Molecular Insight into the Activation of a Plant Disease Resistance Protein

Simon J. Williams

A thesis submitted for the Degree of Doctor of Philosophy

School of Biological Sciences
Faculty of Science and Engineering
Flinders University, South Australia

September 2009

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES VII

DECLARATION VIII

ACKNOWLEDGMENTS IX

ABSTRACT XI

ABBREVIATION LIST XIII

CHAPTER 1: INTRODUCTION 1

1.1 Overview of Plant Innate Immunity 2

1.1.1 Plant innate immunity: a global view 2

1.1.2 Amplitude of resistance 3

1.2 PAMP-Triggered Immunity 6

1.3 Pathogen Effectors 7

1.4 Effector-Triggered Immunity 10

1.4.1 R genes 10

1.4.1.1 The eLRR R genes 10

1.4.1.2 NBS-LRR 11

1.5 The Flax, Flax Rust System 12

1.5.1 The flax rust pathogen 12

1.5.2 Flax R genes and rust effector proteins 13

1.6 Proposed Function of the NSB-LRR R Protein Domains 15

1.6.1 Leucine-rich repeat 15

1.6.2 Nucleotide binding site (NBS) 16

1.6.3 Amino terminal domain 20

1.7 Effector Perception by R Proteins 21

1.7.1 Recognition of effector proteins by R proteins 21

1.7.2 Indirect recognition: the guard and decoy hypothesis 21

1.7.3 Direct recognition 22

1.8 R protein Structure/Function 25

1.8.1 Structure studies of the NB-ARC region of Apaf-1 and CED-4 25

1.8.2 Intramolecular interactions regulate activation and signalling 26

1.8.3 A model of R protein activation 27

1.8.4 Functional studies of STAND proteins support the R protein molecular switch model 28

1.9 Project Aims and Objectives 31

1.9.1 Aims 31

CHAPTER 2: EXPERIMENTAL PROCEDURES 33

2.1 Materials 34

2.1.1 Culture media, solutions and buffers 34

2.1.2 Yeast and bacterial stains used in this study 37

2.1.3 Oligonucleotides 37

2.1.4 Plasmids 38

2.2 General Procedures 39

2.2.1 Preparation of electrocompotent and heat shock *E. coli* 39

2.2.2 Transformation of competent *E. coli* 39

2.2.2.1 Electroporation of DH10B 39

2.2.2.2 Heat Shock of BL21 pLysS 39

2.2.3 Colony PCR 39

2.2.4 Sodium do-decyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) 40

2.2.5 Coomassie staining of proteins in polyacrylamide gels 40

2.2.6 Silver staining of proteins in polyacrylamide gels 40

2.2.7 Western blot 40

2.2.8 Protein concentration 41

2.2.9 Protein quantification 41

2.3 Generation of Mutated Variants of M for Expression in P. pastoris 41

2.3.1 M truncated mutants 41

2.3.1.1 Generation of MΔTIR 41

2.3.1.2 Generation of MΔLRR 42

2.3.2 Site-direct mutagenesis to produce rational point mutations in M 42

2.3.3 Preparation of electrocompetent *P. pastoris* 43

2.3.4 Electroporation of *P. pastoris* 43

2.3.5 Screening and test expression of transformed *P. pastoris* 43

2.4 Growth and Expression of R proteins 44

2.4.1 Shaker flask expression 44

2.4.2 Fermentation expression 44

2.5 Cell Lysis 45

2.5.1 Cell disruptor 45

2.5.1.1 P. pastoris 45

2.5.1.2 Protein expression in E. coli 45

2.5.2 French press 46

2.6 Purification of M, L6 and M Variants 46
2.6.1 Gel filtration (GF) 46
2.6.2 Large scale purification 46
2.6.3 Small scale purification 47

2.7 Cloning, Expression and Purification of AvrM/avrM 47
2.7.1 Cloning of avrMC* into pET vector 47
2.7.2 Transformation of BL21 pLysS with AvrM/avrM 48
2.7.3 Growth and test expression 48
2.7.4 Large scale expression 48
2.7.5 Purification 49

2.8 Biochemical Assays 49
2.8.1 ATPase assay 49
2.8.2 ATP/ADP quantification assays using luminescence 49
2.8.3 M protein quantification using VersaDoc imaging system 50

2.9 Interaction Assays 50
2.9.1 Co-immunoprecipitation 50
2.9.1.1 Immunoprecipitation using Protein G agarose beads 50
2.9.1.2 Immunoprecipitation using protein A microbeads 51
2.9.2 Nucleotide exchange assays 51

2.10 Multiple Sequence Alignment of the NB-ARC Region of M, other R proteins, Apaf-1 and CED-4 51

2.11 Structure Model of the NB-ARC Domain of M 52

CHAPTER 3: PURIFICATION OF THE M AND L6 FLAX RUST RESISTANCE PROTEINS EXPRESSED IN *PICHIA PASTORIS* 53

3.1 Introduction 54

3.2 Results 55
3.2.1 Three-step purification of M and L6 55
3.2.2 Small scale expression and one-step purification 59
3.2.3 Potential issues in the functionality of the purified proteins 62
3.2.4 Full-length M mutants can be purified from *P. pastoris* 66

3.3 Discussion 68
3.3.1 Expression and purification of flax R proteins facilitates *in vitro* studies 68
3.3.2 Physical properties of the flax R proteins cause self-association and aggregation 69
3.3.3 Concluding remarks 70

CHAPTER 4: BIOCHEMICAL INVESTIGATION OF THE NB-ARC DOMAIN OF PURIFIED FLAX RUST R PROTEINS 71

4.1 Introduction 72

4.2 Results 76

4.2.1 M and L6 proteins are purified with ADP bound 76

4.2.2 Mutations within the NB-ARC domain that affect protein function have different nucleotide binding capacities 82

4.2.3 ATP hydrolysis analysis of M proteins 87

4.2.4 Do autoactive M proteins form oligomers? 91

4.3 Discussion 93

4.3.1 ADP bound is the "OFF" state 93

4.3.2 ATP bound is the "ON" state 94

4.3.3 Comparisons with published *in vitro* studies of R proteins 100

4.3.4 Oligomerisation in the ATP bound state? 101

4.3.5 Conclusion 102

CHAPTER 5: A STUDY OF M AND AVRM INTERACTION *IN VITRO* 103

5.1 Introduction 104

5.2 Results 106

5.2.1 Purification of AvrM, avrM and avrMC* 106

5.2.2 Co-immunoprecipitation experiments using purified recombinant proteins support the hypothesis of a direct protein-protein interaction between M and AvrM 111

5.2.3 *In vitro* experiments fail to identify nucleotide exchange in M induced by AvrM interaction 117

5.3 Discussion 119

5.3.1 Purification of flax rust effector proteins 119

5.3.2 Co-IP studies support a direct interaction between M and AvrM 120

5.3.3 The effect of the effector 120

CHAPTER 6: CONCLUSIONS AND FUTURE WORK 123

6.1 Overview 124

6.2 Model of M Protein Activation 125

6.3 A Model Stimulates Experimental Questions for Future Research 127

6.3.1 Determining the *in planta* phenotypes 127

6.3.2 Is ATP hydrolysis and/or oligomerisation involved in M protein function? 127

6.3.3 Interaction and recognition 129

6.3.4 Further characterisation of the NB-ARC domain, including other R proteins 131

6.3.5 Intramolecular interactions and protein structure 132

6.3.6 Intermolecular interactions and signalling 133

6.4 Conclusion 135

APPENDIX 137

Appendix 1: Sequencing Oligonucleotides 137

Appendix 2: Gel Filtration Calibration Curves 138

Appendix 3: Overview of the Truncation/Mutations of M Expressed in P. pastoris 139

Appendix 4: Purification of MΔLRR 140

Appendix 5: Supplementary Information for ATP and ADP Quantification Experiments 141

Appendix 6: AvrM/avrM/avrMC* Alignment 142

Appendix 7: Improved Expression 143

BIBLIOGRAPHY 145

List of Figures and Tables

FIGURE 1.1:	GLOBAL VIEW OF THE PLANT IMMUNE SYSTEM	5
TABLE 1.1:	PATHOGEN RECOGNITION RECEPTORS (PRRS) AND THEIR PAMP TARGETS	9
FIGURE 1.2:	A MODEL OF A FLAX CELL, WITH OR WITHOUT L5, AND ITS RESPONSE TO FLAX RUST CARRYING AVRL567A	14
TABLE 1.2:	NUCLEOTIDE BINDING SITE MOTIF CONSENSUS SEQUENCES	18
FIGURE 1.3:	AN NB-ARC ALIGNMENT	19
FIGURE 1.4:	THE MOLECULAR SWITCH MODEL FOR R PROTEIN ACTIVATION	30
TABLE 2.1:	CULTURE MEDIA	34
TABLE 2.2:	SOLUTIONS AND BUFFERS	35
TABLE 2.3:	YEAST AND BACTERIAL STRAINS	37
TABLE 2.4:	CLONING OLIGONUCLEOTIDES	37
TABLE 2.5:	PLASMIDS USED IN THIS STUDY	38
FIGURE 3.1:	EXPRESSION AND PURIFICATION OF RECOMBINANT M	57
FIGURE 3.2:	EXPRESSION AND PURIFICATION OF RECOMBINANT L6	58
FIGURE 3.3:	SMALL SCALE EXPRESSION AND PURIFICATION OF M FOR BIOCHEMICAL ANALYSIS	61
FIGURE 3.4:	PURIFICATION OF MΔTIR AND GEL FILTRATION ANALYSIS	63
FIGURE 3.5:	IMIDAZOLE AFFECTS MΔTIR MONOMERIC STATE	64
FIGURE 3.6:	GEL FILTRATION COMPARISON OF M, L6, AND L6ΔTIR	65
FIGURE 3.7:	PURIFIED M FULL-LENGTH MUTANTS	67
TABLE 4.1:	THE M AND M-MUTANTS ANALYSED *IN-VITRO*	75
FIGURE 4.1:	M AND L6 PROTEINS AFTER NIA PURIFICATION ARE ASSOCIATED WITH ADP	79
FIGURE 4.2:	GEL FILTRATION FRACTIONS CONTAINING M AND MΔTIR ALSO CONTAIN ADP	80
FIGURE 4.3:	PERCENTAGE OF M AND MΔTIR PROTEINS OCCUPIED WITH ATP OR ADP	81
FIGURE 4.4:	PURIFICATION OF M AND M MUTANTS FOR ATP AND ADP QUANTIFICATION	84
FIGURE 4.5	ADP BINDING IS DEPENDENT ON A FUNCTIONAL P-LOOP	85
FIGURE 4.6:	PUTATIVE AUTOACTIVE MUTANTS ARE PURIFIED WITH ATP BOUND	86
FIGURE 4.7:	ATPASE ACTIVITY IN GF FRACTIONS OF M AND M-MUTANTS	88
FIGURE 4.8:	M AND M^{K286L} PROTEINS HAVE INDISTINGUISHABLE HYDROLYSIS ACTIVITIES	89
FIGURE 4.9:	LONG WASH NIA PURIFICATION DOES NOT REMOVE BACKGROUND ATPASE ACTIVITY	90
FIGURE 4.10:	GF ANALYSIS OF M^{D555V} AND $M^{K286L+D555V}$	92
FIGURE 4.11:	A STRUCTURAL MODEL OF THE NB-ARC DOMAIN OF M	99
FIGURE 5.1:	PURIFICATION OF AVRM	108
FIGURE 5.2:	PURIFICATION OF AVRM	109
FIGURE 5.3:	PURIFICATION OF AVRMC*	110
FIGURE 5.4:	CO-IMMUNOPRECIPITATION OF M AND AVRM	114
FIGURE 5.5:	CO-IMMUNOPRECIPITATION OF M, AVRM AND AVRMC*	115
FIGURE 5.6:	CO-IMMUNOPRECIPITATION OF M, M^{K286L}, AVRM AND AVRMC*	116
FIGURE 5.7:	NUCLEOTIDE EXCHANGE ASSAY	118
FIGURE 6.1:	A WORKING MODEL OF M ACTIVATION	126

Declaration

I certify that this thesis does not incorporate, without acknowledgment, any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where due reference is made in the text.

Simon J. Williams

Acknowledgments

It is with great pleasure that I recognise the important people in my life that have helped bring this thesis to fruition.

Throughout this study my supervisor Dr. Peter Anderson has been an excellent supervisor, superb mentor and great friend. Passion, determination and commitment are words that I would use to describe Peter's quest to further understand the molecular function of an R protein. During my time in the Anderson laboratory I believe I have taken on these attributes towards this topic and this has been critical in the completion of the experiments and presentation of this thesis. Peter (and the Anderson lab), some time ago, decided to take a somewhat difficult scientific path due to Peter's genuine interest and commitment to this work. It has required immense patience and experienced a number of low points; however, I believe our efforts as a laboratory group are beginning to be justified by the scientific breakthroughs that we are now generating. Also I'd like to say a huge thank you, for your advice, prompt feedback and editing assistance, throughout the thesis write-up stage.

I would also like to express thanks to my co-supervisors Dr. Ian Menz and Dr. Simon Schmidt. Without Simon this project would not have been possible and he has provided me with sound advice and encouragement throughout this study. Ian's knowledge of biochemistry is something that I aspire too and without his experience and advice much of this work would not have been achieved.

A number of collaborators have also helped make this work possible. I am particular grateful to Dr. Peter Dodds for supplying me with various scientific materials, advice, opportunities, and encouragement. Also thank you to the Brisbane group; Professor Bostjan Kobe, Dr. Andi Wang and Thomas Ve, for their continued support throughout my thesis and the opportunities that have arisen due to their involvement in this project.

I am appreciative to the many members of the Anderson laboratory for providing an excellent working environment that has stimulated my growth as a researcher. I am particularly grateful for the help and support supplied by Pradeep Sornaraj and Emma deCourcy-Ireland. You both have shown guts and determination to undertake PhD projects that focus on R protein activation and a large amount of trust in the research system that we have been developing. I'm confident that this decision will prove a great one for your future. I am also very thankful to Emma for her assistance in the editing process.

Throughout my time as a student within the School of Biological Sciences I have made many great friends. I would like to mention Mel P, Bart, Drew and Nick for their assistance in my

experimental design and the numerous stimulating scientific discussions that have undoubtedly assisted in my development. I would also like to express gratitude to Keryn. We have shared countless cups of tea and chats throughout the entire PhD period. Your discussions frequently provided a welcome distraction from the realities of a PhD project and lot of laughter.

Thank you also to my many supportive friends, in particular Brad (Bunga); thanks for being a best mate and a lawyer/investment banker. Your generosity at the bar over the last four years and support of a struggling PhD student has been greatly appreciated (and noted), as has your continued friendship and support.

Thank you to my incredibly supportive family and extended family, in particular my parents John and Monica and sister Renée. Your never-ending, emotional and at times, financial support throughout this endeavour has enabled me to complete this thesis.

And finally to my partner in life, Jess. I'm unable to find the words that express my gratitude for your continued love and support over the last 4 years. Together I think we have proven that it is possible to maintain a relationship and two PhD's concurrently. Love you always.

Abstract

A plant's ability to detect an invading pathogen and circumvent a subsequent disease state is essential for its survival. Disease resistance, and the mechanisms behind it, are thus of critical importance. The pioneering work of Harold Flor, using the interaction between flax and the flax rust fungus, *Melampsora lini*, demonstrated that this ability to detect and resist the infection of a specific pathogen rests with two critical genes; a resistance (*R*) gene in the plant and a corresponding avirulence (*Avr*) gene in the pathogen. This, so called 'gene-for-gene' model, has subsequently been shown to apply in many other plant-pathogen interactions and has spawned considerable research efforts directed towards understanding the molecular basis of host-pathogen interactions and the consequential disease resistance response. Using the flax-flax rust pathosystem, and utilising a biochemical approach, this research has endeavoured to further the current understanding of the molecular basis of the interaction between plants and pathogens, with a particular focus on R protein function.

Chapter 3 describes the production of soluble, recombinant flax R proteins, M and L6, using the *Pichia pastoris* expression system. These flax R proteins can be purified from total cell lysates utilising a number of chromatography techniques. Following nickel affinity chromatography, concentration of protein in the presence of imidazole, leads to aggregation. This, however, can be alleviated by lowering the imidazole concentration prior to the protein concentration step. This fine tuning of the purification protocol enabled the expression and enrichment of near full-length and truncated versions of M and L6, and rational point mutations of M.

Utilising this expression and purification system, Chapter 4 presents a detailed functional study of the flax M protein, with particular focus on mutations that cause autoactivity and inactivity. These mutations were generated in the NB-ARC region of M with predicted loss- or gain-of-function consequences, as determined from the results of the *in planta* phenotypes of analogous mutations in other R proteins, in particular the flax L6 protein. Nucleotide quantification of purified wild type M and L6 demonstrated that these proteins are associated with ADP. Analysis of proteins with mutations within the NB-ARC domain demonstrated that this ADP binding is dependent on a functional P-loop in the NB subdomain. Mutations within the MHD motif and motif VIII that are predicted to result in an autoactive *in planta* phenotype, have more ATP associated with purified protein preparations in comparison to wild type. Taken together, these results further support the model that R proteins act as a molecular switch, whereby the

inactive form of the protein is ADP bound, while the active conformation of the protein is ATP bound.

Prior to this study yeast two hybrid analysis had demonstrated that a direct interaction between M and AvrM occurs. To investigate the interaction, and the consequence of interaction, between M and AvrM proteins *in vitro*, an expression and purification protocol was generated for AvrM (and variants) in Chapter 5. Here, a direct protein-protein interaction was supported by co-immunoprecipitation of purified M and AvrM proteins. The interaction that M has with AvrM is dependent on a functional P-loop and therefore presumably requires the presence of a bound nucleotide. The preferred model of R protein activation suggests that interaction with an effector causes the R protein to exchange its bound nucleotide from ADP to ATP. To determine if AvrM could induce nucleotide exchange, ADP/ATP exchange assays were performed, however, the results of this study were inconclusive. It is possible that nucleotide exchange is not the mechanism of activation of the flax M protein; although, it is equally likely that the conditions in the *in vitro* assay were not conducive for exchange to occur or that other proteins are needed to facilitate the exchange event. Whilst this study adds further proof to the theory of a direct interaction between flax rust effectors and their corresponding R proteins, the molecular effect that this event has on the R protein is yet to be understood.

In summary, only a small number of biochemical investigations of R proteins have been published, nevertheless, they have provided highly revealing information regarding R protein function. Utilising an *in vitro* approach, the results from this thesis provide further insight into the function and interaction between flax R proteins, and their effectors. It is hoped that the techniques developed and presented in this thesis will assist, and inspire, future *in vitro* investigations of flax R protein molecular function, and thus contribute to a wider understanding of plant disease resistance.

Abbreviation List

ADP	Adenosine 5'-diphosphate
Apaf-1	Apoptotic protease-activating factor 1
Avr	Avirulence
ATP	Adenosine 5'-triphosphate
CC	Coiled coil
CED-4	Cell death protein 4
CEX	Cation exchange
Co-IP	Co-immunoprecipitation
eLRR	extracellular LRR
ER	Endoplasmic Reticulum
ETI	Effector triggered immunity
GF	Gel filtration
HR	Hypersensitive response
LRR	Leucine-rich repeat
MAPK	Mitogen-activated protein kinase
NB-ARC	Nucleotide-binding adaptor shared by APAF-1, certain *R* gene products, and CED-4
NBS	Nucleotide binding site
NOD	Nucleotide oligomerisation domain
NiA	Nickel-metal ion affinity
NLR	NOD-LRR
PAMPs	Pathogen associated molecular patterns
PRRs	Pattern recognition receptors
PTI	PAMP triggered immunity
R gene / R protein	Resistance gene/Resistance protein
RLK-eLRR	Receptor-like kinases with extracellular leucine-rich repeats
SAR	Systemic acquired resistance
STAND	Signal transduction ATPases with numerous domains
T3SS	Type-III secretory system
TIR	Toll/Interleukin-1 Receptor
TLR	Toll-like receptors
TMV	Tobacco mosaic virus
Y2H	Yeast two hybrid

Chapter 1: Introduction

From this introductory chapter the following manuscript was produced for an invited review, this is currently submitted and in review:

Williams, S. J., Anderson, P. A., Kobe, B., Ellis, J. G. & Dodds, P. N. (2009) The Molecular Basis of Rust Resistance in Flax, *The Americas Journal of Plant Science and Biotechnology,* in review.

1.1 Overview of Plant Innate Immunity

Plants lack an adaptive immune system and rely on the capabilities of individual cells to detect and respond to invading pathogenic agents and thus prevent infection and disease. It is this innate immunity that enables them to survive and thrive in an environment where they encounter a broad range of potential pathogens with diverse life styles. A number of models have been proposed in recent years that attempt to explain how this innate immune system has developed during the evolutionary battle between pathogen and plant (Bent and Mackey, 2007, Jones and Dangl, 2006, Chisholm *et al.*, 2006). This chapter will introduce the data that supports, and in some cases refutes, these models, and will ultimately focus on a crucial plant protein that coordinates disease resistance.

1.1.1 Plant innate immunity: a global view

A plant has several layers of defence that it can employ in order to resist disease from pathogenic agents. These layers can be broadly divided into passive and active defence. Passive defence mechanisms include structural, chemical and biological plant components, such as the leaf waxy cuticle, lignified cell walls and anti-microbial/fungal compounds. These non-targeted defences provide the front line of plant defence and most likely prevent the vast majority of potential pathogen-related infections. Active defence mechanisms are those that are induced by the plant and thus invoke the need for some form of plant cellular machinery to cover roles of pathogen surveillance, detection, signal activation and response by the plant cell. It is this coordinated machinery that controls what is termed the plant innate immune system.

Conceptually, a plant's innate immune system can be divided into two lines of defence. The first line is historically known as basal defence. The basal defence mechanism utilises a broad detection system that is targeted towards pathogen/microbe associated molecular patterns (PAMPs or MAMPs). As a result, this line of defence is now commonly referred to as PAMP or MAMP triggered immunity (PTI). Detection is, in most cases, facilitated by extracellular transmembrane pattern recognition receptors (PRRs) which survey the apoplast on the look out for PAMPs. Upon detection of a molecular pattern, these PRRs are activated and communicate signals across the plasma membrane to the inside of the plant cell, switching on defence related pathways to achieve effective resistance.

Certain, more specialised pathogens have devised infection strategies that utilise effector molecules to aid in plant colonisation. Effectors are secreted from the pathogen and function in the apoplast or enter the plant cell to affect intracellular targets. Effectors from bacteria have been demonstrated to act by suppressing aspects of PTI. In response, however, plants have

evolved a second line of defence where specialised host proteins, known as resistance (R) proteins, can detect these specific pathogen effectors. Aptly named effector-triggered immunity (ETI), but historically known as R gene mediated resistance, these R proteins detect the pathogen and coordinate the disease resistance response. As one would expect, natural selection pressures force pathogens to diversify or shed recognisable effectors to evade ETI. This, inturn, has led to the diversification of R gene specificities to recognise modified or alternative effectors. This cyclical evolutionary battle between plant and pathogen continues to generate many pathogen strains with different combinations of effectors, and variant genotypes of plants with different combinations of R genes. Modern agricultural practices have increased the stakes in this evolutionary battle. A number of elegant and informative explanations of plant innate immunity have been presented in recent reviews and Figure 1.1 demonstrates these pictorially (Jones and Dangl, 2006, Bent and Mackey, 2007, Chisholm et al., 2006).

1.1.2 Amplitude of resistance

ETI is a more targeted plant defence strategy and tends to induce a stronger response than PTI. ETI is typified by the hypersensitive response (HR), which often culminates in programmed cell death of the infected cell, and to a limited extent, neighbouring cells (Greenberg, 1997, Jones and Dangl, 2006, Tao et al., 2003). Early studies of HR identified a number of inducible defence mechanisms. Ion fluxes, such as Ca^{2+}, into the cytoplasm are known to play an early role. Calcium has been shown to be associated with the production of reactive oxygen species (ROS) (causing an oxidative burst), and phytoalexin production (Greenberg, 1997, Heath, 2000). Phytoalexins (plant antibiotics), and inducible defence-related genes, have been linked directly to limiting pathogen growth. Signalling molecules, such as salicylic acid (SA), jasmonic acid (JA), ethylene (ET), nitric oxide (NO) and ROS, have all been shown to contribute to plant defence (Hammond-Kosack and Parker, 2003). Interestingly, a number of these same pathways are involved in PTI, suggesting significant signalling overlap between PTI and ETI (reviewed by (Altenbach and Robatzek, 2007, Schwessinger and Zipfel, 2008)). PTI has been shown to be tightly regulated, and there is growing evidence to suggest that for activation of ETI this negative regulation is released. This could explain, at least in part, the difference in response amplitude between ETI and PTI (Schwessinger and Zipfel, 2008). The importance of cell death during ETI is a contentious issue; however, plants that have been infected by a pathogen that induce ETI have a heightened state of readiness for further attack by the same or different pathogen(s). This general, long lasting defence response is referred to as systemic acquired resistance (SAR) (Gaffney et al., 1993, Ryals et al., 1996, Verberne et al., 2003). This thesis is concerned

predominately with the processes involved in ETI with a focus on R protein activation; however PTI will be discussed in more detail below.

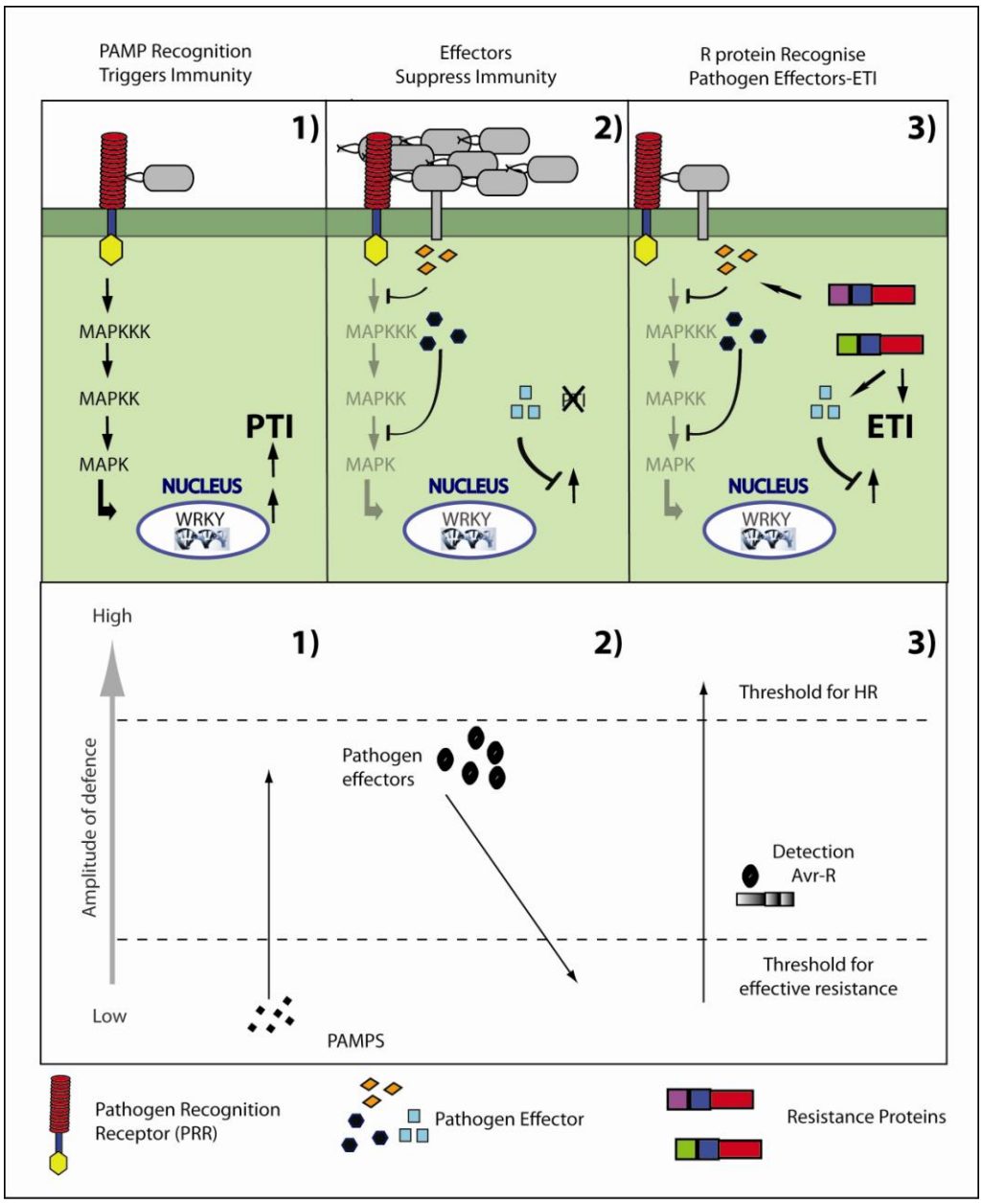

Figure 1.1: Global view of the plant immune system

This figure is adapted from (Chisholm et al., 2006, Jones and Dangl, 2006). It illustrates the potential evolution of resistance (in this case for a bacterial pathogen) combined with the amplitude of defence signalling within the plant. 1) Utilising its PRRs, the plant can recognise PAMPs and signal for the amplification of defence pathways for effective resistance (PTI). 2) Accordingly, pathogens have evolved genes and pathways to secrete proteins (effectors) into the cell where the virulence function enables the pathogen to evade or prevent PTI. 3) In order to maintain a competitive advantage, plants have evolved resistance genes, the products of which detect pathogen effectors and mediate signalling pathways leading to hypersensitive response and effector triggered immunity (ETI). Points 1, 2 and 3 are discussed in Sections 1.2, 1.3 and 1.4 respectively.

1.2 PAMP-Triggered Immunity

Plant PTI is facilitated by the perception capabilities of PRRs, which survey the extracellular space surrounding the cell wall on the look out for PAMPs. To date, only a handful of PRRs have been found (summarised Table 1.1), however, it is anticipated that this represents only a fraction of what exist (Nurnberger et al., 2004). The recognition of two common bacterial elements, flagellin and elongation factor TU (EF-TU), is controlled by different receptor-like kinases with extracellular leucine rich repeats (RLK-eLRR) (Chinchilla et al., 2006, Zipfel et al., 2006). LRRs are protein elements commonly found to facilitate intra- or inter-molecular protein-protein interactions (Kobe and Kajava, 2001). Their importance in pathogen perception and innate immunity in mammals, as a part of the Toll-like receptors (TLR), has been well studied. It is this LRR element found in plant PRRs and some R proteins, that is one of a number of similarities that can be found between human and plant innate immunity (reviewed by (Nurnberger et al., 2004, Rairdan and Moffett, 2007). The RLK-eLRR is, however, not the only class of PRRs. Other receptors identified so far, range from receptor-like proteins (RLP) with short cytoplasmic tails, to soluble proteins expressed extracellularly (Table 1.1). It is likely that such PRRs would associate with other RLKs or RLPs to facilitate further signalling (Zipfel, 2008). This has been demonstrated for the RLK-eLRR named BRI1-assocated kinase 1 (BAK1). BAK1, by interaction with BRI1, mediates the signalling of brassinosteroid hormones, which control many aspects of plant growth and development. Studies have demonstrated that BAK1 is also involved in signalling of both FLS2 and EFR (Chinchilla et al., 2007, Zipfel, 2008). Fundamentally, the role of the PRR is to perceive the potential pathogen and instigate the relay of a signal into the plant cell where it can be acted on.

Once the signal is transmitted to the inside of the cell signalling pathways are required to transduce it to the nucleus, where reprogramming of transcription enables the necessary action to be taken. In *Arabidopsis thaliana* mitogen-activated protein kinase (MAPK) cascades have been demonstrated to transfer the signal after flagellum perception by FLS2, which leads to the activation of WRKY transcription factors that assist in transcriptional reprogramming of the cell (Asai et al., 2002). This cascade facilitates resistance to both bacterial and fungal pathogens (Asai et al., 2002), confirming MAPKs as the signal transducers of PTI. Interestingly, MAPK cascades are also involved in signal transfer in mammalian innate immunity, identifying yet another link between plant and mammalian innate immunity (Nurnberger et al., 2004).

PTI is an ever expanding field of research and a number of recent reviews explore aspects of perception, signal cascades and pathway cross talk between PTI and ETI in more detail (Bent and Mackey, 2007, Nurnberger et al., 2004, Zipfel, 2008). Although PTI provides an important level of protection for a plant, some pathogens have evolved mechanisms to overcome

it. In many cases such pathogens secrete a suite of molecules into the infected plant cell to circumvent PTI and cause disease.

1.3 Pathogen Effectors

Plant pathogens have a wide range of infection strategies, however an emerging commonality between these pathogens is the secretion of proteins (including peptides and small compounds) that alter plant cell structure and function (Hogenhout et al., 2009, Panstruga and Dodds, 2009). These factors are collectively termed effectors. Effectors are generally thought of as pathogen elements that promote virulence during infection and have been found to function in both the apoplast and within the plant cell (reviewed by (Kamoun, 2006, Hogenhout et al., 2009). In cases where the effector is recognised by an R protein, the effector is said to be an avirulence (Avr) protein and its intended virulence function is compromised. The term avirulence, or Avr, is a genetic term that was first used to describe the phenotypic consequence on the pathogen when resistance was triggered in the plant. More recently, the term effector has been used to describe pathogen-derived proteins which include the products of *Avr* genes (Hogenhout et al., 2009). For the purposes of this thesis, the term "effector" will be used to indicate the product of an avirulence or *Avr* gene.

A large proportion of effectors have virulence functions within the plant cell, so understanding the mechanisms used by pathogens to internalise these effectors is very important. Perhaps the most well characterised delivery system to date is that of gram negative bacterium *Pseudomonas syringae*. These bacteria use a type-III secretory system (T3SS) to transport effectors into the cell where they can then attack host targets. A gene cluster, called *Hrp* (hypersensitive response and pathogenicity), encodes the machinery for a T3SS which includes the formation of an infection pilus to enable effector transfer (reviewed by (Büttner and Bonas, 2002, Galan and Collmer, 1999, Staskawicz et al., 2001)). Oomycetes provide the most well studied eukaryotic example of effector translocation. Rather then using a host derived secretory system; these pathogens encode effectors with specialised sequences that facilitate their translocation into a plant cell. Those that undergo plant cell internalisation have two conserved sequence motif, RxLR and EER, which have been implicated in translocation (reviewed by (Birch et al., 2006, Kamoun, 2006, Morgan and Kamoun, 2007)). Recently, both motifs were found to be required for effector delivery by the potato blight pathogen *Phytophthora infestans* (Whisson et al., 2007). The sequenced genomes of a number of oomycetes, including *P. sojae, P. ramorum, and P. infestans,* identified 400, 314 and 425 genes respectively, that encode secreted RxLR-EER classed-proteins (Tyler et al., 2006, Whisson et al., 2007). This

suggests that during infection, oomycetes have the potential to deposit a large and diverse array of effectors into the plant cell to aid in infection and help promote disease.

Research interests have now turned towards defining the potential virulence function of known effectors and also to define their host cell targets (Block et al., 2008). Suppression of host defence mechanisms seems an obvious target and a number of examples of this are emerging. He et al., (2006) demonstrated that two *P. syringae* effectors, AvrPto and AvrPtoB, were capable of suppressing early defence gene transcription and MAPK signalling (He et al., 2006). In fact, suppression of aspects of PTI appears to be a primary function of many characterised effectors (reviewed by (Jones and Dangl, 2006, Hogenhout et al., 2009)). It is also clear that effectors change aspects of plant development (Kay et al., 2007) and more recently a fungal effector was shown to suppress ETI (Houterman et al., 2008). The roles of pathogen effectors in host manipulation come from a range of functions including transcription regulation, protease activity and inhibition, protein degradation and protein phosphorylation (reviewed by (Block et al., 2008)). At present, however, a large number of cloned effectors still have not been characterised with regard to host cell targets and virulence function. Understanding the molecular function of many other effectors, and the common trends in their action (if any), will undoubtedly be essential if we are to fully understand the mechanisms used by different pathogens in plant infection and colonisation.

While effectors enhance infection, they also present potential targets for detection by their hosts. Consequently, an entire branch of plant immunity is based on effector perception for the activation of disease resistance. It is this function that is controlled by R proteins and is the focus of this thesis.

PRR	Structure	Plant	PAMP	Pathogen	Reference
FLS2 (Flagellin Sensing 2)	Receptor Like Kinase with extracellular LRR (RLK-eLRR)	Arabidopsis Tomato *N. benthamiana*	Flagellin	Bacteria	(Chinchilla et al., 2006, Gomez-Gomez et al., 2001)
EFR (EF-TU Receptor)	(RLK-eLRR)	Arabidopsis	Elongation factor-TU (EF-TU)	Bacteria	(Zipfel et al., 2006)
LeEIX 1 and LeEIX2 (Ethylene-inducing xylanase)	Receptor Like Proteins (RLP) with short cytoplasmic tail	Tomato	Xylanase	Fungi	(Ron and Avni, 2004)
chitin oligosaccharide elicitor binding protein (CEBiP)	A transmembrane protein with two extracellular LysM domains and a short cytoplasmic tail	Rice	Chitin	Fungi	(Kaku et al., 2006)
CERK1	A receptor like kinase with three LysM domains	Arabidopsis	Chitin	Fungi	(Miya et al., 2007)
b-glucan-binding protein (GBP)	Soluble protein (no transmembrane domain)	Soybean	Hepta-glucan	Oomycetes	(Fliegmann et al., 2004)

Table 1.1: Pathogen recognition receptors (PRRs) and their PAMP targets

Outline of known PRRs including their plant origin and the PAMP targets and pathogens they defend against.

1.4 Effector-Triggered Immunity

Effector-triggered immunity (ETI) is an inducible defence mechanism mediated by plant disease resistance (R) gene products. The first genetic interpretation of ETI was described by Flor in the 1930s-1950s while investigating the genetic basis of resistance in flax (*Linum usitatissimum*) to the fungal pathogen *Melampsora lini* (Flor, 1956). This investigation led to the development of the gene-for-gene theory; whereby, plant resistance to a specific pathogen is achieved only when a *R gene* in the plant and a corresponding avirulence (*Avr*) gene within the pathogen are present (Flor, 1956, Flor, 1971). The absence of either gene, ultimately leads to the plant's susceptibility to infection and the development of disease symptoms. R genes have been identified that control resistance to an array of pathogens including bacteria, viruses, fungi, oomycetes, nematodes and insects. Identifying and cloning both resistance and avirulence genes and investigating the interplay between the gene products has become a major research objective for plant scientists, and is the focus of this thesis.

1.4.1 R genes

R genes have been reviewed extensively since the cloning of the first resistance gene, *Pto* (Martin *et al.*, 1993), which controls resistance to tomato bacterial speck disease. Currently over 40 isolated R genes from numerous plant species have been cloned (reviewed by (Dangl and Jones, 2001, Ellis et al., 2000, Hammond-Kosack and Jones, 1997, Chisholm et al., 2006)). With such a number of R genes now isolated and cloned, attention has now focussed towards understanding how the proteins they encode perceive effectors and activate signal pathways to achieve ETI. Bioinformatic tools have aided in this objective by assigning R genes into structural classes depending on their domain organisation. Despite some outliers, R genes can be divided into two broad classes; genes that encode products with extracellular leucine-rich repeat (eLRR), presumably involved in the recognition of effectors in the apoplast, and those that encode tri-domain intracellular nucleotide binding (NBS)-LRR proteins, shown to recognise internalised effectors protein (Chisholm et al., 2006).

1.4.1.1 The eLRR R genes

The R genes containing eLRRs can be further subdivided depending on the presence and origin of their intracellular signalling domain (Chisholm et al., 2006). They share similarities in domain organisation with a number of the PRRs (see 1.2), and also draw comparison with pathogen receptors from mammalian innate immunity (Nurnberger et al., 2004). Arguably, the most well characterised example of this broad class are the tomato *Cf*-genes that confer resistance to the

tomato leaf mould pathogen, *Cladosporium fulvum* (reviewed by (Rivas and Thomas, 2005)). This thesis, however, involves the study of an NBS-LRR protein, and it is this class of R proteins that will therefore be the main focus of this chapter.

1.4.1.2 NBS-LRR

The NBS-LRR proteins are generally trimodular and are distinguished by either a coiled-coil (CC) or Toll-interleukin 1 receptor (TIR)-like domain at the N-terminus. The NBS-LRR class is the most predominant class of plant disease R proteins, and the genes which encode NBS-LRR proteins account for approximately 150, 400 and over 500 genes in the *A. thaliana*, poplar and rice genomes, respectively (Monosi *et al.*, 2004, Tuskan *et al.*, 2006, Meyers *et al.*, 2003). It is clear that NBS-LRR proteins are central to disease resistance, as members of other R protein classes have been demonstrated to rely on NBS-LRR proteins for function and/or immunity. For example, the tomato cytoplasmic serine/threonine protein kinase resistance protein, Pto, requires Prf, an NBS-LRR-like protein, for resistance to strains of *P. syringae* (Salmeron et al., 1996). NBS-LRR proteins have also been shown, in two separate studies, to act downstream of recognition in HR-mediated defence. Peart et al., (2005) showed that NRG1, which encodes a CC-NBS-LRR protein, acts downstream of the TIR-NBS-LRR protein N to facilitate tobacco's resistance to strains of tobacco mosaic virus (TMV) (Peart et al., 2005). They also demonstrated that NBS-LRR proteins are required in signal pathways that lead to HR and theorised this maybe a general feature of such R proteins. This idea was supported when *NRC1*, also encoding a CC-NBS-LRR protein, was cloned from tomato. *NRC1* was implicated in HR signalling by the eLRR R protein Cf4, responsible for tomato resistance to strains of *C. fulvum* (Gabriels et al., 2007). *NRC1* is important for resistance to a range of diseases in tomato and has been found to act downstream of a number of R genes including *Pto, Cf-9, Rx and Mi,* as well as the PAMP receptor *LeEIX* (Gabriels et al., 2007). This adds support to the hypothesis raised by Peart et al., (2005) who suggest that any form of resistance protein may require downstream NBS-LRR proteins for cell death signalling (Peart et al., 2005) and also provides a possible link between PTI and ETI (Gabriels et al., 2007).

Given that many pathogens secrete their effector molecules into the cells of their host, R genes that control disease resistance to these pathogens that encode NBS-LRR proteins are expected to function within the host cell. This prediction has been confirmed by a number of localisation studies. *A. thaliana* R proteins RPS2 and RPM1, both CC-NBS-LRR proteins, localise to the inner leaflet of the plasma membrane (Axtell and Staskawicz, 2003, Boyes et al., 1998) while the CC-NBS-LRR barley resistance protein Mla is cytosolic (Bieri et al., 2004).

More recently, nuclear localisation of Mla has been demonstrated and this localisation been shown to be required for resistance against the powdery mildew fungus *Blumeria graminis* f. sp. *hordei* (Shen et al., 2007). A number of NBS-LRR R proteins including N, RRS1-R, Rx and RPS4 have also been shown to localise to the nucleus (Burch-Smith et al., 2007, Tameling and Baulcombe, 2007, Wirthmueller et al., 2007, Deslandes et al., 2003) spawning a number of recent reviews (Liu and Coaker, 2008, Shen and Schulze-Lefert, 2007). The *L* and *M* genes from flax encode a hydrophobic N terminal region that despite being predicted to act as a signal peptide (Schmidt et al., 2007a), may instead facilitate plasma membrane association or membrane anchoring (unpublished data). This is not unlike the case of RPP1A which has an N-terminal element that resembles a signal peptide but instead localises to cellular membranes, the endoplasmic reticulum (ER) and/or Golgi system (Weaver et al., 2006).

NBS-LRR proteins are undoubtedly critical components of the plant immune system and understanding the roles they play in the coordination of the resistance response is a crucial step in our understanding of plant immunity. This study focuses on the flax R proteins, M and L6, which are members of the NBS-LRR protein class, and are responsible for resistance to strains of flax rust (*M. lini)*. Subsequent sections within this chapter will therefore focus primarily on the function of the NBS-LRR R proteins; however, prior to this, the flax-flax rust pathosystem will be introduced.

1.5 The Flax, Flax Rust System

1.5.1 The flax rust pathogen

Rust fungi obtain their nutrients exclusively from living plant cells and are thus defined as obligate biotrophs. The flax rust pathogen, *M. lini,* has an infection strategy that involves a number of steps and complex signalling mechanisms (reviewed by (Heath, 1997, Lawrence et al., 2007)). In short, fungal spores germinate on flax leaves and extend a germ tube out across the leaf until a stomatal entry point is discovered. An appressorium is then formed from which an entry peg extends down between the guard cells. Infection hyphae then grow into the mesophyll layer of the plant leaf, where haustorial mother cells subsequently form. From these cells, haustorial feeding structures penetrate the plant cell wall and invaginate the cell (Heath, 1997). The plant cell membrane is not breached during this process and the region formed between the plant and haustorial membrane is termed the extra-haustorial matrix. Haustoria are specialised feeding structures that are vital for both the acquisition of nutrients from the plant cell, and their conversion into useful metabolites to sustain the fungus (Hahn and Mendgen, 2001, Sohn et al., 2000). This, in effect, is the front line of rust infection and the genetic determinants that both the

rust and its host possess are critical in determining the cross talk between pathogen and host, and thus the outcome of the interaction (Figure 1.2).

1.5.2 Flax R genes and rust effector proteins

Thirty one resistance genes, each conferring resistance to different strains of *M. lini,* have been reported in flax. They are confined to 5 separate loci, designated, K, L, M, N and P. A total of 19 different R genes have now been cloned with all encoding proteins of the Toll-Interleukin 1 Receptor-like, nucleotide binding site, Leucine rich repeat (TIR-NBS-LRR) class (Anderson et al., 1997, Dodds et al., 2001a, Dodds et al., 2001b, Ellis et al., 1999, Lawrence et al., 1995, Lawrence et al., 2009).

Genetic studies have defined approximately 30 *Avr* genes, of which alleles of four have been cloned, sequenced and functionally tested (Dodds and Thrall, 2009). These include *AvrL567, AvrM, AvrP123 and AvrP4* (Catanzariti et al., 2006, Dodds et al., 2004). *Avr* genes are expressed in the rust haustoria and encode small soluble proteins with N-terminal signal peptides, although, unlike the R proteins, there is little sequence similarity between different effector proteins, and no identifying sequence motifs that give an insight into their potential function (Ellis et al., 2007a). They do, as mentioned, have an N-terminal signal peptide indicating that they are secreted from the haustoria into the extra-haustorial matrix. It is anticipated that this sequence would be cleaved from the mature protein following secretion. Once in the extrahaustorial matrix, effectors need to be translocated into the plant cell where they can interact directly, as outlined below, with their corresponding intracellular R protein. Although translocation of AvrL567, AvrM and AvrP4 across the plant plasma membrane has been demonstrated the mechanism by which this occurs is unknown and is currently being investigated (Catanzariti et al., 2006, Dodds et al., 2004). Furthermore, the host targets of these effectors and their function during infection are at this stage unknown.

An extremely important outcome of the study of flax R proteins and their cognate rust effectors has been the discovery that they interact directly (Dodds et al., 2006) (see 1.7.3). This, coupled with the cloning and expression of flax R and flax rust effector proteins, makes the flax-flax rust pathosystem an excellent model to investigate elements of NBS-LRR protein function, including both interaction and activation.

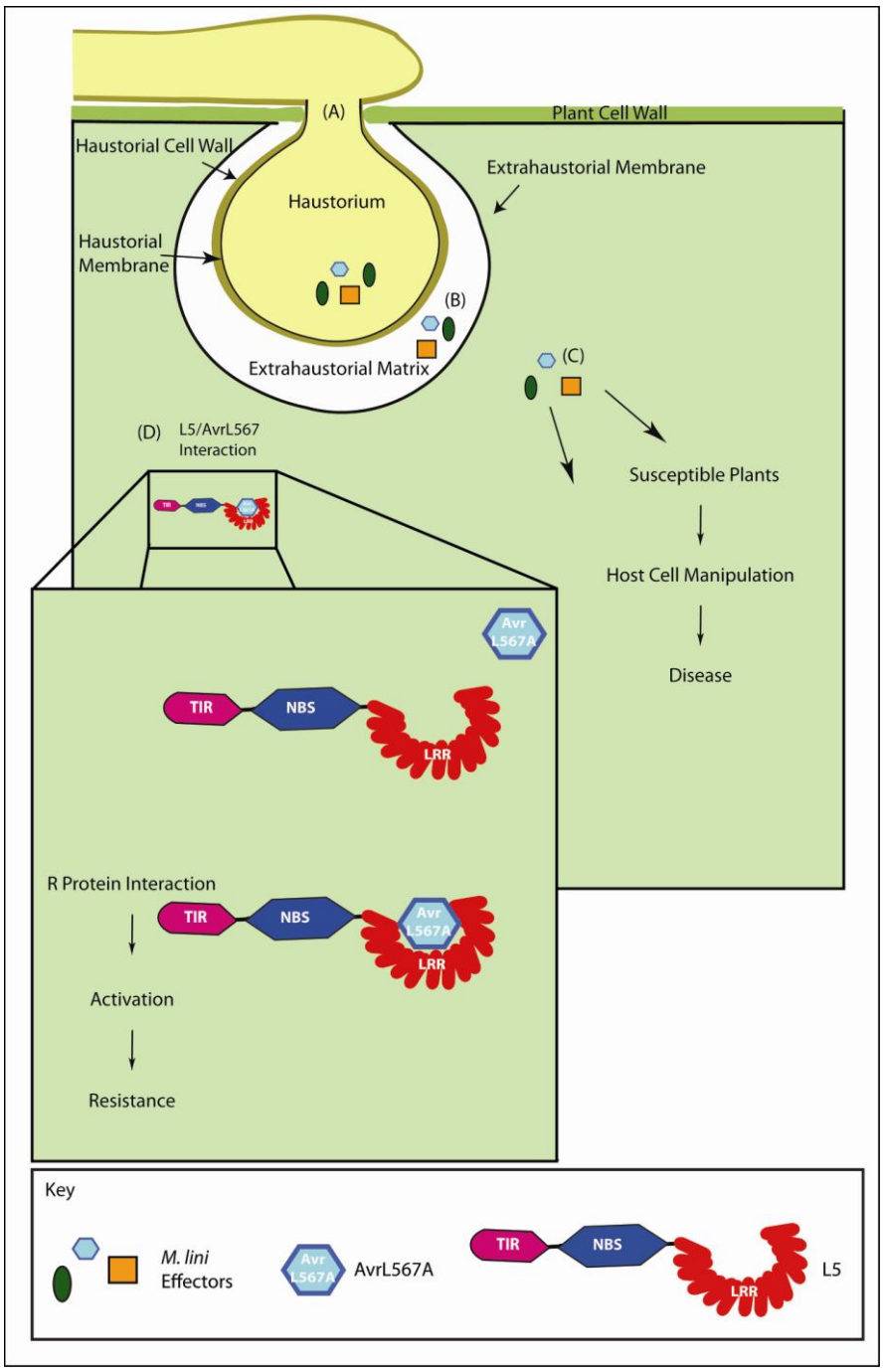

Figure 1.2: A model of a flax cell, with or without L5, and its response to flax rust carrying AvrL567A

A) The haustorial feeding structure penetrates the cell wall and establishes an extrahaustorial matrix. B) Fungal effector proteins are secreted from the haustoria and gain entry to the plant cell by an, as yet, undetermined mechanism. C) In the absence of a corresponding resistance protein, the effectors presumably manipulate host targets to promote disease; however, both the targets and molecular function of effectors are unknown at this point. D) When a corresponding R protein is present, in this case L5, it interacts directly with the effector AvrL567A and activates resistance (see 1.7.3).

NB: The protein domain labels in L5 are abbreviated as in text.

1.6 Proposed Function of the NSB-LRR R Protein Domains

The structural domains predicted within the NBS-LRR R proteins provide the first clue for interpretation of their function. Early hypotheses suggested that the C-terminal LRR is involved in pathogen perception; the central NB domain controls protein activation and the N-terminal CC or TIR enabled transduction of a signal, or signals, to activate the HR and other defence responses. Whilst this assignment of function to separate domains has provided a useful predictive framework for experimental design, it appears to be an oversimplification of how R proteins work, and more intricate interactions between domains are likely to control tertiary protein structure and function.

1.6.1 Leucine-rich repeat

The LRR domain provides a structural platform for protein-protein interactions in a wide range of diverse proteins (Bella et al., 2008, Kobe and Kajava, 2001). The first structure of a LRR protein came from the human and porcine ribonuclease inhibitor (Kobe and Deisenhofer, 1993). It revealed that individual repeats, shown to range from 20-29 residues, make up a structural unit. The structural unit generally consists of β strand and α-helix linked by either a β-turn or loop (Bella et al., 2008, Kobe and Kajava, 2001). The β strand comprises the characteristic repeat motif xxLxLxx (where x can be any amino acid and L indicates a conserved leucine, but can be replaced with a valine, isoleucine and phenylalanine). The overall structure resembles that of a curved solenoid, with parallel β strands on the concave side, and the helical or β turn elements on the convex side, whereby each repeat is a turn of the solenoid (Bella et al., 2008, Kobe and Kajava, 2001). The concave side is generally thought to be the ligand binding interface, and changes to the x residues exposed to the solvent would be predicted to alter binding site dynamics.

Consistent with this idea is the fact that the most variation between R genes and their closely related homolog's, resides within the LRR domain, particularly in sequences encoding the repeat motif (Ellis et al., 2000). With selection for amino acid variation occurring at these sites, it was predicted that the LRR domain provides the most likely site for effector binding. This, however, is the likely scenario for only a few effector-R protein interactions (see 1.7.3).

In a number of studies of NBS-LRR class of R proteins, the LRR has been implicated in numerous functions other than direct effector interaction. LRRs facilitate the direct interaction with the chaperones Hsp90 and Hsp75 and the co-chaperones, protein phosphatase 5, SGT1 and RAR1, which have all been demonstrated to be required in the positive regulation of R proteins in a pre-activated state (reviewed by (Padmanabhan et al., 2009, Shen and Schulze-Lefert, 2007)). Apart from providing interactions with other proteins, the LRR appears to be

important in maintaining intra-molecular interactions and has been implicated in both the negative and positive regulation of activation in a number of R proteins (see 1.8.2) (Hwang and Williamson, 2003, Hwang et al., 2000, Moffett et al., 2002, Rairdan and Moffett, 2006). With such an array of data presented regarding the function of the LRR domain in R proteins, it is clear that both structural and functional studies of purified R proteins would further clarify its role.

1.6.2 Nucleotide binding site (NBS)

NBS domains are common protein elements involved in the activation of proteins through the catalysis of nucleotide hydrolysis and/or the conformational changes in protein structures that are induced by nucleotide binding. R proteins have an NBS domain that contains the hallmark characteristics commonly found in nucleotide binding proteins (Meyers et al., 1999, Traut, 1994). Conserved sequence motifs within NBS domains facilitate catalytic and/or ligand binding activity. R proteins have a number of motifs known to be critical and highly conserved in ATP and GTP binding proteins, such as the P-loop (Kinase 1 or Walker A) and Kinase 2 (Walker B) (Saraste et al., 1990, Walker et al., 1982). Whilst the number of defined motifs varies depending on interpretation of sequences, ten motifs within the NBS region of R proteins have been identified and these are summarised in Table 1.2. Although the biochemical function of these motifs in R proteins remains mostly undefined, many have been demonstrated to have important *in-planta* function, with changes to key residues causing both loss- and gain-of-functions (elicitor-independent) phenotypes (Figure 1.3). Nucleotide binding is clearly critical for R protein function, as mutations to residues in motifs predicted to be involved in this function, particularly within the P-loop, cause loss-of-function.

Interestingly, conserved motifs in R proteins are also conserved in the mammalian apoptotic protease-activating factor 1 (Apaf-1) and *Caenorhabditis elegans* cell death protein-4 (CED-4). Consequently, the R protein NBS domain is most commonly termed the NB-ARC, standing for a nucleotide-binding adaptor shared by APAF-1, certain *R* gene products, and CED-4 (van der Biezen and Jones, 1998a). The NB-ARC nomenclature will be used in the place of NBS for the remainder of this thesis. R proteins have also been included in a much broader protein class known as signal transduction ATPases with numerous domains (STAND) (Leipe et al., 2004). The STANDs are generally signalling hubs with functions ranging from, mediators of cell death and inflammation, to regulators of transcription. Mandatory to inclusion in this classification is the NOD (nucleotide binding and oligomerisation domain) module, or NB-ARC domain in R proteins, which is closely related to that found in the AAA+ ATPases (Leipe *et al.*, 2004). The STANDs include five major clades differentiated by key features and/or conserved motifs within the NOD module. Apaf-1, CED-4 and R proteins are linked to the AP-ATPase clade,

while the sister clade, NACHT (named after numerous nucleotide binding proteins with similar domains (Koonin and Aravind, 2000)), include the animal NLRs (NOD-LRR), a family of proteins which are involved in human innate immunity and inflammation. Similarities have historically been drawn between animal NLRs and R proteins due to their tridomain architecture, whereby a NOD domain is often linked with a C-terminal repeat domain and a proposed N-terminal signalling domain (Rairdan and Moffett, 2007). A number of conserved motifs are defined within the NOD module of the STAND proteins; however, the P-loop (Walker A) and Kinase II (Walker B) motifs are the most highly conserved. It is predicted that the NOD module will form a similar structural architecture in all STAND proteins, and it is proposed that the mechanisms used to activate these proteins are likely to be conserved (Danot *et al.*, 2009). The classification of R proteins in the STAND group is important as it enables parallels to be drawn from the biochemical and structural studies of other STAND proteins, (discussed further in section 1.8) providing a potential guide, without being too subjective, for future biochemical investigations into R protein function.

The first biochemical study of the NB-ARC domain of an R protein came from a pioneering study on the tomato R genes *I-2* and *Mi1*. Recombinant CC-NB-ARC (truncated for the LRR) proteins expressed in, and purified from, *E. coli*, were demonstrated to have ATP binding and hydrolysis capabilities (Tameling et al., 2002). Importantly, when the invariant lysine (Figure 1.3) within the P-loop was mutated to an arginine, both binding and hydrolysis capabilities of the recombinant protein were significantly reduced. The same mutation in the R gene caused a loss of activity *in-planta,* indicating a role for ATP binding and/or hydrolysis in R protein function. ATP binding and hydrolysis activity associated with the P-loop motif has also been shown from *in-vitro* studies of the NB-ARC-LRR region of tobacco N resistance protein (Ueda et al., 2006). The work by Tameling and co-workers and the mutational analysis of others (Figure 1.3) has confirmed that activation of R proteins, at least in part, is achieved by the function of the NB-ARC domain (Tameling et al., 2002).

Domains	**NB (~180aa)**						**ARC1 (~75aa)**	**ARC2 (~105)**		
Motifs	hhGRExE	P-loop	RNBS-A	Kinase II	RNBS-B	RNBS-C	GLPL	Motif VIII	RNBS-D	MHD
Consensus	hhGRExE	GVG**K**TT	FLENIRExSKKHGLEHL QKKLLSKLL (FDLxAWVCVSQxF)	LLVLDDVW	GSRIIITTRD	YEVxxLSEDEA WELFCKXAF	GLPL	SYD	FLHIACFF (CFLYCAL FPED)	MHD

Table 1.2: Nucleotide binding site motif consensus sequences

The NB-ARC spans ~360 amino acid residues and can be separated into three domains based on the crystal structure of Apaf-1 (Riedl et al., 2005). These are designated NB, ARC1 and ARC2. Consensus sequences for the plant NB-ARC-LRR resistance proteins have been interpreted from a number of studies (Leipe et al., 2004, Meyers et al., 1999, Meyers et al., 2003, Pan et al., 2000, Takken et al., 2006, van der Biezen and Jones, 1998a) and are summarised in (Takken et al., 2006). A Kinase-3 domain is absent from the plant R protein nucleotide binding domain and is therefore not shown above, however, the RNBS-B motif shares positional but not sequence similarity with known Kinase-3 domains.

The non-TIR and TIR classes of NBS-LRR genes vary in the RNBS-A and -D motif regions, hence the non-TIR consensus is in brackets.

Bold type in the consensus sequence row is the invariant amino acids of the P-loop- lysine; in the consensus the underlined residues are highly (almost invariantly) conserved and positions of functional interest.

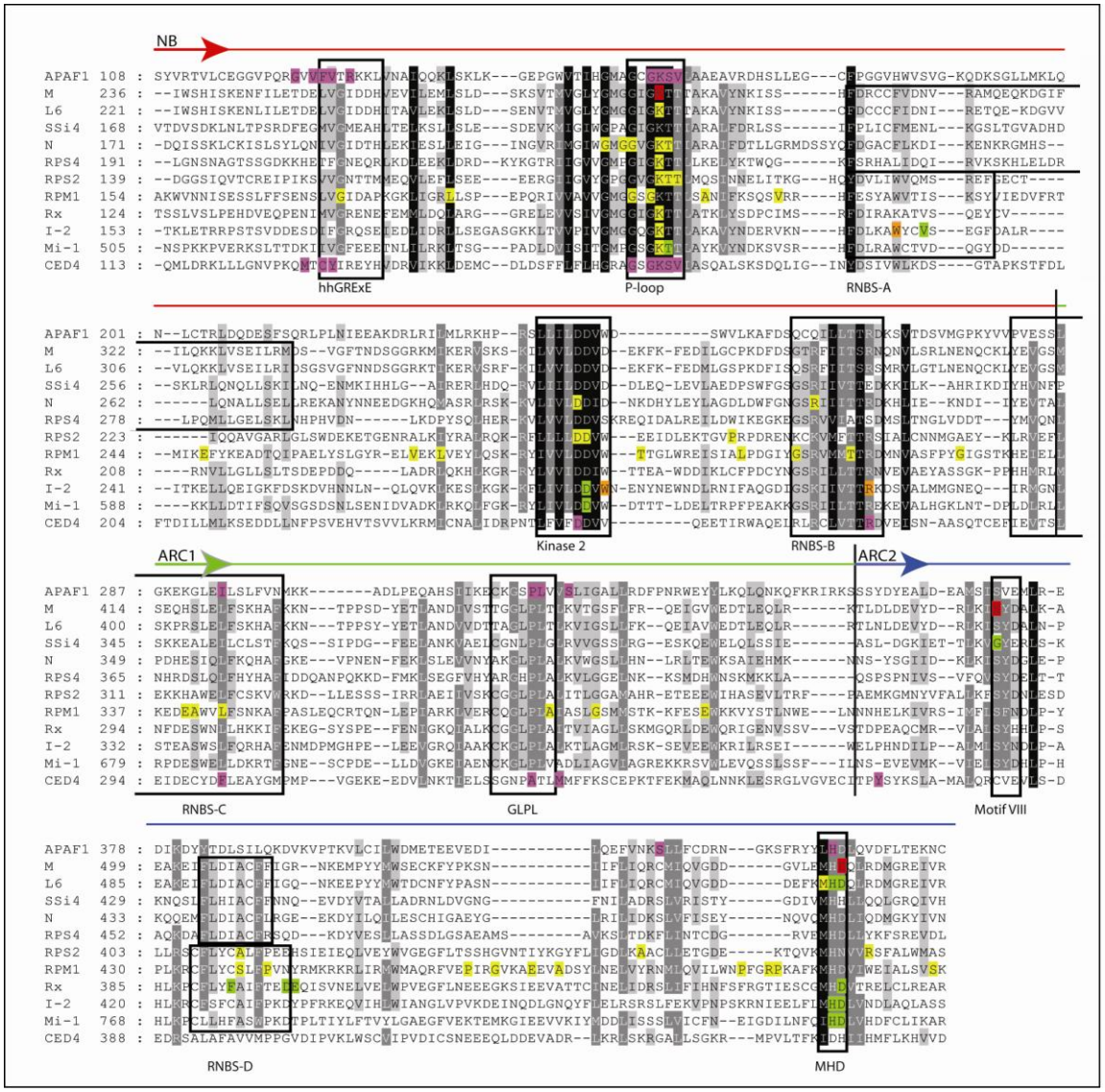

Figure 1.3: An NB-ARC alignment

This figure is adapted from (van Ooijen et al., 2008b). It shows a multi-sequence alignment of the NB-ARC region of numerous R proteins, Apaf-1 and CED-4. TIR-NB-ARC-LRR R proteins included in the alignment are M/L6 (flax), SSi4/RPS4 (*A. thaliana*) and N (tobacco). The non-TIR R proteins include RPS2/RPM1 (*A. thaliana*), Rx (potato) and I-2/Mi-1 (tomato). The predicted R protein subdomain borders, as determined by (Albrecht and Takken, 2006), are defined with a line that runs the length of the subdomain, NB: red; ARC1: green; ARC2: blue. Residues that are involved in coordinating the binding of either ADP or ATP in respective 3D structures of Apaf-1 (1z6t) or CED-4 (2a5y) were determined from the referenced source (Riedl et al., 2005, Yan et al., 2005) and using the ligand interaction program interface on the protein data bank website (http://www.rcsb.org/pdb), are highlighted in purple. Ten motifs are defined (see table 1.2) and are labelled/ boxed. Residues that have been demonstrated to be critical in respective R protein function for these proteins are highlighted; yellow indicates loss-of-function and green indicates gain-of-function. Residues highlighted in orange are predicted to cause loss of function. The sources of this information are listed as follows L6 (Howles et al., 2005), SSi4 (Shirano et al., 2002), N (Dinesh-Kumar et al., 2000), RPS4 (Zhang et al., 2004), RPS2 (Axtell et al., 2001, Mindrinos et al., 1994, Tao et al., 2000), RPM1 (Grant et al., 1995, Tornero et al., 2002), Rx (Bendahmane et al., 2002), I-2, (de la Fuente van Bentem et al., 2005, Tameling et al., 2006, van Ooijen et al., 2008b), Mi-1 (van Ooijen et al., 2008a, van Ooijen et al., 2008b).

1.6.3 Amino terminal domain

As discussed above, the majority of R proteins contain either a CC or TIR domain at their N-terminus. The CC domain is a common structural domain involved in an array of different biological processes including protein-protein interactions (Martin et al., 2003). A number of studies have shown that the CC domain coordinates the interaction between an R protein and host proteins that are targeted by effectors (see section 1.7.2) (Ade et al., 2007, Mackey et al., 2002). The CC is also required for downstream signalling by the *A. thaliana* CC-NBS-LRR protein, RPS5 (Ade et al., 2007, Shen et al., 2007) and is subsequently implicated in both detection and signalling.

The TIR domain is so named due to its similarity with the cytoplasmic signalling domain utilised by Toll and Interluekin-1 (IL-1) like receptor proteins. In these mammalian receptors, the cytoplasmic domain provides a protein scaffold for protein-protein interactions and is indispensable for signalling (Martin and Wesche, 2002, Xu et al., 2000). These sequence similarities suggest that the plant TIR domain may provide the same function, and this has been supported by a number of functional studies. Deletions and point mutations within the TIR, shown to affect *Toll* and IL-1R signalling, also affect the *N*-mediated signalling events that lead to TMV resistance in tobacco (Dinesh-Kumar et al., 2000). Also, over-expression of the TIR domain, including a short C-terminal extension, of the flax R protein L10, and also *A. thaliana's* RPS4 and RPP1A, caused an effector-independent necrotic response in *A. tumefaciens* transient assays in tobacco, and *Nicotiana benthamiana* leaves, respectively (Frost et al., 2004, Swiderski et al., 2009).

Homotypic interactions between TIRs and/or TIR dimerisation seem essential for signalling activation of Toll-like receptors (Takeda and Akira, 2005). TIR-mediated oligomerisation also appears to be involved in at least one member of the TIR-NBS-LRR class of R proteins. Transient expression experiments of the tobacco N disease resistance protein, followed by co-immunoprecipitation, demonstrated that the N protein oligomerises in the presence of the TMV elicitor p50 (Mestre and Baulcombe, 2006). This oligomerisation event was abolished in N protein containing a mutation of the conserved lysine of the P-loop motif. It is not known, however, if this effector triggered R protein oligomerisation is a general feature of the resistance response mediated by the TIR-NBS-LRR class. In the case of the tobacco N protein, the TIR may also play a regulatory role in activation through intramolecular interactions (Ueda *et al.*, 2006) and also in pathogen perception, through interactions with host proteins (Burch-Smith et al., 2007, Caplan et al., 2008).

Collectively, these results indicate that the TIR is likely to play important roles in inter- and intra-molecular interactions and regulation of R protein activity. It also contributes to a growing list of similarities that exist been plant and animal innate immunity. It does, however, further expose a recognised need for the biochemical study of TIR and CC domains, in the context of intact full-length R protein, to elucidate the precise role these domains play in R-mediated resistance.

1.7 Effector Perception by R Proteins

As discussed above, the presence of a particular R gene is vital for resistance to specific pathogens, but what role do the products of R genes play in pathogen detection? R proteins must recognise the presence of a foreign pathogen-derived effector protein. Such recognition must presumably lead to R protein activation, the result of which is the transduction of a signal, or signals, to promote defence responses. The remaining sections will tackle R protein function, covering, with examples, recognition, activation, and signalling.

1.7.1 Recognition of effector proteins by R proteins

Arguably the most important mechanism to decipher in plant disease resistance is the way in which an R protein recognises a pathogen derived effector. Conceptually, the simplest mechanism of interaction between R and effector proteins is a direct one, whereby the R protein acts as a receptor, and the effector protein as a ligand. Whilst this is true in some cases, many R-effector interactions characterised to date recognise the effector's presence by monitoring the integrity of host proteins. In this section, evidence that demonstrates both types of recognition will be introduced.

1.7.2 Indirect recognition: the guard and decoy hypothesis

The idea that an R protein may be guarding a host target of its corresponding effector protein, was first used to explain the dynamics between tomato Pto and Prf, which provide immunity to *P. syringae* carrying AvrPto (van der Biezen and Jones, 1998b). Whilst Pto has been demonstrated to interact directly with AvrPto (Scofield et al., 1996, Tang et al., 1996), Prf is also required for resistance. An array of evidence places Prf very early in the Pto-mediated resistance pathway (Martin et al., 2003, Rathjen et al., 1999). Pto forms a molecular complex with a unique N-terminal region of Prf, and Prf has been shown to have signalling, regulation and recognition capabilities (Mucyn et al., 2006). It is suggested that Prf monitors the integrity of Pto, a target of the virulence activity of AvrPto (van der Biezen and Jones, 1998b). A review by Dangl and Jones (2001) generalised this theory into the guard hypothesis. It predicts that the Avr-Pto product, functioning as an effector, targets a host protein to promote disease. The R protein functions by

protecting the integrity of the host target, thus providing the plant with an effective means of detecting the pathogen and activating an immune strategy (Dangl and Jones, 2001). Support for the guard hypothesis has come from a number of studies in which *P. syringae* is the infectious agent, and effectors target an array of "guarded" host proteins or guardees, (reviewed most recently in (Jones and Dangl, 2006)). This list now includes a *P. syringae* protease AvrPphB which cleaves the *A. thaliana* host protein PBS1 to activate the CC-NBS-LRR protein RPS5 (Ade et al., 2007). PBS1 and RPS5 are required for resistance to AvrPphB, and their interaction is mediated by the CC domain of RPS5 (Ade et al., 2007). Another example is the *A. thaliana* R protein, RPM1, which interacts through its N-terminal domain with the guardee RIN4, (Mackey et al., 2002). RIN4 is the target for the pathogen effector protein AvrRpm1 from *P. syringae*

Recently, the guard model has come under some scrutiny. It has been proposed that the guardee may in actual fact be a decoy protein that mimics an effector target (van der Hoorn and Kamoun, 2008). The "decoy" model has been generated in an attempt to cover some potential inconsistencies associated with the guard model. The authors argue that some effectors have multiple host cell targets, and that pathogen virulence activity may not require alteration of a guardee. A guardee would therefore have opposing selection pressures depending on the presence or absence of an R gene. When an R gene is present, the guardee has selection pressures towards interacting with an effector, however in the absence of an R gene, natural selection would drive the guardee to decrease its binding affinity with an effector. The decoy model has been put forward to potentially solve this evolutionary discrepancy as a decoy protein would evolve to always maintain its interaction with an effector and not be compromised by this interaction (van der Hoorn and Kamoun, 2008). The main difference between a decoy and a guardee is that the decoy is not required in host resistance, and that alteration of the decoy does not result in an enhanced fitness to the pathogen when the R protein is not present (van der Hoorn and Kamoun, 2008). The guard and decoy models may not be mutually exclusive, and both models provide an excellent framework to further decipher the indirect interaction between R and Avr proteins in the future.

1.7.3 Direct recognition

To date, four cases of direct interaction between effectors and NB-ARC-LRR R proteins have been reported, all using the yeast two hybrid (Y2H) system. The R genes involved in these interaction tests include *RRS1-R* from *A. thaliana*, *L5*, *L6* and *L7* from flax, *Pi-Ta* from rice and *N* from tobacco and their corresponding pathogen and the pathogen effectors, *Ralstonia solanacearum,* effector Pop2, *M lini*, effector AvrL567, *Magnaporthe grise*, effector Avr-Pita and tobacco mosaic virus, effector p50, respectively (Deslandes et al., 2003, Dodds et al., 2006, Jia et

al., 2000, Ueda et al., 2006). In all cases, the LRR domain is required for interaction with the effectors. Two of these interactions have been supported by *in vitro* protein binding assays, however in both cases the interaction could only be demonstrated using truncations of full-length proteins containing only the LRR or NBS-LRR domains (Jia et al., 2000, Ueda et al., 2006).

The LRR has been clearly demonstrated to provide pathogen strain-dependent specificity within the gene-for-gene system. The general role of LRRs in protein-protein interactions (see 1.6.1), implicates involvement in effector binding in R proteins that interact directly with their effectors (Dodds et al., 2001b, Ellis et al., 1999, Holt et al., 2003). Within the flax-flax rust system, both genetic and, more recently, *in vitro* interaction studies have strongly implicated the LRR domain with a function of direct recognition of rust effector proteins. Genetic analysis of the *L* locus has shown that there are 12 alleles (encoding entirely TIR-NBS-LRR proteins), conferring at least 10 rust resistance specificities to different strains of *M. lini* (Ellis et al., 2007a, Ellis et al., 1999). Variants, *L6* and *L11*, differ only in sequences within the LRR, yet they recognise distinct strains of *M. lini* (Ellis et al., 1999). Further analysis demonstrated that the majority of those polymorphisms occurred within the repeat motif at the predicted solvent exposed x positions within the LRR consensus, with different polymorphisms being essential for either L6 or L11 specificity (Ellis et al., 2007b). Domain swap experiments between homologues at the *P* locus in flax, also demonstrate the importance of the LRR in specificity. Six amino acid differences between *P2* and *P* were found to be sufficient to distinguish between their resistance specificities. All of the polymorphic residues were localised to variable x residues of the repeat motif in the LRR domain (Dodds *et al.*, 2001b).

It is clear from these results that the LRR domains of flax R proteins play a pivotal role in effector recognition, but what information on this subject can be obtained from the analysis of rust effectors? The AvrL567 gene family encodes twelve highly diverse sequence variants, seven of which return a necrotic response, when infiltrated using the *A. tumefaciens* delivery system into flax leaves containing L5, L6 or L7. This degree of diversity, and difference in specificity, is the likely result of an evolutionary battle, where the rust and its host try to avoid and maintain recognition, respectively (Dodds et al., 2006). Y2H experiments with the different AvrL567 variants and L5, L6 and L7 support the gene-for-gene specificity shown in the *in planta* study, and demonstrate a direct interaction between the gene products (Dodds et al., 2006). Interestingly, a resistance inactive mutation in L6 involving an invariant lysine within the P-loop (also called walker A motif), discussed below as a critical motif found in all ATP binding proteins, prevents interaction with the AvrL567 effector in the Y2H analysis. This suggests that the L6 protein requires a functional nucleotide binding pocket to interact with the AvrL567 effector protein. The integrity of the P-loop has also been shown to be critical in maintaining intra-molecular

interactions in the potato virus X resistance protein, Rx (Moffett et al., 2002). It is therefore conceivable that a disruption in protein structure caused by a mutation in the P-loop motif of L6 may be responsible for the loss of interaction. Similar results have been demonstrated in Y2H experiments involving M and AvrM (P. Dodds, personal communication).

Structural determination of AvrL567A and D proteins revealed that almost all the side chains of the polymorphic residues in the AvrL567 variants were mapped to the outer solvent exposed region of the protein (Wang et al., 2007). Importantly, AvrL567A is recognised by both L5 and L6; however, AvrL567D was only detected by L6, and AvrL567C by neither (Dodds et al., 2006). From this, Wang et al., (2007) made targeted mutations to four polymorphic residues in AvrL567 (residues 50, 56, 90, 96) believed to be critical in this specificity. Using both Y2H, and an *in planta* HR assay, they demonstrated that differing combinations of the four mutations could alter the specificity of the effector. The data also suggested that multiple contacts at distant points in the AvrL567 protein are required for interaction, however single amino acid changes were sufficient to both stabilise and destabilise the interaction (Wang et al., 2007). Armed with this information, a model of the interaction between the L5 LRR domain, and the AvrL567A effector protein was generated, satisfying all the critical interactions that determined specificity (Wang et al., 2007). The structure of the L5 LRR is modelled on other known LRR structures. It is important therefore to stress that the model of L5/AvrL567 interaction is highly speculative.

Resistance specificity in flax is, however, not always confined to the LRR (Luck et al., 2000). Analysis of alleles at the *L* locus in flax demonstrated that alleles with the same LRR region but different TIR and NBS regions can encode different specificities (Ellis et al., 1999, Luck et al., 2000). Clearly interaction between flax rust effectors and flax R proteins is not exclusively confined to the LRR

Ueda et al., (2006) reported that the tobacco N protein directly recognises TMV p50 using Y2H and *in vitro* studies (Ueda et al., 2006). Two recent studies, however, cast some level of doubt over this conclusion. Firstly, co-localisation studies have shown that the TIR domain of N is critical for the interaction and it alone can associate with the p50 effector (Burch-Smith et al., 2007). However, Y2H analysis and *in vitro* pulldown could not demonstrate that the interaction was in-fact direct. A follow-up study has since demonstrated that the interaction between the p50 and N is mediated by a tobacco host protein called NRIP1 (Caplan et al., 2008).

It is clear therefore that care should be taken when interpreting the L5/AvrL567A interaction model. It is likely that many subtleties exist in the R/effector protein interaction in the flax-flax rust system, and for other R/effector systems for that matter. For any predicted interaction between R and effector proteins to be tested, protein purification, followed by structural and biochemical analysis, including *in vitro* R/effector binding studies, is required.

1.8 R protein Structure/Function

1.8.1 Structure studies of the NB-ARC region of Apaf-1 and CED-4

While the structure of the NB-ARC region of an R protein has not been solved, the crystal structure of the NB-ARC domains of Apaf-1 and CED-4 have (Riedl et al., 2005, Yan et al., 2005). The NB-ARC region of Apaf-1, in conjunction with the N-terminal caspase recruitment domain (CARD), was solved in its inactive conformation. The NB-ARC domain was most similar to structures solved for the AAA+ ATPases protein family (Riedl et al., 2005). The NB-ARC region could be separated into four structurally distinct domains. The NB domain produces a three layered α/β domain, also described as a β-sheet flanked by α-helices (Albrecht and Takken, 2006) and the ARC domain was further separated into 3 ARC subdomains; ARC1 generates a four-helix bundle, while ARC2 produces a winged helix fold and the ARC3 domain a second helical domain. ADP was found deeply buried within the structure, bound between the α/β domain, helical domain one and the winged helix domain, with the critical binding residues identified in Figure 1.3 (Riedl et al., 2005).

The structure of CED-4 was solved in a complex with CED-9, a constitutive inhibitor of CED-4. CED-9 binds to an asymmetric dimer of CED-4, however, it recognises only one of the two CED4 proteins (Yan et al., 2005). The CED-4 proteins are ATP bound, which, like ADP in Apaf-1, is deeply buried between the NB, ARC1 and ARC2 domains. It is of interest, however, that very different domain conformations are adopted by Apaf-1 and CED-4 (Takken *et al.*, 2006, Yan *et al.*, 2005). It is suggested that CED-4 is frozen in an active state with ATP bound, although, activation is masked by the binding of the CED-9 inhibitor (Danot et al., 2009).

Alignment studies of R proteins and the NB-ARC region of Apaf-1, predicts that R proteins contain three subdomains within the NB-ARC domain; NB, ARC1 and ARC2 (Figure 1.3) (Albrecht and Takken, 2006). For the tomato R protein, I-2, mapping studies of the NB-ARC region with Apaf-1 and CED-4, and more recent modelling studies with Apaf-1 as the template, have been used to highlight a number of residues likely to be important in R protein function (Takken et al., 2006, van Ooijen et al., 2008b). A large number of loss- and gain-of-function (autoactive) mutations map to areas within the NB-ARC region shown to be important in ATP/ADP interaction in the respective Apaf-1 and CED-4 structures (Figure 1.3). This further supports the role of ATP and ADP binding in R protein function (Takken *et al.*, 2006). Of particular interest is the cluster of autoactive mutations in conserved motifs within the ARC2 region (Figure 1.3). A recent structure/function study of one such motif, the MHD (so named after its consensus), implicates it as an important regulator of R proteins (van Ooijen et al., 2008b). The motif is located towards the C-terminal end of the NB-ARC domain and is the most well

characterised motif in R proteins regarding autoactivity. Manipulations to the conserved histidine or aspartate within the MHD motif have been reported to result in an autoactive phenotype in a number of R proteins (Bendahmane et al., 2002, Howles et al., 2005, van Ooijen et al., 2008b). The MHD is predicted to play a sensory role, coordinating nucleotide binding and controlling the interaction between the subdomains within the NB-ARC domain (van Ooijen et al., 2008b). In the Apaf-1 structure the highly conserved histidine, within this equivalent motif, is shown to form a hydrogen bond with the β-phosphate of the bound ADP (Riedl et al., 2005).

1.8.2 Intramolecular interactions regulate activation and signalling

The structures of Apaf-1 and CED-4 highlight that intramolecular interactions between protein domains are almost certainly involved in maintaining both inactive and active conformations. The potato *Rx* gene (CC-NB-ARC-LRR), which confers resistance to the potato virus X (PVX), has been at the forefront in understanding such interactions in R proteins. Transient expression assays involving the co-expression of separated domains of Rx, (CC-NB-ARC and LRR) and (NB-ARC-LRR and CC) both demonstrated an Avr dependent HR, signifying the reconstitution of resistance, presumably by the reconstitution of a functional R protein (Moffett et al., 2002). Using co-immunoprecipitation reactions, the domains separated by construct design could interact physically, however the CC-NB-ARC and LRR were shown to interact only in the absence of the Avr protein. This indicates that the effector protein is capable of disrupting the CC-NB-ARC and LRR interaction, which may suggest that the LRR is regulating activity (Moffett et al., 2002). This idea was supported in separate studies of *A. thaliana* proteins where in the cases of RPS2, RPS5 and RPP1A, the LRR is likely to be involved in negative regulation of R protein activation. This was the conclusion from the expression experiments of R genes without their LRR that resulted in an autoactive phenotype (Ade et al., 2007, Tao et al., 2000, Weaver et al., 2006). Ironically, the LRR has also been argued to positively regulate resistance for the tomato R protein Mi1.2 (Hwang and Williamson, 2003, Hwang et al., 2000). These differing arguments on domain function further emphasise the need for structural and *in vitro* functional analysis of purified R proteins.

The Moffett group have shown that physical association between the CC-NB-ARC and LRR requires the ARC1 subdomain, but is not dependent on nucleotide binding (Rairdan and Moffett, 2006). The ARC2 subdomain is believed to play an auto-inhibitory role in Rx and is also required for activation (Rairdan and Moffett, 2006). The importance of the ARC2 domain in STAND protein activation has also been reported. In a recent structural review of the crystal structures available for STAND proteins, the conformation changes that underlie protein activation was a major focus (Danot et al., 2009). From this study, It is suggested that the ARC2

domain (WHD) undergoes a 180°C rotation between the closed and open conformations (Danot et al., 2009). This evidence suggests that ARC2 relays the elicitor perception signal to the rest of the NB-ARC domain for protein re-organisation and activation. This idea is somewhat consistent with the localisation of a large number of loss- and gain-of-function mutations within the ARC2 domain in R proteins, discussed above. However, such a hypothesis requires further testing *in vitro* for any STAND-like protein, including plant R proteins.

1.8.3 A model of R protein activation

As discussed above, the NB-ARC domain of R proteins is critical for function and is likely to provide the necessary components required for protein activation. As stated, truncated recombinant versions of the R proteins I-2, Mi1 and N have demonstrated ATP binding and hydrolysis capabilities (Tameling et al., 2002, Ueda et al., 2006). Further analysis of I-2 demonstrated the potential mechanism of activity through the biochemical investigation of two mutations shown to display autoactive phenotypes *in planta* (Tameling et al., 2006). To elaborate, utilising their *E. coli* recombinant expression system, CC-NB-ARC domains of I-2 carrying an aspartate to glutamate change within the kinase 2 motif, and a serine to phenylalanine change within the RNBS-A motif, were expressed and purified (Figure 1.3). The two mutant proteins were found to have similar ATP binding kinetics, however, their hydrolysis activity was compromised when compared to the non-mutated CC-NB-ARC protein (Tameling et al., 2006). This suggested that hydrolysis may not be a requirement for R protein activation and changes in the identity of the bound nucleotide may in-fact control the activation of the protein. The same kinase 2 mutation in the *A. thaliana* protein RPS5 also caused an autoactive phenotype (Ade *et al.*, 2007).

Tameling and co-workers proposed that the NB-ARC domain of R proteins function as a molecular switch to control activation (Takken et al., 2006, Tameling et al., 2006). The functional switch is anticipated to be in an “off” or inactive state with ADP bound, and an “on” or active state when ATP is bound. In such a model, effector perception would theoretically induce nucleotide exchange from ADP to ATP, thus forming a molecular switch. Hydrolysis of the ATP molecule to ADP may enable the R protein to be reset, facilitating possible signal amplification (Figure 1.4). This model, however, differs from that proposed from the only other biochemical study of an NBS-LRR R protein (Ueda et al., 2006). Here the resting state of N is proposed to be ATP bound, and interaction with the TMV elicitor, p50, promotes ATP hydrolysis. Ueda et al., (2006) suggest that hydrolysis is what triggers the defence response (Ueda et al., 2006). So which model is correct? Given that these researchers presented work on different R proteins from different structural classes (CC/TIR), it is possible both models are correct and there exist a fundamental difference

in the activation of R proteins from these classes. However, a number of biochemical studies of STAND proteins have supported the molecular switch model involving ADP/ATP exchange (Figure 1.4), with suggestions that this mode of activation may be generally conserved among this broad class of proteins (reviewed by (Danot et al., 2009)).

1.8.4 Functional studies of STAND proteins support the R protein molecular switch model

As previously mentioned the NOD module (NB-ARC domain in R proteins) is fundamental for the inclusion of a protein into the STAND class and assists its classification within the STAND sub groups. In a recent review it has been predicted that the NOD module forms an architecture that is generally conserved throughout the STAND members, and also suggests that the subsequent mechanics behind the function of the STAND proteins could also be generally conserved (Danot et al., 2009). At this stage, the structures of only four members of the STAND class are available and clearly more structures are required before major generalisations can be made. However, we can tentatively interpret the mechanics of protein activation from biochemical studies of STAND proteins and apply this knowledge to the context of R protein activation. To this end, arguably the most advanced biochemical studies of STAND proteins has come from the investigations of Apaf-1 and MalT, the latter an *E. coli* transcription activator. Biochemical characterisation of both of these proteins is further advanced than that of R proteins and therefore the experimental evidence to support models of activation that lead to signalling are more compelling.

MalT is an *E. coli* transcription activator of the maltose regulon. Like R proteins, the model of MalT activation has been shown to involve an ADP bound autoinhibited state, and an ATP bound active state (Marquenet and Richet, 2007). The evidence for this comes from the purification of the wild type MalT protein in an ADP bound state, while a mutation in the second aspartate of the kinase-2 motif, which is the same respective residue mutated and analysed in I-2 (Tameling *et al.*, 2006), was ATP bound (Marquenet and Richet, 2007). A functional assays demonstrated that this ATP bound form of MalT was hyper-activated (Marquenet and Richet, 2007). MalT is normally activated by the chemical, maltotriose, which was demonstrated experimentally to promote ADP to ATP exchange and the formation of an oligomer. The addition of maltotriose also increased the rate of ATP hydrolysis in wild type MalT, however, the hyper-activated mutant was unable to hydrolyse ATP. Whilst ATP hydrolysis is therefore not critical for transcription activation by MalT, it is critical for the control of activity. It is suggested that ATP hydrolysis returns a STAND protein from an active to an inactive state, and that it may therefore be involved in protein recycling (Marquenet and Richet, 2007).

Apaf-1 is an extremely important mammalian cell death protein which complexes with cytochrome c in the presence of deoxyadenosine triphosphate (dATP) or ATP to form an oligomeric apoptosome. The apoptosome is capable of recruiting and activating procasapse-9, which, inturn activates a caspase-related cell death pathway (Bao and Shi, 2007, Riedl et al., 2005). Some conjecture surrounds the nucleotide bound state of the autoinhibited form of Apaf-1, with structural and biochemical studies suggesting it is dADP/ADP bound (Bao et al., 2007, Riedl et al., 2005), while others present evidence suggesting it is bound to dATP (Kim et al., 2005). Nonetheless, the activated apoptosome constitutes a complex that contains seven Apaf-1 proteins in the dATP bound state bound with cytochrome c (Bao and Shi, 2007, Kim et al., 2005, Zou et al., 1999).

From the experimental evidence presented for both MalT and Apaf-1 activation, it is tempting to speculate upon the similarities in the model proposed for the activation of an R protein (Tameling et al., 2006). At this point some key features of the model in R proteins have been demonstrated experimentally, however, evidence to further support or reject such a model are undoubtedly necessary. This will come from further attempts to express and purify R proteins to enable their biochemical study.

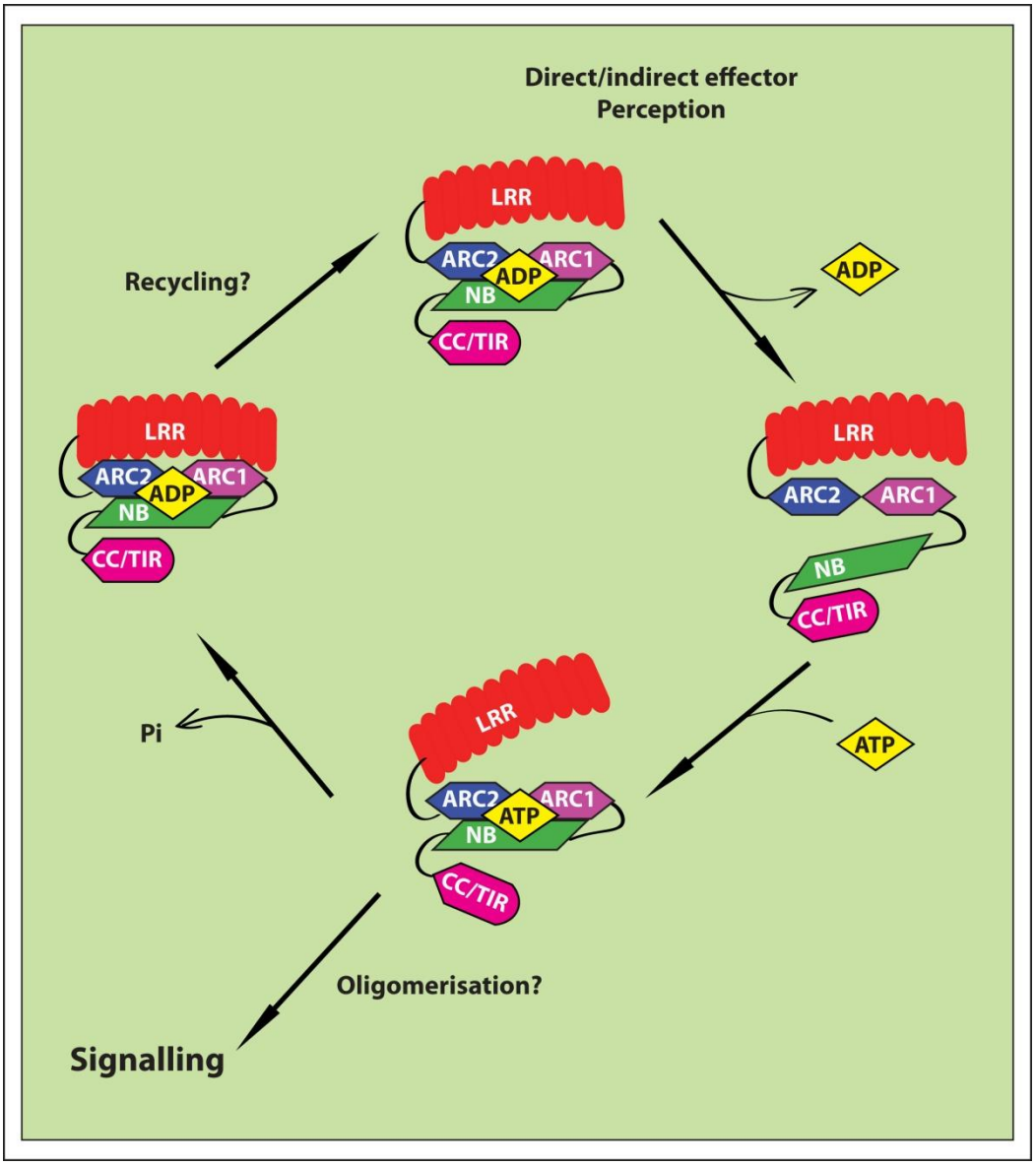

Figure 1.4: The molecular switch model for R protein activation

This figure is adapted from (Takken et al., 2006, Takken and Tameling, 2009, Tameling et al., 2006). In the absence of a pathogen effector, the R protein exists in an inactive, tightly regulated, ADP bound conformation (1). Direct or indirect interaction with an effector protein (2) causes the protein to change conformation, stimulating a more open structure (3). The open conformation is then able to exchange ADP for ATP, resulting in the formation of an active ATP bound conformation that may undergo oligomerisation (4). It is the ATP bound conformation that is active and signals the defence response. Hydrolysis of the bound ATP enables the protein to return to an autoinhibited ADP bound state (1). This could potentially enable protein recycling as a possible method for signal amplification.

1.9 Project Aims and Objectives

Our understanding of the functions of R protein domains and how R proteins are activated has progressed significantly over recent years, there is, however, still much to be determined. The ability to perform biochemical analysis has been confounded by a well recognised difficulty in expressing and purifying functional full-length or near full-length recombinant R protein. This study aimed to help rectify this deficiency and describes a method for the production of near full-length flax R proteins, M and L6. This method has facilitated a biochemical investigation of the flax R protein, M; to further elucidate aspects of its function and interaction with the corresponding flax rust effector protein, AvrM.

1.9.1 Aims

1. Develop and refine recombinant protein expression in, and purification from, *Pichia pastoris,* to enable the production of near full-length flax M and L6 proteins, to facilitate biochemical investigations.
2. Using purified recombinant M protein, investigate the mechanisms of activation and regulation within the NB-ARC domain.
3. Determine the nature of the interaction between the flax M protein and the flax rust effector protein AvrM *in vitro*.
4. Reconstitute *in-vitro,* the activation of the M protein in response to AvrM, and investigate any affect this may have on the nucleotide binding dynamics of the M protein.