease resistance, *PLoS Pathog 8*, e1002752.
[154] Franceschetti, M., Bueno, E., Wilson, R. A., Tucker, S. L., Gomez-Mena, C., Calder, G., and Sesma, A. (2011) Fungal virulence and development is regulated by alternative pre-mRNA 3'end processing in Magnaporthe oryzae, *PLoS Pathog 7*, e1002441.
[155] Giannakopoulou, A., Steele, J. F., Segretin, M. E., Bozkurt, T., Zhou, J., Robatzek, S., Banfield, M. J., Pais, M., and Kamoun, S. (2015) Tomato I2 immune receptor can be engineered to confer partial resistance to the oomycete Phytophthora infestans in addition to the fungus Fusarium oxysporum, *Mol Plant Microbe Interact*.
[156] Gabriels, S. H., Vossen, J. H., Ekengren, S. K., van Ooijen, G., Abd-El-Haliem, A. M., van den Berg, G. C., Rainey, D. Y., Martin, G. B., Takken, F. L., de Wit, P. J., and Joosten, M. H. (2007) An NB-LRR protein required for HR signalling mediated by both extra- and intracellular resistance proteins, *Plant J 50*, 14-28.
[157] Sueldo, D. J., Shimels, M., Spiridon, L. N., Caldararu, O., Petrescu, A. J., Joosten, M. H., and Tameling, W. I. (2015) Random mutagenesis of the nucleotide-binding domain of NRC1 (NB-LRR Required for Hypersensitive Response-Associated Cell Death-1), a downstream signalling nucleotide-binding, leucine-rich repeat (NB-LRR) protein, identifies gain-of-function mutations in the nucleotide-binding pocket, *New Phytol 208*, 210-223.

[158] Winter, G., Lobley, C. M., and Prince, S. M. (2013) Decision making in xia2, *Acta Crystallogr D 69*, 1260-1273.
[159] Kabsch, W. (2010) Xds, *Acta Crystallogr D 66*, 125-132.
[160] Evans, P. (2006) Scaling and assessment of data quality, *Acta Crystallogr D 62*, 72-82.
[161] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Crystallogr D 67*, 235-242.
[162] Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D 66*, 213-221.
[163] Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr D 67*, 355-367.
[164] Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D 60*, 2126-2132.
[165] Rairdan, G. J., Collier, S. M., Sacco, M. A., Baldwin, T. T., Boettrich, T., and Moffett, P. (2008) 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.
[166] Rigden, D. J., Mello, L. V., and Bertioli, D. J. (2000) Structural modeling of a plant disease resistance gene product domain, *Proteins-Structure Function and Genetics 41*, 133-143.
[167] Fenyk, S., Townsend, P. D., Dixon, C. H., Spies, G. B., de San Eustaquio Campillo, A., Slootweg, E. J., Westerhof, L. B., Gawehns, F. K., Knight, M. R., Sharples, G. J., Goverse, A., Palsson, L. O., Takken, F. L., and Cann, M. J. (2015) The Potato Nucleotide-Binding Leucine-Rich Repeat (NLR) Immune Receptor Rx1 is a Pathogen Dependent DNA-Deforming Protein, *J Biol Chem*.
[168] Syson, K., Stevenson, C. E., Rashid, A. M., Saalbach, G., Tang, M., Tuukkanen, A., Svergun, D. I., Withers, S. G., Lawson, D. M., and Bornemann, S. (2014) Structural insight into how Streptomyces coelicolor maltosyl transferase GlgE binds alpha-maltose 1-phosphate and forms a maltosyl-enzyme intermediate, *Biochemistry 53*, 2494-2504.
[169] Stevenson, C. E., Assaad, A., Chandra, G., Le, T. B., Greive, S. J., Bibb, M. J., and Lawson, D. M. (2013) Investigation of DNA sequence recognition by a streptomycete MarR family transcriptional regulator through surface plasmon resonance and X-ray crystallography, *Nucleic Acids Res 41*, 7009-7022.
[170] Saleh, A., Srinivasula, S. M., Acharya, S., Fishel, R., and Alnemri, E. S. (1999) Cytochrome c and dATP-mediated oligomerization of Apaf-1 is

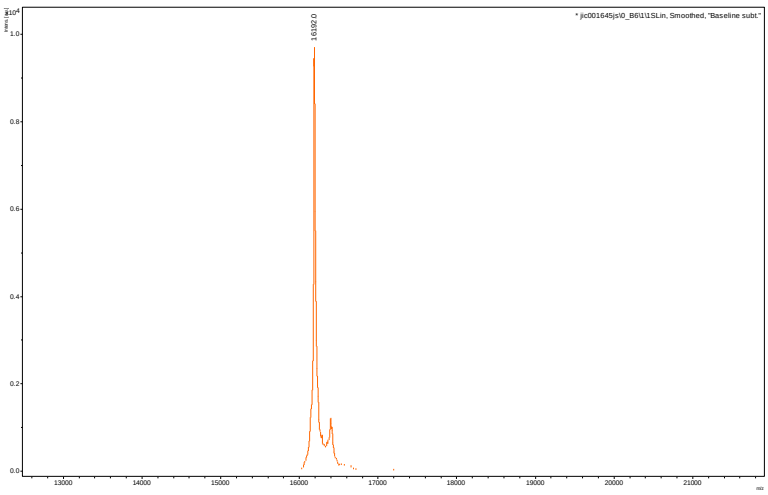
a prerequisite for procaspase-9 activation, *J Biol Chem 274*, 17941-17945.
[171] Jupe, F., Witek, K., Verweij, W., Sliwka, J., Pritchard, L., Etherington, G. J., Maclean, D., Cock, P. J., Leggett, R. M., Bryan, G. J., Cardle, L., Hein, I., and Jones, J. D. (2013) Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations, *Plant J 76*, 530-544.
[172] Slootweg, E. J., Spiridon, L. N., Roosien, J., Butterbach, P., Pomp, R., Westerhof, L., Wilbers, R., Bakker, E., Bakker, J., Petrescu, A. J., Smant, G., and Goverse, A. (2013) Structural Determinants at the Interface of the ARC2 and Leucine-Rich Repeat Domains Control the Activation of the Plant Immune Receptors Rx1 and Gpa2, *Plant Physiol 162*, 1510-1528.
[173] Riedl, S. J., Li, W., Chao, Y., Schwarzenbacher, R., and Shi, Y. (2005) Structure of the apoptotic protease-activating factor 1 bound to ADP, *Nature 434*, 926-933.
[174] Yan, N., Chai, J. J., Lee, E. S., Gu, L. C., Liu, Q., He, J. Q., Wu, J. W., Kokel, D., Li, H. L., Hao, Q., Xue, D., and Shi, Y. G. (2005) Structure of the CED-4-CED-9 complex provides insights into programmed cell death in Caenorhabditis elegans, *Nature 437*, 831-837.
[175] van Ooijen, G., van den Burg, H. A., Cornelissen, B. J. C., and Takken, F. L. W. (2007) Structure and function of resistance proteins in solanaceous plants, In *Annu Rev Phytopathol*, pp 43-72, Annual Reviews, Palo Alto.
[176] Kim, H. E., Du, F., Fang, M., and Wang, X. (2005) Formation of apoptosome is initiated by cytochrome c-induced dATP hydrolysis and subsequent nucleotide exchange on Apaf-1, *Proc Natl Acad Sci U S A 102*, 17545-17550.
[177] Sueldo, D. J. D., H.; de Koster, C.; Joosten, M.; Tameling, W. (2014) Towards generating broad-spectrum resistance to pathogens in plants: Studies on a down-stream signalling NB-LRR of tomato, Wageningen University.

9 Appendix



Appendix Figure 1 Western blot of R3a IP experiment

Full blot from Figure 4.2.5 A. I- input, U- unbound, E-elution.



Appendix Figure 2 Intact MS for R3aCC domain

Intact MALDI mass spectroscopy spectrum for R3a^{CC} purified in Section 3.4. Performed by Gerhard Saalbach

Name	Description	Cloning method	Sequence
R2_CC-F	Forward primer for S. demissum R2 coiled-coil domain	Infusion	aagttctgtttcagggcccgATGGCTGATGCCTTTCTATCATTTGCAG
R2_CCR	Reverse primer for S demissum R2 coiled-coil	Infusion	atggtctagaaagcttaATCCACATATGAGGTAGTTCTCCTCAATGTTC
R2_NBARC-F	Forward primer for S. demissum R2 NBARC domain	Infusion	aagtctgttcagggcccgGACCAGGATTACATTTTTGTTGGACTTCAGG
R2_NBARC-R	Reverse primer for S. demissum R2 NBARC domain	Infusion	atggtctagaaagcttaCTCCAATACCTTTTGTATCGCAAGATCACG
R2_LRR-F	Forward primer for S. demissum R2 LRR domain	Infusion	aagtctgttcagggcccgGTAAACTTCTTTGACATTTATGATCCAAGAAGCC
R2_LRR-R	Reverse primer for S. demissum R2 LRR domain	Infusion	atggtctagaaagcttaCAACATATAATTCCGCTTCAACAGCTCCAC
I2_CC-F	Forward primer for S. lycopersicum I2 coiled-coil domain	Infusion	aagtctgttcagggcccgATGGAGATTGGCTTAGCAGTTGGTGG
I2_CC-R	Reverse primer for S. lycopersicum I2 coiled-coil domain	Infusion	atggtctagaaagcttaAAAATACTCCTTTAAGCCAAGGAGACCAATTTG
I2_NBARC-F	Forward primer for S. lycopersicum I2 NBARC domain	Infusion	aagtctgttcagggcccgGATTCCACGAAACTAGAAACTAGAAGACCTTC
I2_NBARC-R	Reverse primer for S. lycopersicum I2 NBARC domain	Infusion	atggtctagaaagcttaCTGCTCCAATTTGTAGAGGGGTGTC
I2_LRR-F	Forward primer for S. lycopersicum I2 LRR domain	Infusion	aagtctgttcagggcccgTTGAGGACATTGCTCTCCGATACG
I2_LRR-R	Reverse primer for S. lycopersicum I2 LRR domain	Infusion	atggtctagaaagcttaAATATATTTCCAATCGATATTTATGATGGGAATATGAGCAATTTGTGGCC
Rpi-blb2_CC-F	Forward primer for S. bulbocastanum Rpi-blb2 coiled-coil domain	Infusion	aagtctgttcagggcccgATGGAAAAACCGAAAAAGATAATGAAGAAGCAAAC
Rpi-blb2_CC-R	Reverse primer for S. bulbocastanum Rpi-blb2 coiled-coil domain	Infusion	atggtctagaaagcttaGATGAGACCTCTGTTCTTGGAAATGTTC
Rpi-blb2_NBARC-F	Forward primer for S. bulbocastanum Rpi-blb2 NBARC domain	Infusion	aagtctgttcagggcccgGTTGTGAACTCTCCCAAGAAACCAGTTG
Rpi-blb2_NBARC-R	Reverse primer for S. bulbocastanum Rpi-blb2 I2 NBARC domain	Infusion	atggtctagaaagcttaCAACTTTTCCTTTCTTGCTTTTATCAAACAAAAG
Rpi-blb2_LRR-F	Forward primer for S. bulbocastanum Rpi-blb2 LRR domain	Infusion	aagtctgttcagggcccgTGTGATCGGATAAGTTCAAGTGCTCCATC
Rpi-blb2_LRR-R	Reverse primer for S. bulbocastanum Rpi-blb2 LRR domain	Infusion	atggtctagaaagcttaCTTAAATAACGGGATATCCTTCTGGCCAAGG
NRC1_CC-F	Forward primer for S. lycopersicum NRC1 coiled-coil domain	Infusion	aagtctgttcagggcccgATGGTTGATGTAGGGGTTGAATTTCTG
NRC1_CC-R	Reverse primer for S. lycopersicum NRC1 coiled-coil domain	Infusion	atggtctagaaagcttaAGGCTGTCTCTTCTCCACCTCTGC
NRC1_NBARC-F	Forward primer for S. lycopersicum NRC1 NBARC domain	Infusion	aagtctgttcagggcccgCCTGTGGTTGAGGAAGATGATGTGG
NRC1_NBARC-R	Reverse primer for S. lycopersicum NRC1 NBARC domain	Infusion	atggtctagaaagcttaCTCTGTCGTAGCCTCTTGCCAGC
NRC1_LRR-F	Forward primer for S. lycopersicum NRC1 LRR domain	Infusion	aagtctgttcagggcccgGAAAATCTTTTCCATGAAGTAAAATTCGGTGG
NRC1_LRR-R	Reverse primer for S. lycopersicum NRC1 LRR domain	Infusion	atggtctagaaagcttaAGAAGCTGTCTGTACATCAGAATCAGGAGGG
NRC1-like_CC-F	Forward primer for N. benthamiana NRC1-like coiled-coil domain	Infusion	aagtctgttcagggcccgATGGCACATGCAGTACTGAATTTTCCG
NRC1-like_CC-R	Reverse primer for N. benthamiana NRC1-like coiled-coil domain	Infusion	atggtctagaaagcttaAGGACCCTGCTGAGTTGTTTCCG
NRC1-like_NBARC-F	Forward primer for N. benthamiana NRC1-like NBARC domain	Infusion	aagtctgttcagggcccgGCATTGGAGGATGATGAAGTTGTCG
NRC1-like_NBARC-R	Reverse primer for N. benthamiana NRC1-like NBARC domain	Infusion	atggtctagaaagcttaGTTACTAGCCTCCTTTTTGCAGAACTGATGC
NRC1-like_LRR-F	Forward primer for N. benthamiana NRC1-like LRR domain	Infusion	aagtctgttcagggcccgAAATGGCTTTTTCAAGAAGTCAGTCTAACACC
NRC1-like_LRR-R	Reverse primer for N. benthamiana NRC1-like LRR domain	Infusion	atggtctagaaagcttaCTGTGTGGCCTTGGATCCAGC
R3a_CC-F	Forward primer for S. tuberosum R3a CC domain	Infusion	AAGTTCTGTTTCAGGGCCCGGACCGTCTGGCTCCGCATG
R3a_CC-R	Reverse primer for S. tuberosum R3a CC domain	Infusion	ATGGTCTAGAAAGCTTTAGGTGGAGCCAAAGTGTTCTTTCAGAC
R3a_NBARC-F	Forward primer for S. tuberosum R3a NBARC domain	Infusion	AAGTTCTGTTTCAGGGCCCGAAACAAGAAACCCGCACGCCG
R3a_NBARC-R	Reverse primer for S. tuberosum R3a NBARC domain	Infusion	ATGGTCTAGAAAGCTTTAACGCGGCAGAATATTATGCAGCAC
R3a_LRR-F	Forward primer forS. tuberosum R3a LRR domain	Infusion	AAGTTCTGTTTCAGGGCCCGCTGACGAGTCTGCCGCAC
R3a_LRR-R	Reverse primer for S. tuberosum R3a LRR domain	Infusion	ATGGTCTAGAAAGCTTTAGTAGATCGACAGTTCTGACAGACTAGACGG
R3a_CC_GW_F	GW Forward primer for S. tuberosum R3a CC domain	Gateway	AAAAAGCAGGCTCAATGGAA ATC GGC CTG GC
R3a_CC_GW_R	GW Reverse primer for S. tuberosum R3a CC domain	Gateway	AGAAAGCTGGGTTTA TGACGTGGACGGCG
R3a_NB-ARC_GW-F	GW Forward primer for S. tuberosum R3a NBARC domain	Gateway	AAAAAGCAGGCTCAATG AAA CAA GAA ACC CGC ACG
R3a_NB-ARC_GW-R	GW Reverse primer for S. tuberosum R3a NBARC domain	Gateway	AGAAAGCTGGGTTTA CTGTTCCAGTTTATACACGCGGG
R3a_LRR_FulL_GW	GW Forward primer forS. tuberosum R3a LRR domain	Gateway	AAAAAGCAGGCTCACTGCGTACCCTGC
R3a_LRR_FulL_R_GW	GW Reverse primer for S. tuberosum R3a LRR domain	Gateway	AGAAAGCTGGGTTTACATACATTCACGGTCG
FAM-labelled	5'-FAM labelled oligo for anisotropy	-	TGGGTCAACGTGGGCAAAGATGTCCTAGCAATGTAATCGTCTATGACGTT
DNA 40 F	Oligonucleotide used for SPR	-	CTAGCTACTAGCTACAAACTTTCTTGTCTAGTAGCTAGTG
DNA 40 R	Oligonucleotide used for SPR	-	CACTAGCTACTAGACAAGAAAGTTTGTAGCTAGTAGCCTACCCTACGTCCTCCTGC
DNA 30 F	Oligonucleotide used for SPR	-	TACTAGCTACAAACTTTCTTGTCTAGTAGC
DNA 30R	Oligonucleotide used for SPR	-	GCTACTAGACAAGAAAGTTTGTAGCTAGTACCCTACGTCCTCCTGC
DNA 20F	Oligonucleotide used for SPR	-	GCTACAAACTTTCTTGTCTA
DNA 20 R	Oligonucleotide used for SPR	-	TAGACAAGAAAGTTTGTAGCCCTACCCTACGTCCTCCTGC
Sco3204 22 plus 8 F	Oligonucleotide used for SPR	-	TGCTAATACTTGAACTCTCAATCTTTGAAT
Sco3204 22 plus 8 R	Oligonucleotide used for SPR	-	ATTCAAAGATTGAGAGTTCAAGTATTAGCACCTACCCTACGTCCTCCTGC
Sco7681F	Oligonucleotide used for SPR	-	ATAGAAATGATAATCATTTCTATCTAGCACGT
Sco7681R	Oligonucleotide used for SPR	-	ACGTGCTAGATAGAAATGATTATCATTTCTATCCCTACCCTACGTCCTCCTGC
35S promoter	CaMV 35S promoter for sequencing	-	CTATCCTTCGCAAGACCCTTC

Appendix Table 1

List of primers used in this project.